



University College London  
Department of Genetics, Evolution and Environment

# **Traits influence responses to land-use and climate change in terrestrial vertebrates**

**Adrienne Etard**

Supervision: Dr. Tim Newbold, Dr. Alex Pigot

April 28, 2022

Submitted in part fulfilment of the requirements  
for the degree of Doctor of Philosophy in Ecology at University College London

Word count: ~XXXXXX



# Declaration

I, Adrienne Etard, confirm that the work presented in this thesis is my own. Some parts have been conducted in collaboration with other researchers, and the contributions of co-authors to specific Chapters are described in the *Thesis outline of contents* on page 6. All else is appropriately referenced, and where information has been derived from other sources, I confirm that this has been indicated in the Thesis.

A handwritten signature in black ink, appearing to read "Adrienne Etard".

# Acknowledgments

First and foremost, huge thanks to my primary supervisor, Tim Newbold, whose guidance has been pivotal to the completion of this PhD journey. I am extremely grateful to Tim for enabling, trusting and encouraging me throughout, for the countless opportunities, for his availability and optimism.

I am equally grateful to Alex Pigot for welcoming me in his lab, for his cutting-edge insights on my work, and for the weekly Zoom meetings that made remote-working more bearable throughout the various phases of the Covid-19 pandemic.

I would like to extend my thanks to the wonderful people of CBER, both to those with whom I have directly collaborated and to those who have facilitated my work in this journey with insightful conversations. Thank you for making the workplace a friendly, constructive and welcoming environment.

To my friends in London and elsewhere who put up with my shenanigans. In particular, Yumi & Tom. To my family for their spirits and their support, Anne, Jean-François, Vivien, Octave & Caroline.

And, last but not least, a very special thank to Tea Trains, Coffee Breaks & Journal Club Cookies.

# **Abstract**

# **Impact statement**

# Data and code access

## Code access

The main pieces of R code for Chapters 2, 3, 4 and 5 are available at: [https://github.com/AdrienneEtard/PhD\\_thesis\\_code](https://github.com/AdrienneEtard/PhD_thesis_code).

This document was compiled with L<sup>A</sup>T<sub>E</sub>X; the source code and files are available at: <https://github.com/AdrienneEtard/Thesis>

## Data

The data used in this Thesis are freely accessible and sources are referenced throughout. In particular:

- the PREDICTS database (**Hudson2014; Hudson2017**) can be downloaded from: <https://data.nhm.ac.uk/dataset/the-2016-release-of-the-predicts-database>;
- the trait datasets I compiled in Chapter 2 are available at: [https://figshare.com/articles/dataset/Global\\_gaps\\_in\\_terrestrial\\_vertebrate\\_trait\\_data/10075421](https://figshare.com/articles/dataset/Global_gaps_in_terrestrial_vertebrate_trait_data/10075421).

# **Thesis outline of contents, authorship and collaborations**

# **Contents**

# **List of Tables**

# **List of Figures**

# **1 | General introduction**

## **2 | Global gaps and biases in trait data for terrestrial vertebrates**

### **Keywords**

Terrestrial vertebrates; traits; coverage; completeness; taxonomic biases; spatial biases; phylogenetic biases.

### **Abstract**

Trait data are increasingly used in studies investigating the impacts of global changes on the structure and functioning of ecological communities. Despite a growing number of trait data collations for terrestrial vertebrates, there is to date no global assessment of the gaps and biases the data present. Here, I assess whether terrestrial vertebrate trait data are taxonomically, spatially and phylogenetically biased. I compile seven ecological traits and quantify coverage as the proportion of species for which an estimate is available. For a species, I define completeness as the proportion of non-missing values across traits. I assess whether coverage and completeness differ across classes and examine phylogenetic biases in trait data. To investigate spatial biases, I test whether wider-ranging species have more complete trait data than narrow-ranging species. Additionally, I test whether species-rich regions, which are of most concern for conservation, are less well-sampled than species-poor regions. My results show that mammals and birds are well-sampled even in species-rich regions. For reptiles and amphibians (herptiles), only body size presents a high coverage (>80%), as well as habitat related variables (for amphibians). Herptiles are poorly sampled for other traits. The shortfalls are particularly acute in some species-rich regions and for certain clades. Across all classes, geographically rarer species have less complete trait information. Hence, trait information is less available on average in some of the most diverse areas and in geographically rarer species, both critical for biodiversity conservation. Gaps in trait data may impede our ability to conduct large scale analyses, while

23 biases can impact the validity of extrapolations. A short-term solution to the problem is to estimate missing  
24 trait data using imputation techniques, while a longer-term and more robust filling of existing gaps requires  
25 continued data collection efforts.

## 26 **2.1 Introduction**

27 Species traits are fundamental to ecological and evolutionary research. Comparative studies regularly use  
28 trait data across organisms to understand evolutionary processes and species coexistence (**Escudero2016**;  
29 **Zamudio2016**), to investigate global patterns of life forms and functions (**Diaz2016**), or to assess species'  
30 vulnerability to environmental changes (**Bohm2016**; **Pacifici2015**; **Pearson2014**). Because traits influence  
31 species' ability to cope with environmental changes (**Newbold2013**) and underpin species' contributions to  
32 ecosystem processes (**Lavorel2002a**; **Violle2007**; **Wong2018**), they play an increasingly important role in  
33 functional and conservation ecology.

34 Past and recent efforts to collate and release trait data in the public domain have facilitated the development  
35 of trait-based research. For instance, a global trait database has been published for plants (**Kattge2011**).  
36 As of May 2020, data from this database had been used in 297 publications since its release (Activity report,  
37 18/06/2020, <https://www.try-db.org/TryWeb/Home.php>). Such databases hence constitute in-  
38 valuable research tools and have the potential to greatly advance the field.

39 Vertebrates are one of the most studied taxa (**Titley2017**). There are now diverse sources of ecological traits for vertebrate groups (primates: **Galan-Acedo2019**; mammals: ‘PanTHERIA’, **Jones2009**; amniotes: **Myhrvold2015**; amphibians: ‘AmphiBIO’, **Oliveira2017**). These datasets stem from important efforts to collate published estimates of trait data and make them readily available. Trait data have also been made available on online platforms (for instance, the Global Assessment of Reptile Distribution initiative: <http://www.gardinitiative.org/>; the IUCN Red List of Threatened Species: <https://www.iucnredlist.org/>; BirdLife data zone: <http://datazone.birdlife.org/home>).

46 Nevertheless, despite the importance of vertebrate species in global research outputs, there is no single  
47 source for vertebrate ecological traits. Consequently, researchers wishing to conduct comparative studies  
48 across vertebrate groups may have to collate trait data from a range of sources (such as in **Cooke2019a**;  
49 **Cooke2019b** or in **Gonzalez2018**), a time-consuming prerequisite which may be a limiting step of the re-  
50 search process. Indeed, collating data from heterogeneously-formatted sources presents many challenges  
51 (**Schneider2019**), particularly when working across a large number of species. For instance, traits may be  
52 measured differently across datasets; units may be inconsistent; and taxonomic resolution and nomenclature

53 may vary.

54 The lack of a curated, readily available global database for vertebrate ecological traits impedes our ability  
55 to conduct cross-taxon comparative studies at global scales. However, efforts to collate data into a single  
56 database are limited by the availability of underlying data. Because there exist important gaps in biodiversity  
57 knowledge (**Hortal2015**), trait datasets are often incomplete, with many species lacking estimates for many  
58 traits. The incompleteness of ecological trait data at the species level has been termed the ‘Raunkiær  
59 shortfall’ by **Hortal2015**. Furthermore, incomplete trait data are likely to be biased. Biases in trait data may  
60 be the consequence of uneven taxonomic and spatial collection effort, with a set of charismatic or easily  
61 detectable species being more completely sampled. For instance, **Gonzalez-Suarez2012** investigated biases  
62 in global trait information in mammals. They notably found that the availability of mammalian trait data was  
63 geographically and phylogenetically biased, with larger and more widely distributed species being better  
64 sampled. In addition, data availability also differed across IUCN Red List extinction risk categories, with  
65 threatened species (Critically Endangered, Endangered or Vulnerable) being less well sampled for traits than  
66 non-threatened species (Least Concern or Near Threatened).

67 A major issue with incomplete, biased data is the introduction of bias in subsequent analyses. Assessing  
68 the amount of missing data as well as the so-called ‘missingness mechanism’ – whether missing data are  
69 missing at random, as opposed to there being systematic biases in the way missing values are distributed,  
70 see **Baraldi2010** – is an important prerequisite. Indeed, there exist diverse techniques to deal with data  
71 missingness. The simplest one consists of retaining complete cases only by filtering out missing values  
72 (case deletion, see **Nakagawa2008**). Nevertheless, case deletion may lead to biased parameter estimates  
73 and erroneous conclusions when values are not missing at random (**Gonzalez-Suarez2012**). Therefore, it  
74 is critical to determine the most appropriate way to deal with data incompleteness. For instance, previous  
75 studies using terrestrial vertebrate trait data have employed multiple imputation techniques to fill in the gaps  
76 (**Gonzalez-Suarez2012; Cooke2019b**). Yet, imputation techniques could be sensitive to non-randomness  
77 in trait data. Phylogenetic biases (where some clades are under-sampled compared to other clades) could  
78 notably impact the performance of several imputation approaches. It is thus vital to characterise the gaps in  
79 trait data prior to any analysis. However, there has been no study to date investigating global patterns in the  
80 availability of trait data across terrestrial vertebrates.

81 Here, I aim to assess the global state of trait data in terrestrial vertebrates. I focus on a set of traits that are  
82 available across the four classes and that are commonly used by ecologists: body size; litter or clutch size;  
83 longevity; trophic level; activity time; habitat breadth; and a measure of habitat specialisation. I quantify and  
84 compare the gaps in trait data across classes by calculating the coverage of each trait across species, and the

85 completeness of trait estimates for each species (Box 1). I investigate taxonomic, spatial and phylogenetic  
86 biases in trait coverage and completeness.

87 Given that biodiversity research is globally biased towards birds and mammals (**Titley2017**), I hypothe-  
88 sise that herptiles are less well sampled for traits than mammals and birds, having both lower coverage and  
89 completeness.

90 Furthermore, building upon previous studies conducted on mammals (**Gonzalez-Suarez2012**), I hypoth-  
91 esize that species rarity influences completeness, focusing on species geographical range size as one aspect  
92 of rarity. Widely distributed species could be better sampled than narrowly distributed species because their  
93 ranges overlap with more study sites, regardless of their abundance. As such, I test whether species geo-  
94 graphical range size explains trait completeness.

95 It is well established that global research effort is distributed unequally (**ONU2015**), with patterns under-  
96 pinned by various geographical and socioeconomic factors. For instance, countries with higher gross domes-  
97 tic product tend to host a larger number of research institutions (**Martin2012**). The proximity of research  
98 infrastructures and the accessibility of survey sites play an important part in explaining the global distribu-  
99 tion of knowledge (**Hortal2015**). As a result of these factors, biodiversity data gaps tend to be greater in  
100 tropical areas (**Collen2008**). Tropical areas have the greatest species richness, and so these data biases are of  
101 great concern for biodiversity conservation. Whether species-rich regions are systematically under-sampled  
102 for traits compared to species-poor regions is thus important to assess, given the significance of species-rich  
103 areas for global conservation. Here, I investigate spatial biases in trait completeness, hypothesizing that  
104 species-rich areas are on average less well sampled than species-poor areas.

105 Finally, I investigate phylogenetic biases in the trait data. I assess whether particular clades have received  
106 more attention than others by looking for patterns in the distribution of trait completeness across the terminal  
107 branches of phylogenetic trees in each class.

### Box 1. Definitions

*Trait:* Sensu stricto, a characteristic measurable at the level of an individual and that influences organismal fitness or performance (**Violle2007**). In this thesis, I broaden this definition to include ‘ecological’ traits (e.g., the number of habitats used by a species), where the relationship of a species to the surrounding environment needs to be considered. Ecological traits may be estimated by aggregating data across multiple individuals.

*Trait completeness:* For a given species, the proportion of traits for which an estimate is available.

*Trait coverage:* For a given trait, the proportion of species for which an estimate is available.

## 108 2.2 Methods

109 I produced class-specific trait datasets that were made available on figshare (DOI: 10.6084/m9.figshare.  
110 10075421). Data compilation and all analyses were conducted with R version 3.5.1 (**R\_citation**). Dis-  
111 tribution maps were processed using both R and the ArcPy package available in ArcGIS v.10.6 (**ESRI**)  
112 (implemented in Python 2.7; **Python\_citation**).

### 113 2.2.1 Trait data collection

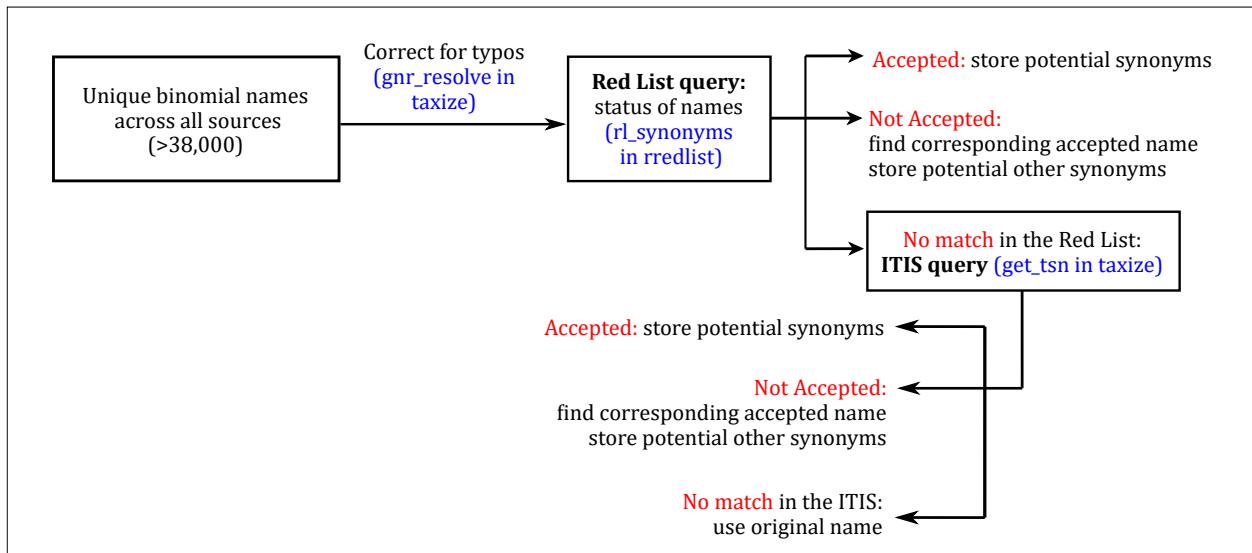
#### 114 Sources and taxonomic matching

115 I used freely accessible secondary sources in my compilation (Table ??), selected for their broad taxonomic  
116 coverage and/or for their frequent use in macroecological studies. Across sources, similar species could  
117 appear under synonymous names. This was a potential problem for matching sources by binomial names.  
118 Indeed, synonymy can artefactually decrease trait coverage, when trait information is not available across  
119 all synonyms. Notably, difficulties arise when species have been divided into several subspecies or when  
120 different subspecies are clumped together. Systematic manual checks could not be applied considering the  
121 scale of the data collection (there were >39,000 unique binomial names across sources). I developed a  
122 procedure aiming at identifying one accepted name for each of the binomial names found across sources.  
123 When I could not find an accepted name, I used the original name. Figure ?? summarizes the main steps;  
124 similar solutions have been used in other large-scale studies (**Cooke2019b**).

125 Briefly, the procedure consisted of extracting synonyms from the IUCN (**IUCN2020**) or from the Inte-  
126 grated Taxonomic Information System (ITIS; <https://www.itis.gov/>), using the rredlist (**rredlist**)  
127 and taxize (**Chamberlain2013**) R packages. One accepted name was assigned to each synonym. I pro-  
128 duced a “Synonym” dataset that I have also made available. I then normalized taxonomy across sources by

129 replacing binomial names with their identified accepted name where applicable.

130 Given that different taxonomic backbones could be used to correct for taxonomy, I make two versions  
131 of the trait compilations available (corrected and not corrected for taxonomy), meaning that users are free to  
132 apply their own corrections; for example, taxonomy could be aligned to that of class-specific sources, such  
133 as The Reptile Database, the American Museum of Natural History's Amphibian Species of the World, the  
134 Mammal Diversity Database or the International Ornithological Congress World Bird List. Datasets cor-  
135 rected for taxonomy contain 11,634 species of birds, 5,381 mammals, 10,612 reptiles and 6,990 amphibians.  
136 Where no taxonomic correction was applied when matching sources, the compiled datasets contain 13,501  
137 birds, 5,791 mammals, 11,012 reptiles and 8,583 amphibians. For more information, see Appendix 2 (S2.1;  
138 Figure ??).



**Figure 2.1: Procedure used to identify the accepted names of species.** I extracted, where possible, the accepted names of species from either the IUCN Red List or the Integrated Taxonomic Information System (ITIS). Figure reproduced from Etard2020.

### 139 Compilation methods

140 For continuous traits, I took the median value within species when multiple estimates were available from  
141 different sources, after removal of any repeated values, which were assumed to represent estimates dupli-  
142 cated across secondary compilations and derived from the same underlying primary sources. Although  
143 intraspecific variation is increasingly being recognized to have important effects on ecological systems  
144 (**Bolnick2011; Siefert2015; DesRoches2018; GonzalezSuarez2012a**), it was not feasible to obtain mea-  
145 sures of intraspecific variability from all sources; therefore, estimates were provided as a single measure for

**Table 2.1: Data sources for each trait.** Abbreviations: A = amphibians; B = birds; BL = body length; BM = body mass; DA = diel activity time; GL = generation length; H = habitat data; LCS = litter or clutch size; L/ML = longevity or maximum longevity; M = mammals; MA = age at sexual maturity; R = reptiles; RS = range size; TL = trophic level. Note. Data sources may contain more traits than shown here. Tick marks in parentheses indicate that the trait was present in the data source but that another closely related trait with a better coverage was used instead. The tilde character (~) before a tick mark indicates that I derived trophic levels from species diet. <sup>1</sup> <http://datazone.birdlife.org/home>; <sup>2</sup> <https://www.iucnredlist.org/resources/spatial-data-download>; <sup>3</sup> <http://apiv3.iucnredlist.org/api/v3/docs##general>. Table reproduced from Etard2020.

Sources	Taxa	Traits									
		BM	BL	L/ML	MA	GL	LCS	TL	DA	RS	H
Oliveira2017	Amphibians	(✓)	✓	(✓)	✓		✓	~✓	✓		
Cooper2008							✓				
Sodhi2008			✓								
Wilman2014	Birds	✓						~✓	✓		
BirdLife <sup>1</sup>		✓				✓				✓	
Jones2009	Mammals	✓	(✓)	(✓)	(✓)		✓		✓		
Kissling2014								✓			
Gainsbury2018								✓			
Wilman2014		✓							✓		
Pacifici2015		✓				✓					
Scharf2015	Reptiles	✓		✓	(✓)		✓	✓	✓		
Vidan2017										✓	
Stark2018		✓		✓			✓		✓		
Schwarz2017							✓				
Novosolov2017		✓						✓			
Novosolov2013							✓				
Slavenko2016		✓									
FeldmanGEB2016		✓									
Meiri2018GEB					✓		✓	✓	✓		
Meiri2015								✓	✓		
Roll2017											✓
Myhrvold2015	B, M, R	✓	✓	✓	(✓)		✓				
IUCN2020 <sup>2</sup>	A, B, M									✓	
IUCN2020 <sup>3</sup>	All										✓

<sup>146</sup> each species. For some species and some traits, measures were provided separately for females and males.

<sup>147</sup> In such cases, I first obtained the mean of these two measures.

<sup>148</sup> Across sources, there were multiple traits related to each of body size and life span. For instance, body

<sup>149</sup> mass and/or body length information could be provided. Different proxies were also available for life span,

<sup>150</sup> such as the age at sexual maturity or generation length. In such cases, I focused on the trait presenting the

151 highest coverage.

152 • **Body size**

153 Adult body mass estimates were compiled for mammals, birds and reptiles. Body length information  
154 was compiled for amphibians, because the coverage for body length was higher than that for body  
155 mass. Body mass and body length are known to scale allometrically, although the allometric relation-  
156 ship differs across amphibian clades (**Santini2018**). In the amphibian dataset, Pearson's correlation  
157 coefficient between log(Body mass) and log(Body length) was 0.71 (data points shown in Appendix  
158 2, S2.2, Figure ??).

159 • **Longevity**

160 I defined longevity as the life span of an individual and maximum longevity as the longest life span  
161 reported. I used closely related traits when longevity/maximum longevity was not available or when  
162 longevity/maximum longevity had a poorer coverage than a related trait. I selected the age at sexual  
163 maturity for amphibians; Pearson's correlation coefficient between log(Age at sexual maturity) and  
164 log(Maximum longevity) was 0.55 (Appendix 2, S2.2, Figure ??). I compiled the generation length  
165 for mammals and birds. The correlation between log(Generation length) and log(Longevity) was 0.74  
166 for mammals and 0.70 for birds (data points shown in Appendix 2, S2.2, Figure ??). Finally, I used  
167 maximum longevity directly for reptiles.

168 • **Litter or clutch size**

169 The number of offspring (litter size) or eggs (clutch size) was compiled directly from the sources and  
170 treated as equivalent across classes. I reported measures of central tendencies provided by the sources  
171 where applicable; otherwise, I calculated range midpoints (mean of smallest and largest reported lit-  
172 ter/clutch sizes).

173 • **Trophic level**

174 In all classes, species were described as omnivores, carnivores or herbivores. For reptiles and mam-  
175 mals, this information was compiled directly from the sources. For amphibians and birds, trophic  
176 levels were not provided. For these two classes, I inferred trophic levels from dietary information  
177 (Table ??). For birds, I used the primary diet (based on food items recorded as composing  $\geq 50\%$   
178 of the diet of a species). Diet for amphibians was described without respect to the percentage use of  
179 food items; simply as a binary record of whether or not food items were used. In both cases, species  
180 recorded to only consume plant-based resources (seeds, nectar, fruit or other plant material) were

181 classified as herbivores. Species consuming only animal resources (invertebrates or vertebrates) were  
182 classified as carnivores. Species consuming a mixture of plant and animal resources were classified as  
183 omnivores.

184 • **Activity time**

185 Species were described as being either nocturnal or non-nocturnal. Despite a higher resolution of  
186 activity time information in some of the sources (e.g., species being described as cathemeral, crepus-  
187 cular or diurnal), I adopted the classification of the source with the lowest resolution (EltonTraits:  
188 **Wilman2014**, for birds), in order to have consistent information across classes. As such, all species  
189 defined as diurnal, cathemeral or crepuscular were classified as non-nocturnal, as opposed to species  
190 classified as strictly nocturnal.

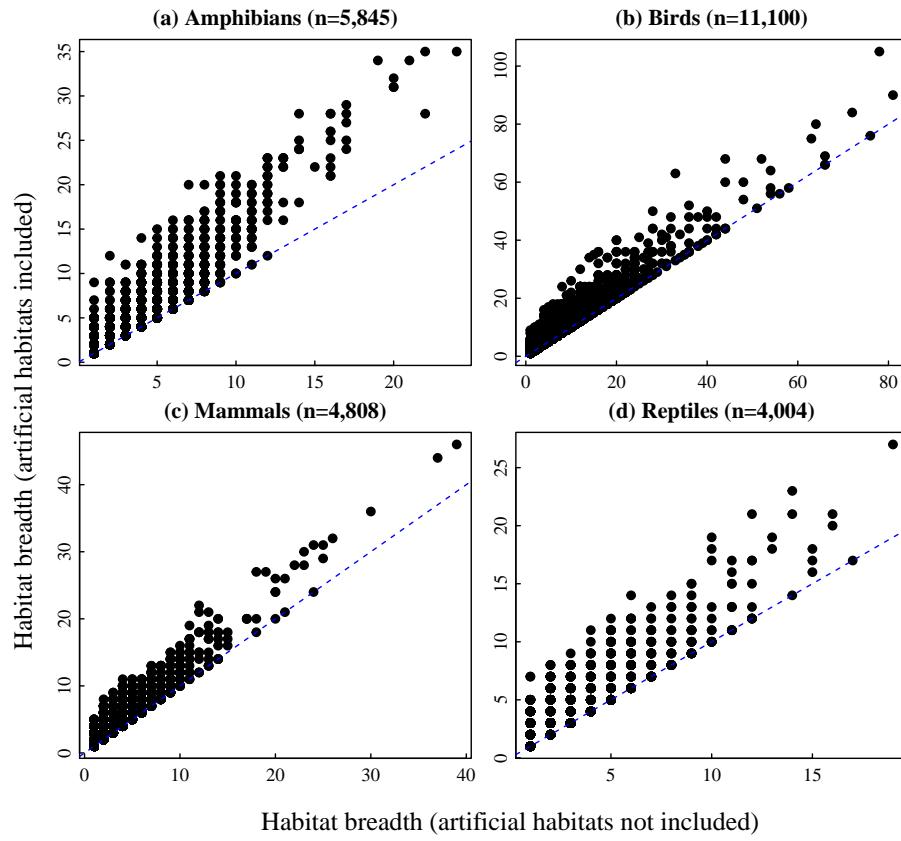
191 • **Habitat breadth**

192 I used IUCN habitat data (**IUCN2020**), which describe species habitat preferences and the suitability  
193 and importance of each habitat. I defined habitat breadth as the number of habitats a species was  
194 known to use, using level 2 of the IUCN Habitat Classification Scheme for description of habitat  
195 types (divided into: Forest, Savanna, Shrubland, Grassland, Wetland, Rocky areas, Caves and subter-  
196 ranean, Desert, Marine, Marine intertidal or coastal/supratidal, Artificial, Introduced vegetation, and  
197 Other/Unknown.) Note that the total number of habitats, determined by including those that qualify as  
198 artificial, correlates positively with the number of natural habitats used (Figure ??).

199 • **Use of artificial habitats**

200 For a species, I recorded whether any artificial habitat was reported to be suitable in the IUCN habitat  
201 data.

202 Finally, the compiled datasets contain an additional column, ‘Note’, where I reported species found to  
203 be extinct or extinct in the wild (EW). I used species Red List status and information from **Meiri2018GEB**  
204 to flag such species. I reported 75 extinct/EW species for mammals, 160 for birds, 34 for amphibians and 53  
205 for reptiles. It is likely that the datasets contain extinct species that I could not flag, because they were not  
206 recorded as extinct in the sources I used.



**Figure 2.2: Number of natural and artificial habitats used by a species against number of strictly natural habitats used by a species.** Pearson's correlation coefficients show a high positive correlation between these two metrics of habitat breadth in all terrestrial vertebrate classes: 0.92 for amphibians (a), 0.95 for birds (b), 0.94 for mammals (c), and 0.90 for reptiles (d).

## 207 Phylogenies

208 I used class-specific phylogenetic trees downloaded on 13 April 2020. For mammals, I used ‘complete’  
 209 trees from **Faurby2018; Faurby2020**, downloaded from <https://zenodo.org/record/3690867#.Xyc5wyhKhPZ>. For amphibians, birds and squamates, I obtained trees from <https://data.vertlife.org/>. The original sources were as follows: **Jetz2012** for birds; **Jetz2018** for amphibians; and **Tonini2016** for squamates. For each class, a distribution of 1,000 trees was available. For plotting purposes, I obtained  
 210 consensus trees using the TreeAnnotator program of the BEAST software (**Bouckaert2019**).  
 211  
 212  
 213

## 214 Species distributions

215 I obtained extent-of-occurrence distribution maps for reptiles from **Roll2017**, available at: <https://datadryad.org/stash/dataset/doi:10.5061/dryad.83s7k> (downloaded 13 April 2020).  
 216  
 217 For mammals and amphibians, species distribution maps were obtained from the IUCN Red List (**IUCN2020**,

218 downloaded 13 April 2020); for birds, they were obtained from BirdLife International (<http://datazone.birdlife.org/species/requestdis>, downloaded 17 April 2020).

220 For amphibians, mammals and birds, I selected areas of extant or probably extant presence only. Additionally, I selected areas where species were resident or present during the breeding season, and I excluded areas occupied during the non-breeding season or where species were considered vagrant.

223 In addition, for all classes, I excluded occupied areas that fell outside the known elevational limits of species, where such data were available. Lower and upper elevational limits were retrieved from the IUCN Red List (queried using the rredlist package) and were available for approximately half of the species (Supporting Information, Appendix 2, S2.3, Figure ??). Decreases in range sizes were observed after cutting distribution maps by the known elevational limits (Appendix 2, S2.3, Figure ??).

## 228 **2.2.2 Investigating gaps and biases in trait data**

229 I used trait coverage and completeness to investigate taxonomic, phylogenetic and spatial biases in the trait data. Table ?? summarizes the sample sizes (number of species) in each of the following analyses. Note 230 that species for which completeness was 0% were included in all analyses (for more details, see Figure ??). 231 Also note that I did not filter out species identified as extinct or extinct in the wild, because they represented 232 a small proportion of the datasets (0.48% for amphibians, 1.4% for both birds and mammals, and 0.50% for 233 reptiles) and also because I could not exclude such species systematically, because it is likely that I did not 234 flag them all.

**Table 2.2: Number of species for each analysis.** All species represented in the trait datasets were included in (1). All species from the class-specific phylogenetic trees or from the distribution maps that matched with species in the trait datasets were included in (2) and (3). *Table reproduced from Etard2020.*

	(1) Taxonomic biases	(2) Phylogenetic biases	(3) Spatial biases
Amphibians	6,990	6,170	5,650
Birds	11,634	8,315	10,802
Mammals	5,381	5,171	5,046
Reptiles	10,612	9,404	9,382

236 **Taxonomic biases**

237 I tested whether completeness varied across taxonomic class using pairwise Wilcoxon rank sum tests. I tested  
238 for the extent and performance of the taxonomic corrections by looking at trait coverage when taxonomic  
239 corrections are applied and when no correction is applied (Appendix 2, S2.4, Figure ??).

240 **Phylogenetic biases**

241 Initially, to assess whether more closely related species were more likely to be similar in trait completeness,  
242 I estimated the phylogenetic signal in completeness with Pagel's  $\lambda$  (**Pagel1999**) in each class. I used a  
243 bootstrapping approach, calculating  $\lambda$  for each of 50 trees randomly sampled in each class (using the phylosig  
244 function of the phytools R package; **Revell2012**). I then estimated the mean and 95% confidence intervals  
245 (95% CIs) of  $\lambda$ . Sample sizes for computing  $\lambda$  (number of species represented in both the phylogenies and  
246 trait datasets) are shown in Table ??.

247 I then plotted within-family median completeness in phylogenetic trees built at the family level, using  
248 the consensus trees. Within-family median completeness was calculated using taxonomic information in the  
249 trait datasets (sample sizes shown in Table ??).

250 **Spatial biases**

251 I first investigated whether wider-ranging species were more likely to be better sampled than narrow-ranging  
252 species. I tested for a relationship between species range size and trait completeness. I fitted a generalized  
253 linear model with a Poisson error distribution (directly using the number of sampled traits, ' $N_{traits}$ ', rather  
254 than the proportion (completeness)). Class was added as a predictor interacting with range size; thus the  
255 model was:

256 
$$N_{traits} \sim \log(\text{Range size}) * \text{Class}.$$

257 Second, I mapped assemblage-level median completeness. Assemblages were characterized at the pixel level  
258 at 50 km<sup>2</sup> resolution. I determined pixel-level composition and richness by stacking species geographical  
259 distributions. I then calculated median completeness across species in each pixel. I show the resulting  
260 maps for herptiles in the main text, and for mammals and birds in Appendix 2 (S2.5, Figure ??; median  
261 completeness was very high across most pixels for mammals and birds). In addition, I provide maps of  
262 assemblage-level mean completeness and standard deviation for all classes in Appendix 2 (S2.5; Figures ??  
263 and ?? show corresponding maps; Figure ?? shows standard deviation against species richness).

I then tested for a spatial correlation between species richness and median completeness. Given that median completeness was very high across most pixels for mammals and birds, I fitted such models for herptiles only. I fitted spatial autoregressive lag models to explain assemblage-level median completeness as a function of species richness (using the function `lagsarlm` of the `spatialreg` package (**spatialreg1**; **spatialreg2**; **spatialreg3**)). Given that responses could vary geographically, I included the biogeographical realm as an interacting factor (using the World Wide Fund for Nature (WWF) ecoregion shapefile to characterise realms, obtained from <https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world>). The considered realms were Afrotopics, Australasia, Indo-Malayan, Nearctic, Neotropics and Palaearctic. To improve normality, I arc-sin square-root transformed completeness values and log-transformed species richness. The `lagsarlm` function allows for a consideration of spatial autocorrelation in the dependent variable by estimating the autoregressive lag coefficient,  $\rho$ , associated with an n-by-n matrix of spatial weights,  $W$ . The final model was:

$$\text{arcsin}(\sqrt{\text{Completeness}}) \sim \log(\text{Species richness}) * \text{realm} + \rho \cdot W \cdot \text{arcsin}(\sqrt{\text{Completeness}}).$$

The value of  $W$  was estimated using the functions `tri2nb` and `nb2listw` of the `spdep` package (**spatialreg3**; **spdep1**). Fitting the model using all grid cells was computationally intractable; therefore, I randomly sampled cells for this analysis (using 30% of the grid cells in each realm). I selected grid cells where species richness was higher than three to avoid sampling issues. I fitted separate models for amphibians and reptiles, because when adding class as an interacting predictor, the same cells (with the same coordinates) might be sampled for multiple classes, whereas the `tri2nb` function does not tolerate duplicated coordinates.

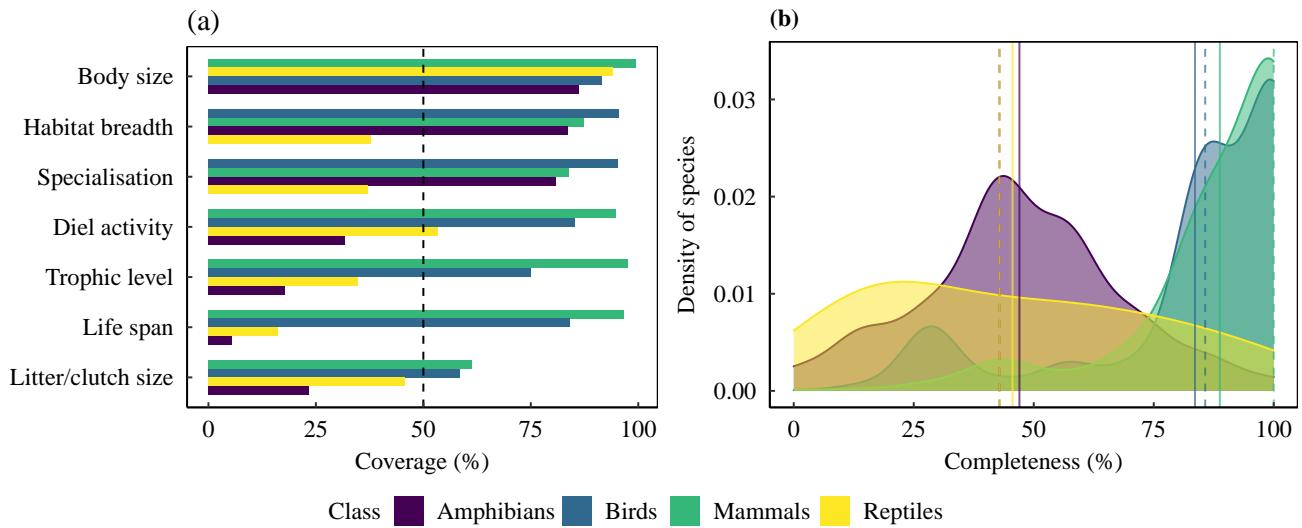
## 2.3 Results

### 2.3.1 Taxonomic biases in trait information

Trait coverage for mammals and birds was overall high (Figure ??(a); mean and median coverage across traits: 89% and 95% for mammals; 84% and 85% for birds). In both cases, litter/clutch size was the trait with the poorest coverage (61% for mammals and 59% for birds). Coverage exceeded 80% for all other traits (except trophic level for birds, at 75% coverage).

Conversely, trait coverage was more variable for herptiles, and poorer overall (Figure ??(a); mean and median trait coverage: 47% and 32% for amphibians, 46% and 38% for reptiles). Coverage exceeded 80% only for body size in both reptiles and amphibians and for habitat related traits in amphibians only. In all other cases coverage was <55%, with very little information available for longevity-related traits.

293 Trait completeness (proportion of non-missing trait values for a species) reflected similar biases (Figure  
 294 ??(b)). The distribution of trait completeness varied significantly among classes (pairwise Wilcoxon rank  
 295 sum test: p-value<0.0001 in all cases). Distributions were highly left skewed in mammals and birds (skew-  
 296 ness: -2 and -1.6). 84% of all mammalian species and 80% of avian species fell in the 80–100% completeness  
 297 range. Moreover, the completeness distribution was moderately right skewed for reptiles (skewness: 0.4),  
 298 and slightly right skewed for amphibians (skewness: 0.02). 56% of all reptiles and 57% of amphibians fell  
 299 in the 0-50% completeness range.



**Figure 2.3: Trait coverage and completeness across species.** (a) I defined coverage as the proportion of species for which an estimate is available for a given trait. The dashed line represents 50% coverage. (b) Trait completeness is the proportion of estimated traits for a species. Here, I show the distribution of completeness. Continuous lines represent the mean trait completeness for each class, whereas dashed lines represent the median trait completeness. Note that there were species with 0% completeness (230 species for amphibians – 3.3% of amphibian species in the trait dataset; 9 for birds – 0.077% of species; 7 for mammals – 0.13% of species; and 161 for reptiles – 1.5% of species). Species with 0% completeness were retained in the datasets when there was information for traits I did not select in the analyses, but no known value for the traits I did select. For instance, the body mass of the amphibian species *Rhinella centralis* was known, but other trait values (including body length) were missing, meaning that *Rhinella centralis* had 0% completeness for the set of traits I considered. *Figure reproduced from Etard2020*.

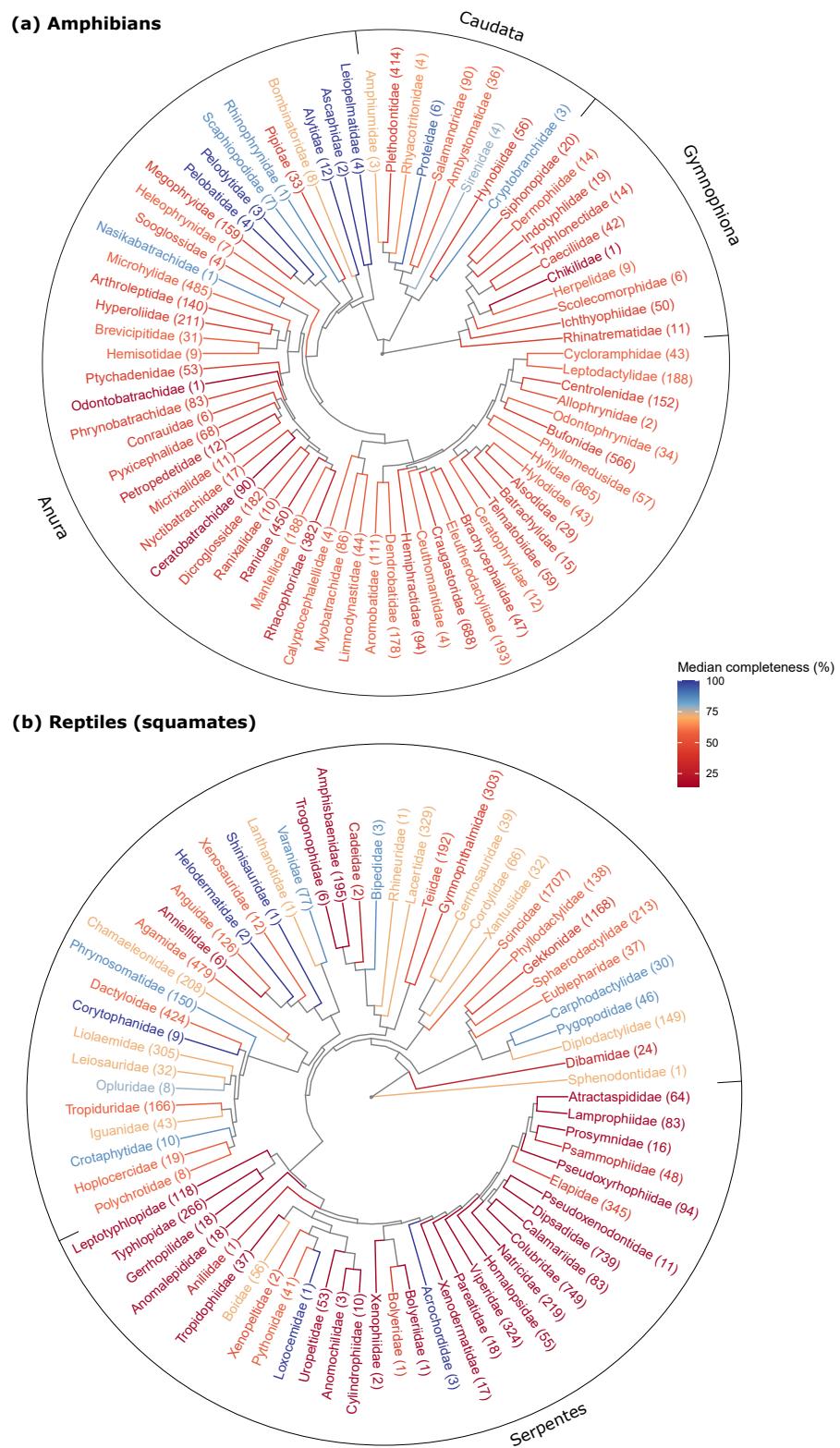
### 300 2.3.2 Phylogenetic biases in trait completeness

301 As expected from the distribution of trait completeness in mammals and birds (Figure ??), within-family  
 302 median trait completeness was high across most tips of the phylogenetic trees (Appendix 2, Figures ?? and  
 303 ??; I present the avian and mammalian phylogenies in the Appendix because there was little variation in  
 304 completeness across tips). For birds,  $\lambda$  was  $0.71 (\pm 0.0053)$ . For mammals,  $\lambda$  was  $0.78 (\pm 0.0035)$ . This  
 305 indicated that, despite completeness generally being high across tips, the sampling was not evenly distributed  
 306 across the phylogeny.

307 In herptiles, clusters of families with similar median trait completeness appeared (Figure ??). In am-

308 phibians, groups of families belonging to the order Anura (frogs) showed both the best and worst median  
309 completeness (Figure ??(a)). The best-sampled families included the tailed frogs of the family Ascaphidae  
310 (two species) and species of the family Leiopelmatidae (four species endemic to New Zealand). The family  
311 Ceratobatrachidae (containing *c.* 90 species occurring in Southeast Asia and in some Pacific islands), the  
312 family Ranidae (true frogs, 450 species considered here) and the family Rhacophoridae (shrub frogs, 382  
313 species considered here) figured among the worst-sampled families. For amphibians,  $\lambda$  was  $0.63 (\pm 0.0039)$ .  
314 In reptiles, most snakes were poorly sampled, whereas families in other suborders appeared to be sampled  
315 better overall (Figure ??(b)). Within snakes, the pythons, boas, the three species of the family Acrochordidae  
316 and the python-like species of the family Loxocemidae were better sampled than other snake families. In  
317 reptiles,  $\lambda$  was  $0.69 (\pm 0.0032)$ . The sampling in herptiles was thus also uneven with regard to the phylogeny.

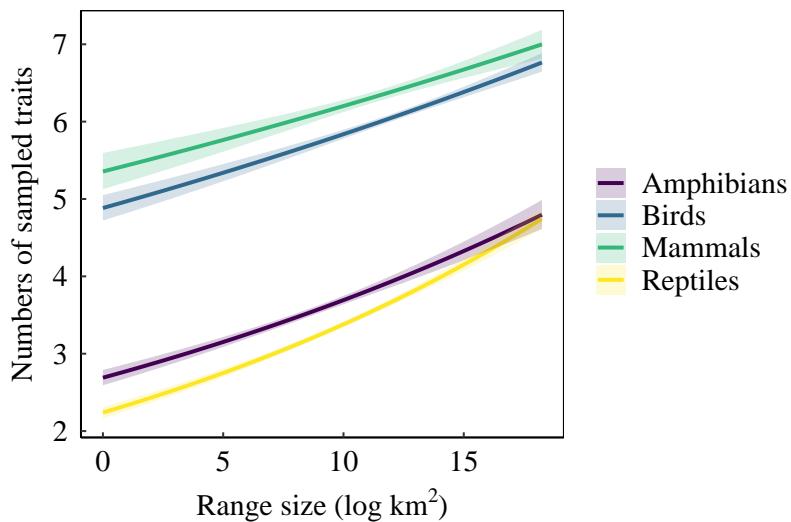
318 It is important to underline that Figure ?? shows within-family median completeness, masking the con-  
319 siderable variation in species richness across families, hence masking potential important variation in com-  
320 pleteness across species within families. For example, in the amphibian family Allophrynididae (three recog-  
321 nized species), the within-family median completeness was 50%; but the dataset comprised two species of  
322 completeness 14% and 86%, respectively. I present similar plots to those in Figure ?? showing the within-  
323 family standard deviation in completeness in Appendix 2 (Figure ??). Within-family standard deviation  
324 tended to increase with within-family species richness (Appendix 2, Figure ??).



**Figure 2.4: Within-family median trait completeness in herptiles.** The number next to each family name represents the number of species included in the calculation of the median. *Figure reproduced from Etard2020.*

### 325 2.3.3 Spatial biases in trait completeness

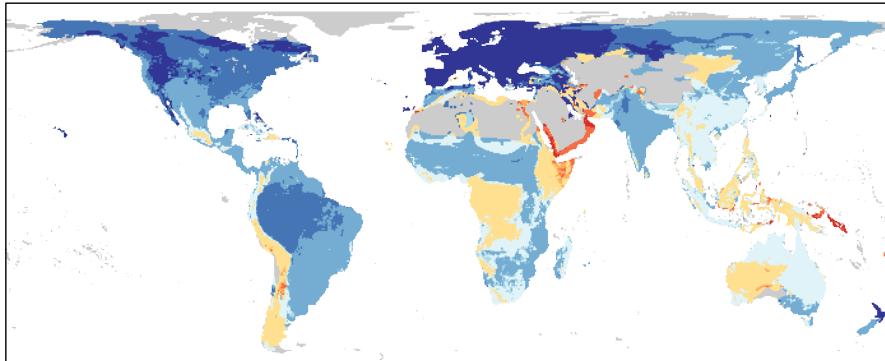
326 Range size was significantly correlated with the number of sampled traits. Larger range sizes were associated  
327 with a higher number of sampled traits (i.e., with higher completeness; Figure ??; Appendix 2, S2.7, Table  
328 ??). Similar results were obtained when using distribution maps not cut by elevational limits (Appendix  
329 2, S2.7, Table ??; Figure ??). The rate of increase was steepest for reptiles, then for amphibians, then for  
330 birds and mammals (slope estimates for birds and mammals were not significantly different from each other;  
331 Appendix 2, Table ??).



**Figure 2.5: Relationship between number of sampled traits and geographical range size.** Models were fitted using a Poisson error distribution. Class was added as a predictor interacting with range size. Rates of increase in number of sampled traits with range size were not significantly different for mammals and birds but differed for reptiles and amphibians, with the steepest rates of increase for reptiles. *Figure reproduced from Etard2020.*

332 There were marked spatial variations in median trait completeness in herptiles (Figure ??). North Amer-  
333 ica and Europe were well sampled for both amphibians and reptiles. Moreover, Southeast Asia and the Congo  
334 basin were on average less well sampled. In other regions, contrasting patterns emerged between amphibians  
335 and reptiles. For instance, median completeness was poorer for amphibians than for reptiles in Australia,  
336 but opposite patterns were observed in South America. As in the phylogenetic analyses, assemblage-level  
337 median completeness could mask potential important variation in completeness within species of a given  
338 assemblage. Assemblage-level mean and standard deviation maps are shown in Appendix 2 (S2.5, Figures  
339 ?? and ??). There was a trend for increasing standard deviation with increasing species richness, with a  
340 larger spread in standard deviation at lower species richness (Appendix 2, section S2.5, Figure ??).

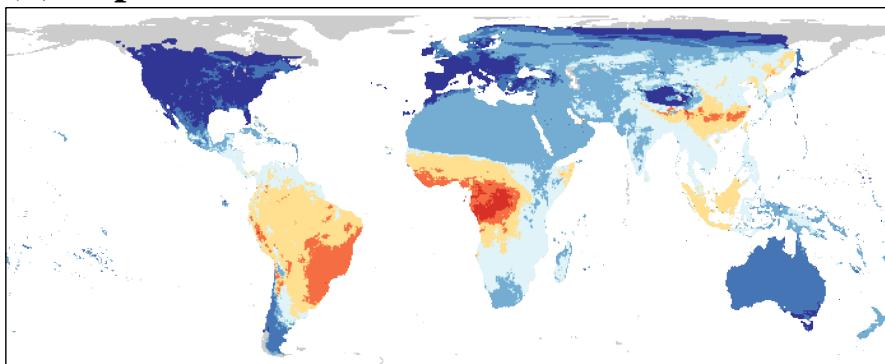
### (a) Amphibians



Median completeness:

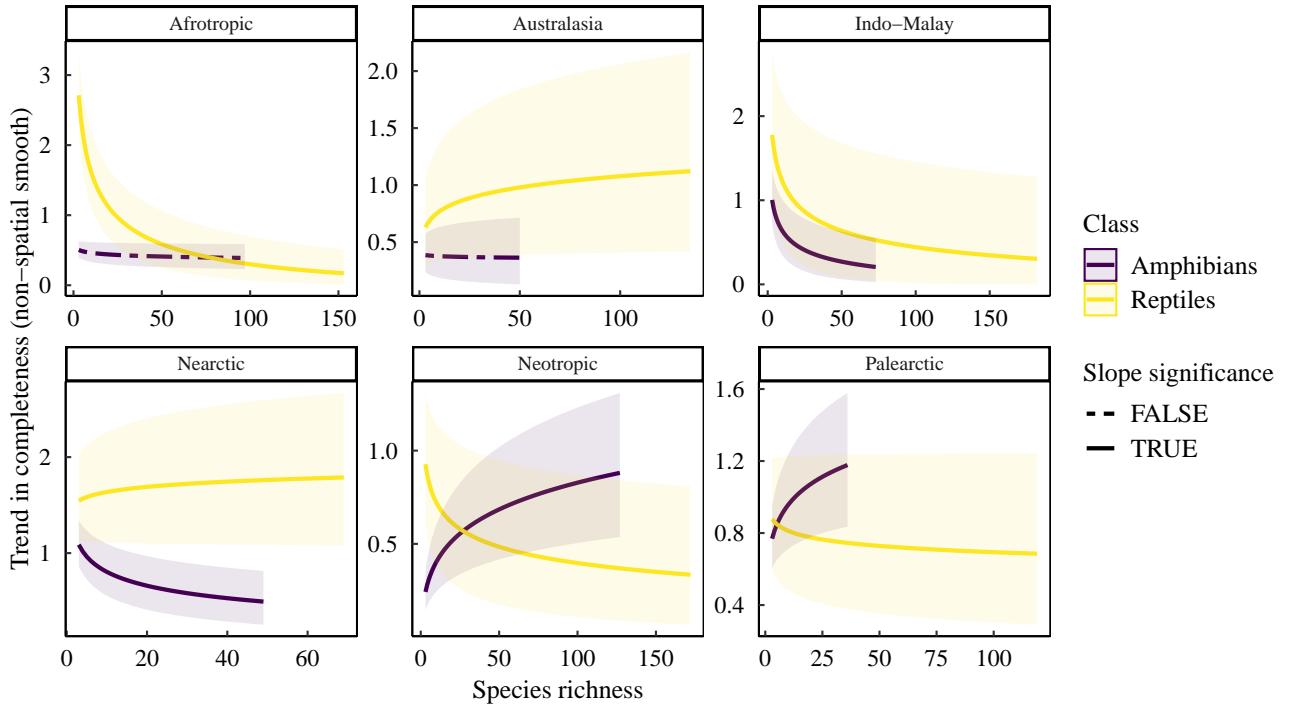
NA
0 – 10%
10 – 20%
20 – 30%
30 – 40%
40 – 50%
50 – 60%
60 – 70%
70 – 80%
80 – 90%
90 – 100%

### (b) Reptiles



**Figure 2.6: Spatial distribution of assemblage-level median trait completeness in herptiles.** Similar maps for birds and mammals are shown in Appendix 2 (S2.5, Figure ??). *Figure reproduced from Etard2020.*

341 Spatial models showed that species richness explained median trait completeness in herptiles in most  
342 realms (Figure ??; Appendix 2, S2.8, Tables ?? and ??); including spatial lags improved the models (am-  
343 phibians:  $\rho = 0.92$ , p-value < 0.0001; reptiles:  $\rho = 0.91$ , p-value < 0.0001). For reptiles, completeness was  
344 negatively correlated with species richness in the most species-rich realms (Afrotropics, Indo-Malayan and  
345 Neotropics) and in the Palaearctic; the relationship was steepest in the Afrotropics and shallowest in the  
346 Palaearctic. In the Australasian and Nearctic realms, completeness tended to increase with species richness.  
347 For amphibians, negative relationships were observed in the Indo-Malay and Nearctic realms, whereas pos-  
348 itive trends were observed in the Neotropics and the Palaearctic. The opposite trends between reptiles and  
349 amphibians observed in the Australasian and Neotropical realms reflected patterns observed on the maps.  
350 The Indo-Malayan was the only realm where median completeness tended to decrease with species richness  
351 for both reptiles and amphibians.



**Figure 2.7: Spatial model trends for herptiles.** The lines represent in-sample predictions ( $\pm$  standard error) for the trend components of the spatial models (trends after accounting for spatial autocorrelation). *Figure reproduced from Etard2020.*

## 352 2.4 Discussion

353 The results of this Chapter illustrate the taxonomic, spatial and phylogenetic dimensions of the knowledge  
 354 gaps in trait data, termed the Raunkiær shortfall by **Hortal2015**. To the best of my knowledge, this work  
 355 constitutes the first comparative assessment of global gaps for terrestrial vertebrate trait data, despite their  
 356 use in numerous studies. I showed that the trait data present important taxonomic, spatial and phylogenetic  
 357 biases, with contrasts in the availability of trait information between, on the one hand, herptiles and, on the  
 358 other hand, birds and mammals.

359 Birds and mammals are globally well sampled for the set of traits I considered, even in the most species-  
 360 rich assemblages. Moreover, the availability of trait information for herptiles is lower overall and phyloge-  
 361 netically and geographically biased. Several factors could interplay to shape these patterns. For instance,  
 362 species that are more easily detectable (for example, wider ranging) and more charismatic are likely to be  
 363 better sampled. Diverse socio-economic predictors could also contribute to geographical biases in trait data  
 364 sampling; global biases in primary data collection are likely to be one of the most important contributors to  
 365 the patterns I highlighted. Nevertheless, biases in the data could have been introduced at later stages, notably  
 366 with the selection of sources and traits. The global compilation I obtained in this chapter reflects, in part, the

367 interest and focus of the secondary data sources I used. It is possible that the addition of new sources from  
368 regional journals or other authorities could diminish spatial biases in the data by increasing coverage for  
369 certain areas. Nevertheless, I argue that by focusing on widely used traits, these results are likely to reflect  
370 the “true” availability of the data in primary sources and that the shortfalls for other, less used traits would  
371 be more pronounced.

372 I believe that the results presented here are robust to taxonomic uncertainty, although taxonomic match-  
373 ing might potentially be improved further using class-specific sources, such as the Reptile Database or Am-  
374 phibiaWeb, for identification of synonyms (but see Appendix 2, S2.9, Figure ??). I have made two versions  
375 of the data compilations available, one in which my own corrections were applied and one using the origi-  
376 nal binomial names of the sources, meaning that users are free to use their own taxonomic backbones and  
377 identify synonyms within the compilations. I believe that taxonomic matching is a recurring issue when  
378 working across thousands of species. Taxonomic synonymy artefactually inflates the numbers of identified  
379 species, potentially lowering trait coverage (whereas clumping subspecies together can have the opposite ef-  
380 fect). Tackling this problem is difficult (**Isaac2004; Jones2012**), notably because there is no global curated  
381 database recording the status of species names, and also because of the nature of taxonomy and the debates  
382 around the species concept (**May2011**). Nevertheless, taxonomic uncertainty can have important conse-  
383 quences. For instance, **Cardoso2017** showed that inaccuracies and errors in species checklists contributed  
384 to the overestimation of plant diversity in the Amazon (but see **Freeman2021**: the relative underdescription  
385 of species in tropical areas compared to temperate areas –‘taxonomic debt’, also referred to as ‘latitudinal  
386 taxonomic gradient’ by the authors– may lead to the underestimation of species richness at low latitudes).

387 Biases in trait data have important implications for conservation planning. Past studies have shown  
388 that narrow-ranged species, for which fewer trait data are available on average, have higher extinction risks  
389 (**Collen2016; Purvis2000; Ripple2017**) and are more negatively impacted by anthropogenic pressures than  
390 wider-ranging species (**Newbold2018a**). Trait information is also less available for herptiles in tropical re-  
391 gions such as the Congo basin, Southeast Asia and South America, which are some of the most diverse  
392 areas of crucial importance for worldwide conservation (**Barlow2018**). Consequently, trait information is  
393 on average less available where potentially more crucial to conservation planning. Indeed, trait information  
394 can be incorporated into vulnerability assessments and, as such, can help to prioritize conservation efforts.  
395 Species traits have been found to mediate species responses to environmental changes across diverse tax-  
396 onomic groups, and thus can inform on the sensitivity of species to anthropogenic pressures (**Flynn2009;**  
397 **Newbold2013; Nowakowski2017**). Traits are now commonly used to estimate species vulnerability or ex-  
398 tinction risks (**Pacifci2015; RamirezBautista2020**). As opposed to trend-based approaches, which rely on

399 historical population trends (changes in abundance or shifts in distributions) to predict species' vulnerability  
400 and extinction risks, trait-based approaches rely on species' intrinsic sensitivity to particular threats. The  
401 appeal of trait-based approaches to extinction risk estimation is that, by providing mechanistic insights, they  
402 diminish the amount of population information needed. If the responses of species to a threat consistently  
403 relate to certain traits, it is possible to generalize patterns across species for which population data are less  
404 available (**Verberk2013**). Integrating traits into vulnerability assessments is hence of particular interest when  
405 field monitoring of species population sizes or distributions is difficult to achieve, but biases in the data could  
406 mean that such information is lacking for some of the most vulnerable species.

407 Traits that influence species responses to environmental changes have been termed 'response traits' (or  
408 'response-mediating traits'; **Luck2012**), as opposed to 'effect traits' that underpin ecosystem functioning  
409 (**Lavorel2002a**). For instance, relative brain size and longevity have been characterized as response traits  
410 in birds (**Newbold2013; Sayol2020**), whereas dietary characteristics (e.g., trophic levels or guilds) are both  
411 response and effect traits. **Hortal2015** highlighted that, for plants, both response and effect traits have been  
412 investigated, whereas for vertebrates the research has been more focused on understanding species responses.  
413 This could be because the way vertebrate traits interact to shape some ecosystem processes has not yet been  
414 characterized well.

415 Ecosystem processes sustained by animals might be harder to quantify and might be influenced by a  
416 combination of traits. The traits compiled in this work are likely to have a role in diverse processes. Never-  
417 theless, there was one important omission, in that I did not compile species diet in this chapter, potentially the  
418 most straightforward trait to link with diverse processes, such as grazing, pollination, scavenging and seed  
419 dispersal. From a practical perspective, I chose traits that had been estimated at least for some of the species  
420 in each class, and that were readily available. Diet was excluded because although estimates were available  
421 for amphibians, birds and mammals, there was no readily available database for reptilian diet. Movement or  
422 dispersal abilities were also excluded because information was not readily available for any class. Although  
423 I expect that species diet and dispersal abilities would present similar sampling biases to the ones presented  
424 in this work, the addition of such traits to the compilation would represent a valuable contribution and would  
425 notably facilitate studies looking at the functional roles of reptiles.

426 For practical reasons, I did not consider intraspecific trait variation. Intraspecific variation has been  
427 shown to have important effects on ecological systems, and a growing body of literature encourages trait-  
428 based research to include intraspecific variability (**Guralnick2016**). There have been several calls to pro-  
429 duce open-access, global trait datasets (**Weiss2019**), including a representation of intraspecific trait variation  
430 (**Kissling2018**). Notably, **Schneider2019** designed a framework to store and share inter- and intraspecific

431 trait data, accompanied by an R package to standardize the data in a proposed format. Such a proposition  
432 could constitute an important step towards the unification of individual datasets into a single, comprehensive  
433 database for ecological trait data.

434 The current spatial and taxonomic gaps in trait data might limit our ability to scale studies up, whereas  
435 biases in the data can affect the validity of extrapolations to groups or areas that are undersampled. More gen-  
436 erally, biases and gaps in biodiversity data can have important implications for ecological studies. Data gaps  
437 can hinder our ability to draw conclusions on observed macroecological patterns. For example, **CHAUDHARY2016**  
438 proposed that marine species richness follows a bimodal distribution, peaking at mid-latitudinal locations,  
439 and argued that these patterns were not underpinned by knowledge gaps in species distributions. Neverthe-  
440 less, **Menegotto2018** attributed the tropical dip in marine species richness to a lack of species distribution  
441 data, explained by lower sampling efforts in tropical areas ('Wallacean' shortfall; **Hortal2015**). Biases and  
442 gaps in trait data could also affect studies in closely related fields, such as functional ecology – for instance,  
443 past studies have shown that functional diversity indices are sensitive to missing data (**Majekova2016**;  
444 **Pakeman2014**) – or community assembly (**Perronne2017**).

445 Ecologists should, therefore, take particular care when designing trait-based studies, because both data  
446 quality and data gaps are likely to influence the results and the generality of the conclusions. There exist  
447 diverse methods to deal with missing trait values, should data missingness be problematic. Complete re-  
448 moval of missing values ('case deletion') is commonly used but presents several issues, because it reduces  
449 sample size and statistical power and introduces potential bias in data subsamples (**Nakagawa2008**). For  
450 example, retaining complete cases only from the trait datasets would generate trait data disproportionately  
451 representative of mammals and birds, which would be problematic for conducting cross-taxon analysis on  
452 terrestrial vertebrates. As such, it is recommended that case deletion be applied only when data are missing  
453 completely at random, which is rarely the case (**Peugh2004**).

454 Alternatives to case deletion consist of filling in the gaps. In recent years, the development of imputa-  
455 tion techniques has provided robust methods to handle missing data. Such imputation techniques have been  
456 used to complete trait datasets in recent studies (**Cooke2019b**). **Penone2014** used a simulation approach  
457 to evaluate the performance of four of these techniques, namely PhyloPars (**Bruggeman2009**), random for-  
458 est algorithms as implemented in R with missForest (**Stekhoven2016**; **Stekhoven2012**), multivariate im-  
459 putation by chained equations (MICE; **micepackage**) and k-nearest neighbour (kNN; **Troyanskaya2001**).  
460 **Penone2014** introduced missing values (10%–80%) in a complete trait dataset of carnivorans and measured  
461 imputation performance in different scenarios. Given that phylogenetic non-randomness in missing trait  
462 values can impact imputation accuracy, **Penone2014** removed values in three different ways (completely

463 at random; with a phylogenetic bias; and with a body mass bias). Out of the four techniques, missForest  
464 and PhyloPars performed best when species phylogenetic position was included as a predictor of missing  
465 trait values. Such imputations appeared to be robust even when trait coverage was as low as 40%, which  
466 might be relevant for many reptilian and amphibian traits. The performance was not significantly affected  
467 by phylogenetic non-randomness of the data. Hence, missForest and PhyloPars appear to be well suited  
468 when traits are phylogenetically conserved, because they allow species phylogenetic position to be included  
469 as a predictor of missing trait values. The study by **Penone2014** highlights that there are robust imputation  
470 techniques allowing to deal with incomplete trait data where biases might otherwise be problematic. Never-  
471 theless, it is important to highlight that some imputation techniques, such as single or mean imputation, can  
472 be problematic because they do not allow an estimation of uncertainty and suffer from a lack of accuracy  
473 (**Nakagawa2008**); indeed, imputation techniques sometimes perform no better than case deletion. More  
474 work should be conducted to assess imputation performance in various contexts (see Johnson 2021), and the  
475 datasets compiled in this chapter might provide an opportunity for such studies.

476 Although robust imputation techniques can be useful for filling gaps in trait datasets, they are no substi-  
477 tute for continued data collection efforts. The results of this chapter show that data are particularly lacking  
478 in herptiles, notably in the Afrotropics, the Neotropics and the Indo-Malayan realms. For these areas, incor-  
479 porating regional databases into existing datasets could contribute to the reduction of global gaps. I believe  
480 that both primary research and subsequent efforts to integrate new data and existing databases are required  
481 if we are to collectively strive towards the unification of trait databases.

482 To conclude, this work constitutes, to my knowledge, the first assessment of the global gaps and biases  
483 in terrestrial vertebrate trait information. I show that herptiles are undersampled compared with mammals  
484 and birds, with important spatial and phylogenetic variability in the availability of trait information. Impu-  
485 tation techniques are one possible solution to these problems. Nevertheless, I believe that primary research,  
486 combined with efforts to complete existing datasets, is the only way to fill the current data gaps genuinely  
487 and robustly. I hope that the compiled trait dataset and these findings can prove useful for guiding further  
488 data collection efforts and for conducting macroecological analyses.

489 **3 | Intensive human land uses negatively af-**

490 **fect vertebrate functional diversity**

491 **Keywords**

492 Land use; land-use intensity; terrestrial vertebrates; functional diversity; traits.

493 **Abstract**

494 Land-use change is the leading driver of global biodiversity loss, thus characterising its impacts on the func-

495 tional structure of ecological communities is an urgent challenge. Using a database describing vertebrate

496 assemblages in different land uses, I assess how the type and intensity of land use affect the functional di-

497 versity of vertebrates globally. I find that human land uses alter local functional structure by driving declines

498 in functional diversity, with the strongest effects in the most disturbed land uses (intensely used urban sites,

499 cropland and pastures), and among amphibians and birds. Both tropical and temperate areas experience im-

500 portant functional losses, which are only partially offset by functional gains. Tropical assemblages are more

501 likely to show decreases in functional diversity that exceed those expected from species loss alone. These re-

502 sults indicate that land-use change non-randomly reshapes the functional structure of vertebrate assemblages,

503 raising concerns about the continuation of ecological processes sustained by vertebrates.

504 **3.1 Introduction**

505 Anthropogenic activities are profoundly transforming global biodiversity. Although multiple pressures act in

506 combination, land-use change currently poses the greatest threat to biodiversity (**Maxwell2016; Newbold2015**).

507 However, not all species respond similarly to land-use change. Traits have been found to explain species' sen-

508 sitivity to land-use change in diverse groups (**Newbold2013; Nowakowski2017; Quesnelle2014; Todd2017**).

509 Previous work has also shown that land-use change leads to non-random modification of assemblage trait  
510 composition (or functional diversity) (**Chapman2018; Colin2018; Flynn2009; LaSorte2018; Newbold2013;**  
511 **Tinoco2018**). Since it is widely acknowledged that biodiversity, and in particular trait diversity, may  
512 promote ecosystem functioning and stability, modification to the trait composition of assemblages could  
513 have far-reaching and adverse impacts on ecological processes (**Hooper2012; Magioli2021; Oliver2015;**  
514 **Tilman1994**).

515 Terrestrial vertebrates support many processes, ranging from pollination (**Ratto2018**), to seed dispersal  
516 to the regulation of lower trophic levels (**Barber2010; Letnic2012; Salo2010; Zhang2018**). However, we  
517 lack a global understanding of how the functional diversity of entire vertebrate assemblages responds to  
518 changes in land use. Most previous studies have been conducted at regional or local scales (**Davison2021**),  
519 but these may not be representative of global patterns. Indeed, recent global syntheses have highlighted how  
520 biodiversity responses can differ substantially between regions and across latitudes, with higher sensitivity  
521 reported for the tropics (**Matuoka2020; Millard2021; Newbold2020**). Another key issue is the taxonomic  
522 coverage of past work. Few studies investigating effects of land use on functional diversity have considered  
523 several vertebrate classes together, and comparative studies remain rare. Thus, how land-use change affects  
524 the functional diversity of local vertebrate assemblages at global scales, and the potential geographical and  
525 taxonomic variation in the effects, still largely remains to be explored.

526 Here, I aim to assess how human land use and land-use intensity affect the functional diversity of  
527 vertebrate assemblages, across and within taxonomic classes. Building on recent work (**Matuoka2020;**  
528 **Millard2021; Newbold2020**), I investigate differences in response between tropical and temperate regions.  
529 I use multiple response metrics to quantify functional diversity. First, functional richness measures the  
530 breadth and variety of trait combinations represented in an assemblage (**Legras2018**). Second, functional  
531 dispersion quantifies how similar species in a given assemblage are in terms of their traits (**Laliberte2010**).  
532 These metrics can mask important alterations of assemblage composition if functional losses are compen-  
533 sated for by functional gains. To address this, I consider pairwise measures between assemblages, to explore  
534 levels of functional loss and functional gain across land uses (Figure ??).

535 To this end, I combine (1) the trait data across terrestrial vertebrates collected in Chapter 2 (also published  
536 in **Etard2020**), with (2) global records of species occurrence in eight land-use types of differing intensity of  
537 use (the PREDICTS database: **Hudson2014; Hudson2017**, Figure ??, Table ??). The PREDICTS database  
538 is currently the most comprehensive database of sampled species occurrence, and for most records also  
539 abundance, across multiple land uses of different land-use intensity. Using the PREDICTS database allows  
540 us to contrast biodiversity metrics among intact land uses (primary-vegetation sites, considered to be the

541 undisturbed reference condition), and all other human land-use types. Specifically, I test the following  
542 hypotheses, both across and within taxonomic classes:

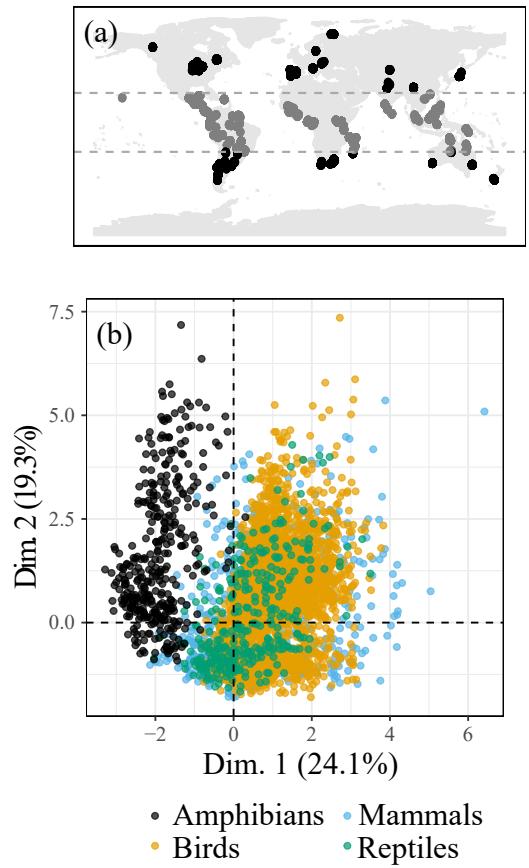
- 543 1. I expect decreases in functional diversity in human land uses compared to primary vegetation, caused  
544 by contractions of occupied trait space. I expect such effects to be more pronounced where land is used  
545 more intensively by humans. This hypothesis builds upon evidence that species with certain traits are  
546 more sensitive to land-use disturbance (**Newbold2013**), meaning that disturbed land uses will retain  
547 only disturbance-tolerant species, more functionally similar to one another. Given the reported higher  
548 sensitivity of tropical assemblages to land-use disturbance, I predict that such effects are stronger in  
549 the tropics.
- 550 2. I hypothesise that decreases in functional diversity in disturbed land uses exceed decreases expected  
551 by chance, given local species loss. Thus, I expect disturbed land uses to promote functional under-  
552 dispersion. Functional under-dispersion occurs when species within an assemblage are more simi-  
553 lar, in term of their traits, than expected by chance (**Cadotte2017; Wong2018**) – or, in other words,  
554 when functional dispersion is lower than expected given local species richness. I predict that under-  
555 dispersion is more likely to occur in the highly disturbed sites, in both temperate and tropical areas.  
556 This hypothesis is based on the idea that species are being removed non-randomly from sensitive areas  
557 of the trait space, and increasingly so with higher disturbance level.
- 558 3. Finally, I expect decreases in functional diversity in human land uses to be driven by high functional  
559 loss, whereby species are being removed from previously occupied areas of the trait space; I expect  
560 no functional gain. This hypothesis is based on the idea that the functional trait space in undisturbed  
561 land uses represents all of the possible regional trait combinations and that species with functional  
562 attributes rendering them unable to persist in altered conditions will be filtered out (**Cornwell2006**).

## 563 3.2 Methods

### 564 3.2.1 Vertebrate assemblages

565 I used vertebrate occurrence data from the PREDICTS database (**Hudson2014; Hudson2017**), a collection  
566 of studies that recorded species occurrence across multiple land uses and land-use intensities. In PREDICTS,  
567 each study contains several sites, which may be clustered into spatial blocks. Assemblage and land-use data  
568 are available at the site level: one site is characterised by a unique land use of given land-use intensity

- Tropical sites (3914)
- Temperate sites (2373)



**Figure 3.1: Overview of the study design and functional metrics.** I used occurrence data for vertebrate species from the PREDICTS database ((**Hudson2014**; **Hudson2017**); 180 studies; 431,170 records; 4,339 species; 6,758 sampled sites). (a) shows the spatial distribution of sites I consider. I combine occurrence data with trait data compiled in Chapter 2 to calculate functional metrics. (b) is a representation of the trait data in two dimensions, plotted across PREDICTS vertebrates. Traits that contributed most to dimension 1 were lifespan (29%) and litter/clutch size (22%), while traits that contributed most to dimension 2 were habitat breadth (47%) and use of artificial habitats (35%). (c) and (d) present the conceptual framework for the calculation of the functional diversity metrics: local measures (c) and pairwise metrics (d). (c) Given a trait space, functional richness is calculated as the hypervolume occupied by the minimum convex hull encompassing all species **Villeger2008**. Functional dispersion is calculated as the mean distance of the species to the centroid,  $g$  **Laliberte2010**. (d) I compute functional loss as the proportion of hypervolume lost from the reference assemblage, and I define functional gain as the proportion of hypervolume of the disturbed assemblage that was gained (proportion of novel trait space in the disturbed assemblage). *Figure reproduced from Etard2021*.

and provides occurrence data for a set of sampled taxa (and the same set of taxa is sought at all other sites within a study). Sites located between  $23.5^{\circ}\text{N}$  and  $23.5^{\circ}\text{S}$  of latitude were considered tropical, and otherwise temperate (Figure ??).

Land uses in PREDICTS were assigned to the following categories, based on the descriptions of the habitat given by the original collectors of the data: primary vegetation (considered to be the undisturbed reference); secondary vegetation; plantation forest; pasture; cropland; urban (considered human, or disturbed;

575 Table ??; **Hudson2014; Hudson2017**). Secondary vegetation is further divided into three categories: ma-  
576 ture, intermediate and young, depending on the stage of recovery of the vegetation. Land-use intensity is  
577 reported as minimal, light or intense, according to criteria that depended on the land-use type in question  
578 (e.g., crop diversity, degree of mechanisation and chemical inputs in cropland, or bushmeat harvesting and  
579 selective logging in primary vegetation; **Hudson2014**). I excluded sites for which the land use could not be  
580 characterised or for which the stage of recovery of secondary vegetation was unclear. As the PREDICTS  
581 database is a collection of independent studies, the design of this study was not balanced: the sample size  
582 varied across land uses (Figures ??, ??), and across taxonomic groups (3103 species of birds; 531 mammals;  
583 379 amphibians; 326 reptiles).

### 584 3.2.2 Functional traits and diversity indices

585 Trait choice is a critical step when calculating functional diversity metrics, which are highly sensitive to  
586 trait selection (**Mouillot2021**). However, trait selection trades off with data availability. Here, a constraint  
587 was to use similar traits across the different classes. I thus used the seven traits compiled in Chapter 2  
588 across terrestrial vertebrates. Most of these traits were available for at least 50% of the species in each class  
589 (except trophic level in amphibians and lifespan in herptiles; Figure ??). In addition, I chose these traits  
590 as they were ecologically relevant, thus I broadened the biological definition of traits (i.e., a characteristic  
591 measurable at the level of an individual) to include measures of habitat breadth and habitat specialisation (still  
592 theoretically measurable at the level of an individual). The final set constituted seven traits that influence  
593 species responses to environmental change: body mass, trophic level, lifespan, litter/clutch size, diel activity,  
594 habitat breadth and use of artificial habitats. These traits related to life-history, habitat specialisation and  
595 use of geographical space (e.g., habitat breadth is a significant predictor of geographical range size in all  
596 classes; Figure ??). Here, I did not consider estimations of dispersal abilities or home range size as these  
597 were available for a small fraction of the species (<3%, **AlexSmith2005; Paradis1998; Sutherland2000;**  
598 **Whitmee2013a**), neither did I include geographical range size which is measured across many individuals,  
599 and hence cannot be considered a trait. As in Chapter 2, I did not consider intraspecific trait variation, thus  
600 assuming no effect of the environment on trait values.

601 Trait coverage was variable among classes and traits, with important gaps for reptiles and amphibians  
602 (Figures ??, ??; Chapter 2; **Etard2020**). I imputed missing trait values using random forest algorithms  
603 (missForest package: **Stekhoven2012, Stekhoven2016**), including traits, taxonomic order and phylogenetic  
604 eigenvectors as predictors (**Debastiani2021; Penone2014**). To further assess the sensitivity of the results  
605 to imputation (see next section), I imputed missing trait values eight times, thereby obtaining eight sets of

606 imputed traits. I randomly selected one imputed trait set for the calculation of functional metrics. Imputations  
607 of missing trait values & imputation performance are detailed in Appendix S3.2 and Appendix S3.4 (and see  
608 Figures ??-??). Post-imputation, continuous traits were log10-transformed (except habitat breadth which  
609 was square-rooted) and z-scored (standardised to unit variance and zero mean). In addition, I also assessed  
610 whether the results were robust to imputation error using a subset of the PREDICTS data considering only  
611 species for which I had complete trait information (see next section).

612 Correlation among traits can be a safeguard against high sensitivity of functional metrics to trait omis-  
613 sion, notably where omitted traits correlate strongly with traits that are already included in the calculation  
614 (**Mouillot2021**). Nevertheless, high multicollinearity among traits has been reported as potentially prob-  
615 lematic for the calculation of functional diversity (**Cadotte2011**). Thus, I verified that the degree of multi-  
616 collinearity among traits was not problematically high (with a threshold of 5 for variance inflation factors,  
617 Table ??). Furthermore, I tested the sensitivity of the results to trait omission, by investigating whether  
618 adding geographical range size in the calculation of functional metrics was likely to affect the results.

### 619 3.2.3 Effects of land use and land-use intensity on FRic and FDis (Hypothesis 1)

620 For each assemblage, I measured functional richness using ‘FRic’ (**Villeger2008**), and functional dispersion  
621 using ‘FDis’ (**Laliberte2010**; Figure ??), from the FD package (**Laliberte2010**; **Laliberte2015**). I assessed  
622 the effects of land use, land-use intensity, and region (temperate versus tropical) on FRic and FDis across  
623 and within taxonomic classes using linear mixed-effects models (lme4 package, **Bates2015**). Land use  
624 and land-use intensity were not ranked in the models. A random intercept of study identity accounted for  
625 variation in experimental design across studies, while a random intercept representing spatial blocks of  
626 sampled sites, nested within study, accounted for spatial structuring within studies. To improve normality and  
627 bound predictions between 0 and 1, I transformed FRic and FDis using an arcsin-square-root transformation.  
628 The best-fitting model was sought using backwards stepwise model selection, starting with the most complex  
629 model that included all two-way interactions among the specified main effects. Model fits were compared  
630 using likelihood-ratio tests at each iteration of the selection procedure.

631 Across vertebrates, the starting models included the effects of land use, land-use intensity and region  
632 (temperate versus tropical). The best-fitting model for FRic was:

$$633 \text{arcsin}(\sqrt{\text{FRic}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Land use : Land-use intensity} + \text{Land use : Region}$$

634

636 For FDis, the best-fitting model did not include interactions between land use and region, but the main  
637 effect of region was retained:

638  $\text{arcsin}(\sqrt{\text{FDis}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Land use : Land-use intensity}.$

639 (Model 1b)

640 To investigate differences in responses across classes, I pooled some of the land uses together, because  
641 otherwise, sample sizes would have been too low. Mature, intermediate and young secondary vegetation were  
642 grouped together as ‘Secondary vegetation’, and cropland and pasture were grouped together as ‘Agricultural  
643 land uses’. The starting models included the effects of land use, land-use intensity, region and taxonomic  
644 class. For FRic, the best model was:

645  $\text{arcsin}(\sqrt{\text{FRic}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Class} + \text{Land use :}$

646  $\text{Land-use intensity} + \text{Land use : Class} + \text{Land-use intensity : Region} + \text{Class : Region}.$

647 (Model 2a)

648 For FDis, regional effects were dropped:

649  $\text{arcsin}(\sqrt{\text{FDis}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Class} + \text{Land use : Land-use intensity} + \text{Land use :}$   
650  $\text{Class} + \text{Land-use intensity : Class}.$

651 (Model 2b)

652 To assess whether the results were robust to imputation error, I used a subset of the PREDICTS data  
653 considering only species for which there were complete trait information (6,212 sites; 442 mammals; 1,975  
654 birds; 78 reptiles; 9 amphibians), and I fitted models again to this data subset. I did not have enough complete  
655 trait data among amphibians to be able to consider this class separately, so I first considered amphibians and  
656 reptiles together (herptiles), and reptiles only. In addition, I complemented this validation with a sensitivity  
657 analysis to variation in imputed values. I calculated FDis and FRic using each of the eight imputed trait  
658 datasets and fitted the previous models to each set. I then qualitatively evaluated the congruence of the  
659 estimates from the different models. Finally, because there tended to be more sites sampled in primary  
660 vegetation than in other land uses (Figures ??, ??), I ran additional sensitivity tests to assess whether the  
661 results were robust to resampling primary vegetation sites to a number equal to 50 (a sample size close to the  
662 median number of sites sampled in land uses other than primary vegetation in both regions (median = 37 for  
663 the temperate subset and 57 for the tropical subset, Figure ??)).

664 **3.2.4 Investigating functional under-dispersion (Hypothesis 2)**

665 To assess whether effects of land use and land-use intensity on FDis differed from what would be expected  
666 by chance given changes in local species richness, I generated null expectations of FDis at each site. I  
667 randomised assemblage composition 500 times, drawing species from the corresponding study's species  
668 pool while maintaining local species richness. For each site, I thus obtained a null distribution for FDis.  
669 Then, I tested whether FDis differed from null expectations using Wilcoxon signed-rank tests. I created  
670 a binary variable which was assigned 1 if FDis was significantly lower than null expectations at a given  
671 site (significant under-dispersion), and 0 otherwise. I investigated how land use, land-use intensity, region  
672 and taxonomic class affected the probability of occurrence of under-dispersion using a generalised linear  
673 mixed-effects model with a binomial distribution of errors. The best-fitting model did not retain any effect  
674 of taxonomic class:

675  $P_{\text{under-dispersion}} \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Land use : Land-use intensity} + \text{Land use :}$   
676  $\text{Region.}$

677 (Model 3)

678 **3.2.5 Functional loss and functional gain (Hypothesis 3)**

679 I calculated the proportion of trait space that was lost in disturbed land uses compared to reference land uses  
680 (functional loss) and the proportion of trait space that was gained in disturbed land uses (functional gain)  
681 (Figure ??(c)), across and within taxonomic classes. I selected studies where at least one site was sampled in  
682 primary vegetation. I then made within study pairwise comparisons between reference assemblages, sampled  
683 in primary vegetation, and disturbed assemblages. In addition, I considered all comparisons between pairs  
684 of primary-vegetation sites, to create reference pairs. I then investigated how land use, land-use intensity  
685 and region affected functional loss and gain across and within taxonomic classes using linear mixed-effects  
686 models, controlling for study identity in the random effects. Across vertebrates, the best-fitting model for  
687 functional loss was:

688  $\text{arcsin}(\sqrt{\text{loss}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Land use : Land-use intensity} + \text{Land use :}$   
689  $\text{Region.}$

690 (Model 4a)

691 For functional gain, one interaction term (land use with region) was dropped:

692  $\arcsin(\sqrt{\text{gain}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Land use : Land-use intensity}.$

693 (Model 4b)

694 When considering the effects of taxonomic class, the best-fitting model for functional loss was:

695  $\arcsin(\sqrt{\text{loss}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Class} + \text{Region} + \text{Land use :}$

696  $\text{Land-use intensity} + \text{Land use : Class} + \text{Land use : Region} + \text{Land-use intensity : Class}.$

697 (Model 5a)

698 For functional gain (Model 5b), the fitted effects were the same as those of Model 2b. More details about  
699 the calculation of functional loss and gain can be found in Appendix S3.5.

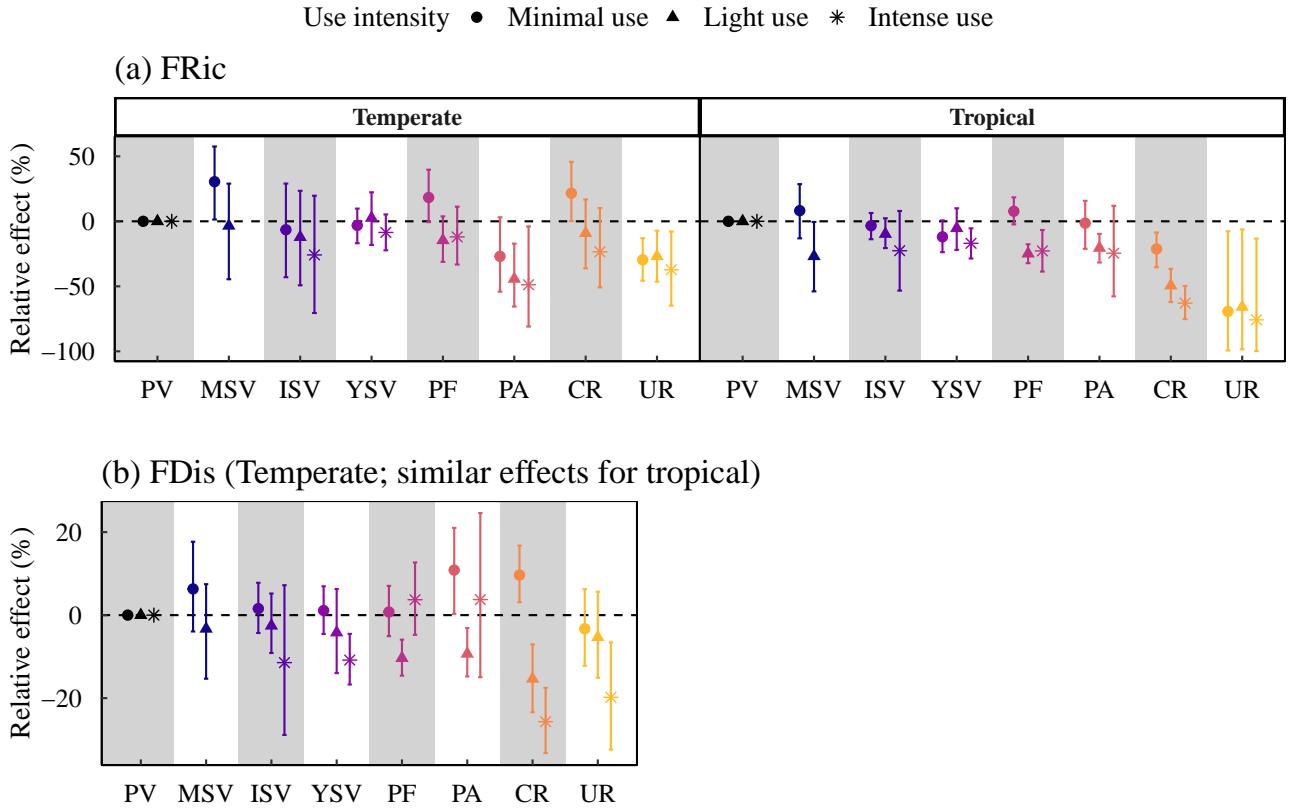
700 All data analyses were conducted using R version 3.5.1 (R Core Team, 2018). I made the code available  
701 on figshare (DOIs: <https://doi.org/10.6084/m9.figshare.14161883> and <https://doi.org/10.6084/m9.figshare.15163926>), as well as the main result datasets (<https://doi.org/10.6084/m9.figshare.15163971>).

### 704 3.3 Results

#### 705 3.3.1 Effects of land use on FRic and FDis

706 Across all vertebrates, land use and land-use intensity significantly affected FRic and FDis (Figure ??). FRic  
707 tended to decrease with increasing disturbance level and higher intensity of land use. For FRic, relative ef-  
708 fects differed between regions (Figure ??(a)). Although declines were overall more important for disturbed  
709 tropical assemblages, significant declines were observed for the temperate assemblages (e.g., a 37% average  
710 decline in intensely used urban areas; a 49% decline in pastoral areas of high land-use intensity). Never-  
711 theless, tropical assemblages typically showed more important reductions in FRic. For instance, declines  
712 averaged 63% for intensely used tropical cropland and 76% for urban areas. For FDis, relative effects were  
713 similar in both regions (Figure ??(b)). The most important average declines were observed for urban assem-  
714 blages of intense use (20% decline), and for lightly- and intensely used cropland (by 15% and 26%). Note  
715 that confidence intervals around the estimated average declines were large in some cases, highlighting some  
716 heterogeneity in the responses.

717 Fitting the same models to the subset of species with complete trait data, I detected important declines in  
718 functional diversity in a number of land uses, showing that the conclusions are robust to trait imputation un-  
719 certainty (for example, FRic declined on average by 75% in intensely used temperate pastoral assemblages;  
720 by 48% for intensely used tropical cropland; and FDis declined by an average 37% in intensely used tropical

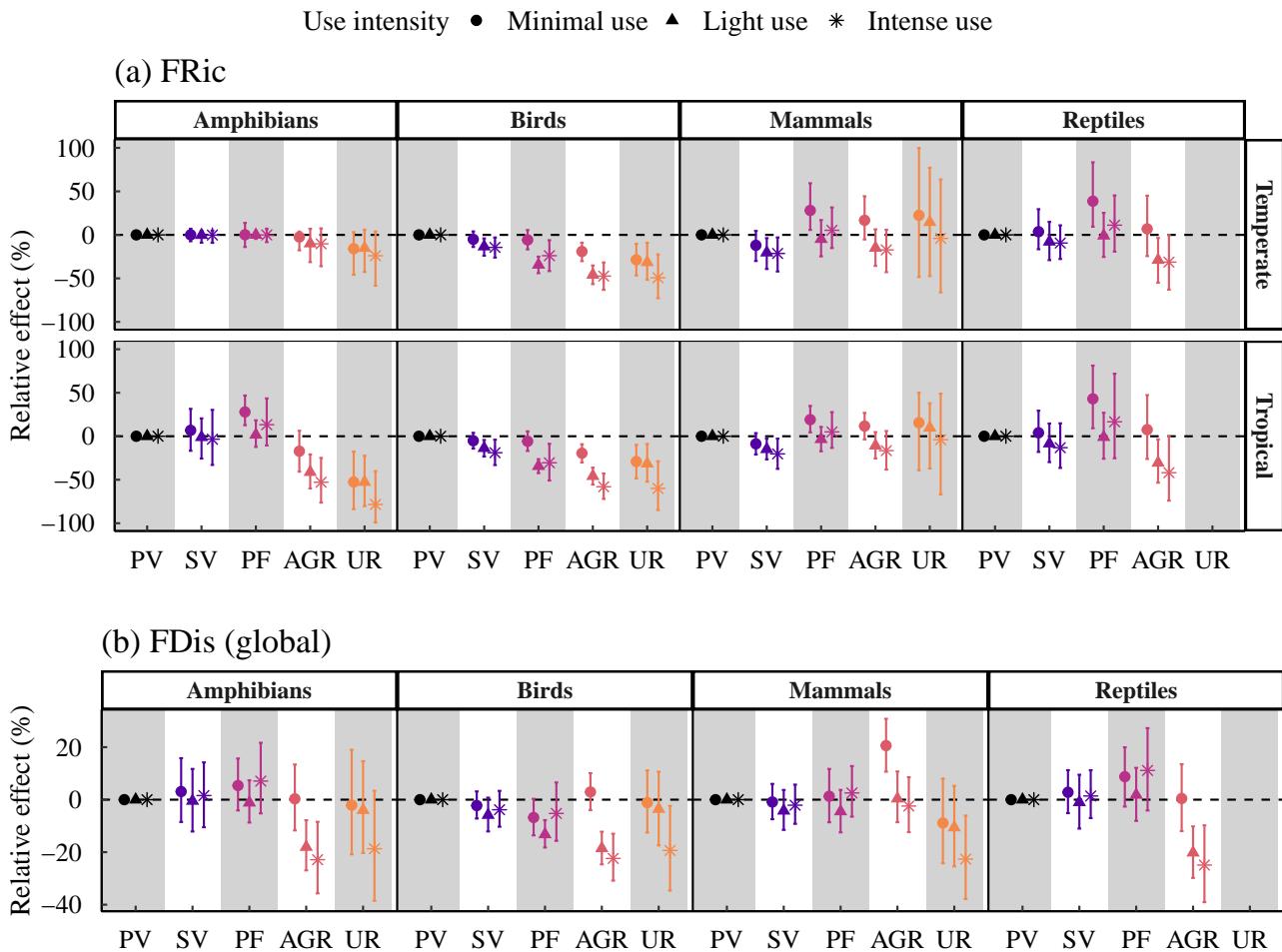


**Figure 3.2: Effects of land use, land-use intensity and region on FRic (a) and on FDis (b) across all PREDICTS vertebrates.** Effects were rescaled with reference to primary vegetation (PV) and are plotted as a % difference relative to PV within each land-use intensity. For FRic, the best-fitting model included interactions between land use and region, while these interactions were dropped for FDis, explaining the similar relative effects in both regions. Error bars represent 95% confidence intervals. MSV: mature secondary vegetation; ISV, intermediate secondary vegetation; YSV, young secondary vegetation; PF, plantation forest; PA, pasture; CR, cropland; UR, urban. Effects for intense use in MSV could not be estimated as there were not enough sampled sites. *Figure reproduced from Etard2021.*

721 urban assemblages; Figure ??). Furthermore, using the subset of species with complete trait data, I found  
 722 that the results were not sensitive to the inclusion of geographical range size as an additional trait (Figure  
 723 ??). Finally, the results were not sensitive to variation across imputed trait values (Figure ??) and were also  
 724 robust to resampling in primary-vegetation sites (Figure ??).

725 Responses of FRic and FDis to land use and land-use intensity differed among taxonomic classes (Figure  
 726 ??). Within-class effects for FDis were similar between regions. The most notable decreases were observed  
 727 in lightly- and intensely used agricultural land uses in amphibians, birds and reptiles; and in intensely used  
 728 urban land uses for birds and mammals. For FRic, the effects in tropical and temperate regions were qual-  
 729 itatively similar in three out of four classes (birds, mammals and reptiles), although effect sizes tended to be  
 730 bigger for tropical assemblages. Birds and reptiles showed reductions in disturbed land uses in both tropical  
 731 and temperate regions, whereas I detected few significant effects for mammals. For birds, the most important

average decline, of 50%, was observed in intensely used tropical urban land uses, while for reptiles I detected significant decreases in lightly- and intensely used agricultural sites (but I could not estimate effects for urban land uses due to the small sample size). Finally, the effects differed between tropical and temperate regions for amphibians, with no significant effects detected across temperate assemblages, but important reductions across tropical agricultural and urban assemblages.



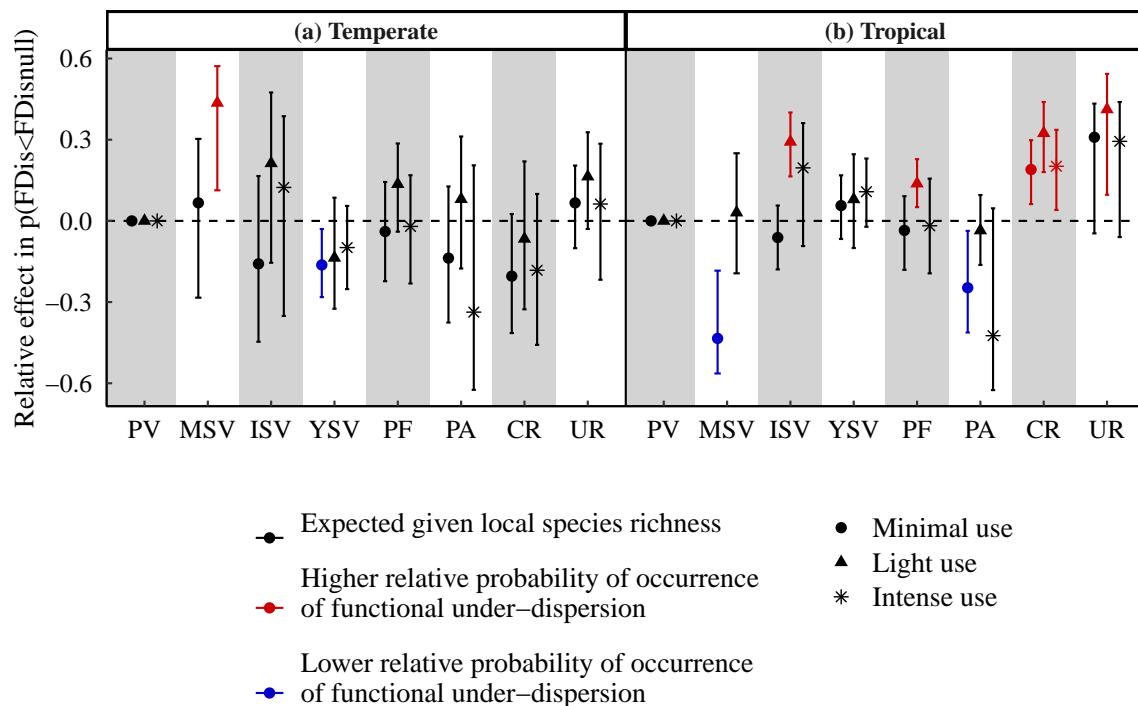
**Figure 3.3: Effects of land use, land-use intensity, region and taxonomic class on FRic (a) and FDis (b).** Effects were rescaled with reference to primary vegetation (PV) and are plotted as a % difference relative to PV within each land-use intensity. Error bars represent 95% confidence intervals. Effects for FRic were estimated from Model 2a, and from Model 2b for FDis. SV, secondary vegetation; PF, plantation forest; AGR, agricultural land uses (pasture and cropland); UR, urban. Effects for reptiles in urban land uses could not be estimated as there were not enough sampled sites. *Figure reproduced from Etard2021.*

Fitting similar models only for species with complete trait data showed that these patterns are unlikely to be affected by imputation uncertainty for birds; for mammals and reptiles, the main results could even be conservative (Figure ??, ??). Indeed, although confidence intervals around the estimates were large, I typically observed larger decreases in functional diversity when using the complete data subset, including

741 an 86% decline in FRic for mammals in intensely used tropical agricultural areas. The results were also  
 742 unaffected by variation across replicate sets of imputed trait values (Figure ??).

### 743 3.3.2 Changes in the probability of occurrence of functional under-dispersion

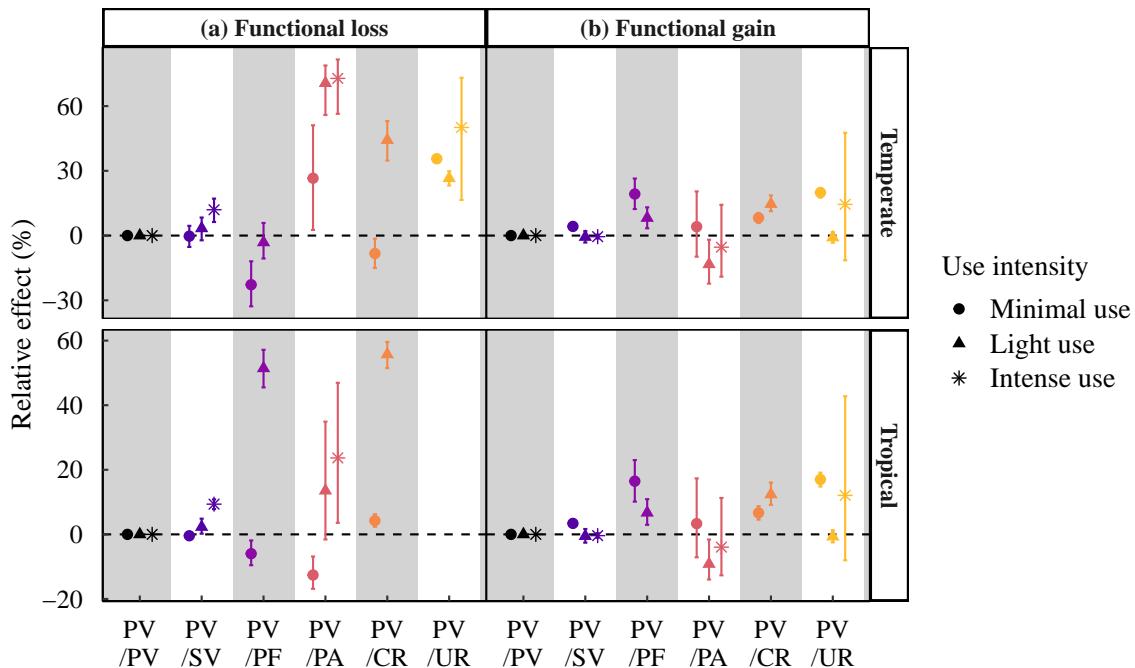
744 Land use, land-use intensity and region significantly affected the probability of occurrence of functional  
 745 under-dispersion across vertebrates. Functional under-dispersion was more likely to occur in tropical crop-  
 746 land of all land-use intensities (Figure ??(b)), as well as in some of the lightly-used land uses (notably urban  
 747 and plantation forest). Contrary to my expectations, and with the exception of tropical cropland, functional  
 748 under-dispersion was not more likely to occur in intensely-used land uses. For minimally-used sites, changes  
 749 in FDis were mostly consistent with that expected given changes in species richness.



**Figure 3.4: Effects of land use, land-use intensity and region on the probability of occurrence of functional under-dispersion.** Error bars represent 95% confidence intervals. PV: primary vegetation; MSV, mature secondary vegetation; ISV, intermediate secondary vegetation; YSV, young secondary vegetation; PF, plantation forest; PA, pasture; CR, cropland; UR, urban. Effects are rescaled and represent the average difference in the probability of occurrence of functional under-dispersion between the reference (PV, probability of functional under-dispersion set at 0 within each land-use intensity) and the disturbed land uses. *Figure reproduced from Etard2021.*

### 750 3.3.3 Functional loss and gain

751 Across and within vertebrate classes, I detected high levels of functional loss, exceeding the natural turnover  
 752 between primary-vegetation sites, both in temperate and tropical regions. Across vertebrates (Figure ??(a)),  
 753 functional loss was notably high in temperate pastures (+27% above reference for minimal use; +73% for  
 754 intense use), temperate urban sites (+27% for light use; +50% for intense use; effects for tropical urban sites  
 755 could not be estimated), temperate and tropical cropland (+44% and +56% respectively for light use; effects  
 756 for intense use could not be estimated). Important levels of functional loss were also observed in tropical  
 757 plantation forest of light use intensity (+51%; effects for the intense use could not be estimated). High levels  
 758 of functional loss were also observed within each class (Figure ??(a)) (although not all effects could be  
 759 estimated because of limited sample sizes, Table ??). The highest losses were observed in agricultural areas  
 760 for amphibians and reptiles, with important losses also observed in temperate urban areas for both birds and  
 761 amphibians (+35% for minimal use; effects for tropical urban areas could not be estimated).

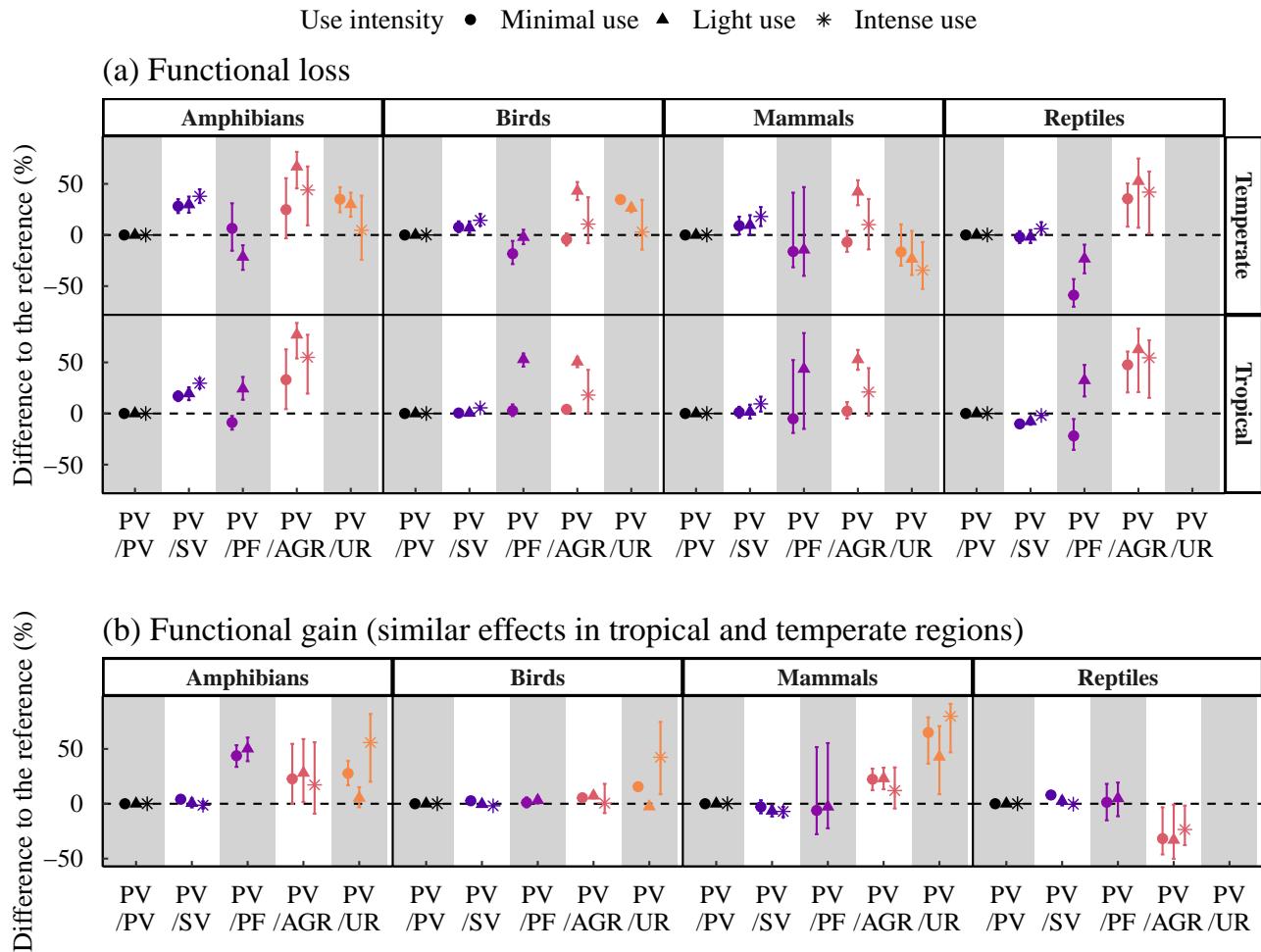


**Figure 3.5: Effects of land use and land-use intensity on (a) functional loss and (b) functional gain across pairs of sites in temperate areas and tropical areas.** PV: primary vegetation; SV, secondary vegetation; PF, plantation forest; PA, pasture; CR, cropland; UR, urban. Error bars represent 95% confidence intervals. Effects are rescaled and represent the average difference in functional loss and gain between the reference pairs (PV/PV, with loss and gain set at 0% within each land-use intensity), and pairs of primary vegetation with disturbed land uses. Negative effects mean that, on average, levels of functional loss (or gain) were lower than functional loss (or gain) observed between pairs of primary-vegetation sites, and not that there were absolute ‘negative losses’ or ‘negative gains’. *Figure reproduced from Etard2021.*

762 Across vertebrates (Figure ??(b)), average functional gain (average proportion of novel trait space in

the disturbed assemblage) was moderate and on average did not exceed 20% in any disturbed land uses. Patterns of functional gain were similar in both regions. The highest functional gains were observed for minimally-used urban sites and plantation forest (range: +16% to +20%). On the other hand, important levels of functional gain were observed in some classes (Figure ??(b)), with the highest functional gain for mammals (+80% in intensely used urban sites).

Diagnostic plots (qq-plots and residual distributions) for the models are shown in Figures ??–???. Overall, the model residuals were appropriately distributed (but with some leptokurtic residual distributions, to which mixed-effect models are generally robust (**Schielzeth2020**)).



**Figure 3.6: Effects of land use, land-use intensity, region and taxonomic class on functional loss and functional gain across pairs of sites.** PV: primary vegetation; SV, secondary vegetation; PF, plantation forest; AGR, agricultural land uses (pasture and cropland); UR, urban. Error bars represent 95% confidence intervals. Effects are rescaled and represent the average difference in functional loss and gain between the reference pairs (PV/PV, with loss and gain set at 0% within each land-use intensity), and pairs of primary vegetation with disturbed land uses. Negative effects mean that, on average, levels of functional loss (or gain) were lower than functional loss (or gain) observed between pairs of primary-vegetation sites, and not that there were absolute ‘negative losses’ or ‘negative gains’. *Figure reproduced from Etard2021.*

771 **3.4 Discussion**

772 Here, I showed that the functional diversity of vertebrate assemblages is negatively impacted in human land  
773 uses, particularly in the most intensely used land types. The results of this Chapter extend previous studies  
774 that have been more taxonomically or geographically restricted (**Flynn2009; Matuoka2020**). **Matuoka2020**  
775 found that the functional diversity of tropical bird assemblages was negatively affected by human distur-  
776 bance, a pattern that did not appear in temperate assemblages. Yet, I found that functional diversity was  
777 negatively affected in both tropical and temperate, with important functional losses in all four vertebrate  
778 classes.

779 Using multiple metrics allowed me to explore different facets of functional diversity. For instance, func-  
780 tional gain could locally offset functional loss in some disturbed land uses. This could indicate that despite  
781 no apparent negative effect on FRic, some disturbed land uses (e.g. lightly-used temperate cropland) could  
782 experience important functional loss, and highlights the importance of using a variety of indicators. This  
783 mechanism could be at play in mammalian assemblages, for which important levels of functional gain were  
784 observed in agricultural and urban sites. Further, functional gain in disturbed land uses could indicate that  
785 disturbances facilitate the introduction of functionally novel species, falling into previously unoccupied parts  
786 of the trait space. This may be because non-native species are more likely to become established in disturbed  
787 assemblages. Previous work has shown that land-use disturbance facilitates biological invasions in island  
788 ecosystems (**Jesse2018; Sanchez-Ortiz2019**), but to my knowledge, this has not been tested specifically  
789 across continental areas for invasive vertebrates (but see **Pysek2010**). It is also possible that disturbed areas  
790 harbour synanthropic species that do not occur in primary vegetation, leading to substantial functional gain.

791 Overall, the negative effects of land use on functional richness tended to be more pronounced in the trop-  
792 ics. This is congruent with past studies that have found tropical biodiversity to be disproportionately sensitive  
793 to human pressures (**Newbold2020; Martins2017**). There are a number of potential explanations for this.  
794 First, it could be that a long history of intense land-use disturbance at large scales in many temperate regions  
795 (e.g. Western Europe; **Stephens2019**) means that biodiversity is now less sensitive to new disturbances,  
796 because the most sensitive species have been filtered out (**Balmford1996; Krauss2010; LeProvost2020;**  
797 **Munteanu2020**). Species unable to cope with such disturbances may have gone extinct in the past, while  
798 the remaining species would be more disturbance-tolerant (**Betts2019**). Tropical regions, historically less  
799 disturbed at large scales, would then contain a higher proportion of disturbance-sensitive species than tem-  
800 perate regions. Consequently, the functional richness in undisturbed tropical sites could be less resilient to  
801 new disturbances. This also highlights that time since land-use conversion could have important impacts

802 on local functional diversity. Although I did not consider the effects of time since land-use conversion in  
803 this work (notably because PREDICTS contained data only for about 22% of the sites), I expect that time  
804 since land-use conversion may affect assemblage composition, and thus, functional diversity, with poten-  
805 tially land-use-specific relationships between time since conversion and functional diversity (e.g., a positive  
806 relationship for recovering secondary vegetation or a negative relationship for urban areas; but I did not de-  
807 tect such effects when using the data subset for which I have information on time since land-use conversion  
808 [see Appendix 3, section S3.8]).

809 Second, it could be that tropical species are intrinsically more sensitive to disturbances than temperate  
810 species because of their evolutionary history. Natural climatic variability experienced by species as well as  
811 species history of exposure to disturbances have been proposed to influence sensitivity to disturbance. For in-  
812 stance, tropical species are, on average, nearer to their climatic limits than temperate species (**Deutsch2008**;  
813 **Sunday2014a**). Tropical species could therefore experience more deleterious effects from interacting drivers  
814 of change, with land-use change bringing about novel climatic conditions pushing them beyond their toler-  
815 ance limits (**Frishkoff2016**; **Williams2020a**).

816 In addition to filtering out sensitive species, land-use change is also expected to modify interactions  
817 among species, thereby influencing species persistence (**Tylianakis2008**; **Valiente-Banuet2015**). Although  
818 I detected a signal of functional under-dispersion (particularly in tropical cropland), which indicates that as-  
819 semblages may be locally structured by environment filtering (**Bregman2015**), it is likely that several assem-  
820 bly rules underpin assemblage composition (**Fournier2016**). For instance, land-use changes could enhance  
821 competition among species, promoting over-dispersion by removing species that share similar resources.  
822 Such opposite signatures of environmental filtering and enhanced competition on functional dispersion could  
823 explain why I did not detect stronger effects of land use on functional under-dispersion occurrence.

824 Studies looking at impacts of global land use on functional diversity computed with species from all  
825 four terrestrial vertebrate classes remain rare. Lack of availability of standardised trait data across terrestrial  
826 vertebrates may have hindered such studies from being conducted in the past. To overcome this problem,  
827 I based the analyses on a large-scale collation of trait data (Chapter 2; **Etard2020**), and I imputed missing  
828 trait values to obtain complete trait datasets in each class. I used random forest algorithms, currently thought  
829 to be one of the most robust technique for missing value imputations in trait datasets (**Debastiani2021**;  
830 **Johnson2021**; **Penone2014**). Replicating the analyses on complete trait data subsets showed that imputation  
831 uncertainty did not affect the main conclusions of this work and that the negative effects of human land uses  
832 were in some cases even stronger when using the complete data subsets. Furthermore, the results were highly  
833 consistent across imputed datasets and so insensitive to variation across imputed values. Although missing

834 value imputation can offer a robust filling of missing entries, this study highlights the existing taxonomic  
835 biases both in trait data availability and in PREDICTS studies, and thus stresses the need to pursue data  
836 compilation efforts, particularly for the least-sampled classes (reptiles and amphibians).

837 Another implication of trait data availability for vertebrates is that the choice of traits was constrained.  
838 **Mouillot2021** showed that functional diversity metrics are sensitive to trait omission and that the sensitivity  
839 to trait omission decreases with increasing levels of correlation among traits. Here, I chose seven traits that  
840 were available across all classes at least for a subset of the species and that have been implicated in shaping  
841 species responses to environmental change. A notable omission was any metric of dispersal ability, which is  
842 likely to influence species' ability to respond to land-use change but is difficult to obtain for most species.  
843 In fact, past studies have shown that dispersal abilities can be predicted from ecological correlates, such as  
844 body mass, diet or geographical range size (**Schloss2012b; Sutherland2000**). Since the results were robust  
845 to the omission of geographical range size, I am confident that the omission of dispersal abilities also does  
846 not affect the conclusions of this work.

847 Functional diversity metrics are often used as a proxy for ecosystem functioning because of the con-  
848 ceptual and mechanistic link between functional 'effect' traits and ecosystem processes (**Lavorel2002a;**  
849 **Violle2007**). In many studies focused on vertebrates, however, functional diversity metrics do not correlate  
850 with a given ecosystem function (**Hatfield2018**). Here, I did not explicitly target given ecosystem functions,  
851 but I argue that evidence of functional loss of vertebrate assemblages indicates that processes sustained by  
852 vertebrates are put at risk by land-use change. My results further show that some disturbed land uses are  
853 more likely to experience functional under-dispersion, particularly tropical cropland and tropical urban ar-  
854 eas, which again indicates a potential imperilment of ecological processes. Indeed, in such cases, decreases  
855 in functional dispersion exceed changes expected from the chance removal of species; such non-random  
856 modifications indicate that certain areas of the functional trait space are more sensitive to land-use distur-  
857 bance. Future work could investigate the impacts of land-use change on particular ecosystem functions. The  
858 integration of trophic information (beyond the trophic levels I used here) to the species-trait dataset could  
859 be an interesting step in that direction, as dietary traits relate to resource use and are, as such, probably  
860 the most straightforward traits to link with ecosystem functions. Furthermore, my results suggest that the  
861 functional loss experienced within a class is unlikely to be compensated for by the persistence of function-  
862 ally similar species in other classes. Indeed, I detected negative effects of human land use on functional  
863 richness in at least three out of four vertebrate classes (amphibians, birds, and reptiles), in accordance with  
864 past studies focusing on each of these groups (**Gallmetzer2015; Marcacci2021; Riemann2017; Sol2020**).  
865 Although overall mammalian functional richness was less affected, high levels of functional gain suggest

866 that the functional composition of mammalian assemblages is heavily modified in disturbed land uses.

867 To conclude, the results of this third chapter highlight the negative impacts of human land uses on mul-  
868 tiple dimensions of functional diversity, within and across terrestrial vertebrate classes, at a global scale.

869 In many disturbed sites, decreases in functional diversity exceed changes expected from species loss alone,  
870 showing that human activities non-randomly reshape ecological assemblages. By intensifying functional  
871 loss and promoting functional under-dispersion, land-use change could have deleterious effects on ecosys-  
872 tem functioning, highlighting the necessity of putting into place effective conservation measures in the face  
873 of anthropogenic change.

874 **4 | Geographical range area, habitat breadth**

875 **and specialisation on natural habitats**

876 **predict land-use responses and climate-**

877 **change sensitivity more consistently than**

878 **life-history and dietary traits in terres-**

879 **trial vertebrates**

880 **Keywords**

881 **Abstract**

882 **4.1 Introduction**

883 **4.2 Methods**

884 **4.3 Results**

885 **4.4 Discussion**

886 **5 | Energetic constraints and trophic group**

887 **explain species persistence in disturbed**

888 **land uses**

889 **Keywords**

890 Land use; land-use intensity; metabolic rates; energetic constraints; energetic requirements; terrestrial verte-  
891 brates; trophic group; occurrence.

892 **Abstract**

893 Land-use change is the primary driver of global biodiversity loss. In terrestrial vertebrates, previous work  
894 has shown that sensitivity to land-use change depends on species traits, but the extent to which energetic  
895 constraints explain species responses to disturbed land uses remains largely unexplored. Here, I investigate  
896 relationships between the energetic requirements of terrestrial vertebrates (estimated from resting metabolic  
897 rates) and land-use change, at two levels of organisation. First, at the assemblage level I hypothesize that  
898 total energetic requirements in disturbed land uses are lower than in undisturbed land uses, assuming that  
899 there is less energy available in these areas overall. Second, after controlling for the effects of body mass and  
900 taxonomy on metabolic rates, I predict that species with relatively lower energetic expenditure are favoured  
901 over species with relatively higher energetic expenditure in disturbed land uses, as resource efficiency will be  
902 beneficial in these resource-poor environments. Because trophic group influences species ability to assimili-  
903 late various types of food, I investigate whether my predictions are consistent among trophic groups (here,  
904 omnivores, carnivores or herbivores). The results challenged both hypotheses. I found that total assemblage-  
905 level energetic requirements did not systematically decrease in disturbed land uses. For instance, I detected

906 significant increases for urban areas in all trophic groups, highlighting that disturbed areas may not be as  
907 energy-poor as I initially assumed. Second, I found a positive effect of metabolic rates (after controlling  
908 for body mass and taxonomy) on species probability of occurrence across all trophic groups for at least one  
909 of the most disturbed land uses I considered (pasture, cropland and urban). Species for which there are  
910 exploitable resources in disturbed environments may benefit from having larger energetic expenditure: they  
911 may display a set of characteristics rendering them more able to cope with disturbances and more able to  
912 acquire available resources, such as higher activity levels or bigger brain sizes. The findings of this Chapter  
913 highlight that land-use change has substantial impacts on vertebrate community metabolism.

## 914 5.1 Introduction

915 Land-use change is currently the strongest driver of global biodiversity declines (**Newbold2015; Maxwell2016**),  
916 with major and long-lasting impacts on the structure and functioning of ecological communities (**Bregman2016;**  
917 **Fukasawa2019; Magioli2021; Marcacci2021**). With land-use change likely to continue to intensify (**Stehfest2019**),  
918 it is vital to put into place conservation and mitigation measures to minimise future losses of biodiversity  
919 and negative impacts on ecosystem functioning. To this end, pressing questions remain as to what renders  
920 species able or unable to cope with human disturbance, and how losses of sensitive species might influence  
921 ecosystem functioning (**Dirzo2014; Young2016**).

922 Land-use change acts as an environmental filter affecting species persistence (**Evans2018; Edwards2021**).  
923 Past studies have shown that sensitivity to land-use change is distributed unevenly across the tree of life  
924 (**Nowakowski2018a**), and across behavioural (**Lowry2013; Samia2015**) and ecological strategies (**Newbold2013;**  
925 **Moller2009; DePalma2015**). For instance, long-lived and large forest specialist birds respond more neg-  
926 atively to land-use change than generalist species (**Newbold2013**). In addition, human land uses impose  
927 energetic constraints on species, by modifying the amount and type of available resources (**Inger2016;**  
928 **ZunigaPalacios2021**). Consequently, land-use change brings about modifications in dietary and foraging  
929 strategies (**Moller2009; Ramesh2017; Seveque2020**), which can have cascading effects on local trophic  
930 structure (e.g., mesopredator release; **Crooks2010**). The impacts of land-use change on biodiversity also de-  
931 pend on land-use intensity (**Newbold2015; Davison2021**), that is, the magnitude of human disturbance in a  
932 given land-use type, which can vary importantly according to local practices and management (**Dullinger2021**).  
933 In particular, land-use intensity affects the types and availability of resources (by homogenising resources  
934 for example), which in turn may affect biodiversity patterns (**Weiner2011**).

935 Energetic requirements are likely to be important predictors of species' ability to cope with a disturbed

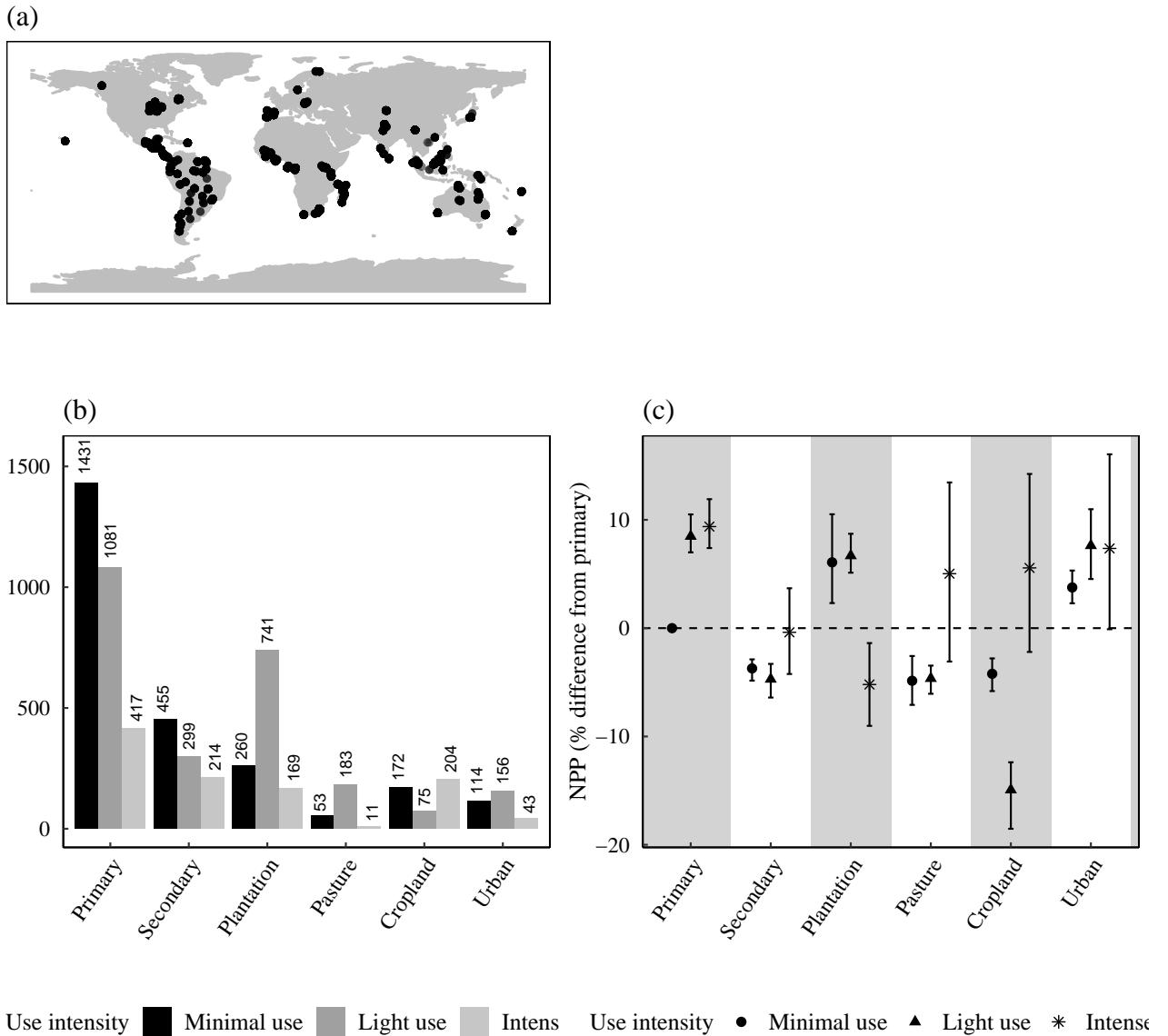
environment where resource availability and resource types are strongly impacted (**Auer2020**). Past research has shown that metabolic rates (the amount of energy required for organismal maintenance and for other processes such as growth and reproduction) are heritable and subject to selection, for instance with species in energy-rich environments having evolved faster metabolic rates than species in less energy-rich environments (**Mueller2001**). However, metabolic rates have received less attention than other life-history traits in studies aiming to understand different species' responses to land-use change. Body mass, known to explain much of the variation in metabolic rates (**White2003; Hudson2013; Bushuev2018**), has been considered in past studies examining effects of land-use disturbance on assemblage composition (**Hevia2017**). Since metabolic rates correlate positively with body mass, the effects of land-use change on local energetic requirements have been indirectly assessed through investigations of changes in body mass composition along land-use gradients (**Newbold2020a; Tinoco2018**). However, to the best of our knowledge, no study has yet directly reported the effects of land-use change on assemblage-level energetic requirements.

Thus, the first objective of this Chapter is to investigate how land-use change impacts assemblage-level energetic requirements in terrestrial vertebrates. To this end, I use metabolic rates as a proxy for maintenance energetic requirements at the species level. I predict that, at the assemblage level, total energetic requirements in disturbed land uses are lower than in undisturbed land uses. This prediction relies on the assumption that there is less energy available overall in disturbed land uses than in undisturbed land uses (Figure ??c), with disturbed land uses being characterised by human appropriation of net primary productivity (**Krausmann2013**). Because species' ability to exploit particular resources and meet energetic demands also depends on food types and diet (**Mendoza2019; McNab1986**), I test whether this prediction holds true within trophic groups (i.e., within herbivores, omnivores, and carnivores). Here, I consider energetic requirements at the assemblage level, not controlling for species body mass, to obtain estimates of total energy expenditure in different land-use types (Figure ??a). Thus, any changes in assemblage-level energetic requirements could be ultimately driven by shifts in the size-spectrum of local assemblages.

Body mass explains much of the variation in metabolic rates, but species with similar body masses still display important variation in metabolic rates (**Mueller2001**). Some of this variation can be explained by taxonomic position (which also accounts for differences in thermoregulatory strategy between endotherms and ectotherms). However, there remains substantial interspecific variation in metabolic rates even after accounting for body mass and taxonomic affiliation (**Mueller2001; White2013**). Food availability and food type might explain this residual variation in metabolic rates among species (**Mueller2001**). Species that evolved in areas with abundant resources assimilate and burn calories at high rates, regardless of their body mass – while species that evolved with limiting resources could process resources at lower rates, irrespectively

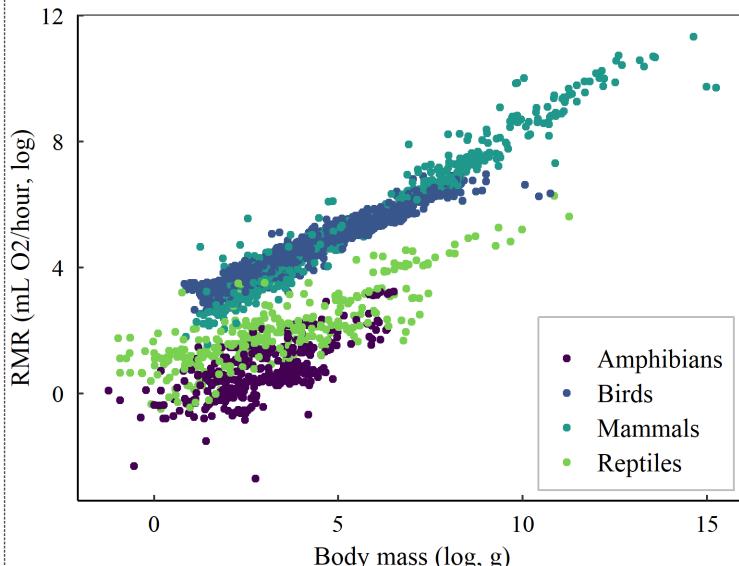
of their body mass (**Mueller2001**). Building upon this idea and the ‘allocation’ principle (i.e., the trade-offs in resource allocation; **Auer2020**), species with lower energetic expenditure (than expected from body mass and taxonomy) may be able to cope with food scarcity better than those with larger energy expenditure (than expected from body mass and taxonomy). Indeed, regardless of their body mass, such species should need fewer resources to meet energetic demands (**Clarke2004a**). Thus, as land-use disturbance modifies the types and abundance of resources, we expect species energetic requirements –after removing the effects of body mass and taxonomic position– to influence responses to land-use change. Yet, whether energetic constraints explain species responses to land-use change in terrestrial vertebrates has not been investigated before (**Hevia2017**). Thus, the second objective of this Chapter is to characterise the influence of energetic constraints on species occurrence probability in disturbed land uses, compared to undisturbed land uses. I approximate energetic constraints at the species level from residual variation in metabolic rates, that is, the variation not explained by body mass or taxonomy. I predict that species with negative deviations (lower rates than expected from body mass and taxonomy) are more likely to persist in disturbed land uses than species with positive deviations (higher rates than expected from body mass and taxonomy; Figure ??b). Given that trophic group can influence resource acquisition and species ability to assimilate various types of food, I investigate whether this prediction is consistent among trophic groups.

To test both predictions, I use a space-for-time substitution approach, obtaining vertebrate assemblage composition data from the PREDICTS database (**Hudson2014; Hudson2017**). I combine this database with trophic-group information (characterising species as either omnivores, herbivores, or carnivores), and with species-level estimates of resting metabolic rates and body mass.



**Figure 5.1:** (a) **Spatial distribution of the sampled sites** from the PREDICTS database for terrestrial vertebrates (6,484 sites); (b) **Number of sites in each land-use and land-use-intensity category**; (c) **Net primary productivity by land use and land-use intensity** (derived from MODIS satellite imagery), with model predictions plotted relative to minimally used primary vegetation (and rescaled with reference to that land-use type). Primary: primary vegetation; secondary: secondary vegetation; plantation: plantation forest.

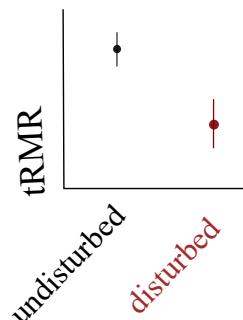
**(a) Assemblage-level:** there is less energy available to utilize by vertebrate assemblages in disturbed land uses than in undisturbed land uses.



**Prediction 1:** within all trophic groups, total assemblage-level **mass-dependent** metabolic rates decrease in disturbed land uses compared to undisturbed land uses.

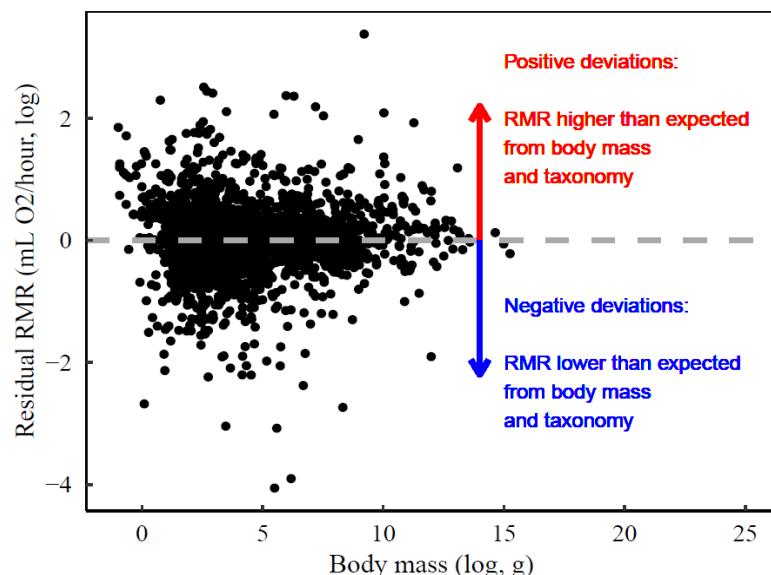
$$\text{Assemblage level:} \\ t\text{RMR} = \sum_i (\text{RMR}_i * a_i)$$

$$\text{Across assemblages - model 1:} \\ \log(t\text{RMR}) \sim \text{LU} + \text{LUI} + \text{TG} + \\ \text{LU:LUI} + \text{LU:TG} + \text{LUI:TG} + \\ \text{LU:LUI:TG} + RE$$



**(b) Species-level:** species with lower energetic expenditure than expected from body mass and taxonomy (negative deviations) do better in disturbed land uses than species with higher energetic expenditure than expected from body mass and taxonomy (positive deviations).

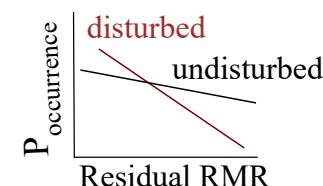
$$\text{residuals of} \\ \log(\text{RMR}) \sim \log(\text{BM}) + \\ (1|\text{Class/Order/Family}))$$



**Prediction 2:** within all trophic groups, species occurrence probability in disturbed land uses is negatively affected by **residual RMR**. The slope of the relationship between occurrence probability and residual RMR is significantly lower than the slope estimate for the undisturbed land uses.

Across species - model 2:

$$P_{\text{occurrence}} \sim \text{LU} + \text{LUI} + \text{resRMR} + \text{TG} + \\ \text{LU:LUI} + \text{LU:TG} + \text{LU:resRMR} + \\ \text{LUI:TG} + \text{LUI:resRMR} + \text{TG:resRMR} + \\ \text{LU:TG:resRMR} + \text{LUI:TG:resRMR} + RE$$



**Figure 5.2: Framework for the predictions and models.** I use resting metabolic rates (RMR) as a proxy for the amount of energy processed by vertebrate species. (a) I hypothesize that there is less energy overall that can be utilized by vertebrate species in disturbed land uses compared to undisturbed land uses. As such, I expect decreases in assemblage-level total metabolic rates ( $t\text{RMR}$ ).  $t\text{RMR}$  is calculated as the sum, over all species occurring at a site, of species RMR ( $\text{RMR}_i$ ) weighted by species abundance ( $a_i$ ). (b) After removing the effects of body mass and taxonomic position, I expect species with lower energy expenditure than expected to do better than species with higher energy expenditure than expected. I predict that the slope of the relationship between residual RMR and occurrence probability is more negative in disturbed land uses than in undisturbed land uses. For both (a) and (b), I test whether the predictions are consistent among trophic groups. LU: Land use; LUI: land-use intensity; TG: trophic group; RE: random effects. In model 1, random effects include the identity of the PREDICTS studies from which the assemblages are derived. In model 2, random effects include study, site, and species identity.

988 **5.2 Methods**

989 **5.2.1 Vertebrate assemblage composition**

990 I obtained vertebrate assemblage composition in different land uses from the PREDICTS database (**Hudson2014**;  
991 **Hudson2017**). The PREDICTS database is a large collection of published studies that measure biodiversity  
992 across different land uses and is one of the most comprehensive global databases of its type. In each PRE-  
993 DICTS study, species occurrence and often abundance were recorded across different sites. Each site was  
994 assigned to one of the following land-use types: primary vegetation (natural, undisturbed vegetation), sec-  
995 ondary vegetation (recovering after complete destruction of primary vegetation), plantation forest (woody  
996 crops), pasture (areas grazed by livestock), cropland (herbaceous crops) and urban (built-up areas). The  
997 land-use categories were assigned based on habitat descriptions from the original studies (**Hudson2014**),  
998 sometimes in consultation with the original study authors. Each site was also classified in terms of land-use  
999 intensity as either minimal, light or intense. The land-use-intensity assignment was also made on the basis  
1000 of the habitat description in the original studies, and depended on criteria specific to each land use (such as  
1001 degree of mechanisation, yield or chemical inputs for cropland; or the amount of green space in urban areas;  
1002 **Hudson2014**).

1003 I subset the PREDICTS database for studies that sampled terrestrial vertebrates, and for which both land  
1004 use and land-use intensity had been characterised. I thus obtained 181 studies for 4,238 species sampled  
1005 across 6,484 sites (Figure ??a). Sample sizes varied across land uses and land-use intensities (Figure ??b).

1006 **5.2.2 Energy availability by land-use type and land-use intensity**

1007 The predictions of this Chapter rely on the assumption that resource types and abundance are modified in  
1008 disturbed environments, with less energy available in disturbed compared to undisturbed land uses over-  
1009 all. To test this assumption, I used terrestrial net primary productivity (NPP) across land uses as a proxy  
1010 for available energy. NPP quantifies the amount of atmospheric carbon fixed by plants and accumulated  
1011 as biomass. NPP estimates were derived using imagery from the Moderate Resolution Imaging Spectrora-  
1012 diometer (MODIS) on board NASA's Terra satellite. NPP estimates were based on a yearly composite of  
1013 measures made at 8-day intervals, captured at 500-m spatial resolution (**Running2015**). I obtained NPP for  
1014 4,062 of the PREDICTS sites used in the analysis (matching the sites to the NPP data using the sampling  
1015 year available in PREDICTS). I fit a linear mixed-effects model (lme4 package, version 1.1-23, **Bates2015**)  
1016 explaining site-level NPP by land use and land-use intensity, with a random intercept accounting for study

1017 identity, to control for differences in experimental design across studies. Model predictions showed that  
1018 NPP decreased significantly in several land uses (e.g., pasture and cropland) compared with the primary  
1019 vegetation reference level, although the strength and in some cases direction of the difference varied among  
1020 land-use and land-use intensity combinations (e.g., increases in urban land uses; Figure ??c).

### 1021 **5.2.3 Resting Metabolic Rates (RMR) & imputations of missing RMR values**

1022 As a proxy for species-level energetic expenditure, I used estimates of the minimum amount of energy re-  
1023 quired for organismal maintenance, i.e., basal metabolic rates (BMR) for endotherms, and resting metabolic  
1024 rates (RMR) for ectotherms. From the literature, I obtained estimates of BMR for 719 species of birds and  
1025 685 mammals, and estimates of RMR for 126 amphibians and 173 reptiles (Appendix 4, Table ??). For  
1026 endotherms, BMR are measured when species are in their thermoneutral zone, that is, when there is little to  
1027 no energy expenditure allocated to thermoregulation. Thus, BMR estimates were derived from lab studies  
1028 that mostly measured oxygen consumption of the organisms at rest under controlled conditions and in the  
1029 thermoneutral zone of the species. For an ectotherm, there is no ‘basal’ metabolic rate, as body temperature  
1030 mainly depends on environmental temperature. Their metabolic rates follow a hump-shaped relationship  
1031 with environmental temperature, highest at an optimal temperature which corresponds to a performance  
1032 peak. To be able to compare endotherms’ BMR with ectotherms’ RMR, **Stark2020** used the metabolic rates  
1033 that correspond to a performance peak for both groups (i.e., BMR in the thermoneutral zone for endotherms,  
1034 and metabolic rates at optimal temperature for ectotherms). Thus, I used the data compiled in **Stark2020** for  
1035 ectotherms, and from the sources specified in Table ?? for endotherms. The units for BMR and RMR were  
1036 standardized to mL of dioxygen consumed per hour (mLO<sub>2</sub>/h). As in **Stark2020**, I henceforth refer to both  
1037 basal and resting metabolic rates as RMR.

1038 For the species occurring in PREDICTS, initial data coverage for RMR was poor (Table ??), necessi-  
1039 tating imputation of missing values. To do so, I first measured the phylogenetic signal in BMR and RMR  
1040 ( $\log_e$ -transformed), using Pagel’s  $\lambda$  (**Pagel1999**), to assess whether metabolic rates were sufficiently phy-  
1041 logenetically conserved to be estimated from species phylogenetic position. I obtained class-specific phy-  
1042 logenetic trees from **Jetz2012** for birds, from **Faurby2018; Faurby2020** for mammals, from (**Tonini2016**)  
1043 for reptiles (squamates), and from **Jetz2018** for amphibians (all downloaded in April 2020). For each class,  
1044 I randomly sampled 100 trees. To account for phylogenetic uncertainty, I calculated Pagel’s  $\lambda$  for each  
1045 sampled tree and reported the median value, as well as the 2.5th and 97.5th percentiles (Table ??).

1046 In addition to being highly phylogenetically conserved (Table ??), RMR correlate strongly with body  
1047 mass (Figure ??a). Thus, I imputed missing values using body-mass information (see next section), phylo-

1048 genetic relationships and taxonomic orders as predictors (**Penone2014**). For each class, I used a consensus  
1049 phylogenetic tree from which I summarised phylogenetic relationships in the form of five phylogenetic eigen-  
1050 vectors. Including more eigenvectors had little impact on the imputed values (results not shown). Consensus  
1051 trees were obtained with the TreeAnnotator programme of the BEAST software (**Bouckaert2014**). Miss-  
1052 ing RMR values were imputed using random forests algorithms implemented in R using the ‘missForest’  
1053 package (version 1.4; **Stekhoven2012**; **Stekhoven2016**).

#### 1054 **5.2.4 Trophic group and body mass information**

1055 I used body mass and trophic group information for terrestrial vertebrates compiled in Chapter 2. Body  
1056 mass was compiled as a single measure at the species level, meaning I was unable to consider intraspecific  
1057 variation. Trophic group described species as either carnivores, omnivores, or herbivores. Because there  
1058 were gaps in the availability of the data, more so for trophic group than for body mass (see Chapter 2), I  
1059 imputed the missing trait values (independently of RMR imputations), then used both imputed and empirical  
1060 body mass values for imputations of missing RMR values. To impute missing body mass and trophic groups,  
1061 I used random forests algorithms (again, using the missForest R package), including as additional predictors  
1062 phylogenetic information, added in the form of 10 phylogenetic eigenvectors (**DinizFilho2012**) following  
1063 **Penone2014**, and also taxonomic order. I considered a wider set of life-history traits in the missing values  
1064 imputations: lifespan, litter/clutch size, habitat breadth and use of artificial habitats (compiled in Chapter  
1065 2). Phylogenetic eigenvectors were extracted from the class-specific phylogenies using the PVR package  
1066 (**Santos2018**).

#### 1067 **5.2.5 Effects of land use, land-use intensity and trophic group on assemblage-level total 1068 RMR (prediction 1; Figure ??a)**

1069 Assemblage-level total RMR (tRMR) was obtained by summing abundance-weighted RMR for the species  
1070 occurring in each site; abundance data were available for 125 of the 181 PREDICTS studies I considered  
1071 (sampling 3,487 species across 4,644 sites). I fitted a linear mixed-effects model to explain  $\log_e$ -tRMR as  
1072 a function of land use, land-use intensity and trophic level, with a random intercept accounting for study  
1073 identity to control for differences in experimental design across studies. I started with a model allowing all  
1074 two-way interactions among the predictors. I then tested whether adding the three-way interaction among  
1075 land use, land-use intensity and trophic level improved the fit of the model, using a likelihood-ratio test.  
1076 The model that included the three-way interaction was retained ( $P \ll 0.01$ ; model 1, Figure ??). In addition,  
1077 because it is well established that resting metabolic rates are influenced by temperature (**Clarke2004a**), I

1078 checked whether including annual mean temperature in the model affected the conclusions. Annual mean  
1079 temperature at each PREDICTS site was estimated from WorldClim version 2.1 (**Fick2017**), using a 2.5  
1080 arc-minute resolution. Adding annual mean temperature did not improve model fit (likelihood-ratio test:  
1081  $P=0.113$ ), thus I did not consider its effects any further.

1082 **Model validation.**

1083 To ensure that imputation uncertainty did not affect the conclusions, I refitted model 1 using the subset of  
1084 species ( $n = 426$ ) from PREDICTS for which there were empirical RMR information (i.e., excluding imputed  
1085 RMR values).

1086 **Disentangling the effects of body mass and abundance on tRMR.**

1087 Since RMR correlates strongly with body mass, changes in tRMR are likely to be driven in part by changes  
1088 in the size-spectrum of ecological assemblages. I fitted an additional model to explain changes in species'  
1089 abundance (given presence) by land use, land-use intensity, trophic group, body mass and their interactions,  
1090 to understand the role of shifts in the body mass of species on observed changes in tRMR (see Appendix 4,  
1091 Figure ??).

1092 **5.2.6 Effects of land use, land-use intensity, trophic group and residual RMR on species  
1093 occurrence probability (prediction 2; Figure ??b)**

1094 To control for the effects of body mass and taxonomy on RMR, I used the residual variation in RMR after  
1095 accounting for these variables, from a linear mixed-effects model fitting  $\log_e$ -RMR as a function of  $\log_e$ -  
1096 body mass with nested random taxonomic effects ( $1|Class/Order/Family$ ; Figure ??). Hence, I used a metric  
1097 that describes how much more energy (positive deviations) or less energy (negative deviations) than expected  
1098 from body mass and taxonomic position a species spends for organismal maintenance. Similar approaches  
1099 have been used in previous papers (**Furness2008**; **Naya2013**). As detailed earlier, I expect species with  
1100 lower residual RMR to do better in disturbed land uses than species with higher residual RMR (prediction  
1101 2; Figure ??b) because, given any body mass, investing less energy in maintenance could contribute to  
1102 persistence in a context of resource scarcity.

1103 To test the second prediction, I fitted a binomial mixed-effects model explaining species occurrence with  
1104 land use, land-use intensity, trophic group and residual RMR. I started with a complete model that included  
1105 all two-way interactions among the main effects. Because I wanted to test whether the second prediction was  
1106 valid for each trophic group, I needed to account for potential differences in the slope of the relationships

1107 between occurrence probability and residual RMR among trophic groups. Thus, I performed a forward  
1108 stepwise selection procedure to test whether adding three-way interactions among (1) land use, trophic group  
1109 and residual RMR and (2) among land-use intensity, trophic group and residual RMR improved model  
1110 fit, using likelihood-ratio tests. The final model included both three-way interactions (Figure ??b; model  
1111 2). I fitted random effects that accounted for species identity, as well as for study and site identity within  
1112 PREDICTS.

1113 **Model validation.**

1114 I checked the phylogenetic signal in the model residuals using Pagel's  $\lambda$  (**Pagel1999**). Non-significant phy-  
1115 logenetic signal in the residuals would indicate that fitting species identity in the model's random effects  
1116 was sufficient to account for residual phylogenetic variation in RMR. Further, to assess the potential effects  
1117 of imputation uncertainty on the results, I again fitted model 2 on the data subset for the 489 species with  
1118 collected empirical RMR values, across 5,948 sites in 151 studies (i.e., excluding imputed values).

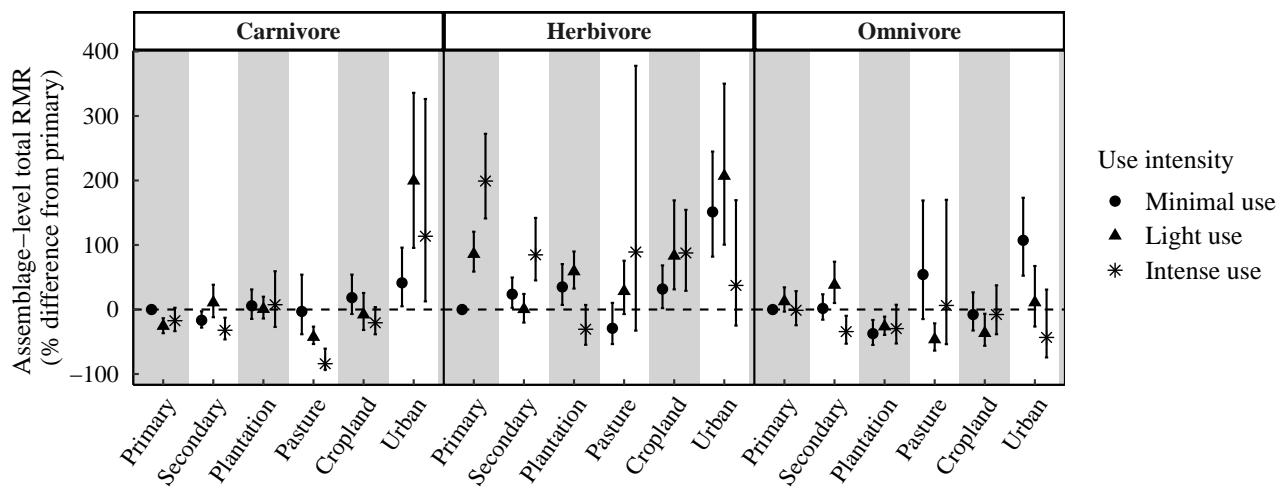
1119 **5.3 Results**

1120 **5.3.1 Effects of land use, land-use intensity and trophic group on assemblage-level total  
1121 RMR**

1122 Land use, land-use intensity, trophic group and their interactions had significant effects on assemblage-level  
1123 total RMR (Figure ??). Overall, and contrary to our expectations, assemblage-level total RMR did not show  
1124 systematic decreases in disturbed land uses. In fact, urban land uses were associated with strong significant  
1125 increases in tRMR in all trophic groups (e.g., a 200% average increase in tRMR in lightly-used urban areas  
1126 for carnivores, compared with primary vegetation levels; +207% on average in lightly-used urban areas for  
1127 herbivores; +107% for minimally-used urban areas for omnivores). In other land uses, responses depended  
1128 on trophic group and land-use intensity. Whilst for herbivores, disturbed land uses were typically associated  
1129 with increases in tRMR, we detected decreases in tRMR for omnivores and carnivores in several land uses,  
1130 most notably in intensely-used pasture for carnivores (-84%). Such effects could reflect changes in the  
1131 size-spectrum of local assemblages (see Appendix 4, Figure ??). For instance, in minimally-used urban  
1132 areas, larger herbivores tended to occur at higher abundances compared to primary vegetation level; and in  
1133 intensely-used pastures, carnivores tended to occur at lower abundances overall (Figure ??).

1134 The model residuals were appropriately distributed (see diagnostic plots in Appendix 4, Figure ??).

1135 Investigating the sensitivity of our results to imputation uncertainty showed that our results and conclusions  
 1136 were robust to the removal of all imputed estimates of RMR (the correlation coefficient was 0.72 between the  
 1137 two sets of model coefficients; Figure ??). Comparing model predictions showed that effects were mostly  
 1138 congruent, although there were some differences (Figure ??). In particular, for herbivores, effect sizes tended  
 1139 to be bigger for the model fitted on empirical data compared with the model that included imputed data. Thus,  
 1140 our main results appear to be conservative if anything. The model fitted on empirical data had larger standard  
 1141 errors, likely due to the reduction in sample size.

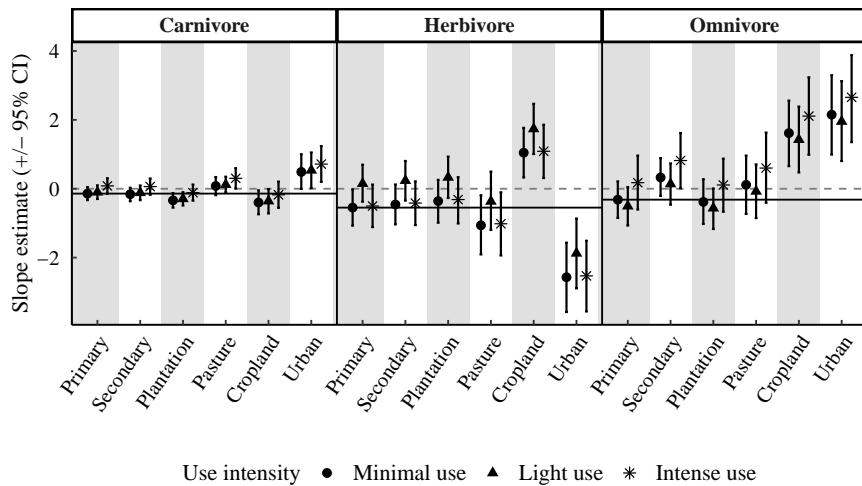


**Figure 5.3: Effects of land use, land-use intensity and trophic group on assemblage-level total RMR.** Model predictions are rescaled with reference to minimally-used primary vegetation, considered to be the undisturbed baseline. Primary: primary vegetation; secondary: secondary vegetation; plantation: plantation forest.

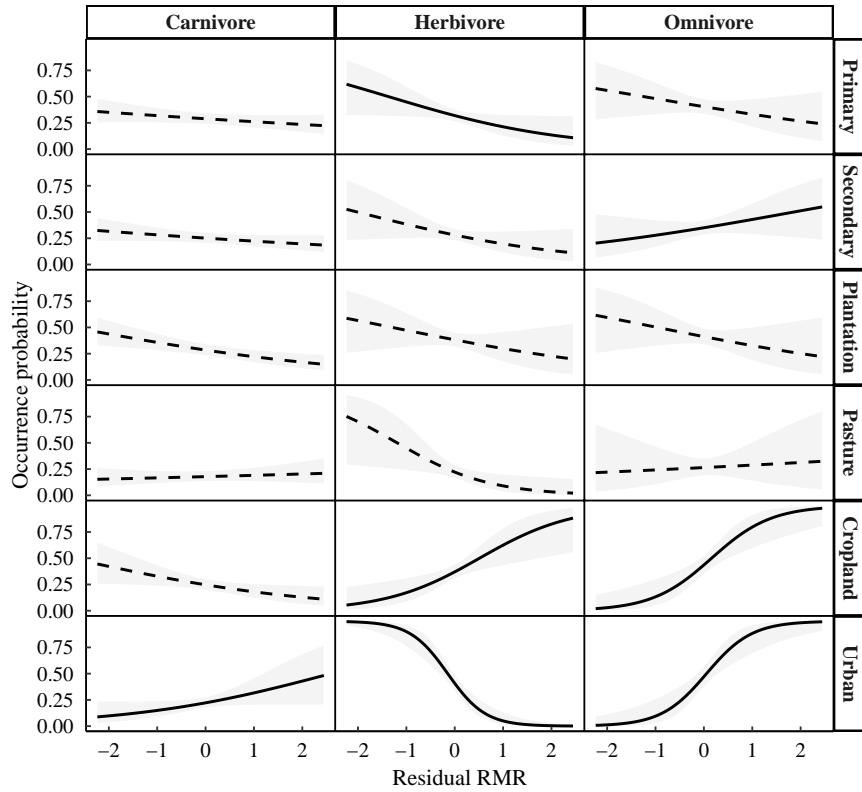
### 1142 **5.3.2 Effects of land use, land-use intensity, trophic group and residual RMR on species' 1143 occurrence probability**

1144 Species' occurrence probability was significantly affected by land use, land-use intensity, trophic level, resid-  
 1145 ual RMR and their interactions (Figures ??, ??). Contrary to our expectations, species with higher residual  
 1146 RMR (relative to their body mass and taxonomic position) tended to do better than species with lower resid-  
 1147 ual RMR in a number of disturbed land uses. Overall, land-use type was more important for determining the  
 1148 relationship between occurrence probability and residual RMR than land-use intensity (Figure ??a).

(a) Slope of the relationships between occurrence probability (log-odds) and residual RMR



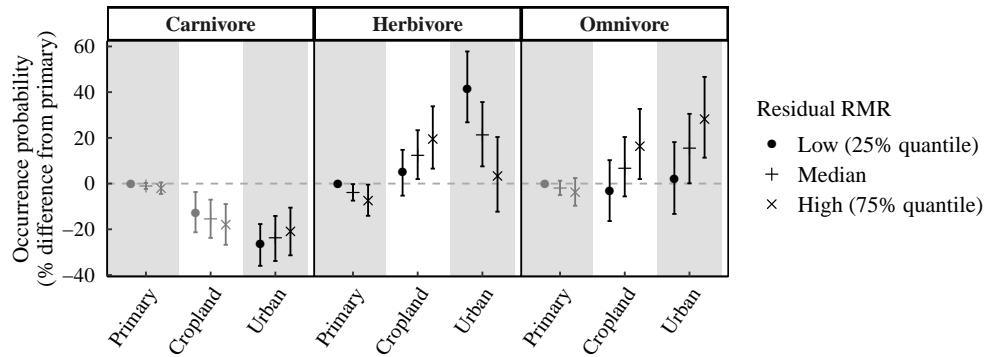
(b) Predicted effects of residual RMR on occurrence probability (minimal use intensity)



**Figure 5.4:** (a) Slope estimates for the relationship between residual RMR and occurrence probability in each land-use type and for the three levels of land-use intensity. The black horizontal line indicates the slope for the reference level (primary vegetation) for minimal land-use intensity. The grey dashed line marks 0. Error bars are 95% confidence intervals. (b) Effect of residual RMR on species probability of occurrence within each trophic level and for each land use-type. I plotted the predictions for minimal land-use intensity only. Solid lines represent significant relationships. Primary: primary vegetation; secondary: secondary vegetation; plantation: plantation forest.

1149 For minimally-used primary vegetation (reference), the model predicted negative effects of residual RMR  
 1150 on species occurrence probability in all trophic levels (but with a significant slope for herbivores only; Figure  
 1151 ??a). However, the directionality of this relationship was reversed in some disturbed land uses in all trophic  
 1152 groups (secondary vegetation, cropland and urban for omnivores; cropland for herbivores; urban for carni-  
 1153 vores), with significant positive slopes, also significantly higher than those observed for primary vegetation  
 1154 (Figure ??a). The only exception was the opposite pattern for urban herbivores (Figure ??b), where residual  
 1155 RMR had a more negative effect on occurrence probability than in minimally-used primary vegetation.

1156 I would like to emphasize that positive effects of residual RMR on occurrence probability in some of the  
 1157 most disturbed land uses (e.g., urban for carnivores) do not mean that there were absolute increases in species  
 1158 occurrence probability in disturbed land uses compared to primary vegetation (and vice-versa). I illustrate  
 1159 this point in Figure ???. For carnivores with a median value for residual RMR, occurrence probability was  
 1160 reduced by an average 24% in urban land uses compared to primary vegetation.



**Figure 5.5: Predicted occurrence probabilities (+/- 95% confidence interval) in primary vegetation (primary) and for the most disturbed land uses where residual RMR was found to have a significant effect on occurrence probability.** For visualisation purposes, I discretised residual RMR in three levels. The predicted probabilities of occurrence were rescaled with reference to primary vegetation for the lowest value of residual RMR (25% quantile). Here, the predictions are plotted for minimal land-use intensity (effects would be similar for light and intense land-use intensities). Black points and error bars are plotted where the relationship between occurrence probability and residual RMR was significant (and dark grey points and error bars represent non-significant trends).

1161 Finally, the model showed some degree of deviation from distributional assumptions (diagnostic plots,  
 1162 Appendix 4, Figure ??). Nevertheless, the model's coefficients were similar when estimated with a Bayesian  
 1163 framework, suggesting that the estimates were robust (Figure ??). The phylogenetic signal in the model  
 1164 residuals was weak and non-significant ( $\lambda=0.004$ ,  $P=0.24$ ). Re-fitting the model using the complete data  
 1165 subset (i.e., excluding imputed RMR estimates) showed that our conclusions are likely robust to imputa-  
 1166 tion uncertainty (Figure ??), with congruent results overall, although there were a few differences in the  
 1167 predictions between the two models – notably, for herbivores in urban land uses (Figure ??).

1168 **5.4 Discussion**

1169 The results of this Chapter provide insights into the relationship between land-use change and energetic  
1170 requirements at two levels of organisation (at the assemblage level and the species level), and contradict  
1171 both my initial predictions. I found that total energetic requirements did not show systematic decreases in  
1172 disturbed land uses. On the contrary, there were strong increases in all trophic groups, most notably in urban  
1173 land uses. Further, in all trophic groups, species with higher energetic expenditure (than expected from body  
1174 mass and taxonomy) tended to be better in some of the most disturbed land uses (that is, cropland and urban)  
1175 than species with lower energetic expenditure (than expected from body mass and taxonomy).

1176 At the assemblage level, the results first highlight the effects of land-use change on vertebrate community  
1177 metabolism. Contrary to my expectations, I found differing effects of land use on total energetic requirements  
1178 among trophic groups, reflecting changes in the size-spectrum of ecological assemblages (i.e., changes in  
1179 the distribution of abundance along the body mass gradient). On the one hand, decreases in total energetic  
1180 requirements, such as observed for carnivores in intensely used pastures, are likely due to overall reductions  
1181 in local abundance (with carnivorous species in pastures perceived as a threat to livestock; **VanEeden2018**;  
1182 or because of increased competition for fewer resources). On the other hand, increases in total energetic  
1183 requirements could reflect higher levels of resource intake in some disturbed land-use types. On the basis  
1184 of net primary productivity, I hypothesized that there would be fewer resources available in disturbed areas.  
1185 However, it could be that the carrying capacity of some disturbed land-use types actually increases for  
1186 some trophic groups compared to that of primary vegetation, because of the presence of novel or more  
1187 abundant food sources. For instance, in urban areas, both wildlife feeding and the presence of anthropogenic  
1188 food sources, such as human refusal or pet food, could lead to an increase in resource availability all year-  
1189 round for species with a non-specialised diet that are able to exploit such food sources (**Fischer2012**). Past  
1190 research on urban carnivores has shown that some species have adapted to urban environments by exploiting  
1191 anthropogenic food sources, and also benefit from physical protection (with some human structures providing  
1192 shelter) and from reduced occurrence of natural enemies (**Bateman2012**). Herbivores and omnivores could  
1193 also benefit from increased resources in urban areas, notably where the urban matrix includes semi-natural  
1194 habitats, which is congruent with observed increases in occurrence probability in minimally used urban  
1195 areas for these trophic groups. Further, anthropogenic ‘bottom-up food forcing’ has been shown to affect  
1196 food-web dynamics, with decreases in predation rates, loss of anti-predator abilities (**Geffroy2020**) and  
1197 mesopredator release (**Fischer2012**). Thus, in urban areas, increases in anthropogenic food sources could  
1198 provide support for a higher abundance of medium-sized species released from predation, increasing total

1199 energetic requirements.

1200 I hypothesized that resource availability would be the main driver behind changes in energetic require-  
1201 ments. However, resource availability likely interacts with other factors to explain the patterns. In particular,  
1202 past work has shown that brain size is an important predictor of species ability to cope with disturbance,  
1203 notably in urban settings (**Sayol2020**). Since brain size scales allometrically with body mass, it is possi-  
1204 ble that I observe increases in total energetic requirements where having larger brains is advantageous for  
1205 resource acquisition, and in turn where resources are both exploitable and abundant enough to sustain the  
1206 larger energetic expenditures associated with bigger brain sizes. In addition, body size, which explains an  
1207 important proportion of the interspecific variation in mobility, likely interacts with characteristics of the land-  
1208 scape (such as fragmentation) to predict species survival in disturbed land uses (**Merckx2018**). Hence, the  
1209 patterns I observe in this Chapter could emerge from interactions and trade-offs among resource availability,  
1210 diet, body size and human management of the landscape.

1211 At the species level, the results underline the role of energetic constraints on species responses to land-  
1212 use change. After controlling for the effects of body mass and taxonomy, I found that residual energetic  
1213 expenditure was a significant predictor of species occurrence probability in disturbed land uses. Contrary  
1214 to my expectations, in several disturbed land uses and in all trophic groups, species with higher residual ener-  
1215 getic expenditure tended to have a higher occurrence probability than species with lower residual ener-  
1216 getic expenditure. It could be that species with lower residual energetic requirements are less well equipped  
1217 than species with higher residual energetic requirements at making use of the available food sources in dis-  
1218 turbed land uses. Species in disturbed land uses may need to display higher levels of feeding innovation  
1219 (**Coogan2018**), have larger brains (**Sayol2020**), or be bolder and more active to make use of the available re-  
1220 sources. In mammals, past research has shown that larger brains are associated with larger residual energetic  
1221 expenditure (**Isler2006**). At the individual level, past research suggests that metabolic rates are linked to  
1222 differences in behaviour, with bolder and more active individuals exhibiting higher metabolic rates than less  
1223 active individuals, with consequences for food acquisition (**Biro2010**). Although I am not aware of similar  
1224 evidence at the species level, I propose that residual metabolic rates interact with species' ecological traits,  
1225 behavioural traits and foraging strategy in influencing species' responses to land use. Among species with  
1226 an adaptable diet, able to make use of the resources available in a disturbed landscape, those with higher  
1227 residual metabolic rates could present a set of behavioural characteristics that render them better at acquiring  
1228 the available resources, hence more able to cope with land-use disturbance. This could be the case for urban  
1229 carnivores and omnivores, as well as herbivores in cropland. Thus, it is possible that disturbed land uses  
1230 favour species with higher residual resting metabolic rates because such species are overall better competi-

1231 tors when faced with disturbance. In addition, release from predation in some disturbed land uses – most  
1232 notably in urban areas – could favour bolder species, as they are less exposed to predation risks. Conversely,  
1233 and in accordance with my initial hypothesis, the results suggest that when there are no or few exploitable  
1234 resources in disturbed areas, spending less energy than expected from body mass and taxonomy might be  
1235 beneficial for persistence (this could be the case for urban herbivores, which include species specialised on  
1236 fruit, nectar, and other plant materials that likely become less abundant and less substitutable in disturbed  
1237 areas).

1238 To summarise, I propose that, for species with adaptable diets, increases in exploitable resources coupled  
1239 with decreases in predation rates leads to increases in overall energetic expenditure, associated with shifts  
1240 towards assemblages composed of a higher abundance of medium-sized, bolder, more active and larger-  
1241 brained species. On the other hand, species with specialised diets might be better able to persist and cope  
1242 with resource scarcity when their energetic expenditure are lower than expected from body mass. These  
1243 ideas could be further tested by considering more refined dietary groups (as those compiled in Chapter 4)  
1244 rather than trophic groups, as diet should be easier to link with resource availability in a given land-use type.  
1245 This could constitute a development for this work.

1246 Finally, I would like to emphasize that energetic budget allocation is a fundamental aspect of organismal  
1247 fitness (**RobertBurger2021**). Past studies suggest that individuals exposed to food scarcity can lower their  
1248 intrinsic energy expenditure to increase their chances of survival. For instance, **Zhang2018a** showed that,  
1249 subjected to food restrictions, the Chinese bulbul (*Pycnonotus sinensis*) was able to lower its basal metabolic  
1250 rate not only passively through body and organ mass reductions, but also by modifying enzymatic activity  
1251 and metabolism in the muscles and liver. Other studies have also found that bird nestlings downregulate  
1252 energetic costs allocated to maintenance when exposed to food shortage (**Brzek2001; Moe2004**), or to dis-  
1253 turbances such as urban noise (**Brischoux2017**). Intraspecific variation in energetic expenditure could affect  
1254 individual responses to land-use change, which in turn could affect species' ability to adapt to human pres-  
1255 sures. However, the lack of spatially explicit estimates of metabolic rates precluded its consideration in this  
1256 Chapter.

1257 To conclude, the findings of this Chapter suggest important effects of land-use change on energetic  
1258 fluxes, and thus potentially on ecosystem functioning. Further research is warranted as to the effects of these  
1259 energetic patterns for ecosystem processes. Indeed, assemblage-level energy fluxes may serve as important  
1260 indicators of change for ecosystem processes, such as decomposition (**Barnes2014**). Understanding these  
1261 impacts is particularly important given increasing land-use change and biodiversity loss across the globe.

1262 **6 | General discussion**

1263 7 | Conclusion

<sup>1264</sup> **Bibliography**

# **Appendices**

## **Appendix 0: Supporting information for Chapter 1**

# **Appendix 1: Supporting information for Chapter 2**

## **S2.1 Taxonomic corrections**

Across the different sources, similar species could appear under different binomial names. This was a problem when matching datasets by species. Moreover, it is possible that within a source, a given species was appearing under two or more different, synonymous names. As such, taxonomic synonymy created duplicated rows for the same species, overall falsely increasing the total number of species and potentially inflating the number of missing trait values. Taxonomic synonymy was hence a major issue. Due to the large number of species across datasets, extensive manual checks could not be applied. The presence of typos in species names had the same effect as synonymy, erroneously duplicating species. We attempted to correct for taxonomy first by correcting for typos, and second by identifying species which were entered under non-accepted names and replacing these with the accepted name. To this end, we developed an automated procedure, complemented with a few manual entries where errors were opportunistically spotted. Such errors in taxonomy were notably spotted when attempting to retrieve trait data for subsets of species, for analyses not directly related to the work conducted here. Taxonomic synonymy was as such checked manually for 91 species (56 birds, 7 mammals and 28 reptiles); in that case, information was extracted from other diverse sources (such as the Reptile Database (<http://www.reptile-database.org/>); Avibase (<https://avibase.bsc-eoc.org/avibase.jsp?lang=EN&pg=home>); AmphibiaWeb (<https://amphibiaweb.org/>); and additional manual checks using the IUCN Red List for mammals). A column in the Synonym dataset mentions where manual checks were applied (in which case the Synonym dataset was manually corrected).

### **Automated procedure and outputs.**

#### **Extracting names from the IUCN Red List and the Integrated Taxonomic Information System (ITIS).**

The objectives of the automated procedure were to (1) extract species synonymous binomial names from the

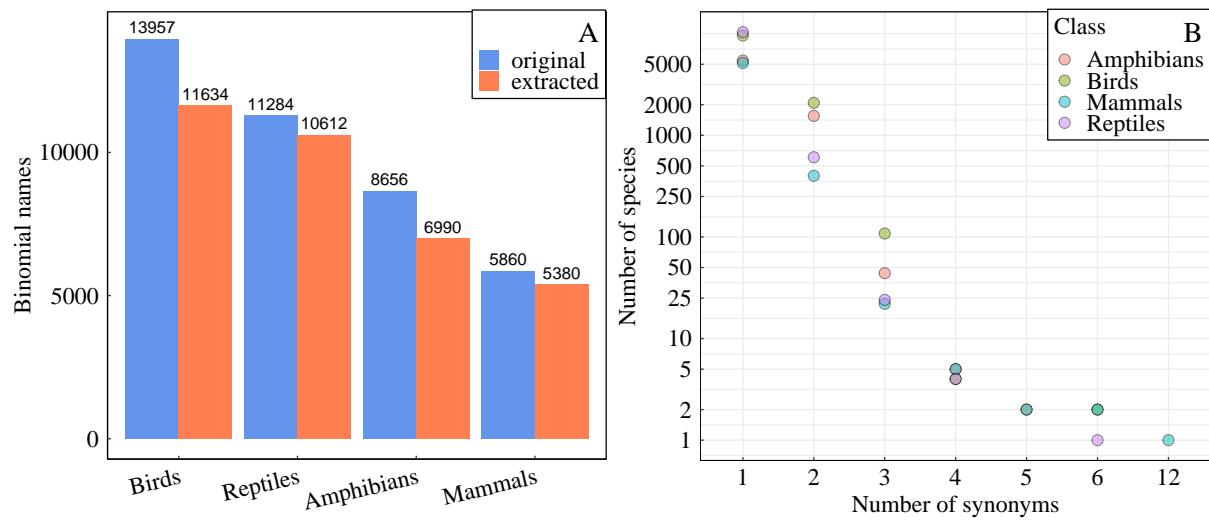
IUCN Red List or from ITIS, using the rredlist (**rredlist**) and taxize (**Chamberlain2013**) R packages; and (2) identify the status of each name (accepted or not accepted). We started by generating a list of all names featuring in any of the sources. These ‘original’ names were corrected for typos (using gnr\_resolve function in the taxize package). Then, the IUCN Red List was queried and any listed synonyms were stored, as well as the status of each synonym (accepted or not accepted). When species were not found in the IUCN Red List, synonyms were extracted from ITIS. When species were not found in ITIS either, corrected names (original names corrected for typos) were used. Family and order designations were extracted using the same procedure and some entries were retrieved from the Global Biodiversity Information Facility taxonomic backbone when not available in the Red List or in ITIS (GBIF, <https://www.gbif.org/tools/species-lookup>).

**NB:** for species entered with the forms *Genus cf.*, *Genus aff.* or *Genus spp.*, the accepted binomial name was left empty.

**Output.** We generated a ‘Synonym’ dataset containing records of binomial names (14,124 recorded names for birds; 8,743 for amphibians; 6,090 for mammals; and 11,678 for reptiles), and for each we recorded their status and their potential synonyms.

**Harmonising taxonomy in trait datasets.** Taxonomy across datasets was finally homogenised by replacing synonyms with a uniquely identified accepted name. As a consequence, the total number of identified unique species decreased (Figure ?? ). The species presenting the highest number of synonyms was the East African mole rat (*Tachyoryctes splendens*), for which we found 12 synonymous names (Figure ??B).

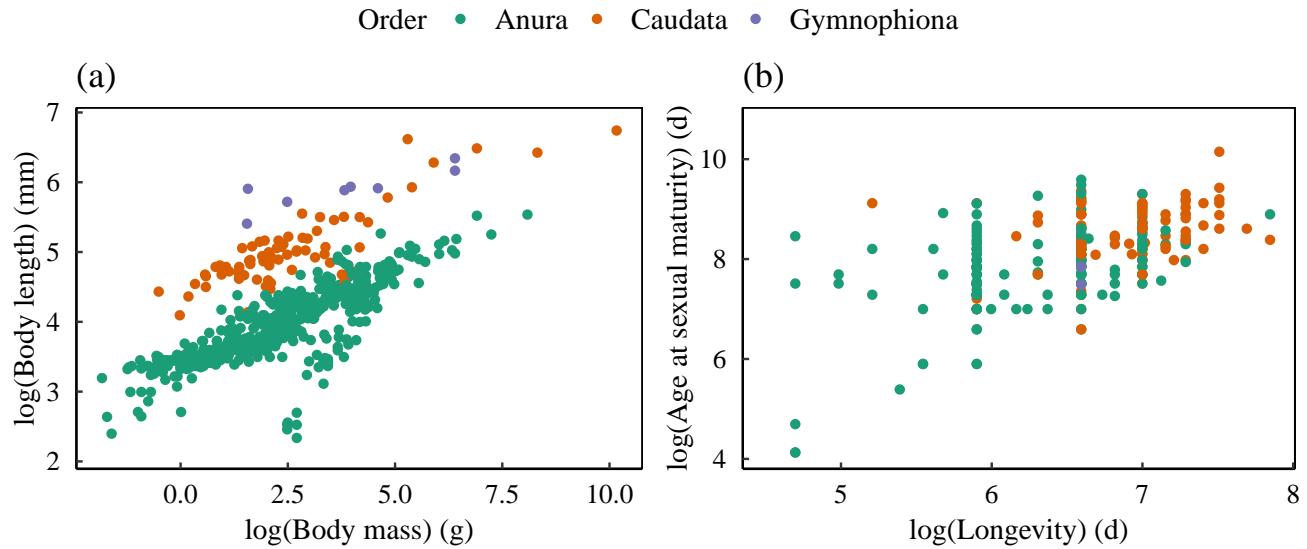
The automated procedure was not perfect, and taxonomic errors are likely to have persisted in the trait datasets. The Red List and the ITIS were not comprehensive taxonomic sources, and for clades with high degrees of synonymy in names, such as reptiles or amphibians, neither the Red List or the ITIS contained enough information. Taxonomy may be further improved by using class-specific sources in an automated procedure.



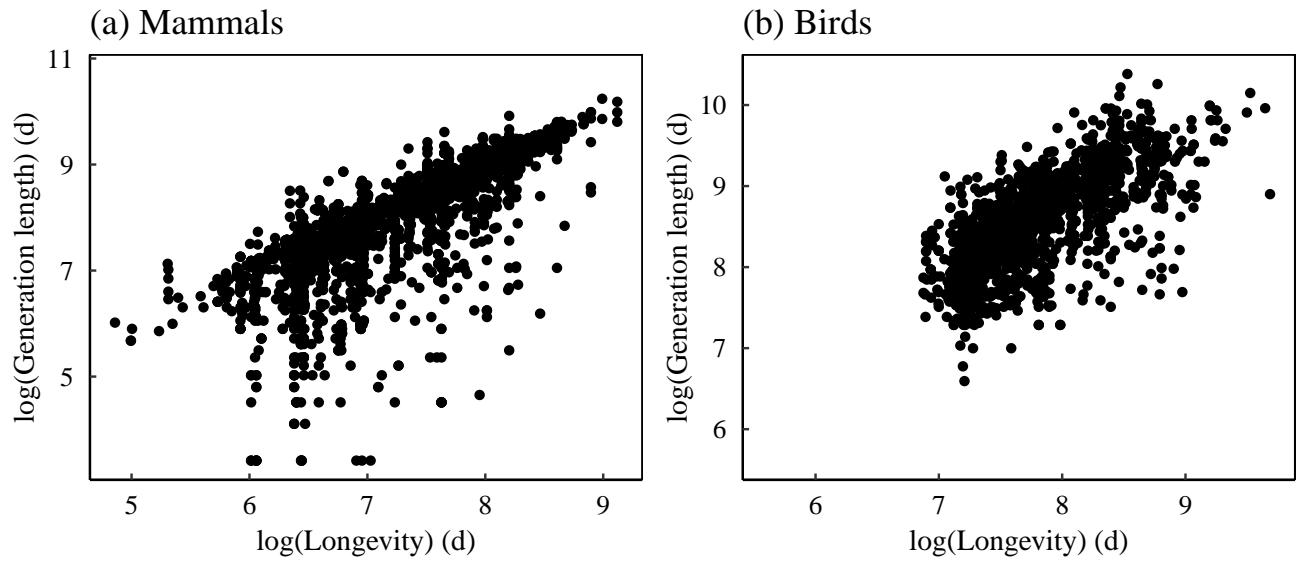
**Figure S2.1: Difference in species number before and after taxonomic correction (A) and distribution of number of synonyms across datasets (B).** (A) shows the number of species (binomial names) extracted from all sources (blue bars), and the number of uniquely identified accepted names (in red). Replacing non-accepted synonyms by one identified accepted name reduced the number of species in all classes, with the largest reduction for birds. (B) shows the distribution of the number of synonymous names. In all four classes, more than 5,000 species were known under one name only. Nevertheless, a large number of species had two identified synonyms (range: 400 species for mammals - 2086 for birds). The most potentially replicated species was the East African mole rat *Tachyoryctes splendens*, for which 12 synonyms were identified.

## S2.2 Additional information for trait compilation

### Correlations among closely related traits

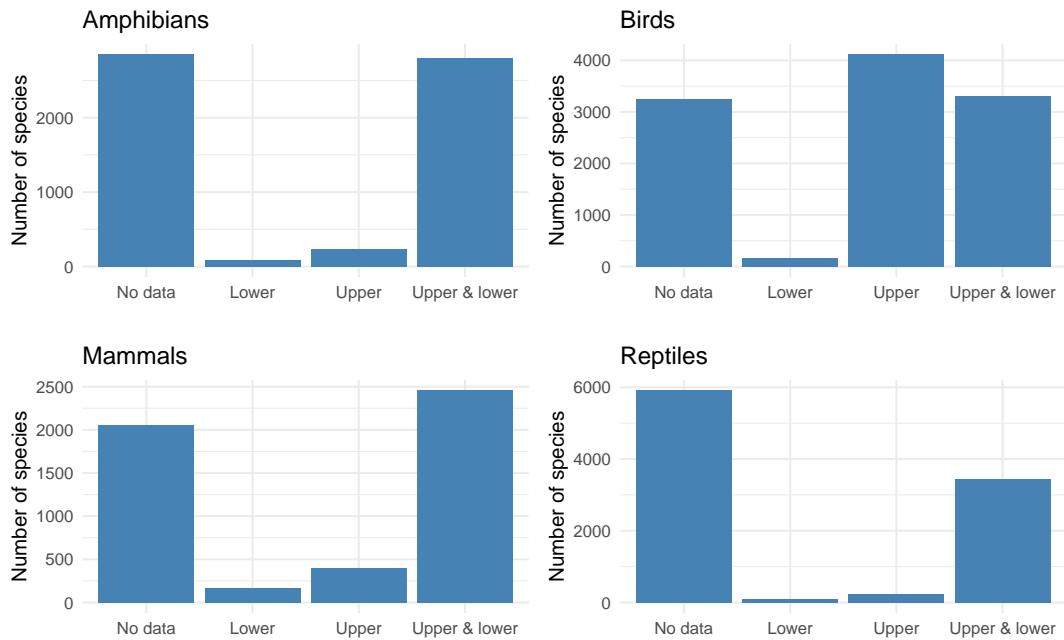


**Figure S2.2:** (a) Body mass versus body length and (b) longevity versus age at sexual maturity in amphibians. The Pearson's correlation coefficient was 0.71 in (a) and 0.55 in (b) (order was not included in these coefficients).

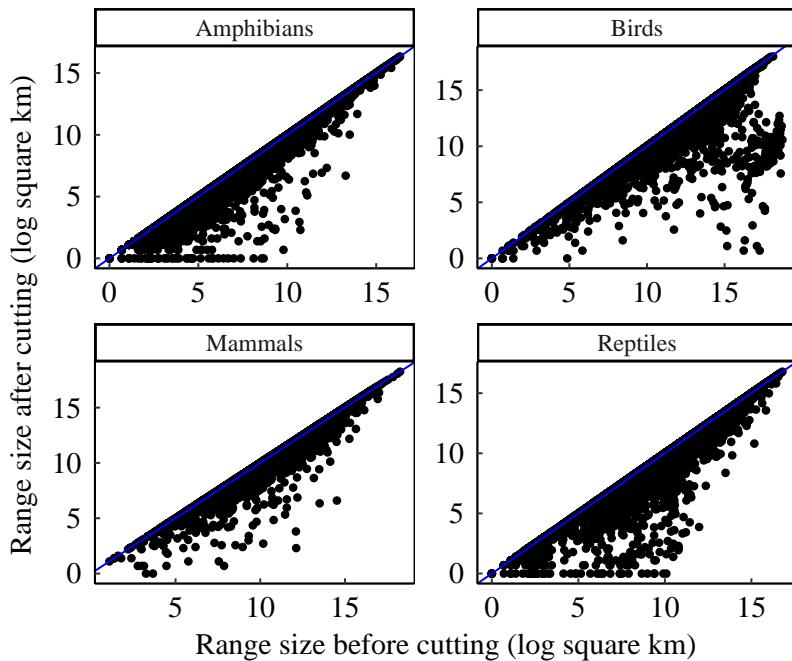


**Figure S2.3:** Generation length versus longevity data in mammals and birds. The Pearson's correlation coefficient was 0.74 in (a) and 0.70 in (b).

### S2.3 Cutting distribution maps by altitudinal limits

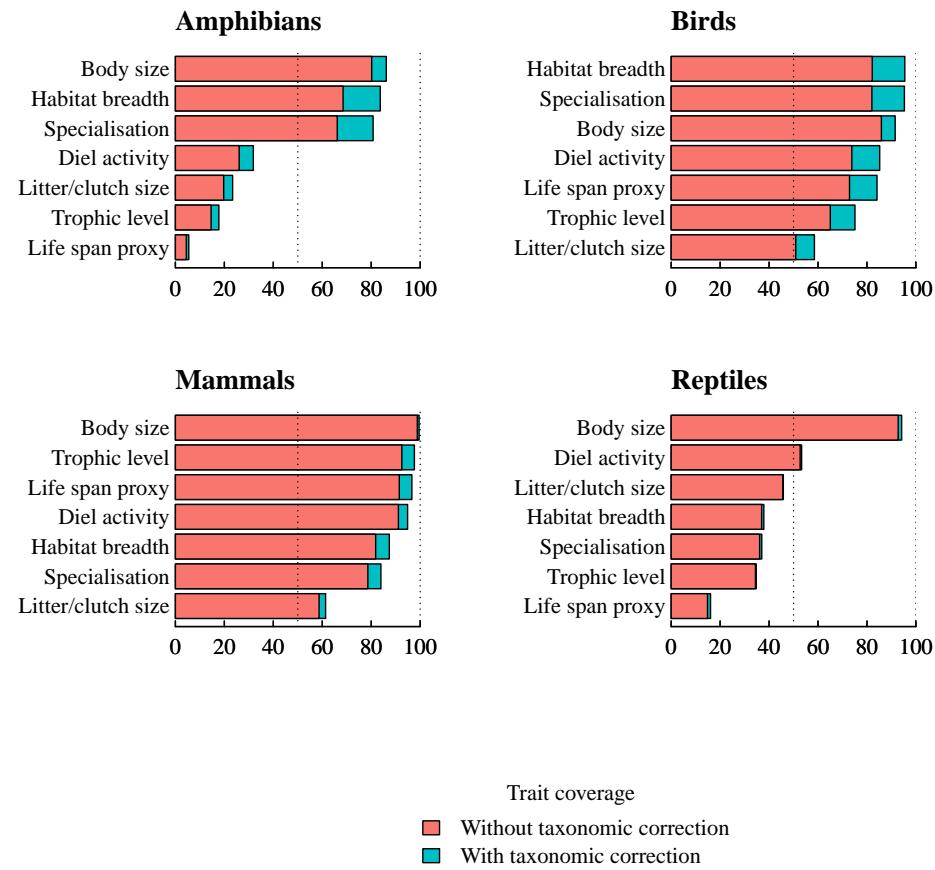


**Figure S2.4: Availability of altitudinal limits across species.** Upper and lower altitudinal limits were extracted from the IUCN Red List.



**Figure S2.5: Range sizes before versus after cutting by altitudinal limits.**

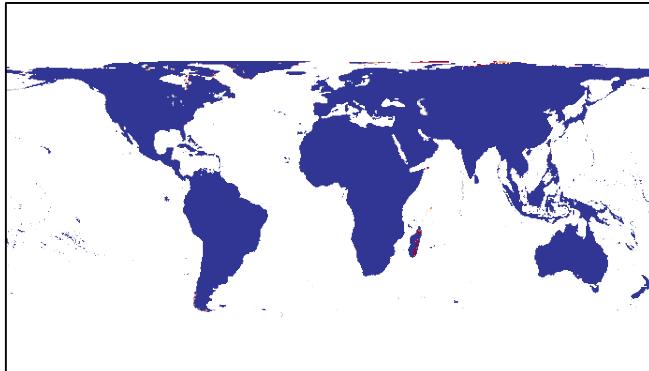
## S2.4 Impact of taxonomic corrections on trait coverage



**Figure S2.6: Trait coverage when no taxonomic correction is applied before matching sources by binomial names, versus when we applied the described procedure.** Identification of synonyms allowed to increase trait coverage in most cases.

## **S2.5 Assemblage-level median, mean and standard deviation of trait completeness (maps)**

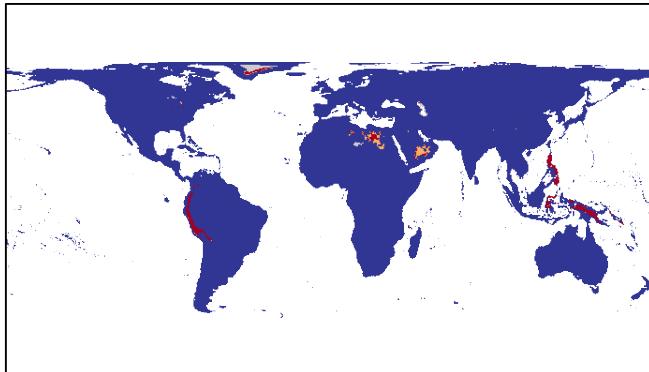
**(a) Mammals**



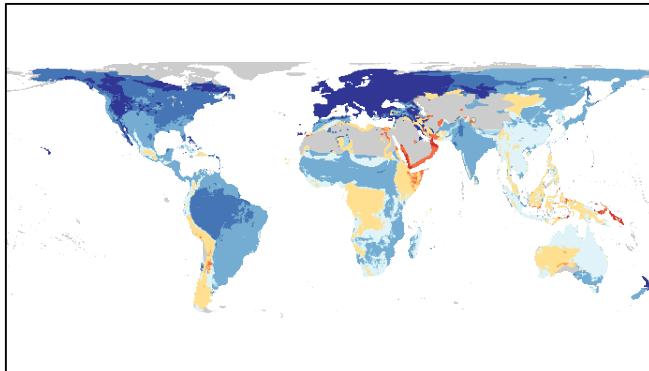
Median completeness  
(birds and mammals):

- NA
- 80 – 90%
- 90 – 91%
- 91 – 92%
- 92 – 93%
- 93 – 94%
- 94 – 95%
- 95 – 96%
- 96 – 97%
- 97 – 98%
- 98 – 99%
- 99 – 100%

**(b) Birds**



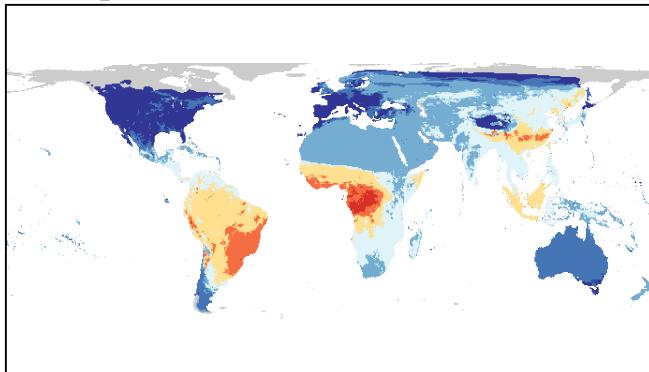
**(c) Amphibians**



Median completeness  
(herptiles):

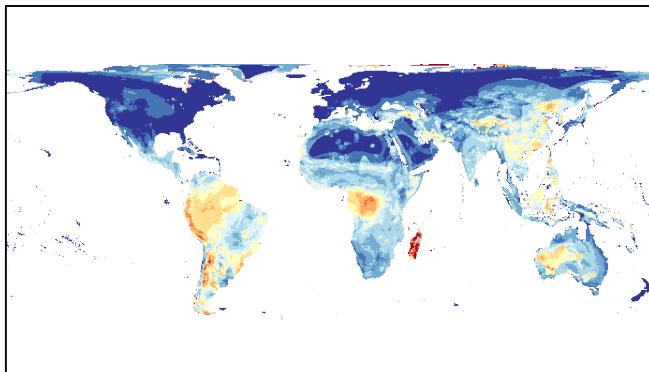
- NA
- 0 – 10%
- 10 – 20%
- 20 – 30%
- 30 – 40%
- 40 – 50%
- 50 – 60%
- 60 – 70%
- 70 – 80%
- 80 – 90%
- 90 – 100%

**(d) Reptiles**



**Figure S2.7: Spatial distribution of assemblage-level median trait completeness in each class.** Note that the color breaks differ for mammals and birds and for herptiles.

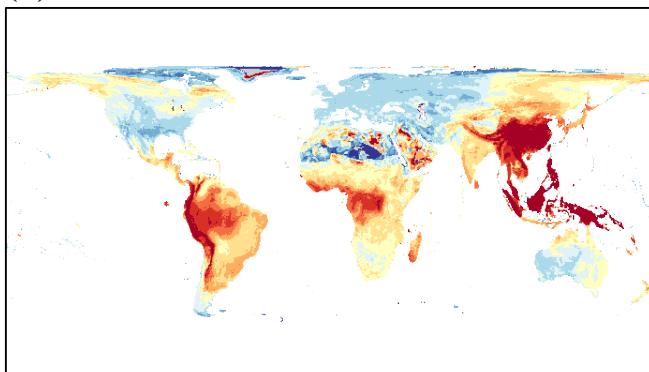
**(a) Mammals**



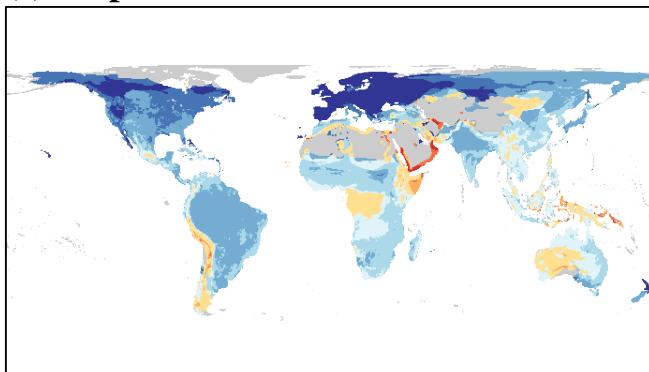
Mean completeness  
(birds and mammals):

- NA
- 80 – 90%
- 90 – 91%
- 91 – 92%
- 92 – 93%
- 93 – 94%
- 94 – 95%
- 95 – 96%
- 96 – 97%
- 97 – 98%
- 98 – 99%
- 99 – 100%

**(b) Birds**



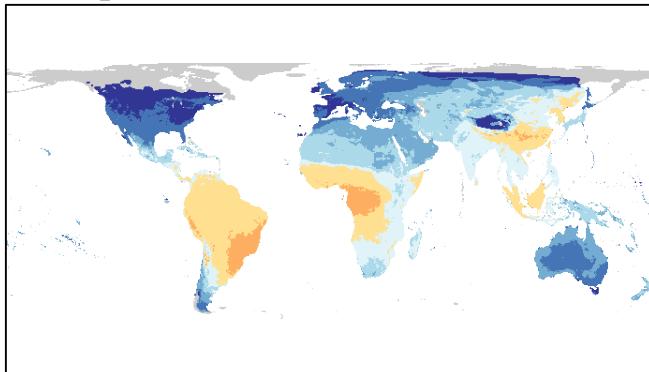
**(c) Amphibians**



Mean completeness  
(herptiles):

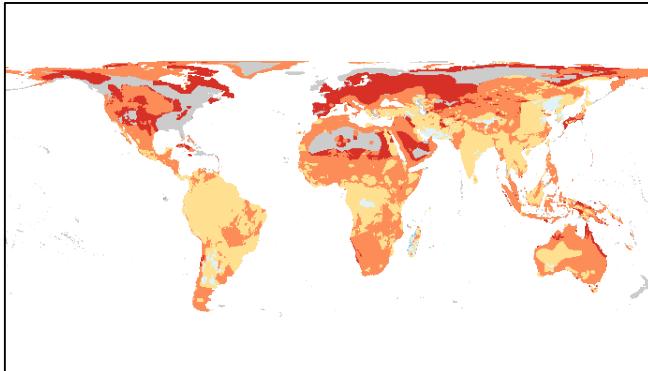
- NA
- 0 – 10%
- 10 – 20%
- 20 – 30%
- 30 – 40%
- 40 – 50%
- 50 – 60%
- 60 – 70%
- 70 – 80%
- 80 – 90%
- 90 – 100%

**(d) Reptiles**



**Figure S2.8:** Spatial distribution of assemblage-level mean trait completeness in each class. Note that the color breaks differ for mammals and birds and for herptiles.

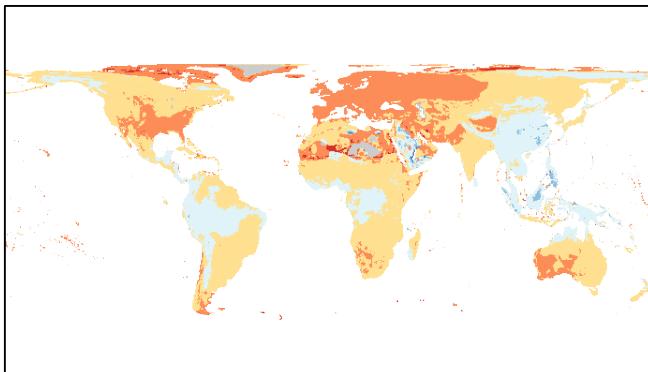
**(a) Mammals**



Standard deviation  
of completeness  
(birds and mammals):

- NA
- 0 – 5%
- 5 – 10%
- 10 – 15%
- 15 – 20%
- 20 – 25%
- 25 – 30%

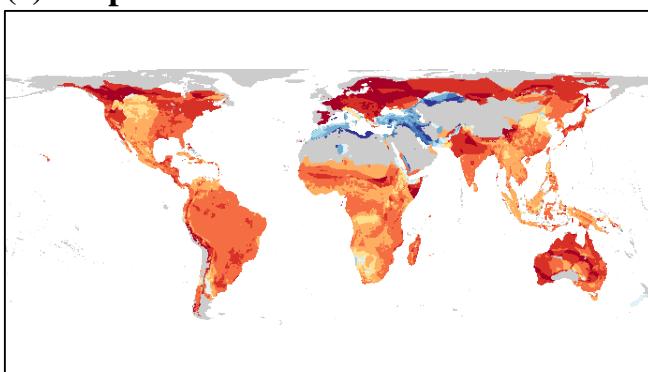
**(b) Birds**



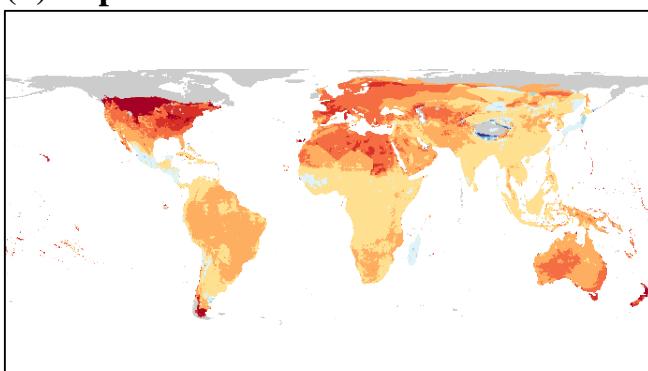
Standard deviation  
of completeness  
(herptiles):

- NA
- 0– 10%
- 10 – 15%
- 15 – 20%
- 20 – 25%
- 25 – 30%
- 30 – 35%
- 35 – 40%
- 40 – 50%
- 50 – 60%
- 60 – 75%

**(c) Amphibians**



**(d) Reptiles**



**Figure S2.9: Spatial distribution of assemblage-level standard deviation of trait completeness in each class.** Note that the color breaks differ for mammals and birds and for herptiles.

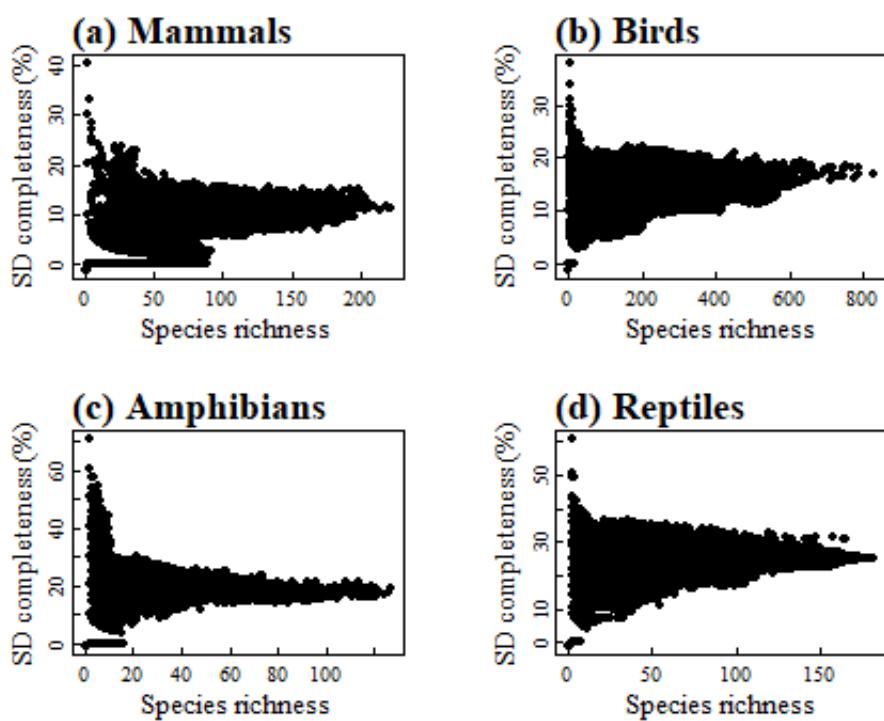
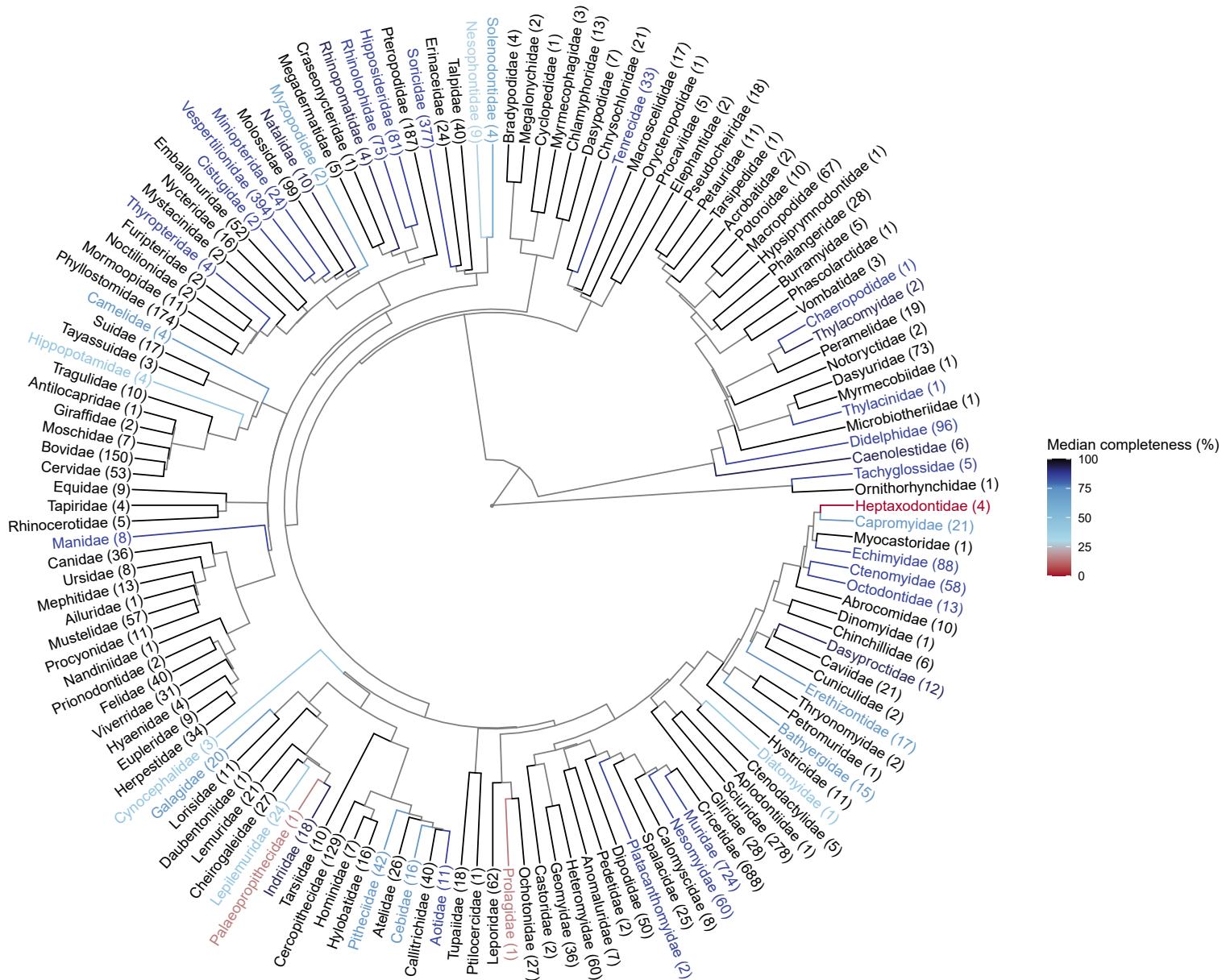
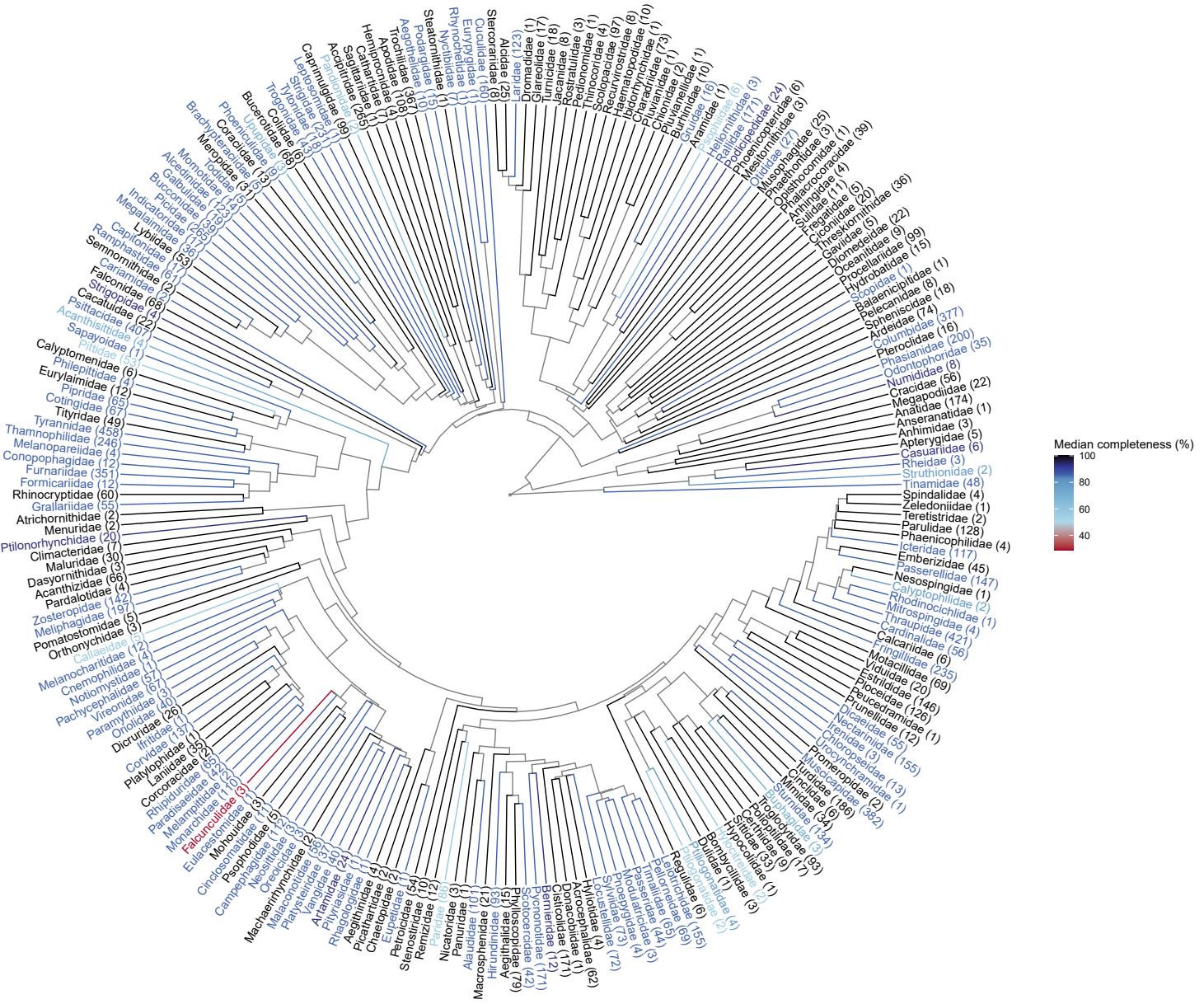


Figure S2.10: Assemblage-level species richness against standard deviation in completeness in each class.

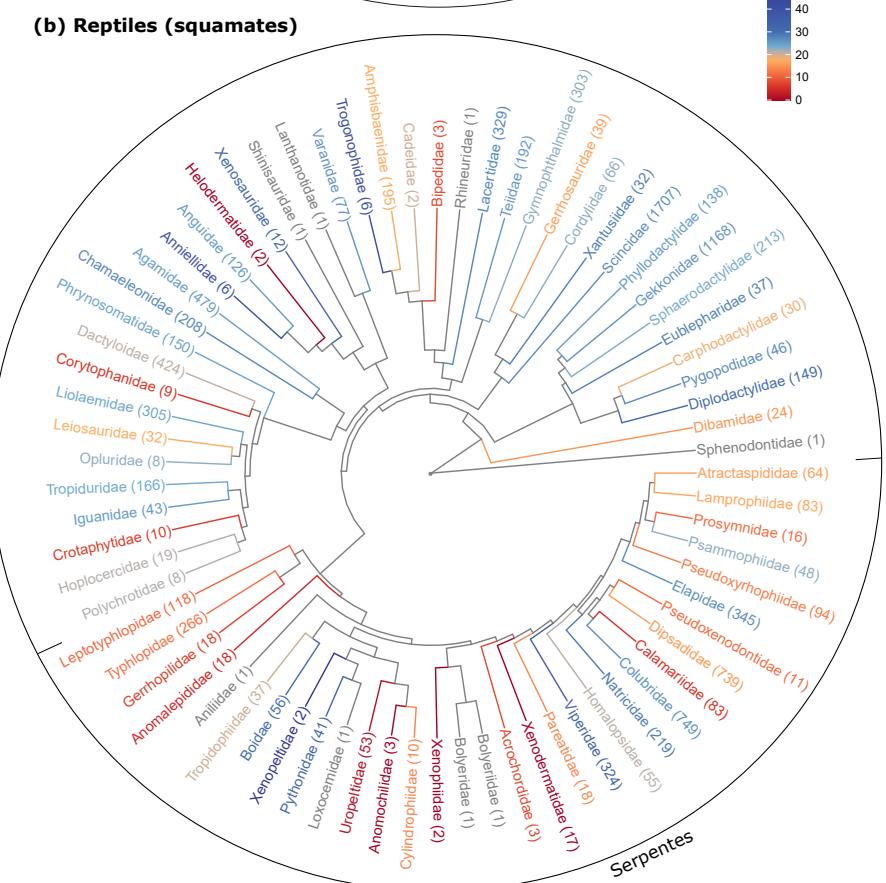
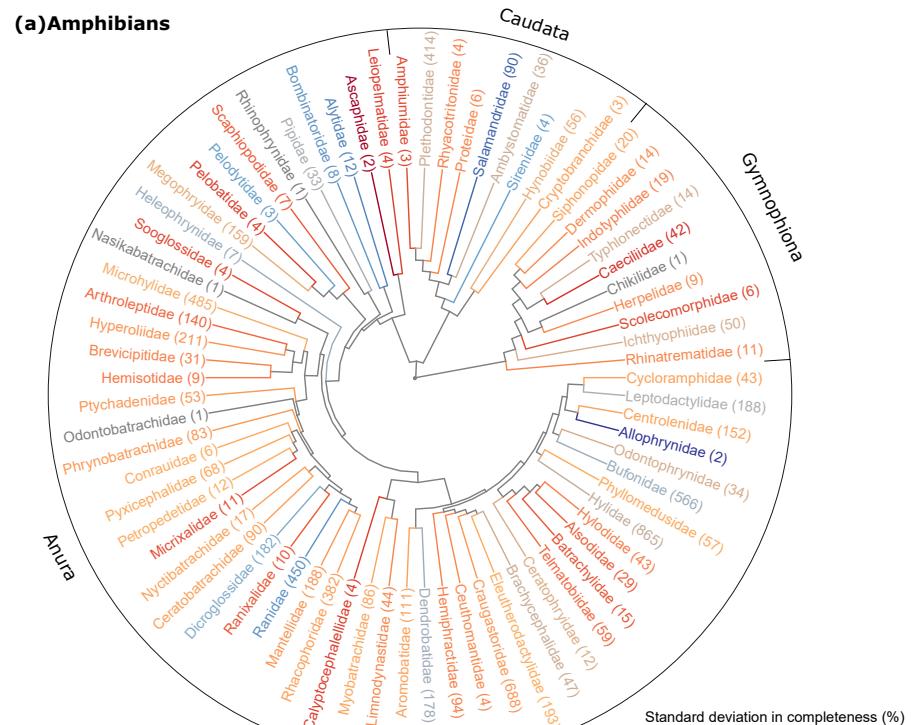
## S2.6 Phylogenetic patterns in trait completeness



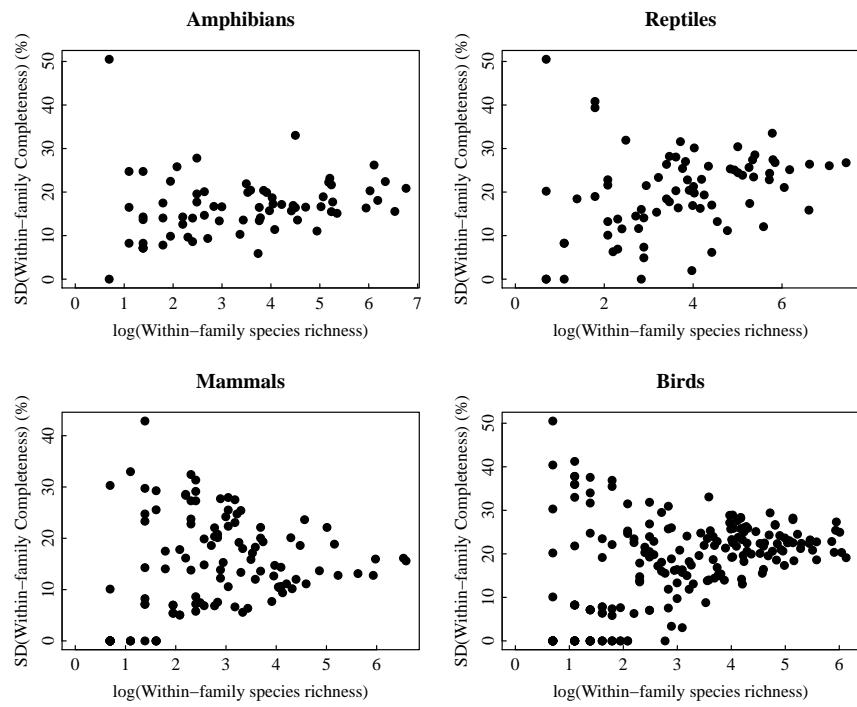
**Figure S2.11:** Within-family median trait completeness in mammals.



**Figure S2.12:** Within-family median trait completeness in birds.



**Figure S2.13:** Within-family standard deviation in completeness (herptiles).



**Figure S2.14: Within-family species richness against the within-family standard deviation of completeness.**

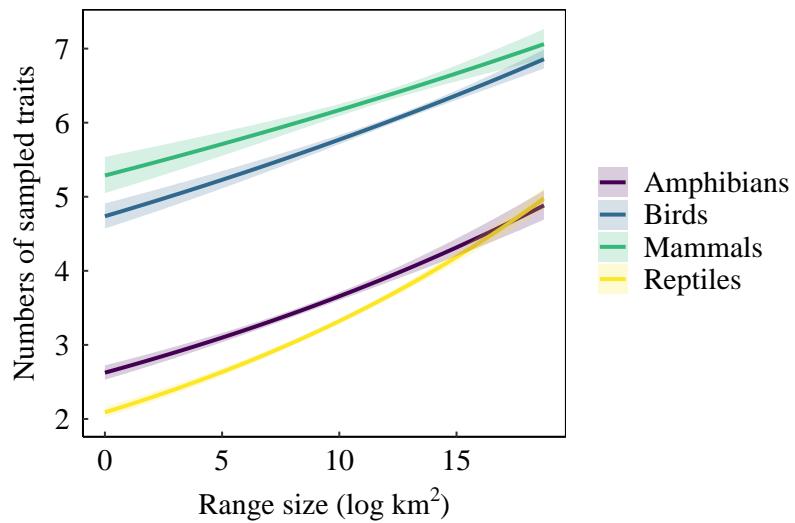
## S2.7 Model coefficients for Range size against Number of sampled traits (Poisson model).

**Table S2.1:** Coefficients of the model investigating whether species range size explained the number of sampled traits. Class was added as an interacting predictor. The reference level for class is mammals. The model was fitted using Poisson error distribution.

	Estimate	Std. Error	z value	Pr(> z )
Intercept	1.678	0.022	76.809	< 2e - 16
log Range Size	0.015	0.002	8.086	6.16e - 16
Class Birds	-0.092	0.028	-3.350	0.000809
Class Amphibians	-0.689	0.029	-24.099	< 2e - 16
Class Reptiles	-0.872	0.027	-31.856	< 2e - 16
log Range Size:Class Birds	0.003	0.002	1.415	0.157
log Range Size:Class Amphibians	0.017	0.003	6.427	1.30e - 10
log Range Size:Class Reptiles	0.026	0.002	11.159	< 2e - 16

**Table S2.2:** Coefficients of the model investigating whether species range size explained the number of sampled traits, *using range maps not cut by altitudinal limits*. Class was added as an interacting predictor. The reference level for class is mammals. The model was fitted using Poisson error distribution.

	Estimate	Std. Error	z value	Pr(> z )
Intercept	1.665	0.023	72.070	< 2e - 16
log Range Size	0.015	0.002	8.167	3.16e - 16
Class Birds	-0.110	0.029	-3.763	0.0002
Class Amphibians	-0.700	0.030	-23.721	< 2e - 16
Class Reptiles	-0.928	0.029	-32.403	< 2e - 16
log Range Size:Class Birds	0.004	0.002	1.840	0.066
log Range Size:Class Amphibians	0.018	0.003	6.564	5.24e - 11
log Range Size:Class Reptiles	0.031	0.002	12.630	< 2e - 16



**Figure S2.15: Relationship between number of sampled traits and geographical range size using distribution maps not cut by altitudinal limits.** Models were fitted using a Poisson error distribution. Class was added as a predictor interacting with range size. Rates of increase were not significantly different for mammals and birds, but differed for reptiles and amphibians, with steeper rates of increase for reptiles overall. Cutting range maps by altitudinal limits had little effects on the results (see Figure 2.5 in Main text).

## S2.8 Spatial models summaries

**Table S2.3: Spatial model summary for amphibians.** The spatial model was fitted to explain assemblage-level median completeness with species richness. Biogeographic realm was added as an interacting explanatory variable.

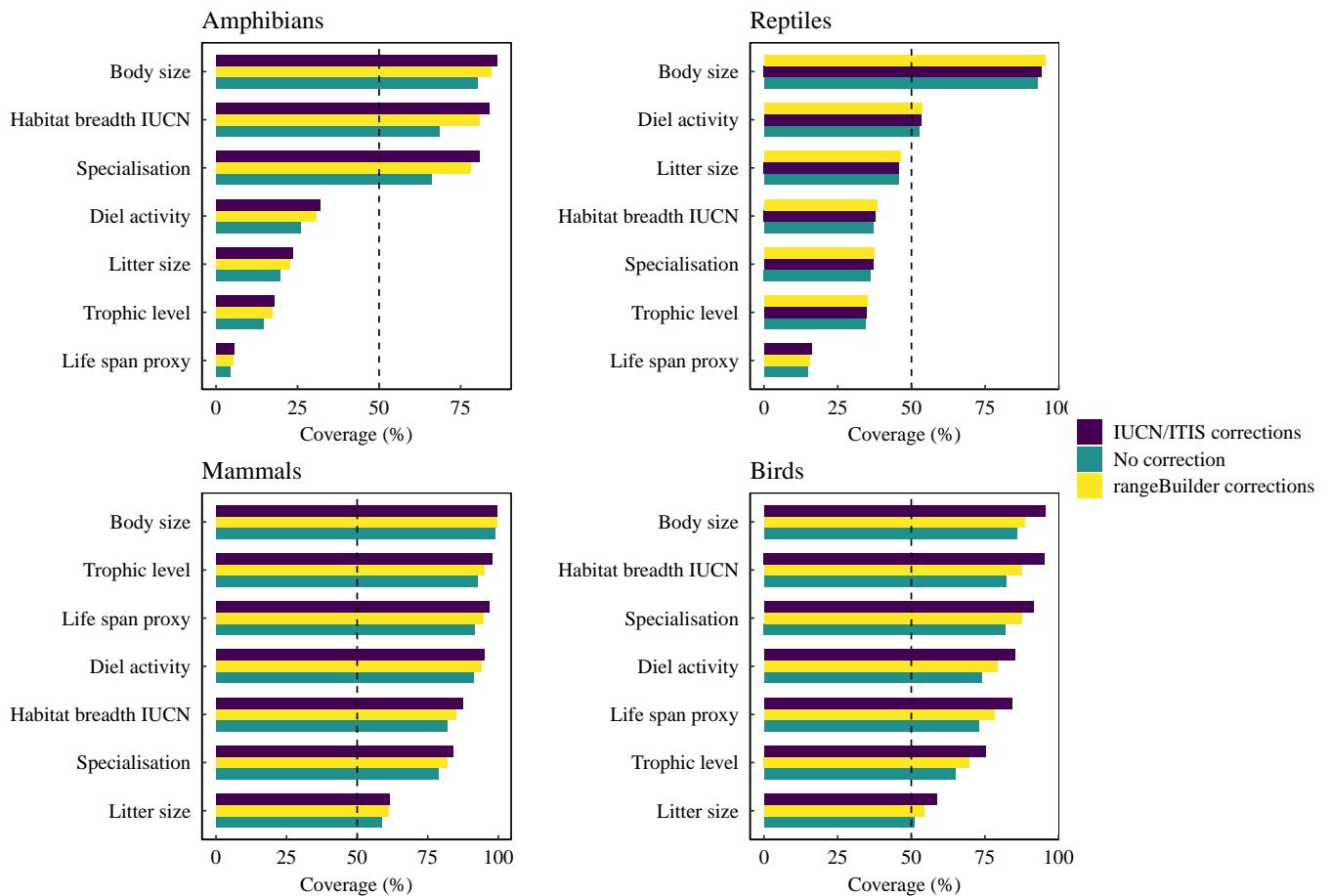
	Estimate	Std. Error	z value	Pr(> z )
Intercept - Realm: Afrotropic	0.0738	0.0064	11.4908	0
log(Species richness)	-0.0025	0.0017	-1.4261	0.1538
Realm: Australasia	-0.0109	0.0095	-1.1453	0.2521
Realm: Indo-Malay	0.0455	0.0119	3.8294	0.0001
Realm: Nearctic	0.0441	0.0082	5.3905	0.000000
Realm: Neotropic	-0.0377	0.0083	-4.5538	0.00001
Realm: Palearctic	0.0047	0.0067	0.6992	0.4844
log(Species richness):Australasia	0.0018	0.0038	0.4789	0.6320
log(Species richness):Indo-Malay	-0.0147	0.0039	-3.7294	0.0002
log(Species richness):Nearctic	-0.0097	0.0030	-3.2003	0.0014
log(Species richness):Neotropic	0.0144	0.0026	5.6454	0.000000
log(Species richness):Palearctic	0.0109	0.0029	3.7358	0.0002

**Table S2.4: Spatial model summary for reptiles.** The spatial model was fitted to explain assemblage-level median completeness with species richness. Biogeographic realm was added as an interacting explanatory variable.

	Estimate	Std. Error	z value	Pr(> z )
Intercept - Realm: Afrotropic	0.2001	0.0144	13.9349	0
log(Species richness)	-0.0316	0.0031	-10.0547	0
Realm: Australasia	-0.1284	0.0189	-6.7851	0
Realm: Indo-Malay	-0.0453	0.0263	-1.7215	0.0852
Realm: Nearctic	-0.0788	0.0140	-5.6366	0.000000
Realm: Neotropic	-0.0932	0.0145	-6.4425	0
Realm: Palearctic	-0.1030	0.0131	-7.8787	0
log(Species richness):Australasia	0.0386	0.0046	8.4019	0
log(Species richness):Indo-Malay	0.0124	0.0061	2.0397	0.0414
log(Species richness):Nearctic	0.0346	0.0038	9.1601	0
log(Species richness):Neotropic	0.0220	0.0034	6.4231	0
log(Species richness):Palearctic	0.0286	0.0033	8.6153	0

## S2.9 Trait coverage and taxonomic matching

Here, we briefly explore the robustness of our work to taxonomic uncertainty by comparing trait coverage obtained with our procedure for taxonomic matching against trait coverage obtained when extracting synonyms from class-specific sources, which could contain more information, notably for herptiles. We aligned taxonomy again using the rangeBuilder R package (**RangeBuilder**), which allows the extraction of accepted names from class-specific sources. Overall, our results are robust to the use of different taxonomic backbones; the main conclusions are likely to be unaffected by taxonomic uncertainty.



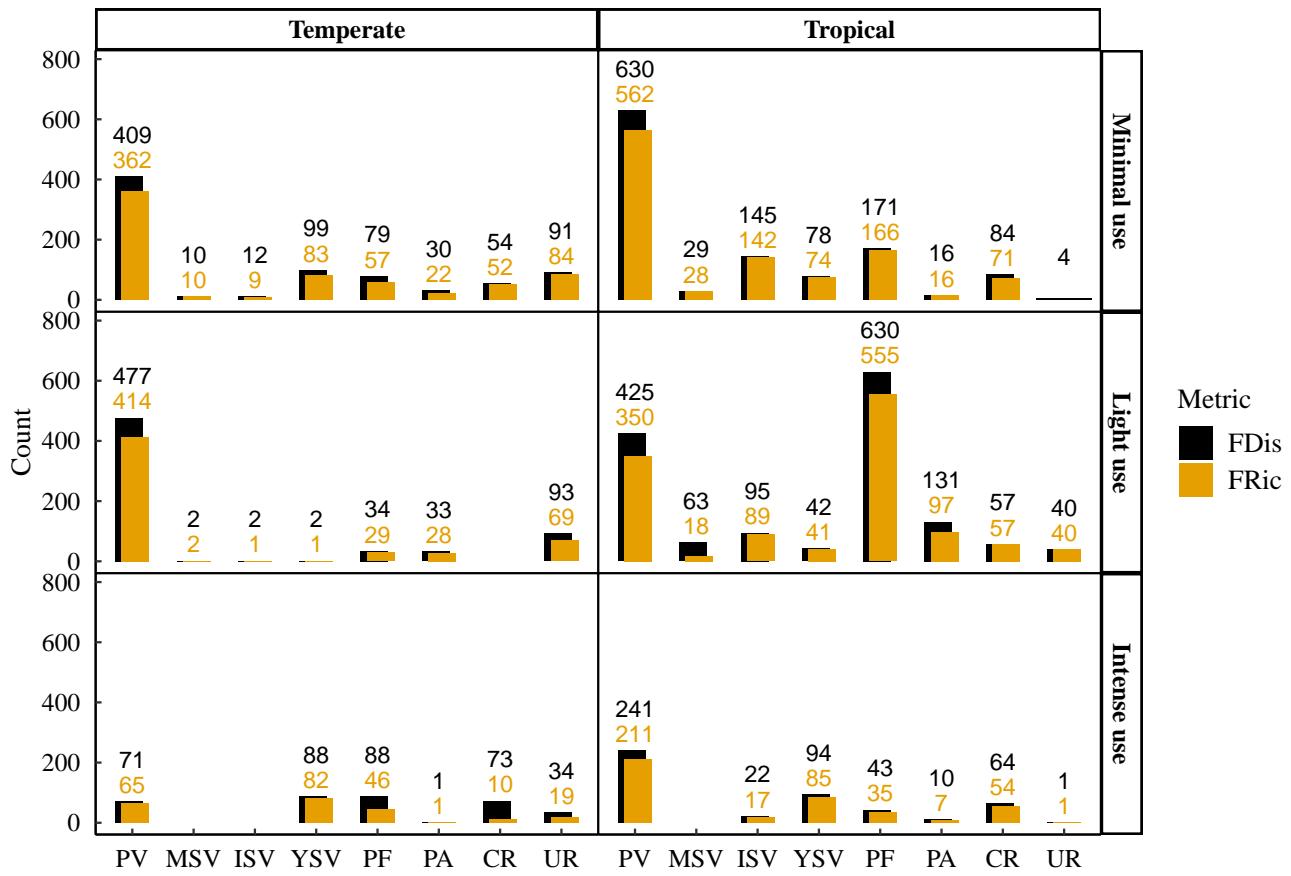
**Figure S2.16: Comparison of trait coverage among datasets corrected for taxonomy in different ways:** using the described procedure (purple bars), datasets corrected using the rangeBuilder package (extraction of synonyms from class-specific sources, yellow bars; **RangeBuilder**) and datasets where no taxonomic correction was applied when matching sources (green bars).

# Appendix 2: Supporting information for Chapter 3

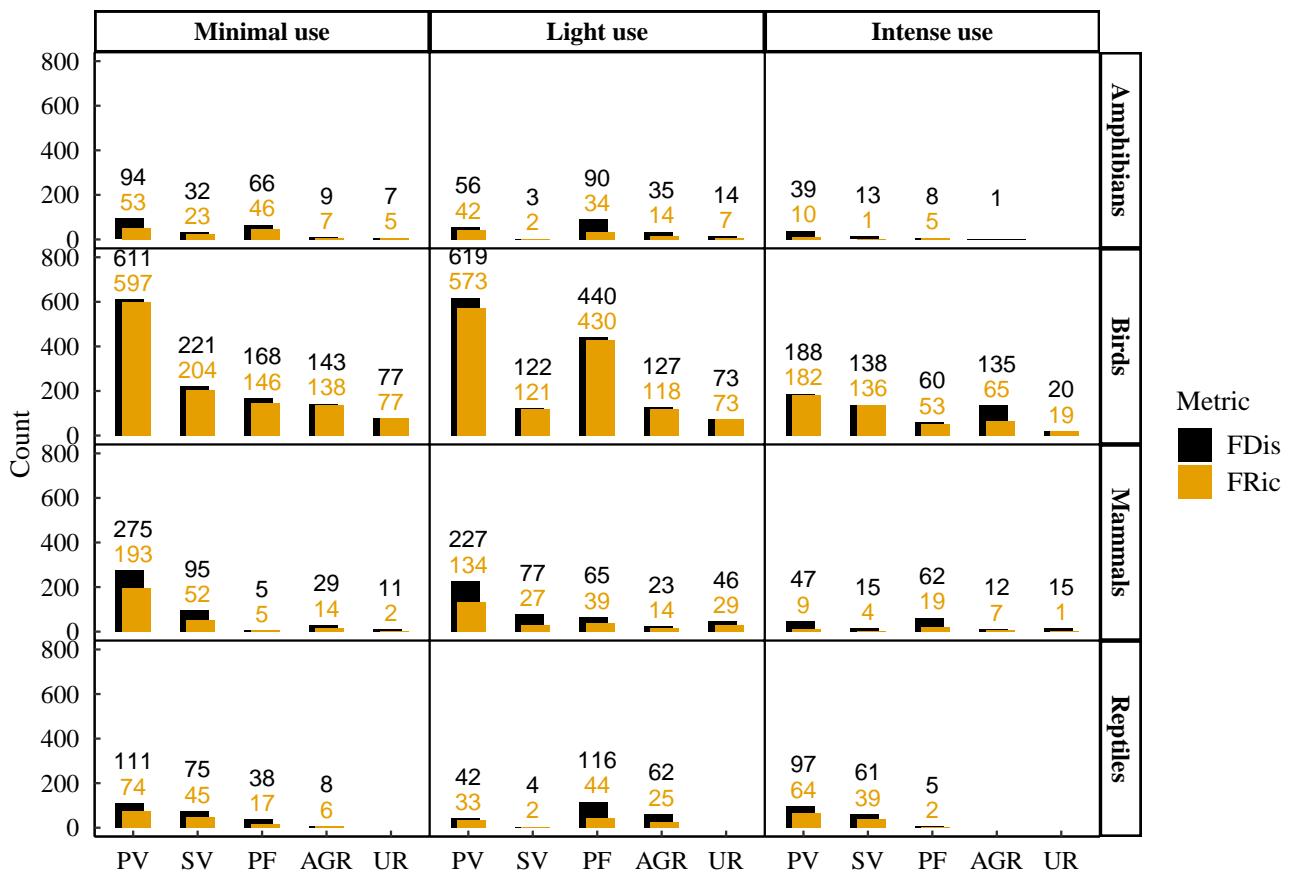
## S3.1 Land-use categories in PREDICTS and sample sizes

**Table S3.1: Land-use categories in the PREDICTS database.** See **Hudson2014; Hudson2017** for more details.

Land-use category	Definition
Primary vegetation	Native vegetation, undisturbed since its establishment under current climatic conditions. No known alterations due to human activities or to extreme natural events.
Mature secondary vegetation	Vegetation recovering after complete destruction of primary vegetation & where succession is near complete – the structure approaches that of primary vegetation.
Intermediate secondary vegetation	Vegetation recovering after complete destruction of primary vegetation at a mid-successional stage.
Young secondary vegetation	Vegetation recovering after complete destruction of primary vegetation at an early successional stage.
Plantation forest	Previously cleared areas planted with crop trees or shrubs grown and harvested for human consumption or for commercial purposes (includes wood, fruit, oil, biofuel, rubber, etc).
Pasture	Areas grazed by livestock, permanently or regularly. Can be improved through cultivation techniques.
Cropland	Previously cleared areas planted with herbaceous crops and harvested for human or animal consumption (including animal feed and crops used in the food industry), or for commercial purposes (e.g., crops grown for the textile industry).
Urban	Previously cleared areas built up by humans. Vegetation is managed for civic or personal purposes.



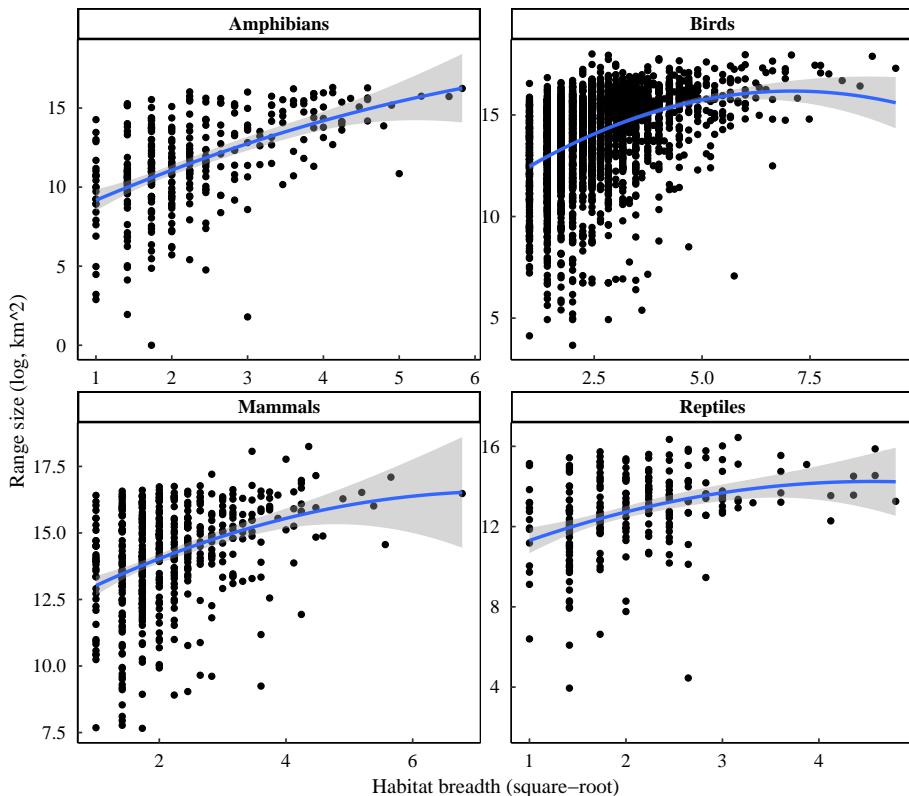
**Figure S3.1: Number of sites in each land use and land-use intensity for which FRic and FDis were calculated, across all species.** The number of sites for FRic can be smaller than the number of sites for FDis because FRic could not necessarily be computed for all assemblages in which FDis was estimated (in assemblages where species richness was 2 or 3, FRic could not be computed).



**Figure S3.2: Number of sites in each land use, land-use intensity and for which FRic and FDis were calculated, within each class.** The number of sites for FRic can be smaller than the number of sites for FDis because FRic could not necessarily be computed for all assemblages in which FDis was estimated (in assemblages where species richness was 2 or 3, FRic could not be computed).

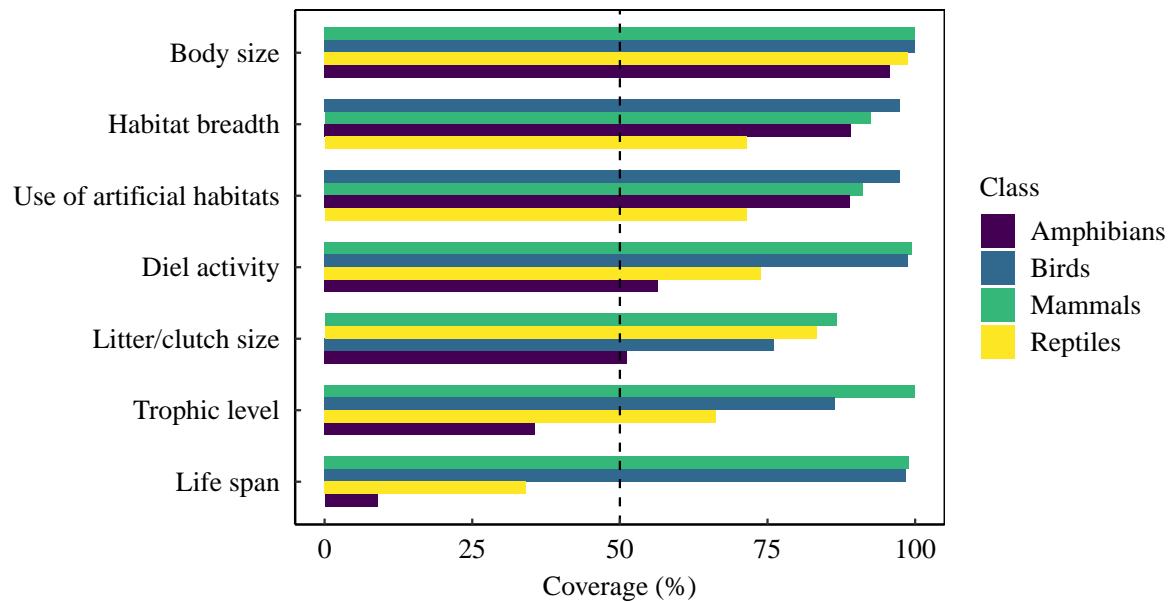
## S3.2 Trait data & imputation of missing trait values

I used the compilation of trait data across terrestrial vertebrates from Chapter 2 (and published in **Etard2020**). I selected seven traits that were available at least for a subset of the species in each class: body mass, trophic level, lifespan, litter/clutch size, diel activity, habitat breadth and use of artificial habitats. These traits related to life-history, habitat specialisation and use of geographical space (for instance, habitat breadth is a significant predictor of geographical range size in all classes, Figure ??).

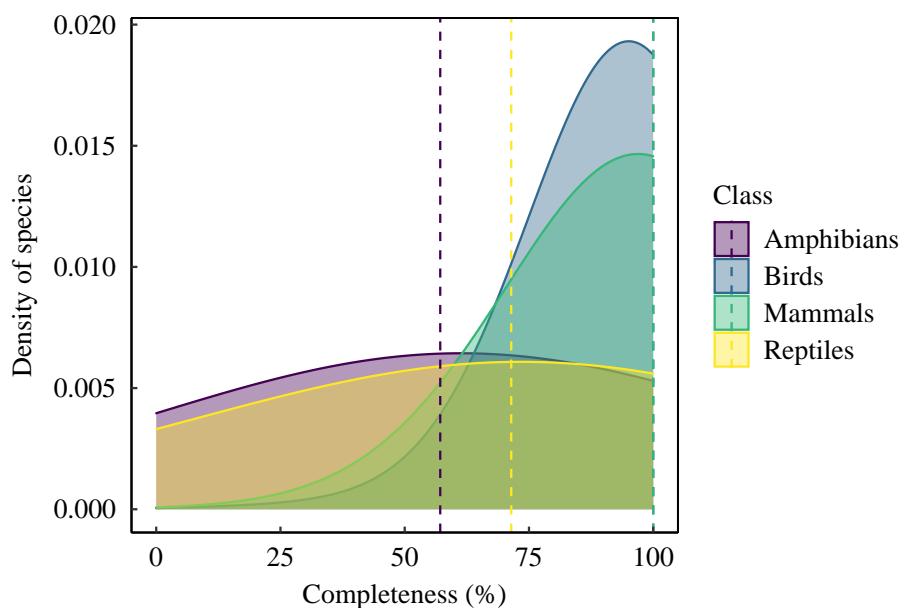


**Figure S3.3: Relationship between habitat breadth and geographical range size across species in each class.** The derivation of geographical range sizes is described in Chapter 2.

Trait-data coverage was highly variable among classes and traits, with important geographical and phylogenetic biases in trait data for reptiles and amphibians (Chapter 2; **Etard2020**; Figure ??, Figure ??). To obtain complete species-trait datasets, I imputed missing trait values. Further, in order to assess the sensitivity of our models to variation in imputed values, I imputed the missing trait values eight independent times. This allowed me to assess the congruence of the model predictions when using the different imputed trait datasets in the analyses.



**Figure S3.4: Trait coverage for the vertebrate species sampled in the PREDICTS database.** For a given trait, coverage is calculated as the percentage of species for which an estimate was available.



**Figure S3.5: Distribution of trait completeness across the vertebrate species sampled in the PREDICTS database.** For a given species, trait completeness is calculated as the proportion of traits for which an estimate was available. Dashed lines represent the median trait completeness.

### S3.2.1 Choice of imputation technique

There exist several imputation techniques (**Etard2020; Penone2014; Debastiani2021; Johnson2021**), such as K-nearest neighbour (**Troyanskaya2001**), multivariate imputation by chained equations (**micepackage**), random forest algorithms (implementable in R with missForest; **Stekhoven2012; Stekhoven2016**), and phylogenetic imputations (implementable in R with PhyloPars, **Bruggeman2009**). **Penone2014** assessed the performance of these four techniques and showed that missForest and PhyloPars performed better when traits were phylogenetically conserved, and when the species phylogenetic position was included as a predictor of missing trait values. PhyloPars can only handle continuous data, while missForest is compatible with mixed-type (including categorical) data. When no phylogenetic information was included, mice was found to be the best method, with fast imputations of mixed-type data (**Penone2014**). Therefore, to assess whether missForest or mice was more appropriate here, I measured the phylogenetic signal in trait data. For continuous traits, I used Pagel's  $\lambda$  (**Pagel1999**), and for categorical traits I used Borges'  $\delta$  (**Borges2018**). Strong phylogenetic signal would indicate that traits are phylogenetically conserved, and hence missForest would be the most suited approach for imputing missing trait values, with the inclusion of species' phylogenetic positions as a predictor.

### S3.2.2 Phylogenetic signal in traits

Across all classes, similar traits were used for calculating functional diversity metrics: body mass, litter/clutch size, lifespan (using different proxies in different vertebrate classes: generation length for birds and mammals, longevity for reptiles, and age at sexual maturity for amphibians), trophic level, diel activity, habitat breadth and use of artificial habitats. In addition, I included some class-specific traits for the imputations, as certain class-specific traits could be useful predictors of other traits (such as body length for instance in amphibians (**Santini2018**)). Table ?? details the traits that were included for the imputations in each class and the phylogenetic signal for each of these traits. Continuous traits were log-10 transformed before assessing Pagel's  $\lambda$  to improve normality. Pagel's  $\lambda$  was estimated using the phylosig function of the phytools package (**Revell2012**), and Borges'  $\delta$  was assessed using code provided by **Borges2018**, available at : [https://github.com/mrborges23/delta\\_statistic](https://github.com/mrborges23/delta_statistic). To test for the significance of  $\delta$ , I generated null distributions of  $\delta$  for each categorical trait by randomising trait vectors 50 times, and calculating  $\delta$  for each randomised vector – following the guidelines proposed by **Borges2018**. I then tested whether the observed medians were greater than the null distributions using one-sided Wilcoxon rank sum tests.

I used class-specific phylogenies to estimate phylogenetic signal, all downloaded on 13th April 2020. Trees from **Faurby2018; Faurby2020** were used for mammals (downloaded from <https://zenodo.org>.

<https://doi.org/record/3690867#.Xyc5wyhKhPZ>). For amphibians, birds and reptiles (squamates only), I downloaded trees from <https://data.vertlife.org/>. Trees were from **Jetz2012** for birds, from **Jetz2018** for amphibians and from **Tonini2016** for squamates. For each class, I downloaded a distribution of 1,000 trees, from which I obtained consensus trees to estimate phylogenetic signal (to that end, I used the TreeAnnotator programme of the BEAST software (**Bouckaert2014**)).

**Table S3.2: Phylogenetic signal in continuous and categorical traits.** BM: body mass; BL: body length; GL: generation length; MA: age at sexual maturity; ML: maximum longevity; L: longevity; LCS: litter/clutch size; HB: habitat breadth; TL: trophic level; DA: diel activity; UA: use of artificial habitats. Continuous traits were log-10 transformed to improve normality before estimating Pagel's  $\lambda$  – except for habitat breadth which was square-rooted. A star indicates a significant signal (p-value<0.05 for the log-likelihood ratio test in the case of  $\lambda$ ; and a significant difference from the simulated null distribution of  $\delta$  for categorical traits). 'NA' indicates traits that were not considered for a given class. All traits showed significant phylogenetic signal, with signals for BM, BL, L, GL, MA and LCS being particularly strong (above 0.8) across the four classes.

Class	Pagel's $\lambda$								Borges' $\delta$		
	BM	BL	GL	MA	ML	L	LCS	HB	TL	DA	UA
<b>Amphibians</b>	0.98*	0.94*	NA	0.85*	0.82*	NA	0.93*	0.99*	18*	3.4*	4.5*
<b>Birds</b>	0.99*	NA	0.97*	NA	NA	NA	0.95*	0.60*	13*	32e3*	1.8*
<b>Mammals</b>	0.99*	NA	0.97	NA	NA	NA	0.99	0.71	26*	17*	1.3*
<b>Reptiles</b>	1.0*	NA	NA	NA	0.94*	0.98*	1.0*	0.52*	6.3*	6.4*	1.4*

### S3.2.3 Implementation of missForest imputations

As phylogenetic signals were strong in many categorical and continuous traits (Table ??), I imputed missing trait values using random forest algorithms, as implemented in R with missForest (**Stekhoven2012**; **Stekhoven2016**). Another advantage of missForest was that, being a non-parametric approach, no prior assumption about data distribution was required. The data were therefore not transformed prior to imputations. In addition, **Penone2014** showed that including phylogenetic information did not decrease the accuracy of imputations for traits that were less phylogenetically conserved, such as habitat breadth in this work.

Phylogenetic relationships were included as additional predictors in the form of phylogenetic eigenvectors (**DinizFilho2012**), extracted from the phylogenies using the PVR package (**Santos2018**). Following **Penone2014**, I included the first ten phylogenetic eigenvectors as additional predictors of missing trait values in each class, enough to minimise imputation error. As not all species were represented in the phylogenies, I also added taxonomic order as a predictor for all species. All traits in Table ?? were included in the imputations. Tuning parameters of missForest were set to ten maximum iterations and to one hundred trees grown in each forest.

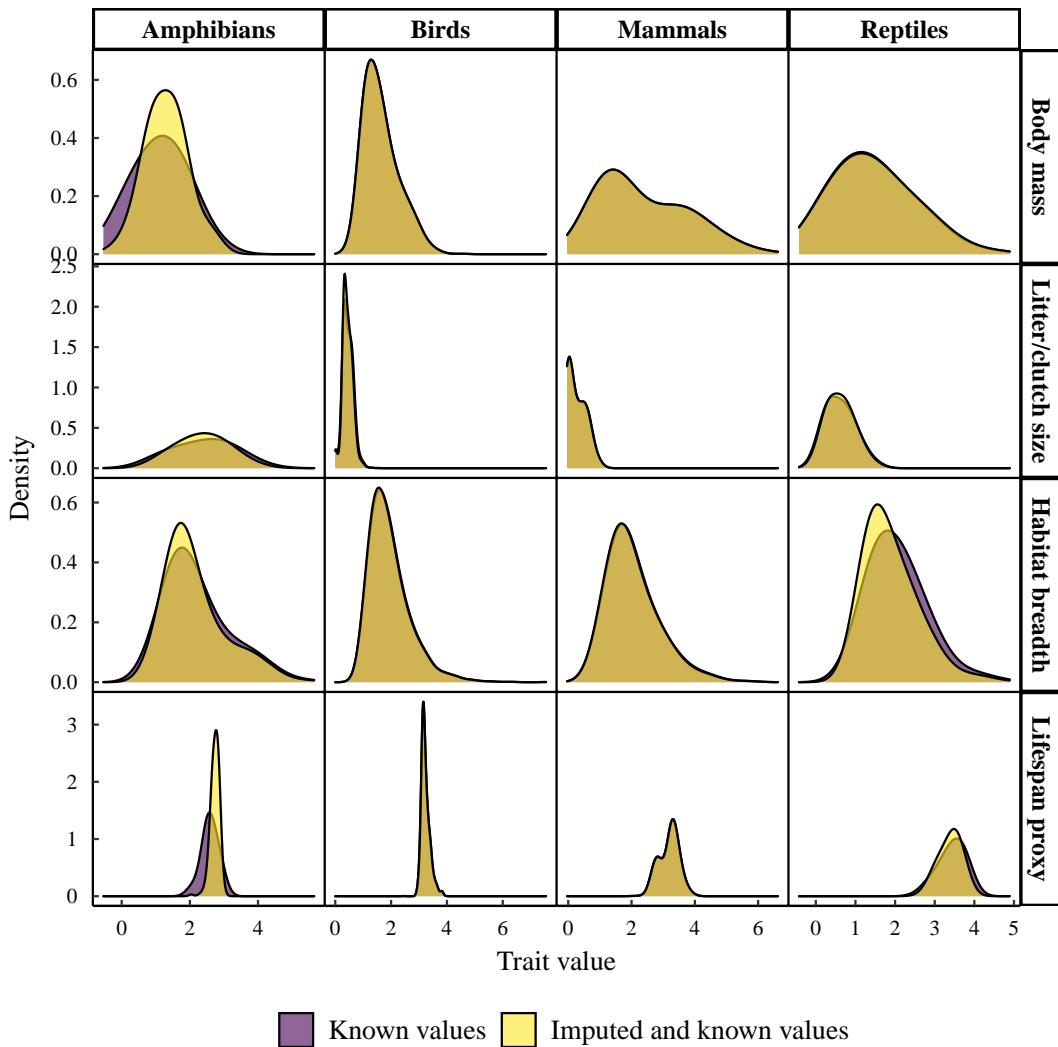
### S3.3 Degree of multicollinearity among traits

Multicollinearity among traits can be problematic when calculating functional diversity indices (**Cadotte2011**). After imputing missing trait values and before estimating functional metrics, I hence assessed whether the degree of multicollinearity among categorical and continuous traits was not problematically high. To that end, I used generalised variance inflation factors (**Fox1992**). Given a regression model, variance inflation factors quantify the overestimation in the variance of estimated regression coefficients due to multicollinearity among the predictors. A GVIF value of 5 or 10 is commonly used as a threshold to select out collinear predictors (**Dormann2013**). I used the stepwise.vif function of the Rnalytica package (<https://github.com/awsm-research/Rnalytica>), with a threshold of 5, to determine the GVIF of each trait. I used the imputed traits from the 8th imputation iteration to assess whether multicollinearity was problematically high. Continuous traits were log-10 transformed (except for habitat breadth which was square-rooted). Multicollinearity across traits was not detected to be problematically high, as all traits had a GVIF value below 2 (Table ??). As such, all seven traits were included in the calculation of functional diversity indices.

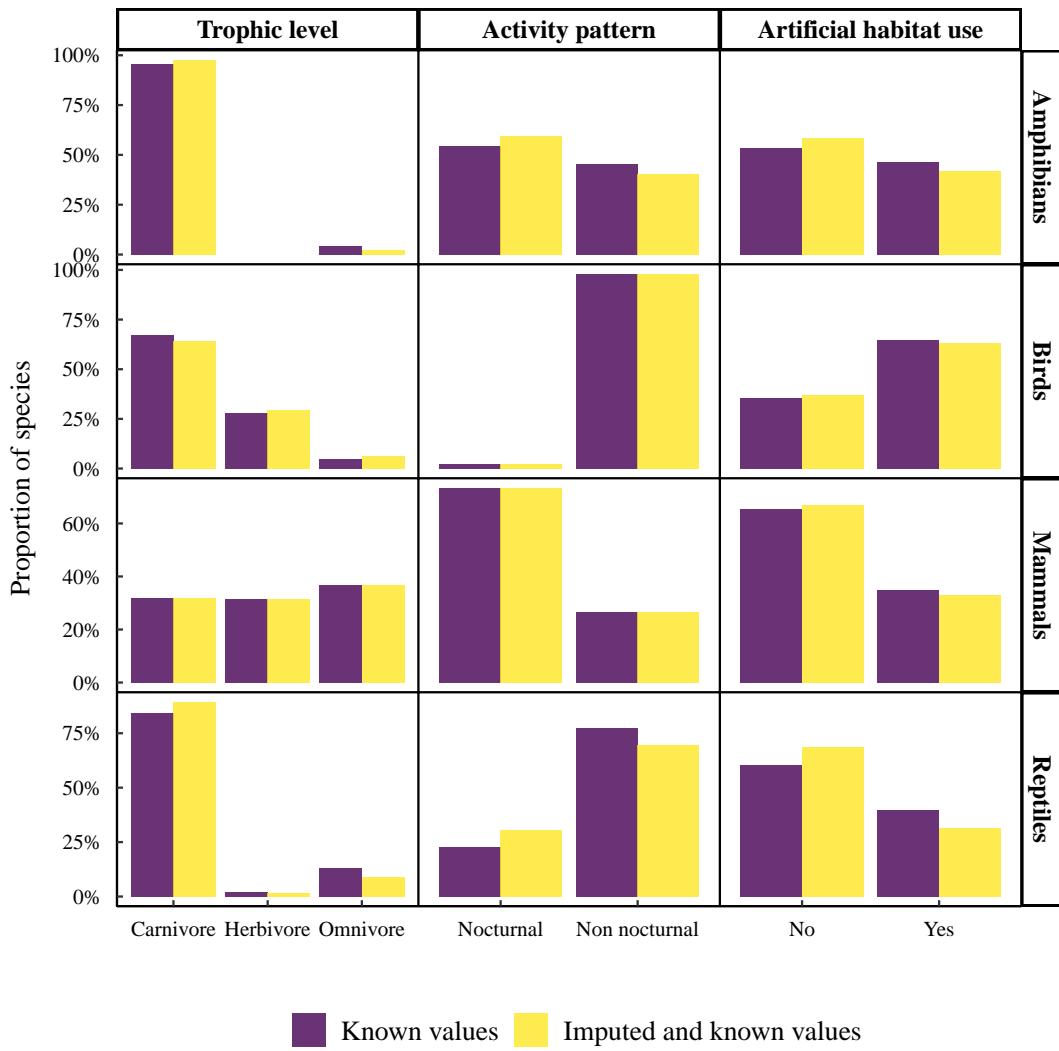
**Table S3.3:** Variance inflation factors across considered (imputed) traits.

Trait	GVIF
Diel activity	1.1
Trophic level	1.3
Use of artificial habitats	1.4
Body mass (log10)	1.5
Habitat breadth (square-root)	1.5
Litter/clutch size (log10)	1.6
Lifespan proxy (log10)	1.7

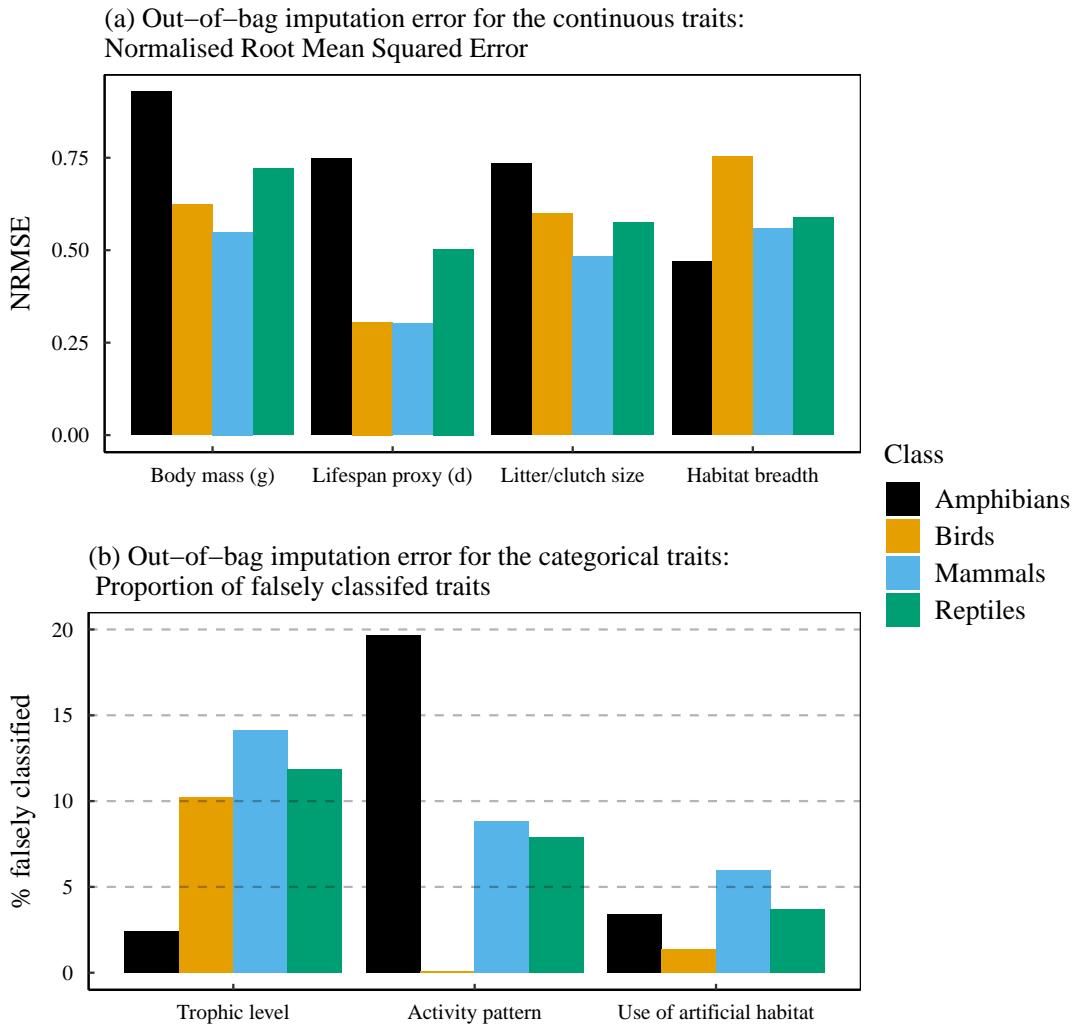
## S3.4 Imputation performance



**Figure S3.6:** Distribution of continuous traits considered in the calculation of the functional diversity metrics (shown as density plots), before and after missing value imputations, in each class and for the species occurring in the PREDICTS database. All traits were log<sub>10</sub>-transformed except Habitat breadth, which was square-rooted.



**Figure S3.7:** Frequency distribution for categorical traits considered in the calculation of the functional diversity metrics (shown as % of total species in each category) before and after missing value imputations, in each class, for the species occurring in the PREDICTS database.



**Figure S3.8: Out-of-bag imputation errors for the continuous traits (a) and categorical traits (b).** For continuous traits, the Normalised Root Mean Squared Error (NRMSE) was obtained by dividing the Mean Squared Error (MSE) by the variance of the known trait distribution, then square-rooting the result. The MSE was returned for each trait by the missForest function (missForest package, **Stekhoven2012; Stekhoven2016**) and corresponds to an out-of-bag error. For categorical traits, the error was estimated as the out-of-bag proportion of falsely classified traits.

### S3.5 Functional loss and functional gain

Across all vertebrates, I estimated functional loss and gain using 84 studies for the tropical subset and 39 studies for the temperate subset (51,514 and 30,470 pairwise comparisons between sites respectively, Table ??). Because of this large number of pairwise comparisons, I did not develop a null modelling approach (if I used 100 randomisations per pair of sites, I would need to compute functional loss and gain for more than 8 million pairs, which would be very computationally demanding). I grouped mature, intermediate and young secondary vegetation together in this analysis. I could not estimate the effects in all land uses (for instance, sample sizes for tropical urban sites were too small).

Within classes, I used 18 tropical studies and 1 temperate study for amphibians; 38 and 21 for birds (respectively); 28 and 9 for mammals; and 11 and 7 for reptiles. As sample sizes differed among pairs of land uses and use I was not able to estimate all effects, notably for the intensely-used land uses.

To calculate functional loss and functional gain, the Gower distance matrix was first subsetted to the species occurring in a given pair of sites (see Chapter 3, Methods, ‘Functional traits and diversity indices’). Cailliez corrections were applied when the distance matrix was not Euclidian (Cailliez corrections consist of applying the smallest positive constant to the distances so as to make them Euclidian (**Cailliez1983**); **ade4** R package (**ade4package**)). I then performed a principal coordinates analysis on the (corrected) Gower distance matrix, retaining the first two axes to reduce the computational load in the calculation of convex hulls. Sites that contained fewer than three functionally different species were excluded (the computation of a convex hull requiring more species in the assemblage than PCoA axes). Then I estimated the volume of trait space occupied by each assemblage of a given pair, as well as the volume of the shared trait space (intersection), from which I derived functional loss and functional gain.

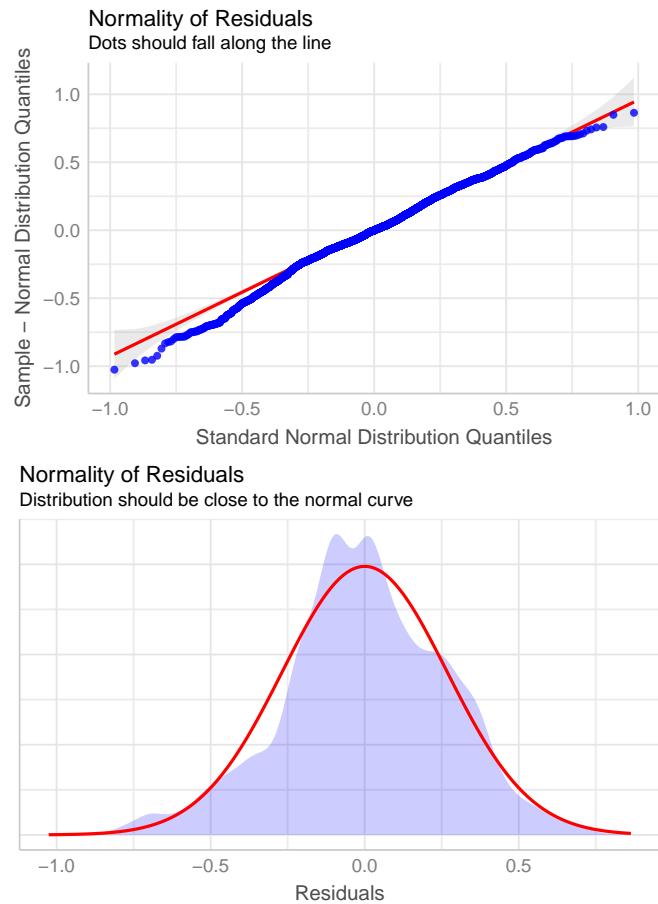
**Table S3.4: Sample sizes (number of pairs of sites) for the calculation of functional loss and functional gain across all species.**

Region	Pairs	Minimal use	Light use	Intense use
Temperate	PV-PV	7626	22546	492
	PV-SV	511	72	–
	PV-PF	9	166	–
	PV-PA	8	40	–
	PV-CR	150	–	–
	PV-UR	6306	1197	7
Tropical	PV-PV	8547	4016	16722
	PV-SV	6584	1124	9713
	PV-PF	580	1378	–
	PV-PA	36	20	22
	PV-CR	1700	1088	–
	PV-UR	–	–	–

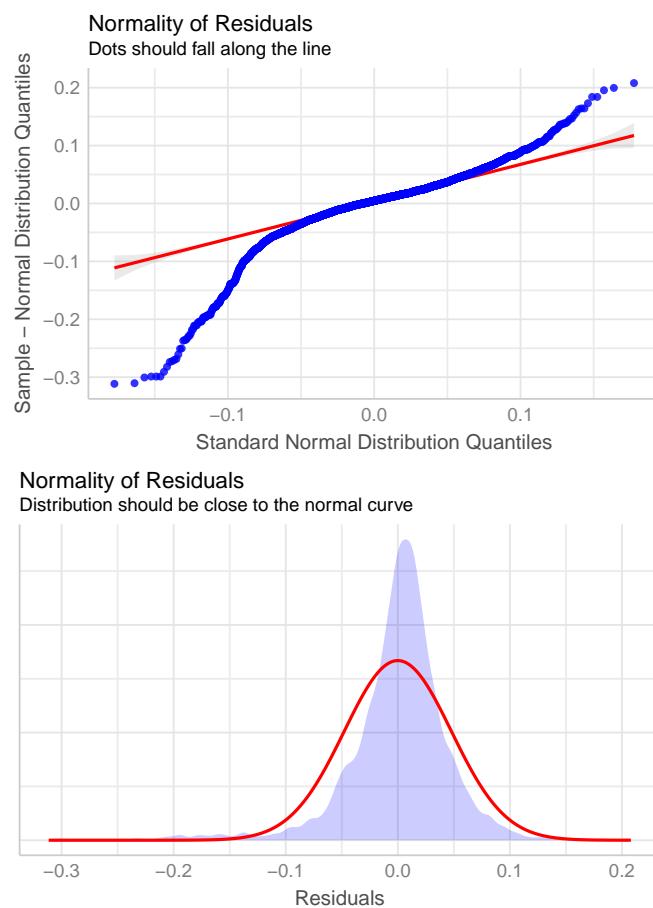
**Table S3.5: Sample sizes (number of pairs of sites) for the calculation of functional loss and functional gain within each class.**

Class	Region	Pair of land uses	Minimal use	Light use	Intense use
Amphibians	Temperate	PV/PV	–	45	–
Amphibians	Temperate	PV/SV	8	70	–
Amphibians	Temperate	PV/AGR	3	–	–
Amphibians	Temperate	PV/UR	6	100	–
Amphibians	Tropical	PV/PV	501	241	307
Amphibians	Tropical	PV/SV	838	–	90
Amphibians	Tropical	PV/PF	422	91	–
Amphibians	Tropical	PV/AGR	1	3	1
Birds	Temperate	PV/PV	7,382	19,300	491
Birds	Temperate	PV/SV	150	1	–
Birds	Temperate	PV/PF	9	166	–
Birds	Temperate	PV/AGR	145	40	–
Birds	Temperate	PV/UR	6,300	992	–
Birds	Tropical	PV/PV	5,059	3,117	9,014
Birds	Tropical	PV/SV	3,491	1,058	5,225
Birds	Tropical	PV/PF	156	994	–
Birds	Tropical	PV/AGR	1,626	1,085	–
Mammals	Temperate	PV/PV	110	3,030	–
Mammals	Temperate	PV/SV	25	–	–
Mammals	Temperate	PV/AGR	5	–	–
Mammals	Temperate	PV/UR	–	105	7
Mammals	Tropical	PV/PV	1,989	637	64
Mammals	Tropical	PV/SV	230	65	8
Mammals	Tropical	PV/PF	2	–	–
Mammals	Tropical	PV/AGR	109	20	21
Reptiles	Temperate	PV/PV	132	2	1
Reptiles	Temperate	PV/SV	250	1	–
Reptiles	Temperate	PV/AGR	5	–	–
Reptiles	Tropical	PV/PV	989	137	5,140
Reptiles	Tropical	PV/SV	1,760	1	3,456
Reptiles	Tropical	PV/PF	–	190	–

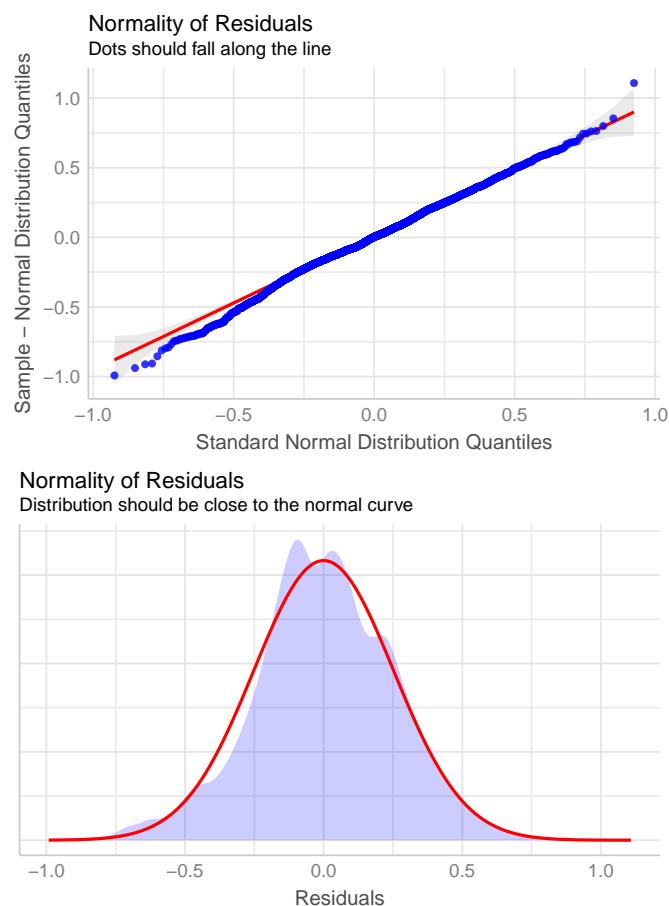
## S3.6 Diagnostic plots



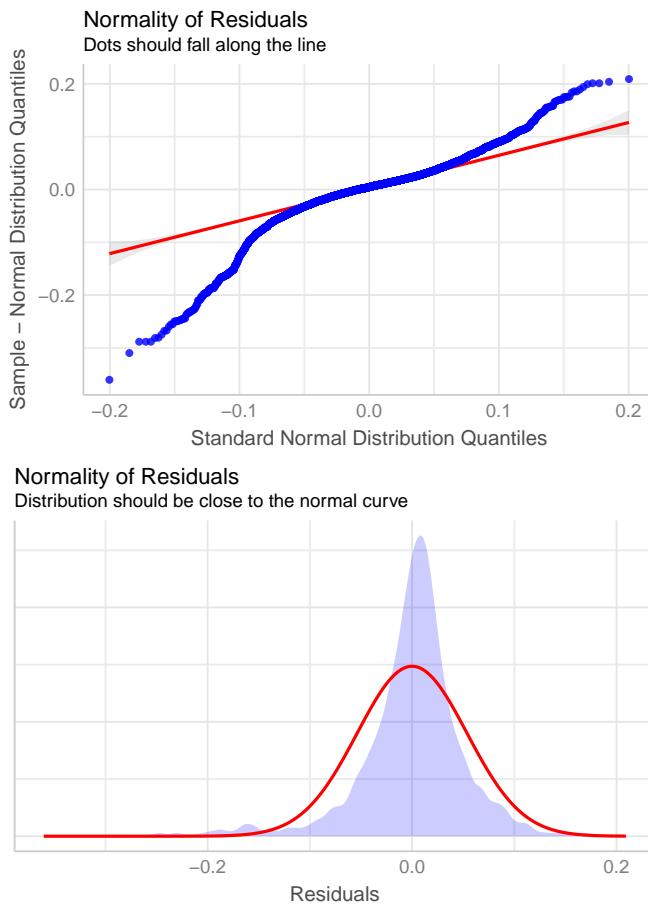
**Figure S3.9:** Diagnostic plots for Model 1a, obtained using the ‘performance’ R package (**performance**).



**Figure S3.10:** Diagnostic plots for Model 1b, obtained using the ‘performance’ R package (**performance**).

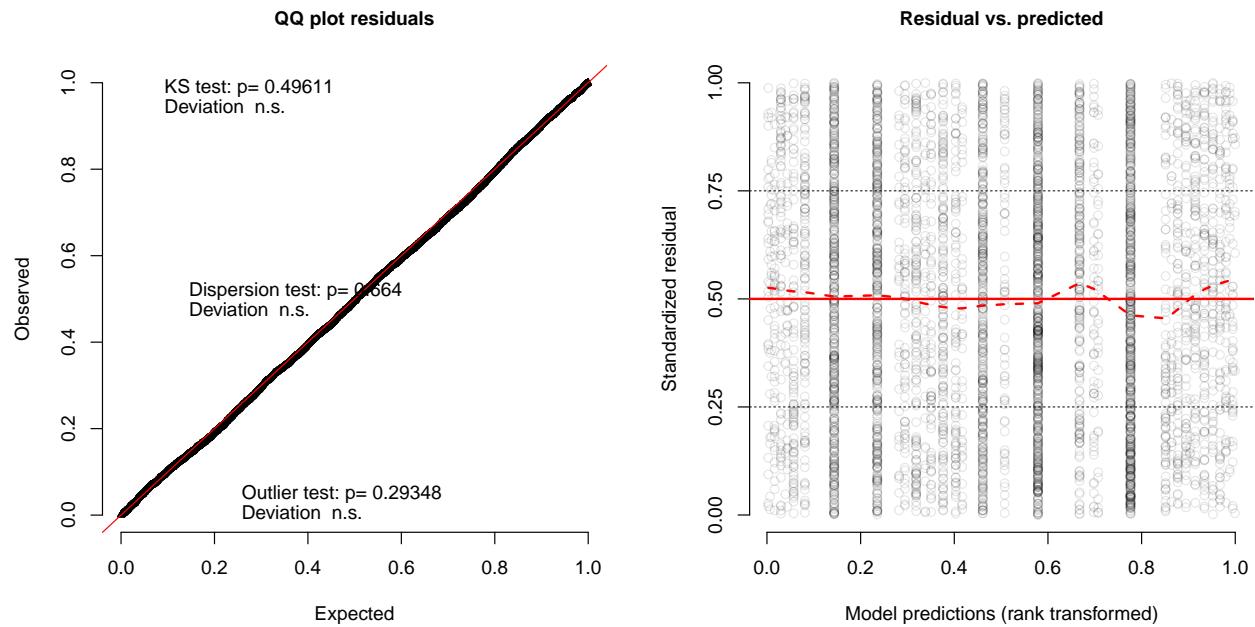


**Figure S3.11:** Diagnostic plots for Model 2a, obtained using the ‘performance’ R package (**performance**).

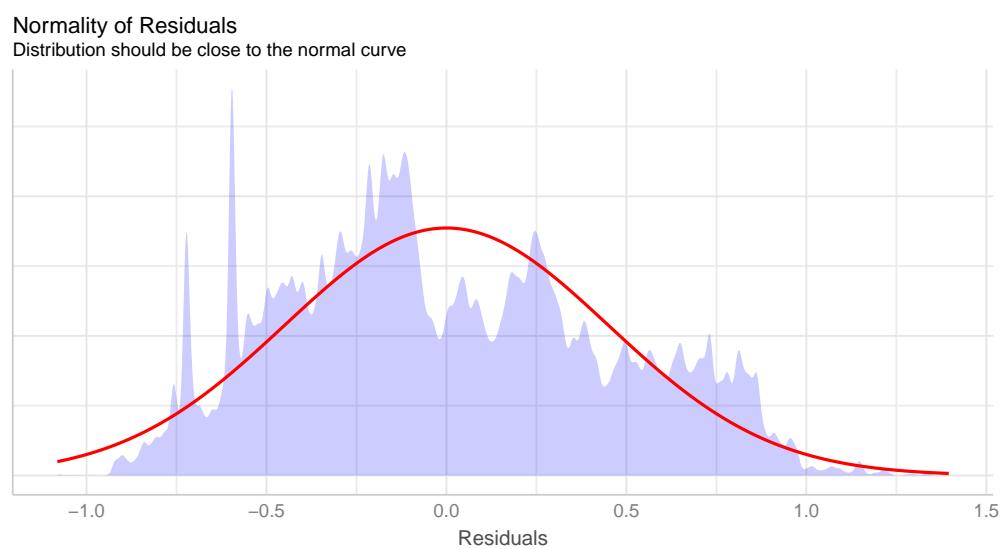
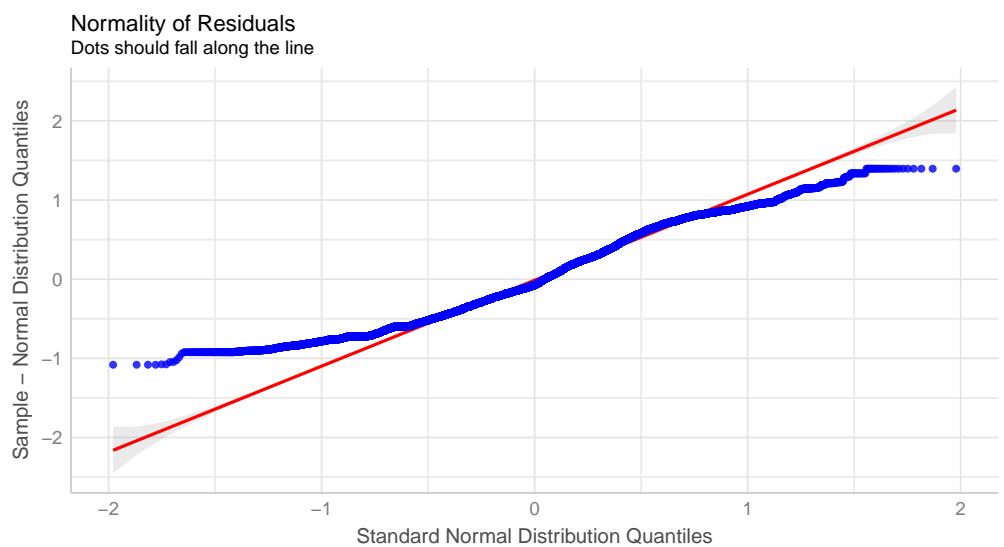


**Figure S3.12:** Diagnostic plots for Model 2b, obtained using the ‘performance’ R package (**performance**).

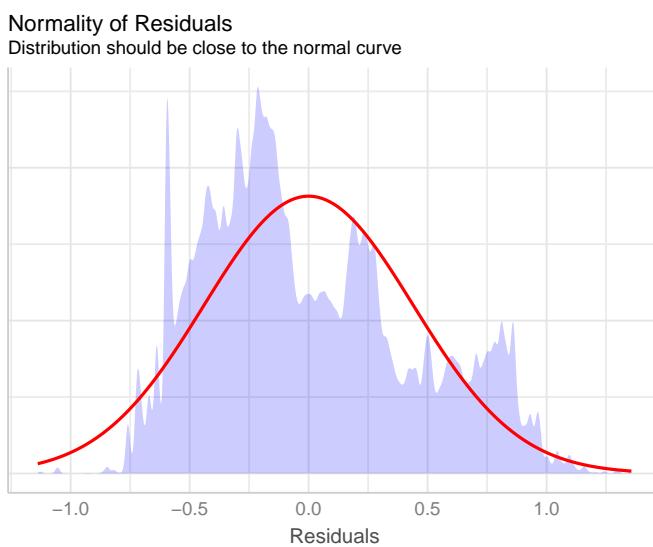
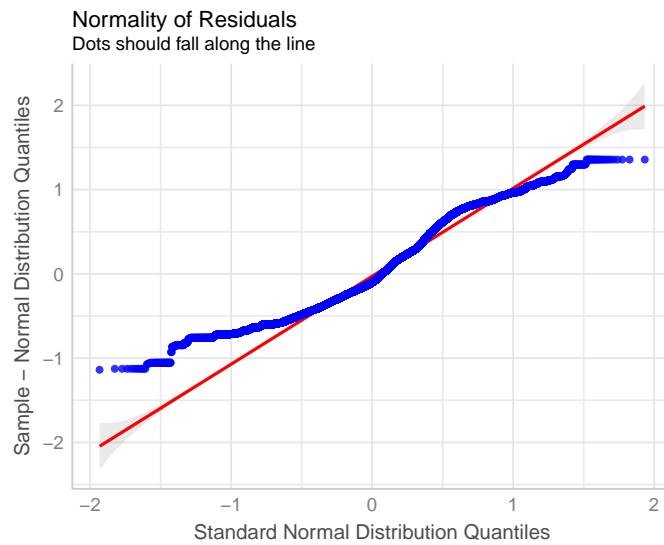
DHARMA residual diagnostics



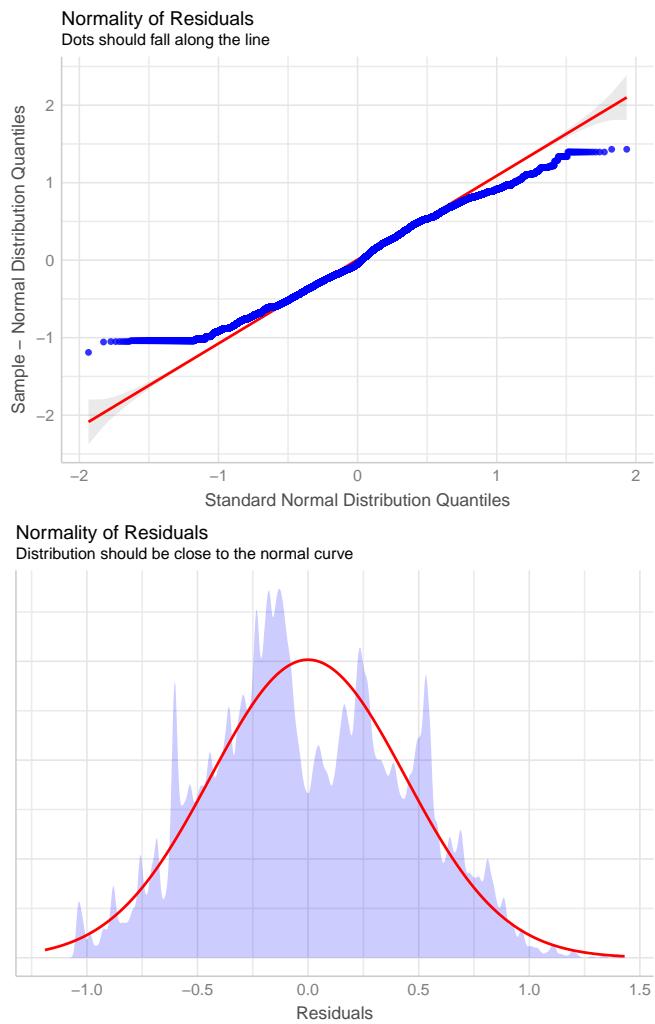
**Figure S3.13: Diagnostic plots for Model 3**, obtained using the ‘DHARMA’ R package (**DHARMA**).



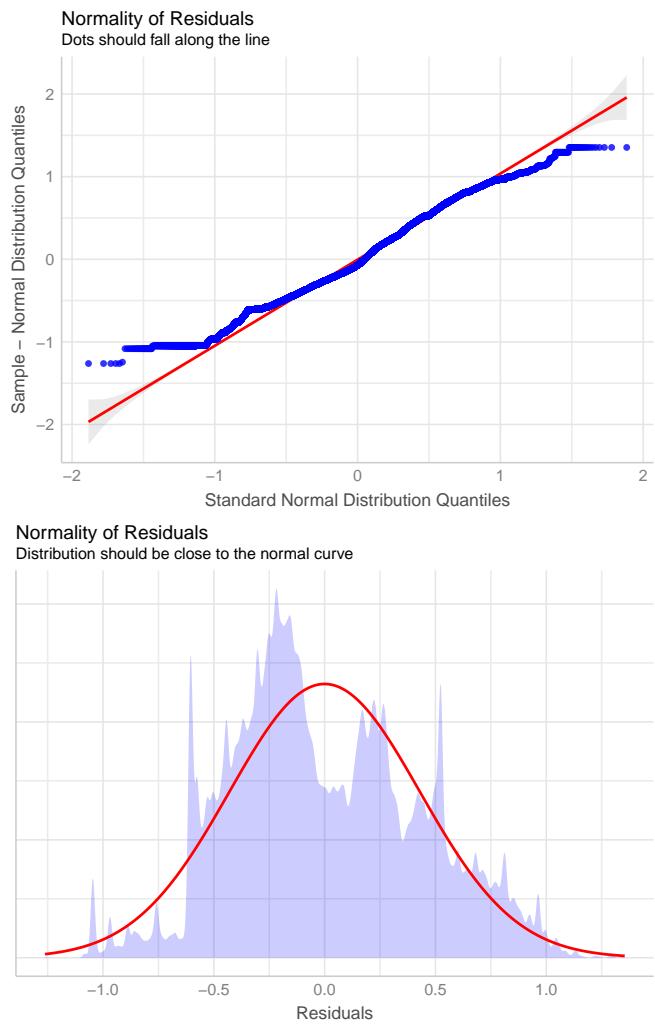
**Figure S3.14:** Diagnostic plots for **Model 4a**, obtained using the ‘**performance**’ R package (**performance**).



**Figure S3.15:** Diagnostic plots for **Model 4b**, obtained using the ‘**performance**’ R package (**performance**).

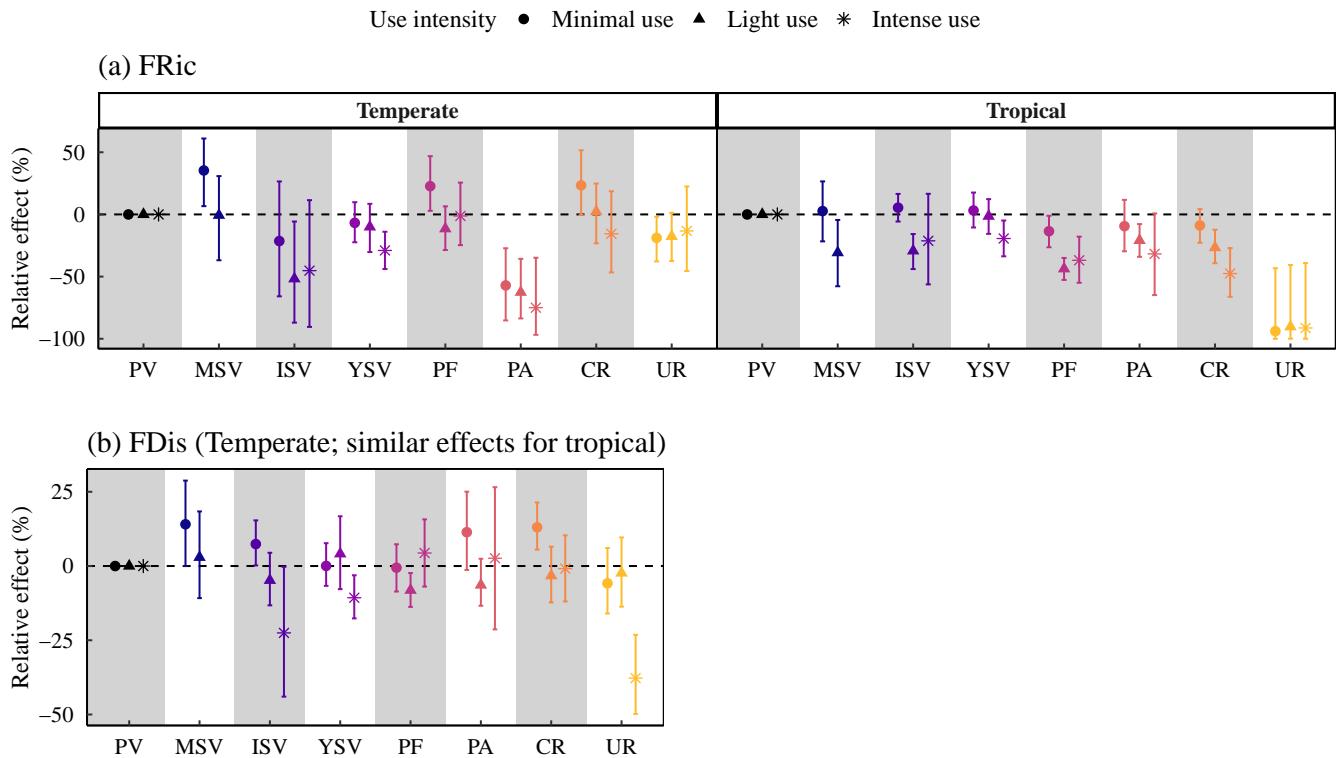


**Figure S3.16:** Diagnostic plots for **Model 5a**, obtained using the ‘**performance**’ R package (**performance**).

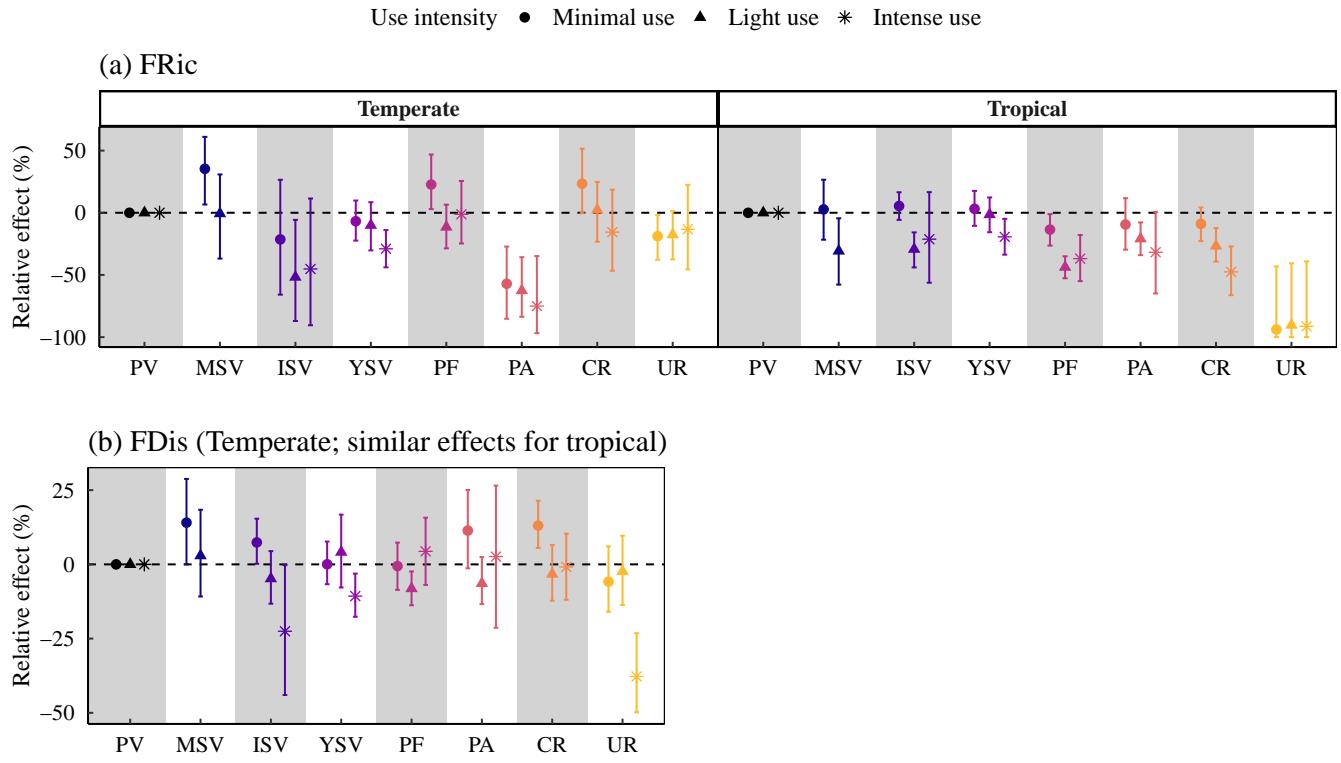


**Figure S3.17:** Diagnostic plots for **Model 5b**, obtained using the ‘**performance**’ R package (**performance**).

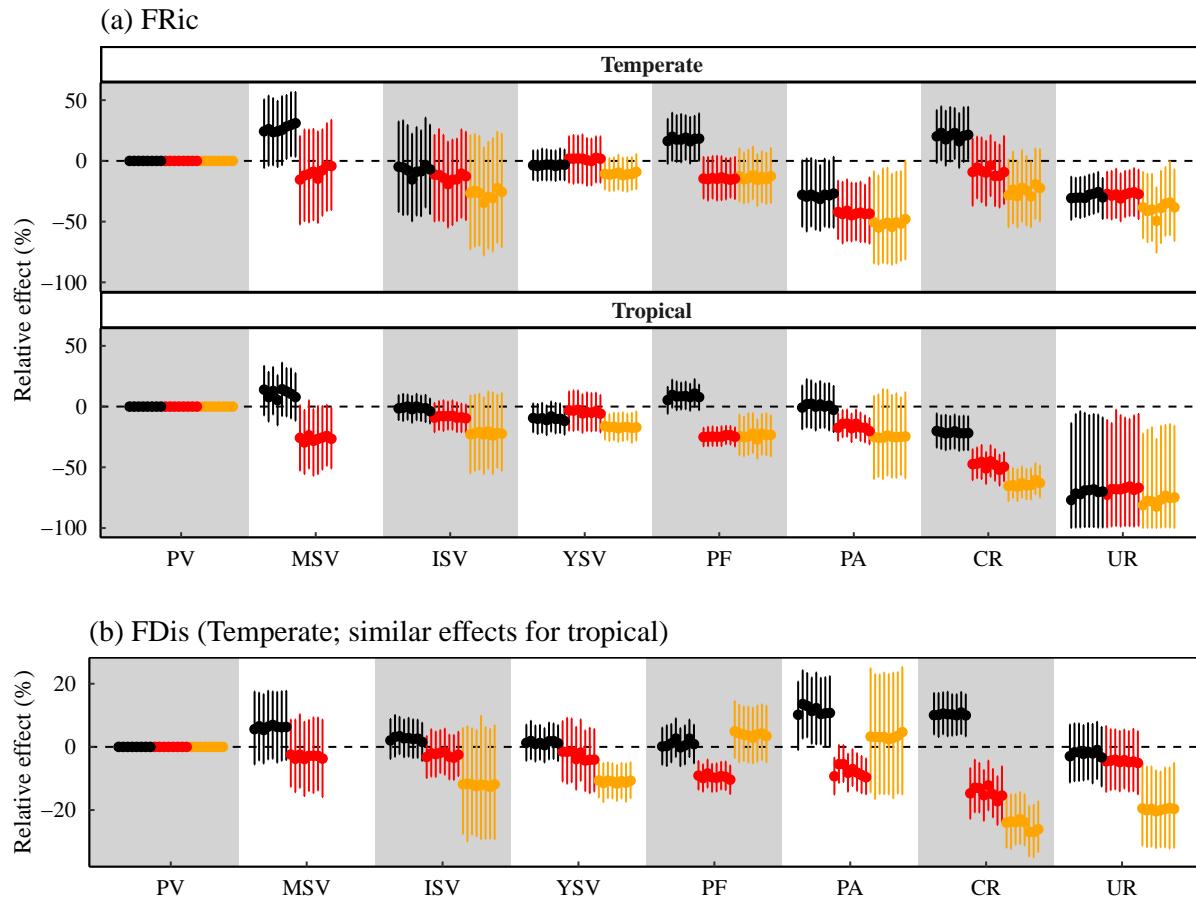
### S3.7 Model robustness



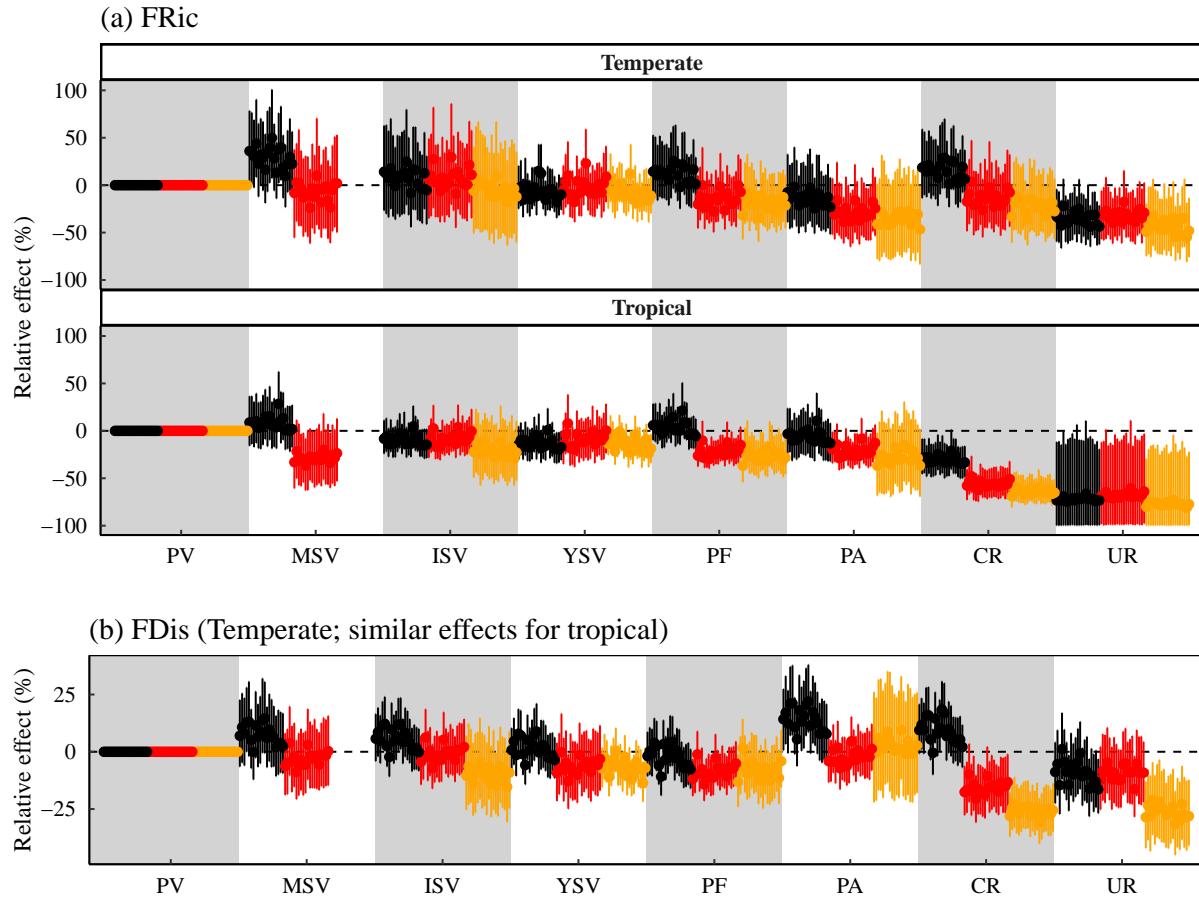
**Figure S3.18: Effects of land use, land-use intensity and region on FRic (a) and FDis (b) across vertebrates, for the subset of species with complete trait data (i.e., excluding species with less than 100% trait completeness).** Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, I fitted Model 1a (see main text, Chapter 3), which included the effects of land use, land-use intensity and region, as well as interactions between land use and land-use intensity and between land use and region. For FDis, I fitted Model 1b, which included effects of land use, land-use intensity and region, and interactions between land use and land-use intensity. Error bars represent 95% confidence intervals. MSV: mature secondary vegetation; ISV: intermediate secondary vegetation; YSV: young secondary vegetation; PF: plantation forest; PA: pasture; CR: cropland; UR: urban. Effects for Intense use in MSV could not be estimated as there weren't enough sampled sites.



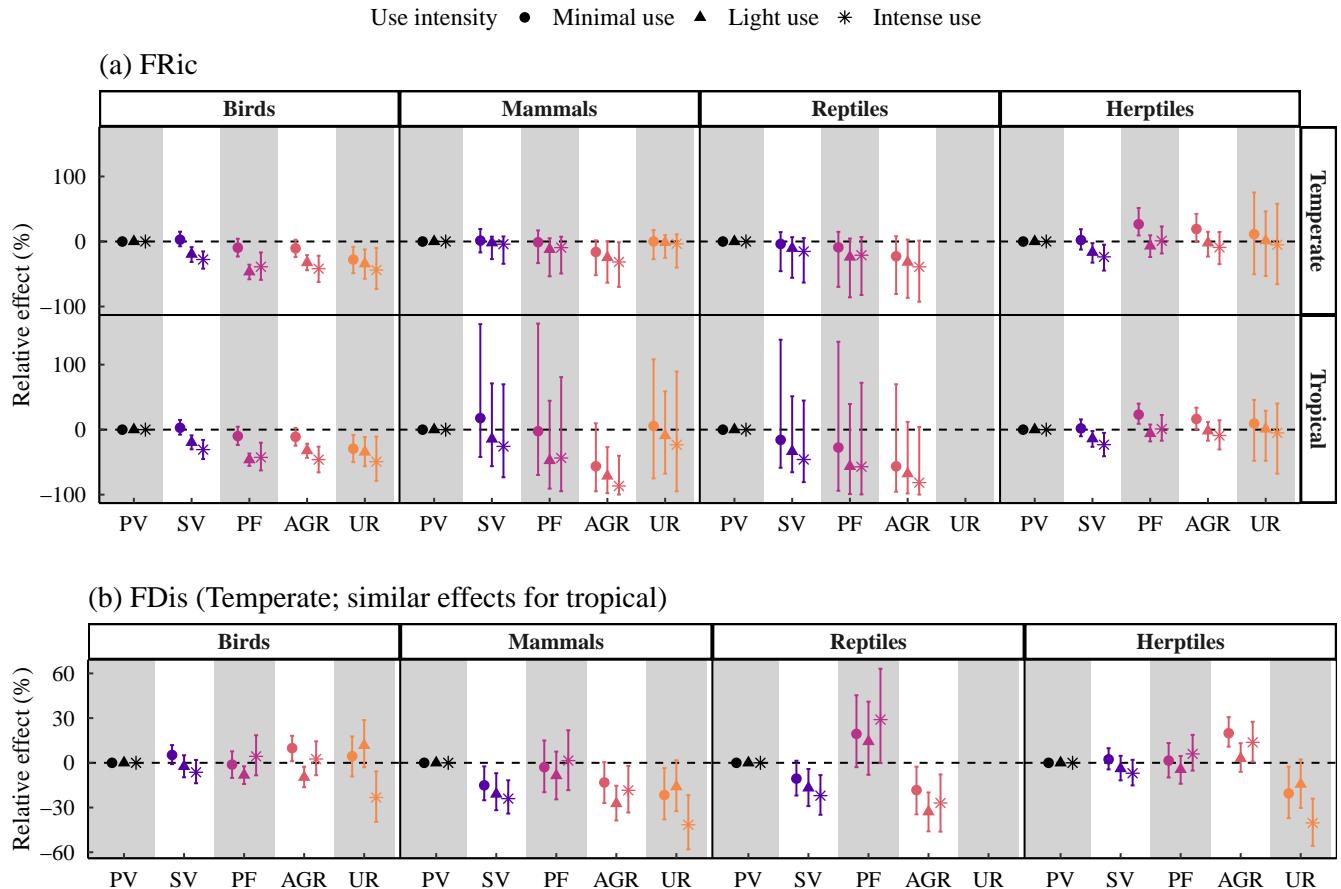
**Figure S3.19: Effects of land use, land-use intensity and region on FRic (a) and FDis (b), for the subset of species with complete trait data (i.e., excluding species with less than 100% trait completeness), with geographical range size as an additional trait considered in the calculation of functional diversity metrics.** Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, I fitted Model 1a (see main text, Chapter 3), which included the effects of land use, land-use intensity and region, as well as interactions between land use and land-use intensity and between land use and region. For FDis, I fitted Model 1b , which included effects of land use, land-use intensity and region, and interactions between land use and land-use intensity. Error bars represent 95% confidence intervals. MSV: mature secondary vegetation; ISV: intermediate secondary vegetation; YSV: young secondary vegetation; PF: plantation forest; PA: pasture; CR: cropland; UR: urban. Effects for Intense use in MSV could not be estimated as there weren't enough sampled sites.



**Figure S3.20: Effects of land use and land-use intensity on FRic (a) and FDis (b), obtained when calculating FRic and FDis with each set of imputed traits (eight in total). Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, I fitted Model 1a (see main text, Chapter 3), which included the effects of land use, land-use intensity and region, as well as interactions between land use and land-use intensity and between land use and region. For FDis, I fitted Model 1b, which included effects of land use, land-use intensity and region, and interactions between land use and land-use intensity. Error bars represent 95% confidence intervals. MSV: mature secondary vegetation; ISV: intermediate secondary vegetation; YSV: young secondary vegetation; PF: plantation forest; PA: pasture; CR: cropland; UR: urban. Effects for Intense use in MSV could not be estimated as there weren't enough sampled sites.**



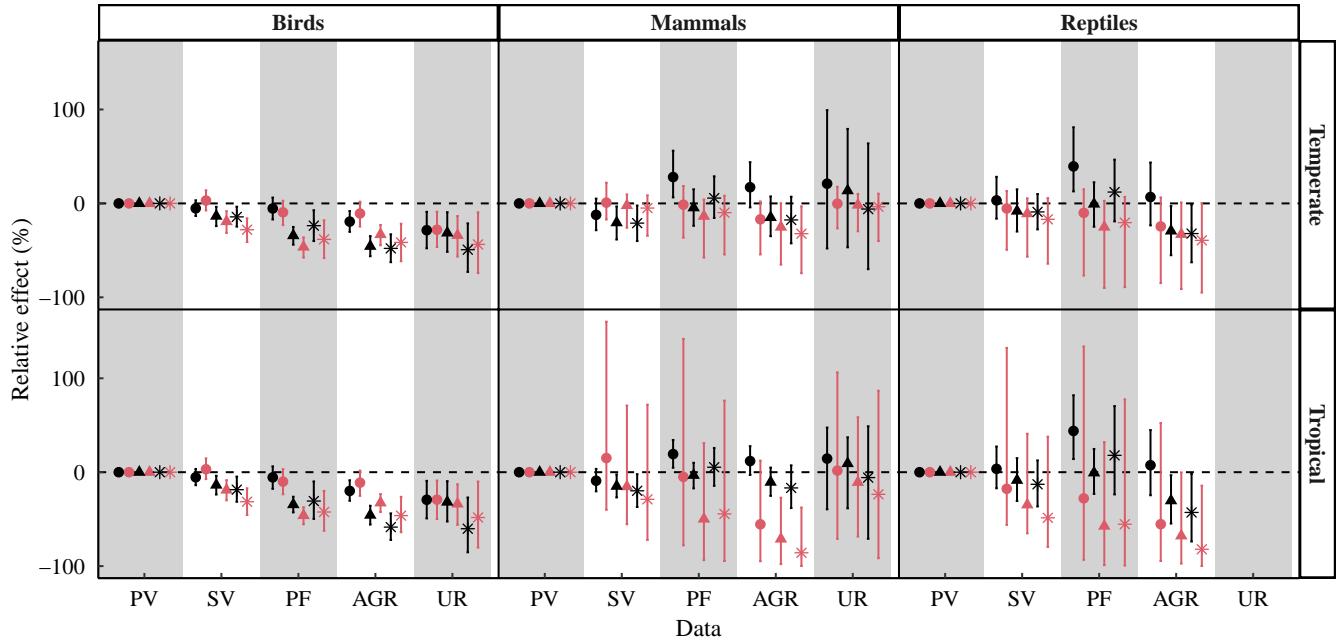
**Figure S3.21: Effects of land use and land-use intensity on FRic (a) and FDis (b), obtained when re-sampling primary vegetation sites twenty independent times.** I fixed the sample size for primary vegetation sites at 50. For FRic, we fitted Model 1a (see main text, Chapter 3), which included the effects of land use, land-use intensity and region, as well as interactions between land use and land-use intensity and between land use and region. For FDis, I fitted Model 1b, which included effects of land use, land-use intensity and region, and interactions between land use and land-use intensity. Error bars represent 95% confidence intervals. MSV: mature secondary vegetation; ISV: intermediate secondary vegetation; YSV: young secondary vegetation; PF: plantation forest; PA: pasture; CR: cropland; UR: urban. Effects for Intense use in MSV could not be estimated as there weren't enough sampled sites.



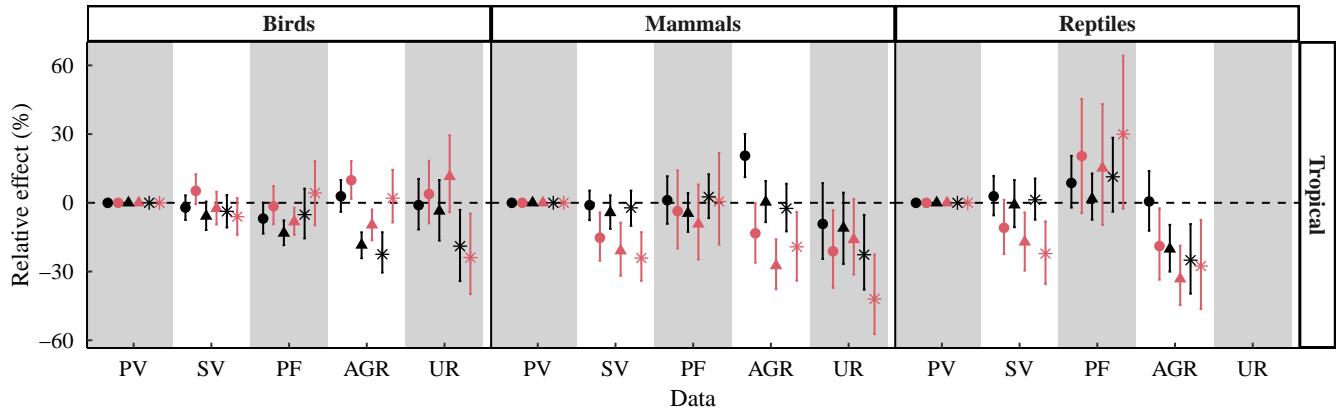
**Figure S3.22: Effects of land use, land-use intensity and taxonomic class on FRic (a) and FDis (c), for the subset of species with complete trait data (i.e., excluding species with less than 100% trait completeness).** Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). I did not include the effects of region here as sample sizes were not large enough for some classes. For FRic, the model included the effects of land use, land-use intensity and class, and interactions between land use and land-use intensity as well as land use and class. For FDis, the model included an additional interaction between land-use intensity and class. Error bars represent 95% confidence intervals. SV: secondary vegetation; PF: plantation forest; PA: pasture; CR: cropland; UR: urban. Effects for reptiles in urban land uses could not be estimated as there weren't enough sampled sites.

—●— complete and imputed trait data —●— complete trait data subset Use intensity • Minimal use ▲ Light use \* Intense use

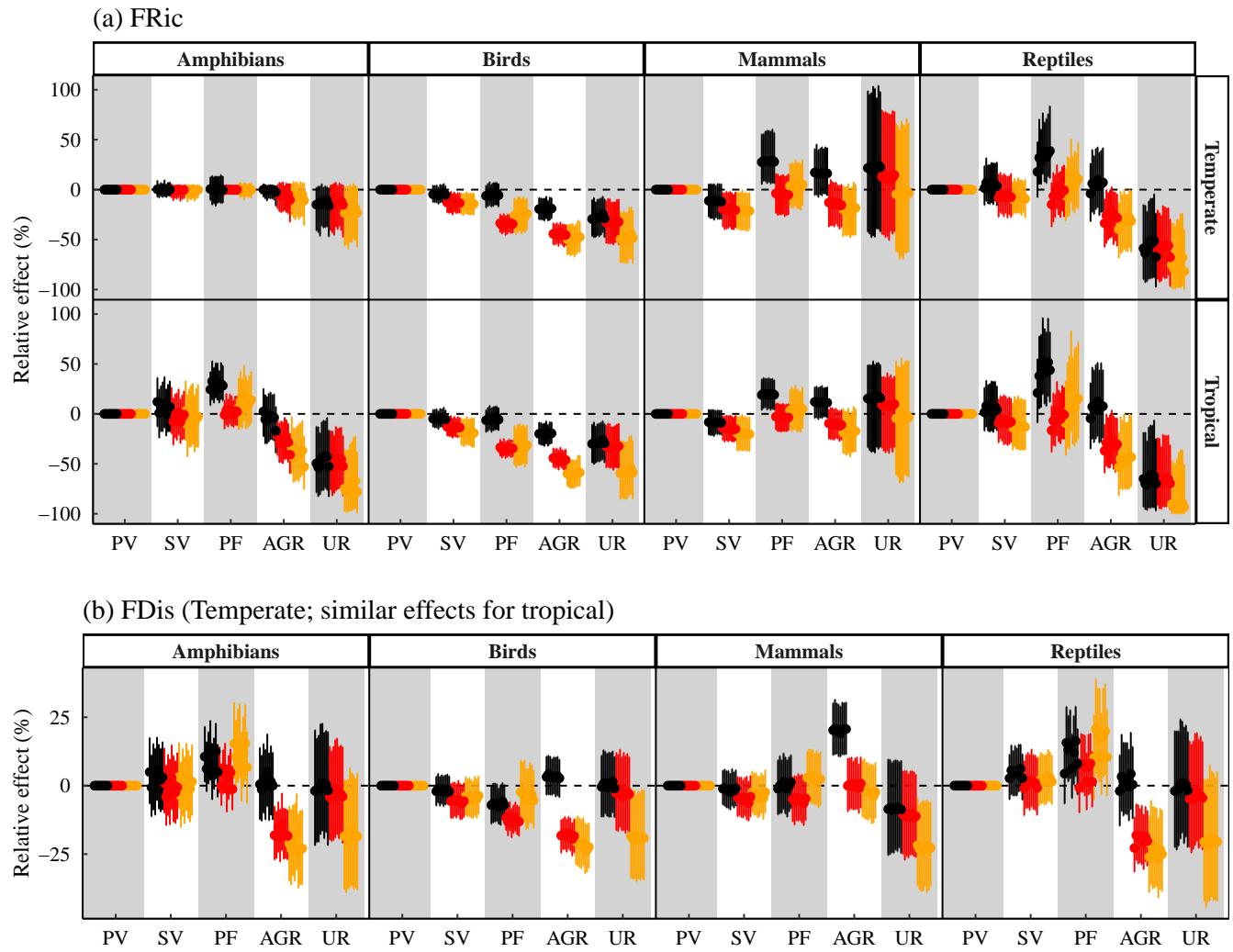
(a) FRic



(a) FDis



**Figure S3.23: Effects of land use, region, land-use intensity and taxonomic class on FRic and FDis obtained with the imputed trait data (black points) or with the complete data subsets (red points).** Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, I fitted Model 2a (see main text, Chapter 3), and I fitted Model 2b for FDis. Error bars represent 95% confidence intervals. SV: secondary vegetation; PF: plantation forest; AGR: agricultural (cropland and pasture); UR: urban. Effects for reptiles in urban land uses could not be estimated as there weren't enough sampled sites.



**Figure S3.24: Effects of land use, region, land-use intensity and taxonomic class on FRic and FDis, obtained when calculating FRic and FDis with each set of imputed traits (eight in total).** Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, I fitted Model 2a (see main text, Chapter 3), and Model 2b for FDis. Error bars represent 95% confidence intervals. SV: secondary vegetation; PF: plantation forest; AGR: agricultural (cropland and pasture); UR: urban. Effects for reptiles in urban land uses could not be estimated as there weren't enough sampled sites.

## S3.8 Model robustness – time since land-use conversion

Time since land-use conversion could have important impacts on assemblage composition and thus, on local functional diversity. I did not investigate these effects because PREDICTS contained data on time since land-use conversion only for about 22% of the sites, considerably reducing samples sizes. Here, I investigated whether the results are likely robust to the inclusion of time since land-use conversion using the subset of sites for which time since land-use conversion was provided. To this end, I found the best-fitting models explaining FRic and FDis, using backwards stepwise selection, starting with complete models that included the effects of land use, time since land-use conversion, region, land-use intensity (for FRic only) and all two-way interactions among these predictors.

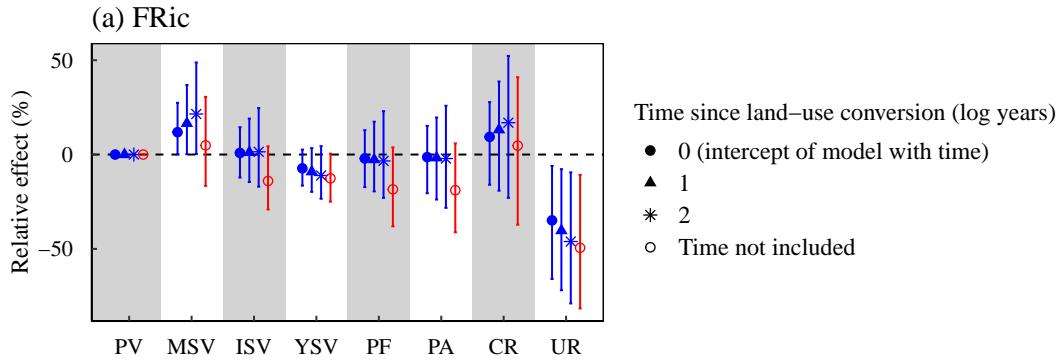
- For FRic, the best-fitting model includes the main effects of land use and time since land-use conversion, but no interaction between these predictors. The model's summary (Table ??) show that time since conversion has a significant negative effect on FRic, but the relationship between FRic and time since land-use conversion is similar in different land uses (as there are no interactions between land use and time since conversion, such that the slopes are similar in different land uses, and so the rate at which FRic decreases with time is similar in different land uses). The intercept is only different for urban land uses (significantly lower). Thus, based on this data subset, I expect time since land-use conversion to have a similar effect in different land uses.

**Table S3.6:** Summary of the model explaining FRic by land use and time since land-use conversion, fitted on the subset of data for which there are information on time since land-use conversion.

	Estimate	Std. Error	t value
Intercept: Primary vegetation	1.156	0.073	15.921
Mature secondary vegetation	0.178	0.093	1.907
Intermediate secondary vegetation	0.018	0.072	0.249
Young secondary vegetation	-0.078	0.051	-1.532
Plantation forest	-0.018	0.082	-0.224
Pasture	-0.005	0.093	-0.054
Cropland	0.133	0.152	0.875
Urban	-0.316	0.133	-2.368
log_Years	-0.094	0.021	-4.566

I then compare this model's predictions with a simpler model that does not account for time since land-use conversion ( $\text{FRic} \sim \text{Land use}$ ). The predictions (Fig. ??) show that including time since land-use conversion does not bias the results, as I find a similar significant effect with both models in

urban land uses, and elsewhere the effects are congruent. Thus, given this data subset, I argue that the results are robust to the inclusion of time since land-use conversion.



**Figure S3.25: Effects of land use on FRic for the model that includes time since land-use conversion (blue points) versus the model that does not take time since land-use conversion into account (red points).**

- For FDis, the best-fitting model includes the main effects of land use, time since land-use conversion as well as interactions between land use and time since land-use conversion (I didn't consider land-use intensity in the starting model because of sample size issues). Nevertheless, the main effect of time since land-use conversion is not significant (Table ??), and the relationship between time since land-use conversion and FDis is not significant in most land uses (except for plantation forest). Thus, I argue the available data do not allow us to properly investigate the relationship between time since land-use conversion and FDis.

**Table S3.7:** Summary of the model explaining FDis by land use and time since land-use conversion, fitted on the subset of data for which I have information on time since land-use conversion.

	Estimate	Std. Error	t value
Intercept: Primary vegetation	0.366	0.011	32.219
Mature secondary vegetation	0.032	0.055	0.577
Intermediate secondary vegetation	-0.015	0.050	-0.298
Young secondary vegetation	0.020	0.015	1.386
Plantation forest	0.074	0.023	3.213
Pasture	-0.017	0.048	-0.346
Cropland	-0.013	0.042	-0.317
Urban	0.031	0.054	0.573
log_Years	-0.004	0.004	-1.186
Mature secondary vegetation:log_Years	-0.005	0.015	-0.335
Intermediate secondary vegetation:log_Years	0.011	0.016	0.650
Young secondary vegetation:log_Years	-0.008	0.007	-1.170
Plantation forest:log_Years	-0.023	0.007	-3.077
Pasture:log_Years	0.010	0.015	0.688
Cropland:log_Years	0.007	0.012	0.620
Urban:log_Years	-0.016	0.022	-0.714

## **Appendix 3: Supporting information for Chapter 4**

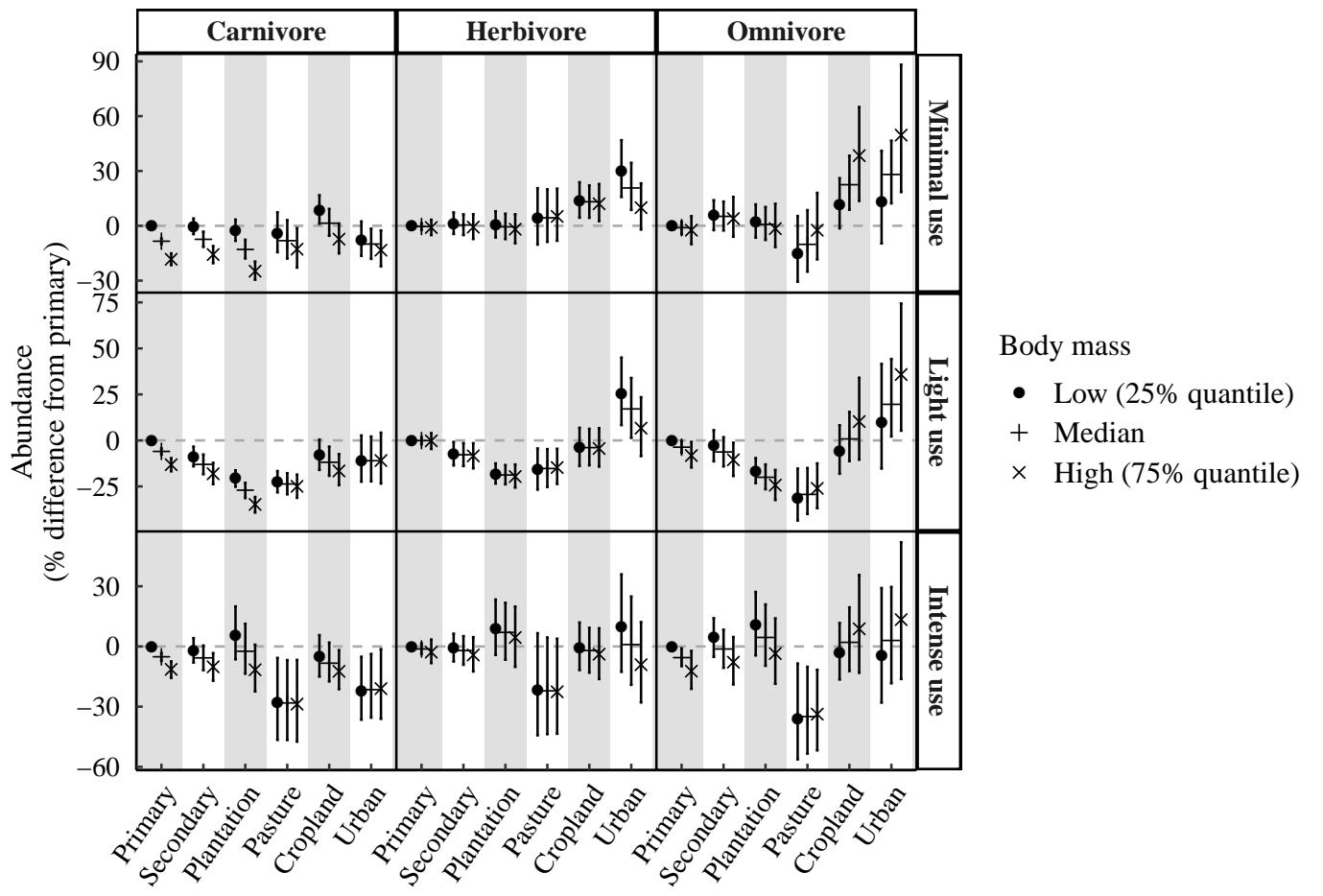
# Appendix 4: Supporting information for Chapter 5

**Table S5.1:** Number of species for which RMR data was collected in each class & data sources, initial RMR coverage for PREDICTS species and phylogenetic signal in RMR.

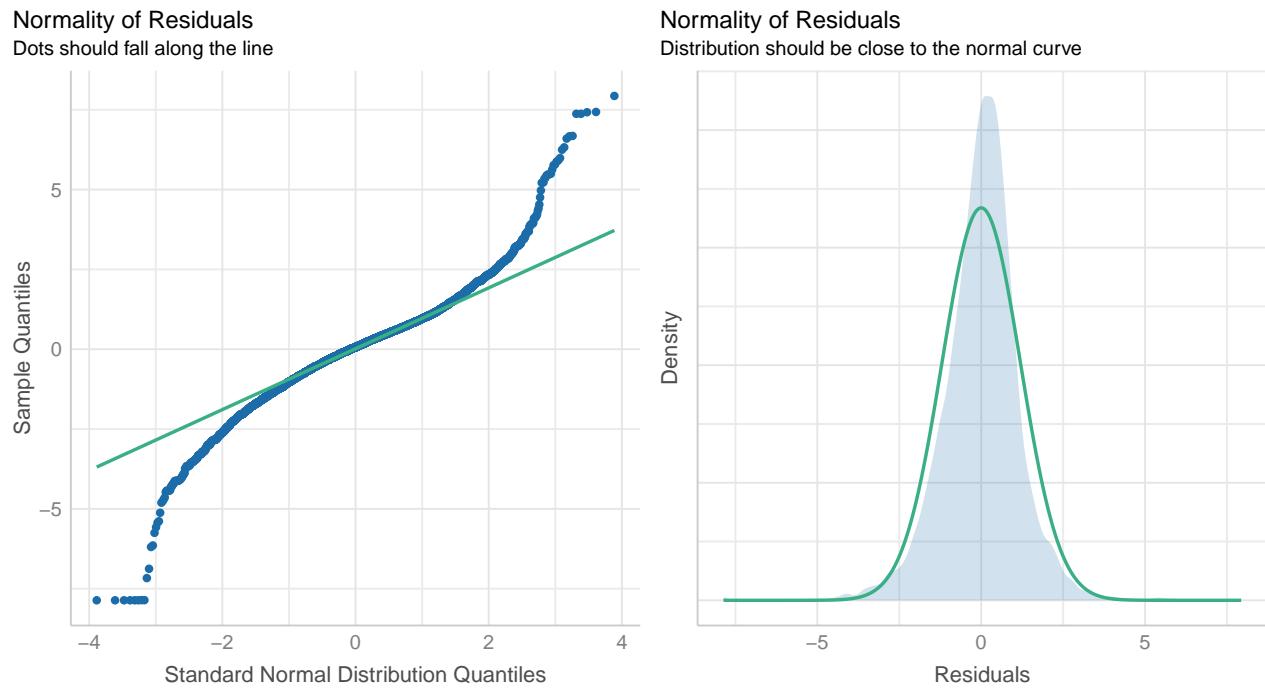
Class	RMR data	Coverage for PREDICTS species	Phylogenetic signal (Pagel's $\lambda$ , $\pm 95\%$ CI)
<b>Amphibians</b>	126 species from <b>Stark2020</b>	16/379 species (4%)	0.89 (0.86-0.91)
<b>Birds</b>	719 species from <b>McNab2009</b> <b>Fristoe2015</b> <b>Londono2015</b> <b>Stark2020</b>	317/3129 species (10%)	0.97 (0.95-0.98)
<b>Mammals</b>	685 species from <b>PanTHERIA (Jones2009)</b> <b>Fristoe2015</b> <b>Stark2020</b>	148/556 species (27%)	0.99 (0.98-0.99)
<b>Reptiles</b>	173 species from <b>Stark2020</b>	24/329 species (7.3%)	0.90 (0.86-0.92)

$$\begin{aligned} \log(\text{Abundance}) = & \text{LU} + \text{LUI} + \text{TG} + \log(\text{BM}) + \\ & \text{LU:LUI} + \text{LU:TG} + \text{LU:log(BM)} + \text{LUI:TG} + \text{LUI:log(BM)} + \text{TG:log(BM)} + \\ & \text{LU:TG:log(BM)} + \text{LUI:TG:log(BM)} \end{aligned}$$

**Figure S5.1:** Formula for the abundance model, used to understand the role of shifts in the body mass of species on observed changes in tRMR (see main text, ‘Disentangling the effects of body mass and abundance on tRMR’). I fitted a model to explain changes in species abundance (given presence) by land use, land-use intensity, trophic group, body mass and their interactions. The model included all two-way interactions among these predictors. To account for potential differences in the slope of the relationship between abundance and body mass among the different trophic groups, I also included two three-way interactions in the model (among land use, trophic group and body mass; and among land-use intensity, trophic group and body mass). Random effects included study, site and species identity. LU: land use; LUI: land-use intensity; TG: trophic group; BM: body mass.

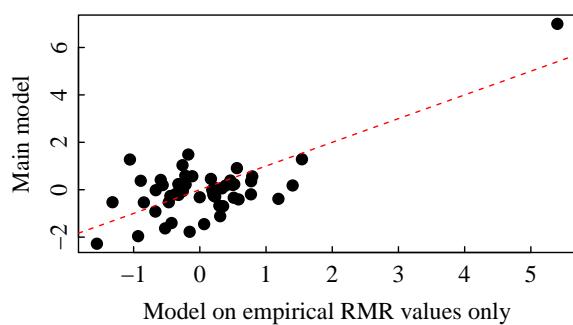


**Figure S5.2: Effects of land use, land-use intensity, trophic group, body mass and their interactions on assemblage-level total abundance**, estimated from the model specified in Fig. S1. The predictions are rescaled with reference to minimally used primary vegetation, considered to be the undisturbed baseline. Primary: primary vegetation; secondary: secondary vegetation; plantation: plantation forest. For visualisation purposes, I plotted the predictions for three body mass levels (but body mass was considered as a continuous variable in the model).

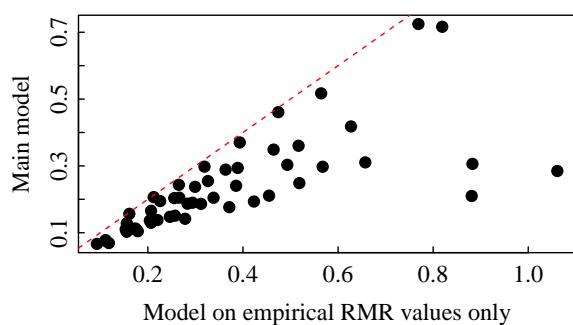


**Figure S5.3:** Diagnostic plots (qq-plot and residual distribution) for the linear mixed-effects model looking at the effects of land use, land-use intensity, trophic group and their interactions on assemblage-level total RMR. The diagnostic plots were obtained with the ‘performance’ R package (**performance**).

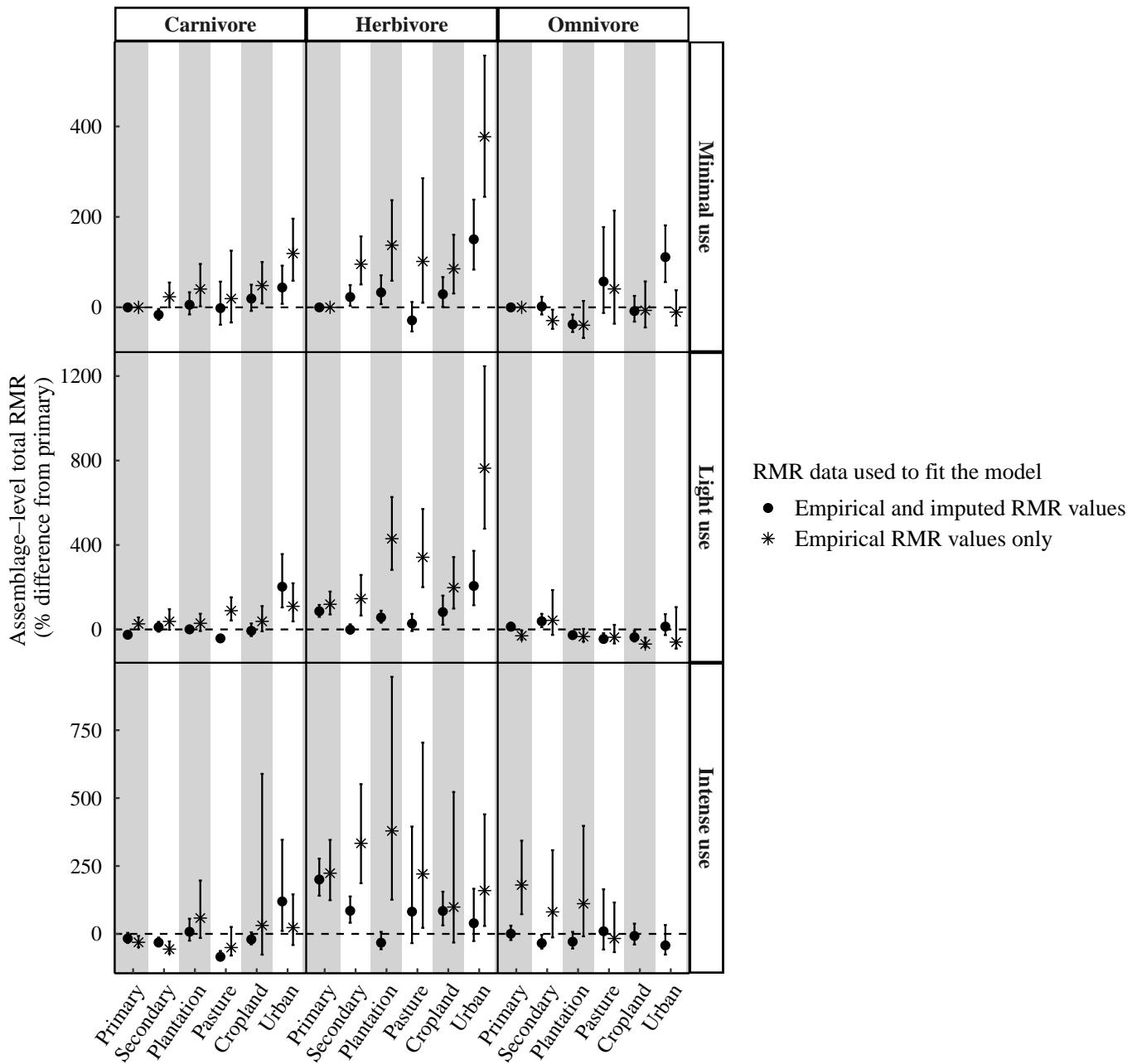
**(a) Coefficient estimates**



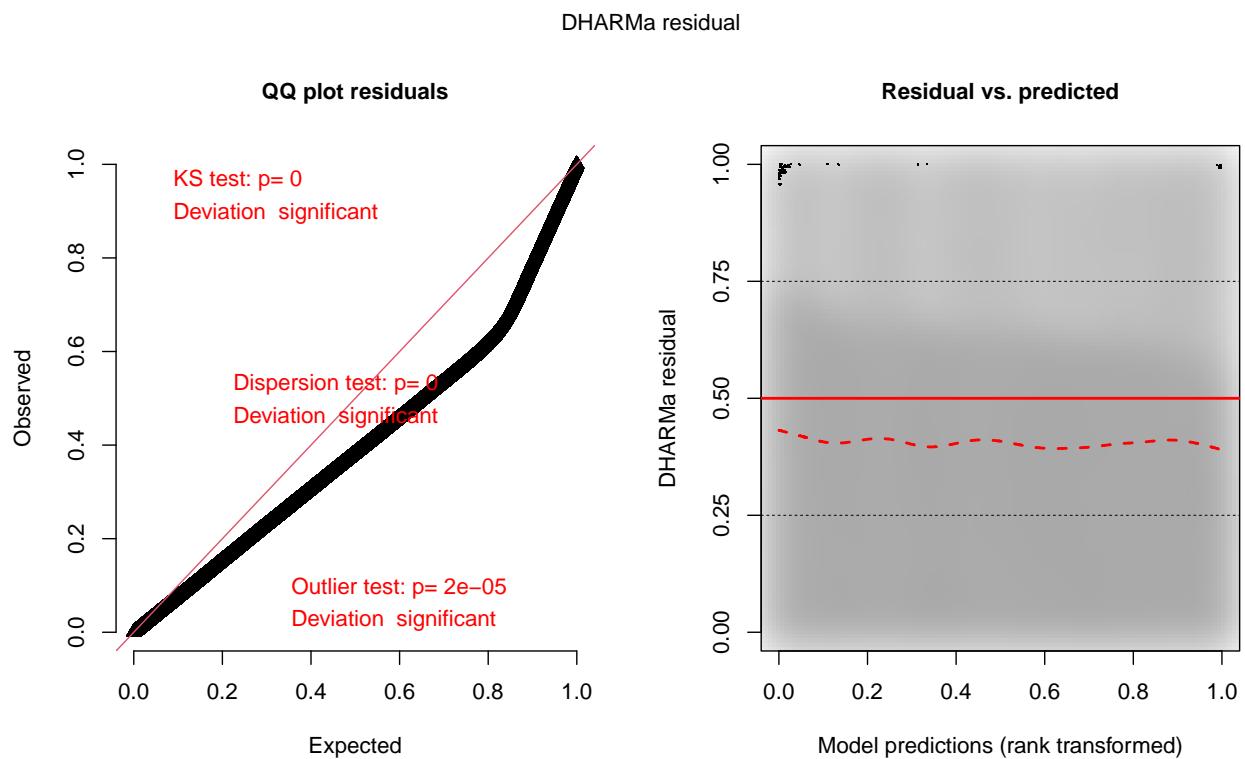
**(b) Standard error estimates**



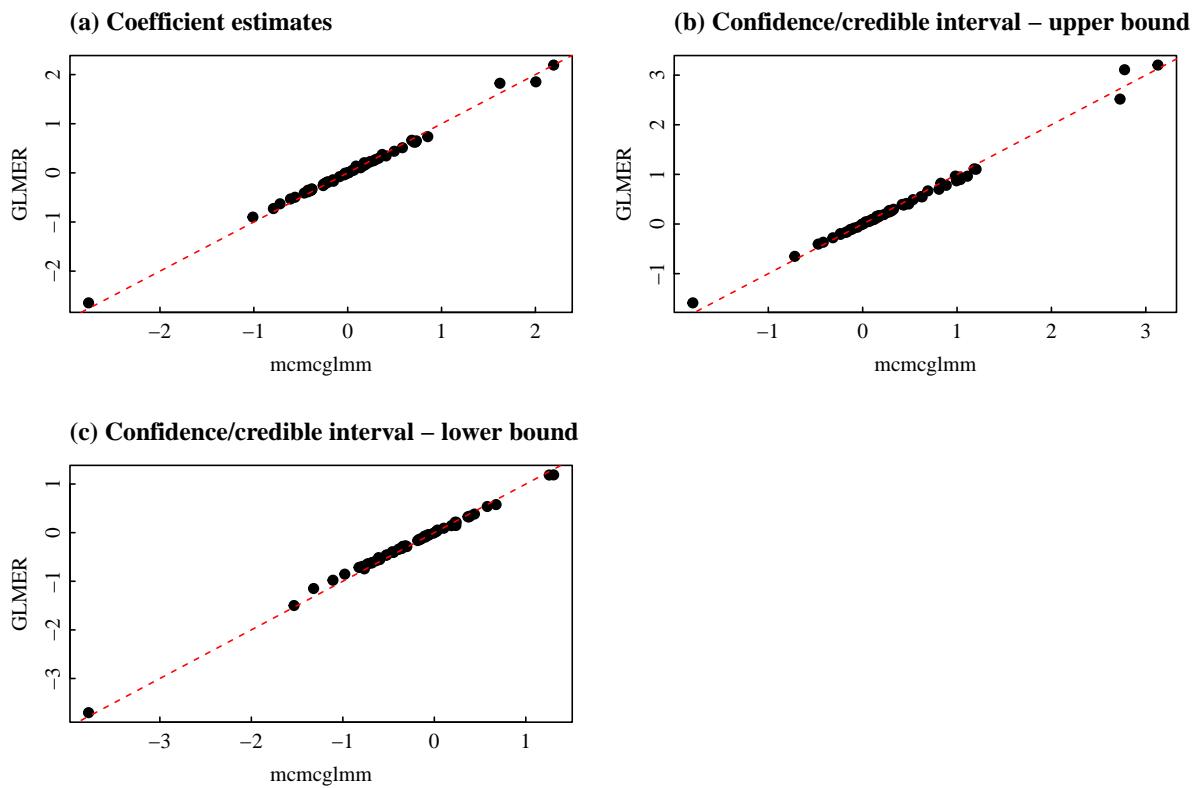
**Figure S5.4: Estimates for the models looking at the effects of land use, land-use intensity, trophic group and their interactions on assemblage-level total RMR.** I plotted the estimates from the model fitted on the empirical and imputed RMR values (presented in the main text) on the y-axis, and the estimates from the model fitted on the empirical RMR values only on the x-axis.



**Figure S5.5: Effects of land use, land-use intensity, trophic level and their interactions on assemblage-level total RMR,** estimated from the model fitted on the empirical and imputed RMR values (presented in the main text) and from the model fitted on the empirical values only. The predictions are rescaled with reference to minimally used primary vegetation, considered to be the undisturbed baseline. Primary: primary vegetation; secondary: secondary vegetation; plantation: plantation forest.

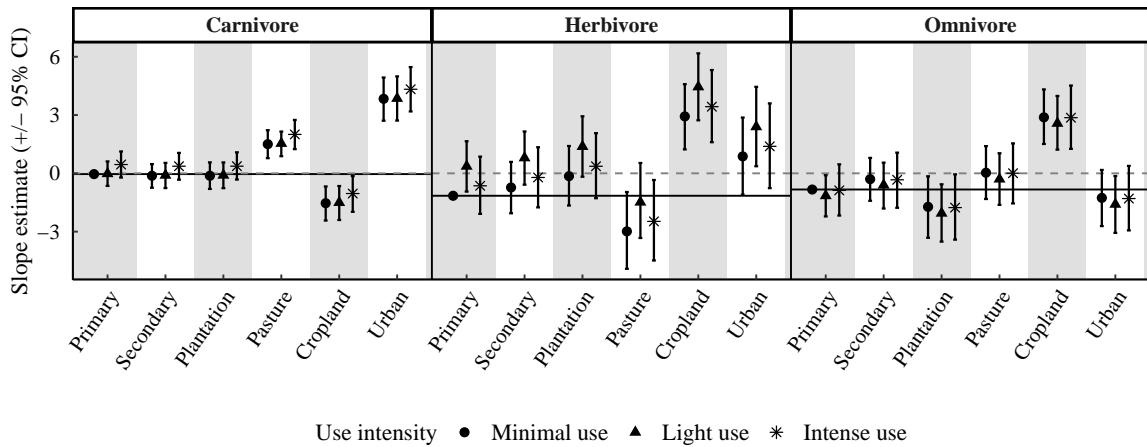


**Figure S5.6:** Diagnostic plots for the generalised mixed-effects model looking at the effects of land use, land-use intensity, trophic group, residual RMR and their interactions on species' probability of occurrence. The diagnostic plots were obtained with the 'DHARMA' R package (**DHARMA**).

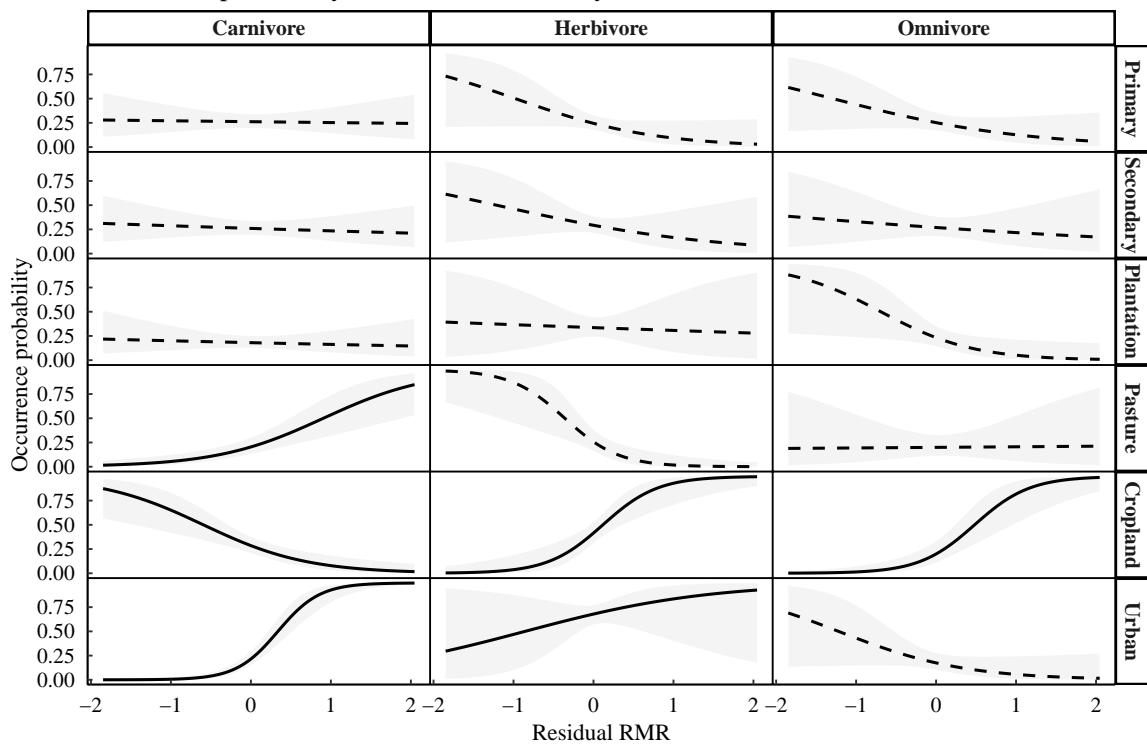


**Figure S5.7:** Model's coefficients from the occurrence model fitted using the `lme4` package (Bates2015) against coefficients from the model fitted using a Bayesian framework with the `MCMCglmm` package (`mcmcglmm`). The models were fitted to investigate the effects of land use, land-use intensity, trophic group and residual RMR on species occurrence probability.

(a) Slope of the relationships between occurrence probability (log-odds) and residual RMR



(b) Predicted effects of residual RMR on occurrence probability (minimal use intensity)



**Figure S5.8:** (a) Slope estimates for the relationship between residual RMR and occurrence probability in each land-use type and for the three levels of land-use intensity, from the model fitted using the empirical RMR values (i.e., excluding imputed RMR values). The black horizontal line indicates the slope for the reference level (primary vegetation) for minimal land-use intensity. The grey dashed line marks 0. Error bars are 95% confidence intervals. (b) Effect of residual RMR on species probability of occurrence within each trophic level and for each land use-type. I plotted the predictions for minimal land-use intensity only. Solid lines represent significant relationships. Primary: primary vegetation; secondary: secondary vegetation; plantation: plantation forest.

## **Bibliography for the Appendices**