



# Validating the transferability of ecological models under global change scenarios with Holocene rock-art

Luís Filipe Oliveira Silveira

Master in Biodiversity, Genetics and Evolution

Department of Biology

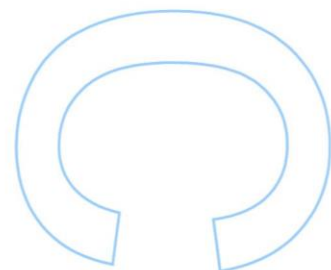
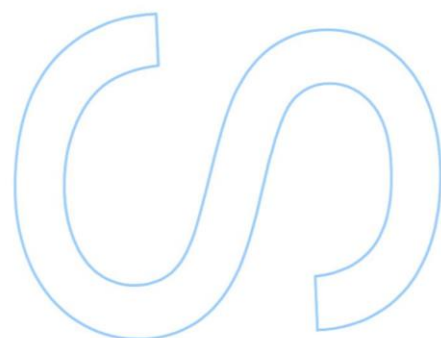
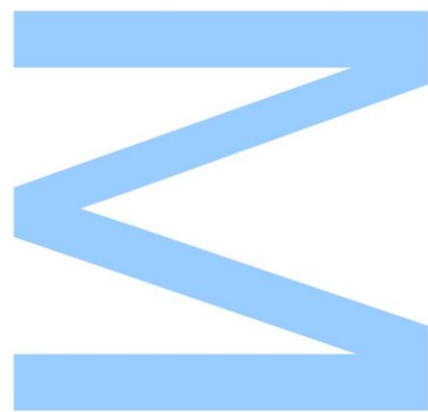
2019

## Supervisor

José Carlos Alcobia Rogado de Brito, Principal Researcher, CIBIO

## Co-supervisor

Hugo Rebelo, Assistant Researcher, CIBIO



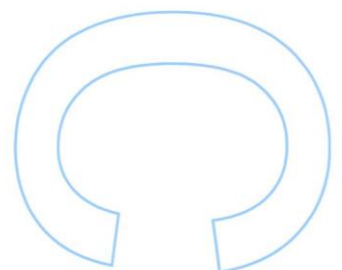
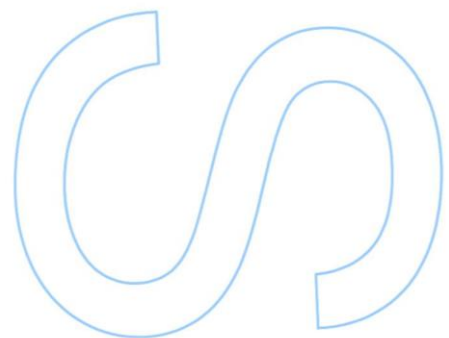
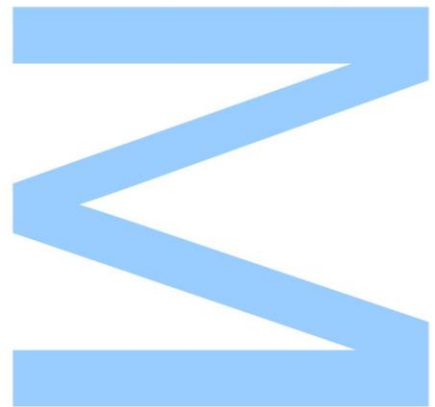




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, \_\_\_\_/\_\_\_\_/\_\_\_\_





## Agradecimentos/Acknowledgments

Em primeiro lugar aos meus orientadores, José Carlos Brito e Hugo Rebelo por me darem a oportunidade de poder desenvolver este trabalho sob a sua orientação. Por toda a disponibilidade e acompanhamento incansável ao longo deste ano e por todos os conhecimentos científicos partilhados imprescindíveis para cumprir esta meta.

Ao Pedro Tarroso, por estar sempre disponível para me ajudar em muitos dos problemas que encontrei durante este trabalho.

To the BIODESERTS group for all the knowledge conveyed during the Journal Club discussions throughout this year.

To the British Museum, specially to Doctor Helen Anderson, for providing me crucial data for this work.

Aos meus colegas de mestrado por todos os bons momentos partilhados durante estes dois anos.

Por fim, aos meus pais, que sem eles nunca aqui teria chegado sem todo o investimento e apoio na minha educação e no meu futuro.



## Abstract

The world is facing a global biodiversity crisis, with an increasing rate of species extinction due to habitat loss and climate change. Predictions of biodiversity responses to global change play a key role to guide conservation efforts. Predictions of climate change effects on biodiversity distribution usually rely on ecological niche-based models due to their ability to predict species responses in unsampled spatial and/or temporal scenarios. Assessing the quality of model transferability across time is critical if they are to be used to predict responses beyond the conditions under which they were trained. However, validating model transferability requires independent data that is impossible to obtain in past or future time scenarios. Rock-art depicting species may provide evidence on paleo-distributions and a unique opportunity to test the robustness of model projections, which may help understating the reliability of future projections. Using rock-art from the Holocene of African savannah elephant (*Loxodonta africana*) and Giraffe (*Giraffa camelopardalis*), this work evaluated the predictive ability of ecological models based in distinct: i) strategies to select pseudo-absence data (minimum distance to observations, random data, and environmentally restricted data); ii) sources of climatic layers (Worldclim and Paleoview) for three time periods (Current, 6,000BP and 7,900BP); and iii) environmental variables (land-cover and bioclimatic datasets). Models were constructed with algorithm Random Forests for the three time periods (Current, 6,000BP and 7,900BP) and projected to all possible combinations of time periods to predict suitable areas for both species. Overall, models calibrated with environmentally restricted pseudo-absences performed better, no major differences were found between models calibrated with different sources of bioclimatic layers, and models coupling land-cover variables with bioclimatic variables performed better than pure bioclimatic or land-cover based models. The study here developed allows providing a series of recommendations related to model transferability in time: 1) When the distributions of studied species are not totally known but the ecological niche traits are well known, it is recommend using environmentally restricted pseudo-absences datasets; 2) When projecting models, it is best to use Worldclim bioclimatic variables given the higher availability of spatial resolutions and GCMs; 3) Coupling land-cover together with bioclimatic variables is recommended, since the association of these variables may reflect accurately environmental changes between calibration and projection periods. Still, given the limitations in past land-cover availability, trade-offs will need to be established between pixel sizes and data sources.

**Keywords:** Africa; Bioclimatic variables; *Giraffa camelopardalis*; Land-cover; *Loxodonta africana*; Paleoview; Pseudo-absences; Random Forests; Range Contraction; Worldclim;

## Resumo

Atualmente, o planeta enfrenta uma crise global de biodiversidade onde cada vez mais espécies se extinguem devido à destruição de habitat e às alterações climáticas. As previsões das respostas da biodiversidade às alterações climáticas são fundamentais para guiar os esforços de conservação. As previsões dos efeitos das alterações climáticas na distribuição da biodiversidade normalmente dependem de modelos baseados no nicho ecológico devido à capacidade em prever respostas das espécies em diferentes cenários espaciais e/ou temporais para os quais não há amostragem. Para determinar a qualidade da transferibilidade dos modelos são necessários dados independentes, os quais são impossíveis de obter para cenários passados ou futuros. Representações de animais na arte rupestre podem fornecer evidências sobre paleodistribuições e uma oportunidade única para testar a robustez das projeções dos modelos, o que pode ajudar a estimar a confiança nas projeções futuras. Utilizando representações de Elefante da savana (*Loxodonta africana*) e de girafa (*Giraffa camelopardalis*) na arte rupestre do Holocénico, este trabalho avaliou a capacidade de previsão destes modelos ecológicos com base em distintas: i) estratégias para gerar dados de pseudo-ausência (distância mínima às observações, aleatoriamente e ambientalmente restrito); ii) fontes de dados climáticos (Worldclim e Paleoview) para três períodos distintos (atual, 6000BP e 7900BP); e iii) variáveis ambientais (ocupação do solo e conjuntos de dados bioclimáticos). Os modelos foram baseados no algoritmo Random Forests para os três períodos (atual, 6000BP e 7900BP) e projetados para todas as combinações possíveis de períodos temporais para prever áreas adequadas para ambas as espécies. Em geral, os modelos calibrados com pseudo-ausências restritas tiveram um melhor desempenho, nenhuma diferença importante foi encontrada entre os modelos calibrados com diferentes fontes bioclimáticas e os modelos que conjugam variáveis de ocupação do solo com variáveis bioclimáticas tiveram um melhor desempenho do que os modelos puramente bioclimáticos ou baseados apenas na ocupação do solo. O estudo aqui desenvolvido permite fornecer uma série de recomendações relacionadas com a transferibilidade no tempo de modelos ecológicos: 1) Quando as distribuições das espécies estudadas não são totalmente conhecidas, mas as características do nicho são bem conhecidas, recomenda-se o uso de pseudo-ausências ambientalmente restritas; 2) Ao projetar modelos, é preferível usar variáveis bioclimáticas do Worldclim, dada a maior disponibilidade de resoluções espaciais e GCMs; 3) É recomendado a conjugação de variáveis de ocupação do solo com variáveis bioclimáticas, uma vez que a associação destas variáveis pode refletir com precisão as mudanças ambientais entre períodos de calibração e projeção. Ainda assim, dadas as limitações da



disponibilidade de variáveis de ocupação do solo no passado, será necessário estabelecer compromissos entre o tamanho do pixel e as fontes de dados.

**Palavras-chave:** África; Contrações da distribuição; *Giraffa camelopardalis*; *Loxodonta africana*; Ocupação do solo; Paleoview; Pseudo-ausências; Random Forests; Variáveis bioclimáticas; Worldclim;

# Table of Contents

Agradecimentos/Acknowledgments .....	1
Abstract .....	3
Resumo .....	4
Table of Contents.....	6
List of Tables.....	7
List of Figures.....	8
List of Abbreviations.....	9
1. Introduction .....	10
1.1 Assumptions and uncertainties of ecological niche-based models.....	13
1.2 Ensemble forecasting .....	17
1.3 Transferability of ecological models .....	18
1.4 Validation of model's transferability .....	19
1.5 Species Niche Contraction.....	22
2. Objectives .....	25
3. Material and Methods .....	26
3.1 Study species.....	26
3.2 Study area.....	26
3.3 Species observations.....	26
3.4 Environmental variables.....	28
3.4.1 Bioclimatic variables.....	28
3.4.2 Land-cover variables.....	34
3.5 Model building strategy .....	36
3.6 Evaluation and comparison of models.....	37
4. Results .....	39
4.1 Distribution of study species .....	39
4.2 Model projections .....	40
5. Discussion.....	50
5.1 Distribution of elephant and giraffe at the Holocene .....	50
5.2 Effects of ecogeographical variables in model transferability .....	51
5.3 Methods to select pseudo-absences for model transfer .....	52
5.4 Sources of uncertainty and limitations.....	54
5.5 Guidelines for model transferability and future research .....	55
6. References.....	56
7. Annex.....	70

## List of Tables

<b>Table 1</b>	Range values (minimum - maximum) of each bioclimatic variable in each climate dataset (Paleoview and Worldclim) in each time period (current, 6,000BP, and 7,900BP).	31
<b>Table 2</b>	Euclidean distances to land-cover types in each time period.	35
<b>Table 3</b>	Number of pseudo-absences generated for each method and for each period.	37
<b>Table 4</b>	Number of observations collected for each species and time period.	40

## List of Figures

<b>Figure 1</b>	The 'BAM diagram', showing a simplified framework for understanding where species will be distributed.	12
<b>Figure 2</b>	The distribution of the Savannah elephant in North Africa during the Holocene.	19
<b>Figure 3</b>	Distribution at about 1900s (points) and in current time (polygons) of the African savannah elephant ( <i>Loxodonta africana</i> ) and the Giraffe ( <i>Giraffa camelopardalis</i> ) in Africa.	21
<b>Figure 4</b>	Rock-art depicting an elephant in Libya	26
<b>Figure 5</b>	Worldclim bioclimatic variables constructed for Current time period	32
<b>Figure 6</b>	Worldclim bioclimatic variables constructed for 6000BP time period	32
<b>Figure 7</b>	Worldclim bioclimatic variables constructed for 7900BP time period	33
<b>Figure 8</b>	Paleoview bioclimatic variables constructed for Current time period	33
<b>Figure 9</b>	Paleoview bioclimatic variables constructed for 6000BP time period	34
<b>Figure 10</b>	Paleoview bioclimatic variables constructed for 7900BP time period	34
<b>Figure 11</b>	Euclidean distances calculated for each biome for current time period.	36
<b>Figure 12</b>	Euclidean distances calculated for each biome for 6000BP time period.	36
<b>Figure 13</b>	Euclidean distances calculated for each biome for 7900BP time period.	38
<b>Figure 14</b>	Modelling strategy diagram	40
<b>Figure 15</b>	Distribution of "Current" (top) and Holocene (bottom) datasets obtained for the Elephant and Giraffe, distinguishing observations used for model training from the full set of available observations.	43
<b>Figure 16</b>	Models calibrated in current time period and projected to 6,000BP time-period.	44
<b>Figure 17</b>	Models calibrated in 6,000BP time period and projected to current time-period.	45
<b>Figure 18</b>	Models calibrated in current time period and projected to 7,900BP time-period.	46
<b>Figure 19</b>	Models calibrated in 7.900BP time period and projected to current time-period.	47
<b>Figure 20</b>	Percentage of observations of each study species in the Holocene (y axis) in each probability of occurrence class (x axis) according to models calibrated in current conditions and projected to both Holocene time periods.	48
<b>Figure 21</b>	Percentage of observations of each study species in the "Current" dataset (y axis) in each probability of occurrence class (x axis) according to models calibrated in both Holocene time periods and projected to current conditions.	49
<b>Figure 22</b>	Percentage of pseudo-absences generated for each study species at the Holocene (y axis) in each probability of occurrence class (x axis) according to models calibrated in current conditions and projected to both Holocene time periods.	50
<b>Figure 23</b>	Percentage of pseudo-absences generated for each study species in current conditions (y axis) in each probability of occurrence class (x axis) according to models calibrated in both Holocene time periods and projected to current conditions.	51

## List of Abbreviations

<b>ENM</b>	Ecological Niche Model
<b>GLM</b>	Generalized Linear Model
<b>GAM</b>	Generalized Addictive Model
<b>ANN</b>	Artificial Neural Networks
<b>ML</b>	Machine Learning
<b>RF</b>	Random Forests
<b>GCM</b>	General Circulation Model
<b>CCSM3</b>	Community Climate System Model ver. 3
<b>MESS</b>	Multivariate Environmental Similarity Surfaces
<b>PA</b>	Pseudo-absence
<b>ROC</b>	Receiver Operating Characteristic
<b>AUC</b>	Area Under Curve
<b>MARS</b>	Multivariate adaptive regression splines
<b>CTA</b>	Classification Tree Analysis
<b>MDA</b>	Model Driven Architecture
<b>BRT</b>	Boosted Regression Trees

# 1. Introduction

We are currently facing a global biodiversity crisis, with an increasing rate of species extinction and habitat loss (Joppa *et al.*, 2016). Although extinction events were always part of the natural history processes, current rates of extinction are about 1000 times higher than the likely rate in the absence of changes caused by humans, and future rates are expected to increase (Pimm *et al.*, 2014). In the last few decades we have witnessed catastrophic declines in both the number and size of populations of many species (Ceballos & Ehrlich, 2002; Pimm, 2008). The huge ongoing population declines and extinctions will likely have negative cascading effects on ecosystem function and services that are critical to sustain humanity (Ceballos *et al.*, 2017). Anthropogenic activities appear as the major reason for the current crisis, together with climate change (Tittensor *et al.*, 2014; Joppa *et al.*, 2016). The steady overall decline of wild species populations size, range and connectivity between suitable habitats, are amplified by human-induced climate change, poaching, land-use transformation, and exotic species introduction (Butchart *et al.*, 2010). While human-induced natural habitat fragmentation is expected to have a more immediate impact on biodiversity, the effects of climate change are likely to become a long-term issue with unpredictable consequences (Hof *et al.*, 2011). These issues create different threats to biodiversity, leading to range contractions and species extinctions (Higgins, 2007). Hence, it is of most importance to understand how climate change interacts with other anthropogenic impacts on biodiversity patterns to predict the future of biodiversity to implement effective conservation plans in advance (Bellard *et al.*, 2012).

Predictions play a key role in alerting researchers and decision makers to potential consequences of climate change: (i) they provide means to aid in the attribution of biological changes to climate change and anthropogenic activities; and (ii) they can support the development of proactive strategies to reduce climate change impacts on biodiversity (Pereira *et al.*, 2010). Attempts to predict climate change effects on biodiversity have relied on ecological niche-based models (Pearson & Dawson, 2003). Through modelling we can define the environmental space that limits the distribution of species and project it against different spatial and temporal scenarios by using associations between known occurrences of species and bioclimatic variables across target landscapes to predict species distributions (Araújo & Peterson, 2012). Several different ecological niche-based modelling (ENMs) methods utilizing different algorithms were developed over the past decades. The impact that the specific method has on model predictions is an important consideration in model applications (Pearson *et al.*, 2006).

Currently, ENMs rely on two main approaches: i) correlative niche models focus on understanding conditions that allow persistence of species by modelling the observed distribution of a species as a function of environmental conditions (Elith & Leathwick, 2009); and ii) mechanistic niche models aim to understand the environmental requirements that make up the fundamental niche of a species by using detailed biophysical modelling approaches (Dormann *et al.*, 2012), where data about niche, dispersal, and sometimes biotic interactions of the entire set of processes leading to the occupation of areas by a species is integrated in models and simulations (Peterson *et al.*, 2015). Most mechanistic models are data-hungry and rely heavily on parameters derived from field observations or empirical relationships. Given the constraints of data availability, obtaining the detailed measurements necessary for robust model calibration is time-consuming and expensive (Peterson *et al.*, 2015), such that mechanistic models have only been successfully built for the most charismatic, well-studied, and/or economically relevant species (Evans *et al.*, 2016). These characteristics limit their utility to support many of the management decisions that model predictions could potentially inform (Bouchet *et al.*, 2019). Thus, information gaps often make correlative descriptions of patterns the only viable pathway for making predictions. Mechanistic and correlative models also share many of the same underlying issues such as: (i) equifinality; (ii) non-stationarity; (iii) model misspecification; and (iv) model complexity (Dormann *et al.*, 2012). Both mechanistic and correlative models are equally valuable, however, in the majority of cases correlative models prevail as the most employable (Bouchet *et al.*, 2019), due to the limitations in obtaining data needed to develop mechanistic approaches.

Regression-based methods attempt to model the variation in species occurrence or abundance within the occupied environmental space, selecting predictors according to their importance (Elith & Leathwick, 2009). Generalized linear models (GLM) are based on an assumed relationship between the mean of the response variable and the linear combination of the explanatory variables (Guisan *et al.*, 2002). Generalised additive models (GAM; Hastie & Robert, 1990) are semi-parametric extensions of GLM. Both GAM and GLM use a link function to establish a relationship between the mean of the response variable and a function of the explanatory variables. The strength of GAM is their ability to deal with highly non-monotonic and non-linear relationships between the response and the set of explanatory variables (Guisan *et al.*, 2002). More recently, the artificial neural networks (ANNs; Lek & Guégan, 1999) have been applied in ecological studies because of its ability to deal with complex datasets, and for the powerful predictive power in the analysis of both linear and nonlinear relationships as well (Tarroso *et al.*, 2012). ANNs can identify and learn correlated patterns between input data and corresponding target values, making it a powerful model algorithm, especially when the underlying relationships in the

data are not known (Lek & Guégan, 1999). Machine-Learning (ML) algorithms are usually seen as more powerful to deal with complex ecological datasets than other methods (Pearson *et al.*, 2002; Olden *et al.*, 2008). Some studies have suggested that ML algorithms can perform better than the regression-based ones (Elith *et al.*, 2006). Together with MaxEnt (Phillips *et al.*, 2006), Random Forest (RF; Breiman, 2001) is considered to be among the most powerful ML algorithms, for common usages and for obtaining robust ensemble models (Elith *et al.*, 2006; Araújo & New, 2007; Wisz *et al.*, 2008; Mi *et al.*, 2017). In this algorithm, random samples from rows and predictors are utilized to build several trees. Each individual tree is constructed from a bootstrap sample and split at each node by the best predictor from a small, randomly chosen subset of the predictor variables. These trees comprising the forest are each grown to maximal depth, and predictions are built by averaged trees through “voting” (Breiman, 2001). RF avoids overfitting by controlling the number of predictors randomly used at each split, using means of out-of-bag samples to calculate an unbiased error rate (Mi *et al.*, 2017). There are several advantages of RF in comparison to other algorithms, such as: (i) great classification accuracy; (ii) ability to model complex interactions amidst predictor variables; (iii) flexibility to execute several types of statistical data analysis, including regression, classification, survival analysis, and unsupervised learning; and (iv) an algorithm for imputing missing data (Cutler *et al.*, 2007). Despite RF being currently on the rise, its potential is still widely underused in conservation, ecological applications and inference (Mi *et al.*, 2017). In contrast, MaxEnt is a widely used method for making predictions, especially from incomplete datasets. It is based on the Maximum Entropy algorithm that estimates the distribution of a species by determining the probability distribution of maximum entropy (Phillips *et al.*, 2006). MaxEnt’s predictive performance is consistently one of the highest among modelling methods (Elith *et al.*, 2006). Some of the advantages include: (i) it requires only presence data and environmental information as input; (ii) it can be applied to species presence/absence data; (iii) it has the ability to incorporate interactions between different variables; (iv) it is based on efficiently deterministic algorithms that guarantee to converge to the optimal (maximum entropy) probability distribution; (v) it avoids overfitting by using regularization parameters; and (vi) it produces continuous outputs, allowing fine distinctions to be made between the modelled suitability of different areas (Phillips *et al.*, 2006). While there are several differences between the modelling algorithms, all niche modelling methods are frequently clouded by several assumptions and uncertainties that hamper their credibility.



## 1.1 Assumptions and uncertainties of ecological niche-based models

The assumption that climate wields control over the distribution of species is supported by evidences from observed trends and fossil records (Soberón & Peterson, 2005). The bioclimatic envelop that limit the multidimensional space where the species is potentially able to maintain populations is defined as ecological niche (Hutchinson, 1957). These ecological requirements are divided into fundamental and realized ecological niches. The fundamental niche corresponds to the geographic region with appropriate set of abiotic factors that impose physiological limits on species' ability to persist in an area (Peterson *et al.*, 1999). The realized niche incorporates abiotic conditions, the effects of interactions with other species (biotic factors), and competitive exclusion (Guisan & Zimmermann, 2000; Soberón & Peterson, 2005). Thus, species will be geographically present where both abiotic and biotic factors are positively combined with area that is within the dispersal capabilities of the species (Soberón & Peterson, 2005; Peterson, 2011; Fig. 1). ENMs assume the concept of ecological niche conservatism (Peterson *et al.*, 1999). Whereby, a relationship between species occurrence and a series of environmental variables is established to predict present distributions of species and projected them to future climates scenarios (Araújo *et al.*, 2005a). Thus, ENMs often discard adaptability and ability of species to endure in refugia, since they are based in the assumption that the ecological niche of species does not change throughout time (Wiens & Graham, 2005). It is known that species range expanded and contracted due to climatic changes throughout the past, adjusting it to the changing environmental conditions, with potential of adaptation to suboptimal future environmental conditions. Therefore, to increase the accuracy of predictions of range shifts, it is important to first identify past refugia. The importance of refugia is recognized (Bennett & Provan, 2008; Ashcroft, 2010) and it is widely used to predict suitable areas that should be conserved to limit the impacts of climate change (Barnosky, 2008; Williams *et al.*, 2008; Rull, 2009). The concept of ecological niche conservatism has been source of discussion for several years. Peterson *et al.* (1999) proposed the concept of ecological niche conservatism, demonstrating that ecological niches show considerable conservatism over evolutionary time periods. Even though recent and short-term events show considerable evidence towards conservatism, long-term events show breakdown of conservatism (Peterson, 2011). Although niche conservatism is widespread, there are methodological limitations. One of these limitations is the environmental data set used to infer niche shifts (Peterson, 2011). Sampling the complete climatic niche where a species can thrive is considered one of the main factors affecting the reliability of ENMs predictions (Maiorano *et al.*, 2013). It its known that longstanding human-induced factors have advocate extinctions in the past, that often

result in reductions of historical distribution (Pimm *et al.*, 2006; Bellard *et al.*, 2012). It was demonstrated that range contraction has an important influence on the predictions of suitable climatic space and species vulnerability trend under climate change scenarios, as it reduces the climatic variability associated with the species niche (Martínez-Freiría *et al.*, 2016). Future estimates are drastically underestimated unless the full historical distribution of the species is included in ENMs (Martínez-Freiría *et al.*, 2016; Faurby & Araújo, 2018). If the realized niche and fundamental niche do not coincide entirely, we cannot hope for any modelling algorithm to characterize the full fundamental niche of a species (Phillips *et al.*, 2006). Thus, forecasts of climate change impacts on biodiversity for many species are unlikely to be reliable without acknowledging past anthropogenic influences on contemporary ranges.

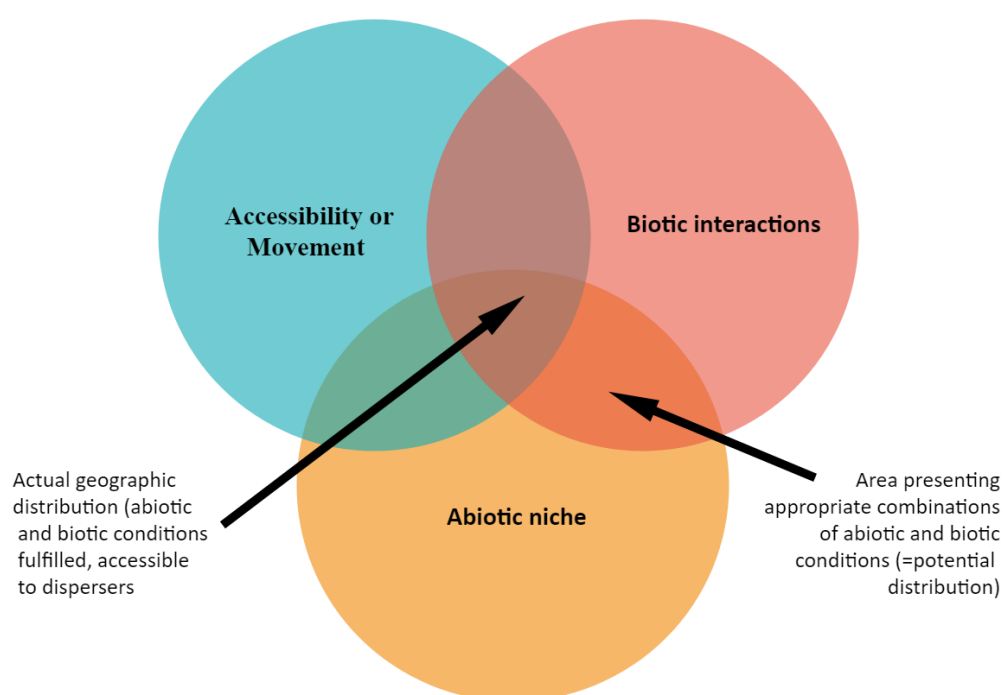


Figure 1 - The 'BAM diagram', showing a simplified framework for understanding where species will be distributed. The combination of biotic factors, abiotic factors and movement factors delimit geographical distributions of species. Adapted from Peterson *et al.* (2011).

Using correlative ENMs to project future distributions assumes that the variables included in models do in fact reflect the niche of a species. Determining all the factors that influence a species niche is an impossible task, and the possibility that unaccounted niche factors may influence the observed distribution has hampered ENMs (Wiens *et al.*, 2009). Utilizing environmental correlates of a specie's distribution to project its future occurrence assumes as well that the current distribution is in equilibrium, where suitable habitat is fully occupied (Soberón & Peterson, 2005). However, suitable places may not be occupied by

several reasons : (i) if recent disturbances such as poaching or habitat destruction wipe out a species from a landscape, (ii) if a species is expanding into areas that have only recently become available, or (iii) if regional population density is not adequate to support colonization of suitable zones (Pulliam, 2000). Moreover, most ENMs assume that each species responds independently to the environmental factors that determine its niche space, and thus its distribution. Therefore species interactions are generally not included in ENMs (Wiens *et al.*, 2009). Another assumption is that individuals will be able to disperse to suitable locations (Pearson & Dawson, 2003). However, if environmental conditions shift more rapidly than species can disperse, they may be forced to persist only in isolated habitat refugia that meet their niche requirements (Loarie *et al.*, 2008). The ability of individuals to disperse to suitable places is not entirely dependent on their inherent dispersal capacity. In many cases the landscape through which species must move to reach suitable places has been fragmented by human actions. This fragmentation breaks habitat connectivity, creating barriers to dispersal and a landscape mosaic of suitable, less suitable, and unsuitable habitat patches (Wiens *et al.*, 2009).

Since ENMs carry plenty of assumptions, several uncertainties are to be expected, mainly derived from: i) input data biases or lack of data; ii) modelling features such as types of algorithms, threshold values and number of pseudo absences; iii) the complexity of natural systems dynamics such as species dispersal ability; iv) reliance on different General Circulation Models (GCM; numerical models representing physical processes in the atmosphere, ocean, cryosphere and land surface) parameters and incorporate different functions to portray the dynamics of atmospheric circulation and ocean effects, they may project different consequences; and v) scale related uncertainties, ranging from distributions not matching environmental factors closely because behavioural interactions may override habitat selection at a very fine resolution, to only the most general environmental relations emerging at coarser resolutions (Barry & Elith, 2006; Elith & Leathwick, 2009; Wiens *et al.*, 2009). All models are sensitive to the quantity and quality of the input data, and so are ENMs. The reliability of occurrence records of species used to create ENMs depends on several factors such as the comprehensiveness of survey coverage, potential biases in recording presences/absences, or the skill and knowledge to identify the species (Scott, 2002). Incomplete spatial coverage or small sample curtails the statistical assurance of correlations and increases the uncertainty of extrapolating distributions to broader areas (Hernandez *et al.*, 2006). Moreover, some surveys collect only presence data, whereas others also document absences. Incorporating absence data may strengthen a ENM because where a species is absent can reveal as much about its niche as where it is present, but “absence” can be a true absence or a failure to record a species that is actually present (Wiens *et al.*, 2009). Absence data is usually poorly

available (specially in poorly sampled areas like remote regions), and it is easily misidentified (Ponder *et al.*, 2001; Phillips *et al.*, 2006) adding another layer of uncertainty to ENMs. Given the rarity of unbiased absence data, researches sometimes resort to modelling techniques that require only presence observations as input data. BIOCLIM (Busby, 1991) for example, has been developed for developing models based in presence-only data. Nonetheless, some pitfalls applied to presence-absence dataset models may also affect the accuracy of presence-only modelling. For instance, the number of occurrence localities used to estimate the parameters of a model is frequently too low, compromising reliability (Phillips *et al.*, 2006). Moreover, comparisons between several modelling methods show that presence–absence models usually perform better than presence-only models (Elith *et al.*, 2006). Hence, models that utilize both presence and absence data are increasingly used when only presence data is available, by creating artificial absences (e.g. pseudo-absences or background points ; Barbet-Massin *et al.*, 2012). Different strategies have been proposed to improve the selection of an appropriate pseudo-absence dataset. Some studies have suggested using pseudo-absence data selected outside a predefined region based on a preliminary model or based on a minimum distance to the presence (Engler *et al.*, 2004; Lobo *et al.*, 2010; Martínez-Freiría *et al.*, 2016). Additionally, the ratio between the quantity of presence data to the quantity of absence data (usually defined as prevalence) used to fit the model has also been shown to influence model accuracy (Mc Pherson *et al.*, 2004). Some studies have provided recommendations and guidelines on the amount of pseudo absences or background points to utilize for better accuracy (Philipps & Dudík, 2008; Barbet-Massin *et al.*, 2012; Liu *et al.*, 2019). However, it is complicated to obtain a general strategy that will perform better for all different taxa and scenarios. Thus, it is wise to incorporate different strategies to generate artificial absences when the implications of the strategy used on the modelling process are unknown.

Inconsistencies or lack of climatic input data is also a significant source of uncertainty that influence model predictions (Soria-Auza *et al.*, 2010). In the last decades, there has been an increasing production of global bioclimatic databases. WorldClim, one of the most popular among researchers, provides interpolated data layers, consisting of long-term average monthly temperature and precipitation values, based on observations from weather stations across the world (Hijmans *et al.*, 2005). Paleoview is another source of climatic layers, which provides a comprehensive way to generate and visualise paleoclimatic data (Fordham *et al.*, 2017). Regional to global scale simulations of temperature, precipitation, humidity and mean sea level pressure can be generated from PaleoView as gridded or time series data for any period during the last 21,000 yr. Modelled

climate reconstructions are based on daily simulation output from the Community Climate System Model ver. 3 (CCSM3). Nonetheless, the use of climatic layers based on interpolated methods is specially problematic in areas with deficient or unevenly distributed network of weather stations, such as tropical forests, desert regions, or mountainous areas (Hijmans *et al.*, 2005; Yi *et al.*, 2011; Fordham *et al.*, 2017). Given these issues, the selection of appropriate bioclimatic data becomes a complex task in distribution modelling. It is important to consider the limitations, and not just assume that the use of a single database is optimal for all regions and scales (Yi *et al.*, 2011; Ashouri *et al.*, 2016). Nonetheless, the criteria utilized for choosing a specific bioclimatic database and its effect on the models and projections have been a rather neglected topic, and only few studies have addressed it (e.g. Bobrowski *et al.*, 2017; Morales-Barbero & Vega-Álvarez, 2019).

## 1.2 Ensemble forecasting

Due to the variability in predictions of ecological niche models according to the method utilized, and since none of the widely used techniques performs universally better than the others (Elith *et al.*, 2006), simultaneously applying several methods within a consensus modelling framework (ensemble model) has been recommended (Buisson *et al.*, 2010). The concept of ensemble modelling is that each individual model carries both some true “signal” about the relationships the model is trying to capture, and some “noise” created by errors and uncertainties. Ensembles combine models with the intention of obtaining better separation of the signal from noise (Araújo & New, 2007). The concept of ensemble is widely used in machine learning, often with complex classifiers built by combining many simple modelling units. The ensemble forecast framework aims to consider the central trend of several niche models, using different methodologies (Marmion *et al.*, 2009), and is currently widely used by ecological niche modellers, often with the same use of pseudo-absences across the different models utilized. Nonetheless, it was shown that the optimal method of generating and using pseudo-absences information differs widely across modelling techniques, thus, the optimal way of utilizing pseudo-absences through an ensemble forecast technique could therefore be to use different pseudo-absence selection strategies (Barbet-Massin *et al.*, 2012). However, most ensemble forecast techniques compare model accuracy either to select the best models or to weight their predictions differently, which can only be done in an unbiased way if the same data were used for all models. Grouping models that share the same way of optimising the usage of pseudo-absences, comparing their model accuracy and selecting the best from each group can potential overcome this problem (Barbet-Massin *et al.*, 2012). Various strategies exist to combine predictions from individual models into an ensemble, the most intuitive of which

is simply taking the mathematical mean or median across predictions, independently of whether such predictions are binary or continuous. More complex approaches involve “weighting”, where predictions of different models are scaled by weights based on some measure of predictive performance (Araújo & New, 2007). These weights are often derived by validating predictions from individual models on some test data. Weighting is thought to improve how well the ensemble predicts (Araújo & New, 2007), although weighted ensembles also require more effort to produce as the individual models need to be validated before they can be combined. Most studies that have explored several models in ensemble framework have mainly attempted (i) to identify the ‘best’ modelling performance among extensive model comparisons; or (ii) to use a consensus approach that summarizes the variability within the ensemble of predictions; (Buisson *et al.*, 2010). Variability in projections of such ensembles, however, has rarely been considered itself, although calls have arisen to improve knowledge of the uncertainty factors that may decrease the reliability of predictions (Barry & Elith, 2006). In the context of climate change, uncertainty in projections becomes even more worrying as additional sources of variability arise at two levels with the use of future climate scenarios (Araújo & New, 2007). Very little attention has been focused on quantifying the variability between the projections obtained from ensembles combining several sources of uncertainty (Buisson *et al.*, 2010). Quantifying the effects of the different sources of uncertainty is crucial, so management and conservation decisions can be effective and taken with knowledge of predictions reliability.

### 1.3 Transferability of ecological models

One of the main interests of researchers in ENMs dwells in their capacity to predict species responses in unsampled spatial or temporal scenarios. Researchers have investigated whether a model developed for a specific region can successfully predict species occurrence in a distinct region (Randin *et al.*, 2006; Peterson *et al.*, 2007), as well as if models developed for a specific time period can predict species occurrence in a distinct temporal scenario with different environmental conditions (Araújo *et al.*, 2005b; Tuanmu *et al.*, 2011). This notion of the cross applicability of models in both space and time has been coined as transferability and it is now widely used to expand the potential of ENMS (Randin *et al.*, 2006). Model transferability is currently an important branch of predictive science that has stemmed mainly from the need to produce ecological forecasts (Bouchet *et al.*, 2019) in face of rapid environmental changes. The transferability of the model to other periods of time allows: (i) the evaluation of the effects that climatic changes have had or will have on species ranges; (ii) to anticipate the effects of the global

environmental changes and its effects on species distributions; and (iii) to pre-emptively design conservation plans (Werkowska *et al.*, 2016). The transference of models to future climate change scenarios was the area that stimulated advances in both conceptual and analytical strategies of ENMs. However, it is impossible to adequately assess model predictions for future scenarios due to the lack of data (Araújo *et al.*, 2005b). Thus, predictions should be considered with relative caution since several sources of uncertainty emerge at each step (Wiens *et al.*, 2009) and proliferate throughout the modelling operation. The transference of a model is not a forthright exercise and requires modellers to handle a conceptual background as well as methods to avoid bias in the results (Werkowska *et al.*, 2016). Transferability should be greater in ENMs fitted to observations that document all dimensions and constraints placed upon the fundamental niche. However, most datasets do not meet this requirement, because species do not always occupy all suitable habitats. While fundamental niches can be expected to stay constant over timescales relevant to management, realized niches will typically vary both spatially and temporally (Yates *et al.*, 2018). This complicates model transfers, particularly when the realized niche becomes a direct function of habitat selection behaviour as it relates to resource availability or tolerance limits (Matthiopoulos *et al.*, 2011). Failures in model transfer can occur if these are calibrated with data do not extrapolate well to novel data (Moreno-Amat *et al.*, 2015). Moreover, unintentional stochastic events in the evaluation data can also result in poor transferability. Overfitting a model often results in loss of model transferability as well. Overfitting may occur for two rather different reasons: (i) weak correlations among variables arise as a result of random noise, and these may be incorrectly interpreted as legitimate relationships; and (ii) when there are statistical associations between predictor and response variables that are real in a given data set but do not occur under a wide range of conditions (Wenger & Olden, 2012). Moreover, there are several fundamental challenges to be surpassed in model transferability in order to achieve more accurate and reliable predictions, such as: (i) which variables make models more or less transferable; (ii) knowing whether models are trait or taxon-specific; (iii) understanding to what extent data quality influences model transferability; (iv) how to sample for an more optimized model transferability; and (v) understanding how the different model algorithms influence model transferability (Yates *et al.*, 2018).

#### 1.4 Validation of model's transferability

For evaluating model transferability it is imperial to have appropriate metrics of prediction accuracy and precision (Fourcade *et al.*, 2018). However, there is still little to no consensus on which metrics are the most appropriate (Randin *et al.*, 2006; Sequeira *et al.*,

2018). Moreover, validation is only possible with independent data, which is often unavailable or insufficient (Yates *et al.*, 2018). In the absence of validation data for a target area, transferability can only be estimated by contrasting predictions with existing expert knowledge or simulations, or benchmarking performance by projecting models into multiple alternative scenarios for which enough data is available (Sequeira *et al.*, 2018). Cross-validation can provide a reasonable approximation of independence, so long as it can be structured to mimic prediction conditions and minimize correlations (Wenger & Olden, 2012). Ultimately, consistent assessments of transferability will require unified and widely applicable standard metrics that enable direct comparisons among studies, systems, and taxa (Sequeira *et al.*, 2018). Nonetheless, approaches for evaluating ecological models do not provide inference into transferability often. As a result, the transferability of a model is often unknown, and the model selected as “best” for a data set may have worse transferability than a rejected one (Wenger & Olden, 2012). Novel approaches for model transferability validation that are independent of model choice and response variable type are of most importance if we want future conservation decision making to be effective in face of an increasingly rapid changing world, especially when resources available for ecological research are usually thin.

Model transferability testing can be a powerful method for model evaluation (Randin *et al.*, 2006). Independent observations from a training data set has been recommended as a more proper method for validating models predictions (Guisan & Zimmermann, 2000). Thus, the use of a temporarily independent testing data set is recommended to assess the temporal transferability of niche model techniques (Boyce *et al.*, 2002; Araújo *et al.*, 2005b). Moreover, how the distribution map relates to reality data, especially in under sampled areas can be used as a metric to assess model performance and generalization. Arguably, if model predictions perform very well there, great progress is provided and usually cost-effectively (Mi *et al.*, 2017). Even though model validations on independent datasets have suggested that in terms of both sensitivity (the proportion of observed positives that were predicted to be positive) and specificity (the proportion of observed negatives that were predicted to be negatives), some of the newer machine-learning and tree-based algorithms have the best overall performance (Thuiller *et al.*, 2003; Elith *et al.*, 2006), the risk of hampering transferability if model complexity is not constrained is a reality in these models (Tuanmu *et al.*, 2011). Assessment of transferability of models is critical if they are to be used in a predictive manner beyond the conditions under which they were trained (Wenger & Olden, 2012). Moreover, variability in projections resulting from the multiplicity of approaches available difficult a clear picture of the future of biodiversity under different scenarios of climatic change (Pereira *et al.*, 2010). Determining how to provide



less unbiased, more accurate and robust predictions is of most importance. However, such predictions are complicated to validate as consequence of lack of reliable presence data for the periods of model projection. Data from museum specimen and published literature are common sources to assess model transferability performance (Graham *et al.*, 2004). Nonetheless, rock art can be found in most regions of the world and has an untapped potential to provide evidence on paleo-distributions and biogeography, particularly on areas or periods where faunal records are scarce (Guagnin *et al.*, 2018). Prehistoric human populations throughout Africa have documented many of the species presence in the environment during the Holocene through paintings and engravings (e.g. Messerli & Winiger, 1992; Masseti, 2010; Guagnin, 2014; Fig. 2). Rock art shows a clear preference for large size vertebrates, which are highly visible in the landscape, particularly large size herbivores, and typically only essential lines (body outline, head, legs, horns and occasionally tails) are depicted (Guagnin, 2014). Therefore, animals such as elephants and giraffes that have a very distinctive body shape are very often depicted and easily identified. Moreover, patterns on the coat of the animals and heard compositions are often depicted and can be used to aid species identification (Guagnin *et al.*, 2018). Many of the engravings also show behavioural traits or movement of the animals. Giraffes, for example are often depicted with their tail curled up, which is typical for running giraffes, and several elephant depictions show elephants in musth (condition in male elephants, characterized by highly aggressive behaviour, and often secrete a thick-tar secretion called temporin from the temporal ducts on the sides of the head; Guagnin, 2014). This suggests that engravers were, at least to some extent, familiar with the depicted species (Guagnin *et al.*, 2018). Some studies have been successful in identifying species depicted in the engravings (Gautier & Muzzolini, 1991; Guagnin *et al.*, 2015, 2018). The availability of such presence records provides a unique opportunity to test the robustness of model projections at Holocene, based in current models.

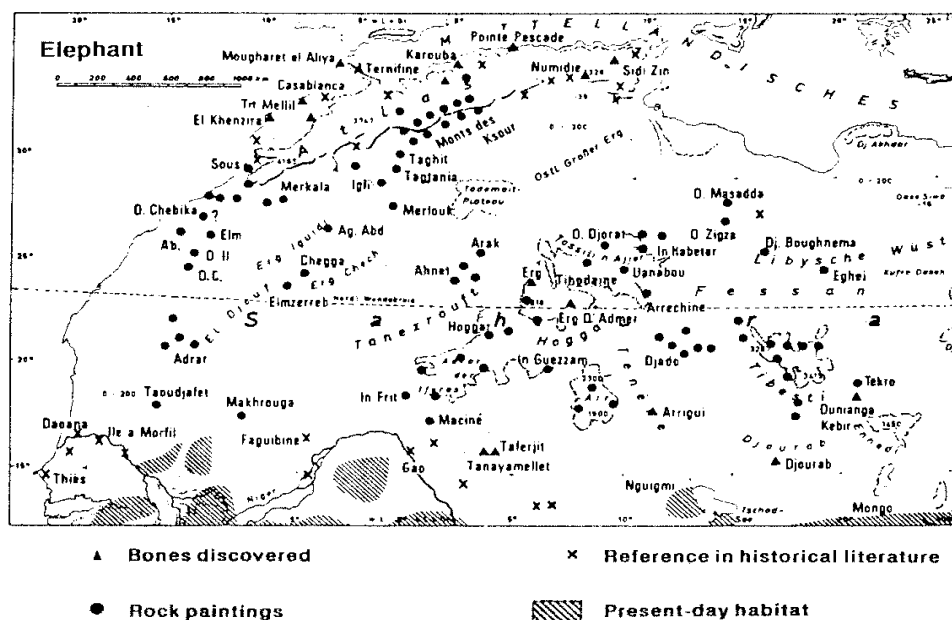


Figure 2 – The distribution of the Savannah elephant in North Africa during the Holocene. Adapted from Messerli & Winiger (1992).

## 1.5 Species Niche Contraction

Africa is a vast continent with an overwhelming amount of biodiversity, but its declining at a fast pace as human populations grow and invade natural habitats. Forecasts of changes in climate projected to Africa are above the global average and the combination with illegal wildlife poaching and habitat fragmentation will likely result in significant range contractions of species and eventually extinctions (Pimm, 2008; Garcia *et al.*, 2012). Moreover, the uncertainties around the predictions of future changes in rainfall make the African continent highly vulnerable to the climatic changes we currently face (Garcia *et al.*, 2012). With the beginning of the 20th century and the increasing human pressure (e.g. land-cover change, poaching), large-sized species range contracted drastically, and many peripheral populations were extirpated (Lindsey *et al.*, 2013). The African savannah elephant (*Loxodonta africana*; hereafter elephant) and the giraffe (*Giraffa camelopardalis*) are emblematic examples of species that have suffered significant habitat loss due to anthropogenic pressures, particularly since the 1970s, resulting in significant contractions of their distributional ranges (Blanc, 2008; Fennessy & Brown, 2010; Fig. 3).

The Elephant is a "keystone" species that plays a crucial role in structuring both plant and animal communities (Blanc *et al.*, 2007). The distribution ranges from the sub-Saharan regions to the south of the African continent, excluding the African tropical forest region (Comstock *et al.*, 2002). Nonetheless, in the past century, it has become increasingly fragmented and it is known to have become extinct in multiple countries (Blanc *et al.*,

2007). West Africa elephant populations have been particularly extirpated, and the distribution has been reduced to tiny scattered fragments (Blanc *et al.*, 2007). Elephants have been the focus of intense conservation studies and debate for a long time, particularly as their numbers have decreased by nearly 50% in the 1980s (Comstock *et al.*, 2002). This species has a vast history of human disturbance. Europe had imported around c. 100–200 tons of ivory per year and by the late 19th century, European ivory imports may have reached 700 tons, representing a potential 60,000 elephants killed per year (Naylor, 2004; Lee & Graham, 2006). During the last century, the availability of modern weapons and road development has made it far easier for people to kill large mammals and by the middle of the 1970s a pick of ivory demand threw elephant populations to instability (Lee & Graham, 2006). Furthermore, armed conflicts coupled with political instability add an extra layer of pressure to populations already repressed by poaching and other anthropogenic activities (Lindsey *et al.*, 2013; Brito *et al.*, 2018).

Giraffes formerly occurred in a wide region across the sub-Saharan Africa but currently the range has drastically contracted and fragmented, especially in West Africa, to small and scattered populations (Fennessy & Brown, 2010). Since the beginning of the 20th century, the distribution and density of giraffe populations has drastically decreased across the entire African continent to a total of 140,000 by the late 1990s and nowadays to a mere 80,000 (Suraud *et al.*, 2012). The distribution of the giraffe until the beginning of the 20th century has essentially depended on the suitable vegetation, permanent water sources, as well as climate. The range extended throughout most northern Africa (Ciofolo, 1995; Suraud *et al.*, 2012). By the end of the 19th century, the giraffe was present across the Sudano-Sahelian zone from Chad to Senegal and only natural geographical barriers kept the species from colonize new areas (Ciofolo, 1995). The turn of the century was the starting point of the species decline. New advancements in firearms allowed extensive and more effective hunting coupled with the developments of agriculture, the deforestation, and the construction of railroads led to population declined and range contraction, being now extinct in most places that otherwise would not be (Fennessy & Brown, 2010).

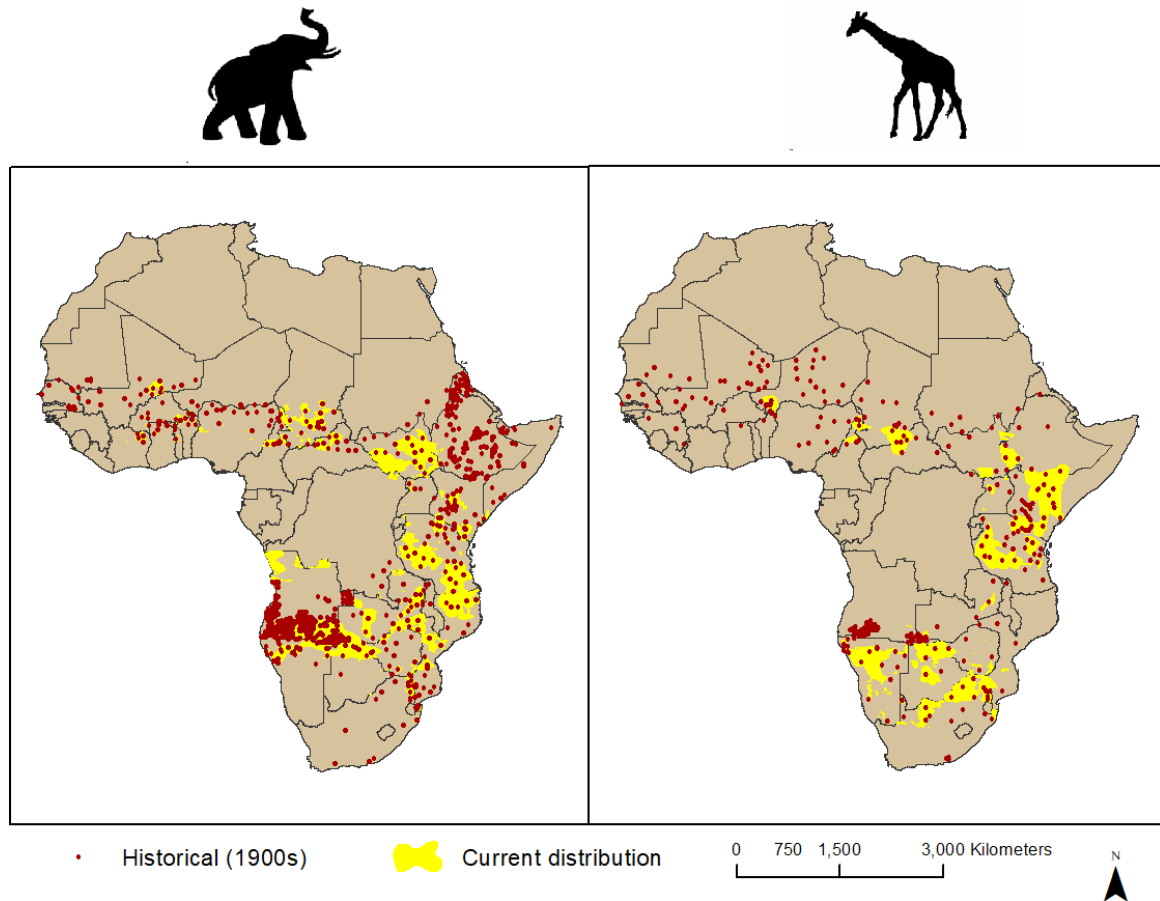


Figure 3 – Distribution at about 1900s (points) and in current time (polygons) of the African savannah elephant (*Loxodonta africana*) and the Giraffe (*Giraffa camelopardalis*) in Africa. Historical distributions adapted from Martínez-Freiría *et al.*, (2016) and current distributions adapted from (Blanc, 2008) and (Muller *et al.*, 2018). The range of Savannah elephant excludes observations of the African forest elephant (*Loxodonta cyclotis*) based on (Roca *et al.*, 2001). The range of Giraffe lumps the observations of the four giraffe species recently described (Fennessy *et al.*, 2016; Winter *et al.*, 2018; O'connor *et al.*, 2019).

## 2. Objectives

This work aims to successfully map the Holocene distribution for the two target species of this study in order to understand how different input data influence ENMs predictive performance and affect their ability to predict suitable areas in different temporal scenarios. Specifically it is aimed to test for differences in distinct: i) strategies to select pseudo-absence data (minimum distance to the presence data, random data, and restricted data); ii) sources of climatic layers for three different time periods (Current, 6000BP, 7900BP) from different sources (Worldclim, Paleoview); and iii) land-cover and climate-based models for the same three time periods. Models will be developed using historical and past (Holocene) presence records from two case-study species, the African savannah elephant (*Loxodonta africana*) and the Giraffe (*Giraffa camelopardalis*). To evaluate each modelling strategy and validate their transferability it will be used a novel approach resorting to rock-art depictions. The results produced by this study are expected to aid researchers selecting appropriate data and modelling methods that can produce more accurate and robust predictions. Hopefully this study may offer new insights and guidelines for the model's methodological approach.

## 3. Material and Methods

### 3.1 Study species

The African savannah elephant (*Loxodonta africana*) and the Giraffe (*Giraffa camelopardalis*) were used as study species. The range of the elephant considered in this study (*Loxodonta africana*) excludes the one of the African forest elephant (*L. cyclotis*) given that they exhibit very different ecological niches; the latter is exclusively present in tropical forest regions while the former is absent from those areas (Roca *et al.*, 2001).

The giraffe considered in this study lumps the four giraffe species, *Giraffa camelopardalis*, *G. tippelskirchi*; *G. giraffa*; *G. reticulata*, recently described (Fennessy *et al.*, 2016; Winter *et al.*, 2018). It was assumed that the ecological niches of these four species are very similar, thus lumping their ranges (O'Connor *et al.*, 2019) into a single taxonomic unit will probably not affect ecological models. Furthermore, allocating the rock-art depictions to any of the current four described giraffe species is an impossible task since phenotypically they are very similar and impossible to distinguish in rock-art depictions.

### 3.2 Study area

To include the global distribution of both model species, the study area encompasses the full African continent, excluding islands.

### 3.3 Species observations

Two datasets were created for each target species, named "Current" and Holocene datasets. The "Current" dataset included both contemporary and historical (early 1900's) observations and was retrieved from Martínez-Freiría *et al.*, (2016). Combining contemporary and historical observations was done in order to account for the complete ecological niche of both species, since the ranges of elephant and giraffe artificially contracted drastically over the last century (due to poaching). Discarding this range contraction would lead to significant underestimation of the current ecological niche, as demonstrated by Martínez-Freiría *et al.*, (2016). The Holocene dataset included observations from ~6,000 yr ago and was generated based in rock-art (Fig. 4) collected from numerous publications with sufficient details to allow species identification with confidence (Breuil, 1923; Gautier, 1935a,b; Bagnold *et al.*, 1939; Murray, 1951; McHugh, 1974; Hall, 1976; Simoneau, 1976; Sayer, 1977; van Lavieren & Esser, 1979; Petit-Maire *et al.*, 1983; Tillet, 1985; Fuchs, 1989; Redford & Redford, 1989; Messerli & Winiger, 1992;

Allard-Huard, 1994; van Albada, 1994a,b,c; Drews, 1995; Ciofalo, 1995; Cremaschi, 1996; Le Pendu *et al.*, 2000; Van der Jeugd & Prins, 2000; De Leeuw *et al.*, 2001; Bond & Loffel, 2001; Masferrer *et al.*, 2001; Ciofalo & Le Pendu, 2002; Parker & Bernard, 2005; Cameron & Du Toit, 2005; Crawford-Cabral & Veríssimo, 2005; Hassanin *et al.*, 2007; Fennessy, 2009; Leroy *et al.*, 2009; Shorrocks & Croft, 2009; Brenneman *et al.*, 2009; Fennessy & Brown, 2010; Bouché *et al.*, 2010; Poilecot *et al.*, 2010; Drake *et al.*, 2011; Nasser *et al.*, 2011), the British Museum (<https://africanrockart.britishmuseum.org/>), and from the Global Biodiversity Facility ([www.gbif.org](http://www.gbif.org)). When available, the coordinates of point localities were extracted and inserted in a database. When coordinates were unavailable, they were estimated in Google Earth using locality descriptions. All observations were georeferenced in the WGS84 projection system. To reduce biases due to spatial autocorrelation in the full set of observations, independent observations for the training dataset were selected from clusters of species occurrence in a grid of 2.5 degrees.

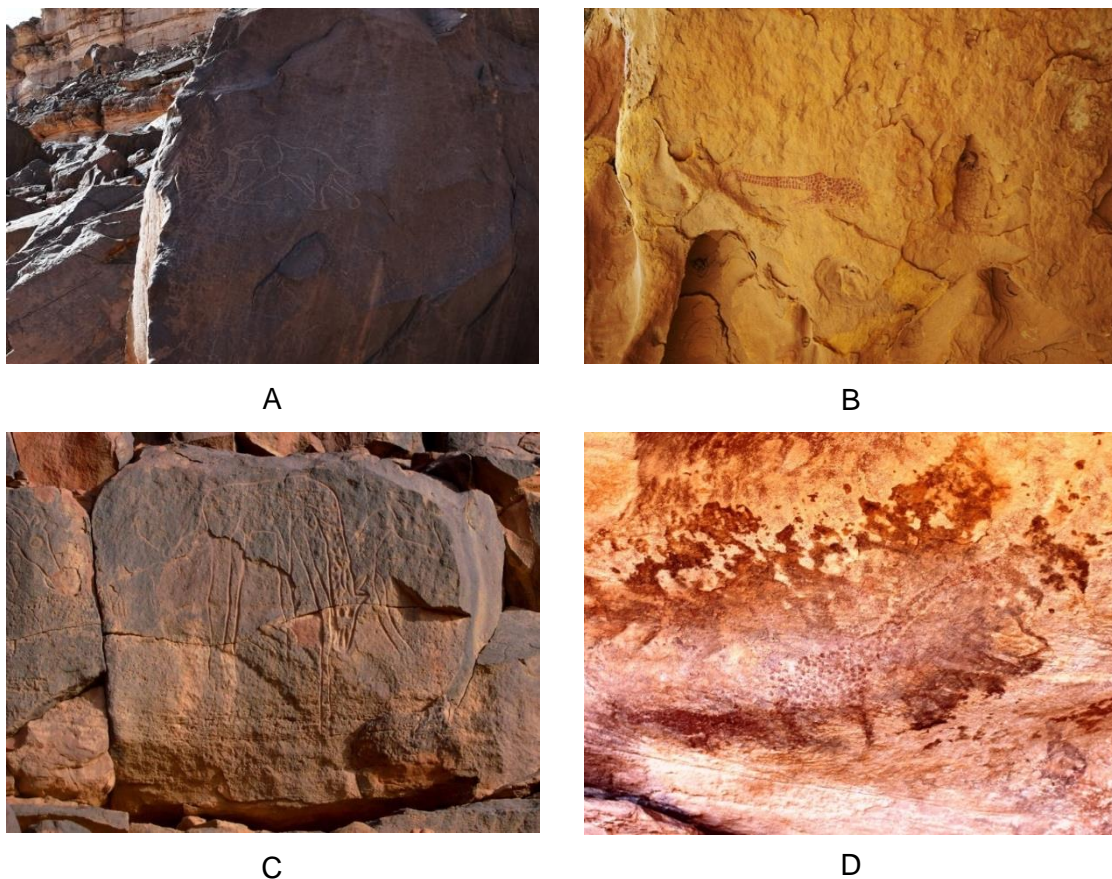


Figure 4 – Rock-art depicting Elephant in Libya (A), and Giraffe in Algeria (B), Libya (C), and Mauritania (D). Photos by J.C. Brito

## 3.4 Environmental variables

### 3.4.1 Bioclimatic variables

To account for probable climate changes occurred in Africa during the species observational records periods, bioclimatic datasets were created for three different time periods (Table 1): Current, 6,000BP and 7,900BP. The current time period considered month-by-month variations in the climate over the period 1960-1990 and data were obtained from Worldclim v1.4 (Hijmans *et al.*, 2005) and Paleoview (Fordham *et al.*, 2017) databases. Spatial upscaling at a resolution of 2.5 degrees was performed for the Worldclim month-by-month data using the Geographical Information System software ArcGIS 10.5v (ESRI, 2010). Upscaling the spatial resolution of Worldclim data was done to match it with the resolution provided by Paleoview, using the bilinear technique (more suitable for continuous data) provided by the Resample tool of ArcGIS. The 6,000BP and 7,900BP time periods corresponded to different phases of the Holocene and considered month-by-month predictions for precipitation and temperature (minimum and maximum) at 100-year intervals taken in 100-year steps, and data were obtained from Paleoview for the single GCM (CCSM3) available. Two bioclimatic variable datasets for different Holocene time periods were used since the dating of the rock-art depictions was uncertain and ranged from the early to mid-Holocene.

The only GCM (CCSM3) provided by Paleoview is not provided by Worldclim. As such, to be able to compare results obtained by models constructed with Paleoview and Worldclim data, it was reconstructed the Holocene bioclimatic variables based in CCSM3 data. Anomalies for both Holocene periods were calculated using ArcGIS 10.5v by subtracting to each pixel the month-by-month climatic predictions obtained from Paleoview to its baseline, following the equations (Ramirez-Villegas & Jarvis, 2010):

$$Anomaly_x = X_{Holocene} - X_{Baseline}$$

$$\text{Temperature: } X_{Holocene} = X_{Baseline} - Anomaly_x$$

$$\text{Precipitation: } X_{Holocene} = X_{Baseline} \times \left| 1 + \frac{Anomaly_x}{X_{Baseline} + 1} \right|$$

These anomalies calculated for 7,900 and 6,000BP were then added independently to Worldclim baseline to obtain month-by-month data for both Holocene periods. For the temperatures (minimum and maximum), the anomalies in degree Celsius were simply added to the value (in degree Celsius) reported in Worldclim baseline. For the precipitation, one millimetre was added to the denominator to account for areas where current precipitation is equal to zero. The absolute value of change relative to the Worldclim



baseline was used in order to avoid negative monthly precipitation values and to maintain homogeneities (Ramirez-Villegas & Jarvis, 2010).

A total of 19 bioclimatic variables for both climatic data sources and for the tree time periods (Figs. 5 to 10) were constructed using the “dismo” package (Hijmans *et al.*, 2015) on R software (Team, 2019). Correlations between bioclimatic variables were tested for both baselines, and nine bioclimatic variables for each dataset were retained (Pearson’s  $R < 0.7$ ; Table S1).

Multivariate Environmental Similarity Surfaces (MESS) were constructed using MaxEnt software v3.4.1 (Phillips *et al.*, 2019) to assess similarities between bioclimatic variables from the current time period and both Holocene periods, and to identify the most dissimilar variables among time periods. In general, climatic variables used for projections between time periods are within the range of the ones used for model training (Fig. S1, S2). The Sahara-Sahel is the region where bioclimatic variables were the most dissimilar between time periods. This is expected since the climatic conditions in these regions during the Holocene are reported to be significantly different from current climate, especially for the Sahara Desert, which during the Holocene appeared to be covered by xerophytic woods/scrubs and grass, while the eastern part remained desert (Claussen & Gayler, 1997).

Table 1 – Range values (minimum - maximum) of each bioclimatic variable in each climate dataset (Paleoview and Worldclim) in each time period (current, 6,000BP, and 7,900BP). Legend: °C – degree Celsius; mm/d – millimetres per day; dim - dimensionless

Bioclimatic variable (code)	Units	Paleoview			Worldclim		
		Current	6,000BP	7,900BP	Current	6,000BP	7,900BP
Annual Mean Temperature (BIO1)	°C	14.35 - 32.87	13.50 - 30.54	13.57 - 30.50	13.29 - 29.59	10.50 - 28.65	9.93 - 27.72
Mean Diurnal Range (BIO2)	°C	1.66 – 41.68	1.67 – 40.69	1.68 – 40.92	5.80 – 19.46	5.56 – 18.54	5.51 – 18.49
Isothermality (BIO3)	°C	31.12 – 85.06	26.78 – 90.09	25.51 – 88.89	29.98 – 88.36	25.78 – 87.45	25.61 – 86.12
Temperature Seasonality (BIO4)	dim	47.48 – 831.55	46.09 – 930.54	50.60 – 993.33	31.70 – 906.41	36.44 – 1018.35	37.79 – 1074.93
Max Temperature of Warmest Month (BIO5)	°C	27.82 – 61.19	26.60 – 61.66	26.21 – 62.59	21.70 – 48.50	19.67 – 49.72	19.36 – 50.98
Min Temperature of Coldest Month (BIO6)	°C	-8.44 – 23.91	-9.20 – 22.82	-9.06 – 22.81	-1.70 – 22.20	-2.63 – 20.89	-3.03 – 20.50
Temperature Annual Range (BIO7)	°C	5.35- 60.19	6.08 – 63.54	5.71 – 65.61	9.70 – 42.70	9.96 – 45.98	9.90 – 47.43
Mean Temperature of Wettest Quarter (BIO8)	°C	13.83 - 36.85	10.36 - 34.52	9.82 - 34.05	9.37 - 36.88	6.57 - 37.74	5.60 -38.20
Mean Temperature of Driest Quarter (BIO9)	°C	12.06 - 32.77	11.4643 - 32.27	10.52 - 32.33	8.05 - 33.97	7.25 - 33.61	7.53 - 34.12
Mean Temperature of Warmest Quarter (BIO10)	°C	18.67 - 36.85	18.26 - 34.91	17.86 - 34.57	15.92 - 37.50	12.10 - 37.81	11.74 - 38.47
Mean Temperature of Coldest Quarter (BIO11)	°C	11.17 - 29.71	9.77 - 29.00	9.27 - 29.23	6.72 - 27.20	5.01 - 27.08	4.50 - 27.24
Annual Precipitation (BIO12)	mm/d	3.06 – 3065.19	7.47 – 2970.40	8.95 – 2905.54	0.00 – 3075.00	0.00 – 2889.57	0.00 – 2823.73
Precipitation of Wettest Month (BIO13)	mm/d	0.88 – 402.54	1.83 – 436.56	1.55 – 429.23	0.00 – 108.00	0.00 – 536.31	0.00 – 514.78
Precipitation of Driest Month (BIO14)	mm/d	0.00 - 191.21	0.00 - 172.71	0.00 - 174.02	0.00 – 108.00	0.00 - 96.71	0.00 - 92.43
Precipitation Seasonality (BIO15)	dim	18.75 - 163.58	19.79 - 195.05	19.00 - 214.63	0.00 - 207.17	0.00 - 264.25	0.00 - 254.42
Precipitation of Wettest Quarter (BIO16)	mm/d	1.38 – 1102.81	3.85 – 1025.61	3.21 – 1009.25	0.00 – 1504.00	0.00 – 1519.46	0.00 – 1459.72
Precipitation of Driest Quarter (BIO17)	mm/d	0.00 - 617.98	0.00 - 529.23	0.00 - 537.81	0.00 – 437.00	0.00 - 350.26	0.00 - 345.03
Precipitation of Warmest Quarter (BIO18)	mm/d	0.24 – 1102.81	1.00 – 955.50	0.27 – 947.13	0.00 – 801.00	0.00 – 981.49	0.00 – 907.21
Precipitation of Coldest Quarter (BIO19)	mm/d	0.00 - 617.98	0.00 - 549.23	0.00 - 562.76	0.00 – 1504.00	0.00 - 1377.71	0.00 - 1301.49

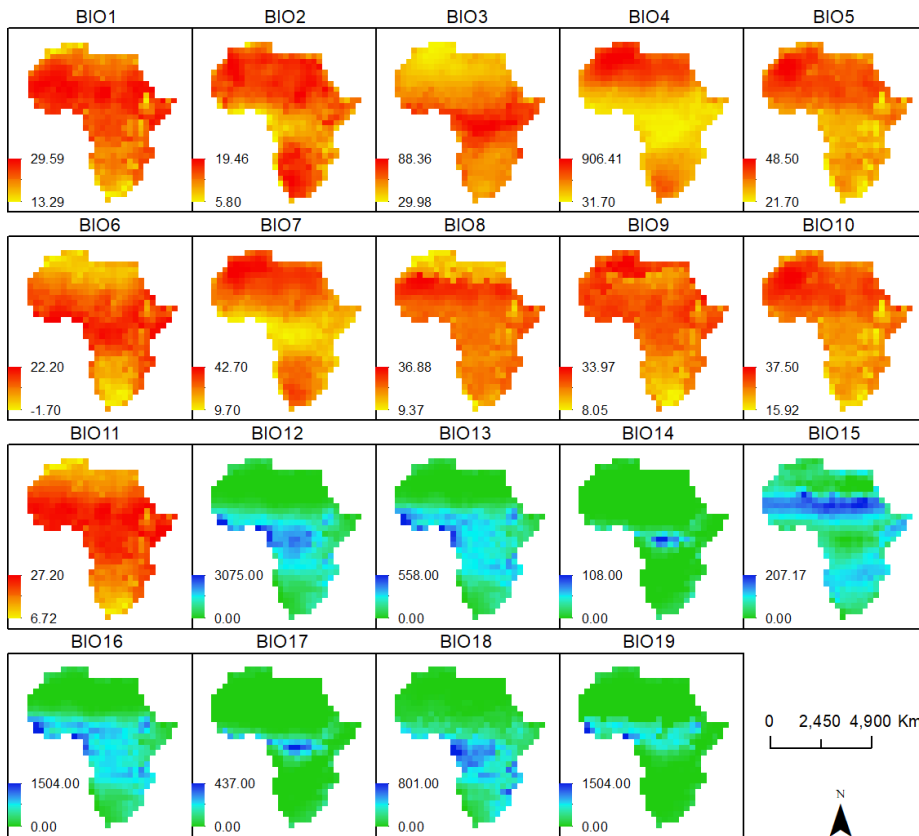


Figure 5 – Worldclim bioclimatic variables constructed for Current time period. See Tab. 1 for units and bioclimatic codes correspondence.

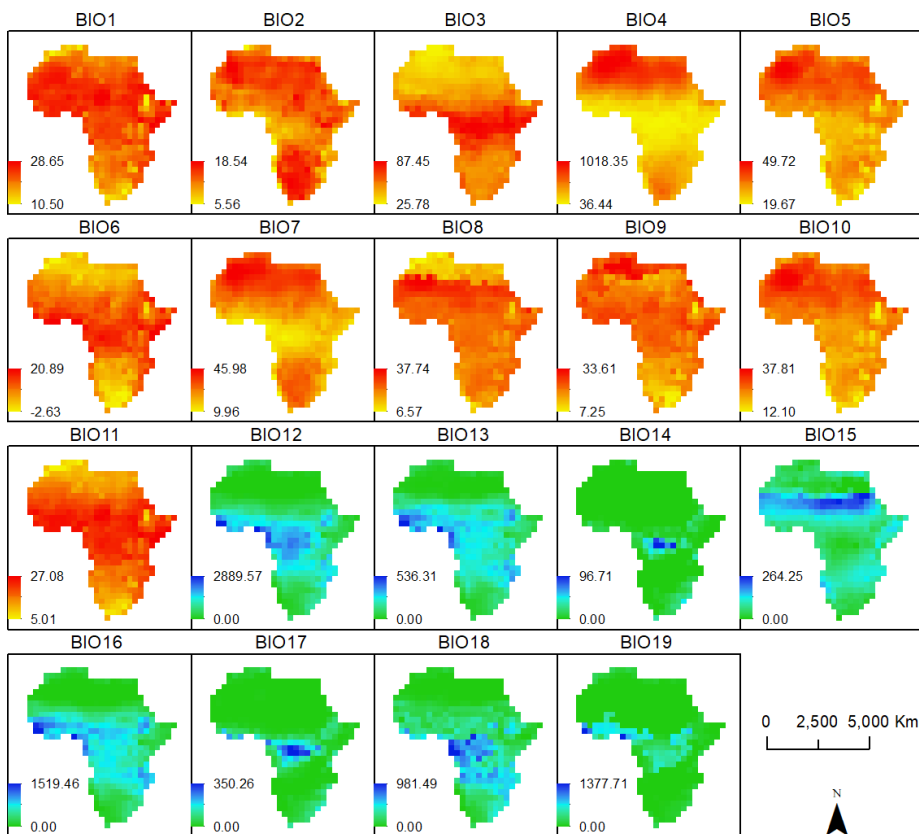


Figure 6 – Worldclim bioclimatic variables constructed for 6,000BP time period. See Tab. 1 for units and bioclimatic codes correspondence.

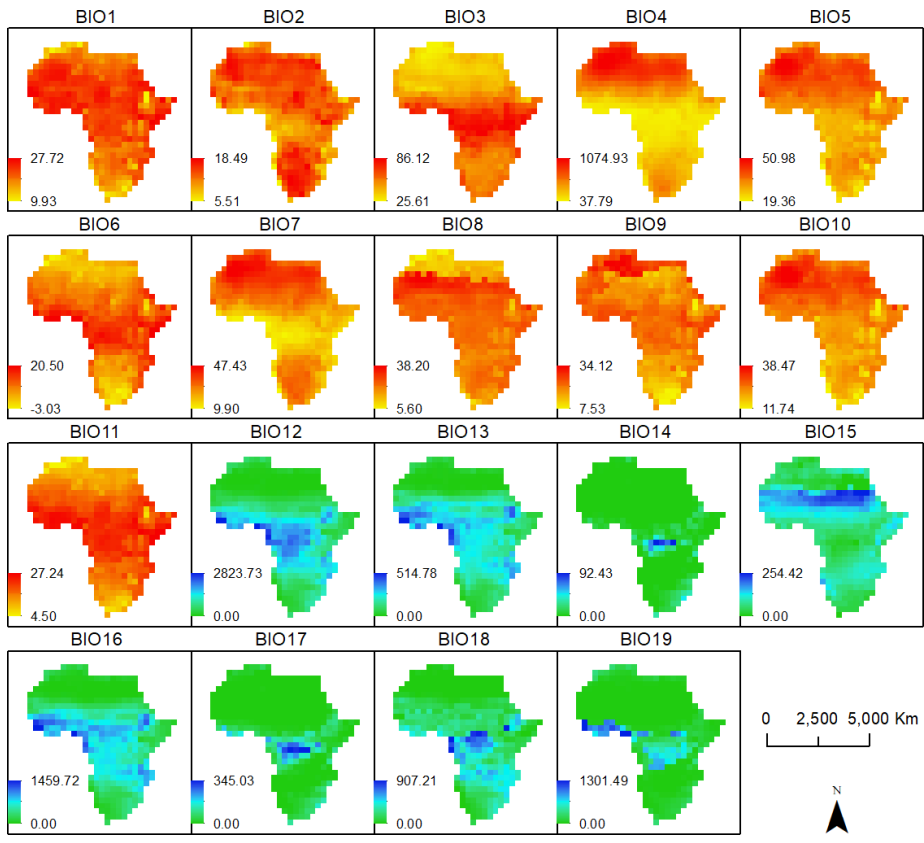


Figure 7 – Worldclim bioclimatic variables constructed for 7,900BP time period. See Tab. 1 for units and bioclimatic codes correspondence.

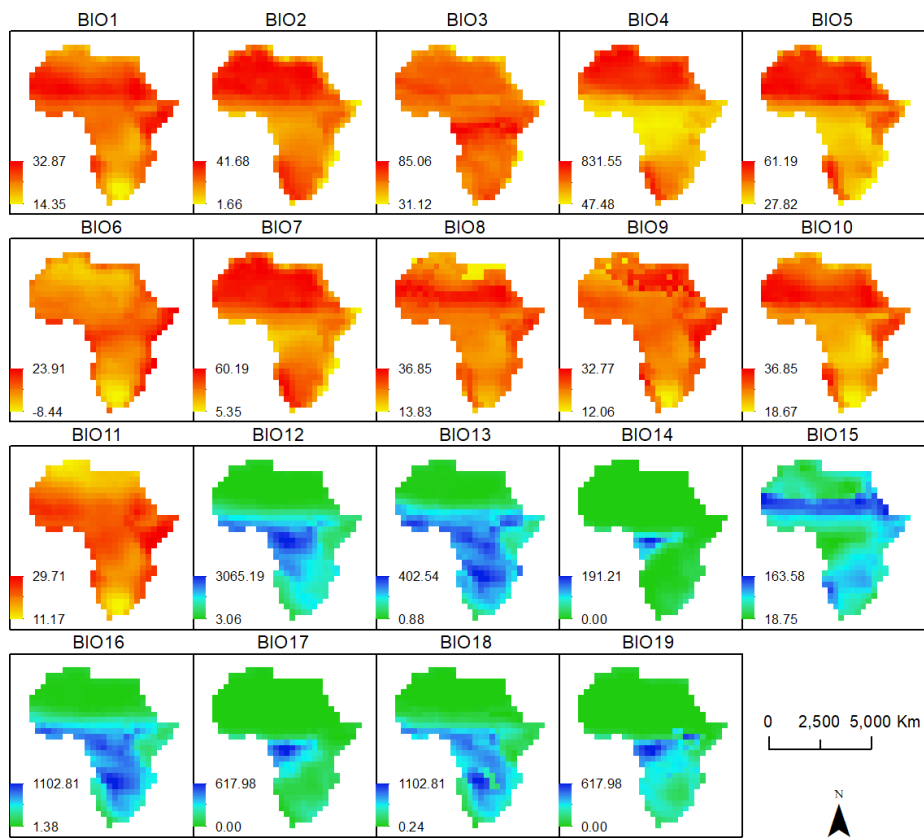


Figure 8 – Paleoview bioclimatic variables constructed for current time period. See Tab. 1 for units and bioclimatic codes correspondence.

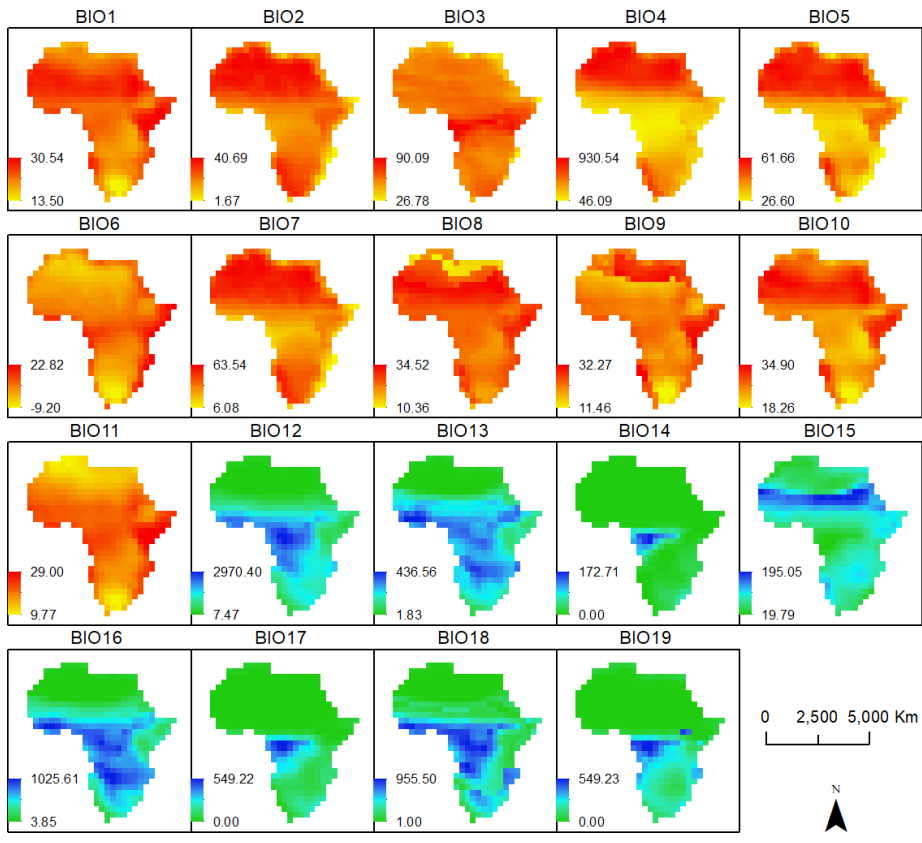


Figure 9 – Paleoview bioclimatic variables constructed for 6,000BP time period. See Tab. 1 for units and bioclimatic codes correspondence.

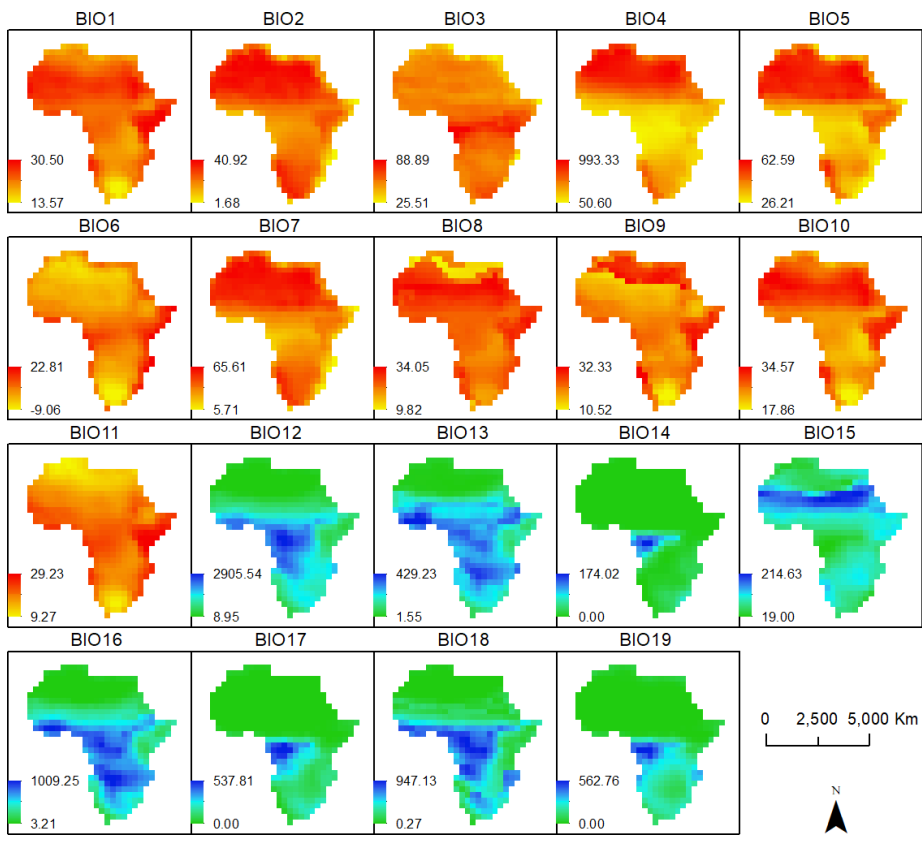


Figure 10 – Paleoview bioclimatic variables constructed for 7,900BP time period. See Tab. 1 for units and bioclimatic codes correspondence.

### 3.4.2 Land-cover variables

Land-cover maps for the African continent were derived for the three time periods (Current, 6,000BP and 7,900BP) from the Global Atlas of Paleovegetation (Fig. S3; Adams & Faure, 1990). The original maps were digitised and georeferenced, and a layer with a resolution of 0.3 degrees was created for each land-cover type. For each land-cover type, Euclidean distances to the pixels with a certain land-cover type were calculated. Finally, layers containing distances to each land-cover type were upscaled to a resolution of 2.5 degrees, using the nearest technique provided by the Resample tool (Table 2; Figs. 11 to 13). All procedures were done using ArcGis 10.5v software.

Table 2 – Euclidean distances to land-cover types in each time period. Units are in degrees.

Land-cover type (Code)	Current	6,000BP	7,900BP
Extreme Desert (LC1)	0.00 - 27.15	0.00 - 27.15	0.00 - 55.52
Grasslands (LC2)	0.00 - 21.03	0.00 - 21.03	0.00 - 18.83
Mediterranean Forest (LC3)	0.00 - 70.85	0.00 - 70.85	0.00 - 70.88
Mediterranean Scrub (LC4)	0.00 - 36.30	0.00 - 36.30	0.00 - 34.47
Montane Forest (LC5)	0.00 - 48.91	0.00 - 48.91	0.00 - 49.31
Savannah (LC6)	0.00 - 23.81	0.00 - 23.81	0.00 - 18.83
Scrub (LC7)	0.00 - 55.37	0.00 - 55.37	0.00 - 57.32
Semi-Desert (LC8)	0.00 - 20.65	0.00 - 20.65	0.00 - 23.92
Tropical Forest (LC9)	0.00 - 28.00	0.00 - 28.00	0.00 - 28.07
Recolonizing Forest Mosaic (LC10)	0.00 - 50.49	0.00 - 50.49	0.00 - 28.15
Woodland (LC11)	0.00 - 48.85	0.00 - 48.85	0.00 - 49.06

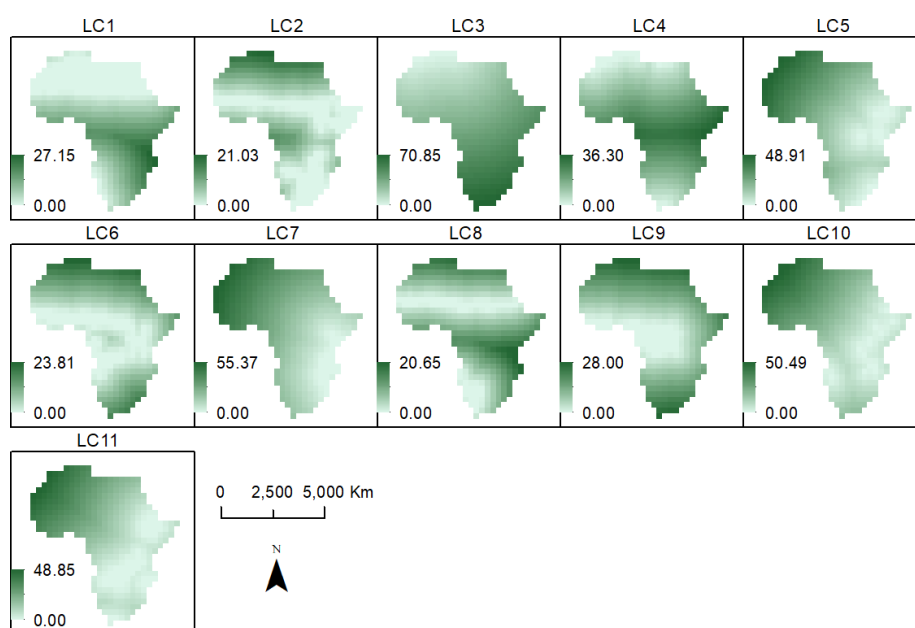


Figure 11 – Euclidean distances calculated for each biome for current time period. Distances are in degrees. See Tab. 2 for land-cover codes correspondence.

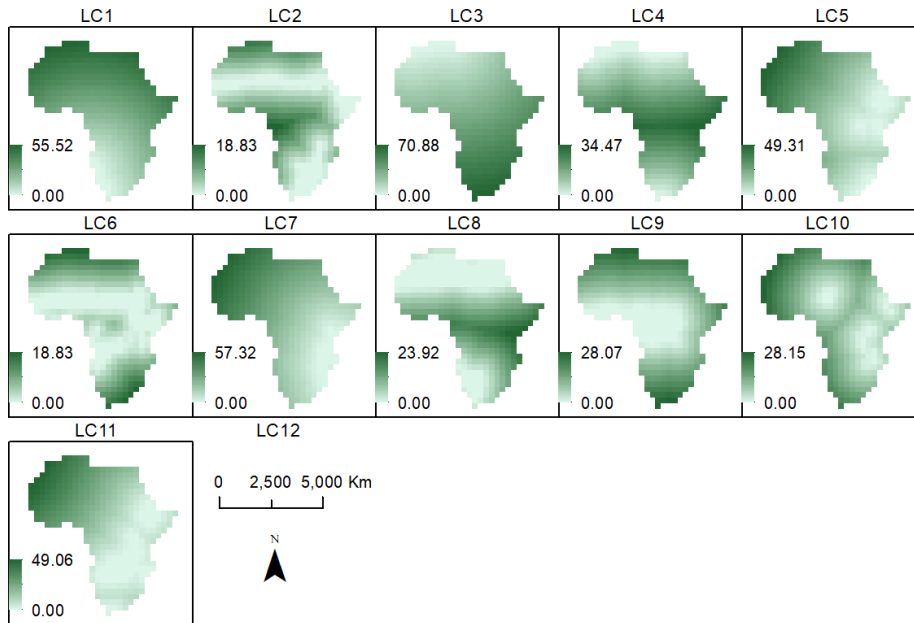


Figure 12 – Euclidean distances calculated for each biome for 6,000BP time period. Distances are in degrees. See Tab. 2 for land-cover codes correspondence

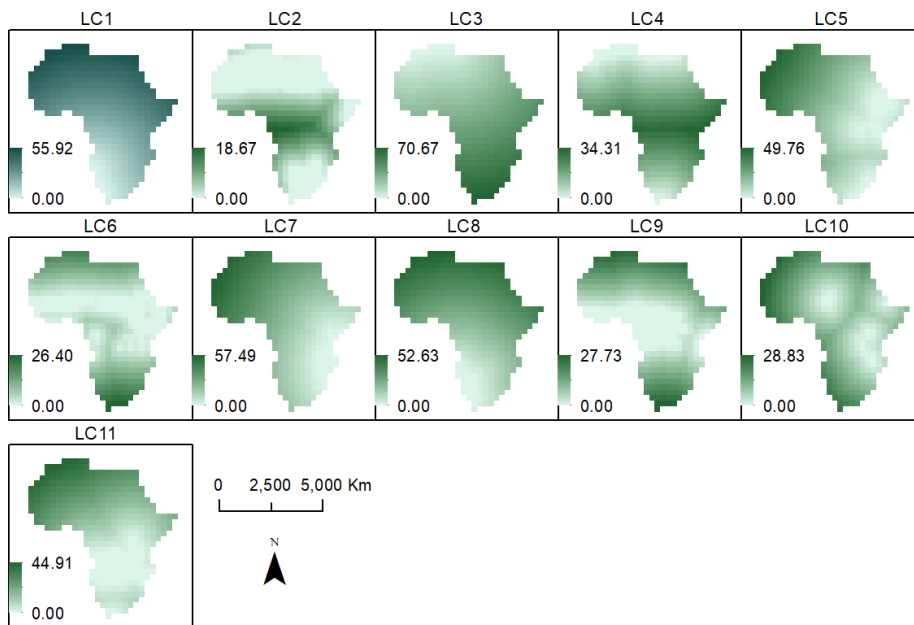


Figure 13 – Euclidean distances calculated for each biome for 7,900BP time period. Distances are in degrees. See Tab. 2 for land-cover codes correspondence

A MESS analysis of land-cover variables shows that they are generally very similar for regions outside the current Sahara-Sahel but not within the Sahara-Sahel (Fig. S4). Biomes in the Sahara-Sahel during early to mid-Holocene were extremely different from current ones especially for the Sahara. At 6,000BP the current Sahara Desert encompassed a semi-desert as well as grasslands and at 7,900BP the Sahara was occupied in its vast majority by grasslands and savannah biomes (Fig. S3).

### 3.5 Model building strategy

Model strategy was planned in order to access the influence of different data inputs on different modelling technique's ability to predict suitable areas for each species in each time period. To access the influence of absence data on the modelling methods, three datasets of pseudo-absences were created with ArcGIS 10.5 software (Table 3; Fig. S5): 1) Disk PA: randomly generated pseudo-absences outside buffers encompassing observations of each species in each time period. Buffer size was set according to home range size estimations, 300 km, for both species (Fennessy, 2009; Fennessy & Brown, 2010); 2) Random PA: randomly generated pseudo-absences in the full study area; and 3) Restricted PA: pseudo-absences created on the tropical forest regions only, since both species are absent from that biome (Roca *et al.*, 2001; O'Conner *et al.*, 2019). A polygon covering the African tropical forest region (Adams & Faure, 1990) was created and then pseudo-absences were selected from each pixel inside the polygon. The number of pseudo-absences in the Disk PA and Random PA datasets was identical to the number of observations in the training datasets, while pseudo-absences in the Ecologically Restricted PA dataset were retrieved from all possible pixels inside the polygon. Important to note while both the disk and random pseudo-absences generated were the same for both Holocene periods, the restricted dataset had to be generated differently for both Holocene periods since Tropical forest regions changed between 6,000BP and 7,900BP. These three pseudo-absence datasets were used on all different modelling procedures performed during this work.

Table 3 – Number of pseudo-absences generated for each method and for each period.

Species	Time period	Disk PA	Random PA	Restricted PA
Elephant	Historical	70	70	52
	Holocene	57	57	82 (6,000BP) and 92 (7,900BP)
Giraffe	Historical	67	67	50
	Holocene	57	57	82 (6,000BP) and 92 (7,900BP)

One modelling algorithm, Random Forests (Breiman, 2001), was used with different datasets and variables to predict suitable areas for both species in the three time periods (Fig. 14). All modelling procedures were done with BIOMOD2 (Thuiller *et al.*, 2009). In each algorithm, one model replicate was developed using all observations selected for model training. All observations were used for training in order to maximize the representation of the niches and because in this work it was not intended to account for the influence of occurrence records on the model. Models were generated using



bioclimatic (Worldclim, Paleoview) or land-cover variables (Landcover) alone, combinations of bioclimatic and land-cover variables (Worldclim and Landcover; Paleoview and Landcover), and all possible combinations of pseudo-absence data (Disk, Random, Restricted). This procedure resulted in 30 different combinations of different data input, for each model algorithm and for each species constructed for three different time periods and then projected to specific temporal scenarios (Fig. 14): i) Models constructed for Current conditions were projected to 6,000BP and 7,900BP; ii) Models constructed for 6,000BP and for 7,900BP were projected to Current conditions; Current replicates were projected to the Holocene time periods (6,000BP and 7,900BP) and vice versa, to check the transferability of species observational datasets.

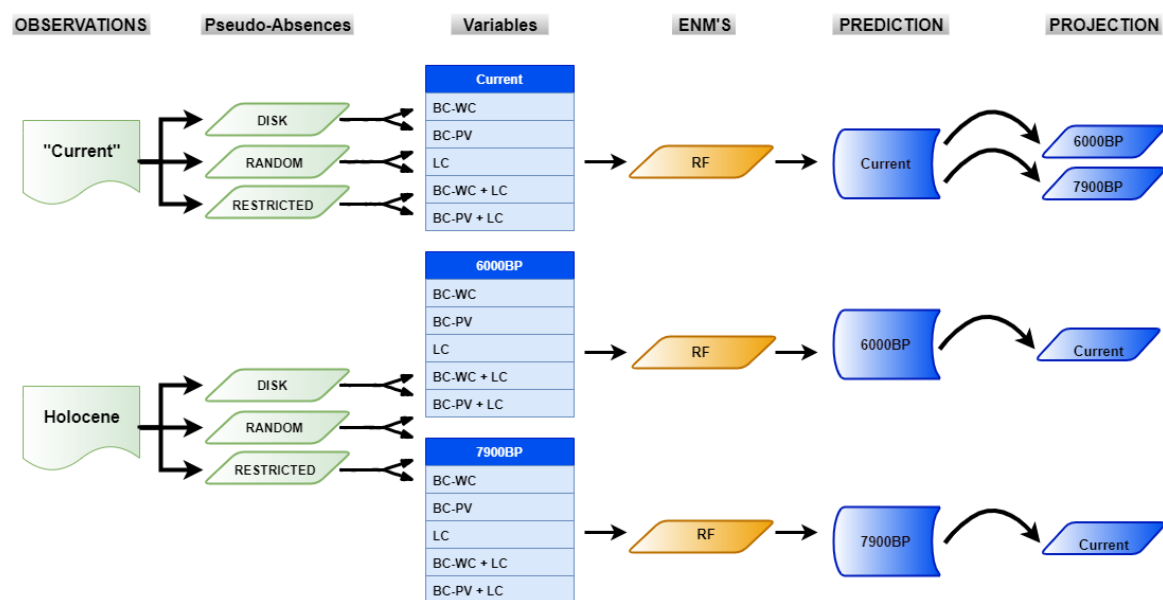


Figure 14 – Modelling strategy diagram. BC-WC: Bioclimatic variables from Worldclim dataset; BC-PV: Bioclimatic variables from Paleoview dataset; LC: Land-cover variables; RF: Random Forest;

### 3.6 Evaluation and comparison of models

The Receiver Operating Characteristic (ROC) was calculated to assess each model discriminatory power constructed in this work. ROC curves are constructed by using all possible thresholds to classify the scores into confusion matrices, obtaining sensitivity and specificity for each matrix, and then plotting sensitivity against the corresponding proportion of false positives. From the ROC, the Area Under Curve (AUC) was calculated. The True Skill Statistic was not calculated since models were calibrated with no test data, and thus values for specificity and sensitivity cannot be obtained. By using all possible thresholds, the need for a selection of a single threshold (which is often arbitrary) is

avoided, and allows evaluation of the trade-off between sensitivity and specificity (Pearce & Ferrier, 2000).

Model Projections were compared with the distributions of the “Current” and Holocene datasets. The observations and pseudo-absence datasets were intersected with the model projection and the respective probabilities of occurrence extracted. Models with projections that better matched the distributions obtained previously were considered as the best ones.

## 4. Results

### 4.1 Distribution of study species

From the Holocene time period, a total of 292 rock-art depictions for the elephant and 273 for the giraffe were collected and mapped (Tab. 4; Fig. 15). Most of these observations were concentrated in the current Sahara desert, with scattered observations recorded in Eastern and Southern Africa. In comparison with the distribution at the Holocene, the “Current” distribution was similar in the Eastern and Southern Africa, but in North Africa both distributions were restricted to Sahel regions (almost absent from the Sahara Desert and absent from the Mediterranean Basin).

Table 4 – Number of observations collected for each species and time period.

Species	Time period	Full set	Model training
Elephant	“Current”	753	70
	Holocene	292	57
Giraffe	“Current”	303	67
	Holocene	273	57

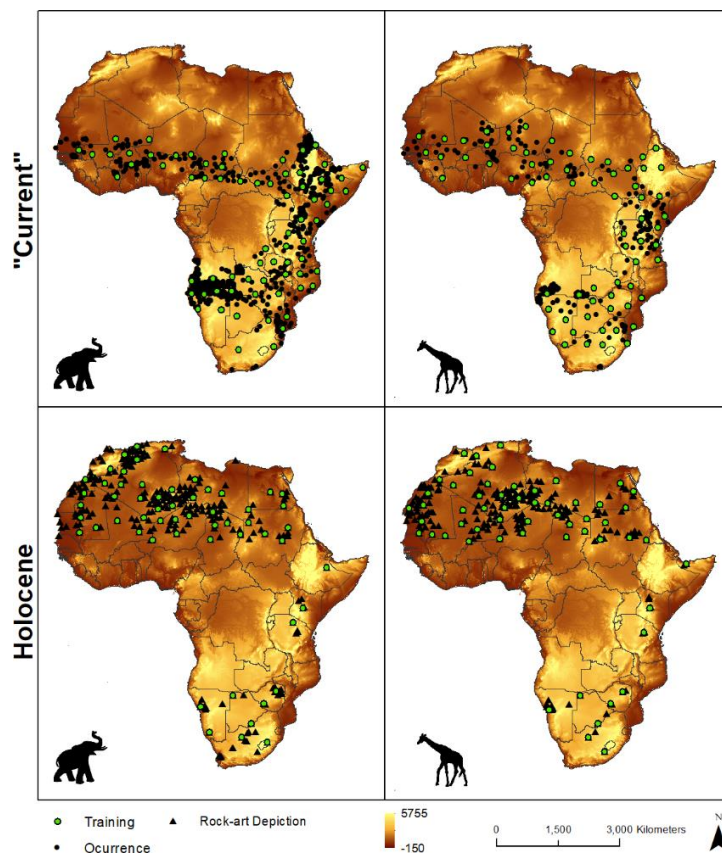


Figure 15 – Distribution of “Current” (top) and Holocene (bottom) datasets obtained for the Elephant and Giraffe, distinguishing observations used for model training from the full set of available observations (Table 4). Gradient represents altitude.

## 4.2 Model projections

Models predictions for current and Holocene time periods scored high AUC values (between 0.9 and 1.0; Tab. S3, S4). Overall, for both study species, the spatial predictions of models calibrated in current conditions and projected for 6,000BP (Fig. 16) and of models calibrated in 6,000BP and projected to current conditions (Fig. 17) were consistent between the different datasets of environmental variables and pseudo-absences. In contrast, predictions of models calibrated in current conditions and projected for 7,900BP (Fig. 18) and of models calibrated for 7,900BP and projected to current conditions (Fig. 19) displayed more variability between the different datasets. The distribution of the percentage of observations (Figs. 20 and 21) and of pseudo-absences (Figs. 22 and 23) in each probability of occurrence class according to model predictions tended to follow the same pattern of spatial predictions: more homogeneous across datasets of environmental variables and of pseudo-absences in models calibrated in current conditions and projected for 6,000BP and vice-versa in comparison to models calibrated in current conditions and projected for 7,900BP and vice-versa.

In both study species, models calibrated with restricted pseudo-absences performed significantly better than the other two pseudo-absence datasets for all model projections performed in this work, and in general the disk pseudo-absences also performed better than the random pseudo-absences. Overall, the projections of models calibrated with disk and random pseudo-absences considerably underperformed in comparison to models based in restricted pseudo-absences. The areas projected with high occurrence probability for both study species based in models calibrated with restricted pseudo-absences were significantly larger across Africa in comparison to projections based in disk or random pseudo-absences, and they were also more consistent between the different environmental variables used to calibrate models in comparison to the other two pseudo-absence datasets. Overall, there were no major differences between the projections from models calibrated with the different sources of environmental variables (Worldclim and Paleoview) in both study species, but minor differences were noticed when projecting models calibrated for the Holocene time periods to current conditions.

In general, when using the restricted pseudo-absences dataset, the models calibrated with land-cover only variables performed similarly to models constructed with climate-only variables when projected to 6,000BP or when calibrated for 6,000BP and projected to current conditions. However, differences were found in models projected to 7,900BP or when calibrated for 7,900BP and projected to current conditions. With the disk and random pseudo-absences datasets, models calibrated with land-cover only variables consistently performed better than models based in climate-only variables.

In general, the best performing projections for both study species and for the three time periods were the projections resulting from models calibrated with land-cover together with climatic variables, which yield consistent predictions and the highest probabilities of occurrence for the occurrences datasets and the lowest probabilities for the pseudo-absences datasets.

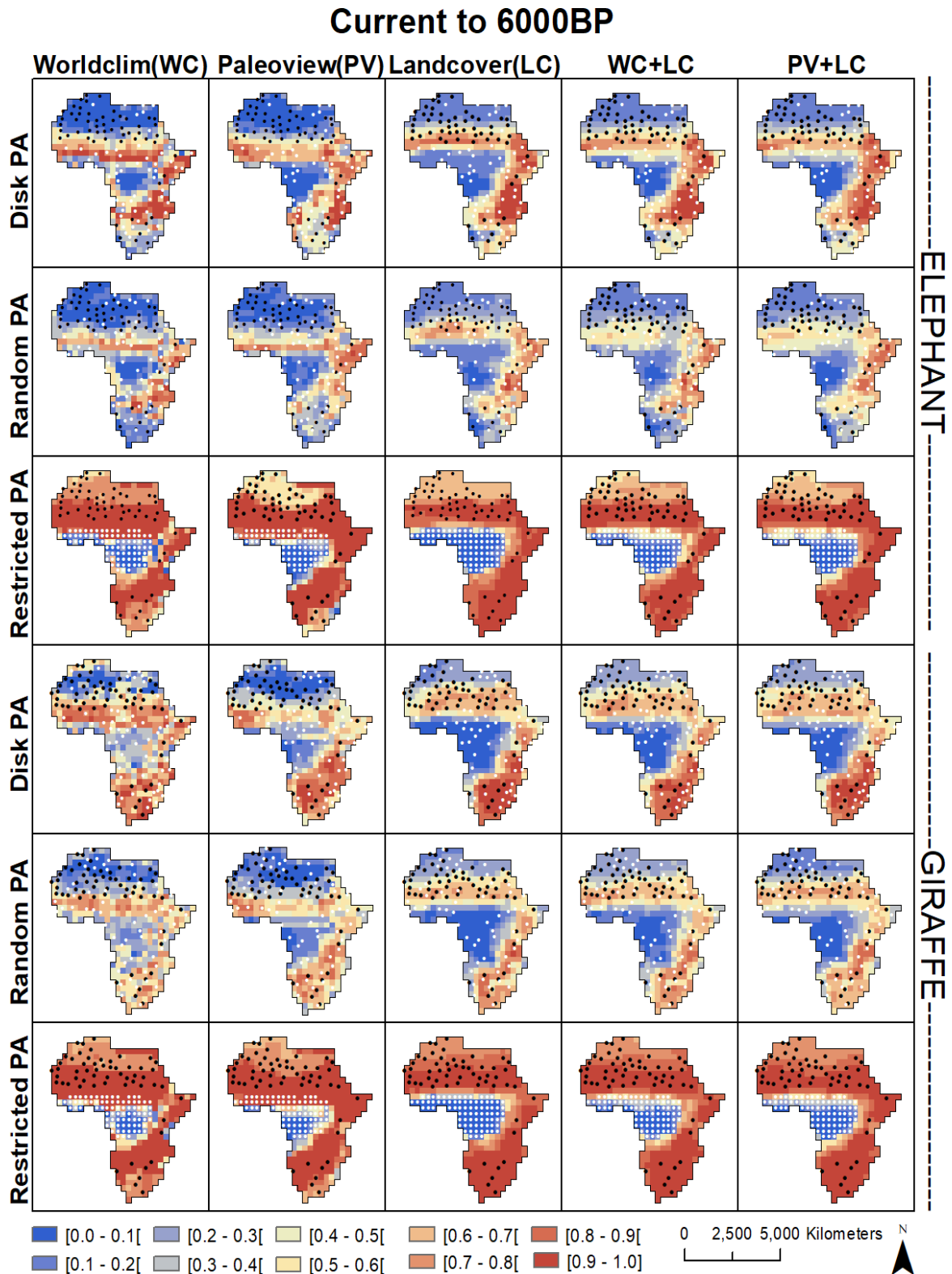


Figure 16 – Models calibrated in current time period and projected to 6,000BP time-period. Black points correspond to occurrences collected for the Holocene inferred from rock-art depictions and white points correspond to the different pseudo-absences generated for the projection time period. Results according to distinct datasets of environmental variables (Worldclim-only, Paleoview-only, Land-cover-only, Worldclim and Land-cover, and Paleoview and Land-cover) and pseudo-absences (Disk, Random and Restricted).

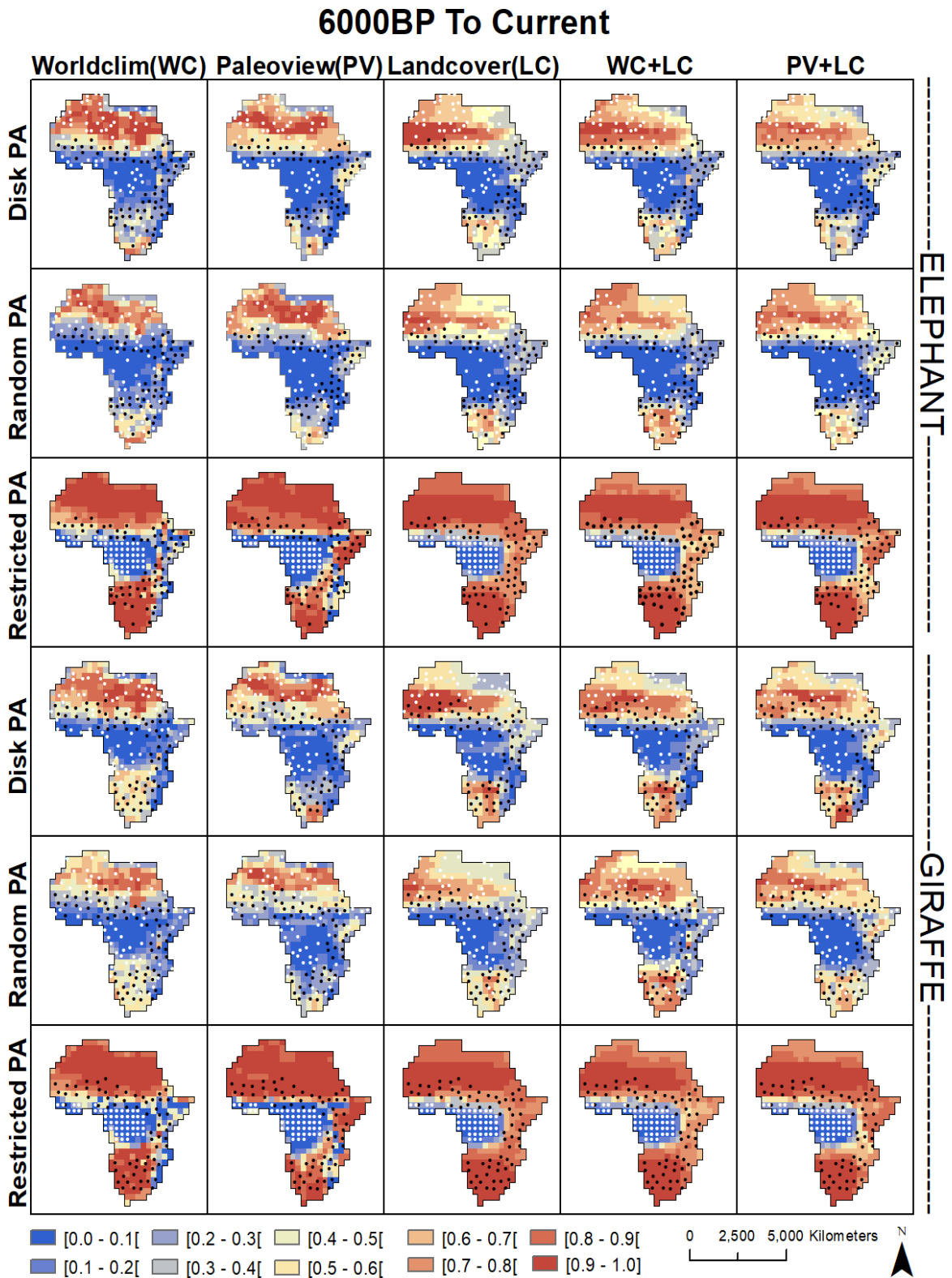


Figure 17 – Models calibrated in 6,000BP time period and projected to current time-period. Black points correspond to occurrences collected for the “current” dataset and white points correspond to the different pseudo-absences created for the projection time period. Results according to distinct datasets of environmental variables (Worldclim-only, Paleoview-only, Land-cover-only, Worldclim and Land-cover, and Paleoview and Land-cover) and pseudo-absences (Disk, Random and Restricted).

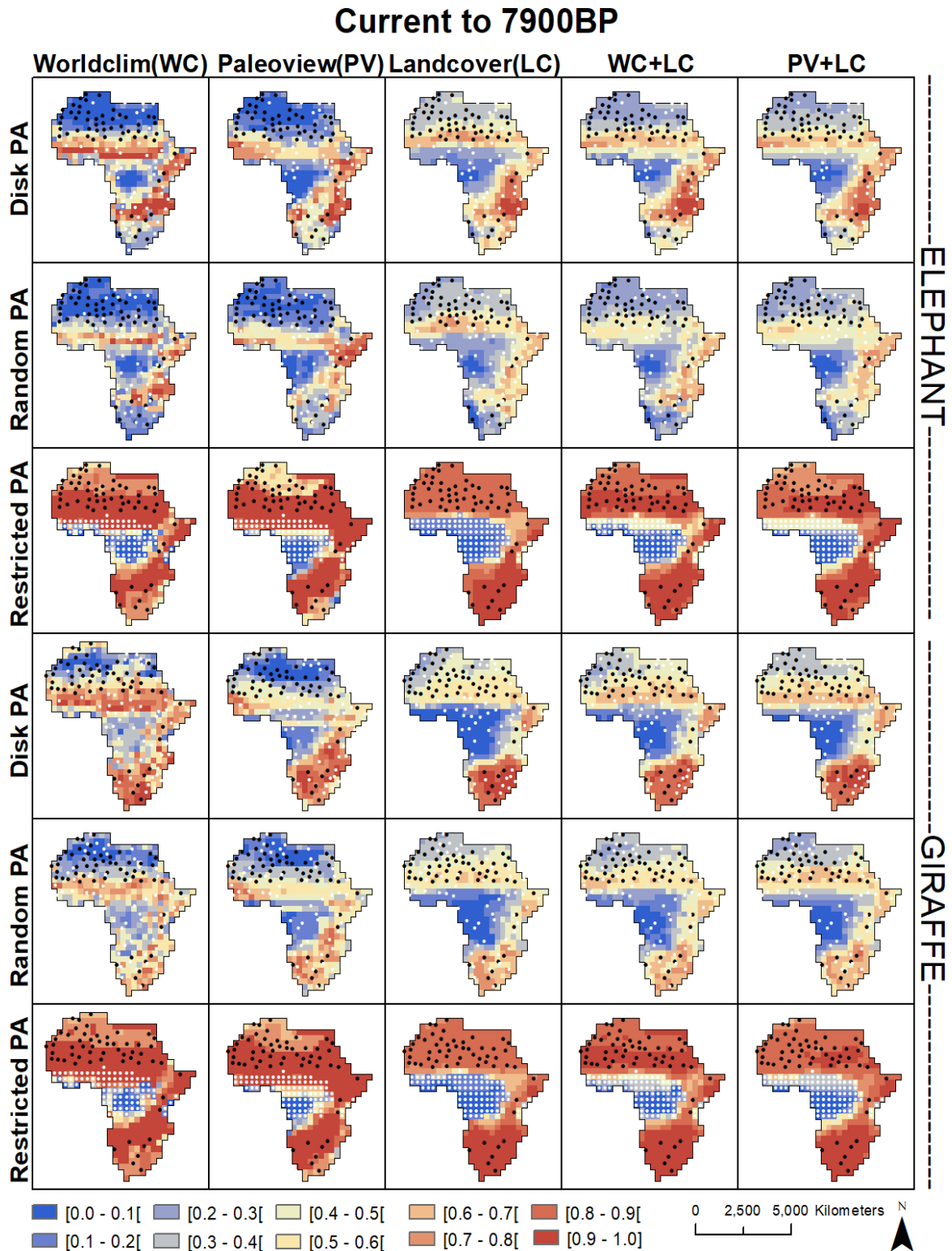


Figure 18 – Models calibrated in current time period and projected to 7,900BP time-period. Black points correspond to occurrences collected for the Holocene inferred from rock-art depictions collected and white points correspond to the different pseudo-absences generated for the projection time period. Results according to distinct datasets of environmental variables (Worldclim-only, Paleoview-only, Land-cover-only, Worldclim and Land-cover, and Paleoview and Land-cover) and pseudo-absences (Disk, Random and Restricted).



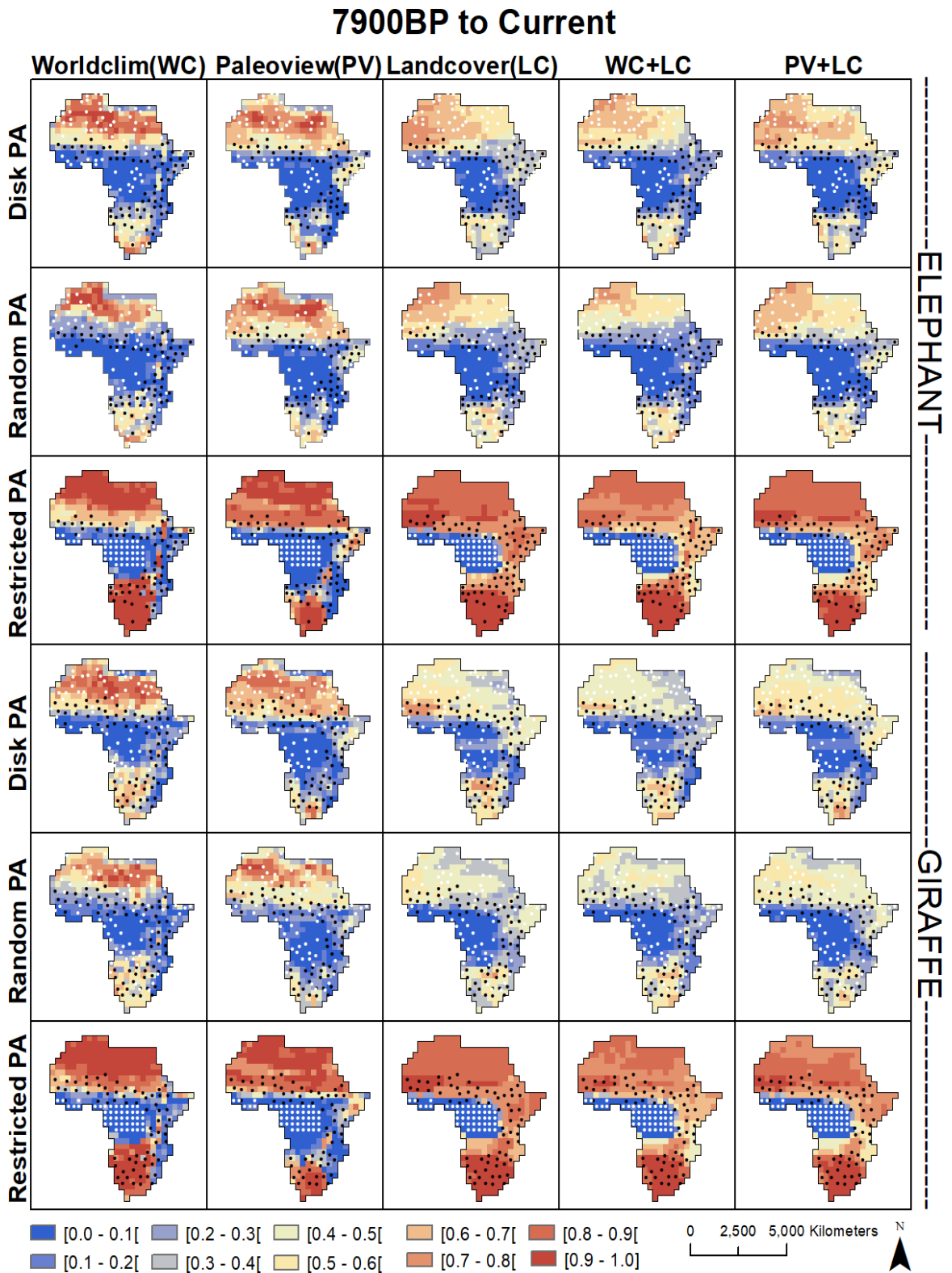


Figure 19 – Models calibrated in 7,900BP time period and projected to current time-period. Black points correspond to occurrences collected for the “current” dataset and white points correspond to the different pseudo-absences created for the projection time period. Results according to distinct datasets of environmental variables (Worldclim-only, Paleoview-only, Land-cover-only, Worldclim and Land-cover, and Paleoview and Land-cover) and pseudo-absences (Disk, Random and Restricted).

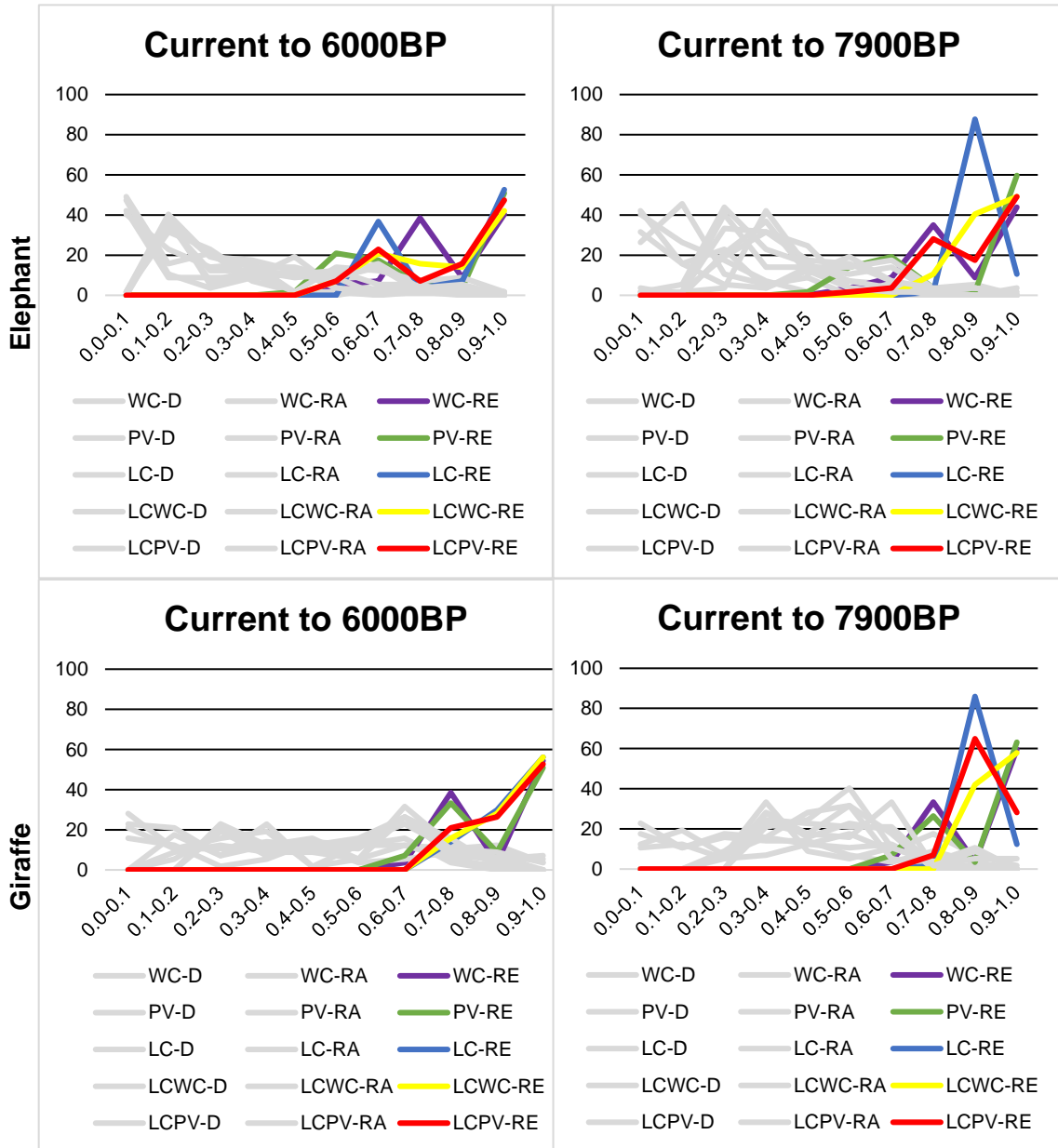


Figure 20 – Percentage of observations of each study species in the Holocene (y axis) in each probability of occurrence class (x axis) according to models calibrated in current conditions and projected to both Holocene time periods according to distinct datasets of environmental variables (WC: Worldclim-only; PV: Paleoview-only; LC: Land-cover-only; LCWC: Worldclim and Land-cover; LCPV: Paleoview and Land-cover) and pseudo-absences (D: Disk; RA: Random; RE: Restricted). Results are individualised only for the cases where the profile is dissimilar to the main profile pattern.

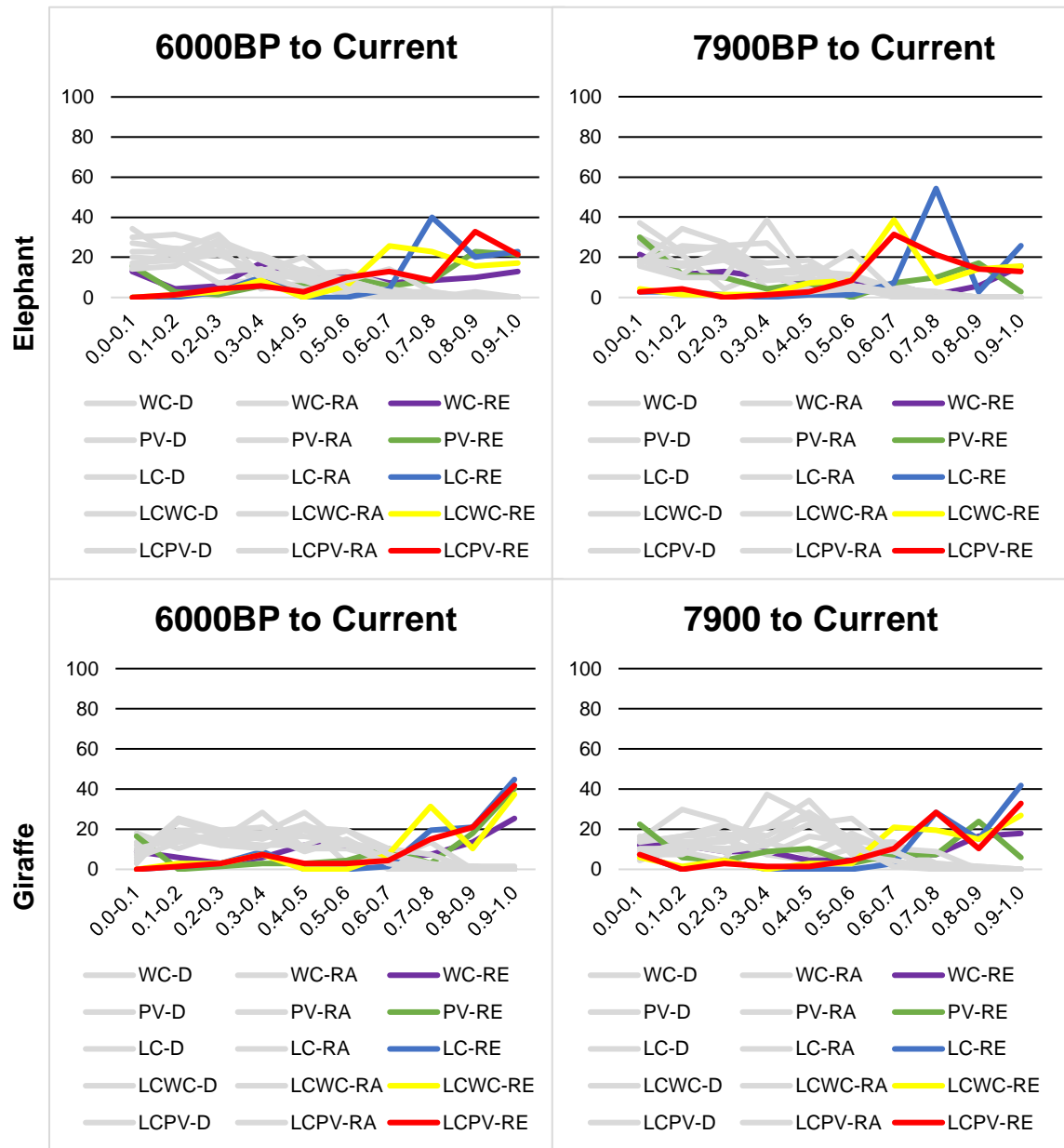


Figure 21 – Percentage of observations of each study species in the “Current” dataset (y axis) in each probability of occurrence class (x axis) according to models calibrated in both Holocene time periods and projected to current conditions according to distinct datasets of environmental variables (WC: Worldclim-only; PV: Paleoview-only; LC: Land-cover-only; LCWC: Worldclim and Land-cover; LCPV: Paleoview and Land-cover) and pseudo-absences (D: Disk; RA: Random; RE: Restricted). Results are individualised only for the cases where the profile is dissimilar to the main profile pattern.

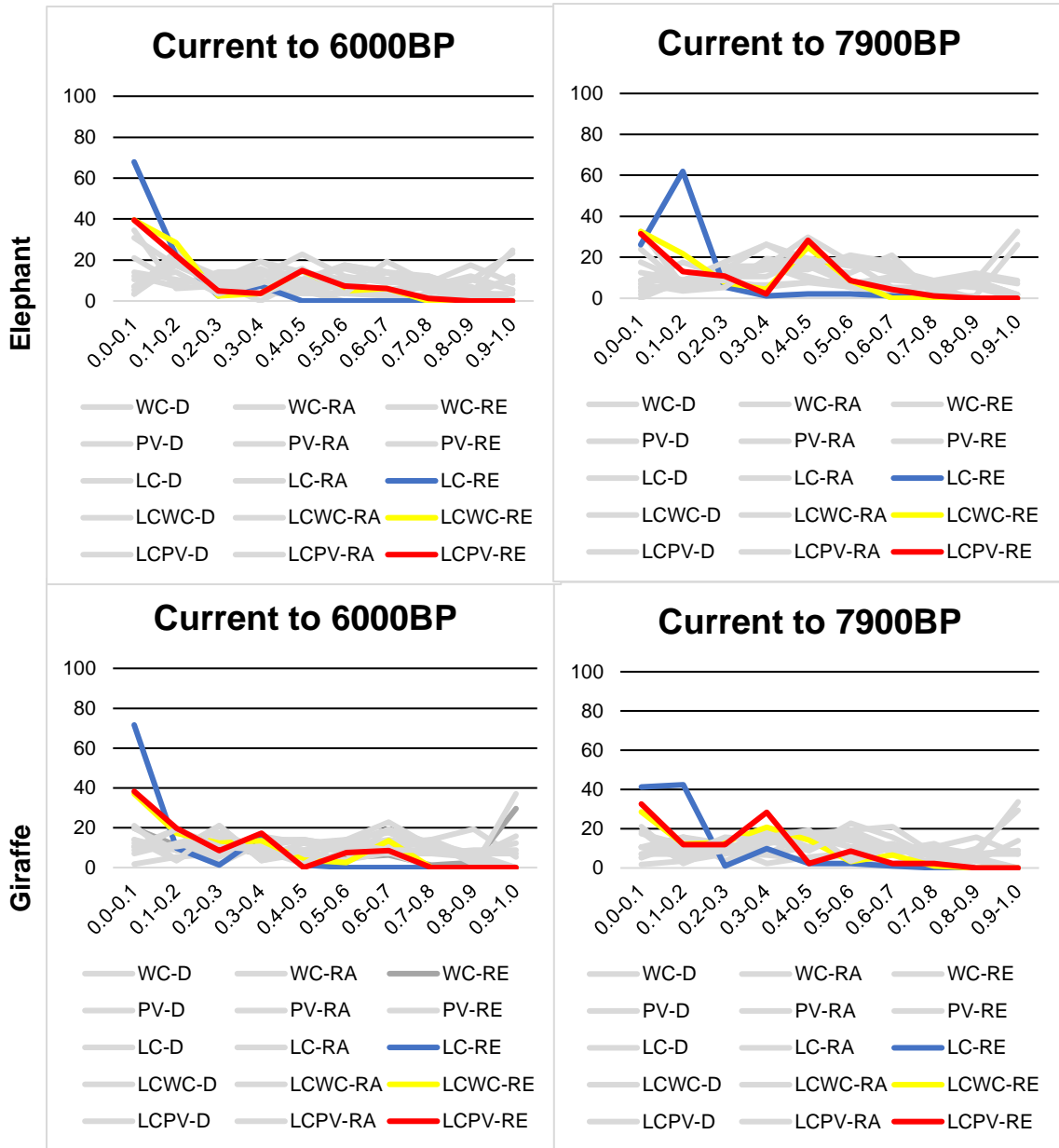


Figure 22 – Percentage of pseudo-absences generated for each study species at the Holocene (y axis) in each probability of occurrence class (x axis) according to models calibrated in current conditions and projected to both Holocene time periods according to distinct datasets of environmental variables (WC: Worldclim-only; PV: Paleoview-only; LC: Land-cover-only; LCWC: Worldclim and Land-cover; LCPV: Paleoview and Land-cover) and pseudo-absences (D: Disk; RA: Random; RE: Restricted). Results are individualised only for the cases where the profile is dissimilar to the main profile pattern.

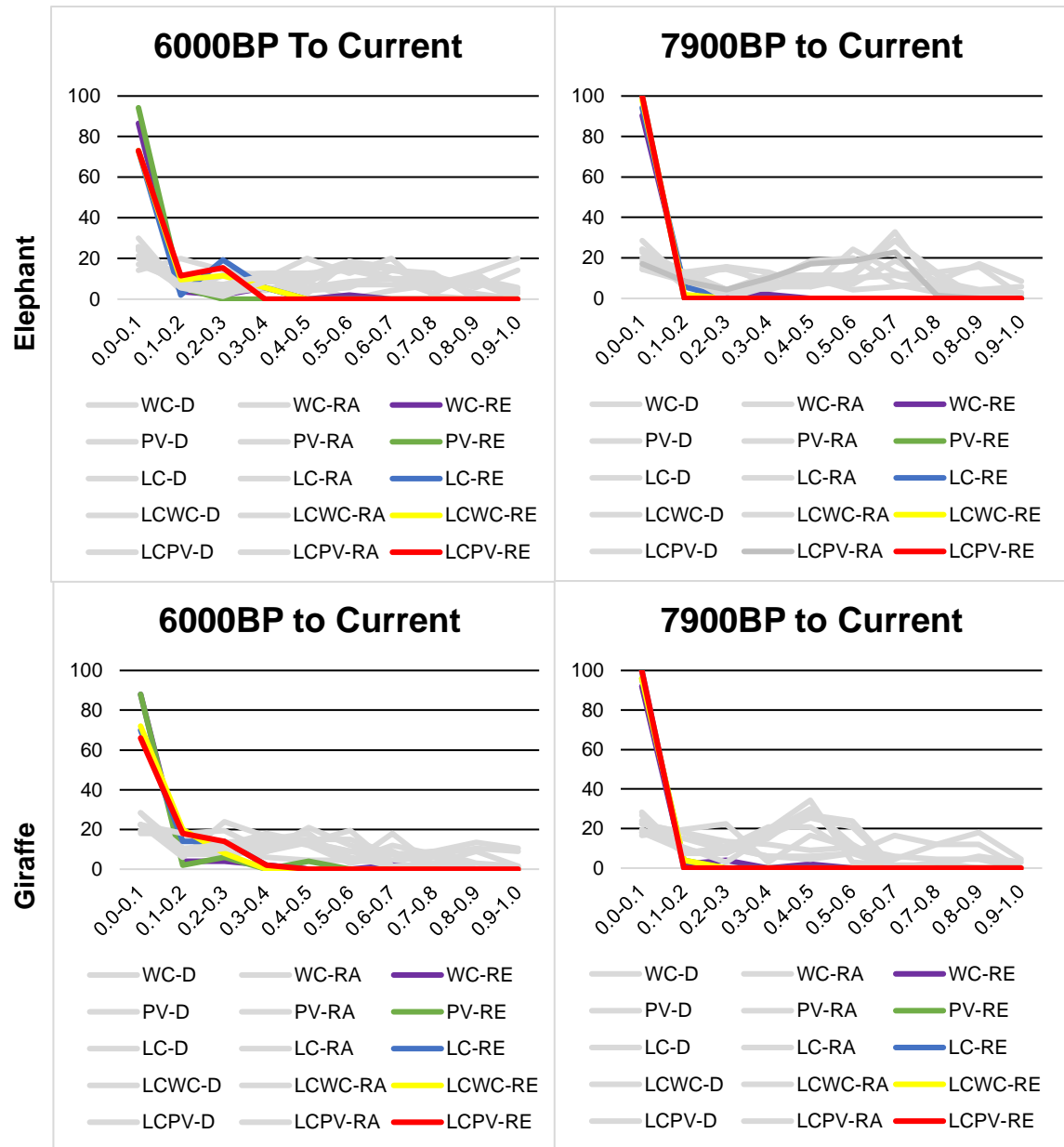


Figure 23 – Percentage of pseudo-absences generated for each study species in current conditions (y axis) in each probability of occurrence class (x axis) according to models calibrated in both Holocene time periods and projected to current conditions according to distinct datasets of environmental variables (WC: Worldclim-only; PV: Paleoview-only; LC: Land-cover-only; LCWC: Worldclim and Land-cover; LCPV: Paleoview and Land-cover) and pseudo-absences (D: Disk; RA: Random; RE: Restricted). Results are individualised only for the cases where the profile is dissimilar to the main profile pattern.

## 5. Discussion

### 5.1 Distribution of elephant and giraffe at the Holocene

Overall, the occurrence data collected in this work from rock-art depictions resulted in a reasonable estimation of the distribution areas and of the environmental conditions associated with the ecological niches of the two studied species. Moreover, the final distributions obtained made sense when considering the drastic differences in both climate and land-cover between current and Holocene time periods: rock-art depictions were collected from localities where the surrounding biomes during the Holocene matched the biomes where both species today occur. Moreover, most of the rock-art depictions were collected in the Sahara, a region where both species were known to occur during the Holocene (Drake *et al.*, 2011).

Clusters of rock-art depictions were stacked on mountain and desert regions. Such clustering is most likely due to the availability of rock surfaces to paint and/or engrave on. Complementarily, mountain areas display higher water availability (in comparison to lowland areas), since river sources are found in mountain tops, which would allow resources for both human, elephant and giraffe populations. No rock-art records were retrieved from regions with dense vegetation cover (i.e. tropical forests). The apparent absence of rock-art from lowland forest areas is probably linked to the difficulty in locating rock-art between dense vegetation and/or to the fewer rock availability of tropical forests in comparison to the open and bare Sahara Desert,

The representation of key morphological aspects (e.g. fur patterns) and behavioural traits (e.g. herds, elephants in musth) in rock-art provided cheap and relatively easy to collect information to identify the presence of elephants and giraffes in past environments. In fact, the present work shows the untapped potential that rock-art depictions can bring into ecological modelling. To date, few studies have tried to identify animal species from rock-art to make inferences on their past distributions. For instance, Guagnin *et al.* (2018), similarly to this work, used rock-art depictions to successfully map the palaeodistribution of Kudu at local scale, and Guagnin *et al.* (2016) used rock-art depictions as a proxy for Holocene environmental change. Still, the potential of rock-art as source of distribution data is mostly limited to the species chosen to be depicted, generally large-sized species and/or species displaying very distinguishable morphological traits. In fact, identifying certain animals can be a challenge due to lack of key characteristics that define them in the depictions. This aspect remains the major drawback when aiming to use rock-art to reconstruct paleodistributions.

## 5.2 Effects of ecogeographical variables in model transferability

The spatial predictions of models based in land-cover or climatic variables were very similar. Such similarity was most likely since, ultimately, climatic processes affect the type of land-cover. However, models suggested that land-cover variables were more representative of the changes occurred between time periods, whereas climatic differences between time periods were not so significant. For instance, in the Sahara Desert, the climatic variables do not mimic the extreme differences occurred in land-cover: while climate variation from the current to the Holocene was overall similar across the region (Figs. 8 to 13), the land-cover was extremely different between time periods (Figs. 14 to 16). Such differences are probably related to the fact that slight increases in precipitation can lead to drastic changes in vegetation cover, and consequently extreme sandy deserts can be replaced with grasslands in relatively short time periods (Navone *et al.*, 2006).

The best performing projections were obtained from models constructed with both climatic and land-cover variables. Previous studies testing whether the integration of land-cover data improves the performance of pure bioclimatic models (using GAM algorithm) for bird species on multiple spatial scales reported that land-cover had a notable contribution to the accuracy of bioclimatic models at finer scales, ranging from 10km to 20km (Luoto *et al.*, 2007). Another study showed that incorporating land-cover data in Artificial Neural Networks models for *Rhynchospora alba* and *Erica tetralix* significantly improved purely climate-driven predictions (Pearson *et al.*, 2004). A study assessing the influence of land-cover and climate on mammal distributions across Europe suggested that land-cover variables, even if not directly related or weakly associated with distributions, are interesting to describe mammal distributions despite not improving the predictive performance when added to bioclimatic models (Thuiller *et al.*, 2004). A possible explanation for the weak contribution of land-cover in the study of (Thuiller *et al.*, 2004) is that land-cover variables used to calibrate the models were very similar to the ones of the projecting period. In the current study, there were extreme differences in land-cover between projection periods, which may explain why models built with bioclimatic and land-cover performed better in comparison to models built with bioclimatic-only or land-cover-only variables.

Overall, models based in Worldclim or Paleoview bioclimatic variables performed very similarly. Very few studies have addressed the effects of different sources of bioclimatic variables in model's performance, which hampers the comparison of the results obtained to the current literature. Morales-Barbero & Vega-Alvaréz (2019) assessed model performance according different sources of bioclimatic layers (Worldclim; CHELSEA;

MERRAclim) and found significant differences between data sources, for instance in probability of occurrence according to environmental variation. In the current study, the lack of difference in model performance according to the two bioclimatic datasets may be related with the larger pixel size used to build models (2.5 degrees). Spatial resolution is known to impact model performance (Thuiller *et al.*, 2003; Seo *et al.*, 2009) and probably, differences between models calibrated with the two data sources would have been detected if finer spatial resolutions were used. Most likely, the large pixel size here used contributed to the dilution of environmental differences between the distinct data sources.

### 5.3 Methods to select pseudo-absences for model transfer

There were major differences in model performance according to the method of selection of pseudo-absences, with models calibrated with environmentally restricted pseudo-absences performing overall better. The methods to generate pseudo-absences in model calibration are known to affect model performance. For instance, in GAMs based in virtual data, model accuracy (based in AUC estimates) increases with larger numbers of pseudo-absences and a ratio of 100:1 of pseudo-absences to presences should be used when presence data availability is limited (Lobo *et al.*, 2010). In MaxEnt models built for 226 species (with number of presences ranging from 2 to around 5,800) for six regions of the world, model accuracy (AUC-based) increased substantially with increasing number of pseudo-absences used (Phillips *et al.*, 2008). Based in different model algorithms (GLM; GAM; MARS; MDA; CTA; BRT and RF) built with virtual data, it was shown that when limited numbers of observations are available (between 30 to 100), model performance increases if pseudo-absences are randomly generated at least two degrees away from any presence point (Barbet-Massin *et al.*, 2012). This method to generate pseudo-absences is similar, to some extent, to the disk pseudo-absence dataset used in the current study (where the threshold distance represented the species home range size). However, in the current study, models calibrated with disk pseudo-absences underperformed and using environmentally restricted pseudo-absences yielded models with larger performance. A possible explanation for this result is that the “Current” observation dataset used in the current study does not represent the full ecological range of conditions where both studies species may occur, due to an artificial reduction of species distributions caused by extensive poaching. Even though the “Current” distribution here used, tried to account for this factor, by incorporating historical (early 20<sup>th</sup> century) and current observations (Martínez-Freiría *et al.*, 2016), there are large areas where both species have likely been extirpated and for which no distribution data are available. For example, although no distribution data are available for elephants in the Mediterranean



region and Sahara desert, both historically or currently, the species was known to be present in these areas in the Roman Period, at about 2,000 yr BP (Pliny the Elder, 23–79 CE; Shaw, 1976). Despite the extreme aridity of the Sahara, if poaching had been absent, most likely elephants would still roam in the region. This is further supported by elephants still residing in the Namib desert, which hosts similar arid conditions to the current Sahara. Both studied species are relatively generalists, occurring in a wide range of environmental conditions (Roever *et al.*, 2012; Muller *et al.*, 2018), being absent from tropical forests mainly due to their large body sizes, resulting in difficulties to move in between very dense and tight vegetation. As such, if the “current” dataset included the complete distribution of the studied species, then the performance of models calibrated with disk pseudo-absences would most likely be very similar to the one of models using restricted pseudo-absences.

Even though the current study did not address the effects of different numbers of pseudo-absences, it is known that they affect model performance. For instance, a study based in virtual species and using ten modelling techniques to build both single and ensemble models with varying levels of prevalence and disparate numbers of training presences, showed relationships between model accuracy and the number of presences and of pseudo-absences or background sites (Liu *et al.*, 2019). In fact, it has been suggested that using identical numbers of pseudo-absences and presences to calibrate random forest models yields the best model performances (Barbet-Massin *et al.*, 2012; Liu *et al.*, 2019). In the current study, although the restricted pseudo-absences dataset (N=from 52 to 92) was not exactly identical to the number of observations (N=from 57 to 70) used to calibrate models, still models calibrated with restricted pseudo-absences outperformed models calibrated with disk or random datasets, which used exactly identical numbers of pseudo-absences and observations.

There were minor differences in model performance when calibrated with disk and random pseudo-absences, most likely related with the large pixel size used. The large pixel size was a consequence of two factors: Paleoview data are available at 2.5 degrees spatial resolution (Fordham *et al.*, 2017) and the home-range size of studied species is above 250 km (Fennessy, 2009; Fennessy & Brown, 2010; Wall *et al.*, 2013). Consequently, the number of pixels available in the study area to select random or disk pseudo-absences was limited, which resulted in spatially similar datasets of pseudo-absences (Fig. S5). Despite the uncertainties associated to the selection of pseudo-absences that represent true absence, models based in presence-absence data consistently outperform presence-only modelling methods (Elith *et al.*, 2006).

## 5.4 Sources of uncertainty and limitations

Using rock-art as source of observational has several associated uncertainties. First, it is unknown if the location of rock-art translates accurately the presence of the depicted animal, as rock-art may represent animals observed in distinct localities from where it was engraved. Still, the large pixel size used in the current work will probably minimise uncertainties relative to the location of the animals against depiction site. Second, the dating of rock-art is rather subjective (Bednarik, 2002), which may affect their use in model transfer between time periods. For this reason, in the current work it was used two time periods of the Holocene which are expected to gather most (if not all) of the rock-art engraving time. This approach minimised the potential effects of dating uncertainties in rock-art depictions.

The large pixel size used implied that established relationships between species presence and environmental variation were only general and superficial in comparison to models based in small spatial resolutions. Models calibrated with large pixel sizes are known to substantially overestimate potentially suitable areas in relation to those using finer scales (Seo *et al.*, 2009). This effect is particularly relevant for species with range sizes under 90,000 km<sup>2</sup>, where there is a chance of identifying inappropriate regions if predictor data used are at or greater than 50x50 km (Seo *et al.*, 2009). Given that the range size of both studied species is well over 90,000 km<sup>2</sup>, it is expected that biases are mitigated. In addition, a trade-off was needed in the current study to account for the uncertainties in the Holocene in the distribution of both species (see previous paragraph) and the availability of bioclimatic and land-cover data (see previous section). In fact, using large pixel sizes in ecological modelling is known to average out uncertainties inherent in the datasets used in model calibration (Wiens *et al.*, 2009).

In the current study, a single modelling algorithm was used, Random Forests. Distinct modelling algorithms have different performances (Elith *et al.*, 2006; Araújo & New, 2007; Wisz *et al.*, 2008) and further analyses could be developed to understand how results will vary between the different methodologies. The current study was based in a single GCM (CCSM3), but different GCMs are known to be a major source of variation in model transferability (Wiens *et al.*, 2009; Goberville *et al.*, 2015; Thuiller *et al.*, 2019). Although the Worldclim dataset provides data from multiple GCMs, the Paleoview dataset only provides data according to the CCSM3 circulation model, which limited substantially the modelling exercises.

## 5.5 Guidelines for model transferability and future research

The study here developed allows providing a series of recommendations related to model transferability in time:

1) When the distributions of studied species are not totally known but the ecological niche traits are well known, it is recommend using environmentally restricted pseudo-absences datasets. Still, this approach requires an intimate knowledge about the studied species, which is often lacking in the literature, but may be possible to use in species with high detectability (such as those used in the current study). Theoretically, when distributions are completely known, disk pseudo-absences datasets would perform very similarly and could be the best solution;

2) When projecting models, it is best to use Worldclim bioclimatic variables given the higher availability of spatial resolutions and GCMs. When models need to be developed based in large pixel sizes (50km or above), apparently, both Worldclim and Paleoview datasets give similar results. Still, using fine spatial resolutions, especially for species with relatively small home-range sizes should be preferable;

3) Coupling land-cover together with bioclimatic variables is recommended, since the association of these variables may reflect accurately environmental changes between calibration and projection periods. Nevertheless, the current availability of coarse land-cover predictions for past time periods greatly constrains its usage in models developed for species with small home-range sizes. Given these limitations, trade-offs will need to be established between pixel sizes and data sources.

In relation to the current study conducted, future additional research efforts should be allocated to assess the effects in model projection of using: (i) different modelling algorithms; (ii) different pixel sizes and their influence on model projection and on pseudo-absences performance; (iii) model replicates based in distinct pseudo-absences datasets; (iv) multiple GCMs to average out uncertainties related to climate models; (v) land-cover layers with better spatial resolution.

## 6. References

- Adams, J.M. & Faure, H. (1990) Global atlas of paleovegetation since the last glacial maximum. <http://www.esd.ornl.gov/projects/qen/adams4.html>.
- Allard-Huard, L. (1994) Les secteurs rupestres du sous continent saharien et du Nil. *Dossiers d'Archéologie*, **197**, 70–83.
- Araújo, M., G. Pearson, R., Thuiller, W. & Erhard, M. (2005b) Validation of species-climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.
- Araújo, M.B. & New, M. (2007) Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, **22**, 42–47.
- Araújo, M.B. & Peterson, A.T. (2012) Uses and misuses of bioclimatic modeling. *Ecology*, **93**, 2712–2724.
- Araújo, M.B., Whittaker, R.J., Ladle, R.J. & Erhard, M. (2005a) Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, **14**, 529–538.
- Ashcroft, M.B. (2010) Identifying refugia from climate change. *Journal of Biogeography*, **37**, 1407–1413.
- Ashouri, H., Sorooshian, S., Hsu, K.-L., Bosilovich, M.G., Lee, J., Wehner, M.F. & Collow, A. (2016) Evaluation of NASA's MERRA precipitation product in reproducing the observed trend and distribution of extreme precipitation events in the United States. *Journal of Hydrometeorology*, **17**, 693–711.
- Bagnold, R.A., Myers, O.H., Peel, R.F. & Winkler, H.A. (1939) An expedition to the Gilf Kebir and 'Uweinat, 1938. *The Geographical Journal*, **93**, 281–313.
- Barbet-Massin, M., Jiguet, F., Albert, C.H. & Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, **3**, 327–338.
- Barnosky, A.D. (2008) Climatic change, refugia, and biodiversity: Where do we go from here? An editorial comment. *Climatic Change*, **86**, 29–32.
- Barry, S. & Elith, J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology*, **43**, 413–423.
- Bednarik, R.G. (2002) The dating of rock art: A critique. *Journal of Archaeological Science*, **29**, 1213–1233.
- Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters*, **15**, 365–377.
- Bennett, K.D. & Provan, J. (2008) What do we mean by “refugia”? *Quaternary Science Reviews*, **27**, 2449–2455.

- Blanc, J.J. (2008) *Loxodonta africana*. The IUCN Red List Of Threatened Species 2008. *Loxodonta africana*, **8235**.
- Blanc, J.J., Barnes, R.F.W., Craig, G.C., Dublin, H.T., Thouless, C.R., Douglas-Hamilton, I. & Hart, J.A. (2007) *African elephant status report 2007: an update from the African elephant database*. Occasional Paper Series of the IUCN Species Survival Commission, No. 60 IUCN / SSC Africa Elephant Specialist Group. IUCN, Gland, Switzerland. vi + 309pp.
- Bobrowski, M., Gerlitz, L. & Schickhoff, U. (2017) Modelling the potential distribution of *Betula utilis* in the Himalaya. *Global Ecology and Conservation*, **11**, 69–83.
- Bond, W.J. & Loffel, D. (2001) Introduction of giraffe changes acacia distribution in a South African savanna. *African Journal of Ecology*, **39**, 286–294.
- Bouché, P., Renaud, P.C., Lejeune, P., Vermeulen, C., Froment, J.M., Bangara, A., Fiongai, O., Abdoulaye, A., Abakar, R. & Fay, M. (2010) Has the final countdown to wildlife extinction in Northern Central African Republic begun? *African Journal of Ecology*, **48**, 994–1003.
- Bouchet, P.J., Peterson, A.T., Zurell, D., Dormann, C.F., Schoeman, D., Ross, R.E., Snelgrove, P., Sequeira, A.M.M., Whittingham, M.J., Wang, L., Rapacciuolo, G., Oppel, S., Mellin, C., Lauria, V., Krishnakumar, P.K., Jones, A.R., Heinänen, S., Heikkinen, R.K., Gregr, E.J., Fielding, A.H., Caley, M.J., Barbosa, A.M., Bamford, A.J., Lozano-Montes, H., Parnell, S., Wenger, S. & Yates, K.L. (2019) Better model transfers require knowledge of mechanisms. *Trends in Ecology & Evolution*, **34**, 489–490.
- Boyce, M., Vernier, P., ENielsen, S. & Schmiegelow, K. (2002) Evaluating resource selection functions. *Ecological Modelling*, **157**, 281–300.
- Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5–32.
- Brenneman, R.A., Louis, E.E. & Fennessy, J. (2009) Genetic structure of two populations of the Namibian giraffe Genetic structure of two populations of the Namibian giraffe , *Giraffa camelopardalis angolensis*. *African Journal of Ecology*, **47**, 720–728.
- Breuil, H. (1923) Station de gravures rupestres d'Aguiet Abderrahman (Sahara Central). *Anthropologie*, **33**, 156–160.
- Brito, J.C., Durant, S.M., Pettorelli, N., Newby, J., Canney, S., Algadafi, W., Rabeil, T., Crochet, P.A., Pleguezuelos, J.M., Wacher, T., de Smet, K., Gonçalves, D.V., da Silva, M.J.F., Martínez-Freiría, F., Abáigar, T., Campos, J.C., Comizzoli, P., Fahd, S., Fellous, A., Garba, H.H.M., Hamidou, D., Harouna, A., Hacha, M.H., Nagy, A., Silva, T.L., Sow, A.S., Vale, C.G., Boratyński, Z., Rebelo, H. & Carvalho, S.B. (2018) Armed conflicts and wildlife decline: Challenges and recommendations for effective conservation policy in the Sahara-Sahel. *Conservation Letters*, **11**, 1–13.

- Buisson, L., Thuiller, W., Casajus, N., Lek, S. & Grenouillet, G. (2010) Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, **16**, 1145–1157.
- Busby, R. (1991) *BIOCLIM – a bioclimate analysis and prediction system*, (ed. by C.R. Margules) and M.P. Austin) CSIRO: Melbourne.
- Butchart, S., Collen, B., Strien, A. Van, Nat, C., Bruno, J., Chanson, J., Galloway, J.N. & Genovesi, P. (2010) Global Biodiversity: Indicators of Recent Declines. *Science*, **5982**, 1164–1169.
- Cameron, E.Z. & Du Toit, J.T. (2005) Social influences on vigilance behaviour in giraffes, *Giraffa camelopardalis*. *Animal Behaviour*, **69**, 1337–1344.
- Ceballos, G. & Ehrlich, P.R. (2002) Mammal population losses and the extinction crisis. *Science*, **296**, 904–907.
- Ceballos, G., Ehrlich, P.R. & Dirzo, R. (2017) Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences*, **114**, 201704949.
- Ciofalo, I. (1995) West Africa 's last giraffes : The Conflict between development and conservation. *Journal of Tropical Ecology*, **11**, 577–588.
- Ciofalo, I. & Le Pendu, Y. (2002) The feeding behaviour of giraffe in Niger. *Mammalia*, **66**, 183–194.
- Claussen, M. & Gayler, V. (1997) The greening of the Sahara during the mid-Holocene: Results of an interactive atmosphere-biome model. *Global Ecology and Biogeography Letters*, **6**, 369–377.
- Comstock, K.E., Georgiadis, N., Pecon-Slattey, J., Roca, A.L., Ostrander, E.A., O'Brien, S.J. & Wasser, S.K. (2002) Patterns of molecular genetic variation among African elephant populations. *Molecular Ecology*, **11**, 2489–2498.
- Crawford-Cabral, J. & Veríssimo, L.N. (2005) The Ungulate fauna of Angola. Systematic list, distribution maps, database report. In: *Estudos Ensaios e Documentos*, 163, 277. Instituto de Investigação Científica Tropical, Lisboa, Portugal.
- Cremschi, M. (1996) The rock varnish in the Messak Settafet (Fezzan, Libyan Sahara), age, archaeological context, and paleo-environmental implication. *Geoarchaeology - An International Journal*, **11**, 393–421.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Kyle, T., Gibson, J., Lawler, J.J., Beard, H. & Hess, T. (2007) Random Forests for classification in ecology Published by : Ecological Society of America. *Ecology*, **88**, 2783–2792.
- De Leeuw, J., Waweru, M.N., Okello, O.O., Maloba, M., Nguru, P., Said, M.Y., Aligula, H.M., Heitkönig, I.M.A. & Reid, R.S. (2001) Distribution and diversity of wildlife in northern Kenya in relation to livestock and permanent water points. *Biological*

- Conservation*, **100**, 297–306.
- Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B. & Singer, A. (2012) Correlation and process in species distribution models: Bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131.
- Drake, N.A., Blench, R.M., Armitage, S.J., Bristow, C.S. & White, K.H. (2011) Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proceedings of the National Academy of Sciences*, **108**, 458–462.
- Drews, C. (1995) Road kills of animals by public traffic in Mikumi National Park, Tanzania, with notes on baboon mortality. *African Journal of Ecology*, **33**, 89–100.
- Elith, J., Graham, C., Anderson, R., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., Peterson, A.T., Phillips, S., Richardson, K., Scachetti-Pereira, R., Schapire, R., Soberón, J., Williams, S., Wisz, M. & Zimmermann, N.E. (2006) Novel methods improve prediction of species' distributions from occurrence data. *European Journal of Biochemistry*, **29**, 129–151.
- Elith, J. & Leathwick, J.R. (2009) Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.
- Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. **41**, 263–274.
- ESRI (2010) ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.
- Evans, M.E.K., Merow, C., Record, S., McMahon, S.M. & Enquist, B.J. (2016) Towards process-based range modeling of many species. *Trends in Ecology and Evolution*, **31**, 860–871.
- Faurby, S. & Araújo, M.B. (2018) Anthropogenic range contractions bias species climate change forecasts. *Nature Climate Change*, **8**, 252–256.
- Fennessy, J. (2009a) Home range and seasonal movements of *Giraffa camelopardalis angolensis* in the northern Namib Desert. *African Journal of Ecology*, **47**, 318–327.
- Fennessy, J., Bidon, T., Reuss, F., Kumar, V., Elkan, P., Nilsson, M.A., Vamberger, M., Fritz, U. & Janke, A. (2016) Multi-locus analyses reveal four giraffe species instead of one. *Current Biology*, **26**, 2543–2549.
- Fennessy, J. & Brown, D. (2010) *Giraffa camelopardalis*. <[www.iucnredlist.org](http://www.iucnredlist.org)>, IUCN 2010. IUCN List of threatened Species. Version 2010.4.,.
- Fordham, D.A., Saltré, F., Haythorne, S., Wigley, T.M.L., Otto-Bliesner, B.L., Chan, K.C.

- & Brook, B.W. (2017) PaleoView: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography*, **40**, 1348–1358.
- Fourcade, Y., Besnard, A.G. & Secondi, J. (2018) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecology and Biogeography*, **27**, 245–256.
- Fuchs, G. (1989) Rock Engravings in the Wadi el Barramiya Eastern Desert of Egypt. *African Archaeological Review*, **7**, 127–153.
- Garcia, R.A., Burgess, N.D., Cabeza, M., Rahbek, C. & Araújo, M.B. (2012) Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Global Change Biology*, **18**, 1253–1269.
- Gautier, A. & Muzzolini, A. (1991) *The life and times of the giant buffalo alias Bubalus/Homoioceras/Pelorovis Antiquus in North Africa*. *Archaeozoologia*, **4**, 39-92.
- Gautier, E.F. (1935a) Anciennes voies du commerce trans-Saharien. *Geografiska Annaler*, **17**, 550–562.
- Gautier, E.F. (1935b) The ancestors of the Tuaregs. *Geographical Review*, **25**, 12–20.
- Goberville, E., Beaugrand, G., Hautekèete, N.C., Piquot, Y. & Luczak, C. (2015) Uncertainties in the projection of species distributions related to general circulation models. *Ecology and Evolution*, **5**, 1100–1116.
- 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.
- Guagnin, M. (2014) Animal engravings in the central Sahara: A proxy of a proxy. *Environmental Archaeology*, **20**, 52–65.
- Guagnin, M., Jennings, R.P., Clark-Balzan, L., Groucutt, H.S., Parton, A. & Petraglia, M.D. (2015) Hunters and herders: Exploring the Neolithic transition in the rock art of Shuwaymis, Saudi Arabia. *Archaeological Research in Asia*, **4**, 3–16.
- Guagnin, M., Jennings, R.P., Eager, H., Parton, A., Stimpson, C., Stepanek, C., Pfeiffer, M., Groucutt, H.S., Drake, N.A., Alsharekh, A. & Petraglia, M.D. (2016) Rock art imagery as a proxy for Holocene environmental change: A view from Shuwaymis, NW Saudi Arabia. *The Holocene*, **26**, 1822-1834.
- Guagnin, M., Shipton, C., el-Dossary, S., al-Rashid, M., Moussa, F., Stewart, M., Ott, F., Alsharekh, A. & Petraglia, M.D. (2018) Rock art provides new evidence on the biogeography of kudu (*Tragelaphus imberbis*), wild dromedary, aurochs (*Bos primigenius*) and African wild ass (*Equus africanus*) in the early and middle Holocene of north-western Arabia. *Journal of Biogeography*, **45**, 727–740.



- 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**, 89–100.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Hall, P. (1976) Priorities for wildlife conservation in North-Eastern Nigeria. *Nigerian Field*, **41**, 99–112.
- Hassanin, A., Ropiquet, A., Gourmand, A.L., Chardonnet, B. & Rigoulet, J. (2007) Mitochondrial DNA variability in *Giraffa camelopardalis*: consequences for taxonomy, phylogeography and conservation of giraffes in West and central Africa. *Comptes Rendus - Biologies*, **330**, 265–274.
- Hastie, T. & Robert, T. (1990) *Generalized additive models*. Chapman and Hall.
- Hernandez, P., Graham, C., Master, L. & Albert, D. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Acquisitions medicales recentes*, **29**, 773–785.
- Higgins, P.A.T. (2007) Biodiversity loss under existing land use and climate change: An illustration using northern South America. *Global Ecology and Biogeography*, **16**, 197–204.
- 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., Phillips, S., Leathwick, J. & Elith, J. (2015) Dismo: Species Distribution Modeling. R Package.
- Hof, C., Levinsky, I., Araújo, M.B. & Rahbek, C. (2011) Rethinking species' ability to cope with rapid climate change. *Global Change Biology*, **17**, 2987–2990.
- Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.
- Joppa, L.N., O'connor, B., Visconti, P., Smith, C., Geldmann, J., Hoffmann, M., Watson, M., Butchart, S.H.M., Virah-Sawmy, M., Halpern, B.S., Ahmed, S.E., Balmford, A., Sutherland, W.J., Harfoot, M., Hilton-Taylor, C., Foden, W., Di Minin, E., Pagad, S., Genovesi, P., Hutton, J. & Burgess, N.D. (2016) Filling in biodiversity threat gaps. *Science*, **352**, 416–418.
- Lee, P.C. & Graham, M.D. (2006) African elephants and human–elephant interactions: implications for conservation. *International Zoo Yearbook*, **40**, 9–19.
- Lek, S. & Guégan, J.F. (1999) Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, **120**, 65–73.
- Le Pendu, Y., Ciofolo, I. & Gosser, A. (2000) The social organization of giraffes in Niger.

*African Journal of Ecology*, **38**, 78–85.

- Leroy, R., de Visscher, M.N., Halidou, O. & Boureima, A. (2009) The last african white giraffes live in farmers' fields. *Biodiversity and Conservation*, **18**, 2663–2677.
- Lindsey, P., Balme, G., Becker, M., Begg, C., Bento, C., Bocchino, C., Dickman, A., Diggle, R., Eves, H., Henschel, P., Lewis, D., Marnewick, K., Mattheus, J., McNutt, J.W., McRobb, R., Midlane, N., Milanzi, J., Morley, R., Murphree, M., Nyoni, P., Opyene, V., Phadima, J., Purchase, N., Rentsch, D., Roche, C., Shaw, J., van der Westhuizen, H., Van Vliet, N., Zisadza, P. Illegal hunting and the bush-meat trade in savanna Africa: drivers, impacts and solutions to address the problem. Panthera/Zoological Society of Lon FAO, don/Wildlife Conservation Society report, New York. 79 pages
- Liu, C., Newell, G. & White, M. (2019) The effect of sample size on the accuracy of species distribution models: considering both presences and pseudo-absences or background sites. *Ecography*, **42**, 535–548.
- Loarie, S.R., Carter, B.E., Hayhoe, K., McMahon, S., Moe, R., Knight, C.A. & Ackerly, D.D. (2008) Climate change and the future of California's endemic flora. *PLoS ONE*, **3**, e2502.
- Lobo, J.M., Jiménez-Valverde, A. & Hortal, J. (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, **33**, 103–114.
- Luoto, M., Virkkala, R. & Heikkinen, R.K. (2007) The role of land cover in bioclimatic models depends on spatial resolution. *Global Ecology and Biogeography*, **16**, 34–42.
- Maiorano, L., Cheddadi, R., Zimmermann, N.E., Pellissier, L., Petitpierre, B., Pottier, J., Laborde, H., Hurdu, B.I., Pearman, P.B., Psomas, A., Singarayer, J.S., Broennimann, O., Vittoz, P., Dubuis, A., Edwards, M.E., Binney, H.A. & Guisan, A. (2013) Building the niche through time: Using 13,000 years of data to predict the effects of climate change on three tree species in Europe. *Global Ecology and Biogeography*, **22**, 302–317.
- Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K. & Thuiller, W. (2009) Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**, 59–69.
- Martínez-Freiría, F., Tarroso, P., Rebelo, H. & Brito, J.C. (2016) Contemporary niche contraction affects climate change predictions for elephants and giraffes. *Diversity and Distributions*, **22**, 432–444.
- Masferrer, N.S., Plaja, J.U., Pujol, J.E. & Salamé, C.S. (2001) *Sluguilla Lawash*, an open air site with rRock Art in the Western Sahara. *Les Premiers Hommes Modernes de La Péninsule Ibérique: Actes Du Colloque de La Commission VIII de l'UISPP* (ed. by J. Zilhão, T. Aubry, and A.F. Carvalho), pp. 281–291. Instituto Português de

Arqueologia, Lisboa.

- Masseti, M. (2010) Holocene mammals of Libya: A biogeographical, historical and archaeozoological approach. *Journal of Arid Environments*, **74**, 794–805.
- Matthiopoulos, J., Hebblewhite, M., Aarts, G. & Fieberg, J. (2011) Generalized functional responses for species distributions. *Ecology*, **92**, 583–589.
- Mc Pherson, J., Jetz, W. & Rogers, D. (2004) The effects of species' range sizes on the accuracy of distribution models. *Journal of Applied Ecology*, **41**, 811–823.
- McHugh, W. (1974) Late prehistoric cultural adaptation in southwest Egypt and the problem of the Nilotic origins of Saharan cattle pastoralism. *Journal of the American Research Center in Egypt*, **11**, 9–22.
- Messerli, B.J. & Winiger, M. (1992) Climate, environmental change, and resources of the African Mountains from the Mediterranean to the Equator. *Mountain Research and Development*, **12**, 315–336.
- Mi, C., Huettmann, F., Guo, Y., Han, X. & Wen, L. (2017) Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ*, **5**, e2849.
- Morales-Barbero, J. & Vega-Álvarez, J. (2019) Input matters matter: Bioclimatic consistency to map more reliable species distribution models. *Methods in Ecology and Evolution*, **10**, 212–224.
- Moreno-Amat, E., Mateo, R.G., Nieto-Lugilde, D., Morueta-Holme, N., Svenning, J.C. & García-Amorena, I. (2015) Impact of model complexity on cross-temporal transferability in Maxent species distribution models: An assessment using paleobotanical data. *Ecological Modelling*, **312**, 308–317.
- Muller, Z., Bercovitch, F., Brand, R., Brown, D., Brown, M., Bolger, D., Carter, K., Deacon, F., Doherty, J.B., Fennessy, J., Fennessy S., Hussein, A.A., Lee, D., Marais, A., Strauss, M., Tutchings, A. & Wube, T. (2018) *Giraffa camelopardalis* (amended version of 2016 assessment). *The IUCN Red List of Threatened Species*, **8235**, e.T9194A136266699.
- Murray, G.W. (1951) The Egyptian climate: an historical outline. *Geographic Journal*, **117**, 422–434.
- Nasseri, N. a., Mcbrayer, L.D. & Schulte, B. a (2011) The impact of tree modification by African elephant (*Loxodonta africana*) on herpetofaunal species richness in northern Tanzania. *African Journal of Ecology*, **49**, 133–140.
- Navone, S., Bargiela, M., Dent, D., Espoz-Alsina, C., Maggi, A., Montaña, E., Morrison, S., Pastor, G., Rosatto, H.G., Salomón, M., Soria, D., Torres, L. & Fidel, R. (2006) *Global Deserts Outlook*, (ed. by E. Ezcurra) UNEP/Earthprint.
- Naylor, R.T. (2004) The underworld of ivory. *Crime, Law and Social Change*, **42**, 261–295.

- O'Connor, D., Stacy-Dawes, J., Muneza, A., Fennessy, J., Gobush, K., Chase, M.J., Brown, M.B., Bracis, C., Elkan, P., Zaberirou, A.R.M., Rabeil, T., Rubenstein, D., Becker, M.S., Phillips, S., Stabach, J.A., Leimgruber, P., Glikman, J.A., Ruppert, K., Masiaine, S. & Mueller, T. (2019) Updated geographic range maps for giraffe, *Giraffa* spp., throughout sub-Saharan Africa, and implications of changing distributions for conservation. *Mammal Review*, mam.12165.
- Olden, J.D., Lawler, J.J. & Poff, L.N. (2008) Machine Learning Methods Without Tears: A Primer For Ecologists. *The Quarterly Review Of Biology*, **83**, 171–193.
- Parker, D.M. & Bernard, R.T.F. (2005) The diet and ecological role of giraffe (*Giraffa camelopardalis*) introduced to the Eastern Cape, South Africa. *Journal of Zoology*, **267**, 203–210.
- Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.
- Pearson, R. & Dawson, T. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.
- Pearson, R.G., Dawson, T.P., Berry, P.M. & Harrison, P.A. (2002) SPECIES: A spatial evaluation of climate impact on the envelope of species. *Ecological Modelling*, **154**, 289–300.
- Pearson, R.G., Terence, T.P. & Liu, C. (2004) Modelling species distributions in Britain: A hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285–298.
- Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P. & Lees, D.C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704–1711.
- Pereira, H.M., Leadley, P.W., Proença, V., Alkemade, R., Scharlemann, J.P.W., Fernandez-Manjarrés, J.F., Araújo, M.B., Balvanera, P., Biggs, R., Cheung, W.W.L., Chini, L., Cooper, H.D., Gilman, E.L., Guénette, S., Hurtt, G.C., Huntington, H.P., Mace, G.M., Oberdorff, T., Revenga, C., Rodrigues, P., Scholes, R.J., Sumaila, U.R. & Walpole, M. (2010) Scenarios for global biodiversity in the 21st century. *Science*, **330**, 1496–1501.
- Peterson, A.T. (2011) Ecological niche conservatism: A time-structured review of evidence. *Journal of Biogeography*, **38**, 817–827.
- Peterson, A.T., Papeş, M. & Soberón, J. (2015) Mechanistic and correlative models of ecological niches. *European Journal of Ecology*, **1**, 28–38.
- Peterson, A.T., Soberón, J. & Sánchez-Cordero, V. (1999) Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265–1267.

- Petit-Maire, N., Celles, J.C., Commelin, D., Delibrias, G. & Raimbault, M. (1983) The Sahara in northern Mali: man and his environment between 10,000 and 3500 years bp. (Preliminary results). *The African Archaeological Review*, **1**, 105–125.
- Philippis, S.J. and Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.
- Phillips, S., Anderson, R. & Schapire, R. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–252.
- Phillips, S.J., Dudík, M. & Schapire, R.E. (2019) [Internet] Maxent software for modeling species niches and distributions.
- Pimm, S. (2008) Biodiversity: Climate change or habitat loss - Which will kill more species? *Current Biology*, **18**, 117–119.
- Pimm, S., Raven, P., Peterson, A., Sekercioglu, C.H. & Ehrlich, P.R. (2006) Human impacts on the rates of recent, present, and future bird extinctions. *Proceedings of the National Academy of Sciences*, **103**, 10941–10946.
- Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M. & Sexton, J.O. (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, **344**, 1246752.
- Pliny the Elder (23–79 CE) *Natural history, volume VIII: books 28-32*. Translated by W. H. S. Jones. Loeb Classical Library 418. Cambridge, MA: Harvard University Press, 1963.
- Poilecot, P., N'Gakoutou, E.B. & Taloua, N. (2010) Evolution of large mammal populations and distribution in Zakouma National Park (Chad) between 2002 and 2008. *Mammalia*, **74**, 235–246.
- Ponder, W.F., Carter, G.A., Flemons, P. & Chapman, R.R. (2001) Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology*, **15**, 648–657.
- Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.
- Ramirez-Villegas, J. & Jarvis, A. (2010) *Downscaling Global Circulation Model outputs: The delta method decision and policy analysis working paper no. 1*. International Center for Tropical Agriculture (CIAT). Cali. CO. 18 p.
- Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.
- Redford, S. & Redford, D.B. (1989) Graffiti and petroglyphs old and new from the Eastern Desert. *American Research Center in Egypt*, **26**, 3–49.
- Roca, A.L., Georgiadis, N., Pecon-Slattery, J. & O'Brien, S.J. (2001) Genetic evidence for

- two species of elephant in Africa. *Science*, **293**, 1473–1477.
- Roever, C.L., van Aarde, R.J. & Leggett, K. (2012) Functional responses in the habitat selection of a generalist mega-herbivore, the African savannah elephant. *Ecography*, **35**, 972–982.
- Rull, V. (2009) Microrefugia. *Journal of Biogeography*, **36**, 481–484.
- Sayer, J.A. (1977) Conservation of large mammals in the Republic of Mali. *Biological Conservation*, **12**, 245–263.
- Scott, J.M. (2002) *Predicting species occurrences : issues of accuracy and scale*. Island Press, Washington, D.C., 868 pp.
- Seo, C., Thorne, J.H., Hannah, L. & Thuiller, W. (2009) Scale effects in species distribution models: Implications for conservation planning under climate change. *Biology Letters*, **5**, 39–43.
- Sequeira, A.M.M., Bouchet, P.J., Yates, K.L., Mengersen, K. & Caley, M.J. (2018) Transferring biodiversity models for conservation: Opportunities and challenges. *Methods in Ecology and Evolution*, **9**, 1250–1264.
- Shaw, B.D. (1976) Climate, environment and prehistory in the Sahara. *World Archaeology*, **8**, 133–149.
- Shorrocks, B. & Croft, D.P. (2009) Necks and networks: A preliminary study of population structure in the reticulated giraffe (*Giraffa camelopardalis reticulata* de Winston). *African Journal of Ecology*, **47**, 374–381.
- Simoneau, A. (1976) Les rhinoceros dans les gravures rupestres du Dra-bani. *Antiquités africaines*, **10**, 7–31.
- Soberón, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species distributional areas. *Biodiversity Informatics*, **2**, 1–10.
- Soria-Auza, R.W., Kessler, M., Bach, K., Barajas-Barbosa, P.M., Lehnert, M., Herzog, S.K. & Böhner, J. (2010) Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. *Ecological Modelling*, **221**, 1221–1229.
- Suraud, J.P., Fennessy, J., Bonnaud, E., Issa, A.M., Fritz, H. & Gaillard, J.M. (2012) Higher than expected growth rate of the Endangered West African giraffe *Giraffa camelopardalis peralta*: A successful human-wildlife cohabitation. *Oryx*, **46**, 577–583.
- Tarroso, P., Carvalho, S.B. & Brito, J.C. (2012) Simapse - simulation maps for ecological niche modelling. *Methods in Ecology and Evolution*, **3**, 787–791.
- Team, R.C. (2019) R: A language and environment for statistical computing.
- Thuiller, W., Araújo, M.B. & Lavorel, S. (2004) Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography*, **31**, 353–361.

- Thuiller, W., Guéguen, M., Renaud, J., Karger, D.N. & Zimmermann, N.E. (2019) Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications*, **10**, 1–9.
- Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009) BIOMOD - A platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.
- Thuiller, W., Lavorel, S. & Araújo, M.B. (2003) Generalized models vs classification tree analysis: Predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, **14**, 669–680.
- Tillet, T. (1985) The Palaeolithic and its environment in the northern part of the Chad basin. *The African Archaeological Review*, **3**, 163–177.
- Tittensor, D.P., Walpole, M., Hill, S.L.L., Boyce, D.G., Britten, G.L., Burgess, N.D., Butchart, S.H.M., Leadley, P.W., Regan, E.C., Alkemade, R., Baumung, R., Bellard, C., Bouwman, L., Bowles-Newark, N.J., Chenery, A.M., Cheung, W.W.L., Christensen, V., Cooper, H.D., Crowther, A.R., Dixon, M.J.R., Galli, A., Gaveau, V., Gregory, R.D., Gutierrez, N.L., Hirsch, T.L., Hoft, R., Januchowski-Hartley, S.R., Karmann, M., Krug, C.B., Leverington, F.J., Loh, J., Lojenga, R.K., Malsch, K., Marques, A., Morgan, D.H.W., Mumby, P.J., Newbold, T., Noonan-Mooney, K., Pagad, S.N., Parks, B.C., Pereira, H.M., Robertson, T., Rondinini, C., Santini, L., Scharlemann, J.P.W., Schindler, S., Sumaila, U.R., Teh, L.S.L., van Kolck, J., Visconti, P. & Ye, Y. (2014) A mid-term analysis of progress toward international biodiversity targets. *Science*, **346**, 241–244.
- Townsend Peterson, A., Papeş, M. & Eaton, M. (2007) Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography*, **30**, 550–560.
- Tuanmu, M.N., Viña, A., Roloff, G.J., Liu, W., Ouyang, Z., Zhang, H. & Liu, J. (2011) Temporal transferability of wildlife habitat models: Implications for habitat monitoring. *Journal of Biogeography*, **38**, 1510–1523.
- van Albada, A. (1994a) De nombreux centres culturels. *Dossiers d'Archéologie*, **197**, 22–33.
- van Albada, A. (1994b) L'univers imaginaire. *Dossiers d'Archéologie*, **197**, 60–69.
- van Albada, A. (1994c) La représentation animaux. *Dossiers d'Archéologie*, **197**, 34–45.
- Van der Jeugd, H.P. & Prins, H.H.T. (2000) Movements and group structure of giraffe (*Giraffa camelopardalis*) in Lake Manyara National Park, Tanzania. *Journal of Zoology*, **251**, 15–21.
- van Lavieren, L.P. & Esser, J.D. (1979) Numbers, distribution and habitat preference of large mammals in Bouda Ndjida National Park, Cameroon. *African Journal of Ecology*, **17**, 141–153.

- Wall, J., Wittemyer, G., Klinkenberg, B., LeMay, V. & Douglas-Hamilton, I. (2013) Characterizing properties and drivers of long distance movements by elephants (*Loxodonta africana*) in the Gourma, Mali. *Biological Conservation*, **157**, 60–68.
- Wenger, S.J. & Olden, J.D. (2012) Assessing transferability of ecological models: An underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, **3**, 260–267.
- Werkowska, W., Márquez, A.L., Real, R. & Acevedo, P. (2016) A practical overview of transferability in species distribution modeling. *Environmental Reviews*, **25**, 127–133.
- Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A. & Snyder, M.A. (2009) Niches, models, and climate change: Assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences*, **106**, 19729–19736.
- Wiens, J.J. & Graham, C.H. (2005) Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 519–539.
- Williams, S.E., Shoo, L.P., Isaac, J.L., Hoffmann, A.A. & Langham, G. (2008) Towards an integrated framework for assessing the vulnerability of species to climate change. *PLoS biology*, **6**, e325.
- Winter, S., Fennessy, J. & Janke, A. (2018) Limited introgression supports division of giraffe into four species. *Ecology and Evolution*, **8**, 10156–10165.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., Elith, J., Dudík, M., Ferrier, S., Huettmann, F., Leathwick, J.R., Lehmann, A., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.C., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S.E. & Zimmermann, N.E. (2008) Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, **14**, 763–773.
- Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., Fielding, A.H., Bamford, A.J., Ban, S., Barbosa, A.M., Dormann, C.F., Elith, J., Embling, C.B., Ervin, G.N., Fisher, R., Gould, S., Graf, R.F., Gregr, E.J., Halpin, P.N., Heikkinen, R.K., Heinänen, S., Jones, A.R., Krishnakumar, P.K., Lauria, V., Lozano-Montes, H., Mannocci, L., Mellin, C., Mesgaran, M.B., Moreno-Amat, E., Mormede, S., Novaczek, E., Opiel, S., Ortuño Crespo, G., Peterson, A.T., Rapacciuolo, G., Roberts, J.J., Ross, R.E., Scales, K.L., Schoeman, D., Snelgrove, P., Sundblad, G., Thuiller, W., Torres, L.G., Verbruggen, H., Wang, L., Wenger, S., Whittingham, M.J., Zharikov, Y., Zurell, D. & Sequeira, A.M.M. (2018) Outstanding challenges in the transferability of ecological models. *Trends in Ecology and Evolution*, **33**, 790–802.
- Yi, Y., Kimball, J.S., Jones, L.A., Reichle, R.H. & McDonald, K.C. (2011) Evaluation of



MERRA land surface estimates in preparation for the soil moisture active passive mission. *Journal of Climate*, **24**, 3797–3816.

## 7. Annex

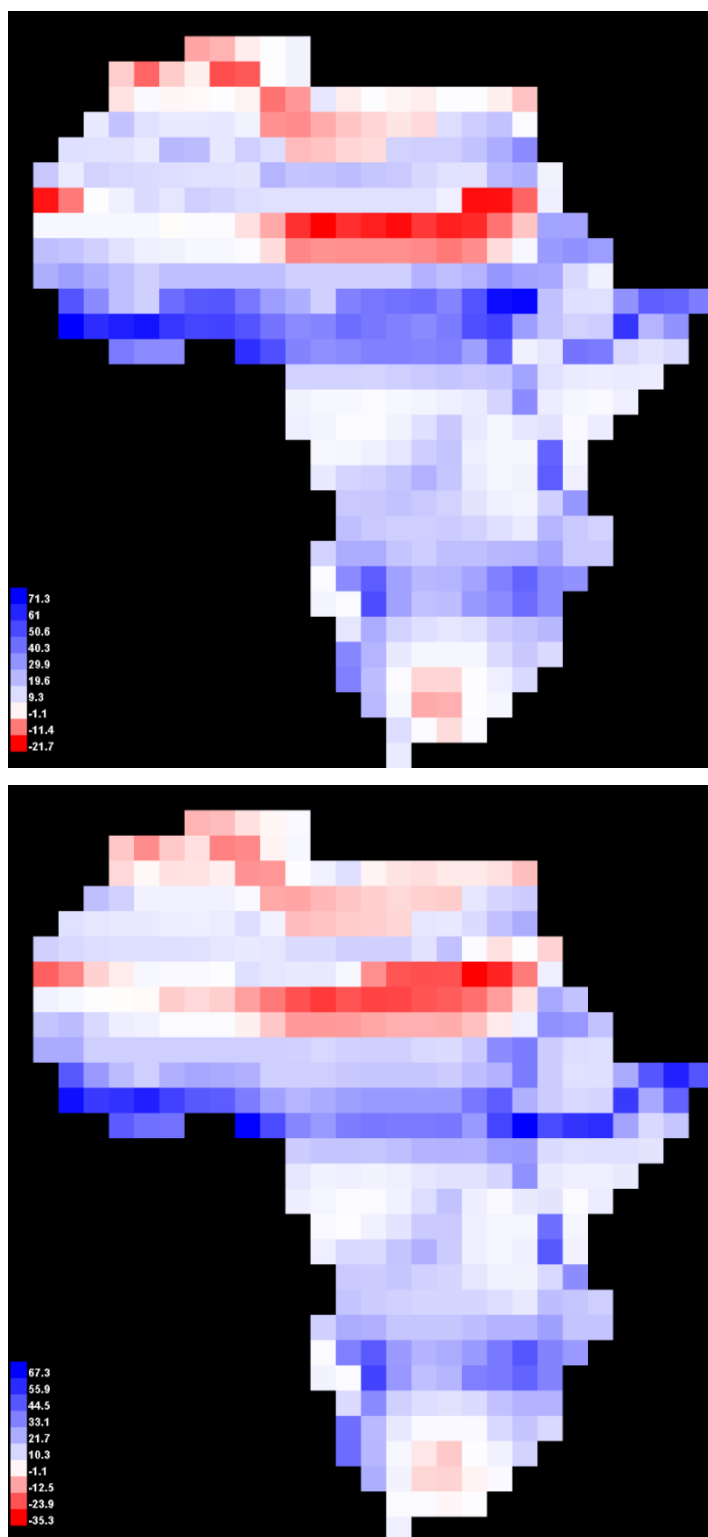


Figure S1 - Comparison of climates between current and 6,000BP (top) and 7,900BP (bottom) conditions, using nine Paleoview bioclimatic variables. Areas in red have one or more environmental variables outside the climatic range present in the model calibration

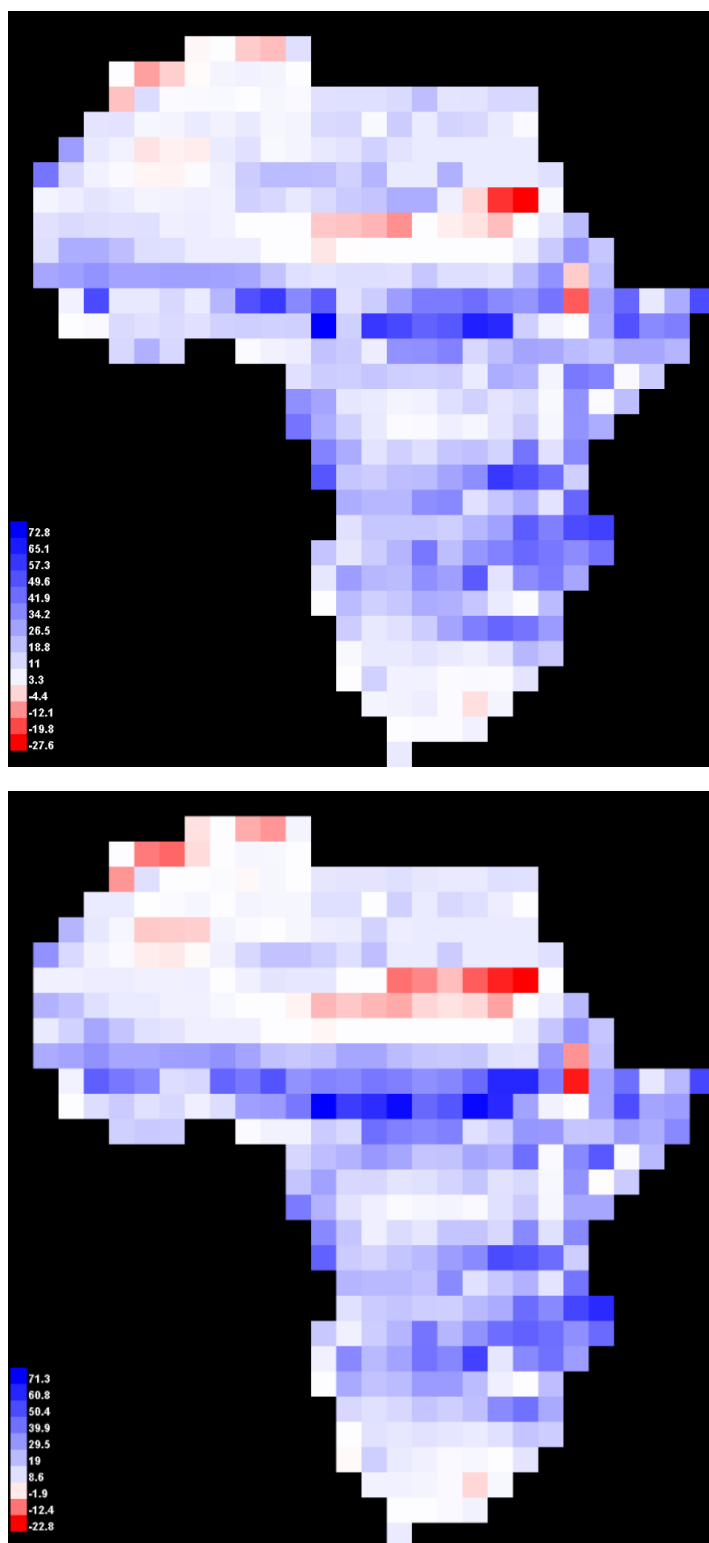


Figure S2 - Comparison of climates between current and 6,000BP (top) and 7,900BP (bottom) conditions, using nine Worldclim bioclimatic variables. Areas in red have one or more environmental variables outside the climatic range present in the model calibration.

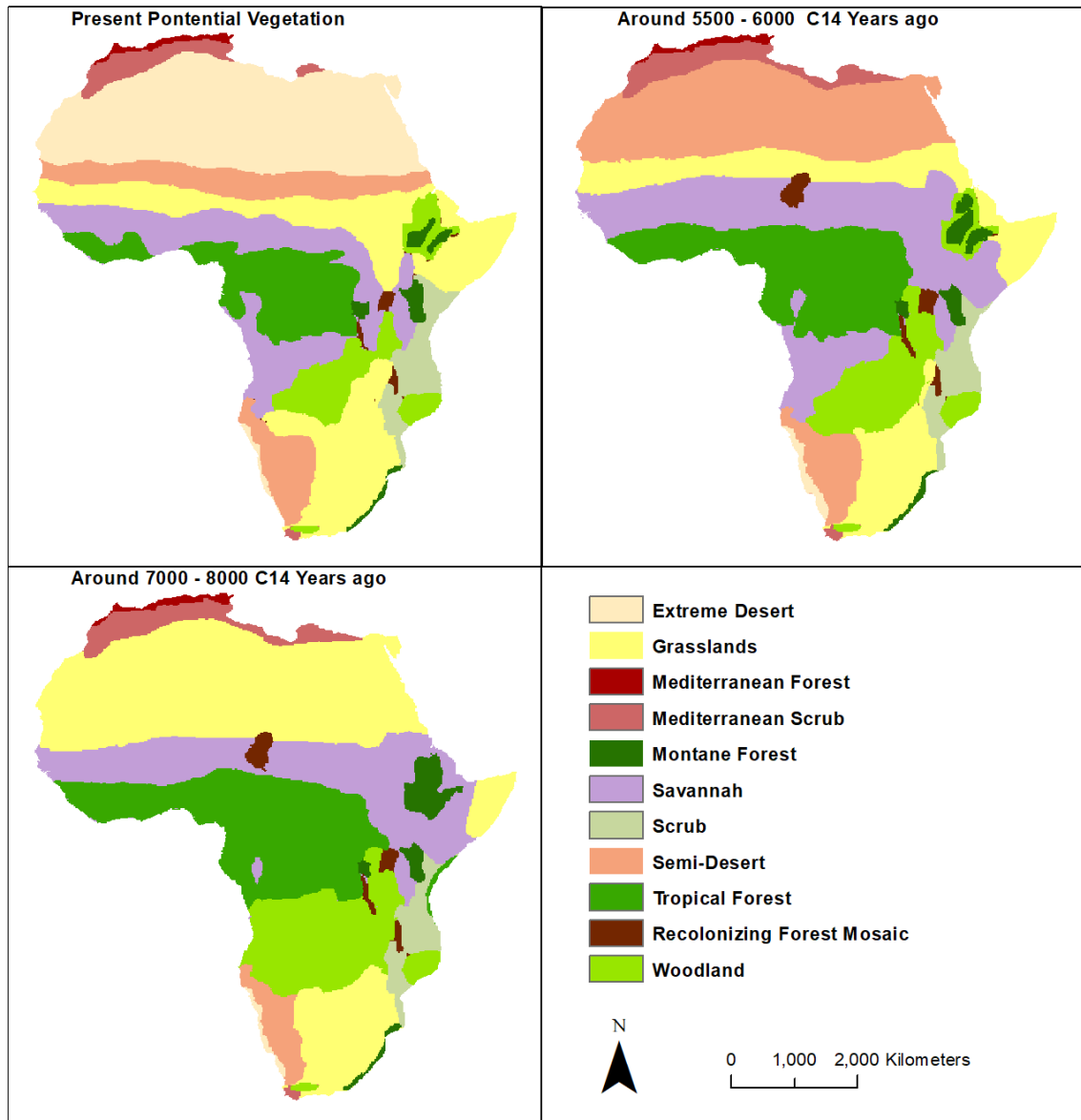


Figure S3 – Distribution of land-cover types in the current and Holocene time periods. Data from Adams and Faure (1990).

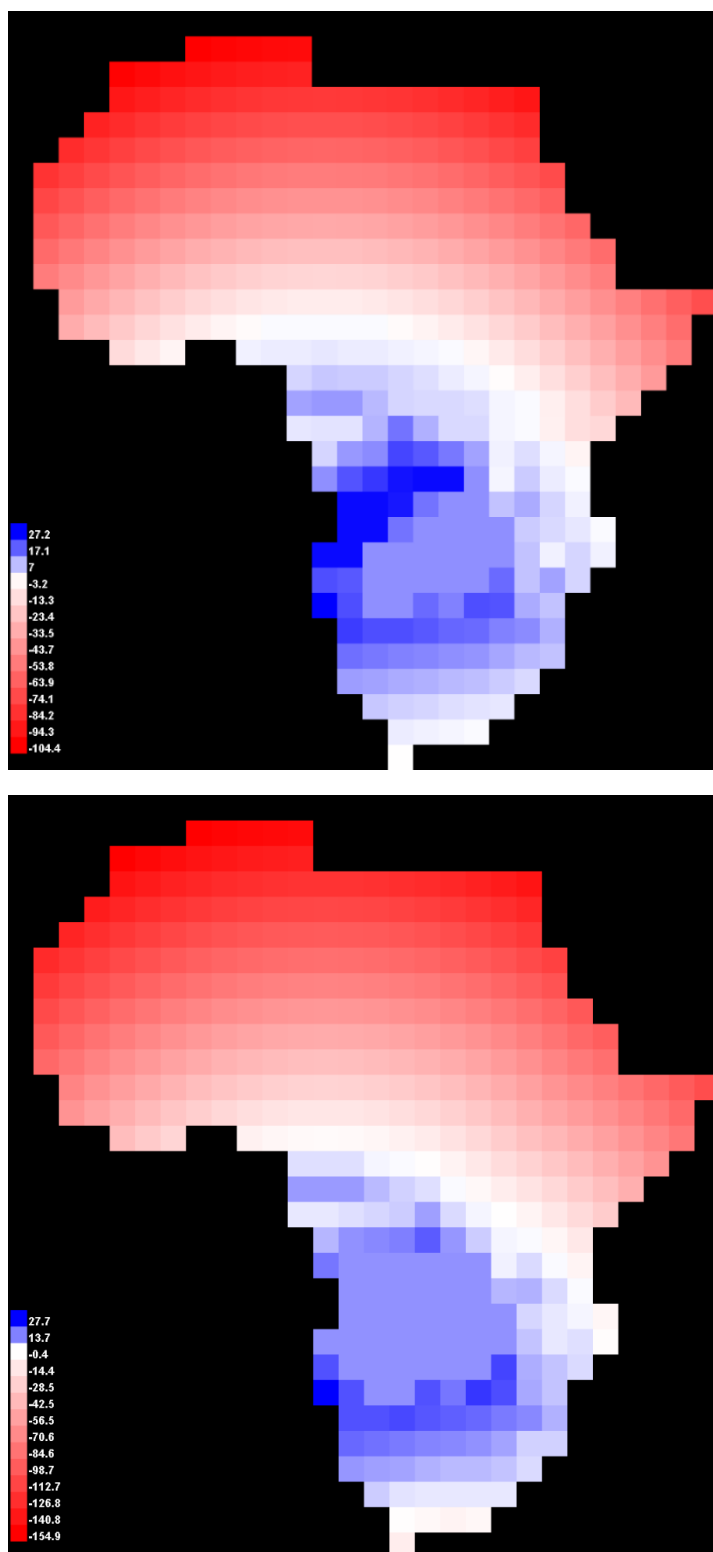


Figure S4 - Comparison of land-covers between current and 6,000BP (top) and 7,900BP (bottom) conditions, using 11 land-cover categories. Areas in red have one or more land-cover variables outside the range present in the model calibration.

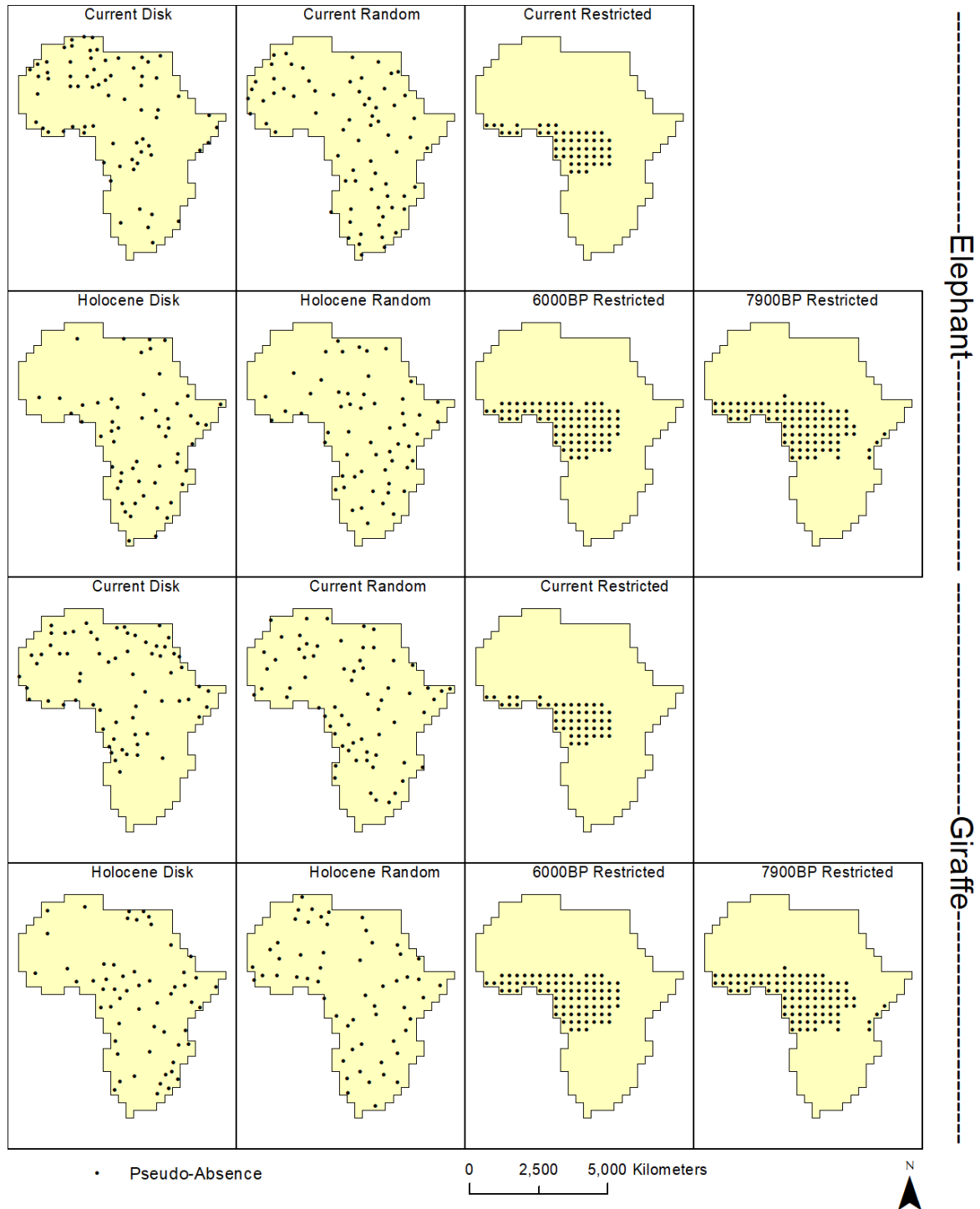


Figure S5 - Distribution of pseudo-absences "Current" (top) and Holocene (bottom) datasets obtained for the Elephant and Giraffe, distinguishing observations used for model training from the full set of available observations (Table 4).

Table S1 – Correlation coefficients between Worldclim bioclimatic variables in current conditions.

BIO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.040																		
3	0.130	-0.502																	
4	-0.150	0.596	-0.919																
5	0.648	0.592	-0.565	0.609															
6	0.681	-0.593	0.706	-0.750	-0.057														
7	-0.074	0.815	-0.879	0.939	0.690	-0.762													
8	0.786	0.116	0.134	-0.153	0.449	0.501	-0.071												
9	0.600	-0.116	-0.050	0.070	0.545	0.434	0.038	0.146											
10	0.760	0.394	-0.486	0.518	0.963	0.116	0.541	0.532	0.622										
11	0.868	-0.276	0.558	-0.613	0.223	0.921	-0.523	0.669	0.487	0.356									
12	-0.011	-0.639	0.752	-0.777	-0.579	0.563	-0.784	-0.055	-0.021	-0.495	0.380								
13	0.086	-0.539	0.678	-0.792	-0.486	0.576	-0.733	0.023	-0.013	-0.420	0.462	0.926							
14	-0.056	-0.465	0.550	-0.387	-0.377	0.343	-0.493	-0.058	0.012	-0.289	0.156	0.572	0.342						
15	0.434	0.229	0.005	-0.178	0.243	0.223	-0.004	0.527	0.069	0.231	0.414	-0.156	0.099	-0.384					
16	0.061	-0.531	0.662	-0.777	-0.490	0.550	-0.717	-0.004	-0.015	-0.430	0.438	0.936	0.989	0.335	0.058				
17	-0.014	-0.523	0.611	-0.454	-0.388	0.426	-0.560	-0.048	0.068	-0.292	0.224	0.666	0.434	0.952	-0.397	0.425			
18	-0.150	-0.607	0.690	-0.695	-0.677	0.404	-0.732	-0.047	-0.240	-0.591	0.206	0.859	0.761	0.527	-0.163	0.756	0.559		
19	0.146	-0.397	0.453	-0.464	-0.187	0.482	-0.471	-0.038	0.244	-0.132	0.371	0.668	0.645	0.397	-0.140	0.666	0.526	0.274	
<b>TOTAL</b>	3	1	4	6	1	4	8	1	0	2	2	5	5	1	0	4	1	4	0

Table S2 – Correlation coefficients between Paleoview bioclimatic variables in current conditions.

BIO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1																			
2	0.348																		
3	0.227	0.347																	
4	0.122	0.760	-0.138																
5	0.644	0.904	0.197	0.728															
6	0.314	-0.764	-0.122	-0.690	-0.470														
7	0.293	0.982	0.193	0.826	0.908	-0.796													
8	0.829	0.377	0.224	0.139	0.576	0.144	0.326												
9	0.554	0.001	0.059	-0.021	0.255	0.402	-0.016	0.261											
10	0.905	0.614	0.131	0.498	0.859	-0.021	0.598	0.801	0.444										
11	0.685	-0.283	0.253	-0.622	-0.020	0.746	-0.368	0.566	0.445	0.364									
12	-0.289	-0.572	0.232	-0.853	-0.668	0.362	-0.629	-0.219	-0.165	-0.578	0.384								
13	-0.247	-0.503	0.132	-0.815	-0.570	0.309	-0.537	-0.130	-0.181	-0.493	0.420	0.921							
14	-0.162	-0.448	0.326	-0.481	-0.497	0.380	-0.521	-0.160	-0.030	-0.344	0.212	0.625	0.403						
15	0.440	0.257	-0.075	0.104	0.391	-0.016	0.276	0.542	0.010	0.482	0.306	-0.240	0.045	-0.466					
16	-0.351	-0.526	0.173	-0.815	-0.648	0.261	-0.568	-0.225	-0.238	-0.603	0.323	0.948	0.971	0.451	-0.082				
17	-0.292	-0.499	0.379	-0.596	-0.617	0.340	-0.584	-0.254	-0.121	-0.501	0.194	0.759	0.551	0.900	-0.452	0.613			
18	-0.413	-0.509	0.176	-0.752	-0.674	0.197	-0.555	-0.259	-0.283	-0.641	0.224	0.918	0.900	0.468	-0.166	0.954	0.637		
19	-0.360	-0.506	0.315	-0.597	-0.626	0.312	-0.577	-0.285	-0.116	-0.537	0.177	0.715	0.596	0.757	-0.340	0.651	0.876	0.668	
<b>TOTAL</b>	2	4	0	7	4	3	4	2	0	3	1	6	3	2	0	4	3	4	3



Table S3 - AUC values calculated for model predictions for the elephant

<b>Current</b>		<b>6000BP</b>		<b>7900BP</b>	
Worldclim					
DISK PA	1.000	DISK PA	0.997	DISK PA	0.998
RANDOM PA	0.999	RANDOM PA	0.995	RANDOM PA	0.997
RESTRICTED PA	1.000	RESTRICTED PA	1.000	RESTRICTED PA	1.000
Paleoview					
DISK PA	0.999	DISK PA	0.998	DISK PA	0.998
RANDOM PA	0.999	RANDOM PA	0.995	RANDOM PA	0.994
RESTRICTED PA	1.000	RESTRICTED PA	1.000	RESTRICTED PA	1.000
Landcover					
DISK PA	0.998	DISK PA	0.998	DISK PA	9.997
RANDOM PA	0.998	RANDOM PA	0.997	RANDOM PA	0.996
RESTRICTED PA	0.999	RESTRICTED PA	1.000	RESTRICTED PA	1.000
Worldclim+Landcover					
DISK PA	0.999	DISK PA	0.947	DISK PA	0.999
RANDOM PA	1.000	RANDOM PA	0.998	RANDOM PA	0.998
RESTRICTED PA	1.000	RESTRICTED PA	1.000	RESTRICTED PA	1.000
Paleoview+Landcover					
DISK PA	1.000	DISK PA	0.999	DISK PA	0.999
RANDOM PA	0.999	RANDOM PA	0.997	RANDOM PA	0.997
RESTRICTED PA	1.000	RESTRICTED PA	1.000	RESTRICTED PA	1.000

Table S4 – AUC values calculated for model predictions for the giraffe

<b>Current</b>		<b>6000BP</b>		<b>7900BP</b>	
Worldclim					
DISK PA	0.993	DISK PA	0.998	DISK PA	0.998
RANDOM PA	0.998	RANDOM PA	0.996	RANDOM PA	0.997
RESTRICTED PA	0.999	RESTRICTED PA	1.000	RESTRICTED PA	1.000
Paleoview					
DISK PA	0.991	DISK PA	0.996	DISK PA	0.998
RANDOM PA	0.996	RANDOM PA	0.994	RANDOM PA	0.994
RESTRICTED PA	1.000	RESTRICTED PA	1.000	RESTRICTED PA	1.000
Landcover					
DISK PA	0.996	DISK PA	0.995	DISK PA	0.996
RANDOM PA	0.996	RANDOM PA	0.998	RANDOM PA	0.997
RESTRICTED PA	1.000	RESTRICTED PA	1.000	RESTRICTED PA	1.000
Worldclim+Landcover					
DISK PA	0.997	DISK PA	0.997	DISK PA	0.998
RANDOM PA	0.998	RANDOM PA	0.998	RANDOM PA	0.998
RESTRICTED PA	1.000	RESTRICTED PA	1.000	RESTRICTED PA	1.000
Paleoview+Landcover					
DISK PA	0.997	DISK PA	0.998	DISK PA	0.998
RANDOM PA	0.996	RANDOM PA	0.998	RANDOM PA	0.997
RESTRICTED PA	1.000	RESTRICTED PA	1.000	RESTRICTED PA	1.000