

1
2
3
4
5
6
7
8
9
10
11
12

Shifted distribution baselines: neglecting long-term biodiversity records risks overlooking potentially suitable habitat for conservation management

Sophie Monsarrat*¹, Peter Novellie¹, Ian Rushworth², Graham Kerley¹

¹*Centre for African Conservation Ecology, Nelson Mandela University, Port Elizabeth, South Africa;*

²*Ezemvelo KZN Wildlife, Pietermaritzburg, South Africa*

* Corresponding author: sophiemonsarrat@gmail.com

13 **ABSTRACT**

14 Setting appropriate conservation measures to halt the loss of biodiversity requires a good
15 understanding of species' habitat requirements and potential distribution. Recent (past few
16 decades) ecological data are typically used to estimate and understand species' ecological niche.
17 However, historical local extinctions may have truncated species-environment relationships,
18 resulting in a biased perception of species' habitat preferences. This may result in incorrect
19 assessments of the area potentially available for their conservation. Incorporating long-term
20 (centuries-old) occurrence records with recent records may provide better information on
21 species-environment relationships and improve the modeling and understanding of habitat
22 suitability. We test whether neglecting long-term occurrence records leads to an
23 underestimation of species' historical niche and potential distribution and identify which species
24 are more vulnerable to this effect. We compare outputs of species distribution models and niche
25 hypervolumes built using recent records only with those built using both recent and long-term
26 (post-1500) records, for a set of 34 large mammal species in South Africa. We find that, while
27 using recent records only is adequate for some species, adding historical records in the analyses
28 impacts estimates of the niche and habitat suitability for fourteen species (41%) in our dataset,
29 and that this effect is significantly higher for carnivores. These results show that neglecting long-
30 term biodiversity records in spatial analyses risks misunderstanding, and generally
31 underestimating, species' niche, which in turn may lead to ill-informed management decisions,
32 with significant implications for the effectiveness of conservation efforts.

33 **Keywords**

34 habitat suitability models, historical ecology, megafauna, natural range, niche, reintroduction, rewilding

35 INTRODUCTION

36 To avert the ongoing human-induced biodiversity decline, scientists have recently called for conservation
37 efforts to be intensified, including through increased habitat protection and restoration [1]. Data on
38 species' distribution patterns and species assemblages are key to identify candidate areas for
39 conservation [2]. However, distribution patterns have been drastically modified by humans, notably
40 through global extinctions and regional to local extirpations [3,4], and thus contemporary patterns do not
41 necessarily reflect species' natural distribution and habitat preferences. Analyses of species distributions
42 that tend to ignore these modifications will likely result in a biased understanding of species'
43 biogeography and ecological requirements and lead to misleading perceptions of the options available for
44 conservation [5,6]. This phenomenon of spatially shifted baseline poses clear challenges for conservation
45 and management. By providing information on species' historic rather than current-day relictual
46 distribution, long-term biodiversity data have the potential to improve our understanding of the
47 biogeography of species and participate in setting appropriate spatial and ecological baselines for
48 environmental conservation and restoration [7].

49 Mammals are one of the most studied taxa, and their current distribution patterns are well known [8].
50 Historic and prehistoric human-driven global and local extinctions have however caused a strong deviation
51 between current and pre-anthropogenic impact diversity patterns, in particular for large terrestrial
52 mammals [4]. In South Africa, habitat loss, competition with livestock and direct exploitation, in particular
53 following European colonization, have resulted in the global extinction of one mammal species [9], and
54 the collapse of large mammal diversity in large parts of the country [10,11]. To halt this decline and restore
55 populations, conservation efforts have focused on establishing protected areas and actively managing
56 large mammal populations through reinforcement - to increase population viability - and reintroductions
57 - to re-establish populations within species' historical range [12]. Defining species' historical distributions
58 and suitable habitat is thus a critical aspect for conservation planning in South Africa [13], as it is for most
59 restoration attempts elsewhere [14]. Among large mammal species, those identified as threatened by the
60 IUCN Red List are a high priority for conservation [15] and large carnivores have an important ecological
61 role and have undergone considerable historic range contractions [16], making them a major focus of
62 conservation and rewilding efforts [17]. It is thus critical to understand the extent to which historical data
63 are needed to inform wildlife conservation and management, for threatened species and large carnivores
64 in particular.

65 Habitat suitability models (HSMs) [18,19] and n -dimensional hypervolumes [20] are two widely-used tools
66 that relate species' occurrences to environmental variables in order to, respectively, map species'
67 potential distribution in geographical space and characterize species' niche in the environmental space.
68 They have notably been used in a conservation and management contexts to improve our knowledge of
69 species ranges, support management plans for species' recovery, prioritize areas for biodiversity
70 protection and predict changes in suitable habitat in response to human impacts [21–23]. HSMs and
71 hypervolume approaches rely on the assumption that the observed geographical distribution of a species
72 reflects its ecological requirements, making them highly contingent on the quality of the occurrence
73 records[19,23]. Range contraction that have affected the array of conditions that the species occupy risk
74 truncating species–habitat relationships [24], thus hindering our ability to estimate species niches and
75 predict the distribution of suitable habitat [6]. Failing to consider past local extinction events may thus
76 misguide conservation efforts by overlooking potentially suitable sites for reintroduction or restrict
77 protection to suboptimal habitats [5]. Despite providing useful information on the historic distribution of
78 species, historical written records and museum specimens have long been overlooked in habitat suitability
79 modeling approaches, being perceived as untrustworthy for their intrinsic biases and limitations [25] (but
80 see [26–28]). The development of methods to address sampling biases in HSMs [29–32] however, provides
81 an avenue for more confident incorporation of these records in spatial modeling analyses, and hence in
82 conservation interventions.

83 Here we investigate how long-term biodiversity records can contribute to setting appropriate baselines
84 for species' distribution. We test the hypothesis that neglecting historical data in niche quantification and
85 HSM approaches leads to biased perceptions of species' historic niche and suitable habitat distribution,
86 and the patterns of potential species richness at the regional level. We focus on large terrestrial mammals
87 in South Africa, for which we have access to a unique dataset of long-term occurrence records spanning
88 the last five centuries, as well as recent (post-1950) occurrence records, for a community of 34 mammal
89 species.

90 **METHODS**

91 **Overview of the approach**

92 We considered two datasets of occurrence - recent records (post-1950, RECENT) and recent + long-term
93 records (post-1500, TOTAL) - to quantify the effect of neglecting long-term occurrence data. We compared
94 results obtained from these two datasets in three different approaches, two that are species based: 1)

95 estimation of the climatic niche in environmental space using n-dimensional hypervolumes [23] and 2)
96 prediction of suitable habitat in the geographic space using habitat suitability models (HSMs) and one at
97 the community level, namely prediction of the distribution of potential species richness using stacked-
98 HSMs. For the species-level approaches, we used two indices that summarize the cost of neglecting long-
99 term biodiversity data and tested how the combination of these indices relate to species' conservation
100 status and diet. For the community-level approach, we investigated spatial differences in predicted
101 potential species richness, notably by comparing predictions between different South African bioregions.

102 **Species data**

103 The general study area, hereafter referred to as South Africa, covers the countries of South Africa, Lesotho
104 and eSwatini (former Swaziland). We considered all extant South African large (> 20 kg) terrestrial
105 mammals, except for species with fewer than 25 long-term observations in the dataset. In total, 34 species
106 were included: 23 from the order Artiodactyla, 6 Carnivora, 4 Perissodactyla and 1 Proboscidea.

107 *Theoretically accessible areas*

108 Barve et al. [33] outline the concept of the theoretically accessible area (the area that is climatically
109 suitable and has been accessible to the species via dispersal over relevant periods of time), and show that
110 restricting a model's training and validation areas to this theoretically accessible area greatly improves
111 HSM performance and provides more accurate predictions of species richness and community
112 composition [33,34]. As an approach to estimating the theoretically accessible area for each species, we
113 identified the bioregions in which the species are known to have occurred historically, based on
114 information on their ecology and interpretation of historical occurrences, and built a polygon using the
115 boundaries of these bioregions. We defined the accessible area for each species as a buffer of 20 km
116 around this polygon, to include ecotone regions where the species could disperse. This option is suggested
117 by Barve et al. [33] to be the most operational compared to more intricate alternatives. We acquired
118 spatial information on bioregions from the 2012 Vegetation Map of South Africa, Lesotho and Swaziland
119 [35].

120 *Modern and historical occurrence records*

121 The long-term occurrence dataset used in this study covers the period 1500 to 1950 and includes records
122 extracted from the historical literature, museum specimens and fossil records. For historical records and
123 museum specimens, we used the database presented in Boshoff et al [10], completed with records from
124 the KwaZulu-Natal, eSwatini and the rest of South Africa, using the same approach and criteria defined in

125 Boshoff et al. [10], so that the dataset covers all of South Africa. The reliability of these records in terms
126 of identification and locality is discussed in Boshoff and Kerley [36] and their spatial, environmental and
127 taxonomic biases in Monsarrat et al. [29] and Monsarrat and Kerley [37].

128 Recent fossil records were obtained from Avery [38], a comprehensive compilation of information on the
129 taxonomy and distribution in time and space of all currently recognized South African fossil mammal. We
130 recovered radiocarbon dating for these record from primary sources and kept only those that were
131 deposited in the period from 1500 to today.

132 Modern (post-1950) occurrence records were provided by the South African National Biodiversity
133 Institute and the Endangered Wildlife Trust, who consolidated and centralized a total of over 460,000 geo-
134 referenced unique occurrence records for South African mammals, from 59 different contributors
135 (governmental institutions, research institutions, non-governmental organizations, the private sector and
136 citizen science projects) [39]. This database formed the basis of the 2016 “Red List of Mammals of South
137 Africa, Lesotho and Swaziland” and as part of this process, data were vetted and underwent several
138 rounds of data cleaning to check accuracy. These data are spatially biased, with the highest densities of
139 records typically found in protected areas [39], artificially increasing spatial auto-correlation of
140 occurrences. This in turn may affect the performance of habitat suitability models built with these data
141 [40]. To reduce the effect of sampling bias and spatial clustering on model performance, we subsampled
142 the modern occurrence dataset using spatial thinning of the data (no occurrence records closer than 0.1
143 degree), as recommended by Boria et al. [41].

144 We considered all occurrence records located outside of a species’ theoretically accessible area to be
145 extralimital and we excluded them from the analyses. Modern extralimital records often correspond to
146 introductions of individuals or populations outside of their historic range, often in suboptimal habitat, and
147 are not informative of the habitat preferences of the species [42]. We however acknowledge that, by
148 using bioregions as the filter for modern records, we may include some records that are outside the
149 historic range, this being due to the relatively unique situation in SA of game translocations for commercial
150 purposes [42].

151 In total, we analyzed 15,315 recent (post-spatial thinning, range 55-1,274) and 5,446 historical (range
152 25-501) records for the 34 species of large terrestrial mammals, covering a total area of ca. 1,270,000
153 km² for these three nations.

154

155 **Environmental data**

156 We considered six bioclimatic variables derived from BioClim [43]: mean annual temperature (BIO1) and
157 annual precipitation (BIO12), describing the average climatic conditions, temperature seasonality (BIO4)
158 and precipitation seasonality (BIO15), describing climatic seasonality and maximum temperature of the
159 warmer month (BIO5) and precipitation of the warmest quarter (BIO18), describing extreme climatic
160 conditions. We also considered topography (TOPO), using altitude data from the ASTER Global Digital
161 Elevation Model (ASTGTM) on <https://lpdaac.usgs.gov> [44]. These variables were chosen, because they
162 were biologically meaningful to predict large mammal species richness in South Africa [45] and because
163 they potentially represent environmental characteristics that limit species' distributions. All
164 environmental variables were estimated at a 0.1 x 0.1 degree resolution, using the *raster* package [46] in
165 R 3.5.1 [47].

166 **Hypervolume analysis**

167 The n-dimensional hypervolume was originally proposed by Hutchinson [48] to describe the fundamental
168 niche of a species, i.e. the environmental space where the species can exist indefinitely. In the modern
169 understanding of the hypervolume function, a set of n variables that represent biologically important and
170 independent axes are identified and the hypervolume is defined by a set of points within this n-
171 dimensional space that reflects suitable values of the variables for the species' persistence [23]. Here, we
172 consider five environmental axes: BIO1, BIO4, BIO12, BIO15 and TOPO, rescaled to a common and
173 comparable scale before the analyses. We used the Gaussian kernel density estimation with the Silverman
174 bandwidth estimator method in the *hypervolume* package [20] in R 3.5.1 [47]. The bandwidth was
175 estimated from the RECENT dataset with the Silverman estimator and the same value was used for the
176 TOTAL dataset, to allow direct comparison.

177 The volume of the hypervolume is approximately linearly proportional to the number of observations in
178 the dataset [20]. To ensure results are insensitive to sample size, we randomly subsampled the TOTAL
179 dataset to have the same number of records as the RECENT dataset. We repeated the process ten times
180 and used averaged hypervolume measures of these ten repetitions in the statistical analyses.

181 **Habitat Suitability Modeling**

182 *Background data*

183 Because the species occurrence records are highly biased spatially [29,39], we addressed the potential
184 effect of sampling bias in the models. To do so, we produced background data with similar geographical
185 bias as the RECENT and TOTAL occurrence datasets, following [30]. We first created a sampling effort
186 raster using a two-dimensional kernel density estimation applied on the occurrence dataset. Background
187 data were then created by sampling without replacement within this raster grid, where the probability of
188 a cell being sampled was proportional to the sampling density values (weighted target group approach,
189 following Sanín and Anderson [49]). We selected the same number of background points as the number
190 of occurrence records, so as to achieve a prevalence of 50%, as advised by Liu et al. [50].

191 *Ensemble modeling*

192 We created ensemble HSM [51] for each species by assembling five statistical methods (GAM, MAXENT,
193 MARS, RF and GBM) to account for inter-model variability, using the *ssdm* package [52]. We ran ten
194 repetitions for each of the algorithms and produced an average of the models' outputs, weighting each
195 model according to its predictive ability. We measured predictive ability with a cross-validation approach,
196 by using a random 70% of the data for calibration of the models (keeping the prevalence constant) and
197 testing their predictive ability on the remainder of the dataset using the True Skill Statistic (TSS) [53]. We
198 repeated this approach ten times for each model and used an average of the predictive accuracy measure.
199 In total, for each species and each dataset, we ran 500 models using five different statistical models, ten
200 repetitions of each algorithm, and ten repetitions of the random-splitting strategy. The outputs of these
201 models are maps of predicted habitat suitability over the study area that provide hypotheses for the
202 potential distribution of species for both datasets. We identified areas where the predicted habitat
203 suitability differs between the RECENT and TOTAL datasets by subtracting the predicted values obtained
204 from the RECENT model to those obtained with the TOTAL model in each cell within the study area. Areas
205 with positive (negative) values are where we underestimate (overestimate) habitat suitability when
206 considering only recent records.

207 **Species richness**

208 Stacked-species distribution models (SSDM) combine multiple individual HSMs to produce a community-
209 level model and predictive maps of potential species richness [54]. We used the *ssdm* package [52] to
210 compute maps of local species richness by summing the probabilities from continuous habitat suitability
211 maps provided by the ensemble HSMs, a method that performs better than stacking methods based on
212 thresholding site-level occurrence probabilities [55]. To highlight areas where the potential species

213 diversity is under- or over-estimated because of neglecting long-term occurrence records in the models,
214 we subtracted the map of species richness produced with the RECENT dataset to the one produced with
215 the TOTAL dataset. We also compared the mean difference in predicted species richness for each
216 bioregion of South Africa.

217 **Statistical analyses**

218 For each species, we considered two indices to summarize the effects of neglecting long-term records on
219 the estimation of climatic niche and habitat suitability: 1) the niche dissimilarity in environmental space
220 (N_{dis}) [20] and 2) the dissimilarity in predicted habitat suitability in the geographical space ($PRED_{dis}$) [56];
221 (see Table 1 for a definition of these indices). For each index, higher values indicate higher disparity
222 between the results obtained with the RECENT and the TOTAL dataset.

223 We rescaled all indices by subtracting the mean and dividing by the standard deviation so that they are
224 comparable and conducted a Principal Component Analysis (PCA) to convert these indices into a one-
225 dimension variable (the first principal component PC1), quantifying the effect of neglecting historical
226 records. We ran a two-way ANOVA with Type II errors to test for differences in PC1 between conservation
227 status (threatened vs non-threatened) and broad diet guilds (herbivores vs carnivores). The conservation
228 status was defined from the IUCN Red List categories [57], where species listed as vulnerable, endangered
229 or critically endangered were considered “threatened”, and “non-threatened” otherwise. We used a
230 linear model to test how the change in mean predicted habitat suitability ($\Delta PRED$, calculated as the
231 proportion difference in mean predicted habitat suitability over the study area when it is estimated from
232 the TOTAL dataset, compared to the RECENT dataset) varies with PC1 values. We also estimated the
233 difference in the ability of HSMs to predict all the known occurrences for the species (ΔB), by measuring
234 the proportional increase (or decrease) in the continuous Boyce index, a threshold-independent evaluator
235 of the ability of HSMs to predict species presences [54], when it is estimated from the TOTAL dataset,
236 compared to the RECENT dataset.

237

238

239

240

241

242

243 **Table 1.** Description of the two indices used in the PCA analysis to quantify the effect of neglecting
 244 historical records on the estimation of climatic niche and suitable habitat.

Index	Name	Estimated from	Formula	Description
N_{dis}	Dissimilarity of climatic niche	5-dimensional hypervolume	$1 - Jaccard\ similarity\ index$	The Jaccard similarity index measures the overlap between niche hypervolumes [20]. N_{dis} is comprised between 0 and 1, with higher values meaning higher dissimilarity.
$PRED_{dis}$	Dissimilarity of predicted suitable environments	HSM	$1 - ESP$	The Expected fraction of Shared Presences (ESP) is a derivation from the Sørensen index of similarity of species' distributions that measures the overlap in predicted habitat suitability [56,58]. $PRED_{dis}$ is comprised between 0 and 1, with higher values meaning higher dissimilarity.

245

246 RESULTS

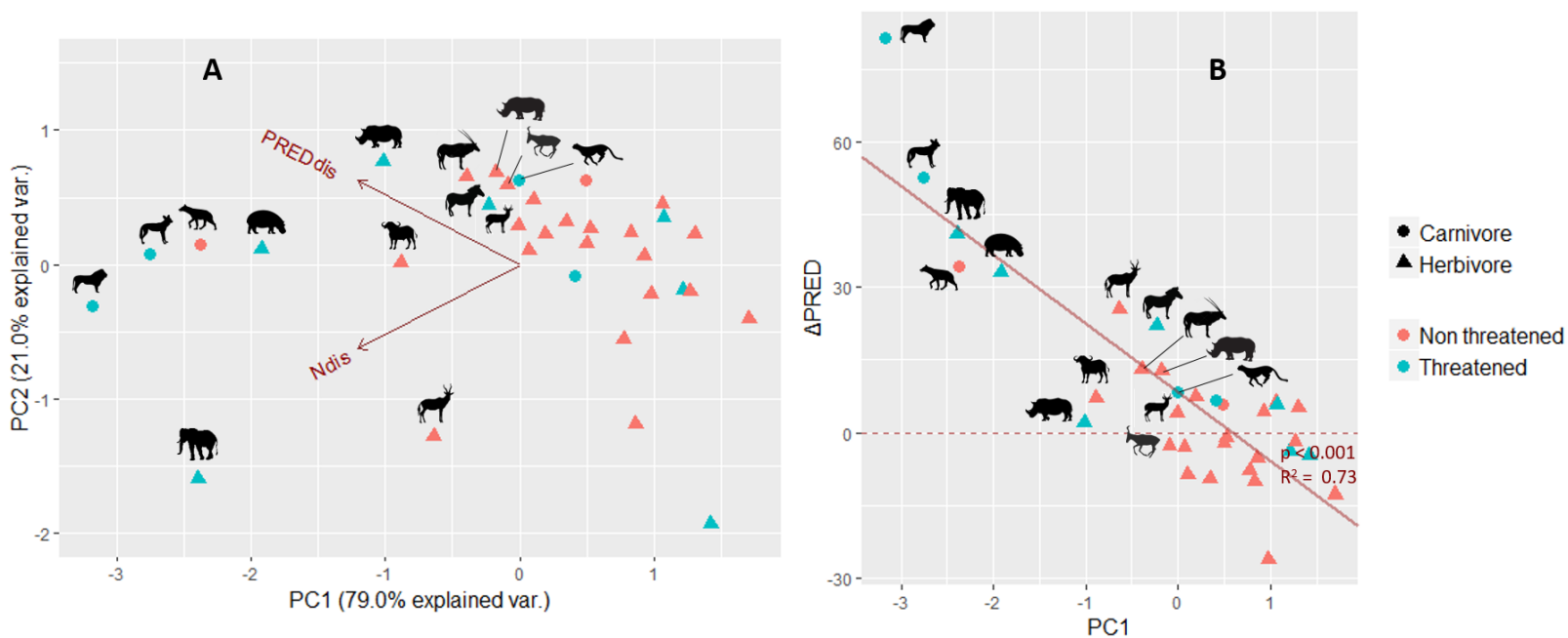
247 The ensemble modeling approach yielded very good agreement between the different modelling
 248 methods, as indicated by low standard deviation around the predicted habitat suitability values
 249 (Supplementary Information 2). For 18 out of 34 species, the inclusion of historical records improved the
 250 ability of the model to predict all known occurrences of the species ($\Delta B > 0$). The highest improvement in
 251 predictive ability was for the Black rhinoceros, Bontebok and African Elephant (ΔB equal to 25%, 16% and
 252 11%, respectively). In contrast, eleven species showed a decrease in predictive ability when historical data
 253 are included in the model, with the gemsbok and blue wildebeest showing the strongest decrease (ΔB
 254 equal to -10% and -8%, respectively). The dissimilarity in climatic niche N_{dis} calculated from the 5-
 255 dimensional hypervolume ranges from 0.09 to 0.52 (mean=0.21 \pm 0.11 SD), with $N_{dis} > 0.40$ for the African
 256 elephant, lion and African wild dog. The dissimilarity in predicted habitat suitability $PRED_{dis}$ ranges from
 257 0.41 to 0.71 (mean=0.57 \pm 0.07 SD), with $PRED_{dis} > 0.70$ for the African wild dog, lion and spotted hyaena
 258 (Table S1 of Supplementary Information).

259 Overall, by combining N_{dis} and $PRED_{dis}$ in a PCA, fourteen species (41% of our dataset) come out as
 260 impacted by neglecting historical records (PC1 < 0, with PC1 explaining 79% of the variance), with five
 261 species identified as the top most impacted: the lion, African wild dog, African elephant, spotted hyaena
 262 and hippopotamus (Fig 1A). Of these five species, four are listed as threatened on the IUCN Red List. Three
 263 are carnivores and the other two are megaherbivores (body mass >1,000 kg). PC1 values were significantly
 264 lower for carnivores compared to herbivores (two-way ANOVA Type II, $F(1,34) = 5.35$, $MSE = 6.60$,
 265 $p = 0.026$), and marginally lower for threatened compared to non-threatened species (two-way ANOVA
 266 Type II, $F(1,34) = 2.26$, $MSE = 2.78$, $p = 0.143$).

267 We found a significant inverse linear relationship between PC1 and the change in mean predicted habitat
268 suitability over the study area Δ PRED ($p < 0.001$, $R^2 = 0.73$) (Fig 1B), i.e. species that are most affected by
269 neglecting historical data have higher mean predicted habitat suitability over their study area when
270 historical records are included in HSMs. The lion, African wild dog, elephant, spotted hyaena and
271 hippopotamus show the largest increase in mean predicted habitat suitability when historical records are
272 included (Δ Pred equal to 81%, 53%, 41%, 34% and 33%, respectively).

273 This results in differences in predicted potential species richness at the community level (Fig 2A) and in
274 the geographic distribution of predicted habitat suitability at the species level (see maps on Fig 2B for the
275 five most impacted species and Supplementary Information for maps of all 34 species). Differences in
276 predicted potential species richness are higher for the Albany Thicket, Fynbos and Savanna Lowveld
277 biomes (Fig 3). The Nama-Karoo has on average very similar predicted species richness with the RECENT
278 or TOTAL dataset, whereas the potential species richness tends to be overestimated in Arid Savanna.

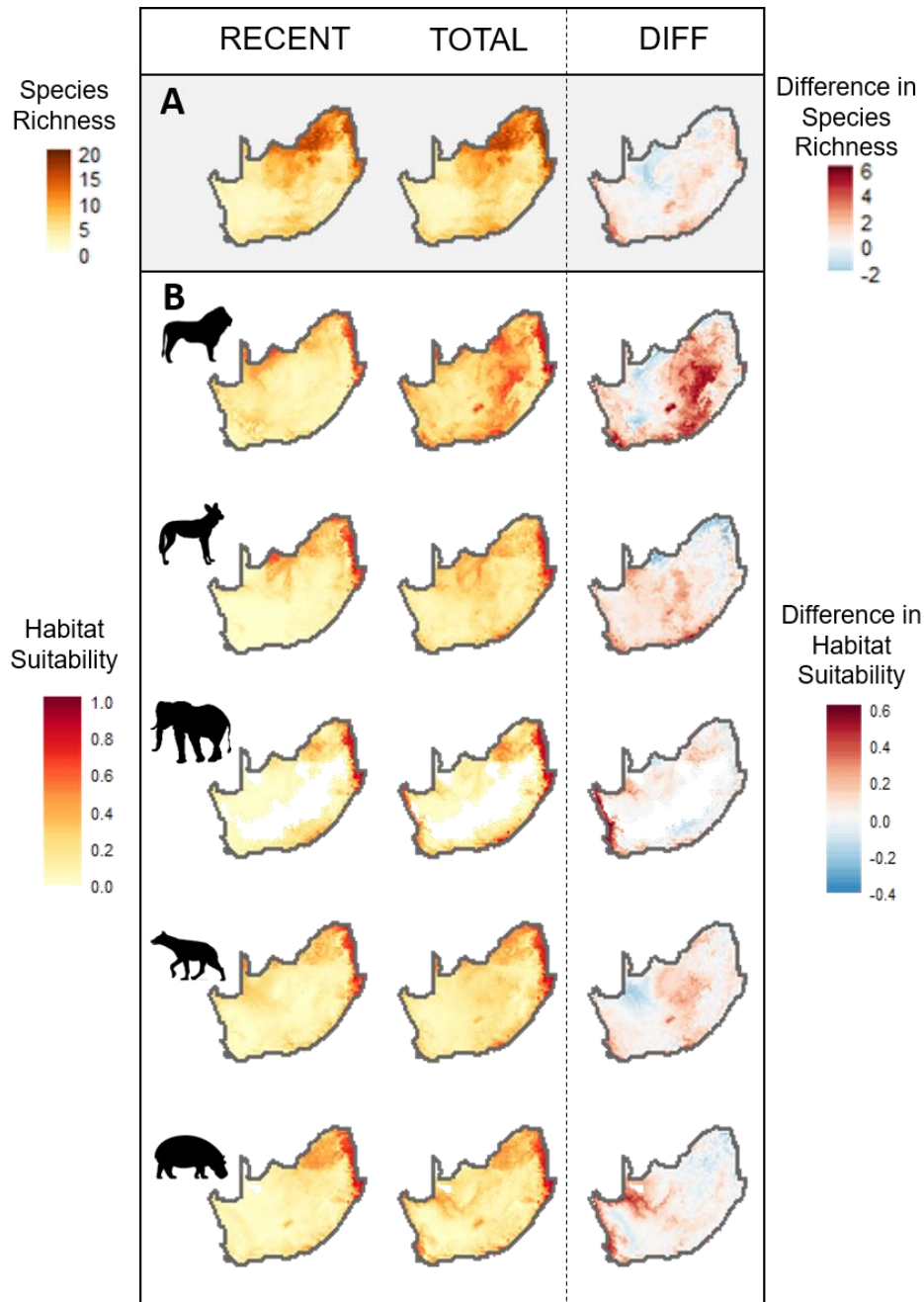
279



280 **Figure 1. Effects of incorporating historical records on the estimation of species climatic niche and**
281 **predictions of habitat suitability.** A) Principal Component Analysis (PCA) of the two indices used to
282 measure discrepancy between estimations of climatic niche (N_{dis}) and habitat suitability ($PRED_{dis}$) with the
283 RECENT and TOTAL datasets, for the 34 species of large mammals considered. Higher values of indices
284 (lower values of PC1) indicate a higher discrepancy. We highlighted (silhouettes) 14 species with negative
285 PC1 values, most affected by neglecting historical records. We differentiate carnivores vs herbivores and

286 threatened vs non-threatened species. The differences between these groups along the first principal
287 component (PC1) are significant for the former ($p=0.026$) and marginal for the latter ($p=0.143$). B) Plot
288 showing the negative relationship between values of PC1 and the proportion difference in mean predicted
289 habitat suitability over the study area when it is estimated from the TOTAL dataset, compared to the
290 RECENT dataset (Δ PRED). Positive values of Δ PRED mean that the habitat suitability is underestimated
291 without using historical records. Species that are most affected by neglecting historical data have
292 increased mean predicted habitat suitability when historical records are included in HSMs. See Table S1
293 in Supplementary Information for a key of silhouettes.

294



295

296 **Figure 2. Effect of incorporating historical records in the spatial prediction of A) species richness and B)**

297 **habitat suitability for the five species most impacted by neglecting historical records. The first column**

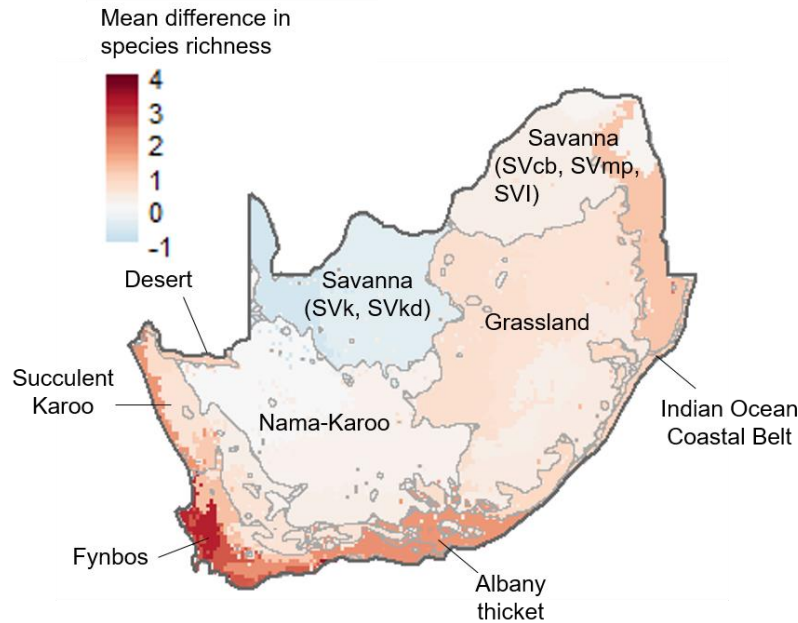
298 **is the prediction of species richness/habitat suitability obtained from the RECENT dataset (post-1950**

299 **records) and the second column is obtained from the TOTAL dataset (historical + recent records). The last**

300 **column is the difference between column 2 and column 1. We highlight the five species most impacted**

301 **through neglecting historical records according to their PC1 score: lion, African wild dog, African elephant,**

302 **spotted hyaena and hippopotamus.**



303

304 **Figure 3. Mean difference in species richness estimated from the TOTAL vs RECENT dataset, calculated**
305 **for each of the bioregions of South Africa.** Darker shades of red indicate that the potential species
306 richness is higher when predicted with the TOTAL dataset than with the RECENT dataset. The main biomes
307 are identified with grey contouring, obtained from simplifying polygons of the 2012 “Vegetation Map of
308 South Africa, Lesotho and Swaziland” [59]. The distinction is made between arid Savannas (SVk: Eastern
309 Kalahari Bushveld Bioregion and SVkd: Kalahari Duneveld Bioregion) and mesic Savannas (SVcb: Central
310 Bushveld Bioregion; SVmp: Mopane bioregion; SVI: Lowveld Bioregion and SVs: Sub-Escarpment Savanna
311 Bioregion) [59].

312

313 DISCUSSION

314 We show that neglecting long-term records can bias estimates of species climatic niche and suitable
315 habitat and underestimate the potential regional species richness. The implications are more severe for
316 carnivore species, and marginally more so for threatened species, for which appropriate conservation
317 actions and management decisions are the most critical. These results have implications for conservation
318 planning and distribution modelling in general, given that globally most mapping of species’ distributions
319 and habitat use exclude long-term occurrence records, and for the conservation and management of
320 South African mammalian fauna. These findings highlight the importance of considering long-term data
321 in modern ecological analyses and may also provide explanatory insights into limits of conservation

322 approaches when they fail to consider appropriate species distribution baselines. We expand on these
323 points below.

324 *Species implications*

325 For several species, we observe only a limited effect of including historical records in the analyses. This
326 indicates that modern occurrence records provide a reasonably good coverage of the climatic conditions
327 found in their historic distribution. This possibly reflects that they have been less impacted by past range
328 contractions, that range contraction did not affect the range of environmental conditions occupied by the
329 species or that they have successfully recovered throughout their historic range, whether by natural
330 recolonization or through active reintroductions. This result highlights the success of conservation efforts
331 in South Africa, where many species have been successfully reintroduced throughout their historic range
332 (e.g. black wildebeest, Cape mountain zebra) [60], with some species even introduced outside their native
333 range (e.g. giraffe, impala) [42] (though extralimital records have been excluded from the analyses and
334 thus the consequences of these introductions were not reflected in this study).

335 In contrast, considering historical data hugely affects estimates of the climatic niche and potential
336 distribution of other species. For these, the geographic distribution of predicted habitat suitability is wider
337 than expected from recent data only and this effect is higher for species of high conservation value. Three
338 of the five most impacted species are carnivores (lion, African wild dog and spotted hyaena) and the two
339 others are megaherbivores (elephant and hippopotamus), all being listed as threatened by the IUCN Red
340 List except the spotted hyaena, which is revealingly listed as near threatened on the Red List of Mammals
341 of South Africa, Swaziland and Lesotho [61]. These species are highly charismatic [62], very sensitive to
342 humans [63], and play important roles in ecosystems[64], thus acting as focal species for management
343 efforts and trophic rewilding initiatives [65]. Analyses based on recent data only will lead to truncated
344 estimates of bioclimatic relationships and underestimations of the extent of suitable areas for
345 conservation. Important suitable areas might be overlooked when selecting appropriate sites for
346 reintroductions and trophic rewilding, and protection efforts might focus on marginal habitat [5,6]. The
347 implications of such missed opportunities on management outcomes and the conservation status of
348 species need to be better understood.

349 *Community implications*

350 At the community level, neglecting historical records underestimates potential regional species richness,
351 with some areas being more impacted than others. In South Africa, the south-western and western parts

352 of the coastline as well as the central Free State and the Eastern Cape provinces have higher potential
353 richness than expected from recent records only. These areas were highly impacted historically, with the
354 establishment of the Cape Colony by the Dutch in the mid-17th century and the subsequent colonization
355 of the interior, with increased pressures from land-use change and direct hunting [11,66]. In most
356 bioregions, overlooking historical records underestimates the potential species richness, with particularly
357 strong effects in the Fynbos and Albany Thicket biomes. These shifted distribution baselines clearly have
358 implications for our understanding of broader biogeographic patterns and processes. As an example, the
359 underestimate of large mammal species richness in the Fynbos biome illustrated here demonstrates that
360 the role of mammals in this biome, traditionally considered to support a low diversity of large mammals
361 [13,68], needs to be reassessed. In addition to having suffered major biodiversity declines in the past
362 [13,67], these areas are also where conservation efforts are thus most likely to be misguided (but see [13]
363 for conservation planning in the Fynbos), which carries major implications for wildlife management and
364 conservation in South Africa.

365 Our study area doesn't cover the full distribution range of some species, meaning that we are only
366 sampling part of the environmental conditions that these species might encounter throughout their range.
367 While this limits the transferability of predictions in space or time, our results remain valid at the regional
368 level because we don't extrapolate outside the environmental space sampled in the occurrence dataset.
369 South Africa is an exceptional ecoregion, with unique climatic regimes, high species richness and
370 endemism [69,70]. Being at the southern margin of some species' global distribution, it is a particularly
371 important area for conservation since it may harbor populations with unique local adaptations that will
372 be critical for species' ability to persist in the face of future climate change [71]. Range contractions that
373 truncate species-climate relationships in this area are thus even more critical for our understanding of
374 species' niche than those occurring at the center of the range.

375 *Implications for informing climate change and invasion risk*

376 HSMs are widely used to forecast species range shifts under contemporary climate change [72] and to
377 assess the geographic risk of species' invasions [73]. In a study investigating the impact of overlooking
378 historical records in HSMs for 36 North American mammal species, Faurby and Araújo found that forecasts
379 of climate change impacts on biodiversity are unlikely to be reliable without acknowledging for past
380 anthropogenic range contraction [74]. Our study provides further evidence that using recent distribution
381 records only can underestimate species' bioclimatic niche, which in turn is likely to provide biased forecast
382 of species' response to climate change. Similarly, this truncated understanding of suitable bioclimatic

383 areas will make us vulnerable to underestimating invasion risk. While this latter aspect is of less relevance
384 for large mammals which are less frequently involved in invasions, the principle is of importance when
385 modelling risk areas for known invasive species.

386 *Setting baselines*

387 Shifting baselines [75] emphasize the need for setting appropriate references when exploring ecological
388 patterns and how these may change, especially for detecting long term processes. We have demonstrated
389 here the occurrence of shifted baselines for the distribution of South African mammals, against which one
390 can assess recent and future shifts in the geographic patterns of this fauna. Such phenomenon can be
391 expected elsewhere, and there is thus a need to study this at a global scale.

392 This study focuses on Pre-European settlement conditions, which are often held up as a relevant baseline
393 from which to define restoration objectives and quantify success [13,76,77]. For South Africa, we consider
394 the 15th century baseline to be appropriate for identifying species' niche since the bulk of human-related
395 impact for extant large mammal species occurred after this period [11] (but see [78]). But this might not
396 hold true in other systems, where human impact on mammal megafauna occurred much earlier [79]. The
397 appropriate baseline should be adapted accordingly, to estimate natural diversity patterns and allow the
398 identification of sites that match the biotic and abiotic needs of the focal species.

399 **CONCLUSION**

400 The recognition that neglecting long-term biodiversity might lead to setting inappropriate spatial
401 baselines is the first step towards a better integration of these data in decision-making for biodiversity
402 conservation and management. Due to the difficulty in collecting historical occurrence records, long-term
403 datasets are not currently available for all taxa or regions. However, with the recent recognition of the
404 value of these datasets for conservation, there is an encouraging development towards assembling long-
405 term biodiversity datasets [e.g. 80–82], including for underrepresented taxa [e.g. 83,84]. The release of
406 global databases of historic distributions [85] is a promising avenue to integrate long-term perspectives
407 in future ecological studies. We join previous calls for international, multidisciplinary effort to compile
408 historical data [86], and urge that, whenever possible, these should be included into conservation and
409 biogeography studies. Unless efforts are made to integrate this historical perspective into biodiversity
410 conservation, shifted distribution baselines risk undermining our efforts to define appropriate protected
411 areas and halt the ongoing biodiversity crisis, as well as appropriately manage biodiversity under global
412 change.

413 **ACKNOWLEDGMENTS**

414 We thank Margaret Avery and Shaw Badenhorst for assistance regarding the archeological records,
415 Matthew Child and the Endangered Wildlife Trust for providing access to the recent occurrence dataset
416 and Ana Rodrigues, for providing comments on an earlier version of the manuscript.

417 **SUPPLEMENTARY INFORMATION**

418 *Supplementary Information 1:* Summary table of results for the 34 species of large mammals included in
419 the analyses.

420 *Supplementary Information 2:* HSMs results and maps for 32 species of large mammals included in the
421 analyses (spatial data for the two species of rhinoceros are not provided for security reasons).

422

423 **REFERENCES**

- 424 1. Dinerstein E *et al.* 2017 An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm.
425 *BioScience* **67**, 534–545. (doi:10.1093/biosci/bix014)
- 426 2. Asaad I, Lundquist CJ, Erdmann MV, Costello MJ. 2017 Ecological criteria to identify areas for
427 biodiversity conservation. *Biol. Conserv.* **213**, 309–316. (doi:10.1016/j.biocon.2016.10.007)
- 428 3. Dirzo R, Young HS, Galetti M, Ceballos G, Isaac NJB, Collen B. 2014 Defaunation in the
429 Anthropocene. *Science* **345**, 401–406. (doi:10.1126/science.1251817)
- 430 4. Faurby S, Svenning J-C. 2015 Historic and prehistoric human-driven extinctions have reshaped global
431 mammal diversity patterns. *Divers. Distrib.* **21**, 1155–1166. (doi:10.1111/ddi.12369)
- 432 5. Kerley GIH, Kowalczyk R, Cromsigt JPGM. 2012 Conservation implications of the refugee species
433 concept and the European bison: king of the forest or refugee in a marginal habitat? *Ecography* **35**,
434 519–529. (doi:10.1111/j.1600-0587.2011.07146.x)
- 435 6. Cromsigt JPGM, Kerley GIH, Kowalczyk R. 2012 The difficulty of using species distribution modelling
436 for the conservation of refugee species - the example of European bison. *Divers. Distrib.* **18**, 1253–
437 1257. (doi:10.1111/j.1472-4642.2012.00927.x)
- 438 7. Willis KJ, Araujo MB, Bennett KD, Figueroa-Rangel B, Froyd CA, Myers N. 2007 How can a knowledge
439 of the past help to conserve the future? Biodiversity conservation and the relevance of long-term
440 ecological studies. *Philos. Trans. R. Soc. B Biol. Sci.* **362**, 175–187. (doi:10.1098/rstb.2006.1977)
- 441 8. Schipper J *et al.* 2008 The status of the world’s land and marine mammals: diversity, threat, and
442 knowledge. *Science* **322**, 225–230. (doi:10.1126/science.1165115)

- 443 9. Kerley GIH, Sims-Castley R, Boshoff AF, Cowling RM. 2009 Extinction of the blue antelope
444 Hippotragus leucophaeus: modeling predicts non-viable global population size as the primary driver.
445 *Biodivers. Conserv.* **18**, 3235–3242. (doi:10.1007/s10531-009-9639-x)
- 446 10. Boshoff A, Landman M, Kerley GIH. 2016 Filling the gaps on the maps: historical distribution
447 patterns of some larger mammals in part of southern Africa. *Trans. R. Soc. South Afr.* **71**, 1–65.
448 (doi:10.1080/0035919X.2015.1084066)
- 449 11. Boshoff AF, Kerley GIH. 2015 Lost Herds of the Highveld: Evidence from the Written, Historical
450 Record. *Afr. J. Wildl. Res.* **45**, 287–300. (doi:10.3957/056.045.0287)
- 451 12. Spear D, Chown SL. 2009 The extent and impacts of ungulate translocations: South Africa in a global
452 context. *Biol. Conserv.* **142**, 353–363.
- 453 13. Kerley GI., Pressey RL, Cowling RM, Boshoff AF, Sims-Castley R. 2003 Options for the conservation of
454 large and medium-sized mammals in the Cape Floristic Region hotspot, South Africa. *Biol. Conserv.*
455 **112**, 169–190. (doi:10.1016/S0006-3207(02)00426-3)
- 456 14. IUCN/SSC. 2013 Guidelines for Reintroductions and Other Conservation Translocations. , viiii + 57
457 pp.
- 458 15. Rodrigues A, Pilgrim J, Lamoreux J, Hoffmann M, Brooks T. 2006 The value of the IUCN Red List for
459 conservation. *Trends Ecol. Evol.* **21**, 71–76. (doi:10.1016/j.tree.2005.10.010)
- 460 16. Wolf C, Ripple WJ. 2017 Range contractions of the world’s large carnivores. *R. Soc. Open Sci.* **4**,
461 170052. (doi:10.1098/rsos.170052)
- 462 17. Wolf C, Ripple WJ. 2018 Rewilding the world’s large carnivores. *R. Soc. Open Sci.* **5**, 172235.
463 (doi:10.1098/rsos.172235)
- 464 18. Guisan A, Zimmermann NE. 2000 Predictive habitat distribution models in ecology. *Ecol. Model.* **135**,
465 147–186.
- 466 19. Guisan A, Thuiller W. 2005 Predicting species distribution: offering more than simple habitat
467 models. *Ecol. Lett.* **8**, 993–1009. (doi:10.1111/j.1461-0248.2005.00792.x)
- 468 20. Blonder B, Lamanna C, Violle C, Enquist BJ. 2014 The n -dimensional hypervolume. *Glob. Ecol.*
469 *Biogeogr.* **23**, 595–609. (doi:10.1111/geb.12146)
- 470 21. Rodríguez JP, Brotons L, Bustamante J, Seoane J. 2007 The application of predictive modelling of
471 species distribution to biodiversity conservation. *Divers. Distrib.* **13**, 243–251. (doi:10.1111/j.1472-
472 4642.2007.00356.x)
- 473 22. Franklin J. 2013 Species distribution models in conservation biogeography: developments and
474 challenges. *Divers. Distrib.* **19**, 1217–1223. (doi:10.1111/ddi.12125)
- 475 23. Blonder B. 2017 Hypervolume concepts in niche- and trait-based ecology. *Ecography* **41**, 1441–
476 1455. (doi:10.1111/ecog.03187)

- 477 24. Thuiller W, Brotons L, Araújo MB, Lavorel S. 2004 Effects of restricting environmental range of data
478 to project current and future species distributions. *Ecography* **27**, 165–172. (doi:10.1111/j.0906-
479 7590.2004.03673.x)
- 480 25. Newbold T. 2010 Applications and limitations of museum data for conservation and ecology, with
481 particular attention to species distribution models. *Prog. Phys. Geogr.* **34**, 3–22.
482 (doi:10.1177/0309133309355630)
- 483 26. Hendricks SA, Sesink Clee PR, Harrigan RJ, Pollinger JP, Freedman AH, Callas R, Figura PJ, Wayne RK.
484 2016 Re-defining historical geographic range in species with sparse records: Implications for the
485 Mexican wolf reintroduction program. *Biol. Conserv.* **194**, 48–57.
486 (doi:10.1016/j.biocon.2015.11.027)
- 487 27. Lentini PE, Stirnemann IA, Stojanovic D, Worthy TH, Stein JA. 2018 Using fossil records to inform
488 reintroduction of the kakapo as a refugee species. *Biol. Conserv.* **217**, 157–165.
489 (doi:10.1016/j.biocon.2017.10.027)
- 490 28. Monsarrat S, Pennino MG, Smith TD, Reeves RR, Meynard CN, Kaplan DM, Rodrigues ASL. 2015
491 Historical summer distribution of the endangered North Atlantic right whale (*Eubalaena glacialis*): a
492 hypothesis based on the environmental niche of a congeneric species. *Divers. Distrib.* **21**, 925–937.
493 (doi:10.1111/ddi.12314)
- 494 29. Monsarrat S, Boshoff AF, Kerley GIH. 2018 Accessibility maps as a tool to predict sampling bias in
495 historical biodiversity occurrence records. *Ecography* (doi:10.1111/ecog.03944)
- 496 30. Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009 Sample selection
497 bias and presence-only distribution models: implications for background and pseudo-absence data.
498 *Ecol. Appl.* **19**, 181–197. (doi:10.1890/07-2153.1)
- 499 31. Fourcade Y, Engler JO, Rödder D, Secondi J. 2014 Mapping Species Distributions with MAXENT Using
500 a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for
501 Correcting Sampling Bias. *PLoS ONE* **9**, e97122. (doi:10.1371/journal.pone.0097122)
- 502 32. Ruete A. 2015 Displaying bias in sampling effort of data accessed from biodiversity databases using
503 ignorance maps. *Biodivers. Data J.* **3**, e5361. (doi:10.3897/BDJ.3.e5361)
- 504 33. Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberón J, Villalobos
505 F. 2011 The crucial role of the accessible area in ecological niche modeling and species distribution
506 modeling. *Ecol. Model.* **222**, 1810–1819. (doi:10.1016/j.ecolmodel.2011.02.011)
- 507 34. Cooper JC, Soberón J. 2018 Creating individual accessible area hypotheses improves stacked species
508 distribution model performance. *Glob. Ecol. Biogeogr.* **27**, 156–165. (doi:10.1111/geb.12678)
- 509 35. South African National Biodiversity Institute. 2012 2012 Vegetation Map of South Africa, Lesotho
510 and Swaziland [vector geospatial dataset]. Available Biodivers. GIS Website Downloaded 08 March
511 2017
- 512 36. Boshoff AF, Kerley GI. 2010 Historical mammal distribution data: how reliable are written records?
513 *South Afr. J. Sci.* **106**, 26–33.

- 514 37. Monsarrat S, Kerley GIH. 2018 Charismatic species of the past: Biases in reporting of large mammals
515 in historical written sources. *Biol. Conserv.* **223**, 68–75. (doi:10.1016/j.biocon.2018.04.036)
- 516 38. Avery DM. 2019 *A Fossil History of Southern African Land Mammals*. Cambridge: Cambridge
517 University Press. See [https://www.cambridge.org/core/books/fossil-history-of-southern-african-](https://www.cambridge.org/core/books/fossil-history-of-southern-african-land-mammals/41969EC1E7739F4775954E6ADA8EA036)
518 [land-mammals/41969EC1E7739F4775954E6ADA8EA036](https://www.cambridge.org/core/books/fossil-history-of-southern-african-land-mammals/41969EC1E7739F4775954E6ADA8EA036).
- 519 39. Child MF, Roxburgh L, Do Linh San E, Raimondo D, Davies-Mostert HT. 2017 *Mammal Red List 2016:*
520 *Introduction and Methodology*. South Africa: South African National Biodiversity Institute and
521 Endangered Wildlife Trust.
- 522 40. Veloz SD. 2009 Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-
523 only niche models. *J. Biogeogr.* **36**, 2290–2299. (doi:10.1111/j.1365-2699.2009.02174.x)
- 524 41. Boria RA, Olson LE, Goodman SM, Anderson RP. 2014 Spatial filtering to reduce sampling bias can
525 improve the performance of ecological niche models. *Ecol. Model.* **275**, 73–77.
526 (doi:10.1016/j.ecolmodel.2013.12.012)
- 527 42. Castley JG, Boshoff AF, Kerley GIH. 2001 Compromising South Africa’s natural biodiversity-
528 inappropriate herbivore introductions. *South Afr. J. Sci.* **97**, 344–348.
- 529 43. Fick SE, Hijmans RJ. 2017 Worldclim 2: New 1-km spatial resolution climate surfaces for global land
530 areas. *Int. J. Climatol.*
- 531 44. NASA/METI/AIST/Japan Spacesystems, U.S./Japan ASTER Science Team. 2009 ASTER Global Digital
532 Elevation Model [Data set]. *NASA EOSDIS Land Process. DAAC* (doi:10.5067/ASTER/ASTGTM.002)
- 533 45. Andrews P, O’Brien EM. 2000 Climate, vegetation, and predictable gradients in mammal species
534 richness in southern Africa. *J. Zool.* **251**, 205–231. (doi:10.1111/j.1469-7998.2000.tb00605.x)
- 535 46. Hijmans RJ. 2014 raster: Geographic data analysis and modeling. R package version 2.2-12.
- 536 47. R Core Team. 2018 *R: A language and environment for statistical computing*. Vienna, Austria: R
537 Foundation for Statistical Computing. See URL <http://www.R-project.org/>.
- 538 48. Hutchinson GE. 1957 Concluding remarks. *Cold Spring Harb. Symp Quant Biol* **22**, 415–427.
- 539 49. Sanín C, Anderson RP. 2018 A framework for simultaneous tests of abiotic, biotic, and historical
540 drivers of species distributions: empirical tests for North American wood warblers based on climate
541 and pollen. *Am. Nat.* **192**, E000–E000.
- 542 50. Liu C, Berry PM, Dawson TP, Pearson RG. 2005 Selecting thresholds of occurrence in the prediction
543 of species distributions. *Ecography* **28**, 385–393.
- 544 51. Araujo M, New M. 2007 Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**, 42–47.
545 (doi:10.1016/j.tree.2006.09.010)

- 546 52. Schmitt S, Pouteau R, Justeau D, de Boissieu F, Birnbaum P. 2017 SSDM : An R package to predict
547 distribution of species richness and composition based on stacked species distribution models.
548 *Methods Ecol. Evol.* **8**, 1795–1803. (doi:10.1111/2041-210X.12841)
- 549 53. Allouche O, Tsoar A, Kadmon R. 2006 Assessing the accuracy of species distribution models:
550 prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**, 1223–1232. (doi:10.1111/j.1365-
551 2664.2006.01214.x)
- 552 54. Ferrier S, Guisan A. 2006 Spatial modelling of biodiversity at the community level. *J. Appl. Ecol.* **43**,
553 393–404. (doi:10.1111/j.1365-2664.2006.01149.x)
- 554 55. Calabrese JM, Certain G, Kraan C, Dormann CF. 2014 Stacking species distribution models and
555 adjusting bias by linking them to macroecological models. *Glob. Ecol. Biogeogr.* **23**, 99–112.
556 (doi:10.1111/geb.12102)
- 557 56. Godsoe W, Case BS. 2015 Accounting for shifts in the frequency of suitable environments when
558 testing for niche overlap. *Methods Ecol. Evol.* **6**, 59–66. (doi:10.1111/2041-210X.12307)
- 559 57. IUCN. 2018 The IUCN Red List of Threatened Species. Version 2018-2. <http://www.iucnredlist.org>.
- 560 58. Godsoe W. 2014 Inferring the similarity of species distributions using Species' Distribution Models.
561 *Ecography* **37**, 130–136. (doi:10.1111/j.1600-0587.2013.00403.x)
- 562 59. Mucina L, Rutherford MC, editors. 2006 *The vegetation of South Africa, Lesotho and Swaziland*.
563 Pretoria: SANBI.
- 564 60. Penzhorn BL. 1971 A summary of the re-introduction of ungulates into South African National Parks
565 (to 31 December 1970). *Koedoe Vol 14 No 1 1971*
- 566 61. Hunnicutt A, Power RJ, Lerm L, Page-Nicholson S, Mills MGL, Camacho G, Dalerum F, Child MF. 2016
567 A conservation assessment of *Crocota crocota*. In *The Red List of Mammals of South Africa,*
568 *Swaziland and Lesotho* (eds MF Child, L Roxburgh, E Do Linh San, D Raimondo, HT Davies-Mostert),
569 South Africa: South African National Biodiversity Institute and Endangered Wildlife Trust.
- 570 62. Albert C, Luque GM, Courchamp F. 2018 The twenty most charismatic species. *PLOS ONE* **13**,
571 e0199149. (doi:10.1371/journal.pone.0199149)
- 572 63. Riggio J, Kija H, Masenga E, Mbwilo F, Van de Perre F, Caro T. 2018 Sensitivity of Africa's larger
573 mammals to humans. *J. Nat. Conserv.* **43**, 136–145. (doi:10.1016/j.jnc.2018.04.001)
- 574 64. Malhi Y, Doughty CE, Galetti M, Smith FA, Svenning J-C, Terborgh JW. 2016 Megafauna and
575 ecosystem function from the Pleistocene to the Anthropocene. *Proc. Natl. Acad. Sci.* **113**, 838–846.
576 (doi:10.1073/pnas.1502540113)
- 577 65. Jarvie S, Svenning J-C. In press. Using species distribution modelling to determine opportunities for
578 trophic rewilding under future scenarios of climate change. , 10.

- 579 66. Pringle JA, Bond C, Clark J. 1982 *The Conservationists and the Killers: The Story of Game Protection*
580 *and the Wildlife Society of Southern Africa*. Bulpin. See
581 <https://books.google.co.za/books?id=LmnPAAAAMAAJ>.
- 582 67. Biggs R, Reyers B, Scholes RJ. 2006 A biodiversity intactness score for South Africa. *South Afr. J. Sci.* ,
583 9.
- 584 68. Skinner JD, Chimimba CT. 2005 *The Mammals of the Southern African Sub-region*. Cambridge
585 University Press. See <https://books.google.co.za/books?id=I6RhVKyFfjkC>.
- 586 69. Olson DM, Dinerstein E. 2002 The Global 200: Priority Ecoregions for Global Conservation. *Ann. Mo.*
587 *Bot. Gard.* **89**, 199. (doi:10.2307/3298564)
- 588 70. Cowling RM, Rundel PW, Desmet PG, Esler KJ. 1998 Extraordinary high regional-scale plant diversity
589 in southern African arid lands: subcontinental and global comparisons. *Divers. Distrib.* , 11.
- 590 71. Rehm EM, Olivas P, Stroud J, Feeley KJ. 2015 Losing your edge: climate change and the conservation
591 value of range-edge populations. *Ecol. Evol.* **5**, 4315–4326. (doi:10.1002/ece3.1645)
- 592 72. Pearson RG, Dawson TP. 2003 Predicting the impacts of climate change on the distribution of
593 species: are bioclimate envelope models useful? *Glob. Ecol. Biogeogr.* **12**, 361–371.
- 594 73. Peterson AT. 2003 Predicting the Geography of Species' Invasions via Ecological Niche Modeling. *Q.*
595 *Rev. Biol.* **78**, 419–433. (doi:10.1086/378926)
- 596 74. Faurby S, Araújo MB. 2018 Anthropogenic range contractions bias species climate change forecasts.
597 *Nat. Clim. Change* **8**, 252–256. (doi:10.1038/s41558-018-0089-x)
- 598 75. Soga M, Gaston KJ. 2018 Shifting baseline syndrome: causes, consequences, and implications. *Front.*
599 *Ecol. Environ.* **16**, 263–263.
- 600 76. White PS, Walker JL. 1997 Approximating Nature's Variation: Selecting and Using Reference
601 Information in Restoration Ecology. *Restor. Ecol.* **5**, 338–349. (doi:10.1046/j.1526-
602 100X.1997.00547.x)
- 603 77. Egan D, Howell EA, Meine C. 2005 *The Historical Ecology Handbook: A Restorationist's Guide to*
604 *Reference Ecosystems*. Island Press. See <https://books.google.co.za/books?id=pTqAYoj4znQC>.
- 605 78. Badenhorst S. 2015 Intensive hunting during the Iron Age of Southern Africa. *Environ. Archaeol.* **20**,
606 41–51. (doi:10.1179/1749631414Y.0000000039)
- 607 79. Sandom C, Faurby S, Sandel B, Svenning J-C. 2014 Global late Quaternary megafauna extinctions
608 linked to humans, not climate change. *Proc. R. Soc. B Biol. Sci.* **281**, 20133254–20133254.
609 (doi:10.1098/rspb.2013.3254)
- 610 80. Clavero M, Delibes M. 2013 Using historical accounts to set conservation baselines: the case of Lynx
611 species in Spain. *Biodivers. Conserv.* **22**, 1691–1702. (doi:10.1007/s10531-013-0506-4)

- 612 81. Turvey ST, Crees JJ, Di Fonzo MMI. 2015 Historical data as a baseline for conservation:
613 reconstructing long-term faunal extinction dynamics in Late Imperial–modern China. *Proc. R. Soc. B*
614 *Biol. Sci.* **282**, 20151299. (doi:10.1098/rspb.2015.1299)
- 615 82. Burgio K, Carlson C, Bond A. 2018 Georeferenced sighting and specimen occurrence data of the
616 extinct Carolina Parakeet (*Conuropsis carolinensis*) from 1564 - 1944. *Biodivers. Data J.* **6**, e25280.
617 (doi:10.3897/BDJ.6.e25280)
- 618 83. McClenachan L, O'Connor G, Neal BP, Pandolfi JM, Jackson JB. 2017 Ghost reefs: Nautical charts
619 document large spatial scale of coral reef loss over 240 years. *Sci. Adv.* **3**, e1603155.
- 620 84. Popejoy T, Randklev CR, Neeson TM, Vaughn CC. 2018 Prioritizing sites for conservation based on
621 similarity to historical baselines and feasibility of protection: Zooarchaeological Baseline. *Conserv.*
622 *Biol.* (doi:10.1111/cobi.13128)
- 623 85. Faurby S, Davis M, Pedersen RØ, Schowaneck SD, Antonelli A, Svenning J-C. 2018 PHYLACINE 1.2: The
624 Phylogenetic Atlas of Mammal Macroecology. *Ecology* (doi:10.1002/ecy.2443)
- 625 86. Clavero M, Revilla E. 2014 Biodiversity data: Mine centuries-old citizen science. *Nature* **510**, 35–35.
626 (doi:10.1038/510035c)
- 627