Michael Holthausen

Distribution modeling of paleofauna in the Western Mediterranean between the Heinrich events H5 and H4

Master's thesis

University of Cologne

September 2018

University of Cologne
Institute of Geography
AG GIS & Remote Sensing

Master's thesis

Distribution modeling of paleofauna in the Western Mediterranean between the Heinrich events H5 and H4

Michael Holthausen

mholthausen@icloud.com

Matriculation No.: 5617049

September 24, 2018

1. Examiner: Prof. Dr. Georg Bareth
2. Examiner: Dr. Christian Willmes

Acknowledgments

Abstract

This master's thesis deals with species distribution modeling (SDM) for eight selected prey animals of the Neanderthals and the anatomically modern human within the framework of the Collaborative Research Center 806 ("*Our way to Europe*"). This research was realized for three methods in three climatically different time slices during the Late Pleistocene. One profile method, one regression method and one machine learning method were used. A model was developed that performs these three methods in order to obtain a potential distribution of the paleofauna in the Late Pleistocene and to link it to the dispersal of humans in this region. The results show that all three methods predict conditions for the presence of the species which may have hunted from both Neanderthals and anatomically modern humans. However, there are differences in the predicted regions between the individual methods for each species. Another task was to determine the best performing method. Based on this work, MaxEnt, a machine learning method, emerged as the best performing method among the applied methods.

Contents

Contents

List of Figures

LIST OF FIGURES

List of Tables

List of Code Listings

List of Abbreviations

AUC Area Under the Receiver Operator Curve (AUROC). vii, 35, 44, 46, 49–52

AUROC Area Under the Receiver Operator Curve. 35, 44

CO Carbon monoxide. 21

COR Correlation. vii, 35, 46

CRC Collaborative Research Centre. 1

csv Comma-separated values. ix, 12, 25

D-O Dansgaard-Oeschger. 11

DEM Digital Elevation Model. vi, 2, 21–24, 47, 52

ENM Ecological Niche Model(ing). 3, 53

EPSG European Petroleum Survey Group Geodesy. 10

FAO Food and Agriculture Organization. 5

GAM Generalized Additive Models. 31

GBIF Global Biodiversity Information Facility. ix, 12

GEBCO General Bathymetric Chart of the Oceans. 20–22

GI11 Greenland Interstadial 11. vi, 2, 10, 11, 17, 19, 21–24, 28, 37, 39, 41–44, 47, 51

GIS Geographisches Informationssystem. 22

GLM Generalized Linear Models. ix, 31, 32, 37, 39, 41–44, 46, 47, 49, 54

GRASS Geographic Resources Analysis Support System. 22

H4 Heinrich event 4. vi, 2, 10, 11, 17, 19, 21–24, 28, 37, 39, 41–44, 47, 51, 52

H5 Heinrich event 5. vi, 2, 10, 11, 17, 19, 21–24, 28, 37, 39, 41–44, 47, 51

IRD Ice-rafted detritus. 11

KG Köppen-Geiger. 34

kya thousand years ago. 1, 10

m a.s.l. Meters above mean sea level. 5, 6

List of Abbreviations

MIS Marine Isotope Stage. 21, 52

NASA National Aeronautics and Space Administration. 20, 21

O Oxygen. vi, 10, 17, 21, 22

prec Precipitation. 18

QGIS QGIS Geographic Information System. 21, 24, 37

ROC Receiver Operating Characteristics. 34, 35

RSF Resource Selection Functions. 3

SAGA System for Automated Geoscientific Analyses. 24

SD Standard Deviation. vii, 46

SDM Species Distribution Model(ing). vi, 2–4, 8, 9, 28–30, 53

SRTM Shuttle Radar Topography Mission. 20, 21

SSB Spatial Sorting Bias. 27, 28, 46, 51

tmax Maximum temperature. 18

tmin Minimum temperature. 18

TSC Terrain Surface Classification. viii, 48

USA United States of America. 13, 17

VIF Variance Inflation Factor. viii, 28, 29, 50

WGS World Geodetic System. 10

1 Introduction

1.1 Main Purpose of the Work

How did the human species make it to Europe? This question and the related cultural-environmental context of the distribution of anatomically modern humans from their origin in Africa to Central Europe is the main objective of the Collaborative Research Centre 806 (CRC 806) "*Our Way to Europe*", a consortium of scientists and researchers from the universities of Cologne, Bonn, and Aachen. Archeological and geoscientific methods are used to understand the ancestral migration routes from Africa to Europe and to reconstruct past climates, cultures, population changes and living conditions (CRC 806 2018a). This work is a joint project of the clusters C1 (Settlement Patterns and Climate Change in the Late Pleistocene of the Western Mediterranean) and Z2 (Data Management and Data Services).

The framework of cluster C1 and this work is the transition from Middle to Upper Paleolithic. The period from 45 to 25 kya is assigned to the demise of the Neanderthals and the beginning of the dispersal of the anatomically modern human in Europe. The Iberian Peninsula and particularly its southern part is seen as a refuge for Neanderthals and is a central subject in the research of the project. A very diverse geography of the Iberian Peninsula and the resulting climatically very distinctive regions favor the research on relationships between the environment and human adaptation (CRC 806 2018b). One of the southernmost boundaries of the steppe-tundra expansion during the Pleistocene is formed by the Iberian Peninsula. It was settled from cold-adapted large mammals, which are also known as the *Mammuthus-Coelodonta* faunal complex (Kahlke 1999), during the coldest time periods in the Pleistocene (Álvarez-Lao & García 2010).

Selected species (Chapter 2.2.1) of these cold-adapted mammals are the subject of this work. The recent distribution based on the data of GBIF Secretariat (2017) forms the basis of the development of a model to compute the potential distribution. Furthermore,

Chapter 2 describes the preparation of the necessary data for specifying the environment in recent and past times, including bioclimatic variables (Hijmans et al. 2005; Willmes et al. 2018), Köppen-Geiger climate classification (Willmes et al. 2017), GebCo data (General Bathymetric Chart of the Oceans 2014) as a Digital Elevation Model (DEM) and Terrain Surface Classification according to Iwahashi & Pike (2007). The use of these data in the context of three diverse methods about species distribution modeling (SDM; Chapter 2.1) is intended to determine a possible distribution of prey for hunters. The spatiotemporal context is limited to the Western Mediterranean of the Late Pleistocene, more precisely Heinrich event H5, Greenland Interstadial GI11, and Heinrich event H4 (Chapter 2.2). The application and comparison of three representatives among the most commonly used methods within the distribution modeling using the paleo-environmental data is described in Chapter 2.3, followed by the results (Chapter 3). The discussion (Chapter 4) also deals with the question which one among the applied methods is the best performing method.

The compilation of the methods is mainly based on the recommendations of Hijmans & Elith (2017), who offer a good framework for the concepts in species distribution modeling. The selection of the species is based on Álvarez-Lao & García (2010), Salazar-García et al. (2013), and de Andrés-Herrero et al. (2018), while the implementation for the description of the environment is based on Dormann et al. (2013) and Beaumont et al. (2016). Elith et al. (2006) and Hijmans (2012) provide helpful hints for the evaluation of the applied methods in order to answer the questions of this work adequately.

1.2 Current Status in Research

In recent years, the access to species' occurrence data, also known as *primary biodiversity data* or *presence data* has increased (Soberón et al. 1996; Graham et al. 2004; Soberón & Peterson 2004). These opportunities are supported by the trend of digitizing the dataset, referencing it with geographic coordinates and simplifying the access to large databases for the observation of species. With this large data pool, a basis for the estimation of ecological niches is given. The additional environmental information needed, such as climate, topography, soils, oceanography, vegetation, and many more, is now available for almost the entire planet and in increasingly finer resolution (Peterson et al. 2011). The work of Grinnell, Hutchinson, and Austin, who became known as pioneers of the ecological niche, is gaining new popularity as ecological and

historical dimensions of biogeography can now be linked. These new enabled tools have been termed *species distribution modeling* (GUISAN & ZIMMERMANN 2000; HIRZEL et al. 2002; GUISAN & THUILLE 2005; ARAÚJO & GUISAN 2006). It is helpful to discuss actual and possible spatial distributions of species in the past, present, and future (PETERSON et al. 2011).

Not only the pool of a big data collection is used to accomplish this goal, but also the principle of machine learning, which was first described by SAMUEL (1959). He specifies machine learning as the generic term for the broad field of artificial generation of knowledge from experience using statistical techniques. An artificial system learns from examples and can generalize these after completion of the learning phase by recognizing patterns and regularities in the provided learning data, without being programmed for a specific task (KOZA et al. 2006). Thus, the system can also evaluate data which were previously unknown as a learning transfer or fail to learn these unknown data, which is referred to as overfitting or undercomputing (DIETTERICH 1995). The principle of distribution modeling uses these technical possibilities.

Since species distribution can be determined for terrestrial, marine, and freshwater environments, this work has been carried out by many biologists over a variety of species. Predominately terrestrial species were the main focus of attention to obtain a map of species density and identify conservation areas of biogeographically important species (HORTAL et al. 2004; BUSE et al. 2007). But also the occurrence of geomorphological structures can be identified with the help of the tool for distribution modeling, as LUOTO et al. (2004) conducted with a climate-based spatial model for the distribution of palsa mires in northern Europe, the most marginal permafrost features at the edge of the permafrost zone. Considering literature, the names of the models can be very different. SDMs can be named also *bioclimatic models*, *climate envelopes*, *ecological niche models* (ENMs), *habitat models*, *resource selection functions* (RSFs), *range maps*, *loosely-correlative models* or *spatial models* (ELITH & LEATHWICK 2009b). In this work the term SDM is used.

SDMs are being used more and more often in archeology as the research is moving from simple narrative biogeographic descriptions and interpretations of fossil data towards a new direction of describing and estimating species distribution in the past, shifts and extinction events (RODRIGUEZ-SANCHEZ & ARROYO 2008) or the effects of climate changes on species distributions and human cultures (MCDONALD & BRYSON 2010; POLLY & ERONEN 2011; WALLS & STIGALL 2011).

In addition to Anatolia, Turkey (Collins et al. 2018), the Iberian peninsula has been the subject of several paleogeographic investigations (Varela et al. 2010; Smith et al. 2013; Becker et al. 2016; de Andrés-Herrero et al. 2018) with the help of SDM. The main purpose of these studies was to link hunter-gatherer economies and lifestyles with the distributions of faunal and floral resources in the Upper Paleolithic an Neolithic.

A helpful tool used in many archeological studies is *R*, which was also applied in the calculations of this work. *R* is a programming language for statistical computing by R Core Team (2018) (version 3.5.0). The methods for SDM are provided in the package *dismo* (version 1.1-4) by Hijmans et al. (2017).

1.3 Study Region: The Western Mediterranean
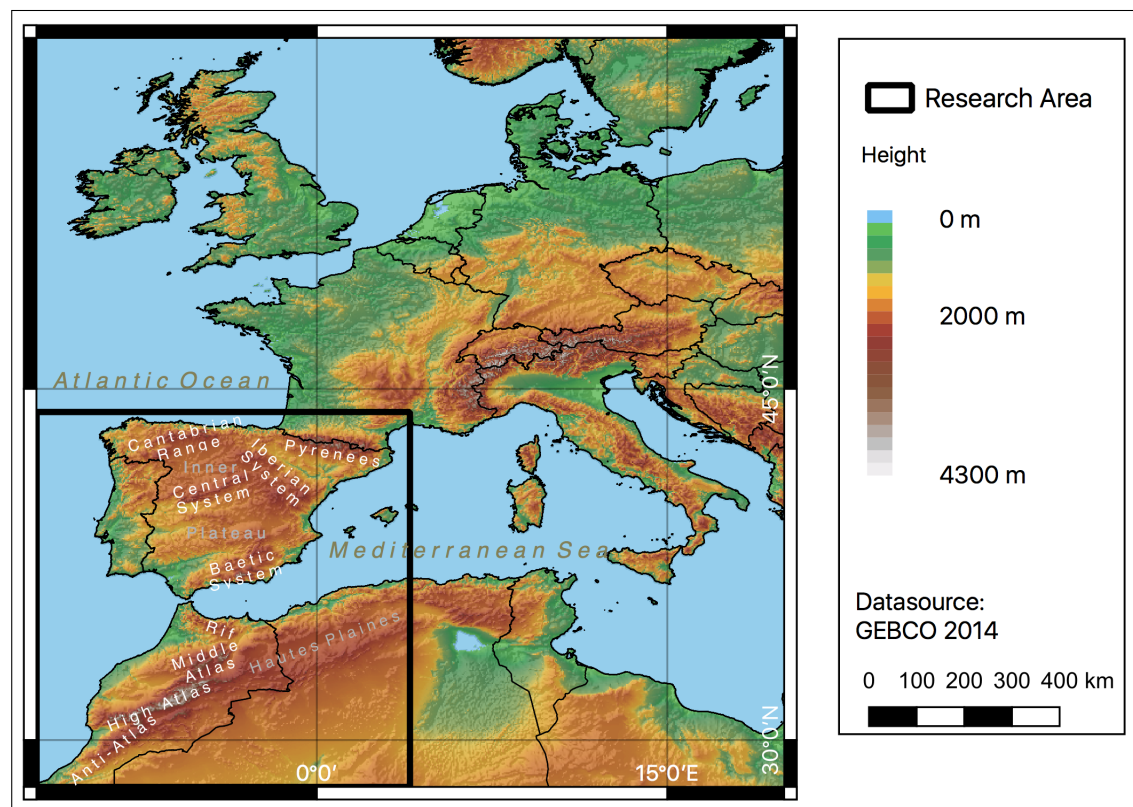
Figure 1: Map of the complete area for modeling and the research area of the Western Mediterranean (black rectangle).

The study area in the Western Mediterranean for this work consists of the Iberian Peninsula, which ranges today from Spain, Portugal, Andorra and Gibraltar in the

north, and Morocco on the African continent in the south (black rectangle in Figure 1). The Iberian Peninsula borders the Mediterranean Sea to the east and south and the Atlantic Ocean to the west and north. Morocco is bordered to the north by the Mediterranean Sea, to the east by Algeria, to the south by Western Sahara, and to the west and northwest by the Atlantic Ocean. The Strait of Gibraltar forms a narrow point between the Atlantic Ocean and the Mediterranean Sea respectively the Iberian Peninsula and Morocco.

The region is characterized by a Mediterranean climate (PEINADO LORCA & MARTÍNEZ-PARRAS 1987; IBÁÑEZ et al. 2013) with mild winters and hot and dry summers (SÁNCHEZ GOÑI et al. 2000). The northen Iberia is located in an Atlantic bioclimatic region (IBÁÑEZ et al. 2013), which is mainly influenced by the winter inter-annual variability mode and furthermore defines the North Atlantic Oscillation (TRIGO et al. 2004). The resulting strong pressure gradient between the Azores high and Icelandic low generates a strengthening and northward displacement of the westerlies over Europe. As a consequence, the Iberian Peninsula experiences different episodes of extreme drought (TRIGO et al. 2004). The geographic position of the Mediterranean Basin makes the region of the Western Mediterranean notably vulnerable to climate change (ADLOFF et al. 2015).

According to the Food and Agriculture Organization of the United Nations (FAO), Acrisols, Cambisols and Regosols are among the most widespread soils in the Iberian Peninsula. Cambisols, in particular, make up a large part of the distribution (FAO 1981). The beginning of soil formation allows intensive agricultural use of this productive soil. In the southwest of the Iberian peninsula, there are some soils of the Acrisol type, which severely restrict agricultural use. The western part of Iberia is associated with soils of type Regosol, which can be used for irrigated farming and grazing (FAO 1998). For Morocco, the FAO (1977) mainly shows Cambisols in the northern part. Leptosols and Calcisols are among the predominant types of soil in the south, but are also prevalent over the entire country (FAO 1977). Lithosols describe shallow soil over hard rock with less rainfed agriculture. The main uses are tree crops and extensive grazing. Calcisols are shallow, low-humus mineral soils with secondary lime enrichment of the semi-desert and predominantly used for low volume grazing of ungulates.

The region is characterized by wide mountain areas in Iberia and Morocco. The border in the northeast to France is formed by the Pyrenees. The central highland (Inner Plateau) of the Iberian Peninsula is crossed by three mountain ranges that strike from northeast to southwest: Cantabrian Range in the north, Central System in the interior of the peninsula and Baetic System in the south which reaches up to 3,483 m a.s.l. at

Mulhacén as the highest peak in the Iberian Peninsula (SIMÓN et al. 2000). Morocco is characterized by the Atlas Mountains that strike in general from northeast to southwest, especially in the northern part of the country. It is subdivided into the Anti-Atlas range in the south, the central High Atlas and the two northern ranges of the Middle Atlas and the Rif. Its highest point is marked by the summit of Toubkal with 4,167 m a.s.l. which is located in the south-western part of the High Atlas range (HUGHES et al. 2011). The southern and southeastern regions of Marocco are less densely populated, as desert areas of the Sahara begin here and are included in the climate region of the warm desert (KOTTEK et al. 2006).

2 Methods and Data

2.1 Species Distribution Modeling

Not only the relationship between ecological niches and geographic distributions of species over space and time can be described as complex. The concept of the ecological niche itself is also assigned to a broad variety and meanings. Often, the term *niche* is used in an attempt to find answers to the questions about the combination of environmental factors that enable a species to exist in a geographical area, as well as what effects the species has on these environmental factors (predictors). HUTCHINSON (1957) defined this space as the fundamental niche. It is described as a hypervolume of environmental variables in which each point corresponds to a state of environment that allows a species to exist indefinitely. The most common applications are based on a species in relation to its area of distribution (*existence* of a species) where the niche is considered as climatic and habitat requirements of the species (GRINNEL 1917). In contrast, ELTON (1927) regards the niche as the functional role of an animal on the community, meaning the local effects. The *existence* is taken for granted, but the focus is set on the *impact* (PETERSON et al. 2011).

The environment of a species consists of extremely diverse factors. These factors have to be distinguished into variables that describe environmental aspects which are influenced by the species and those that affect the fitness of the species without being altered. This creates a multivariate environmental space based on variables that are not dynamically modified (linked) by the species (i.e. climate, topography) and, in contrast, to variables that are dynamically modified (non-linked), such as consumed resources (HARPER 1977; AUSTIN 1980; AUSTIN & SMITH 1989; BEGON et al. 2006). HUTCHINSON (1978) also calls these non-linked variables *scenopoetic*. The variables, described in Chapter 2.2 and used in this work, belong to the scenopoetic variables and construct a multivariate environmental space in which niches are just simple subsets (PETERSON et al. 2011). The usage of scenopoetic variables is limited to climate data and topography data, as other environmental variables have no effect on the distribution of species in an area beyond

the regional scale domain (200-2,000 km). The applied scale in this work belongs to the continental domain (2,000-1,000,000 km; PEARSON & DAWSON 2003).

The term *species distribution* describes the spatial arrangement of a selected species across a defined landscape. The *species distribution model* (SDM) is a spatially explicit, quantitative model that relates occurrence data to environmental characteristics for the selected species (Figure 2). Quantifying the species occurrence related to environmental factors using SDM, enables a large number of approaches for scientific questions where a model-based prediction of the occurrence of a species is helpful (DUARTE et al. 2018). SDM is used widely across terrestrial, freshwater, and marine subjects. Therefore, the modeling requires sometimes extrapolation in space and time (ELITH & LEATHWICK 2009b).
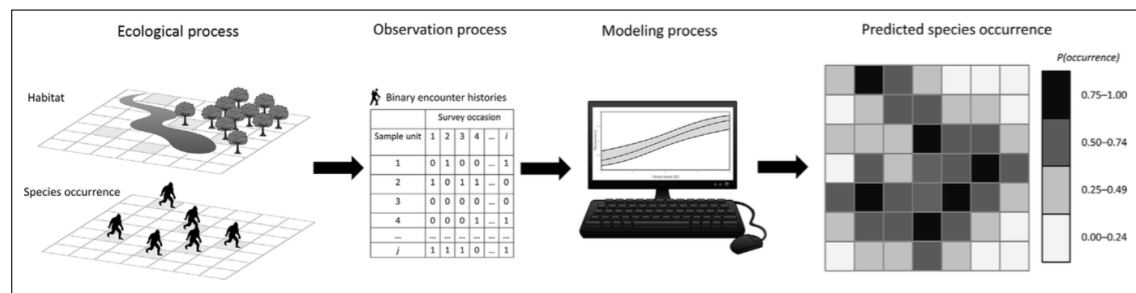
Source: DUARTE et al. (2018)

Figure 2: Visualization of the different components involved when developing species distribution models (SDMs).

For centuries, relationships between the distribution of a species and the physical environment have been identified and recorded. Initially, qualitatively, as described by GRINNEL (1904), the use of numerical models has been growing for quite some time in order to both describe patterns and make predictions (ELITH & LEATHWICK 2009b). The numerical models can be organized in different methods, depending on the type of data used in the model. These include presence-only data, presence/absence data, and replicate detection/nondetection data (DUARTE et al. 2018).

In models for presence-only data a random number of confirmed geographical locations where a species occurs is the minimum required data. These presence-only data can be related to landscape and climatic predictors if they are not overly distorted by sampling or bias (DUARTE et al. 2018). A bias of the data arises whenever an unevenly distributed sampling is present in the data (PETERSON et al. 2011). Methods using this type of data are commonly referred to as *machine learning methods*, because they randomly generate

its own background locations that may or may not overlap with the presence-only data. This will allow the method to estimate the full range of environmental conditions of a species potentially inhabiting the given region. Similar methods which do not generate any background or absence data are classified in the *profile methods* category (Duarte et al. 2018; Hijmans & Elith 2017).

But presence/absence data can also be used to develop SDMs. These data consist of combined successful and unsuccessful occurrences of a species during observation of a random subset or by generating pseudo-absence data points. Like with the presence-only models, the models of these methods are only useful for developing SDMs if the data are not compromised by sampling or bias. Presence/absence data are required in the different methods of *regression models* (Duarte et al. 2018; Hijmans & Elith 2017).

Often there is bias in observational data for a species that results from incomplete observations. For this, the terms "presence-only" and "presence/absence" were replaced for newer models such as *single-season*, *multiseason* and *dynamic occupancy models* with the terms "detection-only" and "detection/non-detection" (MacKenzie et al. 2017). These new type of methods are not used in this work.

Most older studies focus often on ecological aspects or evolutionary biology. The shift in focus to predicting distribution has been driven by advances in model algorithms, greater availability of spatial environmental data, and a growing demand for land management maps and allocation of protected areas. There are two ways the prediction in SDM can be used: a) in model-based interpolation to unsampled sites predictions are applied to new sites but within the same general time slice, and b) to carry out forecasts of new and unsampled regions for future or past climates, which is embraced by the terms extrapolation or forecasting. The data for new regions and time slices must be carefully evaluated, especially for the prediction values that are outside the region of training data (Elith & Leathwick 2009b).

Referred to Elith & Leathwick (2009a), Peterson et al. (2011), and Hijmans & Elith (2017), the first steps of a SDM consist of identifying the ecological process, collecting the necessary training data, and examining it. These include the species occurrence data and the environmental predictor variables that have to be used in order to model the potential habitats and distributions. Topographical and climatic data from different sources, in the form of tables for the individual species or as raster files for the predictor variables, were collated for this work. The following step of this work is the

preparation of the data for the selected study region in order to meet the requirements of the programming language and the respective method.

2.2 Spatiotemporal Context and Preparation of the Data

The raster files used for this work as input variables have a west-east extension of 12°W to 20°E and a north-south extension of 60°N to 28°N. With a cell size of 0.05 this corresponds to a resolution of 640x640 pixels. Both the input and the result maps refer to geographic coordinates (WGS-84) in an equirectangular (plate carrée) projection (EPSG: 4326). This expansion of the region had to be applied because at present times some species have suitable living conditions only outside the region of the Western Mediterranean. All finding spots of the species are located within this extended region. The results will be limited to the actual study area, the Western Mediterranean (Figure 1). Here, the west-east extent covers a range of 12°W to 4°E and a north-south extension of 44°N to 28°N.
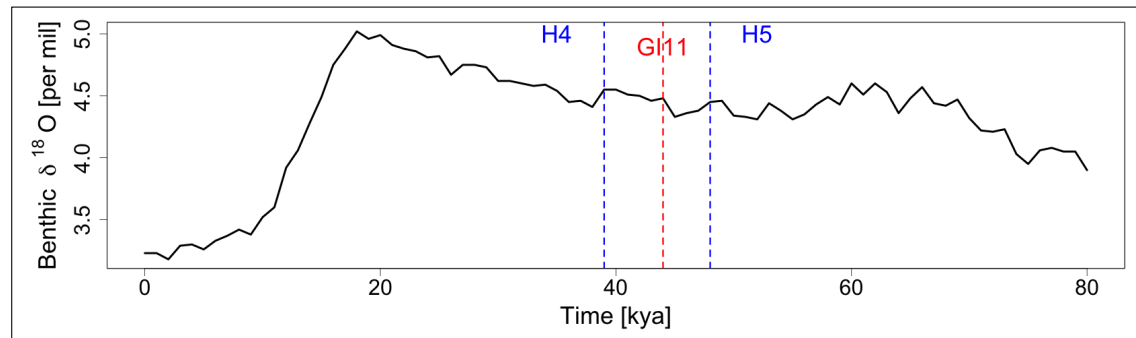
Source: Author's own, based on LISIECKI & RAYMO (2005)

Figure 3: Display of the selected time slices and their assigned $\delta^{18}O$.

Projective time slices (Figure 3) in the Late Pleistocene and the last glacial were chosen in whose the northern hemisphere and the study area experienced climate and environmental changes (SÁNCHEZ GOÑI & HARRISON 2010). With 47 kya, Heinrich event 5 (H5) is the oldest period in this model (HEMMING 2004). It is succeeded by a section of a short-term, relatively warm period within the glacial, the Greenland Interstadial 11 (GI11) with its midpoint at 43 kya (RASMUSSEN et al. 2014). The most recent period of this work is Heinrich event 4 (H4), which is dated to 38 kya (HEMMING 2004).

Heinrich events describe the point of time when, compared to the ambient conditions of the last glacial, extreme cold and dry glacial conditions dominated on the eastern North

Atlantic edge and the Western Mediterranean (HEMMING 2004). HEINRICH (1988) first documented these anomalous layers of ice-rafted detritus (IRD) which may have been derived from the region around Hudson Strait into the North Atlantic. HEMMING (2004) lists three possible mechanisms for the origin of the detritus layers:

- binge-purge cycle of the Laurentide ice sheet,

- jökulhlaup activity from a Hudson Bay lake (a massive flood that takes place when the height of a dam is exceeded by the lake level),

- ice shelf buildup or collapse fed by Hudson Strait.

Dansgaard–Oeschger (D-O) cycles (DANSGAARD et al. 1984) characterize a period of a rapid climate change, marked by warming and followed by a cooling and appear to be related to Heinrich events. Only the initial rapid warming event describes the D-O event. This event and the following slow phase of a subsequent period of cooling is referred to as the Greenland Interstadial (SÁNCHEZ GOÑI & HARRISON 2010).

The paleo environmental features are characterized by a mixed *Quercus* forest during interstadials and semi-desert vegetation during the Heinrich events and D-O stadials (SÁNCHEZ GOÑI et al. 2002; BOUT-ROUMAZEILLES et al. 2007; FLETCHER & SÁNCHEZ GOÑI 2008). The climatic characteristics of the first and last phases of H5 and H4 are described as mild and humid climate in southwestern Europe. The middle phase revealed a cold and dry climate in Iberia, especially the western Iberian Peninsula experienced an extreme continental dryness (SÁNCHEZ GOÑI et al. 2000; FLETCHER & SÁNCHEZ GOÑI 2008).

Sediment sequences of *Los Casares cave-Seno A* in the province of Guadalajara, Spain, contains evidence of both Neandertal and carnivore activity in the warm and humid period of GI11, which represents one of the latest occurrences of Neandertals in the interior from the Iberian peninsula. Total disappearance of Neandertals of interior Iberia is dated to H4. A survival of this human species in Iberia after H4 is limited to regions of the southern coasts (ALCARAZ-CASTAÑO et al. 2017).

2.2.1 Species

Eight prey species, which are not extinct at recent time, were selected for the modeling. Most of them are located in the region of the Western Mediterranean. The number of downloaded finding spots varies between 173 and 57606 (Table 1). The selection of the species is based on the previous research of DE ANDRÉS-HERRERO et al. (2018), SALAZAR-GARCÍA et al. (2013), and ÁLVAREZ-LAO & GARCÍA (2010). All finding spots were downloaded from the Global Biodiversity Information Facility (GBIF).

The GBIF provides its collected data about all types of life on Earth as open access. It was established in 2001 as an international network and research infrastructure and funded by the world's governments with its Secretariat in Copenhagen. With associated data-holding institutions around the world it is possible to share information about location and time of recorded species. Sources include museum specimens collected in the 18th and 19th centuries, as well as geotagged smartphone photos by amateur photographers and provided to the database at the recent time (GBIF Secretariat 2018).

The final lists contain all filtered finding spots of living animals (*human observation* and *living specimen* as basis of record) of the respective species to which geographical coordinates are available. The lists represent the population as of March 3, 2018. The function in Listing 1 from the *dismo* package downloads the species occurrence records from the GBIF database.

Listing 1: Function to download the genus and species related finding spots from GBIF, *Capra pyrenaica* in this example

```
1  gbif_species <- gbif(genus="capra", species="pyrenaica")
```

All selected species represent the cold-adapted mammals of the Iberian Late Pleistocene, which used to inhabit the Iberian Peninsula in its coldest period. Only a few radiocarbon dating have been performed in the past (ÁLVAREZ-LAO & GARCÍA 2010). Furthermore, there is no possibility of obtaining a list of geographical finding spots of the selected species, which has been assigned to the time slices. For this reason, both the species selection and the finding spots, were restricted to living specimen of today's time, which has been done in several studies before (DEPRAZ et al. 2008; RODRIGUEZ-SANCHEZ & ARROYO 2008; SOLOMON et al. 2008; ALBA-SÁNCHEZ et al. 2010).

The data are available as comma separated values (csv) and formatted as proposed by PETERSON et al. (2011). In addition to an identification number, the first two columns contain the longitude and then the latitude value, followed by an indication of presence

or absence. Since this is presence-only data, all occurrences in this column are filled with the number one.

Table 1: The selected species including their order and family for modeling and the imported number of finding spots

Genus & Species	Author & Year	English name	Initial finding spots
Artiodactyla			
Antilocapridae			
Rangifer tarandus	Linnaeus 1758	Reindeer	511
Bovidae			
Bos taurus	Linnaeus 1758	Ox	341
Capra pyrenaica	Schinz 1838	Spanish Ibex	1131
Rupicapra pyrenaica	Bonaparte 1845	Pyrenean Chamois	1072
Cervidae			
Capreolus capreolus	Linnaeus 1758	European Roe Deer	57606
Cervus elaphus	Linnaeus 1758	Red Deer	17697
Suidae			
Sus scrofa	Linnaeus 1758	Wild Boar	14358
Perissodactyla			
Equidae			
Equus caballus	Linnaeus 1758	Horse	173

Rangifer tarandus Linnaeus, 1758, the Reindeer (Figure 4), has a geographical distribution in the circumboreal regions of tundra and taiga in Svalbard, Norway, Finland, Russia, Alaska (USA), Canada, including the most arctic islands, as well as the north of Idaho and the Great Lakes region (USA), Greenland and south to north Mongolia. *Rangifer tarandus* is described as feral in Inner Mongolia and Heilungkiang, China in present time (GRUBB 1993a). Based on WEINSTOCK (1998) the reindeer was a regularly prey of human and non-

Source: iNaturalist.org (2018f)

Figure 4: *Rangifer tarandus* Linnaeus, 1758.

human hunters over a large geographical area from today's region of Ukraine in the east to Great Britain in the west during the last glacial.

Bos Linnaeus, 1758 describes the genus of wild and domesticated cattle (Figure 5) and is divided into five species: *Bos frontalis*, *Bos grunniens*, *Bos javanicus*, *Bos sauvali* and *Bos taurus* (GRUBB 1993a). While *Bos taurus* Linnaeus, 1758 is extinct in the wild, the distribution of the four remaining species is predominantly in Asia and irrelevant to this model. In the defined extent only *Bos taurus* plays a decisive role. *Bos taurus* mainly lived in Europe and western Russia. In the Middle East

Source: iNaturalist.org (2018a)

Figure 5: *Bos taurus* Linnaeus, 1758.

Bos taurus survived at least until the Iron Age and is nowadays held worldwide as a domesticated form (GRUBB 1993a). *Bos primigenius*, the aurochs, is an extinct species of the genus *Bos.* During the late Pleistocene and early Holocene, the aurochs dispersed over large areas of Europe, Asia and North Africa and is regarded as a wild ancestor of modern domesticated cattle (EDWARDS et al. 2010).

The geographical distribution of *Capra pyrenaica* Linnaeus, 1758, the Spanish Ibex (Figure 6), is exclusively limited to the Iberian Peninsula (GRUBB 1993a). The hypothesis of ENGLÄNDER (1986) considers that *Capra pyrenaica* derived from an Alpine Ibex (*Capra ibex*) type, which lived in the Pyrenees between the glaciations of Riss and Würm. Fossil remains of *Capra pyrenaica* have been found throughout the Iberian Peninsula and paleontological remains indicate the presence of the genus *Capra* in the Iberian Peninsula for at least a million

Source: iNaturalist.org (2018b)

Figure 6: *Capra pyrenaica* Schinz, 1838.

years (ALADOS & ESCÓS 1996). According to ALADOS (1985), this fact and findings from various prehistoric paintings suggest a wide distribution, as well as a wide abundance of the Spanish Ibex during the Paleolithic and Neolithic periods.

The geographical distribution of *Rupicapra pyrenaica* Bonaparte, 1845, the Pyrenean Chamois (Figure 7), is limited to northern Spain and the Apennine Mountains in Italy (Grubb 1993a). Chamois are described as "mountain-dwelling, middle-sized ungulates presently distributed over most of the middle- to high-altitude mountain ranges" (Masini & Lovari 1988). The genus *Rupicapra* originated in asian mountain areas west of the Himalayas during the late Miocene. *Rupicapra* dispersed in two different waves of colonization towards North America and western Europe during cold periods of the middle Pleistocene. *Rupicapra pyrenaica* possibly differentiated directly from older representatives in western Europe during the Riss glaciation (Masini & Lovari 1988).

Source: iNaturalist.org (2018g)

Figure 7: *Rupicapra pyrenaica* Bonaparte, 1845.

While *Capreolus capreolus* Linnaeus, 1758, the European Roe Deer (Figure 8), nowadays occurs in Europe to the west of Russia, Turkey, the Caucasis region, northwestern Syria, northern Iraq, and northern Iran, it is considered extinct in Lebanon and Israel (Grubb 1993a). The Roe Deer was already present in Europe at least 600.000 years ago and has since been known from the glacial and the interglacial. The northern border varied in Late Pleistocene, but was distributed up to 10 degrees further north in the western parts compared to its present distribution area, but less distinctive in the eastern part. Within interstadial episodes the distribution area expanded further northward into the Ural Mountains and the mountains of southern Siberia, where the Roe Deer survived in regions of forest and bush vegetation in periglacial forest-steppes. Therefore, the distribution of the Roe Deer was not only restricted to the Western Mediterranean but repeatedly reached regions of central Europe and farther, which is probably due to the favorable climatic conditions during the relatively mild interstadial episodes (Lister et al. 1998; Sommer et al. 2009).

Source: iNaturalist.org (2018c)

Figure 8: *Capreolus capreolus* Linnaeus, 1758.

Cervus elaphus Linnaeus, 1758, the Red Deer (Figure 9), has a wider geographical distribution. In addition to Tunisia and northeastern Algeria, the Red Deer was able to settle in Europe east to the Crimea and the Caucasus. Furthermore, the habitat extends over Turkey, Northern Iran, Central Mongolia, the west and north of China and the Ussuri region (Russia). In Corsica and Sardinia, the Red Deer settled since the Neolithic. The distribution in North America does not matter for this modeling (Grubb 1993a). European red deer in Late Pleistocene were very much larger than members of the same species in recent times or the post-Pleistocene (Beninde 1937; Walvius

Figure 9: *Cervus elaphus* Linnaeus, 1758.

1961). Hominins of the European continent hunted and used the Red Deer as a natural resource since the Middle Pleistocene (Sommer et al. 2008). Clutton-Brock et al. (1982) specified the body size of red deer to reach its peak during the last glaciation and has gradually decreased until present day.

The distribution of *Sus scrofa* Linnaeus, 1758, the Wild Boar (Figure 10), which is relevant for the model, is stated by Grubb (1993a) with North Africa and Europe. Additional occurrences were reported from Southern Russia and from China all the way south to the Middle East and meanwhile as extinct in British Isles and Scandinavia. The population of Corsica and Sardinia and also formerly Egypt and North Sudan are or were old feral origin. The Wild Boar is widespread as feral pop-

Figure 10: *Sus scrofa* Linnaeus, 1758.

ulations among others in Norway, Sweden and South Africa. According to Melis et al. (2006) the number of Wild Boars is strongly and positively correlated with the mean annual temperature, with an expected increase as well in local population densities as an expansion of the geographical range north and north-eastwards. Biogeographically, hunting by wolves seems to have a lower impact on wild boar density than climate and vegetation productivity, although wild boars have been reported as the preferred prey of wolves, especially in the southern regions of Europe.

Already in the classical antiquity the distribution of the domestic horse (Figure 11) is specified up to the European west to Spain. The further distribution into the late 18th Century of *Equus caballus* Linnaeus, 1758 ranges from Poland and Russian Steppes east to Turkestan and Mongolia. Wild population survived in southwestern Mongolia and adjacent Kansu, Sinkiang, and Inner Mongolia (China). But they are domesticated worldwide and feral in Portugal, Spain, France, Greece, Iran, Sri Lanka, Australia, New Zealand, Colombia, Hispaniola, Canada, USA, Galapagos, and other oceanic islands (GRUBB 1993b). *Equus caballus* is able to tolerate a wide range of temperatures and domiciled with more or less open forest but can also be present in a number of different paleoecosystems (ÁLVAREZ-LAO & GARCÍA-GARCÍA 2006).

Source: iNaturalist.org (2018e)

Figure 11: *Equuus caballus* Linnaeus, 1758.

2.2.2 Bioclimatic Data

Environmental data are provided by the bioclimatic variables of WorldClim (HIJMANS et al. 2005) through interpolated climate surfaces for worldwide land areas. Climate data at this very high resolution may be very beneficial for modeling mountain areas and other areas with strong climate gradients. The 30 arc-seconds tiles of WorldClim are provided in GeoTIFF format.

The global climate data collection for current conditions consists of interpolations of observed data and is representative for the climatological reference period from 1960 to 1990. The bioclimatic variables were calculated by average monthly climate data for minimum, mean, and maximum temperature, as well as precipitation.

The bioclimatic variables of time slices H5, GI11 and H4 were derived according to the method of WILLMES et al. (2018), where monthly paleoclimate data are calculated for the last 1.2 million years in a 1,000 year temporal resolution. The method is based on a spatio-temporal interpolation on the basis of $\delta^{18}O$ values between modeled paleoclimate and observed current climate data. An adaptation in terms of extent, resolution, and cell size to the project-related porperties is necessary.

Listing 2: Function to create the bioclimatic datasets

```
1  b <- biovars(prec,tmin,tmax)
```

The function in Listing 2 of the *dismo* package enables the creation of bioclimatic variables (Nix 1986; Booth et al. 2014) from monthly climate data, like the above mentioned paleoclimatedata. The data are imported as raster files of precipitation data (prec), minimum temperature data (tmin), and maximum temperature data (tmax). The results are generated in raster files and consist of 19 variables for each time slice (Table 2). Subsequently, these bioclimatic variables are in the same format as World-Clim, but without having to adjust them in resolution since they have already been generated with the project-specific properties.

Table 2: Description of the 19 bioclimatic variables which are generated from the *dismo* package (Hijmans et al. 2017)

File	Bioclimatic variable	Variable	Description
bio1	Annual Mean Temperature	bio11	Mean Temperature of Coldest Quarter
bio2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	bio12	Annual Precipitation
bio3	Isothermality (bio2/bio7) (* 100)	bio13	Precipitation of Wettest Month
bio4	Temperature Seasonality (standard deviation *100)	bio14	Precipitation of Driest Month
bio5	Max Temperature of Warmest Month	bio15	Precipitation Seasonality (Coefficient of Variation)
bio6	Min Temperature of Coldest Month	bio16	Precipitation of Wettest Quarter
bio7	Temperature Annual Range (bio5-bio6)	bio17	Precipitation of Driest Quarter
bio8	Mean Temperature of Wettest Quarter	bio18	Precipitation of Warmest Quarter
bio9	Mean Temperature of Driest Quarter	bio19	Precipitation of Coldest Quarter
bio10	Mean Temperature of Warmest Quarter		

Mean annual temperature and annual precipitation represent annual trends, annual range in temperature and precipitation impart seasonality and temperature of the coldest and warmest month, and precipitation of the wet and dry quarters indicate extreme or

limiting environmental factors. These bioclimatic variables are often used in species distribution modeling and the similar ecological modeling techniques (DE ANDRÉS-HERRERO et al. 2018; BECKER et al. 2016; FRANKLIN 2010; VARELA et al. 2011) since "climate is the most important variable in reconstructing potential species distribution because it is one important factor that limits the geographic range of different species" (DE ANDRÉS-HERRERO et al. 2018: 199).

2.2.3 Köppen–Geiger Climate Classification

A first climate-induces classification was published by KÖPPEN (1884) followed by several modifications (KÖPPEN 1918; KÖPPEN 1936). Further changes in the middle of the 20th century culminated in the climate classification system of Köppen-Geiger (GEIGER 1954; GEIGER 1961).
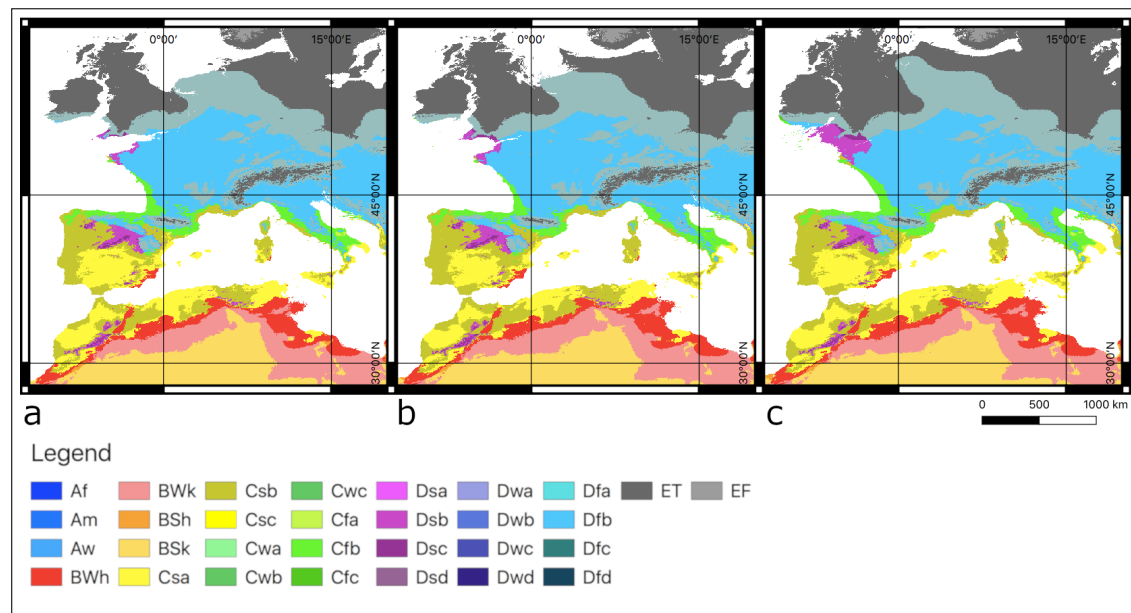
Source: Author's own, based on WILLMES et al. (2017)

Figure 12: Calculations of the Köppen-Geiger climate classification for the selected time slices H5 (a), GI11 (b), and H4 (c).

An updated classification scheme of PEEL et al. (2007) was applied to generate classification maps for recent and past times based on the climate parameters temperature and precipitation described in Chapter 2.2.2 (WILLMES et al. (2017); Figure 12). Six main groups are derived and described by the first letter. These groups are sub-divided by 12

climate types which are described by the second and third letter, based on the combination of temperature and precipitation. As a result, 30 distinct climatic conditions can be specified by this classification.

In all time slices the northern regions are described by the E (polar) and D (cold) climates. The regions of the Iberian Peninsula and Morocco are characterized by D, C (temperate), and B (arid) climates. The A climates (tropical) are not present in the study area. The climate classification belongs to the categorical variables, since it only accepts one status out of limited possible values per cell.

2.2.4 Gebco Data

The 2014 version of GEBCO (General Bathymetric Chart of the Oceans 2014) was used to create a set of topographic data to represent the different sea levels of the individual time slices. GEBCO_2014 contains authoritative data of the most recent digital bathymetric model of ocean floors worldwide. According to WEATHERALL et al. (2015), the bathymetric dataset was merged with topographic data to create a continuous terrain model. As source for land areas digital terrain models of the Shuttle Radar Topography Mission (SRTM30) from U.S. National Aeronautics and Space Administration (NASA) (FARR et al. 2007) were selected. Any depth inconsistencies that could result from vertical reference differences are rather minor, considering the coarse grid resolution of the dataset of 30 arc-seconds.

Several institutions have contributed to the overall dataset. WEATHERALL et al. (2015) list the type of contributors:

- bathymetry surveys over continental shelves

- regional seafloor mapping contributions

- bathymetric data provided in the form of soundings

- multibeam data

- crowd-sourced bathymetry collected by fishery vessels

- satellite altimetry data

Topographical data south of 64°N and north of 60°S, as already mentioned, results from version 2.0 of the SRTM30 elevation model, while areas of ocean floor that could not be captured by the aforementioned contributors were interpolated by satellite-derived

gravity data. SMITH & SANDWELL (1997) and SANDWELL et al. (2003) consider satellite ocean floor mapping as an advantage due to their relatively higher speed and low cost compared to the detection by ships. The sea surface has broad elevations and depressions that can reflect the topography of the ocean floor. Features like seamounts have minor affects on gravity but can trigger tiny variations in ocean surface height. These variations can be mapped with very accurate radar altimeter attached to satellites and compared to the theoretical ocean surface.

The land surface is mapped with data from the images of NASA's SRTM data (FARR et al. 2007). Regions outside the United States of America are released originally with a 3 arc-second sampling (Jet Propulsion Laboratory 2018). For processing in the GEBCO data these recordings were resampled to a resolution of 30 arc-seconds (WEATHERALL et al. 2015).

Table 3: Overview of the δ^{18}O values in per mill (LISIECKI & RAYMO 2005) in the time slices determined by HEMMING (2004) and RASMUSSEN et al. (2014) and the corresponding difference to the current sea level in meters according to ROHLING et al. (2014)

Time slice	δ^{18}O	Sea level	kya
H4	4.41	-103	38
GI11	4.46	-69	43
H5	4.38	-53	47

The dataset must be edited individually for each time slice to get a feasible Digital Elevation Model (DEM). Using the *Raster Calculator* with QGIS Geographic Information System (QGIS) Software in version 3.2 (QGIS DEVELOPMENT TEAM 2018), a mask is first created from the GEBCO dataset to specify the regions of landmass and sea surface. Using the benthic δ^{18}O values by LISIECKI & RAYMO (2005) of each time slice, the sea level can be determined according to ROHLING et al. (2014).

The δ^{18}O value is a measure of the ratio of the stable oxygen isotopes ^{18}O and ^{16}O in per mill. First, UREY et al. (1951) and EPSTEIN et al. (1953) developed a method to use oxygen isotope composition of calcite as paleothermometer. EMILIANI (1955) used oxygen isotope paleothermoetry as a proxy to reconstruct the glacial and interglacial alterations in climate of the Late Pleistocene through foraminifera shells from deep-sea sediments. As a result, he introduced the Marine Isotope Stages (MIS). The selected time slices are classified in MIS 3 and thus correspond with the Weichselian glaciation in northern Europe and Würm in the Alpine region (SVENDSEN et al. 2004). Benthic ^{18}O is determined by isotopic equilibration of CO_2 gas with seawater at a constant temperature as originally conducted by EPSTEIN & MAYEDA (1953). Then, a delta (δ) value is derived

from the difference in the ratio of a sample value compared to the standard (Ravelo & Hillaire-Marcel 2007).

The relative sea level in Table 3 of the Strait of Gibraltar was calculated on the basis of Rohling et al. (2014). Here, the difference of the δ^{18}O measured at planktonic foraminiferal species *Globigerinoides ruber* with the present value is required. This MEDSTACK by Wang et al. (2010) is used to determine the difference of the δ^{18}O values (H5: 1.02, GI11: 1.34, H4: 2.06).

$$y = -54.33006067x \; + \; 2.144129497x^2 \tag{1}$$

In Equation 1, x is the difference of the δ^{18}O values and y indicates the relative sea level of the selected state. These computed values of the relative sea levels were used for the masks.

While land areas in the mask are given the value one, sea surface is marked with zero. Using the respective masks, the *Raster Calculator* generates a DEM from the GEBCO dataset by multiplying all values from the GebCo dataset corresponding to a landmass by one while setting all other values to zero (Figure 13). From the results, a slope and aspect record were additionally derived using the Geographic Resources Analysis Support System (GRASS GIS) Software in version 7.4 by GRASS Development Team (2018). A 3x3 neighborhood around each cell in the raster elevation map is used to determine slope and aspect. Therefore, a determination of the cells at the edges is not possible and these cells are assigned to zero in both the slope and the aspect (Shapiro & Waupotitsch 2015).

GebCo data has been used in numerous projects about species distribution modeling creating a selection of environmental grids to represent conditions at the recent seafloor and land topography, among others Vierod et al. (2014), Becker et al. (2016), and de Andrés-Herrero et al. (2018).

2.2.5 Terrain Surface Classification Data

The creation of terrain surface classification follows the approach of Iwahashi & Pike (2007). Here, the topography represented by DEMs is automatically classified into eight to 16 terrain types (Table 4). For this purpose, morphometric derivations of the geometric values of the gradient slope, local convexity, and surface texture are used. An
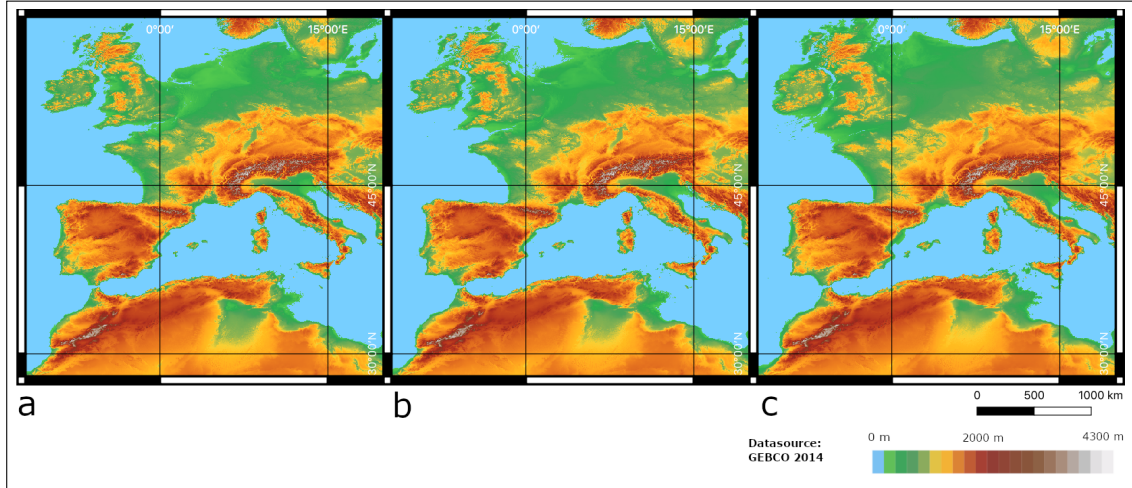
Figure 13: DEMs of the selected time slices H5 (a), GI11 (b), and H4 (c).

unsupervised nested-means-algorithm generates the categories by sequentially combining binarized grayscale raster files of the three variables. The thresholds for the subdivision of the images are randomly set to the mean values of frequency distributions of the input variables to ensure statistically strong classes. While slope and texture can support in the detection of mountains, the convexity of the surface is helpful in differentiating between low-reliefs as alluvial fans, flood plains, and river terraces. These data are used to classify other terrain parameters such as geomorphometrically determined information on landforms and terrain types, divided into classes. These values of the terrain surface classification, in addition to the bioclimatic and topographic values, are assigned to the characteristics of the individual finding spots of the species. So far, the usage of terrain surface classification has not been documented in species distribution modeling.

Table 4: Surface geometry and the corresponding terrain classes according to IWAHASHI & PIKE (2007)

Surface geometry	Terrain Classes			
fine texture, *high convexity*	1	5	9	13
fine texture, *low convexity*	3	7	11	15
coarse texture, *high convexity*	2	6	10	14
coarse texture, *low convexity*	4	8	12	16
	steep	⟷		*gentle*

23

The dataset is derived through the tool *Terrain surface classification* from the System for Automated Geoscientific Analyses (SAGA) by CONRAD et al. (2015) which is delivered with the current version of QGIS. This tool is included in the morphometry subfolder for *Terrain Analysis* and the DEM is required as input variable. Other optional parameters can be slope, convexity and texture. Without specification by the user, these are automatically generated and applied when creating the *Terrain surface classification*. As provided by IWAHASHI & PIKE (2007), the result for 16 classes is generated in a four-neighborhood setting, so there is no data at the edges. For each time slice, a raster file is created from the previously generated DEM. The results for all time slices are mapped in Figure 14. By limiting the amount of classes to 16, this variable is also classified as categorical.
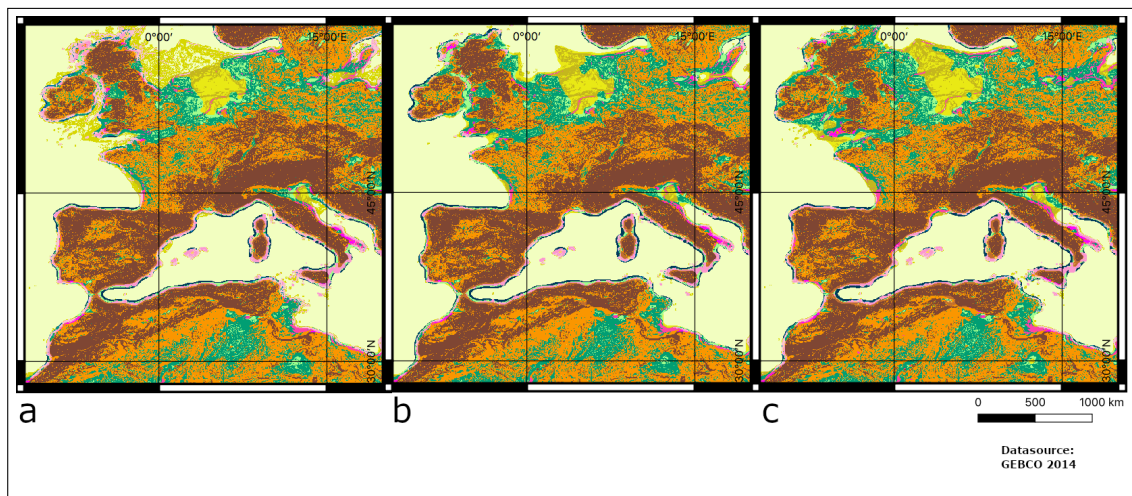
Source: Author's own

Figure 14: Visualization of the terrain surface classification for the selected time slices H5 (a), GI11 (b), and H4 (c). Legend in Table 4.

2.3 Implementations with R

2.3.1 Gathering Relevant Data

The following section describes the four major sections of the model used to determine the distribution of the eight selected paleofauna species in the Western Mediterranean of the time slices H5, GI11 and H4. The process within the model is based on the basics of SDM by ELITH & LEATHWICK (2009a) in general and on HIJMANS & ELITH (2017)

in detail, as shown in Figure 16. The collection and preparation of the data mentioned in chapter 2.2 is to start out on every modeling process. By complying with the applied specifications, the data on the species or environmental data as well as on the other predictor variables (for example, the climate status of past time slices) can easily be exchanged. In the model, the prepared finding spots of the selected species are imported first (Listing 3), e.g. those of the *Capra pyrenaica* as shown in Figure 15.

Listing 3: Import of the csv file to the species finding spots

```
1  file <- "Capra_pyrenaica.csv"
2  species.data <- read.table(file, header = TRUE, sep = ";")
```

In a next step, called the data cleaning, finding spots must be checked for plausibility and accuracy. This step is necessary for all data to avoid incorrect data during import. In particular, this step is important in order to prepare data from an external database. In addition to the accuracy, the structure of the external data can also be checked (HIJMANS & ELITH 2017). Furthermore, duplicates that can result from multiple entries of an observation, are removed in this step. For the purposes of this model, any finding spots that are falsely referenced to water surfaces will be removed. The resulting record will be referred to as `species.clean` in the following code listings. Subsequent subsampling

id	x	y	pres_abs
1	-3.962869	38.215817	1
2	-4.551111	36.949444	1
3	-4.543735	36.953755	1
4	-4.542503	36.964432	1
5	-3.748738	37.244798	1
6	-3.777111	36.742233	1
7	-5.422859	36.684678	1
8	-3.383709	37.094074	1
9	-4.062935	36.872601	1
10	-4.549685	36.628485	1

Source: Author's owniNaturalist.org (2018g)

Figure 15: Extract of the first finding spots to *Capra pyrenaica* and formatted for import.

on the cleaned finding spots corrects any bias (Listing 4), as these are often present in observational data. They might result from differences in the likelihood of finding one species in comparison to another species. In addition, differences in frequency or certain preferences in observation for a particular species, due to a particular expected feature or need for specific research, may lead to biases in different forms (HIJMANS et al. 2001).

Listing 4: Removing bias in occurrence data through sampling

```
1  grid_resolution <- 1
2  r <- raster(species.clean)
3  res(r) <- grid_resolution
4  r <- extend(r, extent(r)+1)
5  species.sampl <- gridSample(species.clean, r, n=1)
```

It must be pointed out that subsampling reduces the number of available finding spots and that regions which have not been sampled before, can not be corrected. On the other hand, it has to be considered that a local density of the finding spots can actually be a true reflection of the relative suitable habitat (Hijmans & Elith 2017). For this reason, no subsampling of a possible bias for *Rangifer tarandus* was carried out because today's habitat in the extended study area is restricted to the northernmost regions only. Furthermore, the recent distribution of *Rupicapra pyrenaica* is limited to the region of the Pyrenees and was also modeled without the removal of any bias.

The environmental data are imported as a description of the environment. These include the bioclimatic variables, the climate classification, the DEM, the slope and aspect and the terrain surface classification data. These data are stacked as predictor variables for the present time and the time slice to be modeled.
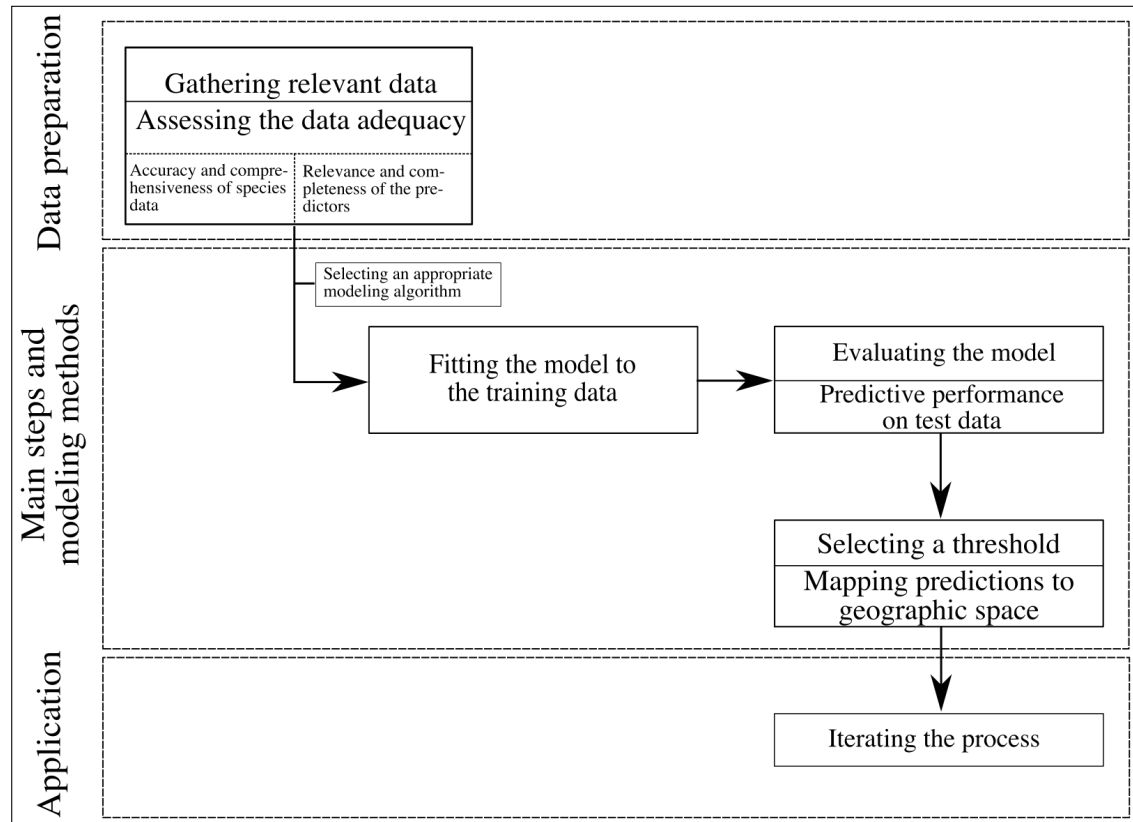
Source: Author's own, based on Elith & Leathwick (2009b)

Figure 16: The main sections of the model to determine the distribution of the selected species.

Many of the first methods only use presence-only data. More modern methods use presence/absence data (Table 5). But even if only presence-only data are available, as in this work, they can be processed in methods for presence and absence data by substituting missing absence data with background data. This background data are not intended to estimate potential absence locations, but rather to characterize the selected environments at those points in the study region. Through this concept of *pseudo-absence* data, a description of the study region is obtained, while the presence data describe under which conditions the species is more present than on average. The function `randomPoints` provided in Listing 5 generates 1000 random points within the specified extent that will be used henceforth as background data. The predictors stack is used as a mask, which allows generating the background points only in areas on land. In the following, the random points are divided into 800 points (80%) for training and 200 points (20%) for testing. The allocation is done randomly by the function `kfold`, whereby each record in a matrix is randomly assigned to a group whose number ranges between 1 and *k* (HIJMANS & ELITH 2017). In this case, 5 groups were created.

Listing 5: Generating random points as background data based on a mask and a partitioning into test and training dataset. A pair-wise distance sampling is done due to spatial sorting bias in the training data

```
1  backg <- randomPoints(predictors_recent, n=1000, ext=ext)
2  colnames(backg) = c( 'lon' ,  'lat' )
3  group <- kfold(backg, 5)
4  backg_train <- backg[group != 1, ]
5  backg_test <- backg[group == 1, ]
6  sb <- ssb(pres_test, backg_test, pres_train)
7  sb
8  ssb_res <- sb[,1] / sb[,2]
9  ssb_res
10 i <- pwdSample(pres_test, backg_test, pres_train, lonlat=TRUE, n
      =1, tr=0.1, warn = TRUE)
11 pres_test_pwd <- pres_test[!is.na(i[,1]), ]
12 backg_test_pwd <- backg_test[na.omit(i), ]
```

Since there is often a bias in these data, too, it must be removed in order to use them in the following steps. After generating background points, the presence data is divided into two random groups, a training and a test set. HIJMANS (2012) suggests a removal of the *spatial sorting bias* (SSB) through a *pair-wise distance sampling* (Listing 5). This is done with a measuring on the difference between a testing dataset and training dataset in the average distance to the nearest point in a reference training dataset. If the indicator

of SSB is close to zero, there is an SSB in the data. A value near to one indicates no SSB (HIJMANS & ELITH 2017). Due to lack of finding spots for the individual time slices from H4, GI11 and H5, only evaluations of these test and training data of a species from the respective method can be delivered and analyzed.

Table 5: Data requirements for the methods used in this work

Method	required data
Bioclim	presence-only
GLM	presence/(pseudo-)absence
MaxEnt	presence-only

With the package *dismo* (HIJMANS & ELITH 2017), which is used in this work, the specific environment values at the presence and background points do not have to be determined manually, since this is done automatically by the provided methods. Furthermore, the methods provided by *dismo* do not use a formula for fitting. Only the regression methods require a fitting by formula in order to identify the dependent and independent variables of the predictor variables (HIJMANS & ELITH 2017). To solve a potential problem with multicollinearity, there are two approaches. The package *usdm* (NAIMI et al. 2014) provides tools to analyze uncertainty in SDM. The package *dismo* allows only a visual check for multicollinearity. This is why the function `vifstep` of *usdm* is used to identify dependent variables from the predictor stack that have a multicollinearity. Using predictor variables in which multicollinearity exists, can cause instability, especially in the regression-type methods (NAIMI et al. 2014). Multicollinearity is described by ALIN (2010) as the condition where two or more predictor variables in a statistical model are linearly related. Twelve of the 19 bioclimatic input variables have a multicollinearity problem (bio5, bio10, bio11, bio16, bio7, bio12, bio17, bio1, bio2, bio18, bio13, bio9) and should not be used from a statistical point of view. This was determined by the variance inflation factor (VIF), which can be derived as shown in Equation 2.

$$1/(1 - r_i^2) \tag{2}$$

The coefficient of determination on all remaining variables for the ith variable is described through r_i^2. All results with a VIF greater than ten indicates variance more than ten times as large as in the case of orthogonal predictors (BELSLEY 1991). Without a deeper ecological understanding, it will not be possible to separate collinear variables by statistical means. (DORMANN et al. 2013). Thus, as a second approach, DORMANN et al. (2013) suggest a general rule, to use those variables that are ecologically relevant, feasible

to collect data on, and are closest to the source of the sequence resource-direct-indirect-proxy. BEAUMONT et al. (2016) identified eight bioclimatic variables, that were selected because of their influence on both physiological and distributional limits of vertebrates (Table 6). In this selection of bioclimatic variables a multicollinearity is also recognizable, but these are the ecologically meaningful predictors. Furthermore, this provides a common set of variables to build consistency between models and support direct comparisons (BEAUMONT et al. 2016). Therefore, these selected bioclimatic variables are used for the further course of action.

Table 6: Bioclimatic variables and their VIF used for modeling

Bioclimatic variable	VIF
Annual Mean Temperature (bio1)	118.8
Temperature Seasonality (bio4)	9.4
Max Temperature of Warmest Month (bio5)	63.3
Min Temperature of Coldest Month (bio6)	29.0
Annual Precipitation (bio12)	131.7
Precipitation Seasonality (bio15)	6.4
Precipitation of Wettest Month (bio13)	68.0
Precipitation of Driest Month (bio14)	22.0

For the SDM of a species, a variety of statistical and mathematical techniques can be used, all designed with one goal in mind: to correlate data about the observed distribution of a species with a set of spatially defined variables. Usually, these are variables that define the environment (GUISAN & ZIMMERMANN 2000; FRANKLIN 2010). There is still ambiguity in the classification of the individual methods in literature. In this work, the classification of the modeling methods provided by HIJMANS & ELITH (2017) is used.

2.3.2 Selecting a Modeling Algorithm and Fitting of the Model

2.3.2.1 Profile Methods

Bioclim is used as an exemplary method for profile methods, because it is integrated in the package *dismo*. In addition, the methods Domain and Mahalanobis are provided in this package, though Bioclim is one of the extensively used methods in SDM and a classic *climate-envelope-model* (BOOTH et al. 2014).

NIX (1986) produced the first version of the Bioclim package as the growing desire came up to generate maps of a multi-dimensional environmental space in which a species may potentially occur. This environmental space is constructed as a kind of bounding

box around the minimum and maximum values of the environmental variables for all occurrences of the species, resulting in a so-called multi-dimensional rectilinear envelope. It has been developed to model species distributions in relation to climatic variables (Figure 17). This assumes that species occurrence is influenced by climate. Categorical variables (such as climate classification and terrain surface classification) can not be processed and must be removed from the predictor variables (Listing 6) before applying the model (HIJMANS & ELITH 2017).

Listing 6: Removal of the categorical variables from recent and past predictor stacks

```
1  pred_nf <- dropLayer(predictors_recent, "iwahashi")
2  pred_nf <- dropLayer(pred_nf, "KG")
3  pred_nf_past <- dropLayer(predictors, "iwahashi")
4  pred_nf_past <- dropLayer(pred_nf_past, "KG")
```

The algorithm set up here calculates the similarity of a location by comparing the values of environmental variables at any given location in the study area with a percentile distribution of the values at known locations of occurrence, the training sites (Figure 17). The closer the value reaches the 50th percentile, which is the median, the more convenient the location is and the higher the probability of occurrence. Since the tails of the distribution will not be differentiated, the 10th percentile is treated equally as the 90th percentile. The scores for each environmental variable are combined to a total probability of occurrence for each site with equal weights for all environment variables (ARAÚJO & GUISAN 2006; BOOTH et al. 2014; HIJMANS & ELITH 2017).

HIJMANS et al. (2017) integrated a transformation of the percentile in the *dismo* package to generate a result that is alike the results of other SDM methods and should therefore be easier to interpret. Usually percentile scores range between zero and one, but here predicted values greater than 0.5 are subtracted from one. Then, the minimum percentile score is calculated across all environmental variables. The final value is subtracted from one and multiplied by two so that all results range between zero and one.

Listing 7: Fitting the environment variables and predicting the potential distribution with Bioclim

```
1  bc <- bioclim(pred_nf, pres_train)
2  pb <- predict(pred_nf_past, bc, ext=ra_ext)
3  e <- evaluate(pres_test_pwd, back_test_pwd, bc, pred_nf)
```

The function in Listing 7 calls the fitting of the environmental data for Bioclim. The values from the predictor variables at the training data do not have to be extracted man-
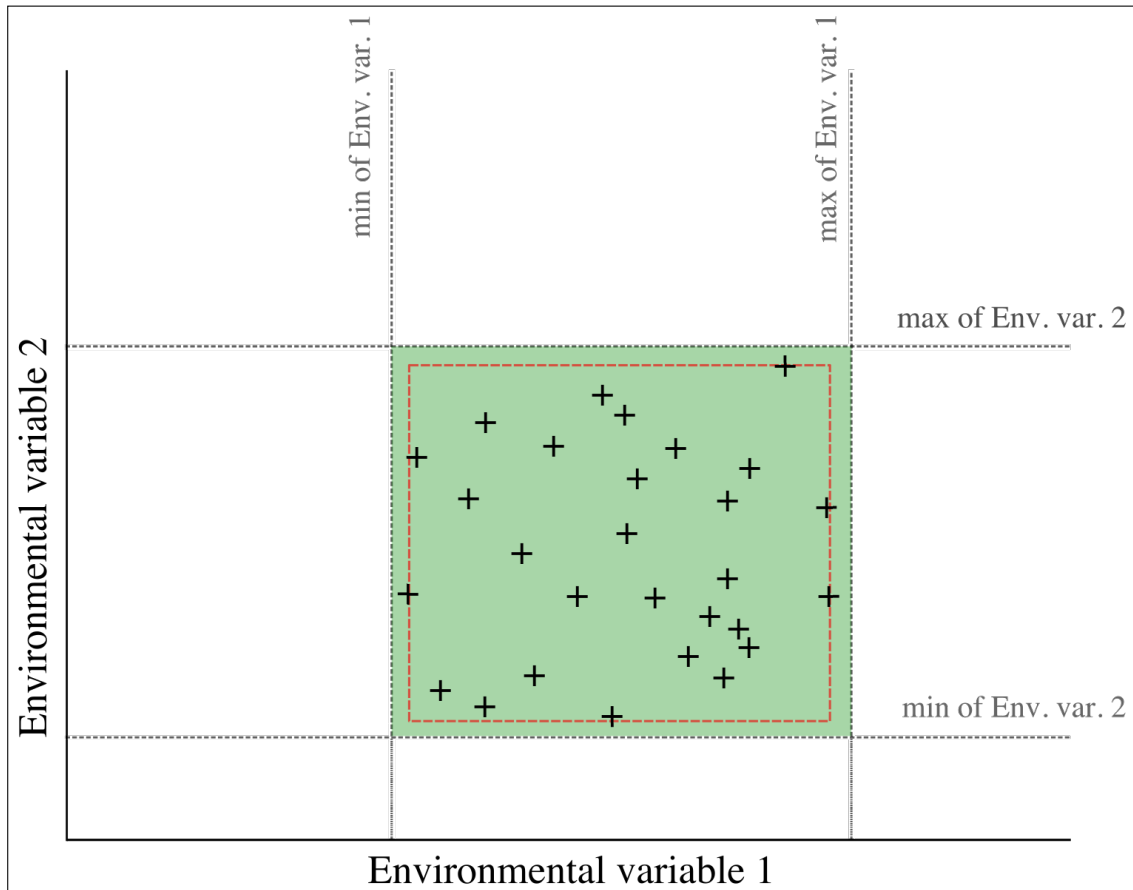
Source: Author's own

Figure 17: Fitting of a Bioclim model using the predictors (green) and the occurrence points (black), reduced to a specified percentile (red) to avoid overpredictive effect of outliers.

ually, as this procedure is taken by Bioclim. The resulting object of class Bioclim (`bc`) is passed to the function `predict`. Further parameters are the environmental variables of the time slice for which a prediction should be made and the extent that should be applied in the resulting raster file. The `evaluate` function provides a cross-validation of the models. Presence points, background points, and the model object are passed a parameter to the function.

2.3.2.2 Regression Methods

There are two approaches that can be used for the methods that are based on regression: the Generalized Regression Models (GLM) and the Generalized Additive Models (GAM). The GLM used in this work is a generalization of the simple linear ordinary least squares

regression models to predict the response variable as a function of multiple predictor variables (HIJMANS & ELITH 2017). Linear regression models assume that a straight line can be used to describe the relationship between the response and predictor variables and a constant change of a predictor leads to a constant change of the response variable. In ecological data, however, this assumption is often violated, which is why these models are extended into the methods called GLMs. Thereafter, the regression models can deal with not normally distributed data (GUISAN et al. 2002).

The GLMs, which find the best performing equation that predicts the possible occurrence of a selected species based on the values of the environmental variables, consist of three important components: the response variable, the combination of continuous predictor variables to represent the environmental suitability and the link function to describe how the mean of the response depends on the linear predictor (Listing 8). The link function provides a transformation of the response, because the relationship between the linear predictors and the response is not linear (HASTIE & TIBSHIRANI 1990).

Source: Author's own

Figure 18: Relationship between predictor data and presence/absence data with a link function for binomial data.
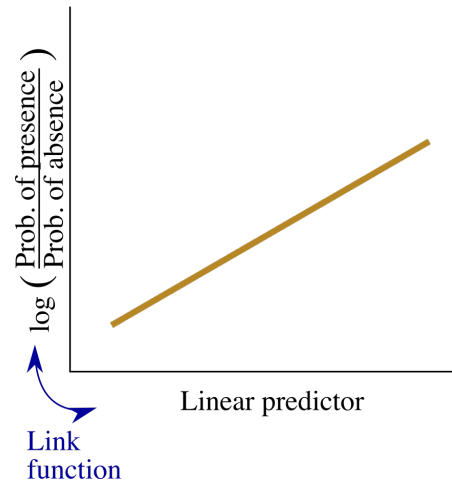
Since binomial GLMs have to get fitted with presence and absence or background data, logistic regression (logit) is used (HIJMANS & ELITH 2017). This is the logarithm of the ratio of probability of presence to the probability of absence (Figure 18). The coefficient can be used to determine the influence of a predictor variable on the potential distribution of a species.

Listing 8: Fitting the environment variables and predict the potential distribution with GLM

```
1  gm1 <- glm(pa ~ bio1 + bio4 + bio5 + bio6 + bio12 + bio15 + bio13
       + bio14 + GebCo_TB_ext + GebCo_TB_ext_slope + GebCo_TB_ext_
       aspect, family = binomial(link = "logit"), data=envtrain)
2  pg <- predict(predictors, gm1, ext=ra_ext)
```

2.3.2.3 Machine Learning Methods

Algorithmic modeling applications are known as machine learning. Their application and development has increased enormously in the past few years (Breiman 2001). MaxEnt (Phillips et al. 2006) is the most widely used method in species distribution modeling (Hijmans & Elith 2017). As a machine learning method MaxEnt iteratively builds multiple models to get predictions or inferences from incomplete information. The aim is to determine a target likelihood suitability by determining the probability distribution of maximum entropy. This includes those distributions that are the most spread out or those who are closest to uniform. The results can only be interpreted as a relative index of environmental suitability rather than predicted probability of occurrence (Phillips et al. 2006). MaxEnt in version 3.4.1 is provided by Phillips et al. (2017) and can be addressed via the *dismo* package (Listing 9).

Listing 9: Fitting the environment variables and predict the suitable environment with MaxEnt

```
1  xm <- maxent(predictors_recent, pres_train, factors=c("iwahashi",
       "KG"), path=".")
2  px <- predict(predictors, xm, ext=ra_ext, progress="")
```

MaxEnt uses only presence data. The algorithm compares the locations where one species exists with all other environments available in the study region. The available environment is defined by sampling a large number of points throughout the study area. In MaxEnt these points are called background points and not pseudo-absence, because they can include locations where the species is known to occur (Phillips et al. 2006).

Phillips et al. (2006) described two important components of the method: a) Entropy, to find the distributions as described above, and b) Constraints, the rules that set the limits to the predicted distributions. They are based on the values of the environmental variables, which in this case are also called features, where the occurrence of a species was observed. By default, MaxEnt allows five types of features, but only two are used in this work (Table 7).

In order to determine the possible distribution of a species, MaxEnt first calculates two probability densities. On the one hand, a probability density is calculated for all presence locations in order to determine the relative likelihood of all features over the range of those points. On the other hand, a probability density is calculated across the entire region. These values are based on the background points.

Table 7: MaxEnt feature types used for the modeling as described by Phillips et al. (2006)

Feature type	Category	Constraint
Linear	Continuous variable (bio1, bio4, bio5, bio6, bio12, bio13, bio14, bio15, slope, aspect)	The mean of the environmental variables should be close to their observed value in known occurrence locations
Categorical	Categorical variable (iwahashi, KG)	The proportion of predicted values in each category should be close to the observed proportion

Thus, the probability of density of the background points describes the available environment within the study region. The probability of density of the presence points describes the respective environment where a species occurs. Subsequently, MaxEnt determines the ratio between these two probability densities. This indicates the relative environmental suitability for the occurrence of a species for each point in the study area. To ward off overfitting, MaxEnt uses the strategy of relaxing the constraints. The model does not use the exact constraints of the environmental variables, but rather intervals around the constraints (Phillips et al. 2006).

2.3.3 Evaluating the Model

In order to obtain a good prediction from the models, one can analyze whether the result makes sense from an ecological point of view. On the other hand, a value can be determined via a statistical analysis. For this purpose, Hijmans & Elith (2017) proposed to generate two further datasets from the cleaned finding spots: a training and a test dataset (Listing 5).

The most common statistic obtained by the evaluation are plots of the Receiver Operating Characteristics (ROC). These characteristics are derived from the *confusion matrix*. It refers to a matrix whose rows summarize the predicted presence and absence, and the columns state the true observed presence and absence status. The respective values in this matrix are specific to individual cells of the model. Cells of an observed distribution where the model predicts absence are called *omission error*. Cells in which the species is unknown but predicted as present, indicate the *commission error*. These values are used to express the sensitivity and specificity of a model, where sensitivity is indicated by *1*

– (omission error rate) and specificity equals *1 – (commission error rate)*. The ROC plots use these characteristics to allow an evaluation of the model. Here, the sensitivity on the *y* axis is plotted against *1 – specificity* on the *x* axis. The two resulting rates are plotted over the series of all predictive values as single points which are connected to form a curve (PETERSON et al. 2011).

The resulting *Area Under the Receiver Operator Curve* (AUROC), or commonly abbreviated as AUC, is a measure of rank-correlation to represent the performance of the model. The AUC values range between zero and one and indicate the ability to rank presence locations higher than a sample of random background data or (pseudo-)absence data (PETERSON et al. 2011). It should be noted that AUCs are generally not comparable among species, because different species potential distributions will cover different areas of a study region. But according to PHILLIPS et al. (2006), it is possible to use this evaluation method in most situations for a model comparison with true presence/absence data.

An AUC value in presence/absence data of 0.5 or below indicates an untrustworthy prediction (HIJMANS & ELITH 2017). AUC values above 0.5 are ranked by SWETS (1988) into poor predictions (0.5 to 0.7), reasonable predictions (0.7 to 0.9) and very good predictions (above 0.9). But these guidelines are not always relevant for ROC plots which are based on presence/background data. The maximum achievable AUC value in this type of data depends on the occurrence of the species and its potential distribution within the study region. This means presence/background AUCs are species- and region specific, but can be used in comparison of performance between different environmental datasets, as it is the case in this work (PETERSON et al. 2011).

Furthermore, the AUC can also be used in the evaluation of training and test data. This will always be in reference to the recent environment and finding spots of the species within a specific modeling method. After that, according to HIJMANS & ELITH (2017), a comparison of the AUC of the test and training data for fitting the model can be performed. The basis is an approach by ELITH et al. (2006) which connects the AUC with the correlation (COR). The correlation between test and training data in the model fitting is calculated as the Pearson correlation coefficient and provided by the function `evaluate` of the package *dismo*. Since a comparison may only be applied between species with removed spatial sorting bias (HIJMANS 2012), the results from *Rangifer tarandus* and *Rupicapra pyrenaica* are excluded for determining the potentially best performing method.

2.3.4 Selecting a Threshold

Another important setting is the threshold. All modeling methods used in this work produce a continuous surface output as a result. PETERSON et al. (2011) suggest to convert continuous model predictions to binary ones. This can be resolved by choosing a threshold value at or above the environment that is predicted as suitable for the selected species. Therefore, the setting of the threshold is determined mainly by the reason of applying the model. The simplest, but also subjective, method is to select a random value, such as 0.5 for a continuous output that ranges from zero to one. In addition, this value cannot be related to the species' ecological requirements and lead to misinterpretations. The more objective methods applied to each continuous output algorithm are often based on criteria applied during model calibration. It must be distinguished whether it is presence/absence or presence-only data that is imported into the model. In the former case, the threshold of *maximizing the sum of sensitivity and specificity* has proved to be the best performing one (LIU et al. 2005). Also for models with presence-only data, this value can be used when working with randomly generated background values if needed (LIU et al. 2016). The threshold is generated in the evaluation object and can be addressed with the function `threshold` and option *spec_sens* in the package *dismo*.

3 Results

3.1 Maps

One of the last points from Figure 16 is the mapping of the previously threshold defined predictions to the geographic space. Geographic space is considered as the study area of the Iberian Peninsula and Morocco. The maps were created with QGIS by importing the raster files generated from the individual methods. Subsequently, the presence and absence cells are differentiated by colors. For each time slice, the results of the individual methods are displayed side by side. Not all 72 distribution maps are depicted in this work, but all prepared maps and the original exported results can be downloaded as supplementary data from the CRC 806 Database (Holthausen 2018). The reference to the respective supplementary data is noted at the first mention of the species (A-H). All maps and results represent the potential distribution or suitable area of the species predicted by the respective methods, based on the previously described settings and environmental variables.

Rangifer tarandus (A1-3, Figure 19)

The predictrions for *Rangifer tarandus* vary greatly between the individual methods. Bioclim shows a dispersion mainly in higher regions of the Iberian System and Cantabrian Range during H5 and only very minor changes in comparison to GI11 and H4. An occurrence in Morocco is not specified and the predicted regions for distribution do not change over time. The predictions of GLM also differ only very marginally over the three time slices. A distribution deviating from Bioclim is given for central regions of the Iberian Peninsula, in particular the Inner Plateau and the Pyrenees show a broad presence. In Morocco, distribution is indicated within central regions of the High and Middle Atlas. MaxEnt does not provide suitable areas for the regions in the Iberian Peninsula and Morocco based on the imported predictors and selected threshold for all time slices.
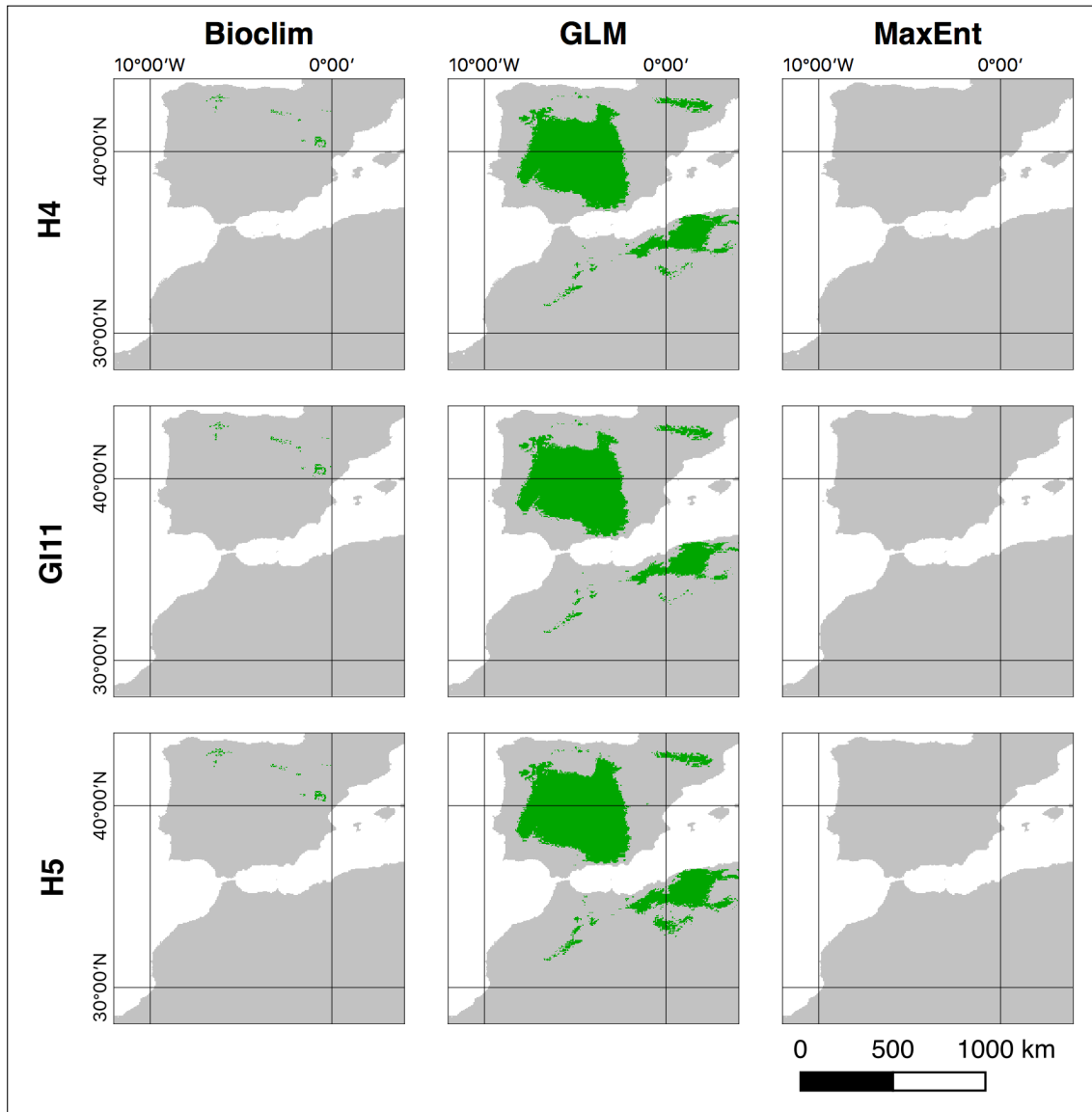
Figure 19: All result maps to *Rangifer tarandus*. Green areas indicate a presence or suitable area and gray indicates a predicted absence of the species.

Figure 19 shows very different predictions between the methods, but similar distributions across the time slices. In the result of MaxEnt, it should be noted that there were no reference points for the study region in the test and training data available and the display is dependent on the selected threshold. The raw values (HOLTHAUSEN 2018) show across all time slices a very small potential suitable area in the region of the

Pyrenees, similar to GLM. Nevertheless, the AUC values of these predictions (Table 8) are similar high in all predictions, as no sampling has been performed here.

Bos taurus (B1-3)

There are a few more similarities among the methods for *Bos taurus*. Bioclim predicts a dispersal during the H5 only along the foots of the Pyrenees. There are only very few changes to GI11, as well as between GI11 and H4. The predicted regions remain unchanged over all time slices. There is never any mention of a distribution in Morocco. GLM deviates from Bioclim predictions and indicates a distribution only in a small region in the northwest of the Iberian Peninsula. Over the three time slices there are only very minor changes, whereby the regions of predicted distribution remain the same. A possible distribution in Morocco is also stated here at no time. MaxEnt shows some overlaps with the Bioclim results, especially in the Pyrenees and at the border of today's Spain and France. Additional suitable regions are indicated at the coast to the Mediterranean Sea in some southeastern regions and the complete coast to the Atlantic Ocean in the north. In Morocco, some suitable areas are specified within the Atlas range. Between H5 and GI11 there are only minor changes. Whereas at H4 almost the entire northern half of the Iberian Peninsula is stated to be suitable. In Morocco suitable areas accrued especially from the coast of the Mediterranean Sea to the ranges of the Middle Atlas.

Capra pyrenaica (C1-3, Figure 20)

The distribution of *Capra pyrenaica* during H5 is indicated by Bioclim with the northwestern Inner Plateau and the adjoining Cantabrian Range as well as in the southeastern Iberian Peninsula from the tails of the Pyrenees across the Iberian System to the Baetic System. In Morocco, the region of the Middle Atlas in particular is characterized as presence. With the transition to GI11 only small changes become visible. Also for H5 only minor changes in the dispersal of the projected regions are recognizable. The predicted distribution rarely changes over the time. GLM, however, predicts in H5 a distribution mainly in the extreme northwest and southwest, as well as in the area of the Strait of Gibraltar. Only a few occurrences are forecasted in the area of the Central System and the Iberian System. Compared to GI11, there are only very few changes and the projected distribution remains the same. The locally predicted distribution is slightly reduced, but there are no changes in the extent. In transition to H4, the locally predicted density of the distribution decreases slightly, but there are no changes in the

extent. MaxEnt predicts suitability for the southeastern regions of the Iberian Penin-
sula, especially in the mountainous regions of the Baetic and Iberian system as well as
the Cantabrian Range in the north. Only a few regions are indicated as suitable in the
Pyrenees. In Morocco, distribution is limited to the regions of the High Atlas, Anti-
Atlas, and Middle Atlas. There is minor change in distribution across all time slices.
The extent predicted as suitable does not change.

Source: Author's own
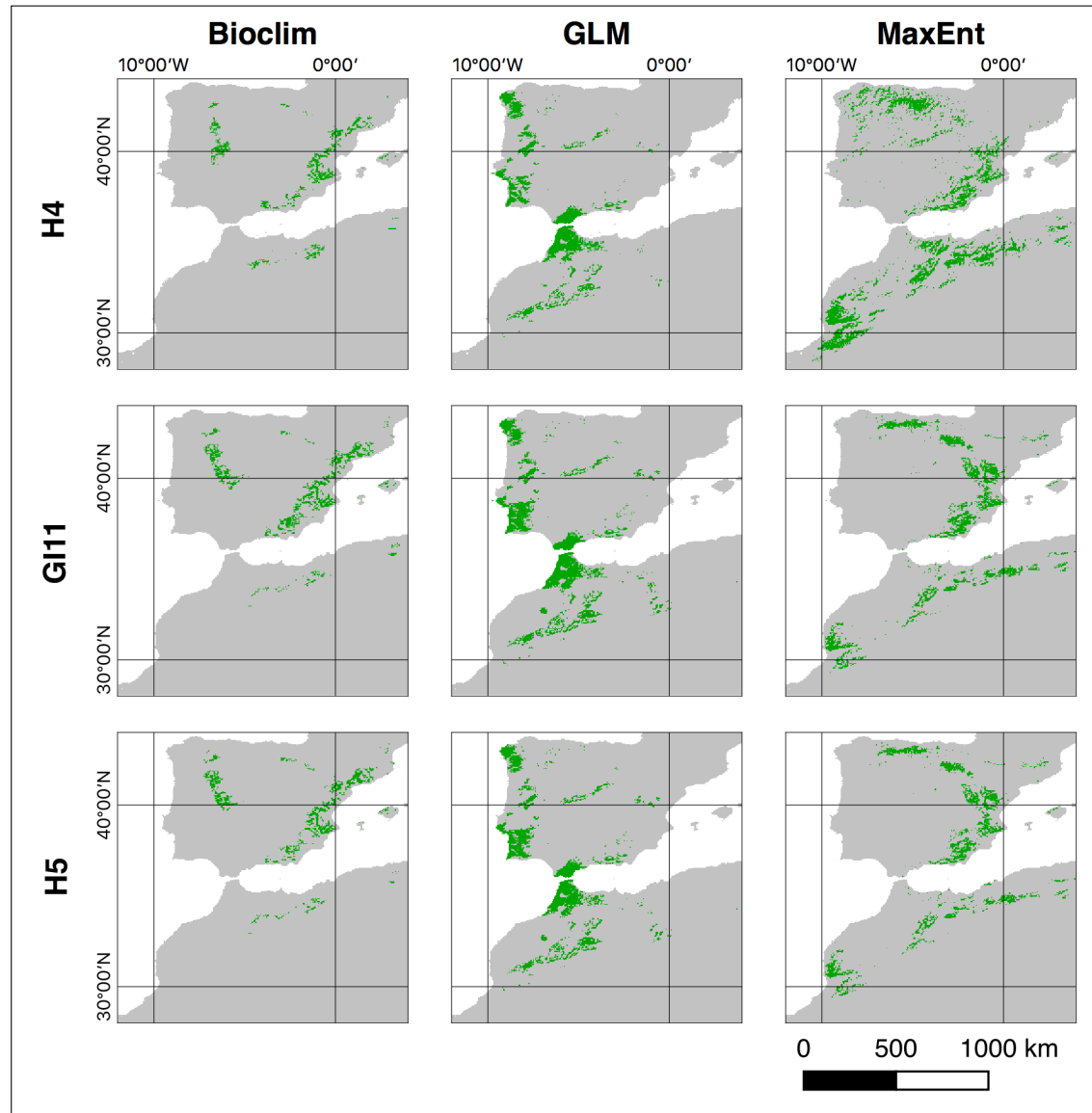
Figure 20: All result maps to *Capra pyrenaica*. Green areas indicate a presence or suitable area
and gray indicates a predicted absence of the species.

The few overlaps of completely different predicted regions can be seen in Figure 20. The AUC values (Table 8) are in a similar value range for MaxEnt and GLM, only Bioclim is in the range below 0.5, although the results show the greatest similarities with MaxEnt.

Rupicapra pyrenaica (D1-3)

As a possible distribution of *Rupicapra pyrenaica* Bioclim specifies the Cantabrian Range in the north of the Iberian Peninsula and the Pyrenees. Excluded are the high-altitude regions of the Pyrenees. Seen over all three time slices, there are only very minor changes in the prediction, the northern mountain regions represent the favor zone. In the change of time slices, differences in the distribution region occur only very occasionally. In Morocco no distribution is indicated. However, GLM predicts the high-altitude mountain areas of the Pyrenees that were left out at Bioclim. There is an overlap in the eastern part of the Cantabrian Range, and in addition the most northwesterly region of the Iberian Peninsula appears to be a potential habitat in H5. In Morocco, only a few regions in the northern Rif and the eastern tails of High Atlas are specified for distribution. In transition to GI11 and H4, there are no significant changes in the distribution. In H5, the prediction from MaxEnt to Bioclim differs only slightly, with no suitable areas for the region between the Cantabrian Range and the Pyrenees. No major change in the suitable regions can be seen in the transition to the time slices GI11 and H5. Like Bioclim, MaxEnt predicts no suitable regions in Morocco.

Capreolus capreolus (E1-3)

Capreolus capreolus has a distribution predicted by Bioclim in the northern regions of the Iberian Peninsula. In particular, these include the Cantabrian Range, the Iberian System and the Pyrenees, except the high-altitude regions. For H5, no presence is predicted in Morocco. In transition to GI11, there are no significant changes in the distribution. During H4 the distribution in the south is reduced and only the northerly and less high-altitude areas of the Cantabrian Range and the Pyrenees are included in the distribution area. An overlap of the distribution from GLM to Bioclim exists only in the northwest and a small region in the Pyrenees. In addition, GLM specifies for H5 and GI11 a potential occurrence in the south of the Iberian Peninsula and in northern Morocco at the Strait of Gibraltar. This distribution prediction also applies to the H4, where the density decreases in the western regions and decreases slightly in the Pyrenees.

MaxEnt has high compliance with Bioclim for suitability in H5, although the potential habitats are less dense and focus on coastal regions in the north on the Atlantic Ocean and northeast on the coast of the Mediterranean See. There are no major differences between H5 and GI11, but H4 has increased suitability in the south, especially along the coasts to the Atlantic Ocean and the Mediterranean Sea.

Cervus elaphus (F1-3)

Bioclim predicts a distribution of *Cervus elaphus* in H5 across almost the entire Iberian Peninsula. Excluded are large areas in the southwest and northwest of the peninsula, as well as the Pyrenees. In Morocco, regions in the southwestern Anti-Atlas, the northeastern Middle Atlas and the Rif Range are indicated as presence in the distribution. With transition to GI11, the distribution regions are changing as they shrink slightly and open up some gaps in the whole area. Whereas in H4, the distribution regions expand again and close the gaps. Changes in distribution within Morocco are minimal. GLM, on the other hand, predicts distribution, especially in the northwest of the Iberian Peninsula. In addition, some areas in the Cantabrian Range, the Central System and the Northern Iberian System are identified as potential habitats. In Morocco, *Cervus elaphus* is characterized as being present during H5 only in northern Rif Range. There are very few changes in the distribution at transition to GI11, both in the Iberian Peninsula and in Morocco. A similar situation occurs in transition to H4, except that the number of potential habitats decreases and the size of the predicted region shrinks. The result from MaxEnt has a greater similarity with GLM to the possible distribution area compared with Bioclim. In H5, a possible distribution is predicted in the Cantabrian Range and the entire northern coastal strip. In addition, a distribution is indicated along the foothills of the Pyrenees to the Mediterranean Sea. From there, the potential distribution area extends to the present Costa Blanca in the south. Furthermore, there are possible regions of *Cervus elaphus* on the coastal strip in the northwestern part of the Iberian Peninsula. In Morocco, the distribution is limited mainly to the foothills of the High Atlas to the coast of the Atlantic Ocean. In GI11, the potential distribution expands into the interior, especially in the north of the Iberian Peninsula. Here, the potential distribution area extends into the Inner Plateau north of the Central System. In H4 an expansion takes place into the south again. The entire coast from today's Lisbon to the north and the entire coast of the Atlantic Ocean in the north of the Iberian Peninsula are considered as possible habitats. Inland, the distribution is predicted from the coast of the Atlantic Ocean in the north to the Central System and the Pyrenees.

Furthermore, the coastal strips to the Mediterranean Sea have to be considered for presence. In Morocco, the coastal strip to the Mediterranean Sea ranks among the suitable area in this time slice.

Sus scrofa (G1-3)

The potential distribution of *Sus scrofa* is specified by Bioclim similar to that of *Cervus elaphus*. In H5, the entire Iberian Peninsula is the preferred area, except for regions in the northwest, southwest, and the Pyrenees. For GI11 only minor changes are visible. In H4 the number of potential habitats is decreasing and less dense, especially in the west of the Iberian Peninsula and Morocco. While until GI11 large areas of the Atlas Ranges could be ranked among the potential distribution area, in H4 the area is restricted to the Hautes Plaines in the northeast, the southeastern foothills of the High Atlas and the transition area between Middle Atlas and the Rif Range. In H5, GLM predicts a dispersal especially in the entire west of the Iberian Peninsula, the Central System and the Inner Plateau, as well as in the Pyrenees. In the south, the Strait of Gibraltar is reached. This continues in Morocco to the Middle Atlas in the south and the Rif Range in the east. In transition to GI11, only very minor changes can be seen in the distribution. In H4, the distribution area is decreasing slightly and more concentrated towards coastal areas. These changes can be observed both in the Iberian Peninsula and in Morocco. MaxEnt has a different distribution of suitable areas in H5. In particular, the northwestern part of the Iberian Peninsula, the coast of the Atlantic Ocean in the north, and several regions in the Cantabrian Range, the Iberian System, and the Pyrenees can be seen as suitable. In Morocco, the area of the coast of the Mediterranean Sea and the coast of the Atlantic Ocean can be ranked among suitable areas. Particularly mountainous regions are not included into the suitable areas. In the change to GI11 no major changes are registered. In H4 this changes a bit. The area of the Cantabrian Range is now a closed suitable area with no gaps. From the west, some overlapping suitable areas are reaching into the interior of the Iberian Peninsula. The high-altitude regions are still not included. Morocco also has a more branched distribution of the potential suitable areas, especially in the north. However, the coastal areas are still among the suitable areas.

Equus caballus (H1-3, Figure 21)

Equus caballus is predicted by Bioclim in H5 throughout the mountainous area of the Iberian Peninsula. Excluded are the high-altitude regions of the Pyrenees and the north-

ern part of the Inner Plateau. In the north and southeast, however, to the coasts of the Atlantic Ocean and the Mediterranean Sea have to be considered as possible distribution areas. A similar picture appears in Morocco, where the distribution covers only the regions of the High Atlas via the Middle Atlas up to the Rif Range. With transition to GI11, there are only minor changes, which are expressed in a small increase in the distribution to the southwest of the Iberian Peninsula. In H4, however, this expansion is undone and the distribution is predicted very similar to H5.

The prediction in GLM for H5 differs significantly to Bioclim. Two regions are identified as potential habitats. These include the plain regions in the northwest of the Iberian Peninsula and the regions along the Strait of Gibraltar, both on the Iberian side and in Morocco. For GI11, the prediction of distribution is the same as for H4. Within the potential regions, individual units shift, but the regions themselves remain the same in extent and location. MaxEnt contradicts these calculations and specifies for H5 a location of suitable areas comparable to Bioclim. In addition, however, the coastal regions in the northwest and southeast of the Iberian Peninsula are classified as suitable areas. Morocco reveals a similar picture. In addition to the coast of the Atlantic Ocean, the regions of the Atlas to the Strait of Gibraltar belong to the suitable areas. The only change in GI11 is that more parts of the Inner Plateau are designated as a suitable area. This decreases with H4 again, but the suitable area intensifies in the northern part of the Inner Plateau. The suitable areas in Morocco show only very minor change.

As shown in Figure 21, there is a very uneven prediction, including in the range of AUC values (Table 8). GLM has the highest value, but deviates significantly from the other two methods, which show more similarities among each other in the potential distribution.

3.2 Statistical Evaluation of the Methods

Table 8 shows the statistical evaluations of the methods for all species. The AUC values from the AUROC plots between the training and the test data are listed. All data are based on the sampled input data, only *Rangifer tarandus* and *Rupicapra pyrenaica* were imported without any sampling. Furthermore, the correlation coefficient (COR) from the evaluation function is indicated. Herewith, according to Elith et al. (2006), the Figure 22 can be generated from the mean values in order to visualize the meaningfulness of the methods. In Figure 22 the lower-left portion indicates the poorly performing

44
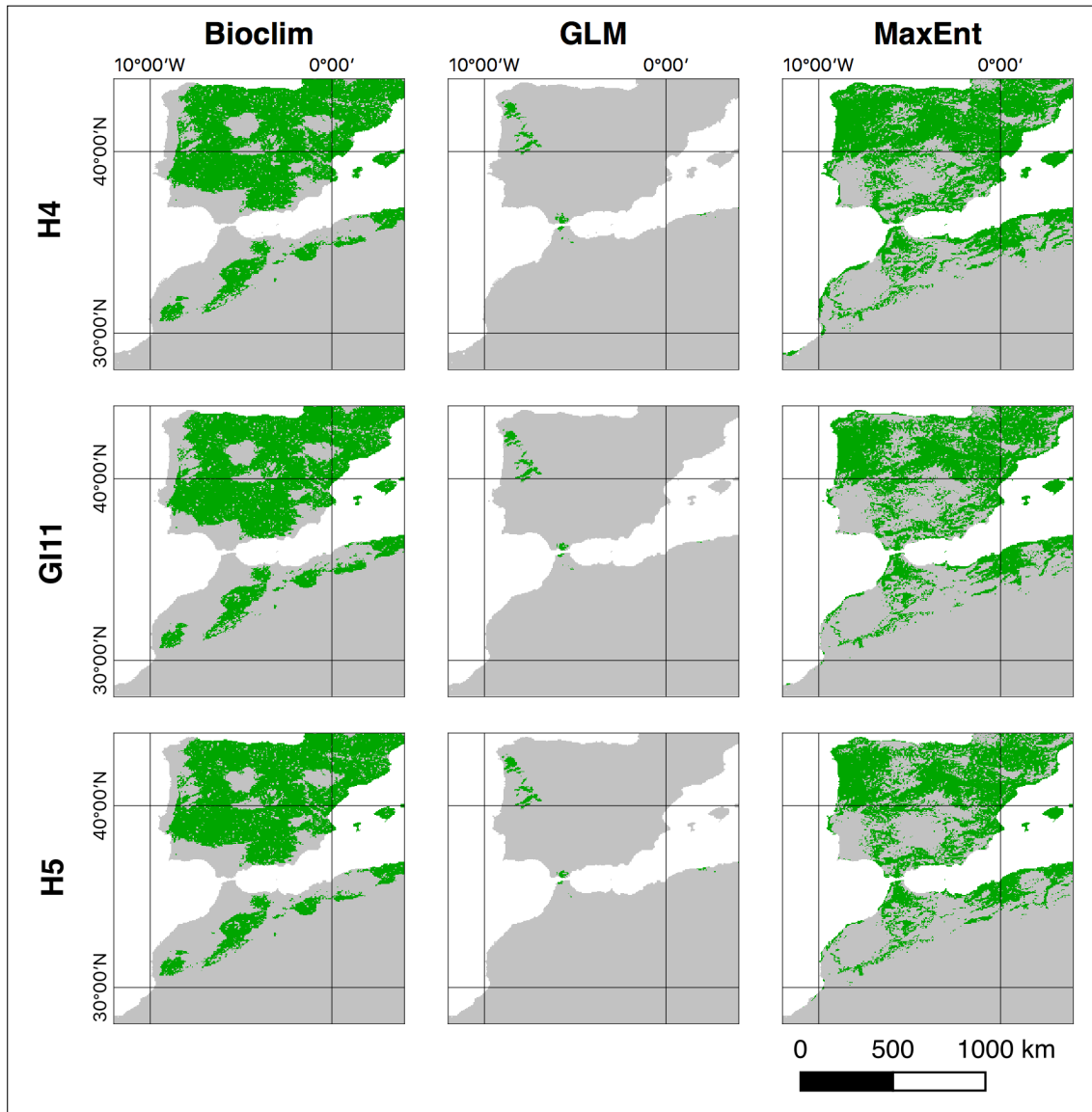
Figure 21: All result maps to *Equus caballus*. Green areas indicate a presence or suitable area and gray indicates a predicted absence of the species.

methods and in the upper-right portion the better performing methods. The mean values used here were obtained from Table 8 and contain all COR and AUC values across all species and all three time slices, grouped by method. As already resumed in ELITH et al. (2006), it can be concluded from these data (Table 8; Figure 22) that the best performing method is MaxEnt which can additionally provide more in-depth

Table 8: Statistical results of the predictions for each method and species, expressed as COR and AUC (*The means are calculated without *Rangifer tarandus* and *Rupicapra pyrenaica*).

Species	Bioclim		GLM		MaxEnt	
	COR	AUC	COR	AUC	COR	AUC
Rangifer tarandus*	0.671	0.978	0.723	0.997	0.967	0.997
Bos taurus	0.200	0.548	0.460	0.776	0.223	0.604
Capra pyrenaica	-0.144	0.444	0.294	0.778	0.478	0.700
Rupicapra pyrenaica*	0.608	0.990	0.818	0.995	0.944	0.999
Capreolus capreolus	0.123	0.628	0.227	0.633	0.249	0.628
Cervus elaphus	0.036	0.540	0.239	0.640	0.279	0.659
Sus scrofa	0.133	0.607	0.281	0.653	0.271	0.682
Equus caballus	0.346	0.723	0.325	0.636	0.527	0.826
Mean	*0.116*	*0.582*	*0.304*	*0.686*	*0.338*	*0.683*

answers referring to the prediction of a distribution. From a statistical point of view, GLM is located not very distant from the MaxEnt value. Bioclim, statistically, provides the worst performing results.

The exclusion of *Rangifer tarandus* and *Rupicapra pyrenaica* does not mean that they provide false predictions. Just the inclusion of these AUC values into the determination of the potentially best method is not permitted, as described in Chapter 2.3.3. All other species used had an SSB which was removed by the pairwise distance sampling. Only two of the AUC values from MaxEnt are higher than GLM, but all are higher than the Bioclim values. And only one AUC from Bioclim is higher than the respective one at GLM. Overall, only four of the results reach the AUC value of 0.7 or above, which would pro-
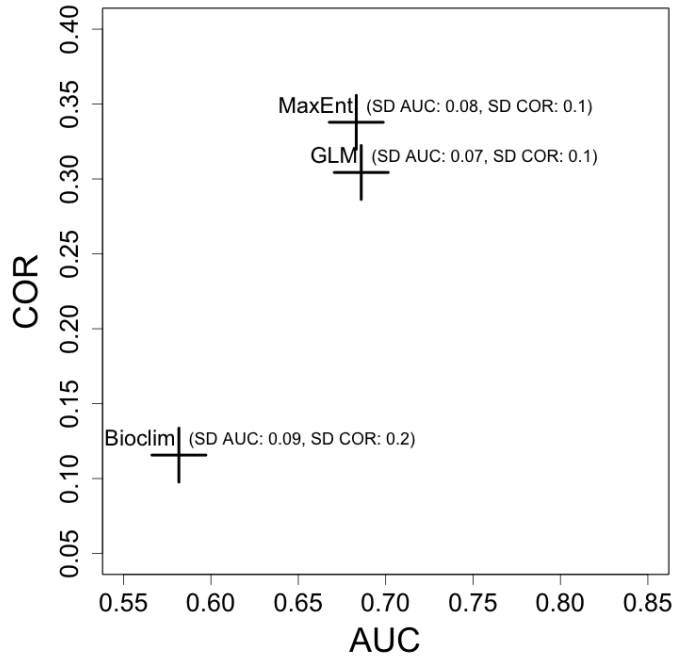
Figure 22: Mean AUC plotted against mean correlation (COR) for all applied modelling methods, summarized across all species (SD = standard deviation).

3 Results

vide a reasonable prediction on
presence/absence data: *Equus caballus* in Bioclim, *Bos tarus* in GLM, *Capra pyrenaica*
in GLM and MaxEnt.

Table 9 gives, in particular for the method MaxEnt, estimates of relative contribution of
selected environmental variables which have been included in the prediction. It should
be noted that the information has to be interpreted with caution, since some of the
predictor variables are correlated.

The table contains entries of a contribution above 10%, the next value below this limit
and any values that can indicate a change between the time slices. It can be seen that the
environmental variable for annual precipitation (bio12) is the dominant environmental
variable for three species (*Cervus elaphus*, *Equus caballus*, and *Sus scrofa*) in all three
time slices. In general, the contributions of precipitation (bio12) or its derivatives (bio13,
bio14, bio15) are always represented by more than 10%. The annual average temperature
(bio1) and its derivatives (bio4, bio5, bio6) also contribute to the calculations, but not
to the same range as the precipitation. The precipitation seasonality (bio15) plays no
role with values always below 10%. The calculations for *Rangifer tarandus* are the
only ones that relate 47% to the categorical variable Köppen-Geiger, which describes
the climate in general. While the DEM only affects *Rupicapra pyrenaica* over all three
time slices (42.7%), in the case of *Capra pyrenaica* in H5 and GI11 the slope is the
highest determining environmental variable. But in H4 the highest contribution changes
to the DEM (28.8%) and slope contributes only marginally with 0.4%. *Sus Scrofa* was
characterized by Melis et al. (2006) with a high correlation to temperature, whereas
MaxEnt shows a relatively large contribution of 60-80% of the annual precipitation
(bio12) to the distribution. The temperature (bio1) contributes only slightly to the
distribution predicted by MaxEnt. The maximum temperature of warmest month (bio5)
is the second largest contribution after the precipitation in H5 and GI11. In H4, the
percentage even drops below 1%. Terrain surface classification and the aspect contribute
only in a minor role.

Table 9: The contribution of the predictors as calculated by MaxEnt per species and time slice in percent (TSC = Terrain surface classification, KG = Köppen-Geiger classification), contributions below 10% are colored in gray

Species	Time slice	bio1	bio4	bio5	bio6	bio12	bio13	bio14	bio15	DEM	Slope	Aspect	TSC	KG
Rangifer tarandus	H4	21	2.4	0.2	6.3	1.5	0	13	5.2	2.7	0.1	0.5	0	47
	GI1	21	2.4	0.2	6.3	1.5	0	13	5.2	2.7	0.1	0.5	0	47
	H5	21	2.4	0.2	6.3	1.5	0	13	5.2	2.7	0.1	0.5	0	47
Bos taurus	H4	3.3	4.9	0.5	0.8	0.2	0.1	67.1	0.7	12.1	2.9	1.2	3.3	3
	GI1	1.7	4.6	1.3	0.9	0.2	0	65.7	1.2	11	2.8	4.1	3.4	3.1
	H5	1.7	4.6	1.3	0.9	0.2	0	65.7	1.2	11	2.8	4.1	3.4	3.1
Capra pyrenaica	H4	0	10	3.3	5.1	19.6	0.8	0.8	3	28.8	0.4	7.7	10.2	10.5
	GI1	0	5.2	2.4	3.8	20.9	0.5	0.2	3.9	18.4	26.6	2	3.9	12.2
	H5	0	5.2	2.4	3.8	20.9	0.5	0.2	3.9	18.4	26.6	2	3.9	12.2
Rupicapra pyrenaica	H4	0.1	15.5	0	3.8	0.3	1	25.9	1	42.7	4.3	0.1	4	1.3
	GI1	0.1	15.5	0	3.8	0.3	1	25.9	1	42.7	4.3	0.1	4	1.3
	H5	0.1	15.5	0	3.8	0.3	1	25.9	1	42.7	4.3	0.1	4	1.3
Capreolus capreolus	H4	0.3	2.7	4.1	0.2	4.3	0	75.4	2.5	2.2	0.4	1.4	4.4	2.2
	GI1	0	3	1.9	0.1	3.4	0	78	1.5	2.9	0.2	1.9	4.3	2.9
	H5	0	3	1.9	0.1	3.4	0	78	1.5	2.9	0.2	1.9	4.3	2.9
Cervus elaphus	H4	1.3	6.5	3.6	0.7	41.4	0.5	29.6	3.6	4.1	1.5	2.4	2.8	2
	GI1	3.9	6.4	6.4	0.6	46.5	0.4	24.9	3.8	1.3	2	0.4	1.3	2
	H5	0.6	7.8	2.1	0.9	49.9	0.1	27	3.7	1.4	0.5	0.6	1.6	4
Sus scrofa	H4	0.6	4	0.8	1.4	60.6	14.4	4.9	4.1	3.6	1.1	1.3	1.2	2
	GI1	1	2.9	3.3	0.9	80.8	0	1.5	2.2	1	0.3	1.6	1.6	2.8
	H5	1	2.9	3.3	0.9	80.8	0	1.5	2.2	1	0.3	1.6	1.6	2.8
Equus caballus	H4	0	8.6	1.2	1.1	34.2	0.5	0.2	0.7	12.4	5.9	5.7	18.4	11.1
	GI1	0	10.4	2.1	1	34.8	0.4	0.6	0.9	9.8	11.1	6	13.7	9.3
	H5	0	10.4	2.1	1	34.8	0.4	0.6	0.9	9.8	11.1	6	13.7	9.3

4 Discussion

Because of these values, it is difficult to decide which method represents the potentially best one. Especially the two high AUC values for *Capra pyrenaica* in MaxEnt and a slightly higher one in GLM but the completely different predicted regions for distribution or suitability do not support the selection.

A meaningful prediction depends primarily on the quality and quantity of the occurrences and predictor variables used in the algorithm. Since no verifiable finding spots assignable to the individual time slices were available for this work, the datasets of the GBIF Secretariat (2017) were used. In general, only those points were used in which a position was specified and categorized as living specimen and human observation. Two potential sources of error emerged: the human observation datasets still contain occasional reports of fossils, although there is a separate category for such reports. In addition, the coordinates stored in the database have an uncertainty, which could lead to errors in the prediction at a finer resolution.

To generate consolidated datasets for the import, the extension of the training area and the occurrence records to almost the complete distribution area of the species, as applied here, is recommended and often used to enhance the reliability of the prediction. The restriction to just one part of the distribution area may lead to a very limited survey of the environmental characteristics and thus to an underestimation of the potential distribution (THUILLER et al. 2004). SÁNCHEZ-FERNÁNDEZ et al. (2011) even suggest to include the world-wide distribution of a species into the training, or at least resorting to a sample of populations that are distributed across the whole environmental range. The region extended here includes almost the entire European region and thus almost all environmental properties for the species, which are mainly distributed in Europe.

Furthermore, the approach proposed by DEPRAZ et al. (2008), RODRIGUEZ-SANCHEZ & ARROYO (2008), SOLOMON et al. (2008), and ALBA-SÁNCHEZ et al. (2010), and used in this work, is not entirely free of doubts. In these studies, where the distribution of a species was projected onto either the future or past climate, it is reported that the

transferability between periods provided adequate results. On the other hand, VARELA et al. (2009) conducted a time transference of model predictions which reveals almost no suitable conditions for their species, despite a widespread presence of fossil remains in the study area during the related period of time. A similar behavior to MaxEnt for *Rangifer tarandus* in this work, which has been identified as present in the northern parts of the Iberian Peninsula at this time (ÁLVAREZ-LAO & GARCÍA 2010). Additionally, as a result, no evaluation of the temporal transference can be performed, just on the basis of the presence data respectively test and training data.

The use of WorldClim's 19 bioclimatic variables (HIJMANS et al. 2005) is mainly due to its easy availability, making it one of the most commonly used predictors (VARELA et al. 2011). This means that the average annual temperature and the annual precipitation or their derivatives are used in most methods for calculating potential distribution. The resulting multicollinearity can be counteracted by determining the VIF in the predictor variables by removing those variables from the predictor stack as described in Chapter 2.3.1. This purely statistical approach is opposed by the ecological approach, which proposes keeping a predictor variable if it is ecologically important for the distribution of the species, resulting in the selected predictors as listed in Table 6

The methods used are based on different approaches and algorithms. These included a) Bioclim, which uses presence-only data and works with the envelope-style method, b) GLM, which requires presence and absence or background data, and is based on a statistical fitting of the data for a theoretical relationship between the appearance of a species and the environment, and c) MaxEnt which also initially is based on presence-only data but generates a large number of background points by sampling throughout the study area in order to characterize the environment. Its aim is to estimate the species' distribution by maximum entropy and to determine a suitable area. This method is ranked among the machine learning methods.

For this reason, the predictions of the methods are sometimes very differing in the prediction of the potentially distributed areas. The geographical comparison between MaxEnt and Bioclim is more consistent than with GLM. In addition, it should be noted that the modeling is done under the conditions of a climate change. While the training takes place with the recent climate, the prediction is applied on three different past climates, with Bioclim always slightly underestimating the range sizes (HIJMANS & GRAHAM 2006).

The comparison of AUC values is criticized (LOBO et al. 2007; PETERSON et al. 2011; JIMÉNEZ-VALVERDE 2012) because the weight of commission errors is much lower than

the weight of omission errors and therefor the AUC is not seen as an appropriate performance measure. In some cases, the comparison is valid after the removal of a bias in the presence data and the spatial sorting bias (SSB) in the test and training data (HIJMANS & ELITH 2017). Additionally, the correlation coefficient is used (ELITH et al. 2006) in a statistical comparison, as conducted in this work. A bias can emerge from the facts that, on the one hand, not all occurrences are reported in the database. On the other hand, in some regions more occurrences may have been reported due to a more active reporting community than in other regions. In the AUC comparison in this work, however, MaxEnt and GLM are closer together and provide better prediction statistics than Bioclim. This is similar to the findings of ELITH et al. (2006). The findings thus far suggest a weak performance of Bioclim in this case, although it is one of the oldest methods in the species distribution and is still widely used (HIJMANS & ELITH 2017).

Another important point is the selection of a threshold. As a result, the methods provide one value per cell that is within a certain range of values to indicate the probability of distribution of the respective cell. These values are called *raw values* and available in the supplementary data (HOLTHAUSEN 2018). The threshold is needed to determine the presence and absence of a species within the study area. Cells below the threshold are considered as absent while cells above are considered as presence. For the determination of broader potential habitats lower values have to be chosen. For a comparison of methods, a more objective value should be used, which is statistically determined during the evaluation. As seen at *Rangifer tarandus* in MaxEnt, this can lead to cells with a very low probability in the *raw values* not being listed in the end result of presence and absence.

The developed model determines possible prey production for the Western Mediterranean region for H5, GI11, and H4 during the Late Pleistocene on the basis of climate data of recent time and the mentioned time slices. The three selected methods differ significantly in the approaches of their algorithms and as a result in the scope in which they project areas of suitable habitats or potential distribution. Even if, as in this work, the same input data are used, the results of varying methods will always be different. The determination of a best performing model, which should describe precisely the environment of a species and project it into other time slices or regions, depends on the objective and the parameterization. Bioclim is very sensitive to outliers, which generally sets the species' environmental tolerance at the most extreme value in the training dataset and does not allow interactions between the predictor variables. All variables are weighted equally and in the case of climate changes the results are slightly underestimated. In

contrast, the methods of machine learning, such as MaxEnt, can identify the highly complex, nonlinear relationships between a species and the environment, and possibly model more efficiently complex patterns, leading to even more accurate predictions such as GLM (THIBAUD et al. 2014). The statistics obtained in this work reflect exactly this statement, and thus from a statistical point of view, MaxEnt in its default regularization will be the best choice for predicting the distribution for the selected species in Late Pleistocene. With a customized model complexity, it may be possible to achieve even better results in the AUC and counteract an overestimation or underestimation (WARREN & SEIFERT 2011).

The evaluation of the contribution of environmental variables, which is generated by MaxEnt, reveals that first of all precipitation is the decisive factor for the distribution of most species. The next higher determining factor, unlike HOF et al. (2012) specified, is the altitude, provided by a DEM. Temperature and terrain surface classification as well as Köppen-Geiger classification have only a moderate contribution to the distribution.

The examined time slices are in the range of the appearance of the anatomically modern human (*Homo sapiens*) in the Iberian Peninsula and the demise of the Neanderthals. This event took place in the transition between Mousterian in the Middle Paleolithic and Aurignacian in the Upper Paleolithic (SEPULCHRE et al. 2007; HOFFECKER 2009). The climate at this time is characterized by abrupt changes with alternating periods between cold and warm and seen as the LGM, especially in the north of the Iberian Peninsula, earlier than in northern Europe (MORENO et al. 2010).

Up to H4, Neanderthals lived in the Cantabrian Range, followed by either an isolation period or a migration into the south (SERRANO et al. 2015). This led to a disappearance from the Cantabrian Range of Neanderthals analogous to environmental changes from closed landscapes to open landscapes and a rapid replacement of the Neanderthals by the anatomically modern human in the north of the Iberian Peninsula (GARCIA GARRIGA et al. 2012; MAROTO et al. 2012). Changes in the ecosystems, which included the Neanderthals, resulted in a lack of resources that culminated in their decline and partially analog to the extinction of associated fauna. These include in particular *Rangifer tarandus* and *Capreolus capreolus*, which are not specified as equally distributed in all applied methods, but have largely predicted potential habitat in the northern regions of the Iberian peninsula (FINLAYSON 2004; STEWART 2005; FINLAYSON et al. 2008). Thus, the anatomically modern human was able to seek refuge during the MIS3 in the northern Iberian Peninsula (SERRANO et al. 2015) and the presence of preferred prey species favored this to a special degree.

It turns out, that a niche can be projected into a geographic space to specify the maximum spatial extent of a species' geographic range because a species cannot live outside of the dimensions of its identified niche. Thus, SDM will always involve ENM. The choice of method and quality should always be dependent on the objective of the modeling, such as the designation of protected areas or the prediction of distributions in the past and the future. SDMs can be very useful for quick and empirical predictions about the spatial arrangement of a species. But an association of environmental suitability predicted from the model with local population density has only a poor relationship (TÔRRES et al. 2012).

It is important to keep in mind that a model is based on a simulated, mathematical simplification of the data, and will never be entirely show the whole truth, but at least a fair enough approximation to gain additional insight about the projected time. Any produced map that can accurately display the results should always be regarded with some skepticism about the predicted distribution. It should be noted that any prediction that is provided by a modeling without information on true absences, is only a hypothesis about similar environments to those of the species being studied, and that these environments are likely to be somewhere between the extremes of the existing fundamental niche and the occupied niche.

5 Conclusion and Outlook

By dint of the *R* package *dismo* (HIJMANS et al. 2017) a potential distribution was modeled for several cold-adapted prey species of Late Pleistocene in the region of the Western Mediterranean. Particular attention was set on the periods of H5, GI11, and H4. Towards the end of this period, the demise of Neanderthals and the spread of the anatomically modern human, *Homo sapiens*, is observed in the Iberian Peninsula.

Three methods regarding distribution modeling were aplied to model and compare the results. After evaluating the literature, the statistical analysis, and a visual assessment of the resulting maps, the method of MaxEnt (PHILLIPS et al. 2006) and the provided software of the same name (PHILLIPS et al. 2017) is the most worthwhile solution. MaxEnt is a modern method, which can be configured by setting parameters and easily addressed using the package *dismo*.

Although Bioclim (NIX 1986) is a very simple and intuitive method, it uses only presence-only data. However, the method is very susceptible to overprediction and offers no option for interactions between the environmental predictor variables. Furthermore, categorical variables can not be applied for calculations. Bioclim's primary focus is on modeling the distribution of a species based on climatic variables and thus assumes that the distribution of a species is influenced by the climate. However, additional continuous variables can be used in the algorithm.

The method of GLM, however, accepts categorical variables and is characterized by a low susceptibility to overfitting. However, there is a certain sensitivity to outliers and the environmental variables should not possess a high correlation, as this can lead to false estimations of the potential distribution. In addition, true absence data of the species or background points are needed for the prediction.

MaxEnt (PHILLIPS et al. 2006), a method from the subject of machine learning, requires presence-only data for the prediction. Furthermore, categorical variables are accepted, the interaction between predictor variables is possible and algorithms are used to prevent

overfitting. It should be noted that the result is an indication of environmental suitability and therefore does not equate to a predicted occurrence of other methods.

Based on the results of all the methods applied, a distribution of the selected prey species in the Western Mediterranean region is observed throughout all time slices. However, despite proven climate variability, the distribution of most species is subject to minor changes in the extent throughout the time. Therefore, modeling results support the presence of Neanderthals and anatomically modern humans, which hunted on these animals, during the Late Pleistocene in the Western Mediterranean. However, it must be pointed out that the reliability of each prediction depends strongly on the quality and quantity of the occurrence data and environmental variables used. For example, the data used by the GBIF Secretariat (2017) have several uncertainties and the climate data used, both for recent and past conditions, is based on models at different stages of development. Furthermore, models can not estimate the full potential distributional area and a region is designated as suitable, which may be smaller than the potential of the species actually allows.

For future modeling, the use of verified, dated, and georeferenced finding spots is recommended in order to evaluate the results statistically, which was not possible here. In this work, only the evaluation of the training of the respective methods could be performed. Further research should determine the best performing quantity of finding spots to obtain a consistent prediction. Furthermore, it is recommended to specialize in one method depending on the initial position and the objective. This includes defining the optimal configuration to avoid overfitting and determining a uniform rule for selecting the threshold in order to display presence/absence results. The value depends on the extent of a potential distribution area to be specified. Also recommended is to perform an evaluation of the new, initially mentioned methods for *single-season*, *multi-season* and *dynamic occupancy models* with detection-only and detection/nondetection data (Chapter 2), which were not content of this work.

6 References

6.1 Literature Sources

Adloff, F., Somot, S., Sevault, F., Jordà, G., Aznar, R., Déqué, M., Herrmann, M., Marcos, M., Dubois, C., Padorno, E., Alvarez-Fanjul, E., Gomis, D., 2015. Mediterranean Sea response to climate change in an ensemble of twenty first century scenarios. Climate Dynamics 45(9-10), 2775–2802.

Alados, C. L., 1985. Distribution and status of the Spanish Ibex (Capra pyrenaica). In: Lovari, S. (Ed.), The Biology and Management of Mountain Ungulates. Croom-Helm, Beckenham, UK, 204–211.

Alados, C. L., Escós, J., 1996. Ecología y comportamiento de la cabra montés. Consideraciones para su gestión. Museo Nacional de Ciencas Naturales (CSIC), Madrid, Spain.

Alba-Sánchez, F., López-Saez, J., Benito-de Pando, B., Linares, J., Nieto-Lugilde, D., Lápez-Merino, L., 2010. Past and present potential distribution of the Iberian Abies species: a phytogeographic approach using fossil pollen data and species distribution models. Diversity and Distributions 16, 214–228.

Alin, A., 2010. Multicollinearity. Wires Computational Statistics 2(3), 370–374.

Álvarez-Lao, D. J., García, N., 2010. Chronological distribution of Pleistocene cold-adapted large mammal faunas in the Iberian Peninsula. Quaternary International 212(2), 120–128.

Álvarez-Lao, D., García-García, N., 2006. A new site from the Spanish Middle Pleistocene with cold-resistant faunal elements: La Parte (Asturias, Spain). Quaternary International 142-143, 107–118.

de Andrés-Herrero, M., Becker, D., Weniger, G.-C., 2018. Reconstruction of LGM faunal patterns using Species Distribution Modelling. The archaeological record of the Solutrean in Iberia. Quaternary International 485, 199–208.

Araújo, M. B., Guisan, A., 2006. Five (or so) challenges for species' distribution modelling. Journal of Biogeography 33, 1677–1688.

Austin, M. P., 1980. Searching for a model to use in vegetation analysis. Vegetatio 42, 11–21.

Austin, M. P., Smith, T. M., 1989. A new model for the continuum concept. Vegetatio 83, 35–47.

Beaumont, L. J., Graham, E., Duursma, D. E., Wilson, P. D., Cabrelli, A., Baumgartner, J. B., Hallgren, W., Esperón-Rodríguez, M., Nipperess, D. A., Warren, D. L., Laffan, S. W., VanDerWal, J., 2016. Which species distribution models are more (or less) likely to project broad-scale, climate-induced shifts in species ranges? Ecological Modelling 342, 135–146.

Becker, D., Willmes, C., Bareth, G., Weniger, G.-C., 2016. A plugin to interface openModeller from QGIS for species' potential distribution modelling. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences III-7, 251–256.

Begon, M., Townsend, C. R., Harper, J. L., 2006. Ecology: From Individuals to Ecosystems. Blackwell Publishing, Oxford, UK.

Belsley, D. A., 1991. Conditioning diagnostics collinearity and weak data in regression. John Wiley & Sons, New York, USA.

Beninde, J., 1937. Zur Naturgeschichte des Rothirsches. Deutsche Gesellschaft für Kleintier- und Pelztierzucht, Leipzig, Germany.

Booth, T. H., Nix, H. A., Busby, J. R., Hutchinson, M. F., 2014. BIOCLIM: the first species distributionmodelling package, its early applications and relevance to most current MAXENT studies. Diversity and Distributions 20(1), 1–9.

Bout-Roumazeilles, V., Nebout, N. C., Peyron, O., Cortijo, E., Landais, A., Masson-Delmotte, V., 2007. Connection between South Mediterranean climate and North African atmospheric circulation during the last 50,000 yr BP North Atlantic cold events. Quaternary Science Reviews 26, 3197–3215.

Breiman, L., 2001. Statistical Learning: The Two Cultures. Statistical Science 16(3), 199–215.

Buse, J., Schroder, B., Assmann, T., 2007. Modelling habitat and spatial distribution of an endangered longhorn beetle - A case study for saproxylic insect conservation. Biological Conservation 137, 425–440.

Clutton-Brock, T. H., Guinness, F. E., Albon, S. D., 1982. Red deer: behavior and ecology of two sexes. Edinburgh University Press, Edinburgh, Scotland.

Collins, C., Asouti, E., Grove, M., Kabukcu, C., Bradley, L., Chiverrell, R., 2018. Understanding resource choice at the transition from foraging to farming: An application of palaeodistribution modelling to the Neolithic of the Konya Plain, south-central Anatolia, Turkey. Journal of Archaeological Science 96, 57–72.

Dansgaard, W., Johnsen, S., Clausen, H. B., Dahl-Jensen, D., Gundestrup, N., Hammer, C. U., Oeschger, H., 1984. North Atlantic climatic oscillations revealed by deep Greenland ice cores. In: Hansen, J. E., Takahashi, T. (Eds.), Climate Processes and Climate Sensitivity. American Geophysical Union, Washington D.C., USA, 288–298.

Depraz, A., Cordellier, M., Hausser, J., Pfenninger, M., 2008. Postglacial recolonization at a snail's pace (Trochulus villosus): confronting competing refugia hypotheses using model selection. Molecular Ecology 17, 2449–2462.

Dietterich, T., 1995. Overfitting and undercomputing in machine learning. ACM Computing Surveys 27(3), 326–327.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36(1), 27–46.

Duarte, A., Whitlock, S. L., Peterson, J. T., 2018. Species Distribution Modeling. In: Reference Module in Earth Systems and Environmental Sciences. Elsevier, 1–10.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K. S., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., Zimmermann, N. E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.

Elith, J., Leathwick, J. R., 2009a. The Contribution of Species Distribution Modelling to Conservation Prioritization. In: Moilanen, A., Wilson, K. A., Possingham, H. P. (Eds.), Spatial Conservation Prioritization: Quantitative Methods and Computational Tools., 70–93.

Elith, J., Leathwick, J. R., 2009b. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics (40), 677–697.

Elton, C. S., 1927. Animal Ecology. Sidgwick and Jackson, London, UK.

Emiliani, C., 1955. Pleistocene temperatures. Journal of Geology 63, 538–578.

Engländer, H., 1986. Capra pyrenaica Schinz, 1838 - Spanischer Steinbock, Iberiensteinbock. In: Niethammer, J., Krapp, F. (Hrsg.), Handbuch der Säugetiere Europas. 2. Aufl. AULA-Verlag, Wiesbaden, Germany, 405–422.

Epstein, S., Buchsbaum, H. A., Lowenstam, H. A., Urey, H. C., 1953. Revised carbonate-water isotopic temperature scale. Geological Society of America Bulletin 64, 1315–1326.

Epstein, S., Mayeda, T., 1953. Variation of O-18 content of water from natural sources. Geochimica et Cosmochimica Acta 4, 213–224.

FAO, 1977. Soil map of the world 1:5 000 000. Volume VI - Africa. United Nations, Paris, France.

FAO, 1981. Soil map of the world 1:5 000 000. Volume V - Europe. United Nations, Paris, France.

FAO, 1998. World Reference Base for soil resources. World Soil Resources Report 84.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shafferand, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. Review of Geophysics 45(2), 1–33.

Finlayson, C., 2004. Neanderthals and Modern Humans. An Ecological and Evolutionary Perspective. Cambridge University Press, Cambridge, UK.

Finlayson, C., Fa, D. A., Espejo, F. J., Carrión, J. S., Finlayson, G., Pacheco, F. G., Vidal, J. R., Stringer, C., Ruiz, F. M., 2008. Gorham's Cave, Gibraltar—The persistence of a Neanderthal population. Quaternary International 181(1), 64–71.

Fletcher, W. J., Sánchez Goñi, M. F., 2008. Orbital- and sub-orbital-scale climate impacts on vegetation of the western Mediterranean basin over the last 48,000 yr. Quaternary Research 70, 451–464.

Franklin, J., 2010. Mapping Species Distributions: Spatial Inference and Prediction. Ecology, Biodiversity and Conservation. Cambridge University Press, Cambridge.

Garcia Garriga, J., Martínez Molina, K., Baena Preysler, J., 2012. Neanderthal Survival in the North of the Iberian Peninsula? Reflections from a Catalan and Cantabrian Perspective. Journal of World Prehistory 25(2), 81–121.

Geiger, R., 1954. Klassifikation der Klimate nach W. Köppen. In: Landolt-Börnstein - Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik, alte Serie. 3rd ed. Springer, Berlin, Germany.

Geiger, R., 1961. Überarbeitete Neuausgabe von Geiger, R.: Köppen-Geiger / Klima der Erde. (Wandkarte 1:16 Mill.) Klett-Perthes, Gotha, Germany.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., Peterson, A. T., 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology and Evolution 19, 497–503.

Grinnel, J., 1904. The origin and distribution of the chestnut-backed chickadee. Auk 21, 364–365.

Grinnel, J., 1917. The niche-relationships of the California Thrasher. Auk 34, 427–433.

Grubb, P., 1993a. Order Artiodactyla. In: Wilson, D. E., Reeder, D. M. (Eds.), Mammal species of the world: a taxonomic and geographic reference. 2nd ed. Smithsonian Institution Press, Washington, USA, 377–414.

Grubb, P., 1993b. Order Perissodactyla. In: Wilson, D. E., Reeder, D. M. (Eds.), Mammal species of the world: a taxonomic and geographic reference. 2nd ed. Smithsonian Institution Press, Washington, USA, 369–372.

Guisan, A., Thuille, W., 2005. Predicting species distribution: Offering more than simple habitat models. Ecology Letters 8, 993–1009.

Guisan, A., Zimmermann, N. E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 147–186.

Guisan, A., Edwards, T. C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157(2-3), 89–100.

Harper, J., 1977. Population Biology of Plants. Academic Press, London, UK.

Hastie, T. J., Tibshirani, R. J., 1990. Generalized Additive Models. Chapman & Hall/CRC, London, UK.

Heinrich, H., 1988. Origin and consequences of cyclic ice rafting in the Northeast Atlantic Ocean during the past 130,000 years. Quaternary Research 29(2), 142–152.

Hemming, S. R., 2004. Heinrich events: Massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint. Reviews of Geophysics 42(1), 1–43.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965–1978.

Hijmans, R. J., Garrett, K. A., Huamán, Z., Zhang, D. P., Schreuder, M., Bonierbale, M., 2001. Assessing the Geographic Representativeness of Genebank Collections: the Case of Bolivian Wild Potatoes. Conservation Biology 14(6), 1755–1765.

Hijmans, R. J., Graham, C. H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global Change Biology 12, 2272–2281.

Hijmans, R. J., 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecological Society of America 93(3), 679–688.

Hijmans, R. J., Elith, J., 2017. Species distribution modeling with R.

Hirzel, A. H., Hausser, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data. Ecology 83, 2027–2036.

Hof, A. R., Jansson, R., Nilsson, C., 2012. The usefulness of elevation as a predictor variable in species distribution modelling. Ecological Modelling 246, 86–90.

Hoffecker, J. F., 2009. The spread of modern humans in Europe. Proceedings of the National Academy of Sciences 106(38), 16040–16045.

Hortal, J., Garcia-Pereira, P., Garcia-Barros, E., 2004. Butterfly species richness in mainland Portugal: predictive models of geographic distribution patterns. Ecography 27, 68–82.

Hughes, P. D., Fenton, C., Gibbard, P. L., 2011. Chapter 76 - Quaternary Glaciations of the Atlas Mountains, North Africa. In: Ehlers, J., Gibbard, P. L., Hughes, P. D.

(Eds.), Quaternary Glaciations - Extent and Chronology. Vol. 15. Developments in Quaternary Sciences. Elsevier, 1065–1074.

Hutchinson, G. E., 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology 22, 415–427.

Hutchinson, G. E., 1978. An Introduction to Population Ecology. Yale University Press, New Haven, USA.

Ibáñez, J. J., Zinck, J. A., Dazzic, C., 2013. Soil geography and diversity of the European biogeographical regions. Geoderma 192, 142–1532.

Iwahashi, J., Pike, R. J., 2007. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology 86(3), 409–440.

Jiménez-Valverde, A., 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Global Ecology and Biogeography 21(4), 498–507.

Kahlke, R. D., 1999. The History of the Origin, Evolution and Dispersal of the Late Pleistocene Mammuthus–Coelodonta Faunal Complex in Eurasia (Largue Mammals). Fenske, Rapid City, USA.

Köppen, W., 1884. Die Wärmezonen der Erde, nach der Dauer der heissen, gemässigten und kalten Zeit und nach der Wirkung der Wärme auf die organische Welt betrachtet. Meteorologische Zeitschrift 1, 351–360.

Köppen, W., 1918. Klassifikation der Klimate nach Temperatur, Niederschlag and Jahreslauf. Petermanns Geographische Mitteilungen 64, 193–203, 243–248.

Köppen, W., 1936. C. In: Köppen, W., Geiger, R. (Eds.), Das geographische System der Klimate.

Kottek, M., Jürgen, G., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift 15(3), 259–263.

Koza, J. R., Bennett, F. H., Andre, D., Keane, M. A., 2006. Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. In: S., G. J., Sudweeks, F. (Eds.), Artificial Intelligence in Design '96., 151–170.

Lisiecki, L. E., Raymo, M. E., 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ^{18}O records. Paleoceanography 20, 1–17.

Lister, A. M., Grubb, P., Sumner, S. R. M., 1998. Taxonomy, morphology and evolution of European roe deer. In: Andersen, R., Duncan, P., Linnell, J. D. C. (Eds.), The European Roe Deer: The Biology of Success. Scandinavian University Press, Oslo, Norway, 23–46.

Liu, C., Berry, P. M., Dawson, T. P., Pearson, R. G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28, 385–393.

Liu, C., Newell, G., White, M., 2016. On the selection of thresholds for predicting species occurrence with presence-only data. Ecology and Evolution 6(1), 337–348.

Lobo, J. M., Jiménez-Valverde, A., Real, R., 2007. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17, 145–151.

Luoto, M., Fronzek, S., Zuidhoff, F. S., 2004. Spatial modelling of palsa mires in relation to climate in northern Europe. Earth Surface Processes and Landforms 29, 1373–1387.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., Hines, J. E., 2017. Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence. Academic Press, San Diego, USA.

Maroto, J., Vaquero, M., Arrizabalaga, Á., Baena, J., Baquedano, E., Jordá, J., Julià, R., Montes, R., Plicht, J. V. D., Rasines, P., Wood, R., 2012. Current issues in late Middle Palaeolithic chronology: New assessments from Northern Iberia. Quaternary International 247, 15–25.

Masini, F., Lovari, S., 1988. Systematics, Phylogenetic Relationships, and Dispersal of the Chamois (Rupicapra spp.) Quaternary Research 30(3), 339–349.

McDonald, G. H., Bryson, R. A., 2010. Modeling Pleistocene local climatic parameters using macrophysical climate modeling and the paleoecology of Pleistocene megafauna. Quaternary International 217, 131–137.

Melis, C., Szafrańska, P. A., Jędrzejewska, B., Bartoń, K., 2006. Biogeographical variation in the population density of wild boar (Sus scrofa) in western Eurasia. Journal of Biogeography (33), 803–811.

Moreno, A., Valero-Garcés, B. L., Jiménez-Sánchez, M., Domínguez-Cuesta, M. J., Mata, M. P., Navas, A., González-Sampériz, P., Stoll, H., Farias, P., Morellón, M., Corella, J. P., Rico, M., 2010. The last deglaciation in the Picos de Europa National Park (Cantabrian Mountains, northern Spain). Journal of Quaternary Science 25(7), 1076–1091.

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A., Toxopeus, A. G., 2014. Where is positional uncertainty a problem for species distribution modelling? Ecography 37(2), 191–203.

Nix, H. A., 1986. A biogeographic analysis of Australian elapid snakes. In: Longmore, R. (Ed.), Atlas of Elapid Snakes of Australia. Australian Flora and Fauna Series Number. Australian Government Publishing Service, Canberra, Australia, 4–15.

Pearson, R. G., Dawson, T. P., 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? Global Ecology and Biogeography 12, 361–371.

Peel, M. C., Finlayson, B. L., McMahon, T. A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System Sciences 11, 1633–1644.

Peinado Lorca, M., Martínez-Parras, J. M., 1987. Castilla-La Mancha. In: Lorca, M. P., Martínez, S. R. (Eds.), La vegetación de España. Universidad de Alcalá de Henares, Alcalá de Henares, Spain, 163–196.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., Araújo, M. B., 2011. Ecological Niches and Geographic Distributions. Monographs in Population Biology 49. Princeton University Press, Princeton, USA.

Phillips, S. J., Anderson, R. P., Schapire, R. E., 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190(3-4), 231–259.

Polly, P. D., Eronen, J. T., 2011. Modeling Pleistocene local climatic parameters using macrophysical climate modeling and the paleoecology of Pleistocene megafauna. In: Ashton, N., Lewis, S., Stringer, C. (Eds.), The Ancient Human Occupation of Britain: Developments in Quaternary Science. Vol. 14, 279–304.

Rasmussen, S. O., Bigler, M., Blockley, S. P., Blunier, T., Buchardt, S. L., Clausen, H. B., Cvijanovic, I., Dahl-Jensen, D., Johnsen, S. J., Fischer, H., Gkinis, V., Guillevic, M., Hoek, W. Z., Lowe, J. J., Pedro, J. B., Popp, T., Seierstad, I. K., Steffensen, J. P., Svensson, A. M., Vallelonga, P., Vinther, B. M., Walker, M. J., Wheatley, J. J., Winstrup, M., 2014. A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy. Quaternary Science Reviews 106, 14–28.

Ravelo, C., Hillaire-Marcel, C., 2007. The use of oxygenand carbon isotopes of foraminifera in paleoceanograph. In: Hillaire-Marcel, C., Vernal, A. de (Eds.), Developments in Marine Geology, vol. 1, Proxies in Late Cenozoic Paleoceanography. Elsevir, New York, USA, 735–764.

Rodriguez-Sanchez, F., Arroyo, J., 2008. Reconstructing the demise of Tethyan plants: climate-driven range dynamics of Laurus since the Pliocene. Global Ecology and Biogeography 17, 685–695.

Rohling, E. J., Foster, G. L., Grant, K. M., Marino, G., Roberts, A. P., Tamisiea, M. E., Williams, F., 2014. Sea-level and deep-sea-temperature variability over the past 5.3 million years. Nature 508, 477–482.

Salazar-García, D. C., Power, R. C., Serra, A. S., Villaverde, V., Walker, M. J., Henry, A. G., 2013. Neanderthal diets in central and southeastern Mediterranean Iberia. Quaternary International 318, 3–18.

Samuel, A. L., 1959. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development 3(3), 210–229.

Sánchez Goñi, M. F., Cacho, I., Turon, J.-L., Guiot, J., Sierro, F. J., Peypouquet, J.-P., Grimalt, J., Shackleton, N. J., 2002. Synchroneity between marine

and terrestrial responses to millenial scale climatic variability during the last glacial period in the Mediterranean region. Climate Dynamics 19, 95–105.

Sánchez Goñi, M. F., Turon, J.-L., Eynaud, F., Gendreau, S., 2000. European Climatic Response to Millennial-Scale Changes in the Atmosphere-Ocean System during the Last Glacial Period. Quaternary Research 54, 394–403.

Sánchez Goñi, M. F., Harrison, S. P., 2010. Millennial-scale climate variability and vegetation changes during the Last Glacial: Concepts and terminology. Quaternary Science Reviews 29(21-22), 2823–2327.

Sánchez-Fernández, D., Lobo, J. M., Hernández-Manrique, O. L., 2011. Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles. Diversity and Distributions 17, 163–171.

Sandwell, D. T., Gille, S. T., Smith, W. H. F., 2003. Sensing Bathymetry from Space. Bathymetry from Space: Oceanography, Geophysics, and Climate, La Jolla, California, USA, Oct. 24–26, 2002, 21–22.

Sepulchre, P., Ramstein, G., Kageyama, M., Vanhaeren, M., Krinner, G., Sánchez-Goñi, M.-F., d'Errico, F., 2007. H4 abrupt event and late Neanderthal presence in Iberia. Earth and Planetary Science Letters 258(1), 283–292.

Serrano, E., Gómez-Lende, M., González-Amuchastegui, M. J., González-García, M., González-Trueba, J. J., Pellitero, R., Rico, I., 2015. Glacial chronology, environmental changes and implications for human occupation during the upper Pleistocene in the eastern Cantabrian Mountains. Quaternary International 364, 22–34.

Simón, M., Sánchez, S., García, I., 2000. Soil-landscape evolution on a Mediterranean high mountain. Catena 39(3), 211–230.

Smith, S. E., Gregory, R. D., Anderson, B. J., Thomas, C. D., 2013. The past, present and potential future distributions of cold-adapted birdspecies. Diversity and Distributions 19, 352–362.

Smith, W. H. F., Sandwell, D. T., 1997. Global Sea Floor Topography from Satellite Altimetry and Ship Depth Soundings. Science 277(5334), 1956–1962.

Soberón, J., Llorente, J., Benítez, H., 1996. An international view of national biological survey. Annals of the Missouri Botanical Garden 83, 562–573.

Soberón, J., Peterson, A. T., 2004. Biodiversity informatics: Managing and applying primary biodiversity data. Philosophical Transactions of the Royal Society of London B 359, 689–698.

Sommer, R. S., Fahlke, J. M., Schmölcke, U., Benecke, N., Zachos, F. E., 2009. Quaternary history of the European roe deer Capreolus capreolus. Mammal Review 39(1), 1–16.

Sommer, R. S., Zachos, F. E., Street, M., Jöris, O., Skog, A., Benecke, N., 2008. Late Quaternary distribution dynamics and phylogeography of the red deer (Cervus elaphus) in Europe. Quaternary Science Reviews 27(7-8), 714–733.

Stewart, J. R., 2005. The ecology and adaptation of Neanderthals during the non-analogue environment of Oxygen Isotope Stage 3. Quaternary International 137(1), 35–46.

Svendsen, J. I., Alexanderson, H., Astakhov, V. I., Demidov, I., Dowdeswell, J. A., Funder, S., Gataullin, V., Henriksen, M., Hjort, C., MichaelHoumark-Nielsen, Hans W. Hubbertenk, Ó. I., Jakobsson, M., Kjæri, K. H., Larsen, E., Lokrantz, H., Lunkka, J. P., Lyså, A., Mangerud, J., Matiouchkov, A., Murray, A., Möller, P., Niessen, F., Nikolskaya, O., Polyak, L., Saarnisto, M., Siegert, C., Siegert, M. J., Spielhagen, R. F., Stein, R., 2004. Late Quaternary ice sheet history of northern Eurasia. Quaternary Science Reviews 23, 1229–1271.

Swets, J. A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.

Thibaud, E., Petitpierre, B., Broennimann, O., Davison, A. C., Guisan, A., 2014. Measuring the relative effect of factors affecting speciesdistribution model predictions. Methods in Ecology and Evolution 5, 947–955.

Thuiller, W., Brotons, L., Araujo, M., Lavorel, S., 2004. Effects of restricting environmental range of data to project current and future species distributions. Ecography 27, 165–172.

Tôrres, N. M., De Marco, P., Santos, T., Silveira, L., Almeida Jácomo, A. T. de, Diniz-Filho, J. A. F., 2012. Can species distribution modelling provide estimates of population densities? A case study with jaguars in the Neotropics. Diversity and Distributions 18(6), 615–627.

Trigo, R. M., Pozo-Vázquez, D., Osborn, T. J., Castro-Díez, Y., Gámiz-Fortis, S., Esteban-Parra, M. J., 2004. North Atlantic oscillation influence on precipitation, river flow and water resources in the Iberian Peninsula. International Journal of Climatology 24(8), 925–944.

Urey, H. C., Epstein, S., Lowenstam, H. A., McKinney, C. R., 1951. Measurement of paleo- temperatures and temperatures of the upper Cretaceous of England, Denmark, and the south-eastern United States. Geological Society of America Bulletin 62, 399–416.

Varela, S., Rodriguez, J., Lobo, J. M., 2009. Is current climatic equilibrium a guarantee for the transferability of distribution model predictions? A case study of the spotted hyena. Journal of Biogeography 36, 1645–1655.

Varela, S., Lobo, J. M., Hortal, J., 2011. Using species distribution models in paleobiogeography: A matter of data, predictors and concepts. Palaeogeography, Palaeoclimatology, Palaeoecology 310(3-4), 451–463.

Varela, S., Lobo, J. M., Rodríguez, J., Batra, P., 2010. Were the Late Pleistocene climatic changes responsible for the disappearance of the European spotted hyena populations? Hindcasting a species geographic distribution across time. Quaternary Science Reviews 29, 2027–2035.

Vierod, A. D. T., Guinotte, J. M., Davies, A. J., 2014. Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. Deep-Sea Research II 99, 6–18.

Walls, B. J., Stigall, A. L., 2011. Analyzing niche stability and biogeography of Late Ordovician brachiopod species using ecological niche modeling. Palaeogeography, Palaeoclimatology, Palaeoecology 299, 15–29.

Walvius, M. R., 1961. A discussion of the size of recent red deer (Cervus elaphus L.) compared with prehistoric specimens. Beaufortia 9(97), 75–82.

Wang, P., Tian, J., Lourens, L. J., 2010. Obscuring of long eccentricity cyclicity in Pleistocene oceanic carbon isotope records. Earth and Planetary Science Letters 290(3-4), 319–330.

Warren, D. L., Seifert, S. N., 2011. Ecological niche modeling in Maxent: the importance of modelcomplexity and the performance of model selection criteria. Ecological Applications 21(2), 335–342.

Weatherall, P., Marks, K. M., Jakobsson, M., Schmitt, T., Tani, S., Arndt, J. E., Rovere, M., Chayes, D., Ferrini, V., Wigley, R., 2015. A new digital bathymetric modelof the world's oceans. Earth and Space Science 2, 331–345.

Weinstock, J., 1998. Late paleolithic reindeer populations in central and western Europe. Antropozoologica 25-26(46), 383–388.

Willmes, C., Becker, D., Brocks, S., Hütt, C., Bareth, G., 2017. High Resolution Köppen-Geiger Classifications of Paleoclimate Simulations. Diversity and Distributions 21(1), 57–73.

Willmes, C., Hölzchen, E., Sommer, C., Rodriguez, J., 2018. Spatio-temporal interpolation of paleoclimate data based on $\delta^{18}O$ climate variability observations. Geophysical Research Abstracts 20.

6.2 Online Sources

Alcaraz-Castaño, M., Alcolea-González, J., Kehl, M., Albert, R.-M., Baena-Preysler, J., de Balbín-Behrmann, R., Cuartero, F., Cuenca-Bescós, G., Jiménez-Barredo, F., López-Sáez, J.-A., Piqué, R., Rodríguez-Antón, D., Yravedra, J., Weniger, G.-C., 2017. A context for the last Neandertals of interior Iberia: Los Casares cave revisited. `https://doi.org/10.1371/journal.pone.0180823`, 2018-8-2.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. `https://doi.org/10.5194/gmd-8-1991-2015`, 2018-7-27.

CRC 806, 2018a. Research profile of the Collaborative Research Centre 806 (Phase3). `http://www.sfb806.uni-koeln.de/index.php/about`, 2018-9-17.

CRC 806, 2018b. Settlement Patterns and Climate Change in the Late Pleistocene of the Western Mediterranean - A Synthesis. `http://www.sfb806.uni-koeln.de/index.php/projects/r-regional-models/c1`, 2018-9-17.

Edwards, C. J., Magee, D. A., Park, S. D. E., McGettigan, P. A., Lohan, A. J., Murphy, A., Finlay, E. K., Shapiro, B., Chamberlain, A. T., Richards, M. B., Bradley, D. G., Loftus, B. J., MacHugh, D. E., 2010. A Complete Mitochondrial Genome Sequence from a Mesolithic Wild Aurochs (Bos primigenius). `https://doi.org/10.1371/journal.pone.0009255`, 2018-8-1.

GBIF Secretariat, 2017. GBIF Backbone Taxonomy. `https://doi.org/10.15468/39omei`, 2018-3-12.

GBIF Secretariat, 2018. What is GBIF? `https://www.gbif.org/what-is-gbif`, 2018-4-1.

General Bathymetric Chart of the Oceans, 2014. GEBCO_2014. `https://www.gebco.net/data_and_products/gridded_bathymetry_data/`, 2018-6-19.

GRASS Development Team, 2018. Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.4. Open Source Geospatial Foundation. `http://grass.osgeo.org`, 2018-7-15.

Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J., 2017. Package 'dismo'. `https://cran.r-project.org/web/packages/dismo/dismo.pdf`, 2018-7-31.

Holthausen, M., 2018. Supplementary data for "Distribution modeling of paleofauna in the Western Mediterranean between the Heinrich events H5 and H4" (master's thesis). `https://doi.org/10.5880/SFB806.44`, 2018-9-12.

iNaturalist.org, 2018a. iNaturalist Research-grade Observations: Bos Linnaeus, 1758. `https://www.inaturalist.org/observations/9371620`, 2018-4-3.

iNaturalist.org, 2018b. iNaturalist Research-grade Observations: Capra pyrenaica Schinz, 1838. `https://www.inaturalist.org/observations/4516436`, 2018-4-3.

iNaturalist.org, 2018c. iNaturalist Research-grade Observations: Capreolus capreolus Linnaeus, 1758. `https://www.inaturalist.org/observations/10015511`, 2018-4-3.

iNaturalist.org, 2018d. iNaturalist Research-grade Observations: Cervus elaphus Linnaeus, 1758. `https://www.inaturalist.org/observations/10708096`, 2018-4-3.

iNaturalist.org, 2018e. iNaturalist Research-grade Observations: Equus caballus Linnaeus, 1758. `https://www.inaturalist.org/observations/10439062`, 2018-4-3.

iNaturalist.org, 2018f. iNaturalist Research-grade Observations: Rangifer tarandus Linnaeus, 1758. `https://www.inaturalist.org/observations/6633128`, 2018-4-3.

iNaturalist.org, 2018g. iNaturalist Research-grade Observations: Rupicapra pyrenaica Bonaparte, 1845. `https://www.inaturalist.org/observations/527454`, 2018-4-3.

iNaturalist.org, 2018h. iNaturalist Research-grade Observations: Sus scrofa Linnaeus, 1758. `https://www.inaturalist.org/observations/2464952`, 2018-4-3.

Jet Propulsion Laboratory, 2018. U.S. Releases Enhanced Shuttle Land Elevation Data. `https://www2.jpl.nasa.gov/srtm/`, 2018-7-14.

Phillips, S. J., Dudík, M., Schapire, R. E., Maxent software for modeling species niches and distributions (Version 3.4.1). `http://biodiversityinformatics.amnh.org/open_source/maxent/`, 2017-11-21.

QGIS Development Team, 2018. QGIS Geographic Information System (QGIS) Software, Version 3.2. Open Source Geospatial Foundation. `http://qgis.osgeo.org`, 2018-7-15.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. `https://www.R-project.org`, 2018-7-31.

Shapiro, M., Waupotitsch, O., 2015. Manual: r.slope.aspect. Open Source Geospatial Foundation. `https://grass.osgeo.org/grass74/manuals/r.slope.aspect.html`, 2018-7-16.

Solomon, S., Bacci Jr., M., Martins Jr., J., Goncalves Vinha, G., Mueller, U. G., 2008. Paleodistributions and comparative molecular phylogeography of leafcutter ants (Atta spp.) provide new insight into the origins of Amazonian diversity. `https://doi.org/10.1371/journal.pone.0002738`, 2018-8-15.

Statement of Authorship

Selbständigkeitserklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden.

Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Michael Holthausen
Köln, 24. September 2018