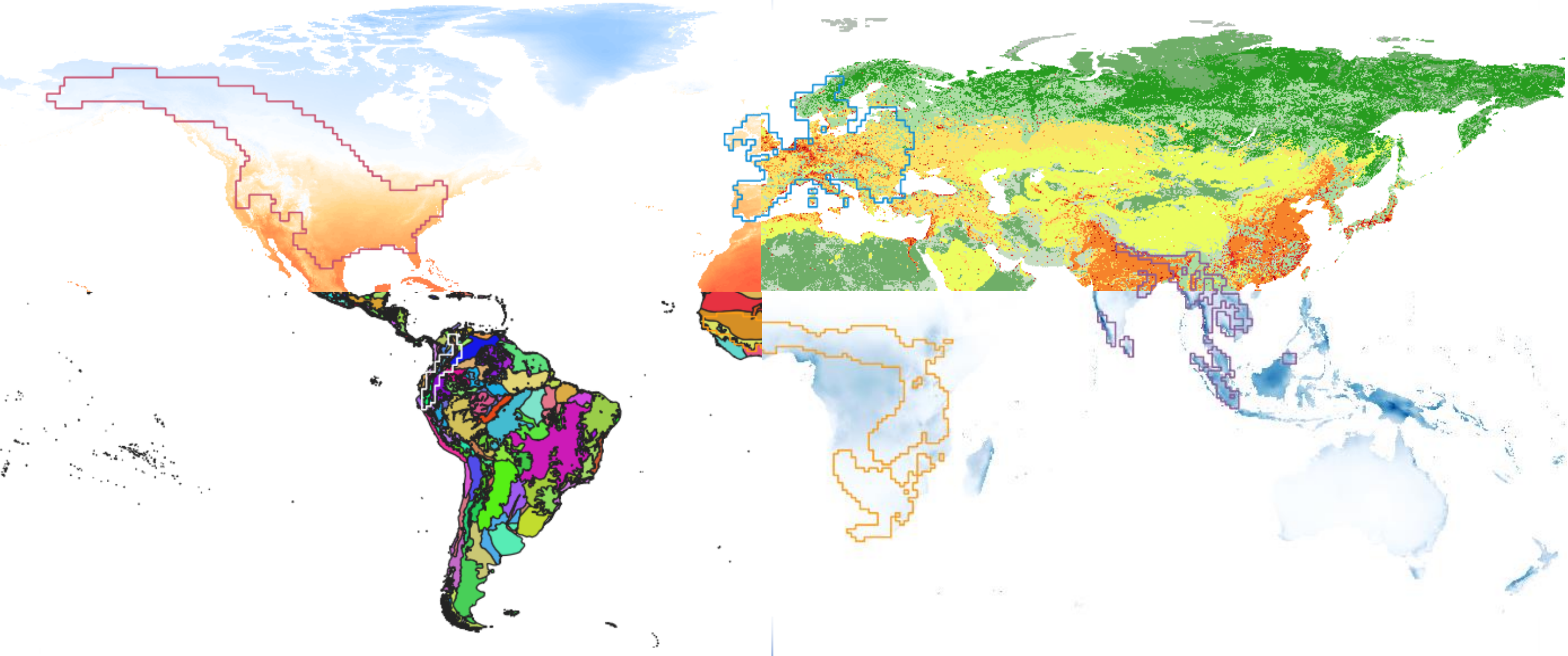
Identifying refugee species by linking global datasets on mammal distribution ranges and human impact

Master Thesis



Anouk Visser
6567479
a.visser3@students.uu.nl

Thesis 30 ETC

MSc Sustainable Development -
Environmental Change and
Ecosystems

Supervisor:
Joris P.G.M. Cromsigt

Co supervisors:
Graham I.H. Kerley
Kees Klein Goldewijk

Utrecht University

Content

Summary 2
Introduction 3
  The problem 3
  Proposed research 3
  Scientific and societal relevance 4
Theoretical background and objectives 5
  Origin of mammals 5
  Extinction drivers and Anthropogenic pressure 6
  Evaluating refugee species 6
  Academic interest in refugee species 7
Methods 9
  Range decline 9
  Diversity decline 9
    Habitat diversity 10
    Environmental diversity 10
  IUCN status 11
  Human activity 12
  Data Analysis to integrate the information on refugee species criteria 12
Results 14
  Range decline 14
  Habitat diversity 14
  Environmental diversity 15
  IUCN status 16
  Anthropogenic pressure 16
  PCA clustering analysis 18
Discussion 23
  Key results summary 23
  Interpretation 23
  Implications 24
  Limitations of this research 26
  Recommendations 26
  Future implications 27
Acknowledgements 28
References 29
Appendix A – PHYLACINE species' range reflection 31

Appendix B – Overview Results All Species 34
Appendix C – Visualisation Methods 36
  Range decline and habitat diversity 36
  Environmental diversity 36
  Human activity 37
Appendix D – Script 39

Summary

A species whose conservation is at risk as a result of a declined fitness due to (pre-) historic anthropogenic confinement to suboptimal habitat are recognised as refugee species. The refugee status potential of only a handful of species have been assessed while a lot more unrecognised refugee species are expected to exist due to the global scale and increasing pressure of the anthropogenic driver. I conducted a global scale initial assessment with existing global datasets to identify potential refugee species among large herbivorous mammals. The assessment approximates the existing first phase refugee species identification assessment of Kerley et al. (2012) and evaluates the species' range reduction, range diversity losses, compromised population fitness and associated anthropogenic pressure. Of the 79 included species 49 species could be assessed of which 18 species show a serious potential of being a refugee species.

My research implies it is possible to use existing global datasets to identify species with an increased refugee status potential, although not all refugee species might be detected. Variables like the range decline and diversity losses are expected to be more efficient in the uncovering of refugee species than other variables, like the current anthropogenic land use. An improved and more elaborately tested version of my developed initial identification assessment could be used for the evaluation of more species.

Introduction

The problem

Humans play, and have played, a significant role in the distribution of animal species around the world. Since the emerging of *Homo sapiens* (~260ky) our hunting activities, expanding presence and later the alteration of land and developing transportation networks have put pressure on species' survival. This pressure has caused an ongoing extinction event and has forced a variety of species to trade their optimal habitat for a less suitable one, negatively affecting the population fitness. When a species' conservation is at risk as a result of a declined fitness due to (pre-)historic anthropogenic confinement to suboptimal habitat, the species is recognised as a refugee species (Kerley et al., 2012).

If the process of human pressure pushing species into (suboptimal) refuge habitat started a long time ago (e.g., going back centuries or even to prehistoric times), the risk is that refugee species are not always recognised as such. This can be caused by the shifting baseline syndrome, which suggests that the perception of natural baselines changes over the generations (Pauly, 1995). Each generation of conservationists or scientists accepts the species' environmental conditions of their time as natural baselines. If anthropogenic changes in habitat use occur over many generations, these conditions may be mistakenly assumed to reflect the species' optimal habitat. The real threat for refugee species occurs when the suboptimal habitats are seen as optimal habitats and treated as priority conservation areas, confining the species to these suboptimal conditions (Kerley et al., 2012).

The ambition to protect biodiversity, as embodied by sustainable development goal 15: life on land, is reliant on the conservation of species. This includes the need to protect populations and their habitat, as has been increasingly done in the last decades (UNEP-WCMC et al., 2018). Despite the growing number of protected areas globally, many species are still declining and continue to be threatened, known as the Protected Area Paradox (Kerley et al., 2020). Additionally, the decline of wilderness area and increasing anthropogenic pressures are seriously affecting conservation opportunities. Conservation management is vital for refugee species due to their reduced fitness. But conservation efforts may result in limited success when species are not recognised as refugees and therefore conserved under suboptimal conditions (Kerley et al., 2020). The suboptimal conditions inadequately provide for the species' needs, regardless of conservation efforts, which may be a part of the explanation of the Protected Areas Paradox (Kerley et al., 2020).

Conservation and protection efforts run on limited means with limited time, which asks for an efficient and proactive approach to conservation. Recognising and better understanding refugee species can facilitate more effective conservation management to sufficiently improve the species' stability and better protect the world's biodiversity.

Proposed research

The aim of this research was to perform a global scale initial assessment of potential refugee species using the by Kerley et al. (2012) outlined assessment framework. The initial assessment method was based on existing global databases, combining and analysing data on current and natural mammal distribution, reconstructed human land use and environmental diversity within the species' current and natural range.

To allow for a feasible analysis within the given time, I focused only on a subset of all animals in my assessment. Species most likely to be refugee species are species that were are most affected by human activity. Therefore, I decided to include extant, herbivorous mammals weighing at least 100kg as an adult.

Larger animals (>100kilogram) are more affected by human activity as they suffer disproportionately from human hunting, have relatively slow life-history traits, are directly affected by habitat loss and degradation while they already need a large range to sustain their diet (Ripple et al., 2016). This is also reflected by the Quaternary extinction event, as this has mainly affected large mammals.

Although human activity affects both herbivores and carnivores, I only wanted to regard one and have decided on herbivores. The link between human pressure and herbivores is less complex, as the effect on carnivores is influenced by a third party, like shared prey.

The aim, assumptions and containments led to the research question:

Can an initial assessment using global databases identify possible refugee species among the world's large herbivorous mammals?

Scientific and societal relevance

The assessment ranks species into classes that indicate a probability of being a refugee species. It provides information on the included species and could help in the evaluation of other species.

Although this initial analysis will not result in a certain refugee status, the results can have a multitude of applications and help obtain more knowledge. First of all, the identified species with a higher potential of being refugee species can be relevant subjects for future in-depth studies. Additionally, the proportion of the resulting probability classes of refugee species may give an indication of the number of possible refugee species that are currently unrecognised.

The used method can also potentially function as enhancer for knowledge on refugee species. It could help in the future assessments of a possible refugee status for other (subsets of) species. It could also help increase our understanding of refugee species, for example by evaluating the link of (additional) data to the refugee status identification and see whether refugee species are typified by certain characteristics. If something sets the refugee species apart, this information can be used to enhance further understanding and identification of refugee species.
Additionally, an enhanced method can be used by conservation planners and managers to improve conservation efforts. It can be used as an initial assessment to verify whether the species they are working with are potentially refugees. Then the species' range confinement and decreased habitat diversity can be evaluated and improved.

Theoretical background and objectives

Origin of mammals

The Phanerozoic Eon (541 Ma - today) is the period of major floral and faunal development and evolution, interrupted by a multitude of mass extinctions leading to the current global diversity. During the Palaeozoic, the earliest of the three Phanerozoic eras, ancestors of mammals started to separate from reptiles and birds, continuing the reptile-mammal transition during the second era, the Mesozoic (Glikson & Groves, 2016a). An asteroid impact 66 million years ago caused the second largest mass extinction, resulting in a loss of 46% of the living genera (Glikson & Groves, 2016b). This event marked the beginning of the Cenozoic era, the Age of Mammals, named after the small mammals amongst the survivors that started to thrive, diversify and expand to fill the ecological niches of the extinct species (Glikson & Groves, 2016b, 2016a). During the late Eocene and early Oligocene some of the mammalian lineages started to increase in size, with some species over time reaching heights of 6 meters (Glikson & Groves, 2016a). During the same time the primates lineage, based in Africa, started to diversify extensively, leading to the origin of the Homo genus around the Pliocene-Pleistocene boundary (Glikson & Groves, 2016a). The *Homo sapiens* is first identified in the late Pleistocene (~260 ka) and stayed in Africa until around 60 ka, when they dispersed to Asia, Australia and Europe, only arriving in the Americas towards the end of the Pleistocene (~14 ka) (Glikson & Groves, 2016a; Smith et al., 2018).

Species extinction and diversification rates were stable during most of the Cenozoic, until sudden extinction events across all continents started to cluster around the Pleistocene-Holocene boundary (Glikson & Groves, 2016a; Smith et al., 2018). These late Quaternary extinctions were unique, as no major cataclysmic event could be identified, the events happened asynchronously around the globe during relatively short periods of time, and the extinction was strongly biased towards large-bodied species in contrast to previous extinctions (Bartlett et al., 2016; Lyons et al., 2016; Smith et al., 2018). Whilst large mammals faced increased extinction rates during the late Quaternary, large body size did not affect extinction vulnerability during previous mass extinction events (Lyons et al., 2016; Smith et al., 2018).

Researchers continue to debate over the causes of this so-called Pleistocene or Late Quaternary extinction event. Previously researchers mainly believed the changing climate to be the extinction cause, but the role of a significant human impact has gradually become more popular (Bartlett et al., 2016; Broughton & Weitzel, 2018; Sandom et al., 2014). Researchers are generally agreeing the extinction event is caused by a combination of both the climatic and human impact, but are still debating the extent of both influences. Recently, the academic consensus seems to shift towards a more significant human impact (Bartlett et al., 2016; Lyons et al., 2016; Sandom et al., 2014; Smith et al., 2018). One of the main recent arguments for humans as the main driver behind the extinctions is the apparent close match between the time of the arrival of modern humans and the geographical occurrence of the extinctions events of large mammals (Glikson & Groves, 2016a; Lyons et al., 2016). When humans arrived in new regions for the first time, larger indigenous mammals went extinct or severely decreased in numbers (Glikson & Groves, 2016a; Sandom et al., 2014). This indicates that large mammals were at least targeted by humans or more vulnerable to their impact (Lyons et al., 2016). The Quaternary extinction continues until this day, and although the average size of mammals has decreased (Smith et al., 2018), large mammals are still more prone to extinction (Lyons et al., 2016).

Extinction drivers and Anthropogenic pressure

The extinction of a species, that is not cataclysmic in nature, is always caused by at least one of three factors: 1) the habitat disappears and the species is unable to migrate to another habitat; 2) the species is outcompeted in their habitat; 3) the species is over-exploited by their predators (Glikson & Groves, 2016a). The final factor in the extinction of a species is often a small remaining population that is very susceptible to genetic and demographic disturbances, known as the small population paradigm (Channell & Lomolino, 2000). Small populations are not as resilient to disturbances and die out more quickly when facing an otherwise mild disruption. The fatal disruption is rarely also the cause of the reduced population, possibly making the population reducing factor(s) a more important contributor in the extinction of a species than the final driver. The declining population paradigm is the identification of those external factors that cause the ranges of populations to decrease and consequently their numbers to decline (Channell & Lomolino, 2000).

Although Late Pleistocene and Holocene climatic changes were drivers of the late Quaternary extinction events, human colonisation is identified as the primal driver (Bartlett et al., 2016). Anthropogenic pressure relates to all three extinction factors as identified by Glikson and Groves (2016a), by expanding their settlements, cutting off migration routes of mammal populations (limiting their response to climatic changes), competing for land use and hunting activities. The climatic changes caused habitats to disappear but it is unlikely that this would have caused this rate of extinction without human interference (Sandom et al., 2014).

Anthropogenic pressure is a broad concept encompassing all direct and indirect negative impacts of humans on the species' persistence. Multiple classifications exist, mostly based on land use change and disruptive activities, often including built environment, resource use and agriculture (Monsarrat et al., 2019). The impact of these anthropogenic pressures varies per area and taxon, as species are affected differently by and have different thresholds for each activity (Monsarrat et al., 2019). Large mammals with predominately plant-based diets are most severely affected by human hunting, habitat loss, habitat degradation and competition with livestock (Monsarrat et al., 2019; Ripple et al., 2016).

The anthropogenic pressures driving the extinction factors are ongoing, forcing species to live in suboptimal habitat while optimal habitat is lost or escaping routes are cut off. If nothing happens to relieve these factors of the declining population paradigm, species may become too vulnerable and face extinction driven by the small population paradigm (Kerley et al., 2012).

Evaluating refugee species

The shifting baseline syndrome can cause current habitat to be mistakenly assumed to reflect the species' optimal conditions (Pauly, 1995), which endangers species and hinders the critical assessment of current conservation efforts (Kerley et al., 2020). Re-evaluating the historic distribution range of a species can help correct these discrepancies and re-establish information about the optimal conditions and habitat of a species (Abicair et al., 2020). When the external pressure leading to the range loss is identified and reduced, a species can be reintroduced into their (newly established) natural habitat, or can be introduced in a new area that meets the criteria of the natural habitat (Abicair et al., 2020; Ali et al., 2018). When the natural (type of) habitat is destroyed or modified by humans, Martínez-Abraín and Jiménez (2016) suggest to search for substitution habitats. These substitution habitats are unintendedly created (semi)artificial habitat types that function comparable to the original habitat.

Studying (re)introduced species is a powerful source of information about refugee species, especially when populations are compared in habitats of varying suitability. Experimentally obtained information on population performance and life history can be used to significantly improve range distribution models for refugee species and enable a transition from the commonly used correlative models towards better fitting mechanistic models (Cromsigt et al., 2012).

Academic interest in refugee species

In recent years, the refugee species concept received increasing academic and some media attention (Abicair et al., 2020; Ali et al., 2018; Kerley et al., 2012; Ruiz-Leotaud, 2020). Research on refugee species mainly consists of case studies of individual species, focusing on topics like: the identification of species as a refugee; the relation between human activity and distribution ranges; the identification of areas that are important for the protection of refugee species; and the reintroduction of species in their natural habitat (Abicair et al., 2020; Kerley et al., 2012; Lea et al., 2016; Monsarrat et al., 2019; Qian et al., 2020).

Anthropogenic pressure is the driver of species becoming refugees, which is an extensive and global disturbance. Still, only a handful of animals has been assessed in terms of their potential of being a refugee species. The fact that human disturbances are so widespread and go back a long time makes the existence of more unrecognised refugee species very probable. To enhance academic understanding of refugee species and to improve their conservation, more (elaborate) research is needed. A developing understanding of refugee species involves progress in their identification, the understanding of when and why they became refugee species and the assessment of their conservation management.

Kerley et al. (2012) have developed a framework for the detection and examination of refugee species, consisting of two phases. The first phase is an initial assessment of potential refugee species based on identifying 1) severe historic declines in species distribution range, 2) slow population growth and/or low densities despite conservation efforts and 3) anomalous habitat use and/or diet. According to Kerley et al. (2012), a refugee species needs to fulfil all three of these criteria. The second phase is a detailed examination of potential refugee species status consisting of a 1) thorough analysis of the species natural history to form hypotheses on how the species became a refugee species and what the species' optimal niche requirements are, and 2) experimental empirical testing of these hypotheses by reintroducing populations.

The initial assessment of this research to identify possible refugee species among the world's large mammalian herbivore species is mainly based on the first phase of the assessment framework by Kerley et al. (2012). I altered the above-listed three criteria that a potential refugee species has to fulfil, added an anthropogenic pressure criterion and transformed them into four specific sub questions that I investigated for each species:

1) How much is the habitat range affected for each species?
2) How much has the environmental and habitat diversity changed for each species, as a proxy for amount of optimal habitat in a species range?
3) What is the current IUCN status, as a proxy for population growth and density, of each species?

4) How much has human pressure increased for the last 12000 years in the habitat range of each species?

The more severe the outcomes of these questions are for a species, the more likely it is for that species to be a refugee.

Methods

The conducted research had a five-part structure: one part for each of the sub questions and one final analysis to answer the main research question. The method is described for each individual part. All the programming was done in the language R.

Species selection
I used a subset of mammal species (n = 79 species) in this research based on the criteria of the species being extant, large (rounded adult weight of ≥100kg) and predominately herbivorous (≥80% plant-based diet). The species selection was performed with data taken from the PHYLACINE dataset of Faurby et al. (2018).

Range decline

To analyse the range losses two distribution ranges are compared: a current distribution range and a potential natural distribution range. Both ranges for each species are a part of the PHYLACINE dataset developed by Faurby et al (2018). The PHYLACINE dataset is developed with the primary motivation to assess human influence on the world's mammalian species (Faurby et al., 2018). They did this by comparing the current known range distribution of all extant mammal species and the potential natural range distribution for all late Quaternary mammalian species (~130,000 years ago). The potential natural range is defined as the current range of species if they had never experienced strong anthropogenic pressures and is called the present natural range (Faurby et al., 2018). The PHYLACINE dataset contains a gridded world map for the two ranges of each species, with each cell indicating the current or present natural presence of a species as true(1) or false(0). The ranges are represented in world maps, divided in 51,120 cells with an equal area of 9,312 km^2. These cells are too large to nicely follow coastlines and can cover a substantial amount of (sea)water, which decreases the accuracy of the range surface. The surface area of the ranges was calculated by overlapping them with the ecoregions map of the *Ecoregions2017* dataset by Dinerstein et al. (2017). In order to overlap the maps, the Coordinate Reference System (CRS) of the *Ecoregions2017* map was converted to the equal area CRS of the PHYLACINE dataset (Behrmann cylindrical equal area). Additionally, the range maps of each species were converted from raster to polygon format. The intersection function calculates the overlapping surface area, of which the sum is the range's total land surface area.

With the determined land surface of the current range and present natural range I calculated the range contraction for each species. The severity of the range decline is represented by the relative surface difference of the two ranges $((Area_{present\_natural} - Area_{current})/Area_{present\_natural})$.

Diversity decline

The loss of habitat diversity is used as a proxy for the loss of optimal habitat availability. The habitat diversity was represented by the habitat type and environmental conditions diversity in both ranges. A decreased range was already expected to result in a less diverse environment, still the severity of the diversity decrease relates to the chances of a species being a refugee.

Habitat diversity

The habitat diversity is expressed by the diversity of ecoregions in the ranges. I chose the Shannon's diversity index as the indicator for the habitat diversity, as it represents both the presence (richness) and the proportion of the ecoregions in both ranges. The idea is that the diversity index offers a balanced presentation of the habitat diversity, as it strongly relates to the richness in the range but is less sensitive to included but unimportant ecoregions with only a small cover area.

The ecoregions are taken from the *Ecoregions2017* dataset of Dinerstein et al. (2017), a map consisting of polygons representing the areas of all 847 global ecoregions, assigned to 14 biomes. The number of ecoregions and their cover in the current and present natural range of each species were calculated by overlapping the polygonised PHYLACINE species' ranges and the Ecoregions2017 map with a projected CRS. I used the intersects function to find the ecoregions present in the range and the intersection function to obtain their surface area. The relative cover area of each ecoregion in a range was calculated by dividing their area by the total range area.

The Shannon's diversity index is calculated (equation 1) for each range with the number of present ecoregions (S) and their relative cover area (p_i). The index has a minimum value of 0 and increases with a higher diversity. The change in habitat diversity was quantified by calculating the difference between the diversity indices of the current and present natural range.

The absolute size of the area is irrelevant in the calculation of the diversity index, as the number and cover ratio of the components are independent of the total area surface. This makes it possible for a smaller range, even with a smaller number of components (ecoregions), to be more diverse.

Eq. 1 $$H = -\sum_{i=1}^{S}(p_i * \ln(p_i))$$

H = Diversity
p_i = Relative area covered by each ecoregion
S = Number of ecoregions

Environmental diversity

The environmental diversity is expressed by the climatic conditions in the ranges. An environmental envelope was created using two bioclimatic variables: the annual mean temperature and the annual precipitation. The climatic data was taken from the *WorldClim* dataset (Fick & Hijmans, 2017), version 2.1 (released in January 2020), containing the monthly climate data and bioclimatic variables over the years 1970 to 2000. The bioclimatic variables BIO1 = Annual Mean Temperature and BIO12 = Annual Precipitation data in the 5 arc minutes format (approximately 85km2 at the equator) were used. The relevant climatic data was extracted by overlapping the polygonised species' range maps with the projected bioclimatic maps. The coupled annual mean temperature and the annual precipitation data of each range is processed into a bagplot (Appendix C).

A bagplot resembles a bivariate boxplot, forming a two-dimensional visualisation of the climatic conditions in the range and was used as a statistical description of the core environmental envelope. A bagplot consists of a center, bag and fence. The center is the depth median: the point with the maximum depth function value. The bag surrounds the center and contains 50% of all points with the largest depth. The fence covers three times the size of the bag and forms the outer limit of the bagplot, thus representing the environmental envelope of the species. All points that fall outside the fence are called outliers and are considered as too divergent to be a part of the environmental envelope.

After the construction of the bagplots for the climatic values in the current and present natural range of each species, the bagplots were overlapped. The extent of the overlap of the two bagplots from each species is directly related to the severity of change in habitat diversity. The bagplot of the climatic conditions in the present natural range (in most cases) completely overlaps the bagplot of the climatic conditions in the current range. The relative area that the current range bagplot overlaps with the present natural range bagplot ($Area_{overlap}/Area_{present\_natural}$) reveals how much of the present natural climatic diversity is lost.

The relative overlap has a value between 0 and 1, which respectively indicates a highly and hardly affected environmental diversity. To indicate the environmental diversity loss instead of the overlap, the percentage complement of the overlap is used (subtracting it from 1).

IUCN status

The species' range data of the PHYLACINE does not include any form of information on species populations, their occurrence or density. No additional large-scale dataset could be found to supplement the population information for large herbivorous mammals or species-specific conservation efforts. Therefore the choice was made to use the IUCN red list threat status as an indicator for the relative population fitness (IUCN, 2020).

The criteria for the IUCN status categories (figure 1) resemble population fitness indicators, as, for each species, they are based on population size, population decrease, extent of occurrence and the size and state of the distribution range (IUCN, 2012).

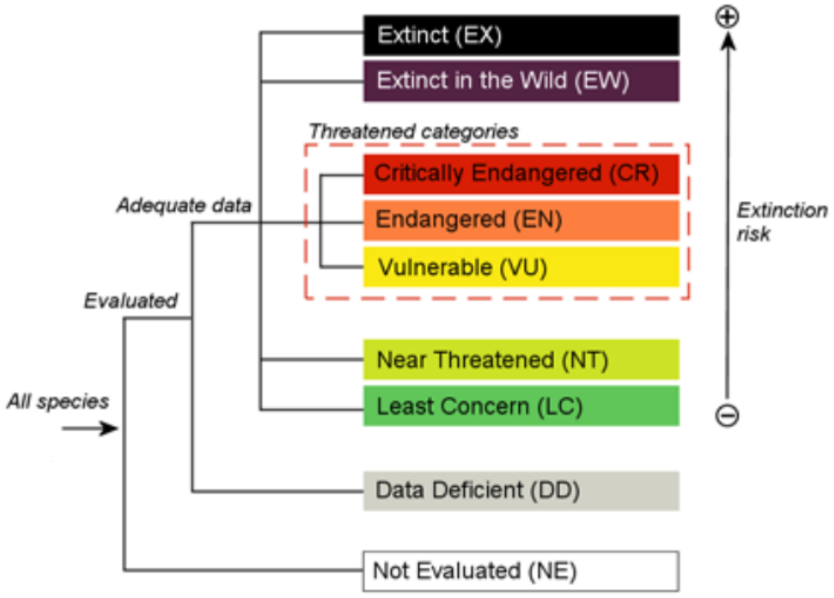


Figure 1 Structure of the IUCN threat categories (IUCN, 2012)

Human activity

To test how likely it is a species' range decline is related to anthropogenic activity, the anthropogenic disturbance was estimated by evaluating the land use and human presence in the ranges of a species. In theory, the lost range of a refugee species should be more targeted by anthropogenic pressure than the current range.

The land use and population density were taken from the History Database of the Global Environment (HYDE) 3.2 database, which contains reconstructed (pre-)historic global land use and human population data for the last 12 000 years (Goldewijk et al., 2017). Land use over time is estimated with a hindcasting approach, combining existing land cover data with information about (pre-)historic human population. A near constant rate of land use per capita is assumed while using changing allocation algorithms for different time periods (Goldewijk et al., 2017).

The 2015 CE data maps on the population density (inhabitants/km^2 per grid cell), cropland area (km^2/grid cell), grazing area (km^2/grid cell) and total built-up area (km^2/grid cell) from the HYDE dataset were used. These come in a 5-arc minute resolution grid, amounting to 85 km^2 grit cells at the equator, and were projected to the equal area CRS from the PHYLACINE dataset. The lost distribution range was created and polygonised for each species with the PHYLACINE data. The anthropogenic data was extracted from the four 2015 CE HYDE raster maps with the overlapping lost and current polygon range maps.

The land use, given as a km^2/cell, was converted to km^2/km^2 (anthropogenic land use/total range) with the total range area. The population density, given as inhabitants/km^2 per cell, was converted to a range wide uniform inhabitants/km^2 total range. The difference in anthropogenic disturbance was measured by the absolute difference of the population density and categorical land use (cropland, grazeland and build environment) in the current and lost ranges of each species.

Data Analysis to integrate the information on refugee species criteria

The assessment was concluded with a species comparison to evaluate their probability of being a refugee species. To this end, I ran a principal component analysis (PCA) combining the results of the above refugee species criteria. A PCA clusters species based on their similarity in terms of multivariate space. So in my case, it clusters species that have similar values for the different refugee species criteria. The species' potential refugee status was further analysed by conducting a k-mean cluster analysis to detect the trends in the results and find clusters of species with similar changes. The formed clusters were studied to classify the species' refugee status potential.

In a PCA the correlation of each variable is determined by calculating the distance between the datapoints and possible axes. Based on the lowest distance sum new axes are formed, known as the principal components, and used as a completely new plot orientation. The result is a graph with the new orientation of the principal components axes and the species scattered in the plot. Correlated species will cluster together in this graph, indicating a similarity, which in this case is the extent to which they meet the different criteria of being a refugee species.

The results used to conduct the PCA were the relative loss of range, diversity index and climatic conditions, the absolute difference in anthropogenic land use and population density, and the qualitative IUCN status. A PCA can only be conducted with quantitative data, making it impossible to include the qualitative IUCN status. While the IUCN status are qualitative in form, they are quantitative in nature as they represent a scale of decreasing fitness. A quantitative form of this IUCN status scale is suitable in a factor analysis, and was converted by simply allocating the status a number of 1 to 6, with 6 being the most severely threatened.

All PCA input values were normalised per variable with a max-min normalisation and scaled to unit lengths to assure equal importance in the analysis.

The k-mean cluster analysis also uses the distance between the species to rank them into clusters. The number of clusters was decided based on the elbow method by plotting the potential number of clusters against their calculated in-cluster variance. The trade-of between the inclusion of more clusters and the added variance shows an elbow point at the point where only little is gained by adding more clusters.

Results

Range decline

About a third of all the species (27 species) did not show any range loss, making it impossible for them to be identified as refugee species in this initial assessment. The same goes for the two species that are extant but lack a recognised current range: these are recognised as officially (*Oryx dammah*) and possibly (*Bos sauveli*) extinct in the wild (Appendix A).

One species (*Ceratotherium simum*) has a larger current range than present natural range. The origin of this surprising range gain lies in the creation of the species' distribution ranges and does not reflect a probable reality (Appendix A). It is not possible in this initial assessment to identify a refugee status probability of this species.

An understanding of the severity of the range decline of the rest of the species (n=49) can be obtained by looking at the relative range loss. The collective results are shown in figure 2. Noticeable is the large number of species on both far ends of the range loss. A total of 12 species lost less than 15% of their present natural range while 14 species endured a significant range loss of more than 85%.

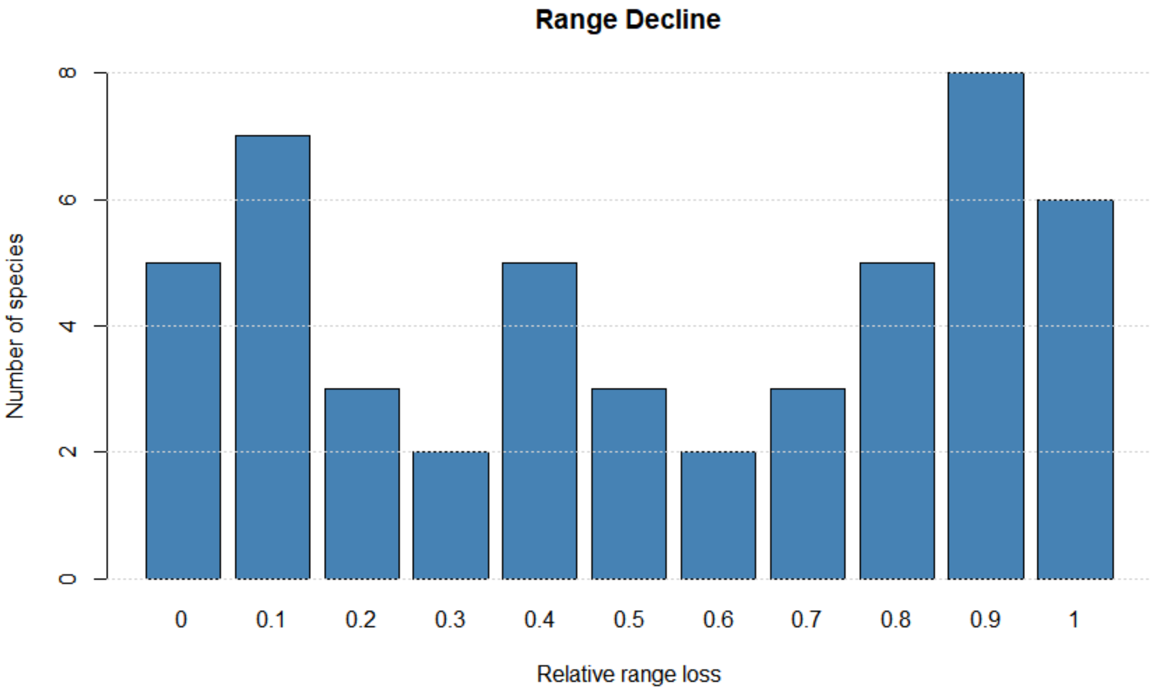


Figure 2 Overview of the fractional loss of the present natural range of all species, rounded to the nearest number (relative range loss = $\frac{Range_{lost}}{Range_{present\ natural}}$*).*

Habitat diversity

The severity of the habitat diversity loss is revealed with the relative Shannon's diversity loss and shown in figure 3. The ecoregions diversity index is shown to be very similar in both ranges for most species. The graph shows an exponential decay in the habitat diversity loss of species.

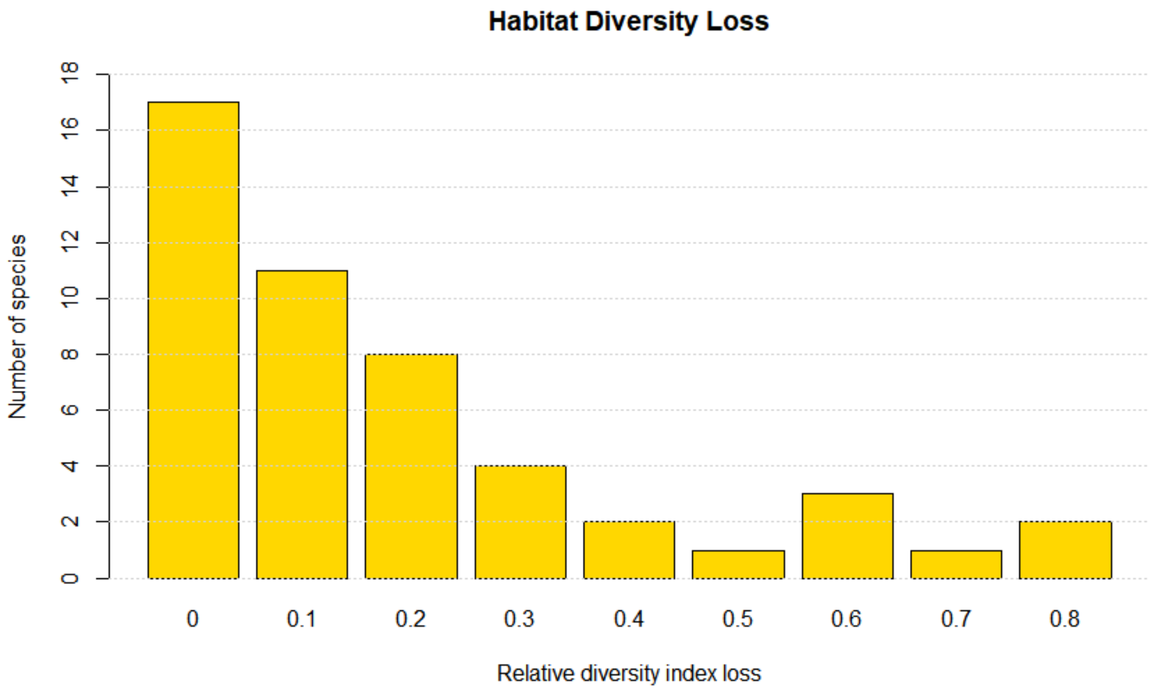


Figure 3 Overview of the habitat diversity losses, represented by the relative change (1 - $\frac{current}{present\ natural}$) in the ecoregion diversity index, rounded to the nearest number.

Environmental diversity

The environmental diversity loss is represented by the relative loss in overlap of the bagplots of the climatic conditions and shown in figure 4. The environmental diversity loss is minor for most species. Only a third of the species lost more than halve of their environmental range diversity.

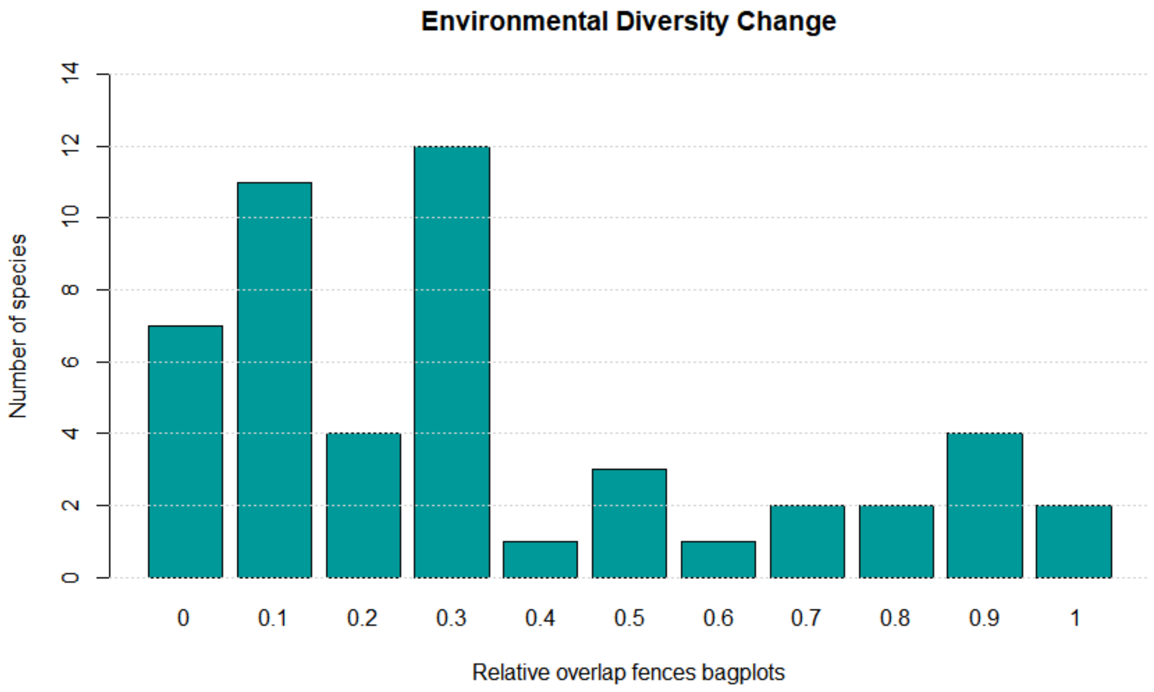


Figure 4 Overview of the environmental diversity losses, represented by the overlap loss of the climatic conditions, rounded to the nearest number.

IUCN status

The number of species with each IUCN threat status is shown in figure 5. About a third of the species fall in the lower concern categories (least concern and near threatened) and the rest in the threatened categories (vulnerable, endangered and critically endangered). Surprisingly there is a species (*Elaphurus davidianus*) recognised as extinct in the wild, which is most likely to be an internal incoherency of the PHYLACINE data base (Appendix A).

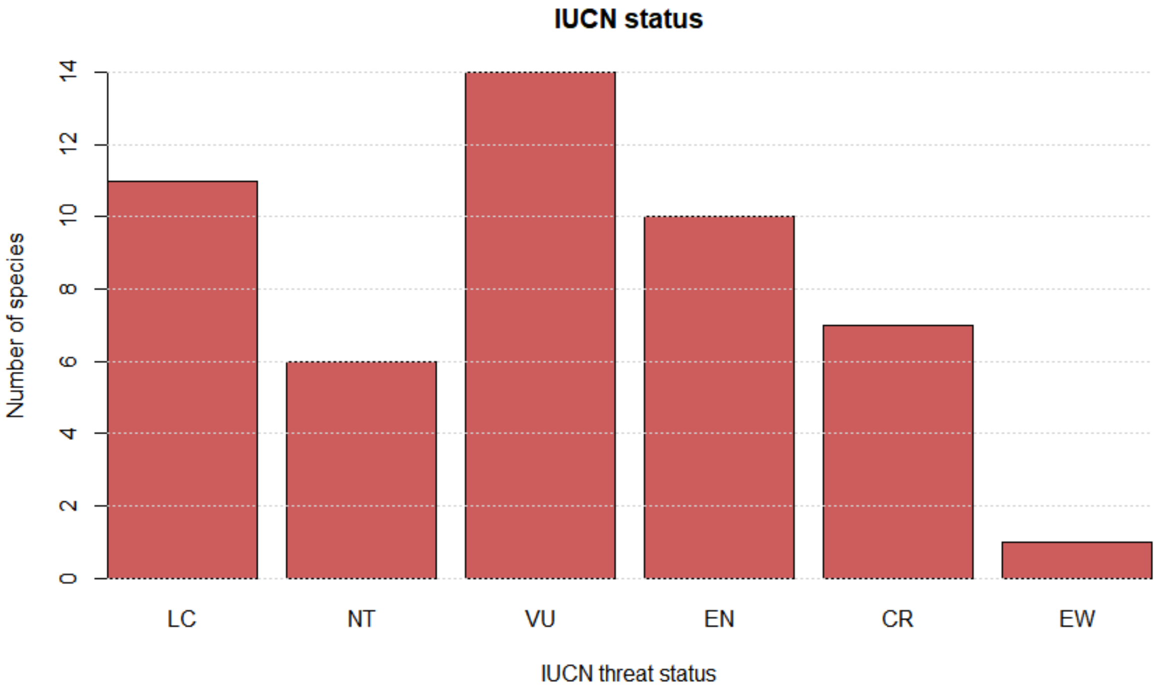


Figure 5 Overview of the IUCN threat status of all species.

Anthropogenic pressure

The change in the human presence in the current and lost range of all species is shown in figure 6. The difference in human population density is shown to be very moderate for most species, with an increase of 0±25 inhabitants/km^2 for almost half of the species. Still the human population density in the lost range of some species is significantly larger than in the current range. On the other hand, for a total of 9 species the human population density in the current range is larger than in the lost range, of which the difference for 4 species is big enough to be shown in the graph.

The change in anthropogenic land use in the current and lost range of all species is shown in figure 7 and 8. Figure 7 shows the categorical land use difference per species, showing varying results for the different categorical land uses in the ranges of each species. For most species the urban area is shown to be larger in the lost range than in the current range, although it only consists of a very small portion of the total anthropogenic land use change. The difference in cropland and grazeland are deciding factors in the total difference in land use, although these are shown to counter each other in some cases.

Figure 8 shows an overview of the sum of the difference in the anthropogenic land use. The lost range consists of more anthropogenic land use for most species, but the difference in the total anthropogenic land use is also shown to be relatively minor for most species. The results resemble a normal distribution with an anthropogenic land use increase of only 10% as the mean. For a total of 10 species there is more anthropogenic land use in the current range than in the lost range, of which the difference for 7 species was big enough to be shown in figure 8.

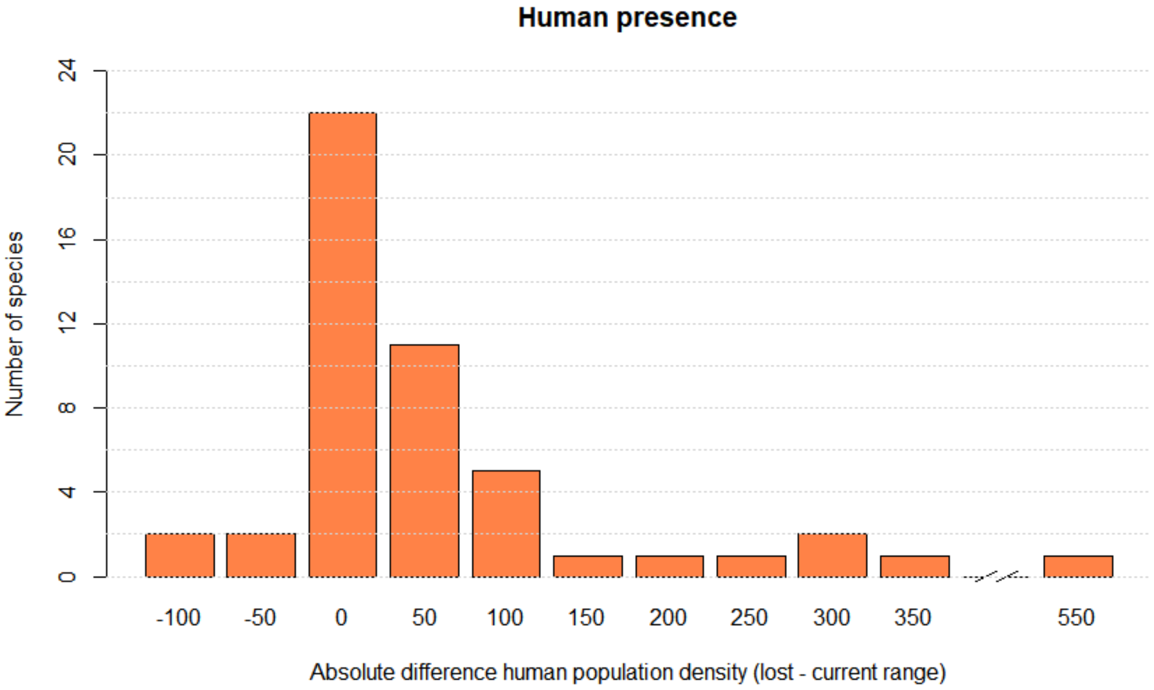


Figure 6 Overview of the human population density difference [average inhabitants/km^2] between the lost and current range of all species

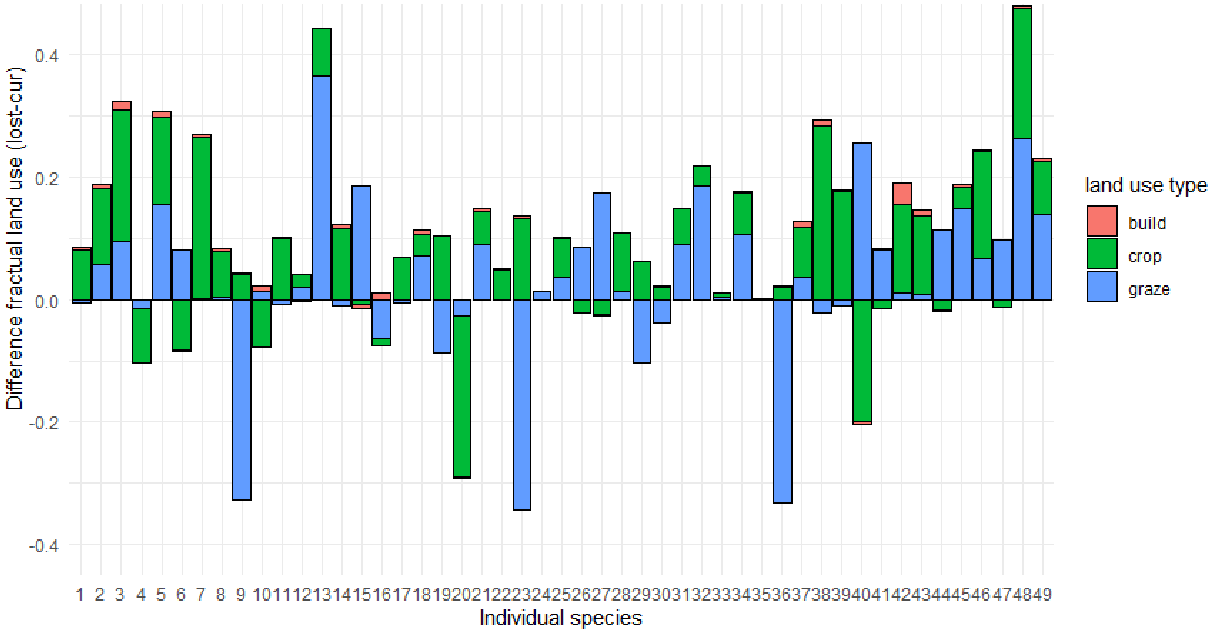


Figure 7 Overview anthropogenic land use difference between the lost and current range of each species for all categories.

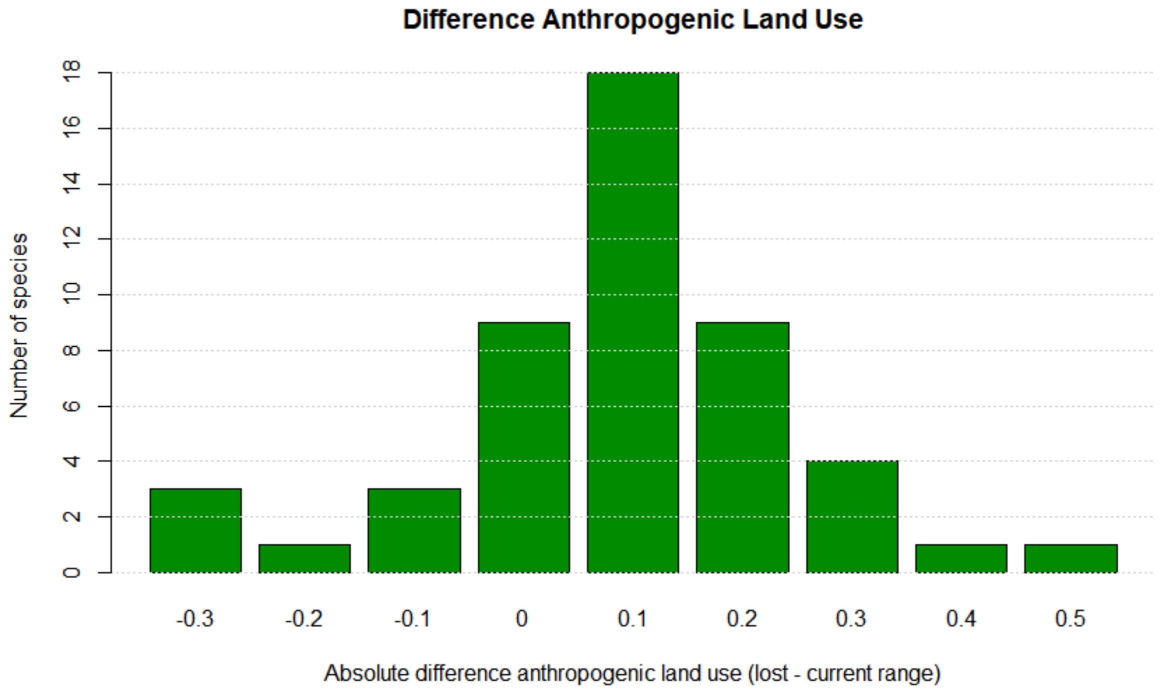


Figure 8 Overview of the total difference in anthropogenic land use [km^2 land use/km^2 range] between the lost and current range of all species

PCA clustering analysis

An overview of the contributions to the three main principal component axis is shown in figure 9. The first principal component axis has a variance of 48% and is mostly influenced by the relative climatic diversity loss, habitat diversity loss and range loss. The second principal component axis has a variance of 17,5% and is mostly influenced by the anthropogenic land use. The third principal component axis has a variance of 17% and is mostly influenced by the difference in human population density.

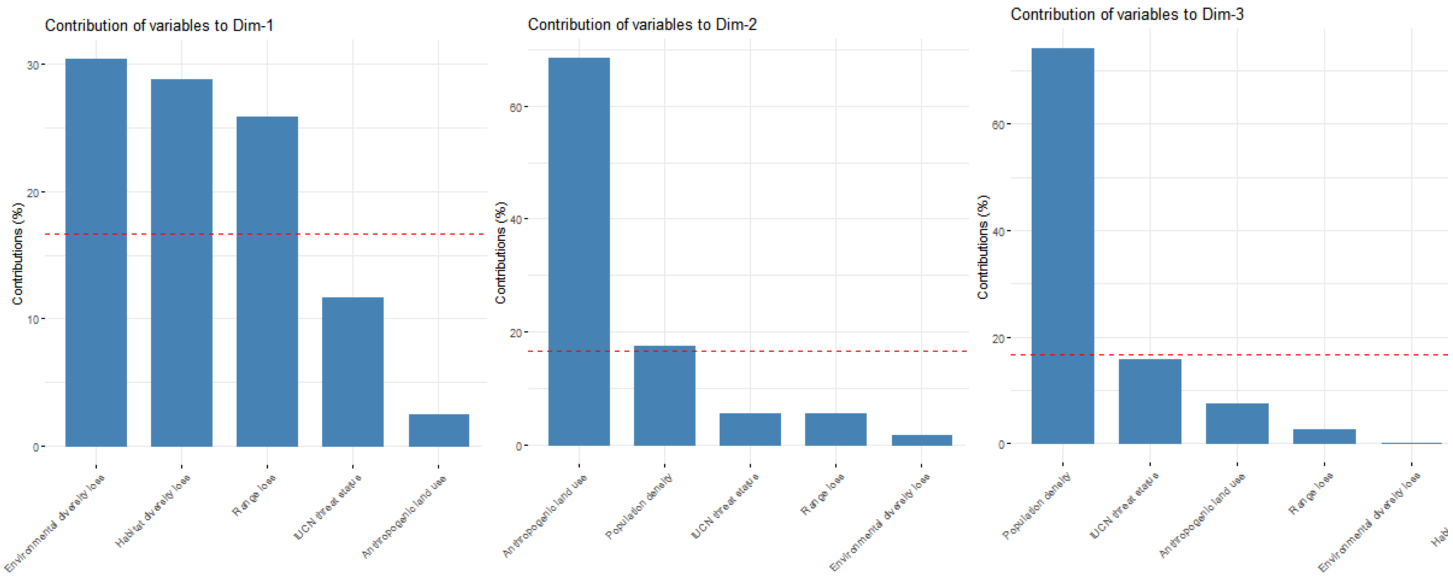


Figure 9 Visualisation of the driving variables in the three constructed main principal component axes

Clusters

The number of clusters was chosen based on the elbow method, which searches for the trade-of balance between the number of clusters and the increase of variability they bring. The elbow in the line marks the point after which the variability change is relatively small compared to the increase of clusters. The potential number of clusters and their increase in variability are plotted in figure 10. In this case there is no easily identifiable elbow, but a cut-off point can be identified at 5 clusters.

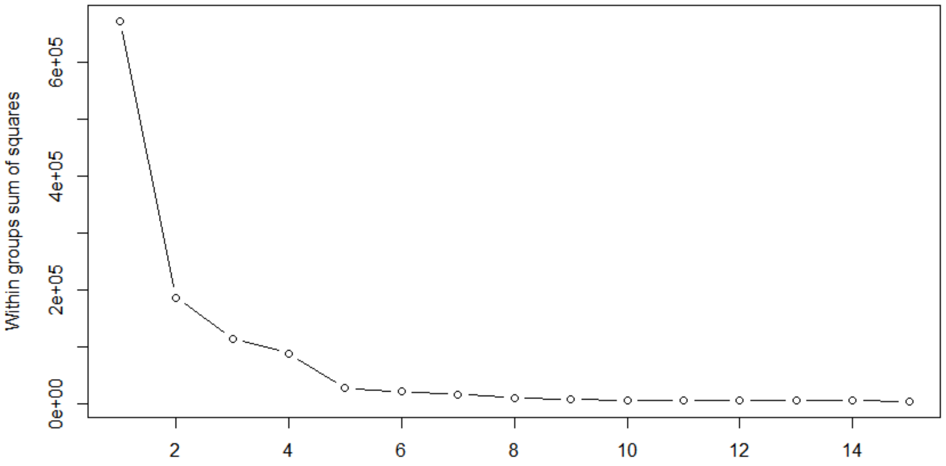


Figure 10 Potential number of clusters and their variability to find the most suitable number of clusters, in this case 5.

The five clusters resulting from the k-mean clustering are colour coded and shown on the three most important principal component axis in figure 11. The clusters are not easily distinguishable and are mainly based on the first principal component axis (climatic diversity, habitat diversity and range loss). Still, the differences are clear enough to differentiate their refugee status probability. Table 1 shows an overview of the clustered species with their parameter values.

The first cluster (black points) mainly differentiates itself based on the large population density in the lost range compared to the current range. The species in this cluster also score high on the first principal component axis, which means there are only small relative diversity and/or range losses. The species in this cluster seem to be affected by human presence, but do not seem to be confined to suboptimal habitat. This combination makes it not likely for this cluster to include refugee species.

The second cluster (dark blue points) differentiates itself with the significantly high range loss, environmental diversity loss and habitat diversity loss. Additionally, all species in this cluster are threatened. Although the human population density and anthropogenic land use difference in the current and present range vary a lot, the species in this cluster have a high refugee status potential.

Species in the third cluster (green points) have a high relative range loss but only moderate diversity losses. For all species the human population density is higher in the lost range than in the current range. Species included in this cluster have a good possibility of being refugee species.

The fourth cluster (light blue points) and fifth cluster (red points) are very similar and have no remarkable differentiations. The species in each cluster have varying values for each variable, with only one or two criteria per species indicating a refugee status potential. The difference between the species in the fourth and fifth cluster is the IUCN threat status. The fourth cluster only contains non-threatened species while all but one of the species in the fifth cluster are threatened. This makes it very unlikely for the fourth cluster to contain refugee species. It is difficult to categorise the species in the fifth cluster, as the cause of their high(er) extinction risk can not be explained with one of the researched variables. There is a chance that (some of) these species are unrecognised refugee species that were not identified in this initial assessment due to the use of data that was incomplete or too general. This cluster is therefore marked as uncertain to include refugee species.

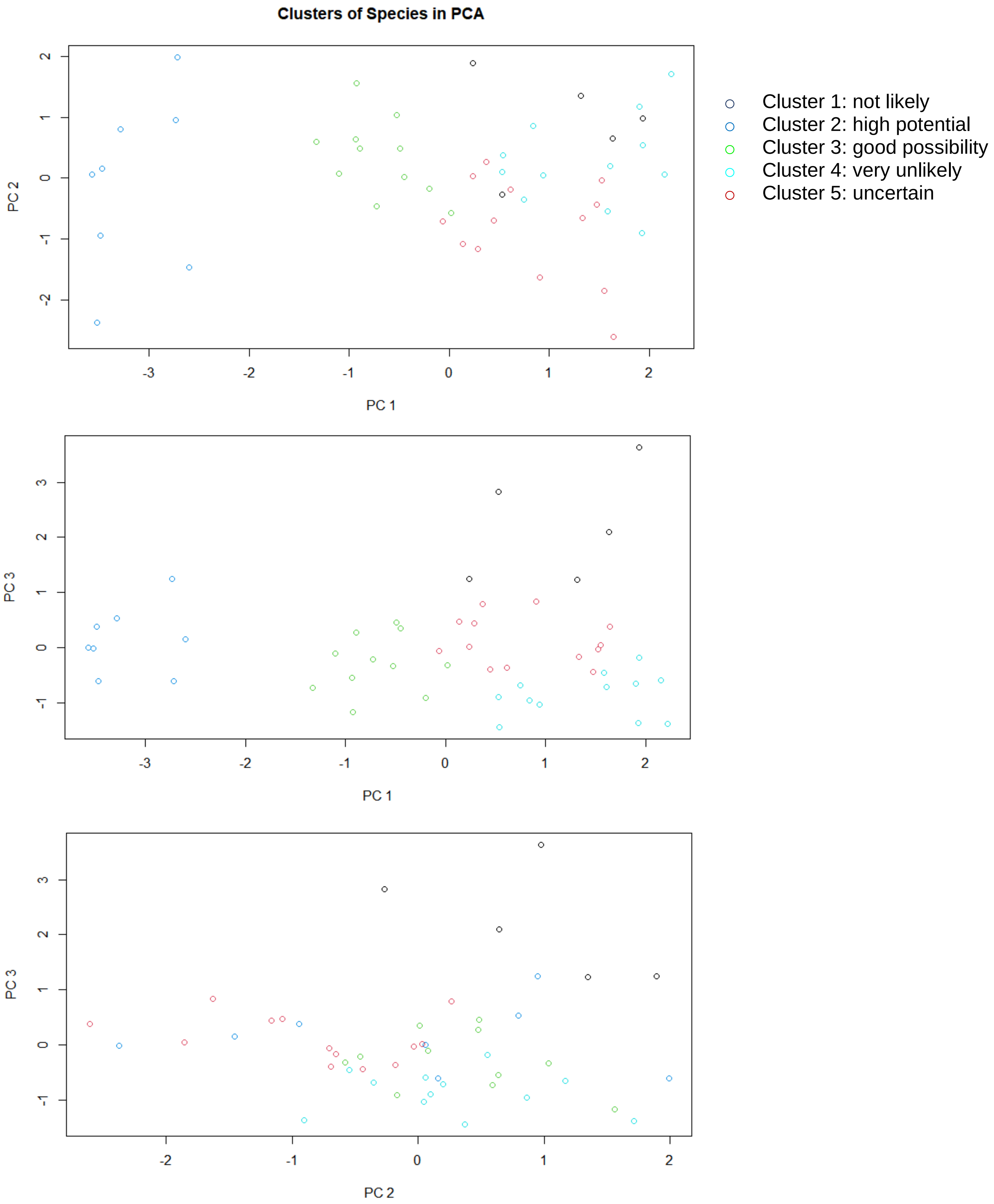


Figure 11 Visualisation of the clustered species resulting from the PCA and k-mean clustering

Table 1 Overview of clustered species and the values of the variables the clusters were based on

Name	**Relative range loss**	**Relative habitat diversity loss**	**Relative environmental diversity loss**	**Absolute population density difference**	**Absolute anthropogenic land use difference**	**IUCN threat number**
Cluster 1: Not likely						
Bos_gaurus	0,390	0,042	0,111	259	0,27	3
Boselaphus_tragocamelus	0,082	0,037	0,248	574	-0,06	1
Choeropsis_liberiensis	0,147	0,306	0,119	345	-0,07	4
Rhinoceros_unicornis	0,880	0,094	0,256	294	0,27	3
Tapirus_bairdii	0,035	0,000	0,114	298	0,19	4
Cluster 2: High potential						
Bison_bonasus	0,998	0,804	0,962	-10	0,00	3
Bos_mutus	0,855	0,580	0,760	21	-0,28	3
Camelus_ferus	0,915	0,592	0,882	12	0,44	5
Dicerorhinus_sumatrensis	0,985	0,502	0,858	195	0,11	5
Elaphurus_davidianus	0,983	0,399	0,936	-108	-0,29	6
Equus_africanus	0,981	0,845	0,757	5	0,05	5
Equus_ferus	0,998	0,644	0,960	43	-0,21	4
Rhinoceros_sondaicus	0,996	0,678	0,888	102	0,13	5
Cluster 3: Good possibility						
Ailuropoda_melanoleuca	0,876	0,186	0,568	121	0,08	3
Bison_bison	0,925	0,247	0,674	18	0,30	2
Bos_javanicus	0,792	0,161	0,288	85	0,08	4
Bubalus_arnee	0,943	0,257	0,382	34	0,09	4
Elephas_maximus	0,755	0,106	0,456	115	0,15	4
Equus_hemionus	0,908	0,362	0,653	30	0,10	2
Hippopotamus_amphibius	0,761	0,171	0,493	23	-0,02	3
Loxodonta_africana	0,654	0,057	0,333	8	0,01	3
Ovibos_moschatus	0,699	0,345	0,300	1	0,00	1
Rucervus_duvaucelii	0,913	0,243	0,496	29	0,17	3
Tragelaphus_derbianus	0,793	0,344	0,302	65	0,24	3
Cluster 4: Very unlikely						
Alcelaphus_buselaphus	0,119	0,056	0,122	106	0,19	1
Alces_alces	0,125	0,048	0,202	75	0,32	1
Cervus_canadensis	0,401	0,060	0,338	-8	0,11	1
Cervus_elaphus	0,553	0,160	0,324	-42	0,17	1
Damaliscus_lunatus	0,732	0,078	0,266	20	0,06	1
Equus_quagga	0,504	0,055	0,215	-4	0,06	2
Hippotragus_equinus	0,271	0,045	0,138	34	0,15	1
Kobus_leche	0,031	-0,005	0,000	11	0,22	2
Sus_scrofa	0,126	0,011	0,022	-89	0,05	1
Syncerus_caffer	0,271	0,019	0,005	10	0,07	2
Ursus_americanus	0,217	0,025	0,046	12	0,48	1
Ursus_arctos	0,443	0,130	0,349	32	0,23	1

Cluster 5: Uncertain

Babyrousa_babyrussa	0,024	0,007	0,000	-16	-0,10	3
Bubalus_mindorensis	0,052	0,111	0,344	2	0,04	5
Diceros_bicornis	0,363	0,046	0,262	20	0,02	5
Equus_grevyi	0,812	0,065	0,124	12	0,01	4
Equus_zebra	0,154	0,015	0,060	-15	0,15	3
Giraffa_camelopardalis	0,565	-0,049	0,275	17	0,11	3
Gorilla_gorilla	0,029	0,031	0,109	34	-0,04	5
Oryx_beisa	0,431	0,222	0,152	33	0,18	4
Ovis_ammon	0,024	-0,001	0,000	21	-0,31	2
Tapirus_indicus	0,540	0,165	0,098	139	0,14	4
Tapirus_pinchaque	0,483	0,127	0,049	-39	0,10	4
Tapirus_terrestris	0,097	0,022	0,116	47	0,19	3
Tremarctos_ornatus	0,174	0,020	0,092	9	0,08	3

Remaining: no range loss

Babyrousa_togeanensis	0					4
Budorcas_taxicolor	0					3
Capra_sibirica	0					1
Capra_walie	0					4
Capricornis_milneedwardsii	0					2
Capricornis_thar	0					2
Cervus_albirostris	0					3
Connochaetes_gnou	0					1
Connochaetes_taurinus	0					1
Damaliscus_pygargus	0					1
Dugong_dugon	0					3
Equus_kiang	0					1
Gorilla_beringei	0					5
Hippotragus_niger	0					1
Hylochoerus_meinertzhageni	0					1
Kobus_ellipsiprymnus	0					1
Lama_guanicoe	0					1
Okapia_johnstoni	0					4
Oryx_gazella	0					1
Rusa_unicolor	0					3
Tragelaphus_buxtoni	0					4
Tragelaphus_eurycerus	0					2
Tragelaphus_oryx	0					1
Tragelaphus_strepsiceros	0					1
Trichechus_inunguis	0					3
Trichechus_manatus	0					3
Trichechus_senegalensis	0					3

Discussion

Key results summary

The aim of this research was to perform a global scale initial assessment of potential refugee species among large herbivorous mammals. The assessment was based on the criteria of range reduction, range diversity loss, compromised population fitness and associated anthropogenic pressure. Of the 79 included species, 27 species did not show any range loss, 2 species have no registered current range and 1 species has a current range that is too inexact to be included. For the 49 remaining species the results show that the relative range losses were very low or high for most species, the diversity losses were minor for most species, two-thirds of the species are recognised as threatened and the anthropogenic land use and population density only increased slightly for most species. The finalisation of the initial refugee status assessment with a PCA did not result in clusters that are so strongly related that they are immediately distinguishable. But with the k-means clustering five clusters were successfully differentiated with distinguishable refugee status potential.

Interpretation

The initial assessment of the refugee status potential is a rough assessment with gradual results and no defined variable threshold for when a refugee status criterion is met. This makes it impossible to define a refugee status probability for the clusters, but does enable a hierarchical ranking in the likelihood of a refugee status potential for the species in a cluster. The clusters were ranked as very unlikely, not likely, good possibility, high potential and uncertain to include refugee species.

The similarity of the different variables is the driver of the species' clustering. To be more precise, the clustering is driven by the similarities of the principal components (PC). This can be seen by looking at the PCA plots with the principal component axis (figure 11), in which it can also be seen that not all variables are equally important in the forming of the final clusters.

Looking at the influence of the different variables on the formed clusters, it is noticeable that the first principal component is the most deciding clustering factor. PC1 consists of three almost equally contributing variables: the relative environmental diversity loss, habitat diversity loss and range decline (figure 9). PC1 is the main factor in forming cluster 2 (high refugee status potential), and an important factor in the distinction of the third cluster (good possibility).

The influence of the diversity losses and range decline on forming the of the primary principal component axis and the clusters is quite significant, and thus the primary driver in the assigned refugee status potential. This major influence falls in line with the first phase refugee status identification framework of Kerley et al. (2012), as two of the three key components of their assessment are represented by the diversity and range losses. The reason the PCA fitted these three variables in one principal component could be influenced by the fact that these all are functions of the current and present natural ranges. The diversity losses are however not simply a result of the range loss.

Principal component 3, mainly contributed by the population density, is an important driver in the forming of cluster 1 (not likely refugee species). The cluster only contains species with a much higher population density in the lost range than in the current range. The second major differentiator of the cluster is the very small range and diversity loss, and is therefore still not likely to include refugee species. Further clustering depends minimally on the population density. The same goes for the IUCN threat status, which is also only decisive in the differentiation between the fourth and fifth cluster. And although the cluster with the highest refugee status probability only consists of species that are threatened, not all species with a good possibility of being a refugee species are threatened. The anthropogenic land use change variable does not seem to be a significant differentiator for any of the clusters.

The lack of a strong correlation of the variables to the refugee status probability could indicate that the variables are not decisive in the identification of refugee species or that the calculated indicators are not a proper representation of the variables. The IUCN threat status for example is used to represent the population fitness criterion in the first phase assessment framework of Kerley et al. (2012): species which maintain slow population growth and/or low densities despite conservation efforts. I still believe the IUCN threat status to be a good representation of the population fitness, but it covers more than the population growth and density. Isolating these parameters might be a more effective in the identification of refugee species. Alternatively, the reflection on the conservation efforts could be vital missing data.

Similar uncertainties exist for the anthropogenic pressure variables. While a range decline caused by human activity is a necessary condition for a refugee status, the driving anthropogenic pressure might not be reflected by the current population density and land use. Historic human activities like hunting, current managed efforts or even the distribution of the human population (density) can be important factors in the identification of a refugee status potential, but are not properly represented in my assessment.

Implications

A small collection of species has been evaluated in existing academic literature on their (partial) refugee status potential, of which some were also included in the initial assessment of this research.

The European bison (*Bison bonasus*) was identified as a refugee species by Kerley et al. (2012). They built a case that the species is a grazer and open habitat species that was historically mistakenly seen as forest species. They argue evidence shows a preference and adaptation for (semi-)open habitats and not for forest habitat. The historical range distribution of the European bison covered most of the European continent and it is hypothesized that the combination of postglacial vegetation changes and anthropogenic pressure avoidance forced the bison into forest habitat. My initial assessment corresponds well with this conclusion, as it identifies the European bison as a part of the cluster with a high refugee status potential based on a large decline in the range extent and diversity. The mentioned anthropogenic pressure driving the Bison to forest habitat is not identified by my initial assessment, which even shows a slightly larger average population density in the current range than in the lost range. Kerley et al. (2012) do however mention a switch in the forest habitat driving anthropogenic pressure from hunting to management. Human management of species is not measured in my initial assessment and neither are historic anthropogenic pressures like historic hunting activities.

The giant panda (*Ailuropoda melanoleuca*) was recently identified as a possible refugee species by Kerley et al. (2020), challenging the assumptions of giant pandas belonging in high altitude forests and being a “bamboo specialist”. They argue the giant panda's predominant habitat of high altitude bamboo forests is the result of a recent (over the past 3,500 years) range contraction and shift that coincides with increasing anthropogenic activity and land use (Kerley et al., 2020). Kerley et al. (2020) also point out the giant panda's digestive system is not specifically adapted to a bamboo-dominated diet and that the species is still vulnerable despite long ongoing intensive conservation efforts. My initial assessment identified the giant panda as a part of the cluster with a good refugee status possibility. It shows the large range and climatic diversity decline and a clearly larger population density in the lost range than in the current range. The small habitat diversity loss is the most important criterion that is not supporting a high refugee status probability. This result is unexpected, as Kerley et al. (2020) highlight the species' retreat to ecologically suboptimal area. As the suboptimal retreat habitat consists of high-altitude forest, the habitat diversity loss discrepancy might very well be the result of the execution of my assessment. The range area of the species in the PHYLACINE dataset is far from detailed and marks a whole cell as a part of the species' range if only a small part of it is actually inhabited. This could result in an inaccurate habitat diversity representation of the range, especially in very diverse areas like mountainous regions.

The Cape mountain zebra (*Equus zebra*) has been identified as a partial refugee species by Lea et al. (2016), stating 12 out of 21 populations can be identified as refugee species based on their population performance and habitat and diet quality. Lea et al. (2016) argue these populations are managed in protected areas with low anthropogenic value but that are equally ecologically unsuitable as habitat for the Cape mountain zebra. My initial assessment actually clusters the Cape mountain zebra with the species that are uncertain to be refugee species. The criteria do not identify a refugee status probability but do indicate a compromised population fitness without identifying a probable cause.

The specifics of this species made it hard for my assessment to accurately assess the refugee status criteria. First of all, the present natural (constructed prehistoric) range does not cover the whole current range, creating an inaccurate representation of the range decline. Lea et al. (2016) argue the habitat quality is too low in a part of the current range to support the species, which could be an explanation of the discrepancy in the current and present natural range. Secondly, the part of the current range with an insufficient quality is mostly undisturbed by humans, creating a shifted representation of the anthropogenic pressure comparison. Finally, the diet quality assessment of Lea et al. (2016) is driven by the occurrence of suitable vegetation, which is not properly represented by the ecoregions or climatic conditions.

In conclusion, I argue that my approach is not detailed enough to register changes in habitat characterisations that are important in the identification of (partial) refugee species. The used ecoregion diversity change is too general to pick up important habitat suitability characteristics like vegetation differences. Suspected (partial) refugee species from existing academic literature that based the refugee status assessment on more detailed habitat diversity evaluations (like the Giant panda by Kerley et al. (2020) and the African elephant by Moolman et al. (2019)) are also picked up by my assessment, but on a lower probability level.

My assessment is also not elaborate enough in the assessment of anthropogenic pressure, especially when the driving anthropogenic pressure was mostly relevant in historic context. My human influence assessment is limited to the current situation and was not developed to assess the arguably more important historic anthropogenic dynamics.

Limitations of this research

The assessment heavily depends on the current and present natural range of each species. The only variable included in the PCA that was independent of the ranges was the IUCN threat status. In this research these ranges were taken from the PHYLACINE dataset, which are rather coarse range indicators and did not all come without complications (Appendix A). Since the final ranges have such a direct and possibly strong influence on the results of the research, the choices and assumptions made in the construction of the ranges could be drivers in the resulting refugee status potential.

This becomes more of a concern when the constructed ranges are unexpected or based on unclear information. An example is a present natural range that does not completely cover the current range. The expectation is that the constructed present natural range includes the whole current range of a species, with the exception of species that are introduced in new areas. This is however not the case for the *Equus zebra* and *Diceros bicornis*, which influences the results of my assessment and potentially altered the cluster they are classified in. Another example of an unclearly constructed range is a present natural range based on 'suitable' habitat. No explanation could be found on how exactly the present natural ranges are (partly) constructed based on suitable habitat. Most of these present natural ranges that are (partly) based on suitable habitat are significantly bigger than the current range, which forms a big contrast to the minimally expanded ranges of species solely based on information from the IUCN.

A key component of my initial global assessment of refugee species status was the use of general information from global datasets to approximate the assessment criteria listed by Kerley et al. (2012). The obvious downside of this approach is that this information does not include species specific characteristics beyond the ranges. Species specific sensitivity to (certain) human activity and optimal habitat conditions like vegetation diversity could not be included in my assessment while these characteristics might be crucial in the identification of refugee species. Therefore, the clusters should be seen as an indication of a possible refugee status and the species should be individually evaluated for an actual refugee status conclusion. The species in the more probable refugee status clusters are more interesting, but the species in less likely refugee status clusters could still very well be (partial) refugee species.

The assessment of the change in environmental diversity in this research implies a near constant climate over the last ~12000 years. This is a major simplification that can result in a false environmental diversity loss representation, especially for species whose range decline was (pre-)historically influenced by climatic shifts. A blind spot for the historical climatic habitat conditions could result in a misidentification of the refugee status (potential) and an underestimation of the severity of the effects of a suboptimal habitat. Awareness is especially important in light of future climatic changes.

Recommendations

The results of the assessment point out species that are interesting for a future individual in-depth refugee status evaluation. A fourth of the included species (18/79) are shown to have a serious potential of being a refugee species. Although the included subset of species was deliberately chosen based on their higher possibility of including refugee species, this is still a considerable number. It highlights the importance of more research on this relatively new and unknown concept to better understand and protect wildlife.

The method of the assessment is shown to be effective in a rough classification of refugee status potential, at least for most of the academically suggested refugee species. The initial assessment could be applied to more species and might be a useful component in elaborate refugee status assessments. It could also be used to expand our understanding of refugee species by evaluating the relation of additional data to the concluded refugee status potential. New relations might be uncovered as refugee species could be typified by certain currently unknown characteristics.

The method could also be useful out of the academic scope by conservation planners and managers to enhance conservation efforts. It can be used as an initial assessment to verify whether the species they are working with are potentially refugees. Then the species' natural history can be (re)assessed to lift limiting range confinements and restore expected habitat diversity losses. When implementing the restorations the species should be researched to test the developed hypothesis.

The initial assessment method is also shown to have multiple aspects that can be improved. One of the most important aspects would be a further evaluation of the relation between the refugee status potential and the chosen variables of the IUCN treat status and anthropogenic pressure. These might need to be changed to better fit the refugee status criteria or even replaced with stronger criteria indicators. More detailed species' population fitness information might need to be used where possible. The research on which the threat status is based could include information on population numbers, trends or densities and existing conservation efforts.

The anthropogenic pressure was only evaluated for the year 2015 CE in this research, not fully utilizing the potential of the in the HYDE dataset included (pre-)historic data. Including historic anthropogenic pressure, especially with historic range data of species, could offer a more complete assessment of the refugee status potential. A shown possible link between the range decline and human activity would be a valuable addition to the assessment.

The habitat diversity loss could be improved making a more detailed habitat diversity assessment. This could be realised by looking at the vegetation diversity or richness.

Future implications

The results and the reflection of what they mean and implicate show how relatively unexamined the refugee species concept still is and how important the knowledge and insight is for species understanding and protection. So many species still need to be evaluated and uncovered as possible refugee species, while the true conservation efforts only starts at the moment a hypothesis has been developed on how and why a species became a refugee. That is the moment alterations can be designed, implemented and evaluated to ensure the preservation of the individual species and the general biodiversity.

The continuous decline of species despite the growing protection efforts, the changing climate and decline of wilderness area make it an urgent matter to expand the refugee species knowledge and management inclusion. My refugee status potential assessment might be a step in this direction.

Acknowledgements

First of all, I want to thank my supervisor Joris Cromsigt for guiding me through this project. His enthusiasm surrounding the fascinating refugee species concept and his extensive feedback on my questions, considerations and work were extremely valuable to me.

Secondly, I want to thank my additional (co)supervisors Graham Kerley and Kees Klein Goldewijk for their input, feedback, discussions and advise. I feel extremely lucky to have had three knowledgeable people that looked out for me, encouraged me and helped me find direction in the development of my assessment and thesis in general.

References

Abicair, K., Manning, A. D., Ford, F., Newport, J., & Banks, S. C. (2020). Habitat selection and genetic diversity of a reintroduced 'refugee species.' *Animal Conservation*, *23*(3), 330–341. https://doi.org/10.1111/acv.12550

Ali, A. H., Kauffman, M. J., Amin, R., Kibara, A., King, J., Mallon, D., Musyoki, C., & Goheen, J. R. (2018). Demographic drivers of a refugee species: large-scale experiments guide strategies for reintroductions of hirola. *Ecological Applications*, *28*(2), 275–283. https://doi.org/10.1002/eap.1664

Bartlett, L. J., Williams, D. R., Prescott, G. W., Balmford, A., Green, R. E., Eriksson, A., Valdes, P. J., Singarayer, J. S., & Manica, A. (2016). Robustness despite uncertainty: regional climate data reveal the dominant role of humans in explaining global extinctions of Late Quaternary megafauna. *Ecography*, *39*(2), 152–161. https://doi.org/10.1111/ecog.01566

Broughton, J. M., & Weitzel, E. M. (2018). Population reconstructions for humans and megafauna suggest mixed causes for North American Pleistocene extinctions. *Nature Communications*, *9*(1), 1–12. https://doi.org/10.1038/s41467-018-07897-1

Channell, R., & Lomolino, M. V. (2000). Trajectories to Extinction: Spatial Dynamics of the Contraction of Geographical Ranges. In *Source: Journal of Biogeography* (Vol. 27, Issue 1).

Cromsigt, J. P. G. M., Kerley, G. I. H., & Kowalczyk, R. (n.d.). *The difficulty of using species distribution modelling for the conservation of refugee species — the example of European bison*. *18*(11), 1253–1257. https://doi.org/10.1111/j

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., … Saleem, M. (2017). Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience*, *67*(6), 534–545. https://academic-oup-com.proxy.library.uu.nl/bioscience/article/67/6/534/3102935#87788289

Faurby, S., Davis, M., Pedersen, R., Schowanek, S. D., Antonelli, A., & Svenning, J. C. (2018). PHYLACINE 1.2: The Phylogenetic Atlas of Mammal Macroecology. In *Ecology* (Vol. 99, Issue 11, p. 2626). Ecological Society of America. https://doi.org/10.1002/ecy.2443

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Glikson, A. Y., & Groves, C. (2016a). Cenozoic biological evolution (by Colin Groves). In *Modern Approaches in Solid Earth Sciences* (Vol. 10, pp. 45–67). Springer International Publishing. https://doi.org/10.1007/978-3-319-22512-8_3

Glikson, A. Y., & Groves, C. (2016b). Phanerozoic life and mass extinctions of species. In *Modern Approaches in Solid Earth Sciences* (Vol. 10, pp. 45–67). Springer International Publishing. https://doi.org/10.1007/978-3-319-22512-8_2

Goldewijk, K. K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene - HYDE 3.2. *Earth System Science Data*, *9*(2), 927–953. https://doi.org/10.5194/essd-9-927-2017

IUCN. (2012). *IUCN RED LIST CATEGORIES AND CRITERIA*. www.iucn.org/publications

IUCN. (2020). *IUCN Red List of Threatened Species*. 2020-2. https://www.iucnredlist.org/

Kerley, G. I. H., Beest, M., Cromsigt, J. P. G. M., Pauly, D., & Shultz, S. (2020). The Protected Area Paradox and refugee species: The giant panda and baselines shifted towards conserving species in marginal habitats. *Conservation Science and Practice*, *2*(6). https://doi.org/10.1111/csp2.203

Kerley, G. I. H., Kowalczyk, R., & Cromsigt, J. P. G. M. (2012). Conservation implications of the refugee species concept and the European bison: king of the forest or refugee in a marginal habitat? *Ecography*, *35*(6), 519–529. https://doi.org/10.1111/j.1600-0587.2011.07146.x

Lea, J. M., Kerley, G. I., Hrabar, H., Barry, T. J., & Shultz, S. (2016). Recognition and management of ecological refugees: A case study of the Cape mountain zebra. *Biological Conservation*, *203*, 207–215. https://doi.org/10.1016/j.biocon.2016.09.017

Lyons, S. K., Miller, J. H., Fraser, D., Smith, F. A., Boyer, A., Lindsey, E., & Mychajliw, A. M. (2016). The changing role of mammal life histories in Late Quaternary extinction vulnerability on continents and islands. *Biology Letters*, *12*(6), 20160342. https://doi.org/10.1098/rsbl.2016.0342

Martínez-Abraín, A., & Jiménez, J. (2016). Anthropogenic areas as incidental substitutes for original habitat. *Conservation Biology*, *30*(3), 593–598. https://doi.org/10.1111/cobi.12644

Monsarrat, S., Jarvie, S., & Svenning, J. C. (2019). Anthropocene refugia: Integrating history and predictive modelling to assess the space available for biodiversity in a human-dominated world. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 374, Issue 1788, p. 20190219). Royal Society Publishing. https://doi.org/10.1098/rstb.2019.0219

Moolman, L., Ferreira, S. M., Gaylard, A., Zimmerman, D., & Kerley, G. I. H. (2019). The decline of the Knysna elephants: Pattern and hypotheses. *South African Journal of Science*, *115*(3–4). https://doi.org/10.17159/sajs.2019/4805

Pauly, D. (1995). Anecdotes and the shifting baseline syndrome of fisheries. In *Trends in Ecology & Evolution* (Vol. 10, Issue 10, p. 430). Elsevier Current Trends. https://doi.org/10.1016/S0169-5347(00)89171-5

Qian, T., Chi, Y., Xi, C., Li, Z., & Wang, J. (2020). Changes in the historical and current habitat ranges of rare wild mammals in China: A case study of six taxa of small-to large-sized mammals. *Sustainability (Switzerland)*, *12*(7), 1–19. https://doi.org/10.3390/su12072744

Ripple, W. J., Chapron, G., López-Bao, V., Durant, S. M., Macdonald, D. W., Lindsey, P. A., Bennett, E. L., Beschta, R. L., Bruskotter, J. T., Campos-Arceiz, A., Corlett, R. T., Darimont, C. T., Dickman, A. J., Dirzo, R., Dublin, H. T., Estes, J. A., Everatt, K. T., Galetti, M., Goswami, V. R., … Zhang, L. I. (2016). Saving the World's Terrestrial Megafauna. In *BioScience* (Vol. 66, Issue 10). http://bioscience.oxfordjournals.org

Ruiz-Leotaud, V. (2020). *Protected areas' location may hinder conservation efforts of refugee species*. Sea Around Us. https://phys.org/news/2020-05-areas-hinder-efforts-refugee-species.html

Sandom, C., Faurby, S., Sandel, B., & Svenning, J.-C. (2014). Global late Quaternary megafauna extinctions linked to humans, not climate change. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1787), 20133254. https://doi.org/10.1098/rspb.2013.3254

Smith, F. A., Smith, R. E. E., Lyons, S. K., & Payne, J. L. (2018). Body size downgrading of mammals over the late Quaternary. *Science*, *360*(6386), 310–313. https://doi.org/10.1126/science.aao5987

Timmins, R. J., Burton, J., & Hedges, S. (2016). *Bos sauveli (Kouprey)*. https://www.iucnredlist.org/species/2890/46363360

UNEP-WCMC, IUCN, & NGS. (2018). *Protected Planet Report 2018*. www.unep-wcmc.org

Appendix A – PHYLACINE species' range reflection

The research and analysis are highly dependent on the current and present natural range from the PHYLACINE dataset. Since the present natural range is not a measurable variable, the decisions and assumptions made for its reconstruction are a key driver in the refugee status assessment.

The current distribution ranges provided in the PHYLACINE dataset are based on the spatial dataset on mammal distribution of the IUCN, Version 2016-3 (Faurby et al., 2018). The current range includes areas marked by the IUCN as 'reintroduced' and excludes areas marked as 'introduced', 'extinct', or 'probably extinct' (Faurby et al., 2018).

The present natural range of the PHYLACINE dataset is researched and estimated for each individual species, based on information provided by the IUCN, other literature and suitability assessments (Faurby et al., 2018). The present natural range is not an accurate representation of the (pre-)historical distribution range, as it is intended to be a current range for species if they had never experienced strong anthropogenic pressures. Faurby et al. (2018) grouped the modifications of the present natural range in nine categories, their order reflecting the decreasing certainty. Of these nine categories only six are used for the included species (table 2).

Reflecting on the species-specific construction of the current and present natural range, Faurby et al. (2018) made some decisions that are less favourably for this research or not strongly explained or reported. An overview of these concerns and the species it accounts for is shown in table 3.

Suitability

A number of species had their present natural range (partly) based on 'suitable habitat', without explanation of what that assessment entails. The expansion based on suitable habitat is often quite extensive, which without supporting argumentation seems a little implausible. It especially forms a big contrast to the minimally expanded ranges of species solely based on information from the IUCN.

Current range not included in present natural range

Overall, the idea is that the current range of a species is a part of the range where it would have lived without the presence of human pressure. An exception can be area where a species is introduced, as is the case for the *Bos javanicus.* There are three species of which a large part of the current range is not included in the present natural range.

For the *Ceratotherium simum* this discrepancy has a technical reason, as the IUCN distribution range covers entire countries for this species instead of specific areas. This does however make it impossible to include the species in this assessment.

No explanation from Faurby et al. (2018) could be found for the *Equus zebra* and *Diceros bicornis*. The discrepancy does affect their results in this assessment, but was not severe enough to exclude them completely.

Extinct in the wild

Two species are recognised as extinct in the wild (*Oryx dammah* and *Elaphurus davidianus*). The first has no the current range and is not suitable to be included in the initial refugee probability assessment. The IUCN does not recognise a current range for the *Elaphurus davidianus* either, but there still is a current range in the PHYLACINE dataset.

One other species has no current range (*Bos sauveli) while it is recognised as critically endangered. The extinction of species is believed to be almost certainly inevitable (Timmins et al., 2016), causing the habitat to be marked as possibly extinct, which is not included in the construction of the current range.*

Missing source

A large part of the species had their present natural based on literature, of which most the source is sited. Three species mis the information on which source the decisions were based.

Missing current range

As the current range excludes areas labelled by the IUCN as possibly extinct, there is no current range for the *Bos sauveli*, making it impossible to include the species in this assessment.

Table 1 The six relevant ways through which species' present natural ranges were created from current ranges, in order of decreasing certainty.

PHYLACINE category number	Modification	Number of researched species per category
0	Expansion of current ranges by including areas coded as extinct or possibly extinct by the IUCN	33
2	Expansions of current ranges for species with known historical distributions based on literature	24
3	Merging of disjunct current ranges, if the disjunction was likely anthropogenic, by filling suitable habitats in the gap	3
4	Expansion of current ranges for island endemic species to cover the whole island (<150,000 km2)	1
5	Expansion of current ranges for species with known or highly suspected range declines to suitable areas continuous with current ranges.	17
8	Estimation of present natural ranges for species with unknown historic distribution based on species-specific ecological knowledge and late-Quaternary geography.	1

Table 2 Overview of the concerns about the by Faurby et al. (2018) *constructed ranges and the species to whom it applies.*

Concern in constructed ranges	**Species**
Current range is not fully included in the present natural range	*Diceros bicornis* *Equus zebra*
Present natural range is fully based on unknown suitability assessment	*Alcelaphus buselaphus* *Bos javanicus* *Damaliscus lunatus* *Equus hemionus* *Hippotragus equinus* *Ovis ammon*

	Tragelaphus derbianus
Present natural range is partly based on unknown suitability assessment	*Ailuropoda melanoleuca* *Alces alces* *Bos gaurus* *Equus ferus* *Giraffa camelopardalis* *Hippopotamus amphibius* *Ovibos moschatus* *Ursus arctos* *Syncerus caffer*
Source not cited	*Boselaphus tragocamelus* *Rucervus duvaucelii* *Sus scrofa*
Incorrect current range	*Ceratotherium simum*

Appendix B – Overview Results All Species

Name	*Area present natural range [million km^2]*	*Area current range [million km^2]*	*Relative range loss (lost/pres.nat)*	*Ecoregions DI present natural range [-]*	*Ecoregions DI current range [-]*	*Relative ecoregions DI loss*	*Climatic overlap loss fences [-]*	*IUCN threat status*	*Population density lost range [inh/km^2]*	*Pop. density current range [inh/km^2]*	*Population density difference (lost-cur)*	*Difference (lost-cur) cropland [km^2/km^2]*	*Difference (lost-cur) grazeland [km^2/km^2]*	*Difference (lost-cur) built area [km^2/km^2]*	*Difference total anthr. land use [km^2/km^2]*
Ailuropoda_melanoleuca	1953614	242122	0,124	2,406	1,958	0,186	0,432	VU	302,7	182,0	120,8	0,081	-0,007	0,003	0,077
Alcelaphus_buselaphus	9829412	8656623	0,881	3,383	3,194	0,056	0,878	LC	151,7	46,0	105,7	0,124	0,057	0,007	0,188
Alces_alces	29134149	25481200	0,875	4,048	3,853	0,048	0,798	LC	84,8	10,1	74,6	0,214	0,094	0,015	0,323
Babyrousa_babyrussa	14414	14072	0,976	0,860	0,854	0,007	1	VU	11,3	26,9	-15,6	-0,088	-0,015	-0,001	-0,104
Bison_bison	10185780	763614	0,075	3,889	2,929	0,247	0,326	NT	20,6	2,7	18,0	0,141	0,156	0,008	0,305
Bison_bonasus	14434494	27937	0,002	3,483	0,682	0,804	0,038	VU	43,3	53,1	-9,8	-0,083	0,081	-0,002	-0,004
Bos_gaurus	4479160	2731316	0,610	3,641	3,489	0,042	0,889	VU	436,1	177,5	258,6	0,262	0,002	0,006	0,270
Bos_javanicus	3394200	706295	0,208	3,342	2,805	0,161	0,712	EN	169,9	84,9	84,9	0,074	0,003	0,005	0,082
Bos_mutus	6093847	884675	0,145	3,577	1,502	0,580	0,240	VU	21,0	0,4	20,6	0,042	-0,326	0,001	-0,283
Boselaphus_tragocamelus	2810712	2580401	0,918	2,845	2,739	0,037	0,752	LC	1000,6	426,4	574,2	-0,077	0,014	0,008	-0,056
Bubalus_arnee	3846854	220425	0,057	3,485	2,589	0,257	0,618	EN	285,2	251,0	34,2	0,100	-0,009	0,001	0,092
Bubalus_mindorensis	12641	11988	0,948	0,654	0,582	0,111	0,656	CR	244,0	242,1	1,9	0,021	0,020	-0,002	0,039
Camelus_ferus	5453759	465618	0,085	3,162	1,291	0,592	0,118	CR	13,8	2,0	11,8	0,076	0,364	0,001	0,441
Cervus_canadensis	22464798	13450140	0,599	4,594	4,318	0,060	0,662	LC	32,0	39,6	-7,6	0,116	-0,011	0,007	0,111
Cervus_elaphus	9901946	4425855	0,447	3,553	2,983	0,160	0,676	LC	75,8	117,6	-41,8	-0,007	0,185	-0,008	0,171
Choeropsis_liberiensis	337373	287680	0,853	1,495	1,038	0,306	0,881	EN	439,3	94,3	344,9	-0,013	-0,063	0,011	-0,065
Damaliscus_lunatus	11814852	3168210	0,268	3,260	3,008	0,078	0,734	LC	55,9	35,7	20,2	0,068	-0,005	0,000	0,063
Dicerorhinus_sumatrensis	5066672	77810	0,015	3,313	1,648	0,502	0,142	CR	264,2	69,5	194,7	0,033	0,071	0,008	0,113
Diceros_bicornis	11963319	7616056	0,637	3,443	3,285	0,046	0,738	CR	58,3	38,4	19,9	0,103	-0,086	0,000	0,016
Elaphurus_davidianus	2065301	34445	0,017	1,815	1,091	0,399	0,064	EW	482,4	590,4	-108,1	-0,262	-0,027	-0,002	-0,291
Elephas_maximus	10418831	2552728	0,245	4,165	3,725	0,106	0,544	EN	318,6	203,5	115,1	0,055	0,089	0,004	0,148
Equus_africanus	6955537	130990	0,019	2,540	0,393	0,845	0,243	CR	45,2	40,7	4,5	0,048	-0,001	0,002	0,049

Equus_ferus	77575331	167623	0,002	5,443	1,936	0,644	0,040	EN	54,7	11,5	43,2	0,131	-0,343	0,005	-0,208
Equus_grevyi	1098343	206891	0,188	1,636	1,529	0,065	0,876	EN	48,2	36,6	11,6	0,000	0,013	0,000	0,013
Equus_hemionus	11236031	1035181	0,092	3,891	2,483	0,362	0,347	NT	50,3	20,4	29,8	0,063	0,036	0,003	0,102
Equus_quagga	7229634	3583258	0,496	3,219	3,042	0,055	0,785	NT	43,5	48,0	-4,4	-0,022	0,084	-0,001	0,061
Equus_zebra	743023	628231	0,846	2,545	2,506	0,015	0,940	VU	3,4	18,2	-14,9	-0,024	0,174	-0,002	0,148
Giraffa_camelopardalis	8849370	3852428	0,435	3,028	3,176	- 0,049	0,725	VU	50,3	33,7	16,6	0,096	0,013	0,000	0,109
Gorilla_gorilla	961633	933695	0,971	1,679	1,627	0,031	0,891	CR	54,5	21,0	33,6	0,062	-0,104	0,000	-0,043
Hippopotamus_amphibius	22406051	5345752	0,239	3,999	3,314	0,171	0,507	VU	81,8	58,7	23,1	0,020	-0,038	0,003	-0,016
Hippotragus_equinus	9948062	7250486	0,729	2,982	2,848	0,045	0,862	LC	72,6	38,6	33,9	0,059	0,089	0,002	0,149
Kobus_leche	605304	586679	0,969	1,927	1,936	- 0,005	1	NT	20,3	9,1	11,3	0,033	0,184	0,001	0,218
Loxodonta_africana	24105720	8329921	0,346	3,712	3,501	0,057	0,667	VU	50,1	42,3	7,8	0,007	0,004	0,001	0,011
Oryx_beisa	2751252	1566753	0,569	2,484	1,933	0,222	0,848	EN	70,6	37,4	33,2	0,069	0,106	0,001	0,176
Ovibos_moschatus	7311404	2198768	0,301	3,140	2,058	0,345	0,700	LC	0,7	0,0	0,7	0,001	0,001	0,000	0,001
Ovis_ammon	5855194	5715509	0,976	3,394	3,397	- 0,001	1	NT	32,1	10,7	21,4	0,019	-0,331	0,003	-0,308
Rhinoceros_sondaicus	2903028	12508	0,004	3,419	1,099	0,678	0,112	CR	298,6	197,0	101,7	0,080	0,037	0,010	0,127
Rhinoceros_unicornis	1005736	121061	0,120	2,478	2,245	0,094	0,744	VU	680,0	386,0	294,0	0,283	-0,022	0,008	0,270
Rucervus_duvaucelii	2027849	176935	0,087	2,705	2,047	0,243	0,504	VU	527,6	498,2	29,4	0,176	-0,010	0,002	0,168
Sus_scrofa	35649121	31169370	0,874	4,940	4,884	0,011	0,978	LC	53,0	142,4	-89,4	-0,199	0,255	-0,004	0,051
Syncerus_caffer	13649647	9944245	0,729	3,569	3,500	0,019	0,995	NT	60,1	49,7	10,3	-0,014	0,081	0,001	0,067
Tapirus_bairdii	1062515	1025008	0,965	3,225	3,224	0,000	0,886	EN	367,1	68,8	298,2	0,145	0,011	0,033	0,189
Tapirus_indicus	1149485	528246	0,460	2,618	2,186	0,165	0,902	EN	225,8	87,0	138,9	0,126	0,009	0,009	0,144
Tapirus_pinchaque	593937	307308	0,517	2,754	2,404	0,127	0,951	EN	46,0	84,6	-38,6	-0,016	0,114	-0,003	0,095
Tapirus_terrestris	13709307	12374935	0,903	3,725	3,644	0,022	0,884	VU	68,0	20,8	47,2	0,034	0,148	0,005	0,186
Tragelaphus_derbianus	3195572	661178	0,207	1,980	1,299	0,344	0,698	VU	77,1	11,9	65,1	0,175	0,065	0,004	0,244
Tremarctos_ornatus	1711270	1412755	0,826	3,365	3,299	0,020	0,908	VU	58,5	49,4	9,1	-0,012	0,096	0,000	0,084
Ursus_americanus	17363242	13589398	0,783	4,221	4,114	0,025	0,954	LC	28,5	16,8	11,7	0,210	0,263	0,005	0,477
Ursus_arctos	57939493	32272731	0,557	4,924	4,282	0,130	0,651	LC	47,6	15,3	32,3	0,088	0,137	0,004	0,230

Appendix C – Visualisation Methods

Range decline and habitat diversity

The polygonised ranges from the PHYLACINE dataset were overlapped with the *Ecoregions2017* map of Dinerstein et al. (2017) of the world's ecoregions to extract the surface area of the present ecoregions in the ranges of all species (figure 12).

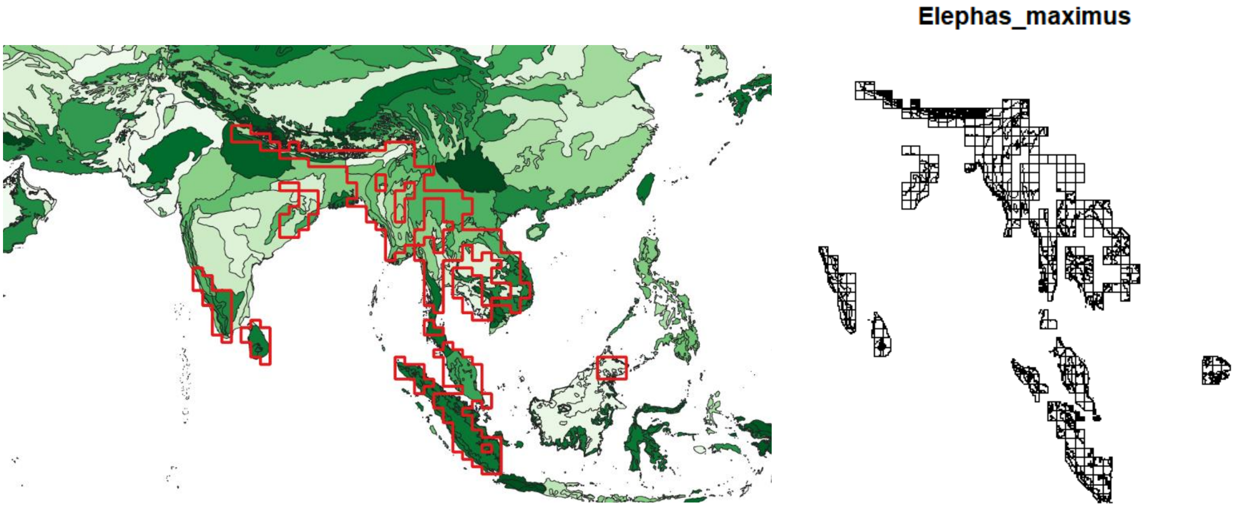


Figure 12 Visualisation of the calculated overlap of the ecoregions with the current range of the Elephas maximus. On the left the ecoregions map is shown with an overlapping outline of the current range of the Elephas maximus. On the right the cut-out of the overlapped range is shown, which is used to calculate the cover of the ecoregions in the range cells.

Environmental diversity

The polygonised ranges from the PHYLACINE dataset were overlapped with the *WorldClim* dataset of Fick & Hijmans (2017) to extract the climatic conditions in the ranges of all species (figure 13). The climatic conditions were used to construct bagplots (figure 14), which were overlapped to calculate the climatic diversity loss (figure 15). The overlap of the fences was considered to be the best representation of the overlap loss.

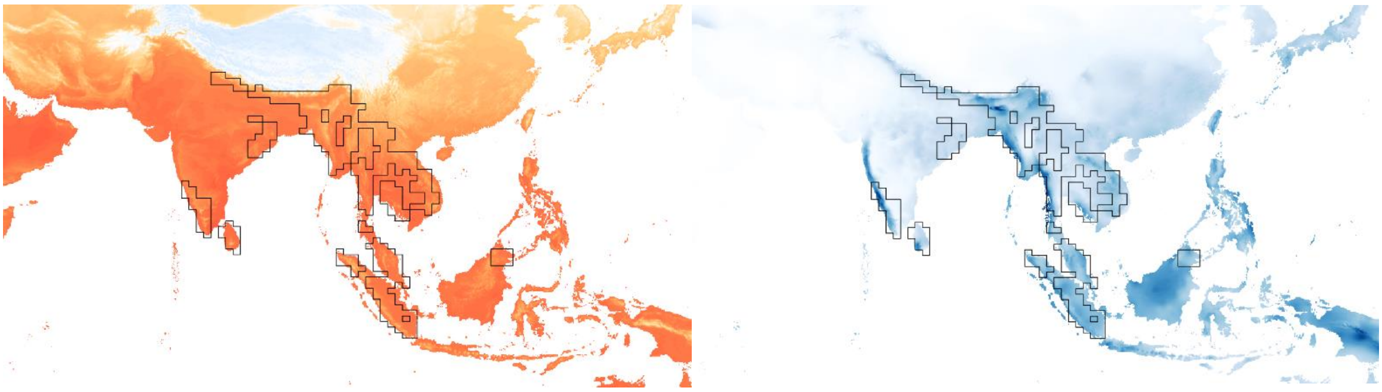


Figure 13 Visualisation of the overlap of the polygonised current range of the Elephas maximus and the bioclimatic conditions. The overlap with the bioclimatic variable BIO1 = Annual Mean Temperature data is shown on the left and the overlap with the bioclimatic variable BIO12 = Annual Precipitation data is shown on the right.

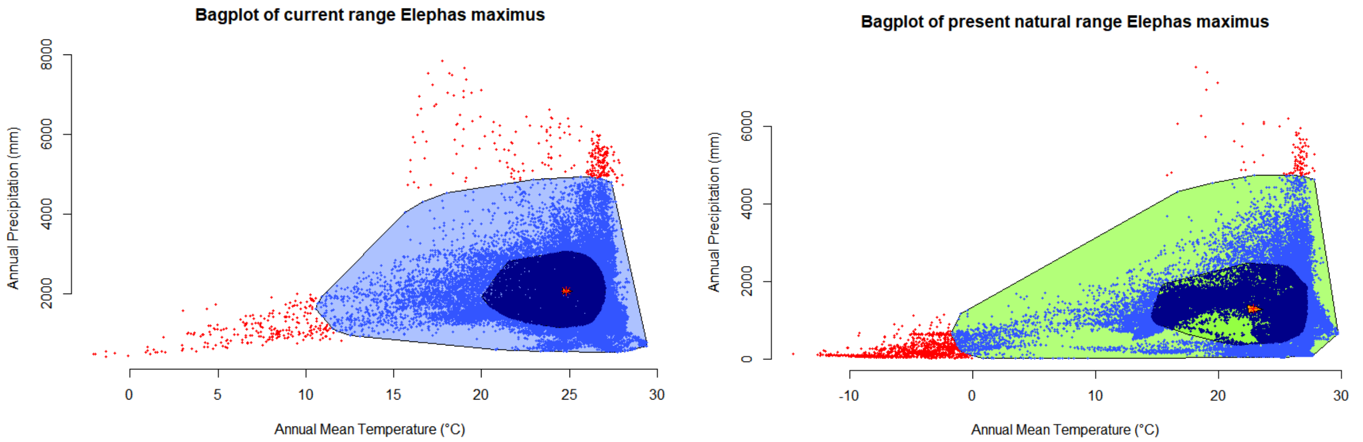


Figure 14 Bagplot of the bioclimatic conditions in the current (left) and present natural (right) range of the Elephas maximus*. The outer polygon depicts the fence, the inner polygon the bag and the smallest area the center. All points are included in the figure, with the outliers coloured red.*

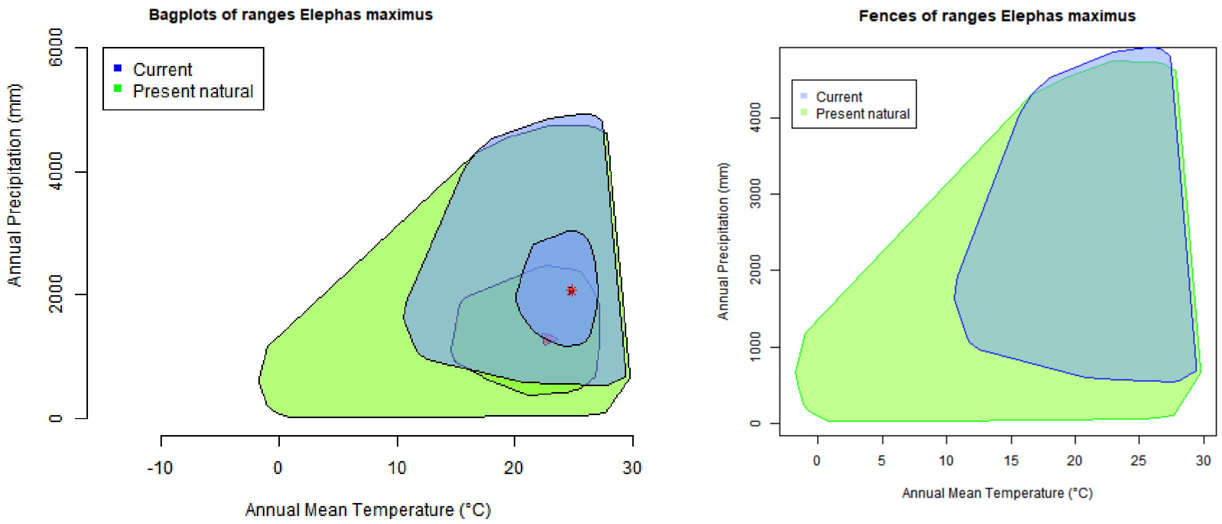


Figure 15 Visualisation of the overlap of the bagplots of the Elephas maximus*. The overlap of the bagplots of the current and present natural range (left) and the overlap of the fences (right).*

Human activity

The polygonised ranges from the PHYLACINE dataset were overlapped with the History Database of the Global Environment (HYDE) 3.2 database of Goldewijk et al., (2017) to extract the anthropogenic land use and population density in the ranges of all species (figure 16).

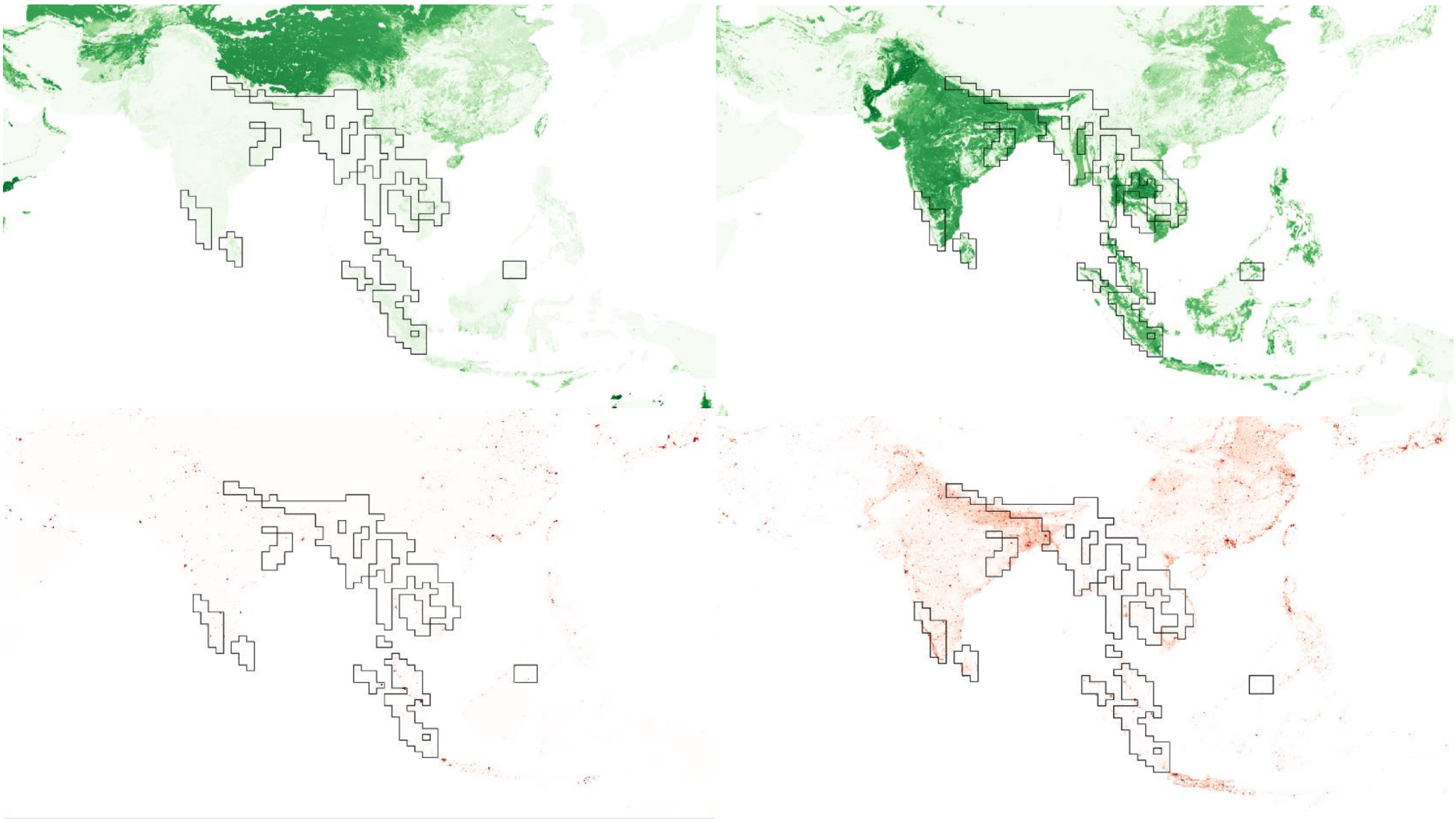


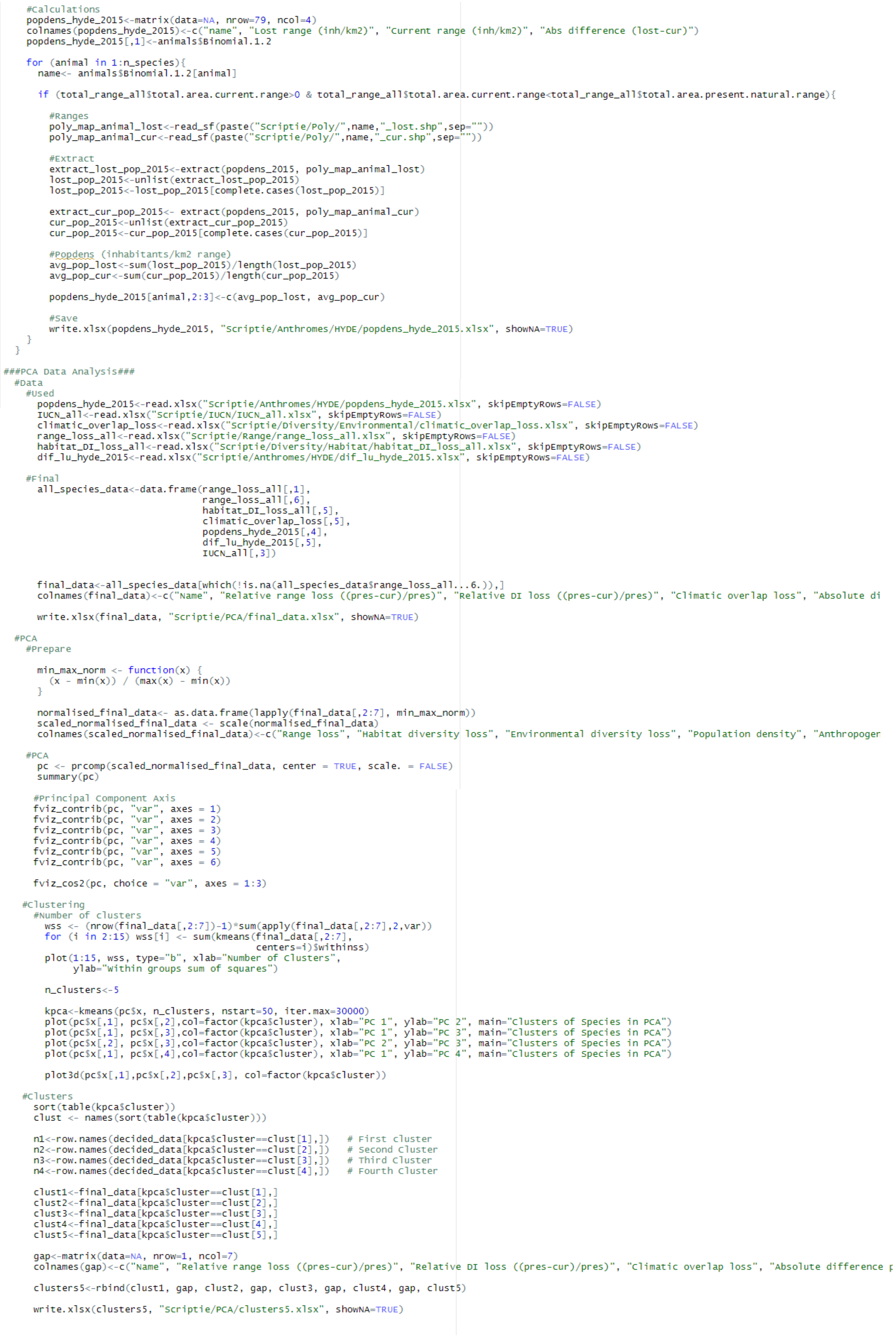
Figure 16 Visualisation of the overlap of the polygonised current range of the Elephas maximus and the HYDE 2015 data on land used for grazing (top left), for cropland (top right), for build area (bottom left) and the population density (bottom right).

Appendix D – Script

```r
library(maptools)
library(ape)
library(raster)
library(openxlsx)
library(ape)
library(plotrix)
library(sf)
library(rgdal)
library(spatialEco)
library(geosphere)
library(PBSmapping)
library(spex)
library(rgeos)
library(prob)
library(aplpack)
library(rlist)
library(mrfDepth)
library(data.table)
library(FactoMineR)
library(factoextra)
library(stats)
library(rgl)


###Species Selection###
  mam <- read.csv("Scriptie/Data/PHYLACINE/Trait_data.csv", fileEncoding = "UTF-8", stringsAsFactors = F)
  mam$IUCN.Status.1.2 <- factor(mam$IUCN.Status.1.2, levels=c("EP", "EX", "EW", "CR", "EN", "VU", "NT", "LC", "DD"))

  animals<- mam
  animals<- animals[animals$Diet.Plant >= 80, ]
  animals<- animals[animals$IUCN.Status.1.2 != "EX", ]
  animals<- animals[animals$IUCN.Status.1.2 != "EP", ]
  animals$Mass.g<-round(animals$Mass.g, digits=-3)
  animals<- animals[animals$Mass.g >= 100000, ]

  n_species<-nrow(animals)
 #Ranges species
  maps.current <- paste0("Scriptie/Data/PHYLACINE/Current/Current/", animals$Binomial.1.2, ".tif")
  maps.pres_nat <- paste0("Scriptie/Data/PHYLACINE/Present_natural/Present_natural/", animals$Binomial.1.2, ".tif")

  # Ranges per species
  cells_in_range<-matrix(data=NA, nrow=n_species, ncol=4)
  colnames(cells_in_range)<-c("Name", "Number of cells in current range", "Number of cells in present natural range", "Number of cells in lost range")

  for (animal in 1:n_species){
    name<- animals$Binomial.1.2[animal]

    map_animal_cur <- raster(maps.current[animal])
    map_animal_pres<- raster(maps.pres_nat[animal])
    map_animal_lost<-map_animal_pres-map_animal_cur
    cells_animal_cur<-rasterToPoints(map_animal_cur)
    cells_animal_pres<-rasterToPoints(map_animal_pres)
    cells_animal_lost<-rasterToPoints(map_animal_lost)
    cells_in_range[animal,1]<-name
    cells_in_range[animal,2]<-sum(cells_animal_cur[,3])
    cells_in_range[animal,3]<-sum(cells_animal_pres[,3])
    cells_in_range[animal,4]<-sum(cells_animal_lost[,3])

    if (cells_in_range[animal,2]>0 & cells_in_range[animal,2]<cells_in_range[animal,3]){

      #polygons all ranges
      map_animal_cur_lim<-map_animal_cur
      map_animal_cur_lim [map_animal_cur_lim  < 1] <- NA
      map_animal_mask_cur <- mask(map_animal_cur, map_animal_cur_lim)
      poly_map_animal_cur<- rasterToPolygons(map_animal_mask_cur)

      map_animal_pres_lim<-map_animal_pres
      map_animal_pres_lim [map_animal_pres_lim  < 1] <- NA
      map_animal_mask_pres <- mask(map_animal_pres, map_animal_pres_lim)
      poly_map_animal_pres<- rasterToPolygons(map_animal_mask_pres)

      map_animal_lost_lim<-map_animal_lost
      map_animal_lost_lim [map_animal_lost_lim  < 1] <- NA
      map_animal_mask_lost <- mask(map_animal_lost, map_animal_lost_lim)
      poly_map_animal_lost<- rasterToPolygons(map_animal_mask_lost)

      shapefile(poly_map_animal_cur, filename=paste("Scriptie/Poly/",name,"_cur.shp",sep=""))
      shapefile(poly_map_animal_pres, filename=paste("Scriptie/Poly/",name,"_pres.shp",sep=""))
      shapefile(poly_map_animal_lost, filename=paste("Scriptie/Poly/",name,"_lost.shp",sep=""))
    }
  }
  write.xlsx(cells_in_range, "Scriptie/n_cells_in_range_all.xlsx", showNA=TRUE)

###Habitat diversity (ecoregions)###
  ecoreg<-read_sf('Scriptie/Data/Ecoregions2017/Ecoregions2017.shp')
  crs_match<-(crs(read_sf(paste("Scriptie/Poly/",name,"_cur.shp",sep=""))))
  proj_eco<- st_transform(ecoreg, crs_match)#"+proj=cea +lat_ts=30 +lon_0=0 +x_0=0 +y_0=0 +datum=WGS84 +units=m +no_defs ")#crs(poly_map_animal_cur))
  proj_eco<-st_buffer(proj_eco$geometry, dist=0)


  # Prep
  habitat_diversity_all<-matrix(data=NA, nrow=nrow(animals), ncol=5)
  colnames(habitat_diversity_all)<- c("Name", "number of ecoregions cur", "DI ecoregions cur",
                                      "number of ecoregions pres", "DI ecoregions pres")
  total_range_all<-matrix(data=NA, nrow=nrow(animals), ncol=4)
  colnames(total_range_all)<-c("Name" ,"total area current range", "total area present natural range", "total area lost range")

  # Calculation loop
  for (animal in 1:n_species){
    name<-animals$Binomial.1.2[animal]

     if (as.numeric(cells_in_range[animal,2])>0 & as.numeric(cells_in_range[animal,2])<as.numeric(cells_in_range[animal,3])){

      #Current
      poly_map_animal_cur<-read_sf(paste("Scriptie/Poly/",name,"_cur.shp",sep=""))  #Map
      overlap_cur<-st_intersection(proj_eco, poly_map_animal_cur)                  #Overlap
      matching_cur<- st_intersects(poly_map_animal_cur, proj_eco)                  #Overlap

      overlap_area_cur<-st_area(overlap_cur)/1000000                               #Area/part
      total_area_cur<-sum(overlap_area_cur)                                        #Area total

      #Ecoregions

      all_areas_cur<-data.frame(unlist(matching_cur), unlist(overlap_area_cur))

      ecoregions_in_range_cur<- unique(all_areas_cur[,1])#[(which(ecoreg$BIOME_NAME[unlist(matching_cur)]!="N/A")),1])
      n_ecoregions_in_range_cur<-length(ecoregions_in_range_cur)

      area_ecoregions_cur<-matrix(data = NA, nrow = n_ecoregions_in_range_cur, ncol = 1)

      for (i in 1:length(ecoregions_in_range_cur)){
        num<- which(all_areas_cur$unlist.matching_cur.==ecoregions_in_range_cur[i])

        area_eco_cur<-0
        for (n in 1:length(num)){
          area_eco_cur<-area_eco_cur+as.numeric(all_areas_cur[num[n],2])
        }
        area_ecoregions_cur[i]<-area_eco_cur
      }

      ecoregions_relative_area_cur<-area_ecoregions_cur/total_area_cur
```

```r
      #Diversity
      Diversity_Index_eco_cur<- -(sum(ecoregions_relative_area_cur*(log(ecoregions_relative_area_cur, base = exp(1)))))

      #Table
      habitat_diversity_species_cur<-data.frame(name,
                                                n_ecoregions_in_range_cur, Diversity_Index_eco_cur,
                                                total_area_cur)

      #Present natural
      poly_map_animal_pres<-read_sf(paste("Scriptie/Poly/",name,"_pres.shp",sep=""))  #Map
      overlap_pres<-st_intersection(proj_eco, poly_map_animal_pres)                   #Overlap
      matching_pres<- st_intersects(poly_map_animal_pres, proj_eco)                   #Overlap

      overlap_area_pres<-st_area(overlap_pres)/1000000                                #Area/part
      total_area_pres<-sum(overlap_area_pres)                                         #Area total

      #Ecoregions

      all_areas_pres<-data.frame(unlist(matching_pres), unlist(overlap_area_pres))

      ecoregions_in_range_pres<- unique(all_areas_pres[,1])#[(which(ecoreg$BIOME_NAME[unlist(matching_pres)]!="N/A")),1])
      n_ecoregions_in_range_pres<-length(ecoregions_in_range_pres)

      area_ecoregions_pres<-matrix(data = NA, nrow = n_ecoregions_in_range_pres, ncol = 1)

      for (i in 1:length(ecoregions_in_range_pres)){
        num<- which(all_areas_pres$unlist.matching_pres.==ecoregions_in_range_pres[i])

        area_eco_pres<-0
        for (n in 1:length(num)){
          area_eco_pres<-area_eco_pres+as.numeric(all_areas_pres[num[n],2])
        }
        area_ecoregions_pres[i]<-area_eco_pres
      }

      ecoregions_relative_area_pres<-area_ecoregions_pres/total_area_pres


      #Diversity
      Diversity_Index_eco_pres<- -(sum(ecoregions_relative_area_pres*(log(ecoregions_relative_area_pres, base = exp(1)))))

      #Table
      habitat_diversity_species_pres<-data.frame(name,
                                                 n_ecoregions_in_range_pres, Diversity_Index_eco_pres,
                                                 total_area_pres)

      #Combine
      #Add lost range
      poly_map_animal_lost<-read_sf(paste("Scriptie/Poly/",name,"_lost.shp",sep=""))  #Map
      overlap_lost<-st_intersection(proj_eco, poly_map_animal_lost)
      overlap_area_lost<-st_area(overlap_lost)/1000000                                #Area/part
      total_area_lost<-sum(overlap_area_lost)                                         #Area total

      #Overview
      total_range_all[animal,]<-c(name, total_area_cur, total_area_pres, total_area_lost)
      habitat_diversity_all[animal,]<-c(name,
                                        n_ecoregions_in_range_cur, Diversity_Index_eco_cur,
                                        n_ecoregions_in_range_pres, Diversity_Index_eco_pres)
    }
  }

  habitat_DI_loss_all<-data.frame(animals$Binomial.1.2,
                                  as.numeric(habitat_diversity_all[,3]), as.numeric(habitat_diversity_all[,5]),
                                  (as.numeric(habitat_diversity_all[,5])-as.numeric(habitat_diversity_all[,3])),
                                  ((as.numeric(habitat_diversity_all[,5])-as.numeric(habitat_diversity_all[,3]))/as.numeric(habitat_diversity_all[,5])))
  colnames(habitat_DI_loss_all)<-c("Name", "DI eco cur", "DI eco pres", "Abs DI eco","Rel DI eco")

  write.xlsx(habitat_diversity_all, "Scriptie/Diversity/Habitat/habitat_diversity_all.xlsx", showNA=TRUE)
  write.xlsx(total_range_all, "Scriptie/total_range_all.xlsx", showNA=TRUE)
  write.xlsx(habitat_DI_loss_all,"Scriptie/Diversity/Habitat/habitat_DI_loss_all.xlsx", showNA=TRUE)

###Range loss###
  total_range_all<-read.xlsx("Scriptie/total_range_all.xlsx", skipEmptyRows=FALSE)

  abs_range_loss<-as.numeric(total_range_all[,3])-as.numeric(total_range_all[,2])
  rel_range_loss_pres<- abs_range_loss/as.numeric(total_range_all[,3])

  range_loss_all<-data.frame(animals$Binomial.1.2, total_range_all[,2:4], abs_range_loss, rel_range_loss_pres)
  colnames(range_loss_all)<-c("Name", "Current range area", "Present natural range area", "Lost range area",
                              "Absolute range loss (pres-cur)", "Relative range loss (lost/pres)")

  write.xlsx(range_loss_all, "Scriptie/Range/range_loss_all.xlsx", showNA=TRUE)

###Environmental Diversity###
  #Data
  biome1<-raster("Scriptie/Data/wc2.1_5m_bio/wc2.1_5m_bio_1.tif")
  biome12<-raster("Scriptie/Data/wc2.1_5m_bio/wc2.1_5m_bio_12.tif")

  proj_bio_1<-projectRaster(biome1, crs= "+proj=cea +lon_0=0 +lat_ts=30 +x_0=0 +y_0=0 +datum=WGS84 +ellps=WGS84 +units=m +no_defs")
  proj_bio_12<-projectRaster(biome12, crs= "+proj=cea +lon_0=0 +lat_ts=30 +x_0=0 +y_0=0 +datum=WGS84 +ellps=WGS84 +units=m +no_defs")

  total_range_all<-read.xlsx("Scriptie/total_range_all.xlsx", skipEmptyRows=FALSE)

  #Bagplot Calculations
  area_bagplot_all<-matrix(data=NA, nrow=n_species, ncol=11)
  colnames(area_bagplot_all)<-c("Name", "Bag Cur", "Bag Pres", "Bag Overlap", "Bag Rel Cur", "Bag Rel Pres",
                                "Fence Cur", "Fence Pres", "Fence Overlap", "Fence Rel Cur", "Fence Rel Pres")

  for (animal in 5:n_species){
    name<-animals$Binomial.1.2[animal] #total_range_all$Name[animal]

    if (!is.na(total_range_all$total.area.current.range[animal])){ #& as.numeric(total_range_all$total.area.current.range[animal])<as.numeric(total_range_al

      #Climatic data
      poly_map_animal_cur<-read_sf(paste("Scriptie/Poly/",name,"_cur.shp",sep=""))
      poly_map_animal_pres<-read_sf(paste("Scriptie/Poly/",name,"_pres.shp",sep=""))

      selection_1_cur<- extract(proj_bio_1, poly_map_animal_cur)
      selection_1_pres<- extract(proj_bio_1, poly_map_animal_pres)
      selection_12_cur<- extract(proj_bio_12, poly_map_animal_cur)
      selection_12_pres<- extract(proj_bio_12, poly_map_animal_pres)
      all_points_1_cur<-unlist(selection_1_cur)
      all_points_1_pres<-unlist(selection_1_pres)
      all_points_12_cur<-unlist(selection_12_cur)
      all_points_12_pres<-unlist(selection_12_pres)

      n_point_cur<-length(all_points_1_cur)
      n_point_pres<-length(all_points_1_pres)

      points_cur<-data.frame(all_points_1_cur, all_points_12_cur)
      points_cur<-points_cur[which(!is.na(points_cur$all_points_1_cur) & points_cur$all_points_1_cur!=""), ]

      points_pres<-data.frame(all_points_1_pres, all_points_12_pres)
      points_pres<-points_pres[which(!is.na(points_pres$all_points_1_pres) & points_pres$all_points_1_pres!=""), ]

      bag_pres<-compute.bagplot(points_pres)
      bag_cur<-compute.bagplot(points_cur)

      #Values bagplot and comparison

      #polygons
      poly_bag_cur <- SpatialPolygons(list(Polygons(list(Polygon(rbind(bag_cur$hull.bag, bag_cur$hull.bag[1,]))), ID=1)))
      poly_bag_pres<- SpatialPolygons(list(Polygons(list(Polygon(rbind(bag_pres$hull.bag, bag_pres$hull.bag[1,]))), ID=1)))

      poly_fence_cur <- SpatialPolygons(list(Polygons(list(Polygon(rbind(bag_cur$hull.loop, bag_cur$hull.loop[1,]))), ID=1)))
      poly_fence_pres<- SpatialPolygons(list(Polygons(list(Polygon(rbind(bag_pres$hull.loop, bag_pres$hull.loop[1,]))), ID=1)))

      poly_bag_inters<-gIntersection(poly_bag_cur,poly_bag_pres)
      poly_fence_inters<-gIntersection(poly_fence_cur,poly_fence_pres)

      #Areas
      area_bag_cur<-gArea(poly_bag_cur)
      area_bag_pres<-gArea(poly_bag_pres)
      area_fence_cur<-gArea(poly_fence_cur)
      area_fence_pres<-gArea(poly_fence_pres)
      area_bag_overlap<-gArea(poly_bag_inters)
      area_fence_overlap<-gArea(poly_fence_inters)
```

```r
      rel_bag_cur<-area_bag_overlap/area_bag_cur
      rel_bag_pres<-area_bag_overlap/area_bag_pres
      rel_fence_cur<-area_fence_overlap/area_fence_cur
      rel_fence_pres<-area_fence_overlap/area_fence_pres

      area_bagplot_all[animal,]<-c(name, area_bag_cur, area_bag_pres, area_bag_overlap, rel_bag_cur, rel_bag_pres,  area_fence_cur,
                                   area_fence_pres, area_fence_overlap, rel_fence_cur, rel_fence_pres)

      write.xlsx(area_bagplot_all, "Scriptie/Diversity/Environmental/area_bagplot_all_uncut.xlsx", showNA=TRUE)
    }
  }

  climatic_overlap_loss<-data.frame(animals$Binomial.1.2, (1-area_bagplot_all[,11]))
  colnames(climatic_overlap_loss)<-c("Name", "Overlap loss (fence rel presnat)")

  write.xlsx(climatic_overlap_loss, "Scriptie/Diversity/Environmental/climatic_overlap_loss.xlsx", showNA=TRUE)

###IUCN threat status###

  older_IUCN<- animals$IUCN.Status.1.2
  IUCN_status<- read.xlsx("Scriptie/IUCN status.xlsx", skipEmptyRows=FALSE)    #Updated IUCN

  IUCN_number<-matrix(data=NA, ncol=1, nrow=n_species)
  for (IUCN in 1:n_species){
    if (IUCN_status[IUCN,4]=="LC"){
      IUCN_number[IUCN]<-1
    }else if (IUCN_status[IUCN,4]=="NT"){
      IUCN_number[IUCN]<-2
    }else if (IUCN_status[IUCN,4]=="VU"){
      IUCN_number[IUCN]<-3
    }else if (IUCN_status[IUCN,4]=="EN"){
      IUCN_number[IUCN]<-4
    }else if (IUCN_status[IUCN,4]=="CR"){
      IUCN_number[IUCN]<-5
    }else if (IUCN_status[IUCN,4]=="EW"){
      IUCN_number[IUCN]<-6
    }
  }

  IUCN_all<-data.frame(IUCN_status$name,IUCN_status$status,IUCN_number)
  write.xlsx(IUCN_all, "Scriptie/IUCN/IUCN_all.xlsx", showNA=TRUE)

###Anthromes###
  #HYDE
    #Land Use
      #Data
        total_range_all<-read.xlsx("Scriptie/total_range_all.xlsx", skipEmptyRows=FALSE)

        crop_2015<-raster('Scriptie/Anthromes/HYDE/2015/cropland2015AD.asc')
        crs(crop_2015)<-"+proj=longlat +datum=WGS84 +no_defs"

        graze_2015<-raster('Scriptie/Anthromes/HYDE/2015/grazing2015AD.asc')
        crs(graze_2015)<-"+proj=longlat +datum=WGS84 +no_defs"

        urban_2015<-raster('Scriptie/Anthromes/HYDE/2015/uopp_2015AD.asc')
        crs(urban_2015)<-"+proj=longlat +datum=WGS84 +no_defs"

      #Calculations
        avg_lu_hyde_2015<-matrix(data=NA, nrow=79, ncol=9)
        colnames(avg_lu_hyde_2015)<-c("name", "crop lost (km2/km2)", "graze lost (km2/km2)", "urban lost (km2/km2)", "total lost (km2/km2)",
                                      "crop cur (km2/km2)", "graze cur (km2/km2)", "urban cur (km2/km2)", "total cur (km2/km2)")
        avg_lu_hyde_2015[,1]<-animals$Binomial.1.2

        for (animal in 1:n_species){
          name<- animals$Binomial.1.2[animal]

          if (total_range_all$total.area.current.range>0 & total_range_all$total.area.current.range<total_range_all$total.area.present.natural.range){

          #Ranges
            poly_map_animal_cur<-read_sf(paste("Scriptie/Poly/",name,"_cur.shp",sep=""))
            poly_map_animal_lost<-read_sf(paste("Scriptie/Poly/",name,"_lost.shp",sep=""))

          #Lost Range
            #Values (extract)
              extract_lost_crop_2015<-extract(crop_2015, poly_map_animal_lost)
              lost_crop_2015<-unlist(extract_lost_crop_2015)
              lost_crop_2015<-lost_crop_2015[complete.cases(lost_crop_2015)]

              extract_lost_graze_2015<- extract(graze_2015, poly_map_animal_lost)
              lost_graze_2015<-unlist(extract_lost_graze_2015)
              lost_graze_2015<-lost_graze_2015[complete.cases(lost_graze_2015)]

              extract_lost_urban_2015<- extract(urban_2015, poly_map_animal_lost)
              lost_urban_2015<-unlist(extract_lost_urban_2015)
              lost_urban_2015<-lost_urban_2015[complete.cases(lost_urban_2015)]

              lost_lu_2015<-data.frame(lost_crop_2015, lost_graze_2015, lost_urban_2015)
              colnames(lost_lu_2015)<-c("Cropland", "Graze", "Urban")

            #Land use (km2 used/km2 range)
              crop_land_lost<-sum(lost_lu_2015[,1])/as.numeric(total_range_all$total.area.lost.range[animal])    #avg km2 land/km2 range
              graze_land_lost<-sum(lost_lu_2015[,2])/as.numeric(total_range_all$total.area.lost.range[animal])
              urban_land_lost<-sum(lost_lu_2015[,3])/as.numeric(total_range_all$total.area.lost.range[animal])

              avg_lu_hyde_2015[animal,2:5]<-c(crop_land_lost, graze_land_lost, urban_land_lost, sum(crop_land_lost, graze_land_lost, urban_land_lost))

          #Current Range
            #Values (extract)
              extract_cur_crop_2015<-extract(crop_2015, poly_map_animal_cur)
              cur_crop_2015<-unlist(extract_cur_crop_2015)
              cur_crop_2015<-cur_crop_2015[complete.cases(cur_crop_2015)]

              extract_cur_graze_2015<- extract(graze_2015, poly_map_animal_cur)
              cur_graze_2015<-unlist(extract_cur_graze_2015)
              cur_graze_2015<-cur_graze_2015[complete.cases(cur_graze_2015)]

              extract_cur_urban_2015<- extract(urban_2015, poly_map_animal_cur)
              cur_urban_2015<-unlist(extract_cur_urban_2015)
              cur_urban_2015<-cur_urban_2015[complete.cases(cur_urban_2015)]

              cur_lu_2015<-data.frame(cur_crop_2015, cur_graze_2015, cur_urban_2015)
              colnames(cur_lu_2015)<-c("Cropland", "Graze", "Urban")

              #Land use (km2 used/km2 range)
              crop_land_cur<-sum(cur_lu_2015[,1])/as.numeric(total_range_all$total.area.current.range[animal])   #avg km2 land/km2 range
              graze_land_cur<-sum(cur_lu_2015[,2])/as.numeric(total_range_all$total.area.current.range[animal])
              urban_land_cur<-sum(cur_lu_2015[,3])/as.numeric(total_range_all$total.area.current.range[animal])

              avg_lu_hyde_2015[animal,6:9]<-c(crop_land_cur, graze_land_cur, urban_land_cur, sum(crop_land_cur, graze_land_cur, urban_land_cur))

          #Save
            write.xlsx(avg_lu_hyde_2015, "Scriptie/Anthromes/HYDE/average_lu_hyde_2015.xlsx", showNA=TRUE)
        }
      }

      #Difference land use
      dif_crop_2015<-as.numeric(avg_lu_hyde_2015[,2])-as.numeric(avg_lu_hyde_2015[,6])
      dif_graze_2015<-as.numeric(avg_lu_hyde_2015[,3])-as.numeric(avg_lu_hyde_2015[,7])
      dif_urban_2015<-as.numeric(avg_lu_hyde_2015[,4])-as.numeric(avg_lu_hyde_2015[,8])
      dif_lu_2015<-as.numeric(avg_lu_hyde_2015[,5])-as.numeric(avg_lu_hyde_2015[,9])

      dif_lu_hyde_2015<-data.frame(animals$Binomial.1.2,dif_crop_2015,dif_graze_2015,dif_urban_2015,dif_lu_2015)
      colnames(dif_lu_hyde_2015)<-c("Name", "Difference cropland", "Difference grazeland", "Difference build area", "Difference total anthropogenic land use
      write.xlsx(dif_lu_hyde_2015, "Scriptie/Anthromes/HYDE/dif_lu_hyde_2015.xlsx", showNA=TRUE)

  #Population Density
    #Data
      total_range_all<-read.xlsx("Scriptie/total_range_all.xlsx", skipEmptyRows=FALSE)

      popdens_2015<-raster('Scriptie/Anthromes/HYDE/2015/popd_2015AD.asc')
      crs(popdens_2015)<-"+proj=longlat +datum=WGS84 +no_defs"
```



```
    #Calculations
    popdens_hyde_2015<-matrix(data=NA, nrow=79, ncol=4)
    colnames(popdens_hyde_2015)<-c("name", "Lost range (inh/km2)", "Current range (inh/km2)", "Abs difference (lost-cur)")
    popdens_hyde_2015[,1]<-animals$Binomial.1.2

    for (animal in 1:n_species){
      name<- animals$Binomial.1.2[animal]

      if (total_range_all$total.area.current.range>0 & total_range_all$total.area.current.range<total_range_all$total.area.present.natural.range){

        #Ranges
        poly_map_animal_lost<-read_sf(paste("Scriptie/Poly/",name,"_lost.shp",sep=""))
        poly_map_animal_cur<-read_sf(paste("Scriptie/Poly/",name,"_cur.shp",sep=""))

        #Extract
        extract_lost_pop_2015<-extract(popdens_2015, poly_map_animal_lost)
        lost_pop_2015<-unlist(extract_lost_pop_2015)
        lost_pop_2015<-lost_pop_2015[complete.cases(lost_pop_2015)]

        extract_cur_pop_2015<- extract(popdens_2015, poly_map_animal_cur)
        cur_pop_2015<-unlist(extract_cur_pop_2015)
        cur_pop_2015<-cur_pop_2015[complete.cases(cur_pop_2015)]

        #Popdens (inhabitants/km2 range)
        avg_pop_lost<-sum(lost_pop_2015)/length(lost_pop_2015)
        avg_pop_cur<-sum(cur_pop_2015)/length(cur_pop_2015)

        popdens_hyde_2015[animal,2:3]<-c(avg_pop_lost, avg_pop_cur)

        #Save
        write.xlsx(popdens_hyde_2015, "Scriptie/Anthromes/HYDE/popdens_hyde_2015.xlsx", showNA=TRUE)
    }
  }

###PCA Data Analysis###
  #Data
    #Used
      popdens_hyde_2015<-read.xlsx("Scriptie/Anthromes/HYDE/popdens_hyde_2015.xlsx", skipEmptyRows=FALSE)
      IUCN_all<-read.xlsx("Scriptie/IUCN/IUCN_all.xlsx", skipEmptyRows=FALSE)
      climatic_overlap_loss<-read.xlsx("Scriptie/Diversity/Environmental/climatic_overlap_loss.xlsx", skipEmptyRows=FALSE)
      range_loss_all<-read.xlsx("Scriptie/Range/range_loss_all.xlsx", skipEmptyRows=FALSE)
      habitat_DI_loss_all<-read.xlsx("Scriptie/Diversity/Habitat/habitat_DI_loss_all.xlsx", skipEmptyRows=FALSE)
      dif_lu_hyde_2015<-read.xlsx("Scriptie/Anthromes/HYDE/dif_lu_hyde_2015.xlsx", skipEmptyRows=FALSE)

    #Final
      all_species_data<-data.frame(range_loss_all[,1],
                                   range_loss_all[,6],
                                   habitat_DI_loss_all[,5],
                                   climatic_overlap_loss[,5],
                                   popdens_hyde_2015[,4],
                                   dif_lu_hyde_2015[,5],
                                   IUCN_all[,3])


      final_data<-all_species_data[which(!is.na(all_species_data$range_loss_all...6.)),]
      colnames(final_data)<-c("Name", "Relative range loss ((pres-cur)/pres)", "Relative DI loss ((pres-cur)/pres)", "Climatic overlap loss", "Absolute di

      write.xlsx(final_data, "Scriptie/PCA/final_data.xlsx", showNA=TRUE)

  #PCA
    #Prepare

      min_max_norm <- function(x) {
        (x - min(x)) / (max(x) - min(x))
      }

      normalised_final_data<- as.data.frame(lapply(final_data[,2:7], min_max_norm))
      scaled_normalised_final_data <- scale(normalised_final_data)
      colnames(scaled_normalised_final_data)<-c("Range loss", "Habitat diversity loss", "Environmental diversity loss", "Population density", "Anthropoger

    #PCA
      pc <- prcomp(scaled_normalised_final_data, center = TRUE, scale. = FALSE)
      summary(pc)

     #Principal Component Axis
     fviz_contrib(pc, "var", axes = 1)
     fviz_contrib(pc, "var", axes = 2)
     fviz_contrib(pc, "var", axes = 3)
     fviz_contrib(pc, "var", axes = 4)
     fviz_contrib(pc, "var", axes = 5)
     fviz_contrib(pc, "var", axes = 6)

     fviz_cos2(pc, choice = "var", axes = 1:3)

   #Clustering
     #Number of clusters
       wss <- (nrow(final_data[,2:7])-1)*sum(apply(final_data[,2:7],2,var))
       for (i in 2:15) wss[i] <- sum(kmeans(final_data[,2:7],
                                            centers=i)$withinss)
       plot(1:15, wss, type="b", xlab="Number of Clusters",
            ylab="Within groups sum of squares")

       n_clusters<-5

       kpca<-kmeans(pc$x, n_clusters, nstart=50, iter.max=30000)
       plot(pc$x[,1], pc$x[,2],col=factor(kpca$cluster), xlab="PC 1", ylab="PC 2", main="Clusters of Species in PCA")
       plot(pc$x[,1], pc$x[,3],col=factor(kpca$cluster), xlab="PC 1", ylab="PC 3", main="Clusters of Species in PCA")
       plot(pc$x[,2], pc$x[,3],col=factor(kpca$cluster), xlab="PC 2", ylab="PC 3", main="Clusters of Species in PCA")
       plot(pc$x[,1], pc$x[,4],col=factor(kpca$cluster), xlab="PC 1", ylab="PC 4", main="Clusters of Species in PCA")

       plot3d(pc$x[,1],pc$x[,2],pc$x[,3], col=factor(kpca$cluster))

   #Clusters
     sort(table(kpca$cluster))
     clust <- names(sort(table(kpca$cluster)))

     n1<-row.names(decided_data[kpca$cluster==clust[1],])   # First cluster
     n2<-row.names(decided_data[kpca$cluster==clust[2],])   # Second Cluster
     n3<-row.names(decided_data[kpca$cluster==clust[3],])   # Third Cluster
     n4<-row.names(decided_data[kpca$cluster==clust[4],])   # Fourth Cluster

     clust1<-final_data[kpca$cluster==clust[1],]
     clust2<-final_data[kpca$cluster==clust[2],]
     clust3<-final_data[kpca$cluster==clust[3],]
     clust4<-final_data[kpca$cluster==clust[4],]
     clust5<-final_data[kpca$cluster==clust[5],]

     gap<-matrix(data=NA, nrow=1, ncol=7)
     colnames(gap)<-c("Name", "Relative range loss ((pres-cur)/pres)", "Relative DI loss ((pres-cur)/pres)", "Climatic overlap loss", "Absolute difference p

     clusters5<-rbind(clust1, gap, clust2, gap, clust3, gap, clust4, gap, clust5)

     write.xlsx(clusters5, "Scriptie/PCA/clusters5.xlsx", showNA=TRUE)
```