Predictive Modelling of Terrestrial Reptile Species Richness for Conservation in Saudi Arabia

Abdulaziz Subhi Alatawi, BSc, MSc

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

February, 2020

Dedication

To my parents

To my wife & family

To those who have taught me

Abstract

Studying species richness distribution patterns depends on data quality, which imposes challenges for conservation action. Species distribution modelling is a powerful tool to overcome some of these hurdles, and can gives us ecological and biogeographical insights about species distributions patterns and habitat suitability even in poorly sampled habitats.

The main objective for this thesis was to model the distribution of 62 terrestrial reptile (lizard and snake) species to obtain the pattern of species richness and habitat suitability across Saudi Arabia. I used both species distribution models (SDMs) built for Egypt (a better-studied area) and then spatially transferred into Saudi conditions, as well as models built with local data. Maxent software was used to interpolate (based on Saudi data) and transfer (based on Egyptian data) models. The accuracies of these models were then evaluated on-the-ground using data from field work conducted in the summer of 2017 and 2018 in Tabuk Province. The detectability and occupancy of each species in the field was estimated using PRESENCE software. I then assessed the conservation value of the proposed and current Protected Area network in Saudi (managed by the Saudi Wildlife Authority) and identified potential new important areas of conservation planning under various different assumptions (socioeconomic and uncertainty analysis) using Zonation software.

The maximum species richness of reptiles was predicted to occur in the central plateau, north-western borders, south-western areas, and in the coastal areas of Saudi Arabia. I identified areas across Saudi Arabia that are considered to be under-sampled relative to these predictions. The transferred Egyptian models yielded very effective predictions, analogous to those generated by local independent models based on Saudi data. Some species were predicted better by the transferred Egyptian model than the equivalent Saudi model. Model accuracy across species as estimated using the independent field data varied among species and correlated positively with accuracy estimated using partitioning of the original

Saudi data, but validation using independent data gave lower accuracy estimates overall than data partitioning. Ground-truthing validation showed that the transferred Egyptian model was able to predict presences better than the Saudi model built from local data, and that the Saudi model was better at capturing the predicted areas of high species richness. Species detectability in the field was weakly positively correlated with model accuracy calculated from the new independent field data, hinting that models work better for species which are easy to survey. The directed field validation successfully led us to record species within Saudi Arabia that had not been recorded before, and extend the known range for a species that is located completely outside the Protected Areas and had not been recorded for 40 years. I found that the conservation value of current and proposed Saudi's PAs was significantly better than the areas outside them across all Zonation models. I proposed possible extensions of some PAs to cover top-ranking areas predicted to be suitable for reptile conservation.

SDM predictions appear to be a very effective tool for exploring spatial patterns of species diversity, especially when there is a paucity of data. When testing SDMs, using new data, collected independently of the data used to build the models, seems to be the best practice to evaluate model predictive performance, but caveats with such data must be considered as they may not be entirely statistically independent. Incorporating different scenarios in conservation assessments and planning can help increase its reliability in identifying potential important areas for conservation.

Acknowledgements

Foremost, all my praise and gratitude to my Lord "Allah" for his countless blessings on me whose facilitates this work and make it achievable.

I would like to express my honest gratitude and appreciation to my supervisors, Dr. Tom Reader and Prof. Francis Gilbert for their valuable guidance and advice through this project. THANK YOU, from the heart, for your assistance and support that made my PhD's journey an unforgettable experience. Thank you for your continued encouragement and for the scientific atmosphere that you made during this period. Thank you for the knowledge that you have taught me and shared with me. I confidently say that I learn so many things, besides research and science, from both of you which will definitely help me through my personal and academic future endeavours.

I would like to express my special thanks and appreciation to my parents for their sacrifices, supplications, and love that help and enthusiasm me to complete my higher education. A special thanks to my wife for her continuous support and understanding through my study. Being abroad for a few years to finish my master and PhD is truly dedicated to them.

I am grateful for the people who have assisted me during this project. I especially thank my friend in the field Mr. Abdullah Tolba for his important contribution to this project through his assistance and friendship during the field work. I thank Dr. Ahmad El-Gabbas for his advice with modelling and assistance running analysis. I thank Dr. Sherif Baha El Din for his valuable assistance with reptile species identification. I thank Dr. Emad Kaky for the advice with modelling.

I am thankful and appreciated for the University of Tabuk and the University of Nottingham for funding this work and providing all the needs to make it possible.

Table of Contents

Table of Figures

Table of Tables

1 Chapter One: Introduction and Background

1.1 The niche concept

Joseph Grinnell (1917) was the scientist who coined the term niche to describe physical environmental conditions that are suitable for species, allowing them to reproduce and survive (Grinnell, 1924; Vandermeer, 1972; Leibold, 1995). The Grinnellian niche includes all possible "limiting" abiotic factors (e.g., rainfall and temperature, etc.,) (Vandermeer, 1972; Leibold, 1995). Elton (1927) added to the niche concept, describing the niche as the role or the place that a species plays in its environment, now known as the Eltonian niche (Vandermeer, 1972; Leibold, 1995; Soberón, 2007). Finally, Gause (1934) introduced the concept now known as the "*competitive exclusion principle"*. Competitive exclusion states that "*two species with same ecological requirements cannot coexist together and one of them will exclude the other*" (Hardin, 1960; Vandermeer, 1972), and hence implies that species with identical niches cannot coexist.

The famous contribution from Hutchinson (1957) formalized the concept of the niche by dividing it into the fundamental and realized niche (Vandermeer, 1972; Leibold, 1995; Pulliam, 2000; Soberón and Peterson, 2005). Hutchinson (1957) defined the fundamental niche as "a "*n-dimensional hypervolume*" within which the environmental abiotic conditions support a species' presence, persistence, and survival in the absence of competition" (Vandermeer, 1972; Leibold, 1995; Pulliam, 2000; Soberón and Peterson, 2005; McInerny and Etienne, 2012). Additionally, Hutchinson described the realized niche as the remaining small portion/part of the fundamental niche in which environmental conditions allow a species to persist in the presence of biotic factors (Hutchinson, 1957; Vandermeer, 1972; Leibold, 1995; Pulliam, 2000; Soberón and Peterson, 2005; McInerny and Etienne, 2012). In other words, if biotic factors are excluded, the species is expected to occupy its fundamental

niche (potential distribution), but if there are biotic limitations, such as those associated with competition or predation, the species is restricted to its realized niche (actual distribution) (Pearson, 2008). This subdivision of the niche concept based on Hutchinson has been applied extensively to explain a wide range of phenomena (Soberón and Peterson, 2005).

At the time, Hutchinson did not realize that this reformulation of the niche concept would be a hotspot of discussion, leading to confusion about the concept (Vandermeer, 1972; Leibold, 1995; Soberón and Peterson, 2005; Kearney, 2006; Colwell and Rangel, 2009). Many scientists have revisited and reviewed the concept since this time (Vandermeer, 1972; Leibold, 1995; Guisan and Zimmermann, 2000; Pulliman, 2000; Soberón and Peterson 2005; Soberón, 2007; Pearson, 2008; Hirzel and Lay, 2008; Colwell and Rangel, 2009; McInerny and Etienne, 2012). It is crucial to distinguish between the realized niche and the fundamental niche to predict accurately the distribution of species in the wild (Guisan and Zimmermann 2000). Pearson (2008) explained how environmental and geographical space could be used to help visualize the fundamental and realized niche.

The "niche" is still an ambiguous term because there is no a clear definition and it is used differently in the literature (Guisan and Zimmermann, 2000; Kearny, 2006; Pearson, 2008; Elith and Leathwick, 2009; Franklin, 2009; McInerny and Etienne, 2012). However, whatever the concept, the goal is always the same: to understand how the surrounding environment and habitats influence species distribution either in the presence of biotic factors or their absence (Franklin, 2009).

1.2 Dispersal limitation and source-sink dynamics

Ecological communities are controlled by two types of process that are very important in shaping communities: local processes affected by the surrounding environments, and regional processes which includes the dispersal of species between different sites (Shurin, 2000). The

basic idea behind dispersal limitation is that not all species are capable of reaching suitable habitats that provide their essential requirements (Pulliam, 1988; Pulliam, 2000; Guisan and Thuiller, 2005). Species with high dispersal ability are expected to inhabit a wide range of habitats, and species with limited dispersal ability mostly have spatially autocorrelated distributions (Pulliam, 2000; Guisan and Thuiller, 2005).

The distribution pattern of species through geographical space and time is controlled by biotic interactions (competition, predation, etc.,) and abiotic interactions (environmental conditions, topography, etc.,), but constrained by accessibility or mobility (e.g., landscape barriers) (Figure 1.1; Soberón and Peterson, 2005). Dispersal limitation is important in directly influencing relationships between species and local communities, constraining distribution patterns, and affecting species-area curves as well (Pulliam, 2000; Soberón and Peterson, 2005; Shen *et al.*, 2009). Species may be recorded as absent from suitable habitats that satisfy their ecological requirements (Pulliam, 1988; Pulliam, 2000; Guisan and Thuiller, 2005). While limited dispersal may prevent species from occupying all the different habitats that are suitable, immigration can also allow persistence of populations in unsuitable habitat. Thus dispersal ability has an effect on the accuracy of SDMs, which may limit its application (Miller and Holloway, 2015).

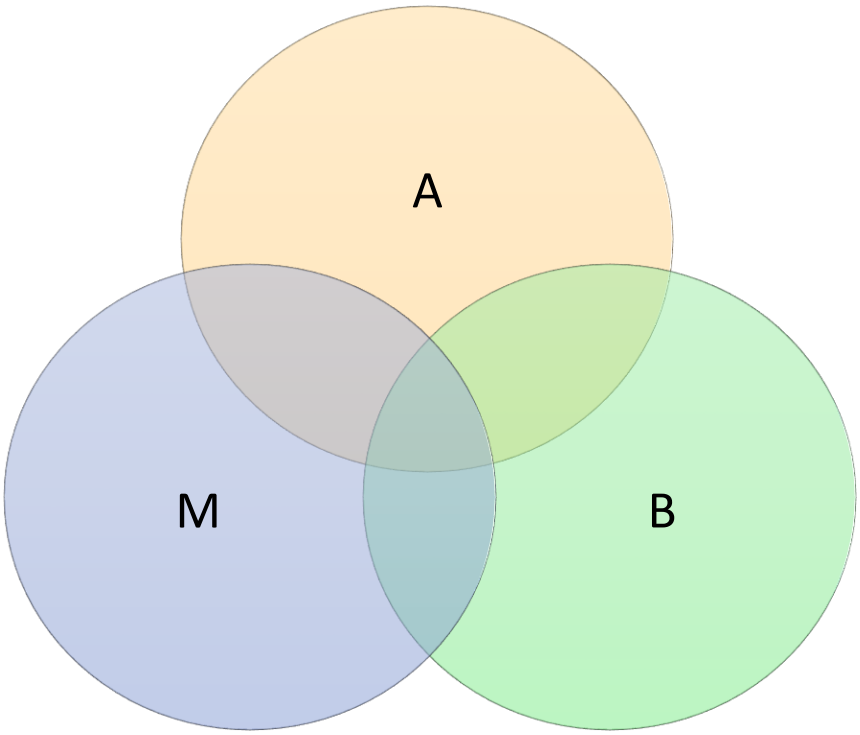


Figure 1.1: Factors which control species distribution. The orange area (A) represents the conditions of abiotic factors that constitutes the fundamental niche. The purple area (M) represents accessibility and mobility of a species without limitation. The green area (B) represents the conditions in which biotic factors allow persistence. The central intersection area is where the habitat is both environmentally suitable and accessible, and hence is where we expect to observe species in nature (modified and cited from Soberón and Peterson, 2005).

A proper understanding of source-sink dynamics (and more generally metapopulation dynamics) is required in order to apply the niche concept to explain species distributions (Watkinson and Sutherland, 1995; Pulliam, 2000). Pulliam (1988) classified habitats into two categories based on their quality and occupied species (Watkinson and Sutherland, 1995; Pulliam, 2000). "Source" habitats provide suitable environmental conditions and support viable populations where the reproductive rate exceeds the mortality rate. In contrast, in "sink" habitats populations are not viable because the mortality rate exceeds the reproductive rate (Pulliam, 1988; Watkinson and Sutherland, 1995; Pulliam, 2000). When sources and sinks are linked via movement, populations in sink habitats depend on immigration to survive and persist (Pulliam, 1988).

A detailed understanding of the species characteristics, habitat types, and factors that influence species movement and limit dispersal must be taken in account in understanding distributions (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005).

1.3 Species distribution models

The spatial locations where species have been recorded are strongly correlated with their surrounding physical environments; attempts to use this relationship to model the distribution are known generally as 'species distribution modelling' (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005; Pearson, 2008). Many terms have been used to describe this relationship, which leads to confusion about which term is the proper one to use (Sillero, 2011; Peterson and Soberón, 2012). For example, the terms "species distribution model" "ecological niche model" "habitat suitability model" "environmental model" "bioclimatic envelope" have all been used in the literature (Guisan and Thuiller, 2005; Hirzel and Lay, 2008; Pearson, 2008; Elith and Leathwick, 2009; Franklin, 2009; Sillero, 2011; Peterson *et al.*, 2011; Peterson and Soberón, 2012). Practitioners of distribution modelling should know

what is being modelled, and what question/purpose is behind the research, in order to choose the proper term and method that fits the study domain (Kearney, 2006; Franklin, 2009; Peterson and Soberón, 2012). However, species distribution modelling (SDM) is the most widely acceptable and preferable term to use as it appears in most of the scientific literature (Pearson, 2008; Franklin, 2009), and this will be the term used in this thesis.

1.4 Type of species distribution models predictions

There are three types of predictions that can be made using distribution models. First, the model can be used to interpolate, which means predicting the habitat suitability of new unvisited sites within the environment extent/range used to calibrate the model. Second, the model can be used to extrapolate, which means to predict into new conditions beyond the environmental conditions for which the original model was calibrated (e.g., new climatic conditions). Third, the model can be used to transfer, which means to predict into new landscapes different from where model was calibrated (Elith and Leathwick, 2009; Peterson *et al.*, 2011; Duque-Lazoa *et al.*, 2016). Hence, species distribution models may be used to identify three parts of a species distribution that are very important to consider: areas that are near to occurrence records where the species is expected to be present; portions of the realized niche that are unknown; and unoccupied areas of the fundamental niche (see Pearson, 2008).

1.5 Correlative species distribution models

There are two types of distribution models, namely *correlative* and *mechanistic* distribution models. Correlative distribution models are the most popular (Kearney, 2006; Pearson, 2008; Kearney and Porter, 2009; Buckley *et al.*, 2010) and the subject of this thesis. Correlative distribution models aim to predict sites that are known to have similar environmental

conditions that are suitable and support species presences using data on previous known locations of species occurrence (Guisan and Zimmermann, 2000; Pearson, 2008). Mechanistic distribution models incorporate the association between biological performance (survival, growth and reproduction) of species and key environmental variables, in order to predict fitness and hence the distribution of habitat in which populations are expected to persist (Pearson, 2008; Kearney and Porter, 2009; Buckley *et al.*, 2010).

Species distribution models have become one of the most active areas of research in ecology and biogeography (Peterson and Soberón, 2012) with hundreds of publications (Lobo *et al.*, 2010; Brotons, 2014). The ability of these models to generate robust predictions of species richness patterns and habitat suitability for remote regions with incomplete data information make them an important conservation tool (Guisan and Thuiller, 2005). Chapter 2 provides a review of correlative SDMs.

1.6 Transferability (generality) of the model

Transference of species distribution modelling is a new technique that is being increasingly used. The basic idea relies on predicting the potential distribution of species in landscapes different from those in which data were gathered (Randin *et al.*, 2006; Elith and Leathwick, 2009; Peterson *et al.*, 2011; Werkowska *et al.*, 2017; Yates *et al.*, 2018). Transference represents a solution for scenarios where biological data are very scarce and few studies have been conducted, especially areas that are remote and inaccessible (Yates *et al.*, 2018). Chapter 3 provides a review of spatial transference using SDMs.

1.7 Mechanistic species distribution models

Broadly speaking, mechanistic distribution models relate the fitness of a species with its surrounding environment to map the fitness components of a species, and thus supply a causal explanation of the distribution (Kearney, 2006; Kearney and Porter, 2009; Buckley *et al.*, 2010). Mechanistic distribution models require a deep understanding of the effect and the magnitude of biotic and abiotic factors on species characteristics and responses (Kearney, 2006; Kearney and Porter, 2009; Buckley *et al.*, 2010). They are potentially very valuable because they provide causal explanations for ecological phenomena that incorporate fundamental principles (Kearney, 2006; Kearney and Porter, 2009). However, correlative models are more widely applicable because they are flexible and straightforward to apply, using even quite limited datasets for species whose biology is poorly known, and can produce useful output in the form of distribution maps. In contrast, mechanistic distribution models require a lot of time, expertise and data, together with a detailed study of the focal species.

1.8 Software and species distribution models

Modern satellite imagery (remote sensing), and Geographic Information System methods mean that digital environmental variables/layers (temperature, rainfall, precipitation, etc.,), topography (aspect, slope, etc.,), and other related variables can be easily downloaded, and plotted (Guisan and Zimmermann, 2000; Salem, 2003; Foody, 2008; Swenson, 2008). As a result, manipulating, storing, and analyzing data has become an easy task from small to global scales (Salem, 2003; Foody, 2008; Swenson, 2008; Elith and Leathwick, 2009). However, despite these developments, predicting species distributions (either through space or time) is still a complex and challenging process because of particular species characteristics (Guisan and Zimmermann, 2000). For instance, species dispersal and mobility may constrain the ability to produce reliable distribution models (Sinclair *et al.*, 2010; Urban

et al., 2016). Furthermore, interactions among species normally are ignored and rarely included in distribution modelling (Sinclair *et al.*, 2010). Most species distribution models have implicitly assumed that the species is in an equilibrium state with its local environment; however, this is not the case in reality for many or even most species (Guisan and Zimmermann, 2000; Sinclair *et al.*, 2010), as in source-sink dynamics, for example. Additionally, the response of a species to environmental variation should be defined by fitness responses, which can change during adaptation to new environmental conditions; adaptation is not usually taken into account in the output of the models (Urban *et al.*, 2016).

1.9 Data inputs to calibrate SDMs

To calibrate species distribution models, two types of input are needed: biological data in the form of species occurrence records, and environmental data for the same region. Species occurrence records reflect either the spatial or temporal distribution, and usually consist of collations of personal records, online resources, citizen observations, museum collections, large surveys and atlases. Nowadays, it is possible to locate the potential distribution of targeted species in any place in the world if the record locations are accurate. The environmental data describe the landscape, climate and topography over an area of interest (Pearson, 2008; Franklin, 2009). Each species distribution model needs to take in account the limitations of these types of data: a model prediction is only as good as the data used to build the model (Pearson, 2008).

1.9.1 Biological data

There is a huge number of species occurrence records and many of these records are available online, although the records for many species are sparse (Graham *et al.*, 2004; Newbold, 2010). However, most of these occurrence records are collected with either spatial or

temporal biases, and therefore caution should be used when building species distribution models using these records (Graham *et al.*, 2004; Newbold, 2010, Newbold *et al.*, 2010). For example, most species have been recorded without systematic sampling, and observations are biased towards roads, cities and other locations that are easy to observe (Graham *et al.*, 2004; Newbold, 2010).

The Global Biodiversity Information Facility is a well-known online resource (GBIF: http://www.gbif.org) which includes information about species occurrences, as well other relevant information. Similarly, VertNet is another website that includes a massive amount of information. It has 125 contributing publishers of records, including universities, many museums, and natural organizations: VertNet claims to have 21,580,767 records for different taxa (http://www.vertnet.org, accessed 25, February 2020). Species records can be obtained from personal collections, which mostly depend on opportunistic observation, and natural history surveys (Pearson, 2008). Finally, museum collections and herbaria are some of the richest places for original species records and specimens, with numerous data from all around the world (Graham *et al.*, 2004; Newbold, 2010). Using data from museum records to calibrate species distribution models should be done with caution because museum data usually contain errors and biases, and often do not explain the locations precisely (Newbold, 2010). Collating, georeferencing and estimating location errors (Wieczorek *et al.*, 2004) for museum records is very laborious. However, data obtained from museums are too valuable to ignore (Newbold *et al.*, 2010; Pardo *et al.*, 2013).

1.9.2 Environmental data

Environmental data should be carefully selected because they can directly impact the model and its outputs (Guisan and Zimmermann, 2000; Austin, 2002; Guisan and Thuiller, 2005; Peterson *et al.*, 2011). There are plenty of sources of environmental variables. WorldClim

(http://www.worldclim.org/) is one of the biggest of the environmental resources, used in many studies (more details in Hijmans *et al.*, 2005).

Austin (1980) proposed and classified environmental gradients/variables into three types that may control species distribution. First, there are resource gradients which reflect variation in resources that are consumed by species, such as water and nutrients. Secondly, there are direct gradients which have physiological and behavioral effects, such as temperature. Finally, there are indirect variables (such as aspect and elevation), which may have no direct physiological impact on species performance but work via their impact on direct factors (Austin, 2002). Austin (2002) categorized environmental gradients into proximal and distal types. Proximal gradients refer to situations in which the responses of species are directly to the gradients (i.e. direct and resource gradients), while distal gradients refer to cases where the responses of species are indirect. Usually, models built using direct and proximal gradients are most likely to capture species distribution accurately, and hence make reliable predictions that can be widely applied, while distal gradients have limited value and are best applicable locally (Austin, 2002). Environmental variables have different impacts on species at different spatial scales (Guisan and Zimmermann, 2000; Pearson and Dawson, 2003; Peterson *et al.*, 2011). For instance, a coarse resolution is usually used for climate variables and topography, while a fine resolution is used for food availability and other biotic variables (Pearson and Dawson, 2003; Peterson *et al.*, 2011).

1.10 Some applications of species distribution models

Species distribution models have been used for many groups of animals and plants and can produce very reliable predictions/descriptions of distributions in many cases (Guisan and Zimmermann, 2000; Raxworthy *et al.*, 2003; Wintle *et al.*, 2005; Arntzen, 2006; Pearson *et*

al., 2007; Trisurat *et al.*, 2012; Guisan *et al.*, 2013; Yang *et al.*, 2013; Yi *et al.*, 2016; Kaky and Gilbert, 2016).

1.10.1 Conservation applications

Species distribution models play an important role in the conservation process, and are used to help make decisions relating to the protection of species and the maintenance of biodiversity, especially for threatened and poorly known species and habitats (Ferrier, 2002; Franklin, 2009; Sinclair *et al.*, 2010; Ferraz *et al.*, 2012; Guisan *et al.*, 2013; Liu *et al.*, 2013; Marshall *et al.*, 2014). However, most species distribution modelling studies have not directly addressed conservation priorities (Cayuela *et al.*, 2009; Guisan *et al.*, 2013), and the need for more work that focuses particularly on applying distribution models in conservation planning has been emphasized (Cayuela *et al.*, 2009; Ferraz *et al.*, 2012; Guisan *et al.*, 2013).

There are many useful applications of species distribution models in conservation. First, models can be used to identify habitats with diverse or rare communities of organisms (Raxworthy *et al.*, 2003; Guisan *et al.*, 2013; Liu *et al.*, 2013; El-Gabbas *et al.*, 2016; Kaky and Gilbert, 2016). It is essential to protect such habitats to maintain global biodiversity and ecosystem services (Chapin *et al.*, 2000). Well-designed species distribution models can distinguish between high-quality habitats and poor-quality habitats, and to identify regional biodiversity "hotspots" (Kremen *et al.*, 2008; Ko *et al.*, 2009).

A second important conservation application of distribution models is in the selection and delineation of reserve areas (Leach *et al.*, 2013; El-Gabbas *et al.*, 2016; Kaky and Gilbert, 2019). The selection of suitable reserve areas is not an easy task because the potential reserve areas must provide adequate biological conditions to allow the persistence of a variety of taxa (Margules and Pressey, 2000). So far, species distribution models along with reserve-selection algorithms (e.g., Zonation software; Moilanen *et al.*, 2014) show reliable

results in identifying and locating potential reserve areas (Leach *et al.*, 2013; El-Gabbas *et al.*, 2016; Kaky and Gilbert, 2019). For instance, a recent study by Lessmann *et al.* (2014) used species distribution models to predict the potential suitable locations for reserve areas in Ecuador for 809 vertebrate and plant species, finding that there is a huge gap in the current network of reserves and that it is necessary to double the current size of reserves to protect existing biodiversity adequately.

A third application of distribution modelling is in planning the translocation of species (Guisan *et al.*, 2013; Liu *et al.*, 2013). Guisan *et al.* (2013) discussed three possible uses for distribution models in the translocation of species: identifying habitats suitable for immediate translocation; identifying new suitable habitats under different climate conditions (assuming future climate change); and identifying possible effects of translocated species on native species. For example, Liu *et al.* (2013) built distribution models for 523 vertebrate species in Victoria (Australia) and then emphasized in their conclusion that their distribution model could be used in planning the translocation of species into different suitable habitats.

There is a range of other useful applications, such as in identifying areas that are suitable for reintroduction programs (Ardestani *et al.*, 2015; D'Elia *et al.*, 2015), guiding new field surveys and finding unknown species populations (Raxworthy *et al.*, 2003; Williams *et al.*, 2009; Peterman *et al.*, 2013; West *et al.*, 2016; Rhoden *et al.*, 2017; Tronstad *et al.*, 2018), and resource management, risk assessment and predicting disease outbreaks (Allen *et al.*, 2006; Blackburn *et al.*, 2007; Václavík *et al.*, 2010).

1.10.2 Invasive species

There is a growing concern about the economic, agricultural, and biological effects of invasive species across the world (Mack *et al.*, 2000; Pimentel *et al.*, 2001, 2005; Thuiller *et al.*, 2005b). The factors that may lead to invasions into new regions have been reviewed

(Mack *et al.*, 2000; Pimentel *et al.*, 2005). Invasive species may compete strongly with native species, destroy their habitats, or act as natural enemies, all of which may threaten their persistence (Mack *et al.*, 2000; Thuiller *et al.*, 2005b; Pimentel *et al.*, 2005; Ficetola *et al.*, 2007). Because modern species distribution models can predict suitable habitats outside the natural range, they can be used to predict directly the future geographical distribution of invasive species (Peterson, 2003; Thuiller *et al.*, 2005b) which eventually can help controlling them.

1.10.3 Coping with climate change

Climate change is perhaps the biggest threat to global biodiversity, causing dramatic changes in habitats through loss and fragmentation (Thuiller, 2004; Sinclair *et al.*, 2010; Kaky and Gilbert, 2017). More specifically, the responses of species to this modification are different depending on their biological and ecological characteristics; climate change can reshape the distribution of both animals and plants (Berry *et al.*, 2002; Kaky and Gilbert, 2017). The use of species distribution models to study the impacts of future climate conditions allows us to identify species at risk, and environmental variables of particular importance (Kaky and Gilbert, 2017). For instance, Thuiller *et al.* (2005a) studied the effect of climate change for 1,350 plant species in Europe and concluded that by 2080, most of these plants may become threatened or vulnerable because of the climate change.

In summary, species are vulnerable to disturbances in their local habitats caused by a range of anthropogenic activities, and it is a priority for ecologists to understand this vulnerability. Well-designed species distribution models based on robust predictor variables can provide a clear ecological insight into the range of environmental conditions that species can tolerate, their habitat preferences, their response to disturbance, and play an important

role in conservation planning (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005; Pearson, 2008; Elith and Leathwick, 2009; Franklin, 2009).

1.11 Testing species distribution models

Testing the accuracy of species distribution models is a critical part of the modelling process (Pearce and Ferrier, 2000). Testing the accuracy of model has been known by other terms such as evaluation, validation and predictive performance (Pearson, 2008). Many statistical methods have been developed to test the accuracy of the model without the need for independent data such as resubstitution, cross-validation, jackknifing, and bootstrapping (Verbyla and Litvaitis, 1989; Pearce and Ferrier, 2000; Araújo *et al.*, 2005; Pearson, 2008). These methods often involve splitting the original data into two parts: calibration data, also known as training data, and test data, also known as evaluated data (Verbyla and Litvaitis, 1989; Fielding and Bell, 1997; Pearce and Ferrier, 2000; Araújo *et al.*, 2005; Pearson, 2008). However, using data that were collected independently from the calibration data provides the most robust test of model performance, avoiding some of the drawbacks of methods involved in the splitting of data (Fielding and Bell, 1997; Pearce and Ferrier, 2000; Araújo *et al.*, 2005; Newbold *et al.*, 2010).

If a model's outcome is not binary (which is usually the case), a threshold probability of occurrence or habitat suitability can be used to convert model outputs from continuous into binary presence-absence data (Pearson, 2008; Liu *et al.*, 2005). Another use for the threshold is to test the accuracy of the model based on indices obtained from a confusion matrix (Liu *et al.*, 2005). Choosing an appropriate value of the threshold is hard and sensitive to the prevalence of species (the frequency of species occurrence in occupied sites), especially with presence-only data (Elith *et al.*, 2011). There are both subjective and objective methods for threshold selection (Liu *et al.*, 2005). For example, the threshold (subjective) of occurrence is

sometimes set arbitrarily at 0.5, and sometimes at 0.3, and 0.05 for species presence (Liu *et al.*, 2005). In the objective approach, the threshold is used to increase the agreement between predicted and observed distribution (Liu *et al.*, 2005). Choosing a value for the threshold is an important final step during the model-building process (Liu *et al.*, 2005).

The confusion matrix can be used on the assumption that predictions in the form of continuous habitat suitabilities can be converted to binary presence/absence data using a threshold above which a species is predicted to be present (Liu *et al.*, 2005). There are many test statistics for testing the accuracy of the model that are derived from the confusion matrix (Tables 1.1 & 1.2; Fielding and Bell, 1997; Pearce and Ferrier, 2000; Liu *et al.*, 2005; Pearson, 2008). The confusion matrix is mainly used to calculate (a) "sensitivity" (the model's ability to predict presence correctly), (b) "commission" error rate (the proportion of sites in which a species is predicted to be present but is actually recorded as absent), (c) "omission" error rate (the proportion of sites in which a species is predicted to be absent but is actually recorded as present), and (d) "specificity" (the model's ability to predict absence correctly). Additionally, the confusion matrix is used to calculate the correct classification rate (which uses both sensitivity and specificity) (Fielding and Bell, 1997; Pearce and Ferrier, 2000; Allouche *et al.*, 2006; Pearson, 2008).

If a distribution model is sensitive to the methods that have been used to convert the prediction from continuous into binary, then it is recommended to use an alternative statistical method known to be threshold-independent, such as the "Area Under the Curve" (AUC). The AUC is derived from the "Receiver Operating Curve" (ROC) (Fielding and Bell, 1997; Pearce and Ferrier, 2000; Pearson, 2008; Peterson *et al.*, 2011; Hanberry and He, 2013). The ROC is a two-dimensional graph that helps to visualize and organize classifiers according to their performance (Fawcett, 2006). In other words, the ROC is a graphical representation all the possible values of the thresholds (Pearson, 2008). This measure plots

sensitivity against (1-specificity) (Fielding and Bell, 1997); it uses all the components of confusion matrix (Pearson, 2008). The main purpose of the ROC is to describe an association between sensitivity and specificity, and it provides a measure of accuracy independent of prevalence (Pearce and Ferrier, 2000; Allouche *et al.*, 2006). An AUC of 0.5 means that the model is no better than random at predicting presence, and a value of 1 means a perfect model in predicting presence (Pearce and Ferrier, 2000; Pearson, 2008). The AUC is interpreted as the probability that a randomly chosen species presence site will be more highly ranked than a randomly chosen absence (Pearce and Ferrier, 2000; Pearson, 2008).

Choosing one of the evaluation methods depends on the purpose of the study (determining the potential or actual distribution) and the available data (presence/absence or presence only) (Pearson, 2008). Unfortunately, most occurrence data are presence-only data which may not produce as robust a model as presence-absence occurrence data (Zaniewski *et al.*, 2002; Brotons *et al.*, 2004); however, many software methods have been developed to allow presence-only data to be used to calibrate distribution models (such as Maxent; Phillips *et al.*, 2006; Baldwin, 2009). Chapter 4 reviews methods of testing species distribution models in detail.

Table 1.1: The four possible outcomes from confusion matrix.

	Presence is recorded	Absence is recorded
Presence is predicted	True positive (a) (sensitivity)	False positive (b) (commission)
Absence is predicted	False negative (c) (omission)	True negative (d) (specificity)

Table 1.2: Some evaluation statistics derived from the confusion matrix. The formula abbreviations are explained in the main text and Table 1.1.

Measure	**Formula**
Accuracy	(a + d) / (a + b + c + d)
Specificity	d/b+d
Sensitivity	a/a+c
False positive rate (commission error rate)	b/(d+b)
False negative rate (omission error rate)	c/(a+c)
Kappa	$\frac{\frac{[(a+d)-((a+c)(a+b)(b+d)(c+d))]}{n}}{\frac{[n-((a+c)(a+b)+(b+d)(c+d))]}{n}}$
True skill statistic (TSS)	Sensitivity + Specificity-1

1.12 Terrestrial reptiles in Saudi Arabia

In comparison with other regions, especially temperate, polar and wet tropical areas, the reptile fauna of Saudi Arabia forms a diverse and functionally important part of the terrestrial ecosystem. Reptiles are the second-largest vertebrate group in the country, after the birds (432 spp), with a list of 103 species having been recorded, 25% more than the mammals (79 spp) (AbuZinada *et al.*, 2004). Saudi Arabia occupies most of the Arabian Peninsula, encompassing a wide range of topographic and climatic conditions (AbuZinada *et al.*, 2004; Vincent, 2008; Gosling *et al.*, 2011), and as a result, spatial variation in reptile species richness and diversity in Saudi Arabia is expected to be considerable. Unfortunately, contemporary knowledge about reptile distributions and their overall biogeography in Saudi Arabia is poor, and much work is required to understand fully their distribution in space and time (AbuZinada *et al.*, 2004). So far, published data and museum records for Saudi reptiles are scarce, and the few surveys conducted thus far have typically been biased towards easily accessible areas, near major cities, and have focussed on certain taxa of interest to particular researchers. Such bias almost certainly means that the community is under-sampled, and species new to science may have been entirely missed. Hence, current knowledge about Saudi terrestrial reptiles is likely to be incomplete taxonomically and biogeographically. The conservation status of some species has been classified as Vulnerable, Near Threatened, and Data Deficient, with major threats caused by habitat loss and hunting (Cox *et al.*, 2012). From a conservation perspective, this sounds worrying as most of these habitats are fragile, and any potential damage can be irreversible with the majority of ongoing conservation actions established to protect other taxa (e.g., mammals) whose conservation challenges can be quite different. Therefore, due to their abundance, diversity, adaptation to the local environment and wide distribution, modelling and mapping terrestrial reptile distributions is important in understanding the challenges of conservation in Saudi Arabia. This involves

modelling their habitat suitability and potential conservation prioritization areas across Saudi Arabia, along with investigating the potential effects of environmental factors.

Thus far, there have only been two attempts (to my knowledge) to estimate the distribution of reptile diversity in Saudi Arabia. Cox *et al.* (2012) performed a regional assessment for IUCN Red-listing of the majority of the reptiles of the Arabian Peninsula by overlaying simple range-distribution polygons for each species. The authors in this assessment did not use computational methods to model reptile distributions, and they did not evaluate the importance of possible environmental predictors. A detailed species distribution model was developed by Wilms *et al.* (2011) for just one species in Saudi Arabia - *Uromastyx aegyptia microlepis* - as part of their studies on its thermal biology and activity patterns.

A study similar to mine was conducted in Egypt by El-Gabbas *et al.* (2016), modelling Egyptian reptiles. Egypt is a well-studied region in terms of its herpetology in comparison with Saudi Arabia. My study added important new information to a larger and partially unknown area, and its results are of interest to anyone studying this arid region. I also used new techniques of transference of spatial predictions coupled with ground-truthing validation and occupancy analysis, and incorporated extra layers of information when prioritizing the Saudi Arabian landscape, aiming to increase the robustness of the study outcomes. These will be of more general interest to conservation scientists.

1.13 Thesis outline

The main goal of this thesis is to understand habitat suitability for terrestrial reptile species across Saudi Arabia, and the spatial distribution pattern of reptile species richness. There is clear paucity of data, and limited information on the ecology and biogeography of terrestrial reptiles across the country. Thus understanding their current distribution, assessing ongoing/proposed conservation actions, and identifying sites for potential new conservation areas based on reliable distribution models and reserve-selection software are important.

The following outline gives more detail of each chapter along with its main questions/objectives.

- Chapter 2 applies species distribution models to predict habitat suitability for reptiles in Saudi Arabia, and the current spatial distribution of species richness. Data describing species occurrences for 62 terrestrial reptile species were specifically collated for this project, using museum collections and the relevant literature. Maxent software was then used to produce an average habitat suitability map for each species. All these maps were summed together using ArcGIS to produce a final map of potential species richness. This chapter analyses which environmental variables contribute to the models. This chapter has been published in the *Journal of Arid Environment* (Alatawi, A. S., Gilbert, F. & Reader, T. 2020. Modelling terrestrial reptile species richness, distributions and habitat suitability in Saudi Arabia. *Journal of Arid Environments*, 178**,** 104153. https://doi.org/10.1016/j.jaridenv.2020.104153).

- Chapter 3 uses the transference technique to build reptile distribution models for Egypt and then spatially transfer them into the conditions of Saudi Arabia. The data

for 71 well-studied Egyptian terrestrial reptile species were used in this chapter. Maxent produced an average habitat suitability map for each transferred species, and then all these maps were summed together using ArcGIS to produce a final transferred species richness for Saudi Arabia. Two indices were used to test the transferability between the two regions for the in-common species. This chapter discusses possible factors that affect the transferability of SDMs between the two separated areas.

- Chapter 4 tests the predictions of the species distribution models using newly collected data from the field. The fieldwork in this chapter was guided by the spatial predictions of the local Saudi model (Chapter 2) and the transferred Egyptian model (Chapter 3). The presence/absence data collected from the field were also used to test the accuracy of both models by calculating confusion matrix statistics. Detectability and occupancy were estimated for each detected species. The predicted and observed species richness from each model were also investigated.

- Chapter 5 assesses the effectiveness and the conservation value of current and proposed Saudi Protected Areas (PAs) that are managed by the Saudi Wildlife Authority, using Zonation software. Different layers were optimized in Zonation to identify top-ranking areas including the impact of socioeconomic data, uncertainty analysis, and the forced inclusion of PAs into the top fraction: all of these manipulations are aimed at robust conservation planning and assessment. This chapter also tests if the current and proposed Saudi PAs have greater conservation value than the areas outside them.

- Finally, Chapter 6 briefly brings together all the main findings of the thesis in an organized thread, in order to gain a general meaningful overview of Saudi reptile ecology before and after the work described in this thesis. It discusses the richness patterns and the conservation value results, along with a short discussion of the benefits of using species distribution models for conservation and as an exploration tool in this case study from Saudi Arabia.

1.14 References

AbuZinada, A., Robinson, E. R., Nader, I. & Al Wetaid, Y. 2004. *First Saudi Arabian national report on the Convention on Biological Diversity*. The National Commission for Wildlife Conservation and Development, Riyadh.

Allen, C. R., Johnson, A. R. & Parris, L. 2006. A framework for spatial risk assessments: potential impacts of nonindigenous invasive species on native species. *Ecology and Society*, 11**,** 39.

Allouche, O., Tsoar, A. & Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43**,** 1223-1232.

Araújo, M. B., Pearson, R. G., Thuiller, W. & Erhard, M. 2005. Validation of species–climate impact models under climate change. *Global Change Biology*, 11**,** 1504-1513.

Ardestani, E. G., Tarkesh, M., Bassiri, M. & Vahabi, M. R. 2015. Potential habitat modeling for reintroduction of three native plant species in central Iran. *Journal of Arid Land*, 7**,** 381-390.

Arntzen, J. W. 2006. From descriptive to predictive distribution models: a working example with Iberian amphibians and reptiles. *Frontiers in Zoology*, 3**,** 1-11.

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

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157**,** 101-118.

Baldwin, R. 2009. Use of maximum entropy modeling in wildlife research. *Entropy*, 11**,** 854-866.

Berry, P. M., Dawson, T. P., Harrison, P. A. & Pearson, R. G. 2002. Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. *Global Ecology & Biogeography*, 11**,** 453-462.

Blackburn, J. K., McNyset, K. M., Curtis, A. & Hugh-Jones, M. E. 2007. Modeling the geographic distribution of *Bacillus anthracis*, the causative agent of anthrax disease, for the contiguous United States using predictive ecologic niche modeling. *The American Journal of Tropical Medicine and Hygiene*, 77**,** 1103-1110.

Brotons, L. 2014. Species distribution models and impact factor growth in environmental journals: methodological fashion or the attraction of global change science. *PLoS ONE*, 9**,** e111996.

Brotons, L., Thuiller, W., Araújo, M. B. & Hirzel, A. H. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27**,** 437-448.

Buckley, L. B., Urban, M. C., Angilletta, M. J., Crozier, L. G., Rissler, L. J. & Sears, M. W. 2010. Can mechanism inform species' distribution models? *Ecology Letters*, 13**,** 1041-1054.

Cayuela, L., Golicher, D., Newton, A., Kolb, M., de Alburquerque, F. S., Arets, E., Alkemade, J. & Pérez, A. 2009. Species distribution modeling in the tropics: problems, potentialities, and the role of biological data for effective species conservation. *Tropical Conservation Science*, 2**,** 319-352.

Chapin, F. S., III, Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., Hooper, D. U., Lavorel, S., Sala, O. E., Hobbie, S. E., Mack, M. C. & Díaz, S. 2000. Consequences of changing biodiversity. *Nature*, 405**,** 234-242.

Colwell, R. K. & Rangel, T. F. 2009. Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences*, 106**,** 19651-19658.

Cox, N. A., Mallon, D., Bowles, P., Els, J. & Tognelli, M. F. 2012. *The Conservation Status and Distribution of Reptiles of the Arabian Peninsula*. Cambridge, UK and Gland, Switzerland: IUCN, and Sharjah, UAE: Environment and Protected Areas Authority.

D'Elia, J., Haig, S. M., Johnson, M., Marcot, B. G. & Young, R. 2015. Activity-specific ecological niche models for planning reintroductions of California condors (*Gymnogyps californianus)*. *Biological Conservation*, 184**,** 90-99.

Duque-Lazo, J., van Gils, H., Groen, T. A. & Navarro-Cerrillo, R. M. 2016. Transferability of species distribution models: The case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. *Ecological Modelling*, 320**,** 62-70.

El-Gabbas, A., Baha El Din, S., Zalat, S. & Gilbert, F. 2016. Conserving Egypt's reptiles under climate change. *Journal of Arid Environments*, 127**,** 211-221.

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.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. & Yates, C. J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17**,** 43-57.

Elton, C. S. 1927. *Animal Ecology*. Sidgwick & Jackson, London.

Fawcett, T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters*, 27**,** 861-874.

Ferraz, K., Ferraz, S., Paula, R., Beisiegel, B. & Breitenmoser, C. 2012. Species distribution modeling for conservation purposes. *Natureza & Conservação*, 10**,** 214-220.

Ferrier, S., 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Systematic Biology*, 51**,** 331-363.

Ficetola, G. F., Thuiller, W. & Miaud, C. 2007. Prediction and validation of the potential global distribution of a problematic alien invasive species-the American bullfrog. *Diversity and Distributions*, 13**,** 476-485.

Fielding, A. H. & Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24**,** 38-49.

Foody, G. M. 2008. GIS: biodiversity applications. *Progress in Physical Geography*, 32**,** 223-235.

Franklin, J. 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press. Cambridge, UK.

Gause, G. F. 1934. *The Struggle for Existence.* Williams & Wilkins, Baltimore.

Global Biodiversity Information Facility. Accessed at http://www.gbif.org, 17 November 2016.

Gosling, S. N., Dunn, R., Carrol, F., Christidis, N., Fullwood, J., Gusmao, D., Golding, N., Good, L., Hall, T., Kendon, L., Kennedy, J., et al. 2011. *Climate: Observations, Projections and Impacts: Saudi Arabia*. The Met Office, UK.

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 & Evolution*, 19**,** 497-503.

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

Grinnell, J. 1924. Geography and evolution. *Ecology*, 5**,** 225-229.

Guisan, A. & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8**,** 993-1009.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., Regan, T. J., Brotons, L., et al. 2013. Predicting species distributions for conservation decisions. *Ecology Letters*, 16**,** 1424-1435.

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

Hanberry, B. & He, H. 2013. Prevalence, statistical thresholds, and accuracy assessment for species distribution models. *Web Ecology*, 13**,** 13-19.

Hardin, G. 1960. The competitive exclusion principle. *Science*, 131, 1292-1297.

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.

Hirzel, A. H. & Lay, G. 2008. Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, 45**,** 1372-1381.

Hutchinson, G. E. 1957. Concluding Remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, 22**,** 415-427.

Kaky, E. & Gilbert, F. 2016. Using species distribution models to assess the importance of Egypt's protected areas for the conservation of medicinal plants. *Journal of Arid Environments*, 135**,** 140-146.

Kaky, E. & Gilbert, F. 2017. Predicting the distributions of Egypt's medicinal plants and their potential shifts under future climate change. *PLoS ONE*, 12**,** e0187714.

Kaky, E. & Gilbert, F. 2019. Allowing for human socioeconomic impacts in the conservation of plants under climate change. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology*, DOI: 10.1080/11263504.2019.1610109.

Kearney, M. 2006. Habitat, environment and niche: what are we modelling? *Oikos*, 115**,** 186-191.

Kearney, M. & Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, 12**,** 334-350.

Ko, C., Lin, R, Ding, T., Hsieh, C. & Lee, P. 2009. Identifying biodiversity hotspots by predictive models: a case study using Taiwan's endemic bird species. *Zoological Studies*, 48**,** 418-431.

Kremen, C., Cameron, A., Moilanen, A., Phillips, S., Thomas, C., et al. 2008. Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Science*, 320**,** 222-226.

Leach, K., Zalat, S. & Gilbert, F. 2013. Egypt's Protected Area network under future climate change. *Biological Conservation*, 159**,** 490-500.

Leibold, M. A. 1995. The niche concept revisited: mechanistic models and community context. *Ecology*, 76**,** 1371-1382.

Lessmann, J., Munoz, J. & Bonaccorso, E. 2014. Maximizing species conservation in continental Ecuador: a case of systematic conservation planning for biodiverse regions. *Ecology and Evolution*, 4**,** 2410-2422.

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

Liu, C., White, M., Newell, G. & Griffioen, P. 2013. Species distribution modelling for conservation planning in Victoria of Australia. *Ecological Modelling*, 249**,** 68-74.

Lobo, J. M., Jimenez-Valverde A. & Hortal. J. 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography*, 33**,**103-114.

Mack, R. N., Simberloff, D., Lonsdale, W., Evans, H., Clout, M. & Bazzaz, F. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, 10**,** 689-710.

Margules, C. R. & Pressey, R. L. 2000. Systematic conservation planning. *Nature*, 405**,** 243-253.

Marshall, C. E., Glegg, G. A. & Howell, K. L. 2014. Species distribution modelling to support marine conservation planning: the next steps. *Marine Policy*, 45**,** 330-332.

McInerny, G. J. & Etienne, R. S. 2012. Ditch the niche–is the niche a useful concept in ecology or species distribution modelling? *Journal of Biogeography*, 39**,** 2096-2102.

Miller, J. A. & Holloway, P. 2015. Incorporating movement in species distribution models. *Progress in Physical Geography*, 39**,** 837-849.

Moilanen, A., Pouzols, F. M., Meller, L., Veach, V., Arponen, A., Leppanen, J. & Kujala, H. 2014. *Zonation – Spatial Conservation Planning Methods and Software*. Version 4. User Manual. C-BIG Conservation Biology Informatics Group Department of Biosciences University of Helsinki, Finland.

Newbold, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, 34**,** 3-22.

Newbold, T., Reader, T., El-Gabbas, A., Berg, W., Shohdi, W. M., Zalat, S., Baha El Din, S. & Gilbert, F. 2010. Testing the accuracy of species distribution models using species records from a new field survey. *Oikos*, 119**,** 1326-1334.

Pardo, I., Pata, M. P., Gomez, D. & Garcia, M. B. 2013. A novel method to handle the effect of uneven sampling effort in biodiversity databases. *PLoS ONE*, 8**,** e52786.

Pearce, J. & Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133**,** 225-245.

Pearson, R. G. 2008. Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*. Available at http://ncep.amnh.org.

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

Pearson, R. G., Raxworthy, C. J., Nakamura, M. & Peterson, A. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34**,** 102-117.

Peterman, W. E., Crawford, J. A. & Kuhns, A. R. 2013. Using species distribution and occupancy modeling to guide survey efforts and assess species status. *Journal for Nature Conservation*, 21**,** 114-121.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M. & Araújo, M. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press.

Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *Q. Rev. Biol.*, 78**,** 419-433.

Peterson A. T. & Soberón J. 2012. Species distribution modeling and ecological niche modeling: Getting the concepts right. *Natureza & Conservação*, 10**,** 102-107.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190**,** 231-259.

Pimentel, D., McNair, S., Janecka, J., Wightman, J., et al. 2001. Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture, Ecosystems and Environment*, 84**,** 1-20.

Pimentel, D., Zuniga, R. & Morrison, D. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, 52**,** 273-288.

Pulliam, H. R. 1988. Sources, sinks, and population regulation. *American Naturalist*, 132**,** 652-661.

Pulliam, H. R. 2000. On the relationship between niche and distribution. *Ecology Letters*, 3**,** 349-361.

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.

Raxworthy, C. J., Martínez-Meyer, E., Horning, N., Nussbaum, R. A., Schneider, G. E., Ortega-Huerta, M. A. & Peterson, A. T. 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, 426**,** 837-841.

Rhoden, C. M., Peterman, W. E. & Taylor, C. A. 2017. Maxent-directed field surveys identify new populations of narrowly endemic habitat specialists. *PeerJ*, 5**,** e3632.

Salem, B. B. 2003. Application of GIS to biodiversity monitoring. *Journal of Arid Environments*, 54**,** 91-114.

Sillero, N. 2011. What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling*, 222**,** 1343-1346.

Sinclair, S. J., White, M. D. & Newell, G. R., 2010. How useful are species distribution models for managing biodiversity under future climates? *Ecology and Society*, 15**,** 8.

Shen, G., Yu, M., Hu, X-S., Mi, X., Ren, H., Sun, I-F. & Ma, K. 2009. Species–area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity. *Ecology*, 90**,** 3033-3041.

Shurin, J. B. 2000. Dispersal limitation, invasion resistance, and the structure of pond zooplankton communities. *Ecology*, 81**,** 3074-3086.

Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10**,** 1115-1123.

Soberón, J. & Peterson, T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, 2**,** 1-10.

Swenson, N. G. 2008. The past and future influence of geographic information systems on hybrid zone, phylogeographic and speciation research. *Journal of Evolutionary Biology*, 21**,** 421-434.

Thuiller, W. 2004. Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, 10**,** 2020-2027.

Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T. & Prentice, I. C. 2005a. Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 102**,** 8245-8250.

Thuiller, W., Richardson, D. M., Pyšek, P., Midgley, G. F., Hughes, G. O. & Rouget, M. 2005b. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology*, 11**,** 2234-2250.

Trisurat, Y., Bhumpakphan, N., Reed, D. H. & Kanchanasaka, B. 2012. Using species distribution modeling to set management priorities for mammals in northern Thailand. *Journal for Nature Conservation*, 20**,** 264-273.

Tronstad, L. M., Brown, K. M. & Andersen, M. D. 2018. Using species distribution models to guide field surveys for an apparently rare aquatic beetle. *Journal of Fish and Wildlife Management*, 9**,** 330-339.

Urban, M. C., Bocedi, G., Hendry, A. P., Mihoub, J.-B., Pe'er, G., Singer, A., Bridle, J. R., Crozier, L. G., De Meester, L., Godsoe, W. Gonzalez, A., et al. 2016. Improving the forecast for biodiversity under climate change. *Science*, 353**,** aad8466.

Václavík, T., Kanaskie, A., Hansen, E. M., Ohmann, J. L. & Meentemeyer, R. K. 2010. Predicting potential and actual distribution of sudden oak death in Oregon: prioritizing landscape contexts for early detection and eradication of disease outbreaks. *Forest Ecology and Management*, 260**,** 1026-1035.

Vandermeer, J. H. 1972. Niche theory. *Annual Review of Ecology and Systematics*, 3**,** 107-132.

Verbyla, D. L. & Litvaitis, J. A. 1989. Resampling methods for evaluating classification accuracy of wildlife habitat models. *Environmental Management*, 13**,** 783-787.

VertNet. Accessed at http://vertnet.org/ on 25, February 2020.

Vincent, P. 2008. *Saudi Arabia: An Environmental Overview*. Taylor & Francis Group, UK.

Watkinson A. R. & Sutherland W. J. 1995. Sources, sinks and pseudo-sinks. *Journal of Animal Ecology*, 64**,** 126-130.

Werkowska, W., Márquez, A. L., Real, R. & Acevedo, P. 2017. A practical overview of transferability in species distribution modeling. *Environmental Reviews*, 25**,** 127-133.

West, A. M., Kumar, S., Brown, C. S., Stohlgren, T. J. & Bromberg, J. 2016. Field validation of an invasive species Maxent model. *Ecological Informatics*, 36**,** 126-134.

Wieczorek, J., Guo, Q. & Hijmans, R., 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18**,** 745-767.

Williams, J. N., Seo, C., Thorne, J., Nelson, J. K., Erwin, S., O'Brien, J. M. & Schwartz, M. W. 2009. Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, 15**,** 565-576.

Wilms, T. M., Wagner, P., Shobrak, M., Rödder, D. & Böhme, W. 2011. Living on the edge? - On the thermobiology and activity pattern of the large herbivorous desert lizard

Uromastyx aegyptia microlepis Blanford, 1875 at Mahazat as-Sayd Protected Area, Saudi Arabia. *Journal of Arid Environments*, 75**,** 636-647.

Wintle, B., Elith, J. & Potts, J. 2005. Fauna habitat modelling and mapping: a review and case study in the lower Hunter Central Coast region of NSW. *Austral Ecology*, 30**,** 719-738.

Yang, X-Q., Kushwaha, S. P., Saran, S., Xu, J. & Roy, P. S. 2013. Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills. *Ecological Engineering*, 51**,** 83-87.

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., et al. 2018. Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33**,** 790-802.

Yi, Y-J., Cheng, X., Yang, Z-F. & Zhang, S-H. 2016. Maxent modeling for predicting the potential distribution of endangered medicinal plant (*H. riparia* Lour) in Yunnan, China. *Ecological Engineering*, 92**,** 260-269.

Zaniewski, A. E., Lehmann, A. & Overton, J. M. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, 157**,** 261-280.

2 Chapter Two: Modelling terrestrial reptile species richness, distributions and habitat suitability in Saudi Arabia

Abstract

Species distribution modelling is a powerful tool that can gives us ecological insights about species distributions, and potential effects of environmental factors, in poorly known habitats. For the first time the distribution of terrestrial reptiles in Saudi Arabia was modelled, and environmental factors that affect their current distribution and richness investigated. Reptiles are a major vertebrate group in Saudi Arabia and protecting them should be a priority for conservation in such an arid environment. Temperature was the most important of eleven predictors. Maximum species richness of reptiles was predicted in the central plateau, north-western borders, and in coastal areas of Saudi Arabia. Overall, the predicted and the observed patterns of species richness followed a similar pattern. My analysis revealed that large scattered parts of Saudi Arabia are considered under-sampled in terms of sampling efforts of terrestrial reptile species. My results represent the most comprehensive description of terrestrial reptile diversity distributions and habitat suitability in Saudi Arabia to date.

Keywords

Habitat suitability, Maxent, Saudi Arabia, species distribution modelling, species richness, terrestrial reptiles

2.1 Introduction

How and why species are distributed in space and time is considered a fundamental scientific question, especially in ecology (Rushton *et al.*, 2004; Guisan and Thuiller, 2005; Brotons, 2014). Humans have observed the unique relationship that connects living organisms with their surrounding environment (Guisan and Thuiller, 2005; Elith and

Leathwick, 2009), and for more than a century scientists have tried to come up with theoretical or experimental explanations for spatial and temporal patterns of presence/absence and abundance (Elton, 1927; MacArthur and Wilson, 1967). Today, many important scientific applications depend on understanding this relationship, such as natural resource management, predicting the impacts of climate change and invasive species, and conservation planning more generally (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005; Peterson *et al.*, 2011). However, despite considerable progress in methods for understanding spatial and temporal patterns, our knowledge of contemporary levels of biodiversity and the distribution of species richness in many parts of the world remains poor (Mora *et al.*, 2011; Ficetola *et al.*, 2013).

Species richness is a key measure of biodiversity and its study can illuminate important ecological phenomena, such as the species-area relationship, and rules for community assembly (Brown, 2007). Species richness is connected with the quality and quantity of available habitat, and habitat degradation can adversely affect species richness (Yi *et al.*, 2016). The distribution of species richness is influenced by many biotic and abiotic factors; understanding and measuring these factors and how they influence the distribution of species is critical for proper conservation planning (Brown, 2007). However, it is not an easy task to model the spatial distribution of the richness of species, especially across large-scale areas (Pineda and Lobo, 2009). The gap in current knowledge about the distribution of species richness and the suitability of habitats could be fixed with enough biological data obtained from surveys, and accurate evidence of spatial patterns in environmental variables; these can be combined to produce predictive maps of species richness (Lobo *et al.*, 2002; Graham and Hijmans, 2006). These descriptive maps of species richness are considered a basic tool in conservation decision-making (Lobo *et al.*, 2002; Graham and Hijmans, 2006).

Techniques for modelling species distributions have developed rapidly in the last two decades (Guisan and Zimmerman, 2000; Guisan and Thuiller, 2005; Elith and Leathwick, 2009). Correlative species distribution models (SDMs) (also called ecological niche modelling, more details in Sillero, 2011) usually work by combining known species occurrence records with spatial layers of digital environmental data, and using statistical techniques to estimate geographical distribution and habitat suitability for species over an area of interest (Guisan and Zimmerman, 2000; Guisan and Thuiller, 2005; Sillero, 2011). The main objective is to generate new predicted distribution maps that capture and explain the species-environment association (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005) which, if presence-only models are used, effectively represents habitat suitability for the target species (see Sillero, 2011). These maps show how habitat suitability varies through space and time, by identifying locations in the study area that have environmental conditions similar to those where the species has been found (Pearson, 2008; Sillero, 2011). However, despite considerable progress in the development of these techniques, predicting species distributions is still a complex and challenging process (Merow *et al.*, 2014).

Many advanced modelling techniques/tools and quantitative methods have been developed and applied in distribution modelling, including various forms of regression analysis (generalized linear models [GLMs], generalized additive models [GAMs], and machine-learning methods Maximum Entropy [Maxent], and artificial neural networks [ANN]) (Peterson *et al.*, 2011). All of these methods can be used to model potential species richness, but each uses a different process to generate the final distribution map. Some modelling techniques require presence-absence data (e.g., GLM) while others require only presence data as they can generate background points during modelling (e.g., Maxent) (Phillips *et al.*, 2006). Choosing which method to apply depends on the available data and the purpose of the study (Elith and Leathwick, 2009).

Reptiles are widely distributed around the world and are considered important components of local ecosystems and biodiversity (AbuZinada *et al.*, 2004; Carranza *et al.*, 2018). Due to their ectothermy, reptiles are sensitive to the thermal characteristics of their environment (Wilms *et al.*, 2011; Carranza *et al.*, 2018), and they demonstrate various special adaptations as a result (Metallinou *et al.*, 2015). Because their response to temperature variation is relatively well understood, they can be used as indicators to assess the impact of global warming on the environment (Wilms *et al.*, 2011). Today, many reptiles are seriously threatened by many factors, such as habitat loss and conversion, invasive species and collection for the pet trade (Cox *et al.*, 2012), all of which can negatively affect their spatial distribution.

The spatial distribution of reptiles is affected by various climate and topographical factors. Different studies have reported a variety of factors contributing to the explanation of reptile distributions, including precipitation (Fattahi *et al.*, 2014; Sanchooli, 2017), temperature (Sillero and Carretero, 2013, Javed *et al.*, 2017), altitude (El-Gabbas *et al.*, 2016) and vegetation cover (Fattahi *et al.*, 2014). However, among all these factors, temperature appears to dominate (directly and/or indirectly), which is not surprising since it is well known in affecting daily activities and reptile biology (Huey, 1982; Wilms *et al.*, 2011). The extinction risk of most reptiles has not been evaluated properly yet (Carranza *et al.*, 2018).

Reptiles are a major group of vertebrates in Saudi Arabia, with a list of 103 species having been recorded (AbuZinada *et al.*, 2004). Reptiles are the second most diverse terrestrial vertebrate group in Saudi Arabia after birds (432 spp.), with 25% more species than the mammals (79 spp.) (AbuZinada *et al.*, 2004). General knowledge about reptile ecology in Saudi Arabia is poor, and much work is required to understand fully their distribution in space and time (AbuZinada *et al.*, 2004). Saudi Arabia occupies most of the Arabian Peninsula, encompassing a wide range of topographic and climatic conditions (AbuZinada *et*

al., 2004; Vincent, 2008; Gosling *et al.*, 2011). The richness and occurrence of reptiles in Saudi Arabia are expected to vary from one location to another in response to this variation. Unfortunately, published data and museum records for Saudi reptiles are scarce. The few surveys of Saudi reptiles conducted thus far have typically been biased towards easily accessible areas, near major cities, and have focussed on certain taxa of interest to particular researchers. Such bias may mean that the community is under-sampled, and species new to the science may have been entirely missed. Therefore, mapping and modelling of reptile distributions, and studies of the factors that may affect them, are urgently required before relevant conservation measures can be enacted.

Thus far, there have only been two attempts (to my knowledge) to estimate the distribution of reptile diversity in Saudi Arabia. Cox *et al.* (2012) performed a regional assessment for IUCN Red-listing of the majority of the reptiles of the Arabian Peninsula. By overlaying simple range-distribution polygons for each species, they predicted that high species-richness locations in Saudi Arabia would include the long chain of south-western mountains, the coast of the Eastern Province and a few locations in northern areas. Locations expected to have the lowest species richness were in the Empty Quarter (the Rub' al Khali), and central southern areas. A more detailed species distribution model was developed by Wilms *et al.* (2011) for just one species in Saudi Arabia - *Uromastyx aegyptia microlepis* - as part of their studies on its thermal biology and activity patterns.

My objective in this paper is to model the distribution of reptiles in general, using presence-only data, and map their potential richness, and their habitat suitability in Saudi Arabia using Maxent. Maxent is capable of capturing complex non-linear response curves using a set of "features" (e.g. linear, quadratic, product, threshold; see Merow *et al.*, 2013; Phillips *et al.* 2017). The distribution predictions made by Maxent are known to perform well, its regularization procedure obviating over-fitting (Merow *et al.*, 2013). I built models

of the spatial distribution of individual reptile species as a function of environmental variables in order to understand the magnitude of the impact of these variables on the species concerned. I then combined these models to produce two maps to describe predicted patterns of diversity across Saudi Arabia. One map shows the summed probability of species occurrence, and the other shows the expected species richness assuming a threshold rule to convert continuous habitat suitability values into a binary prediction of suitable/not suitable. The predictions that I produce represents the most comprehensive attempt to understand the current distribution of terrestrial reptiles in Saudi Arabia to date.

2.2 Methods & Materials

2.2.1 Study area and presence records

Saudi Arabia is a large [2,149,690 km^2] hyper-arid country (Figure 2.1) (Vincent, 2008) with a list of 103 reptiles (at least 85 of them being terrestrial lizards and snakes) that have been recorded, comprising 60 lizards, 34 snakes, and 9 turtles (AbuZinada *et al.*, 2004). There are two main resources for the records used in this thesis. First, I obtained records of occurrences of all terrestrial reptiles (lizards and snakes only) from museums (VertNet website; accessed at http://vertnet.org/ February 07, 2017). Second, I read carefully a long list of relevant literature to extract records of terrestrial reptiles in Saudi Arabia. Extensive georeferencing of locations was carried out to ensure precision using all possible online resources to locate the targeted locations. For this, I used mainly Google Earth, ArcMap, online GeoNames websites, and Gazetteer to locate the sites. Locations with vague descriptions were deleted. A total of 3,943 point records of 62 reptile species (41 lizards and 21 snakes) were collated for use in distribution modelling. The currently accepted names of the reptiles were checked using The Reptiles Database (www.reptile-database.org), and Catalogue of Life (ITIS; www.catalogueoflife.org).

However, as is expected with museum data and non-systematic surveys (Newbold, 2010), the data show an unavoidable bias toward cities, road networks, and easily accessible areas. To minimise the impact of bias in my model predictions, I employed a target-group bias file (Figure S2.8) (Phillips *et al.*, 2009). This uses the sum of all the records of the taxon of interest as an estimate of sampling effort, smoothed using a Gaussian kernel estimation function (using the SDMtoolbox 2.0 in ArcGIS: Brown *et al.*, 2017). The bias file is used in Maxent to promote the selection of more background points from biased locations. It is a common approach to correct sampling bias used in distribution modelling and is known to improve model accuracy (Phillips *et al.*, 2009).

2.2.2 Environmental variables

The environmental variables used in this study were downloaded from WorldClim Version 2.0 (19 bio-layers and 12 indices of solar radiation), and Version 1.4 (altitude) (www.worldclim.org) (Hijmans *et al.*, 2005, Fick and Hijmans, 2017). Because of the massive extent of the study area and relatively low number of records, environmental variables were downloaded at 30-sec resolution (~1 km^2). The 30 arc-sec resolution was chosen because i) reptiles have relatively poor dispersal ability, ii) the georeferencing is good enough to match the selected resolution; and iii) changing to a larger size of pixel does not greatly influence the results (Guisan *et al.*, 2007). Stepwise selection of environmental variables was performed by calculating the Variance Inflation Factor (VIF) using the library *usdm* in R (Naimi *et al.*, 2014; R Development Core Team, 2018), in order to reduce multicollinearity among variables. Among the 32 original variables, only 11 with a VIF less than 10 were retained for modelling (Table 2.1; Figure S2.9). Vegetation cover was also ignored because it is very sparse in such an environment, and has been shown to be a poor predictor of reptile distributions in similar studies (see El Gabbas *et al.*, 2016 for Egypt). All the environmental

variables were prepared by clipping them to the size of the study area and converting them to ascii files using ArcGIS software (Version 10.3.1). I also investigated collinearity between all variables using a Pearson correlation matrix (Table S2.2).

The WorldClim data has global coverage, typically with a high resolution, providing information about the environment at different grain sizes, including coarse and fine resolution. The data are freely and easily available online, and can be manipulated easily in various software programs. Most importantly, the WorldClim data are specifically designed to produce meaningful biological variables believed to influence physiology and hence species distributions. The data are widely used in ecological modelling research, particularly species distribution modelling studies. However, WorldClim data are interpolated from existing weather-station data, and not all areas have good coverage of weather stations (cf. maps in Hijmans *et al*., 2005), including Saudi Arabia. Saudi Arabia has relatively a small number of weather stations mostly located near or within the cities, which are far from randomly located. Thus as with Egypt, where the distribution of weather stations is perhaps even more biased (El-Gabbas *et al.*, 2016), interpolation errors in the environmental predictors needs to be borne in mind when using them to model species distributions.

2.2.3 Species richness distribution modelling and evaluation

I created distribution models using Maxent Version 3.4.1, downloaded from https://biodiversityinformatics.amnh.org/open_source/maxent/ (Phillips *et al.*, 2006). Maxent identifies areas that have similar values on environmental variables to those at the locations of recorded occurrences. By default, Maxent uses 10000 background points chosen randomly from the study area. This is a reasonable number given the size of the study area and its range of environmental conditions (Phillips *et al.*, 2009). It is the most popular software to model species distribution with presence-only data (Merow *et al.*, 2013). I chose Maxent to run the

distribution modelling because: i) it is very robust, producing reliable models from presence-only data; ii) it works with low numbers of records; and iii) it appears to be relatively insensitive to sample bias (see Baldwin, 2009). Maxent has been shown to perform very well in comparisons with other modelling approaches (Baldwin, 2009; Merow *et al.*, 2013).

I chose features and settings in Maxent that maximized the area under curve (AUC), on the basis of extensive testing. I therefore used: product and hinge features; ten subsampled replicates; 1000 iterations; and the new default *cloglog* output (Phillips *et al.*, 2017). 'Product' refers to the product of pairs of continuous environmental predictors, allowing interactions between predictor variables in Maxent (Phillips *et al.*, 2006; Phillips and Dudik, 2008). 'Hinge' generalizes linear and threshold features (Merow *et al.*, 2013), modelling complex relationships in order to minimize overfitting in the training data, allowing more realistic species responses to the environment to be modelled (Phillips and Dudik, 2008; Phillips *et al.*, 2017). I used the new version of Maxent which by default uses a *cloglog* transformation to represent the habitat suitability across the study area, recommended for its more meaningful output and theoretical justification (Phillips *et al.*, 2017). The replicates were produced by partitioning the data into 90% training and 10% testing subsets; and I used the '10% training presence' as the threshold rule to convert continuous habitat suitability into dichotomous suitable/unsuitable values (Liu *et al.*, 2005). The permutation importance table was used to determine the importance of each variable and the degree it contributed to the model. I also considered other approaches to test the model, including partitioning data into 70% training and 30% testing subsets, and cross-validation with ten replications (see supplementary materials) to investigate the final patterns of habitat suitability distribution.

2.2.4 Model evaluation

To evaluate the accuracy of each model, I used the threshold-independent Area Under the Curve ("AUC") calculated from the "Receiver Operating Curve" (ROC) (Fielding and Bell, 1997; Pearce and Ferrier, 2000). The ROC curve plots sensitivity against 1-specificity for all possible values of the threshold habitat suitability above which the habitat is assumed to be suitable (Fielding and Bell, 1997). The AUC can be interpreted as the probability that a randomly chosen presence site will be more highly ranked than a randomly chosen absence one (Pearce and Ferrier, 2000; Merow *et al.*, 2013).

The ten subsampled replicate maps for each species were reduced to a single average map by taking the arithmetic mean, or by creating a consensus binary map (see below). Then I combined the species distribution predictions for each modelled species, creating two maps summarising the distribution of terrestrial reptile diversity in Saudi Arabia: i) a probability map which shows the sum of the predicted habitat suitabilities for all species in each grid cell; and ii) a binary map which shows the "predicted potential species richness" based on the conversion of the continuous surface of habitat suitability for each species into a dichotomous surface (suitable = 1 and not suitable = 0). Both maps were created in ArcGIS (10.3.1) using the Raster Calculator and the Reclassify Tool. To create the probability map, all the averaged ascii maps for each species were added together to produce a final map of potential species richness. To create a consensus binary map from the 10 replicate maps for each species, the average of ten ascii thresholds for each species were added together, and a species was considered to be present in a pixel if more than five replicates of the model predicted its presence. Then, all the species binary maps were added together to create a final consensus species-richness map.

2.2.5 Identifying areas of low sampling effort

I measured the degree of sampling in areas that probably have been under-sampled, and areas with low/moderate to high sampling effort in relation to the predicted habitat suitabilities, by calculating the difference between predicted and observed species richness at each grid square cell. To do so, 50x50-km grid cells were created across the study area and the observed species richness calculated for each grid cell. Next, each individual thresholded species map (i.e., the consensus binary map at a resolution of ~ 1km) was checked and if any suitable cell was detected. Then 50 x 50 cell was marked as positive. The species maps at the large cell size were then added together to generate a new thresholded species richness map that matches the observed richness resolution.

2.3 Results

2.3.1 Model performance and variables contribution

The distribution models provided good predictions for most reptile species, with AUCs ranging from 0.68 to 0.99 (Figure 2.2). For 22 reptile species, models had excellent performance, with mean AUC > 0.90. The number of records used to create the best performing models varied considerably in size (10-79), and there was a significant negative relationship between the number of observations and mean AUC (Pearson's correlation: n = 62, r = -0.326, P = 0.01). The species with highest mean AUC was *Scincus hemprichii* (0.998), and the species with lowest was *Stenodactylus doriae* (0.679).

The contribution of each environmental predictor to models for all 62 reptile species modelled is shown in Figure 2.3. Temperature variables were the most frequently retained as significant predictors of the spatial distribution of terrestrial Saudi reptiles. Overall, bio2 (mean diurnal temperature range) was most frequently the highest contributor to the models (20 species). The frequency of other predictors contributing to the models was moderate (see

Figure 2.3 & Table 2.1). Altitude had a low contribution to the models, while solar radiation in September never contributed significantly (Figure 2.3 & Table 2.1).

2.3.2 Current patterns of species richness

The current spatial distribution of potential species richness and habitat suitability (Figures 2.4 & 2.5) shows that large areas of Saudi Arabia are expected to have moderate to high species richness of reptiles. Overall, the two maps show a broadly similar pattern. The central plateau of Saudi Arabia has high predicted species richness. This area covers from south of Riyadh Province up to Al-Qaseem Province. There is moderate to high predicted species richness in the area between the central plateau and the Eastern Province (eastern coast and surrounding areas). There are some fragmented locations in other parts of Saudi Arabia that have high predicted species richness, including the extreme north-west, particularly in Tabuk Province. Finally, the distribution models also predict that western areas surrounding Jeddah city, the long Asir mountain chain, and the south (Jizan and Najran) have moderate species richness. The predicted pattern declines from the northern edge of the Empty Quarter to the border with Oman, and richness is predicted to be low between the central region and the western coast, apart from in a few scattered locations. The northern border of Al-Jawf Province along the Jordanian border also has low predicted diversity except in a few fragmented locations. Interestingly, the other two partitioning approaches generated similar potential species richness and habitat suitability distributional patterns (see Figures S2.10 & S2.11).

For the individual species, Maxent predicted distributions very well despite variable sample sizes. Several groups of reptile species had similar distribution patterns (see supplementary material Figure S2.12). For example, Bosk's Fringe-fingered Lizard (*Acanthodactylus boskianus*) (n = 222 records), and Arabian horned viper (*Cerastes*

gasperettii) (n = 155) were predicted to have high habitat suitability in the central plateau, along the coastlines, and in fragmented locations in the north and south. Other species such as Sandfish skink (*Scincus scincus*) (n = 62) and Small-spotted Lizard (*Mesalina guttulata*) (n = 42) had medium to high habitat suitability across most of the country. Even though in general the Empty Quarter had low habitat suitability, for some species it was predicted to be very suitable; for example, the Arabian sand skink (*Scincus mitranus*) (n = 172) and Arabian sand boa (*Eryx jayakari*) (n = 96).

2.3.3 Sampling effort across Saudi Arabia

Overall, the predicted and the observed species richness broadly followed similar patterns. The map of their differences (Figure 2.6) shows that a large portion of the extreme north-west, the west coast to south-western areas, the majority of the central areas, and the east coast and its surroundings appear to have been greatly under-sampled because the predicted species richness recorded species richness was noticeably higher than the observed value. In contrast, areas corresponding to small differences between the predicted and the observed species richness were identified as above the central areas toward the borders with Kuwait, Iraq, and Jordan, with scattered locations in the mid-west, and the whole of the south-west.

2.4 Discussion

My distribution models appear to predict the distribution of reptiles in Saudi Arabia well, since 35 % of species had excellent performance (AUC > 0.90) and only 4 % were poor (AUC < 0.70). Diurnal temperature range was the most effective environmental predictor in explaining the current distribution of reptiles. The models suggest that terrestrial reptile habitat suitability and their potential richness is highest in the eastern part of the central

plateau of Saudi Arabia, in the mountain chains of the southern border, and around the coasts, especially in north-western and eastern areas. The degree of under-sampling was identified in relatively large parts of Saudi Arabia from large differences between the observed and the predicted species richness.

Knowing which environmental variables contribute most to distribution models for reptiles in Saudi Arabia is important in such an extreme environment, where diurnal temperature ranges are large, and drought is prevalent most of the year. Ecologically, it is unsurprising that diurnal temperature range contributed most to the models, given the ectothermy of reptiles (Huey, 1982; Wilms *et al.*, 2011) and the nature of the environment (AbuZinada *et al.*, 2004; Vincent, 2008). Temperature is expected to be a limiting factor in reptile distribution because of their behavioural and physiological characteristics (Huey, 1982; Wilms *et al.*, 2011). My maps predict that reptile species richness in Saudi Arabia is expected to be high in areas with moderate diurnal temperature ranges, but lower in areas which experience very high temperatures.

Variables in Table 2.1 have different effects on species distribution. Temperature appears to be a dominant variable in controlling terrestrial reptile distribution. For example, besides its effect on thermoregulation, temperature can affect reptile sex ratio (Bickford *et al.*, 2010). Solar radiation works as a source of energy, impacting the thermal balance of reptiles directly, and indirectly through its influence on food availability. Thus temperature and solar radiation variables in Table 2.1 are believed to have both direct and indirect effects on terrestrial reptile distributions. Precipitation also influences reptile distributions mostly through its indirect effects on the food supply and provision of microhabitats, both of which affect growth rate and fitness of reptiles (Bickford *et al.*, 2010). Thus the precipitation variables listed in Table 2.1 should influence reptile distribution indirectly. Finally, elevation is another variable known to influence reptile distributions, almost certainly indirectly

through its effects on temperature and precipitation. There are some caveats when selecting environmental variables based on removing collinearity. Among the 32 variables, there might be factors that are also important to include in the modelling process but which were removed by the collinearity criteria followed here. Another suitable procedure might be either to include all bioclimatic variables and investigate the performance of each in detail, or to select them a priori based on prior knowledge of their potential effects on terrestrial reptiles in these habitat conditions.

Other studies of reptiles have found similar results to ours in different parts of the world, even where typical environmental characteristics are very different from those in Saudi Arabia. For example, Javed *et al.* (2017) found that mean diurnal temperature range (bio2) was one of best predictors of the distribution of the Indian Golden Gecko (*Calodactylodes aureus*) in India. In Eastern Asia and Taiwan, Ananjeva *et al.* (2015) found that temperature variables were amongst the most effective predictors of the distribution of Square-headed Cat Snake (*Boiga kraepelini*). In Europe, Sillero and Carretero (2013) found that temperature was an important predictor variable in their model of the distribution Carbonell's wall lizard (*Podarcis carbonelli*) on the Iberian Peninsula.

The predicted current distribution of reptiles generated by Maxent suggests that large areas of Saudi Arabia have habitats that are reasonably suitable for reptiles, even though they have not been surveyed. However, the predicted areas with highest potential species richness tend to be concentrated around locations with many known species occurrences. Sillero and Carretero (2013) reported similar results because their model predicted larger suitable areas over the Iberian Peninsula than currently known for *P. carbonelli*, but relatively concentrated around observed occurrences. In my model, predicted reptile species richness is highest in the centre and around the coast of Saudi Arabia. A similar coastal pattern was reported in North Africa by Kaliontzopoulou *et al.* (2008), and around the coastline of the Arabian Peninsula by

Cox *et al.* (2012). Along these coastal areas in Saudi Arabia there are major cities (e.g., Jeddah, Jizan, Al-Khoabr, Al-Jubil), near which there are a lot of occurrence records.

The predicted pattern of high reptile diversity and habitat suitability around the coast of Saudi Arabia is in-line with the findings of El-Gabbas *et al.* (2016), who modelled reptile species richness and habitat suitability in Egypt. The authors found that some coastal areas of Egypt had high predicted species richness, especially around the Sinai Gulf and Mediterranean Sea. My model predicted that some habitats in Saudi Arabia would be suitable for at least 50 out of 62 species, which is similar to the findings of El-Gabbas *et al.* (2016), who found that some locations in Egypt would be suitable for at least 52 out of 75 modelled species.

As well as predicting high potential species richness in coastal areas, my models predicted high potential species richness in the central areas of the Riyadh and Al-Qassem areas. The areas with high predicted potential species richness are often near centres of human population. This is a pattern seen in other studies of terrestrial reptiles in arid environments (e.g. in Oman by Carranza *et al.*, 2018), and could an artefact of sampling bias. However, I have attempted to control for sampling bias by using a bias file. Alternatively, the association between centres of human habitation and predicted reptile species richness may reflect the fact that human activity is associated with more food-rich or otherwise suitable environments, either because humans choose to live in areas which are benign or productive, or because humans influence reptile distributions positively by acting as a source of suitable habitat or resources (e.g. farmland, which can be exploited by reptiles) (Tytar *et al.*, 2015; Carranza *et al.*, 2018). A study by Tytar *et al.* (2015) modelled reptile distributions in the Western Podillya (Ukraine), and found that the human "footprint" contributed significantly (positively) as a predictor in the distribution model. Carranza *et al.* (2018) in Oman found some areas with high predicted species richness around the capital Muscat; having sampled

thoroughly, they concluded that reptiles are not badly affected by the presence of human settlements and may actually benefit. Finally, the combination of low to moderate altitude and suitable temperatures most of the year along the coasts (AbuZinada *et al.*, 2004, Vincent, 2008), which are relatively densely populated by humans, may provide good conditions for terrestrial reptiles in Saudi Arabia, including both generalist and specialist species.

The analysis of under-sampling suggests that even areas that are easily accessible and not isolated have received low sampling effort in terms of terrestrial reptile detection (Figure 2.7), which unfortunately represents large parts of the Saudi Arabian landscape (Figure 2.6). This is a very important result that can help direct future sampling effort to provide better systematic conservation planning. A study by Sillero *et al.* (2009) used a similar approach to identify potential under-sampled areas in the Iberian Peninsula (see also Sillero *et al.* 2014).

I found that altitude in Saudi Arabia was not an important predictor of reptile distributions, which contradicts the findings of a similar study in Egypt (El-Gabbas *et al.*, 2016). However, climate-related variables are generally known to influence reptiles more than topographical variables (Guisan and Hofer, 2003). According to Chettri *et al.* (2010), species richness of reptiles typically gradually declines with elevation (no species were recorded above 3000 m in their study), but the distribution of reptiles in Saudi Arabia may be restricted by factors other than climate and topography, such as interspecific competition (Chettri *et al.*, 2010). Locations at high altitude in Saudi Arabia are not well sampled, and this makes it difficult to speculate about the reason why altitude is not an important predictor in this study. This underlines the need for more ground-truthing surveys in less sampled areas.

Since most of Saudi Arabia lacks systemic surveys for reptiles, and most of the available records were gathered from opportunistic observations (e.g., museum records), some species may have been completely overlooked, including perhaps species that are new to science. My model predictions cover large areas that have not been surveyed at all (to my

knowledge), emphasising the need for more ground-truthing surveys to test predictions and reveal the true extent of Saudi reptile diversity. For example, I have found that the Empty Quarter is not very suitable for terrestrial reptiles, in line with Cox *et al.* (2012), who reported similar patterns about the Empty Quarter. However, more sampling and ground-truthing is required to establish if diversity is really that low.

I have used a target-group bias file in Maxent to minimize the effect of the usual biases in museum data (Phillips *et al.*, 2009). Biased observations are known to have an effect on the predictive quality of the model (Phillips *et al.*, 2009; Fourcade *et al.*, 2014). Corrections for any bias depend on the type and intensity of bias (Fourcade *et al.*, 2014) in addition to species characteristics (Hernandez *et al.*, 2006), and the best methods are still being actively researched (Phillips *et al.*, 2009; Fourcade *et al.*, 2014).

The resources that I used to collect information about Saudi reptile distributions did not contain records of species absences. Using presence-only data to calibrate distribution models has some known drawbacks (see Zaniewski *et al.* 2002; Brotons *et al.* 2004) which may limit model performance (Brotons *et al.* 2004). Importantly, presence-only methods probably over-estimate species occurrence, because locations predicted to be suitable may not in fact be occupied, as a result of limited species dispersal. As a result, using presence-absence data is strongly recommended whenever available (Brotons *et al.* 2004). However, presence-only records often the only available information about species occurrences, and these are still informative about the true underlying distribution (Zaniewski *et al.* 2002). Despite using presence-only data, Maxent has been shown to perform well, generating predictive models even with biased data and small sample sizes (see Hernandez *et al.* 2006; Pearson *et al.* 2007; Wisz *et al.* 2008). In the absence of systematic surveys of the reptile fauna across the country, the Saudi data that I collected represent valuable information, and presence-only distribution modelling provides the current best option for describing and

understanding patterns of species richness. Because of the tendency for this approach to over-estimate actual species occurrence, I must treat my predictions as likely upper-estimates of actual species richness. Only extensive systematic field sampling will prove whether these estimates are correct.

It is important to note that large portions of Saudi Arabia lack surveys, and the total number of observations upon which my model predictions are based is rather small. This is well known to limit the performance of models (Hernandez *et al.*, 2006). The majority of my records were obtained from museums and the literature, and show signs of bias towards easily accessible locations and species that are easy to identify and catch. Although I have attempted to minimise the effect of sampling bias, it is unlikely that my predictions are unaffected by these issues.

Even though there is a general lack of detailed information about reptiles in Saudi Arabia (AbuZinada *et al.*, 2004), they do represent a good case study because: i) they are very diverse and represent a large portion of vertebrate diversity; ii) they include endemic, vulnerable and unusual species; and iii) they are relatively well known and easy to sample, compared with many other taxa. It is important to encourage local environmental agencies and Saudi herpetologists to fill the knowledge gaps by establishing more reliable and more accurate datasets that are updated regularly. In particular, there should be systematic surveys of the different areas of the country, including locations have not been surveyed before. I consider my modelling exercise as the first step towards a more detailed understanding of the spatial distribution of reptile diversity, and biodiversity more generally, in this understudied part of the world.

2.5 Conclusion

Because efforts to study species distributions and patterns of species richness are unevenly distributed in many parts of the world, some locations are expected to have more species richness and uniqueness than currently reported (e.g., the south-western Arabian Peninsula; Ficetola *et al.*, 2013). Reptiles are the major vertebrate taxon in the desert ecosystem in Saudi Arabia. Ensuring their conservation and long-term persistence should be a priority. In my case, even with unevenly distributed occurrence data, I was able to build informative spatial models to describe the likely current distribution of reptiles in Saudi Arabia. Distribution modelling studies such mine can provide important insights into patterns of diversity, in Saudi Arabia and elsewhere. I hope that this study will stimulate detailed future studies, including ground-truthing, to fill the gap in our knowledge about local terrestrial reptile distributions. Species distribution modelling has proven useful in many different subject areas, facilitating fieldwork and conservation planning, and helping us to understand the likely consequences of climate change and the threat of invasive species (Guisan and Thuiller, 2005; Pearson, 2008; Peterson *et al.*, 2011). Distribution modelling has potential to contribute to these goals in Saudi Arabia, and may be vital in future studies such as those testing the value of networks of protected areas or seeking to ameliorate the impact of global environmental change.

2.6 References

AbuZinada, A., Robinson, E. R., Nader, I. & Al Wetaid, Y. 2004. *First Saudi Arabian national report on the Convention on Biological Diversity*. The National Commission for Wildlife Conservation and Development, Riyadh.

Ananjeva, N. B., Golynsky, E. E., Lin, S-M., Orlov, N. L. & Tseng, H-Y. 2015. Modeling habitat suitability to predict the potential distribution of the Kelung Cat Snake *Boiga kraepelini* Steineger, 1902. *Russian Journal of Herpetology*, 22**,** 197-205.

Baldwin, R. A. 2009. Use of maximum entropy modeling in wildlife research. *Entropy*, 11**,** 854-866.

Bickford, D., Howard, S. D., Ng, D. J. J. & Sheridan, J. A. 2010. Impacts of climate change on the amphibians and reptiles of Southeast Asia. *Biodivers. Conserv.* 19**,** 1043-1062.

Brotons, L. 2014. Species distribution models and impact factor growth in environmental journals: Methodological fashion or the attraction of global change science. *PLoS ONE*, 9**,** e111996.

Brotons, L., Thuiller, W., Araujo, M. B. & Hirzel, A. H. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27**,** 437-448.

Brown, J. L., Bennett, J. R. & French, C. M. 2017. SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ*, *5***,** e4095.

Brown, R. L., Jacobs, L. A. & Peet, R. K. 2007. Species Richness: Small Scale. *eLS*, 1-8. John Wiley & Sons Ltd.

Carranza, S., Xipell, M., Tarroso, P., Gardner, A., Arnold, E. N., Robinson, M. D., et al. 2018. Diversity, distribution and conservation of the terrestrial reptiles of Oman (Sauropsida, Squamata). *PLoS ONE*, 13**,** e0190389.

Catalogue of Life. 2018. http://www.catalogueoflife.org/ (accessed on 15 March 2018).

Cox, N. A., Mallon, D., Bowles, P., Els, J. & Tognelli, M. F. 2012. *The Conservation Status and Distribution of Reptiles of the Arabian Peninsula*. Cambridge, UK and Gland, Switzerland: IUCN, and Sharjah, UAE: Environment and Protected Areas Authority.

Chettri, B., Bhupathy, S. & Acharya, B. K. 2010. Distribution pattern of reptiles along an eastern Himalayan elevation gradient, India. *Acta Oecologica*, 36**,** 16-22.

El-Gabbas, A., Baha El Din, S., Zalat, S. & Gilbert, F. 2016. Conserving Egypt's reptiles under climate change. *Journal of Arid Environments*, 127**,** 211-221.

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.

Elton, C. 1927. *Animal Ecology*. Macmillan Company, New York.

Fattahi, R., Ficetola, G. F., Rastegar-Pouyani, N., Avci, A., Kumlutaş, Y., Ilgaz, Ç. & Yousefkhani, S. H. 2014. Modelling the potential distribution of the Bridled Skink, *Trachylepis vittata* (Olivier, 1804), in the Middle East. *Zoology in the Middle East*, 60**,** 208-216.

Ficetola, G. F., Bonardi, A., Sindaco, R. & Padoa-Schioppa, E. 2013. Estimating patterns of reptile biodiversity in remote regions. *Journal of Biogeography*, 40**,** 1202-1211.

Fick, S. E. & Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37**,** 4302-4315.

Fielding, A. H. & Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24**,** 38-49.

Fourcade, Y., Engler, J. O., Rödder, D. & Secondi, J. 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS ONE*, 9**,** e97122.

Gosling, S. N., Dunn, R., Carrol, F., Christidis, N., Fullwood, J., Gusmao, D., Golding, N., Good, L., Hall, T., Kendon, L., Kennedy, J., et al. 2011. *Climate: Observations, Projections and Impacts: Saudi Arabia*. The Met Office, UK.

Graham, C. H. & Hijmans, R. J. 2006. A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, 15**,** 578-587.

Guisan, A., Graham, C. H., Elith, J., Huettmann, F. & the NCEAS Species Distribution Modelling Group. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, 13**,** 332–340.

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

Guisan, A. & Hofer, U. 2003. Predicting reptile distributions at the mesoscale: Relation to climate and topography. *Journal of Biogeography*, 30**,** 1233-1243.

Guisan, A. & Thuiller, W. 2005. Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8**,** 993-1009.

Hernandez, P. A., Graham, C. H., Master, L. L. & Albert, D. L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29**,** 773-785.

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.

Huey, R. B. 1982. Temperature, physiology, and the ecology of reptiles. In *Biology of the Reptilia*, vol. 12 (eds Gans, C., & Pough, F.H.). Academic Press, London, pp. 25-91.

Javed, S. M., Raj, M. & Kumar, S. 2017. Predicting potential habitat suitability for an endemic gecko *Calodactylodes aureus* and its conservation implications in India. *Tropical Ecology*, 58**,** 271-282.

Kaliontzopoulou, A., Brito, J. C., Carretero, M. A., Larbes, S. & Harris, D. J. 2008. Modelling the partially unknown distribution of wall lizards (*Podarcis*) in North Africa: ecological affinities, potential areas of occurrence, and methodological constraints. *Can. J. Zool*, 86**,** 992-1001.

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

Lobo, J. M., Lumaret, J. P. & Jay-Robert, P. 2002. Modelling the species richness distribution of French dung beetles (Coleoptera, Scarabaeidae) and delimiting the predictive capacity of different groups of explanatory variables. *Global Ecology and Biogeography*, 11**,** 265-277.

MacArthur, R. H. & Wilson, E. O. 1967. *Theory of Island Biogeography*. Princeton University Press.

Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., McMahon, S. M., Normand, S., Thuiller, W., Wüest, R. O., Zimmermann, N. E. & Elith, J. 2014. What do we gain from simplicity versus complexity in species distribution models? *Ecography*, 37**,** 1267-1281.

Merow, C., Smith, M. J. & Silander, J. A. 2013. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, 36**,** 1058-1069.

Metallinou, M., Červenka, J., Crochet, P. A., Kratochvíl, L., Wilms, T., Geniez, P., Shobrak, M. Y., Brito, J. C. & Carranza, S. 2015. Species on the rocks: Systematics and biogeography of the rock-dwelling *Ptyodactylus* geckos (Squamata: Phyllodactylidae) in North Africa and Arabia. *Molecular Phylogenetics and Evolution*, 85**,** 208-220.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. & Worm, B. 2011. How many species are there on earth and in the ocean? *PLoS Biol*, 9**,** e1001127.

Naimi, B., Hamm, N., Groen, T. A., Skidmore, A. K. & Toxopeus, A. G. 2014. Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37**,** 191-203.

Newbold, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, 34**,** 3-22.

O'Donnell, M. S. & Ignizio, D. A., 2012, Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey Data Series 691, 10 p.

Pearce, J. & Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133**,** 225-245.

Pearson, R. G. 2008. Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*. Available at http://ncep.amnh.org.

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Peterson, A. T. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34**,** 102-117.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M. & Araújo, M. 2011. *Ecological niches and geographic distributions*. Princeton University Press.

Phillips, S. J. & Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31**,** 161-175.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190**,** 231-259.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. 2017. Opening the black box: an open-source release of Maxent. *Ecography*, 40**,** 887-893.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. & Ferrier, S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19**,** 181-197.

Pineda, E. & Lobo, J. M. 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, 78**,** 182-190.

Reptiles Data-Base. 2017. http://www.reptile-database.org/ (accessed on 20 October 2017).

R Development Core Team. 2018. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rushton, S. P., Ormerod, S. J. & Kerby, G. 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology*, 41**,** 193-200.

Sanchooli, N. 2017. Habitat suitability and potential distribution of *Laudakia nupta* (De Filippi, 1843) (Sauria: Agamidae) in Iran. *Russian Journal of Ecology*, 48**,** 275-279.

Sillero, N. 2011. What does ecological modelling model? A proposed classification of ecological niche models based on their underlying methods. *Ecological Modelling*, 222**,** 1343-1346.

Sillero, N., Brito, J. C., Skidmore, A. K. & Toxopeus, A. G. 2009. Biogeographical patterns derived from remote sensing variables: the amphibians and reptiles of the Iberian Peninsula. *Amphibia-Reptilia*, 30**,** 185-206.

Sillero, N., Campos, J., Bonardi, A., Corti, C., Creemers, R., Crochet, P., et al. 2014. Updated distribution and biogeography of amphibians and reptiles of Europe. *Amphibia-Reptilia*, 35**,** 1-31.

Sillero, N. & Carretero, M. A. 2013. Modelling the past and future distribution of contracting species. The Iberian lizard *Podarcis carbonelli* (Squamata: Lacertidae) as a case study. *Zoologischer Anzeiger - A Journal of Comparative Zoology*, 252**,** 289-298.

Tytar, V., Sobolenko, L., Nekrasova, O. & Mezhzherin, S. 2015. Using ecological niche modeling for biodiversity conservation guidance in the Western Podillya (Ukraine): Reptiles. *Vestnik zoologii*, 49**,** 551-558.

VertNet. http://portal.vertnet.org/search, (downloaded on February 07, 2017).

Vincent, P. 2008. *Saudi Arabia: An Environmental Overview*. Taylor & Francis Group, UK.

Wilms, T. M., Wagner, P., Shobrak, M., Rödder, D. & Böhme, W. 2011. Living on the edge? - On the thermobiology and activity pattern of the large herbivorous desert lizard *Uromastyx aegyptia microlepis* Blanford, 1875 at Mahazat as-Sayd Protected Area, Saudi Arabia. *Journal of Arid Environments*, 75**,** 636-647.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A. & NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14**,** 763-773

Yi, Y-J., Cheng, X., Yang, Z-F. & Zhang, S-H. 2016. Maxent modeling for predicting the potential distribution of endangered medicinal plant (*H. riparia* Lour) in Yunnan, China. *Ecological Engineering*, 92**,** 260-269.

Zaniewski, A. E., Lehmann, A. & Overton, J. M. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, 157**,** 261-280.

Table 2.1: Bioclimatic variables with a Variance Inflation Factor (VIF) of less than 10 which were used to build the reptile distribution models using Maxent. The third column shows the averaged percent contribution of each environmental variable to the final distribution models of all the species. Most variable descriptions were modified from O'Donnell and Ignizio (2012).

Variable	Meaning	% contribution in the models	Description	unit	Correlation of each variable with observed species richness (Pearson's r)
Alt	Altitude	4.8	height above sea level	m	-0.0653
Bio_02	Mean Diurnal Range	32.2	calculated from the monthly mean of (max temp - min temp) as an index of daily temperature fluctuation	°C	0.0339
Bio_03	Isothermality	14.5	calculated as (100 x BIO2/BIO7); an index of the daily temperature fluctuation relative to the annual temperature fluctuation. BIO7 is the mean annual temperature range	%	-0.0012
Bio_08	Mean Temperature of Wettest Quarter	3.2		°C	0.0346
Bio_09	Mean Temperature of Driest Quarter	1.6		°C	0.1011
Bio_14	Mean Precipitation of Driest Month	4.8		mm	-0.0864
Bio_15	Mean Precipitation Seasonality	12.9	the average coefficient of variation in precipitation over a year	%	0.0482
Bio_19	Mean Precipitation of Coldest Quarter	12.9		mm	-0.0114
Srad_04	Mean Solar radiation in April	3.2		kJ m^{-2} day^{-1}	-0.0168
Srad_06	Mean Solar radiation in June	9.7		kJ m^{-2} day^{-1}	0.0377
Srad_09	Mean Solar radiation in September	0		kJ m^{-2} day^{-1}	-0.0047

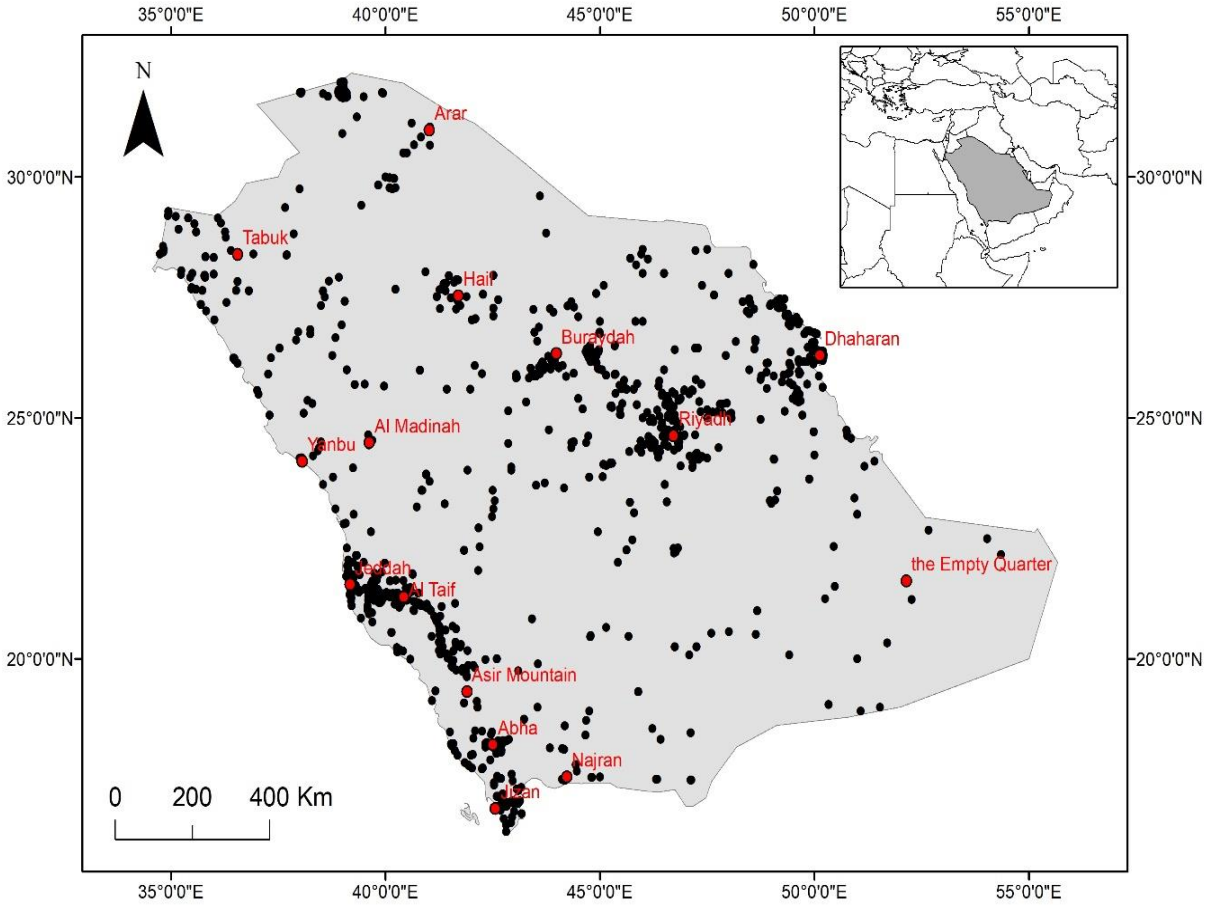


Figure 2.1: The locations in Saudi Arabia of the observations of 62 terrestrial reptile species (n = 3,943 observations in total) that were used to build distribution models.

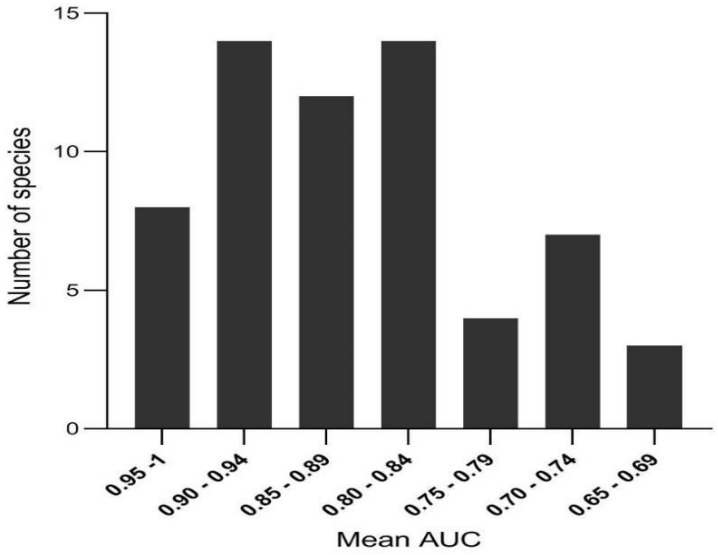


Figure 2.2: The frequencies of mean AUC values for reptile distributions models in Saudi Arabia

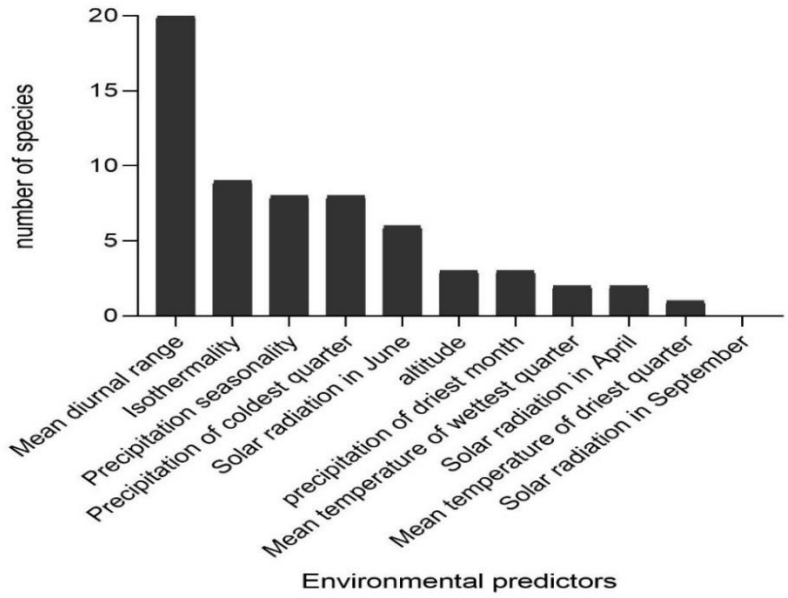


Figure 2.3: The frequencies with which each environmental predictor was the most important contributor to the models of 62 Saudi reptile species. The variable contributions were assessed using permutation importance in the Maxent output.

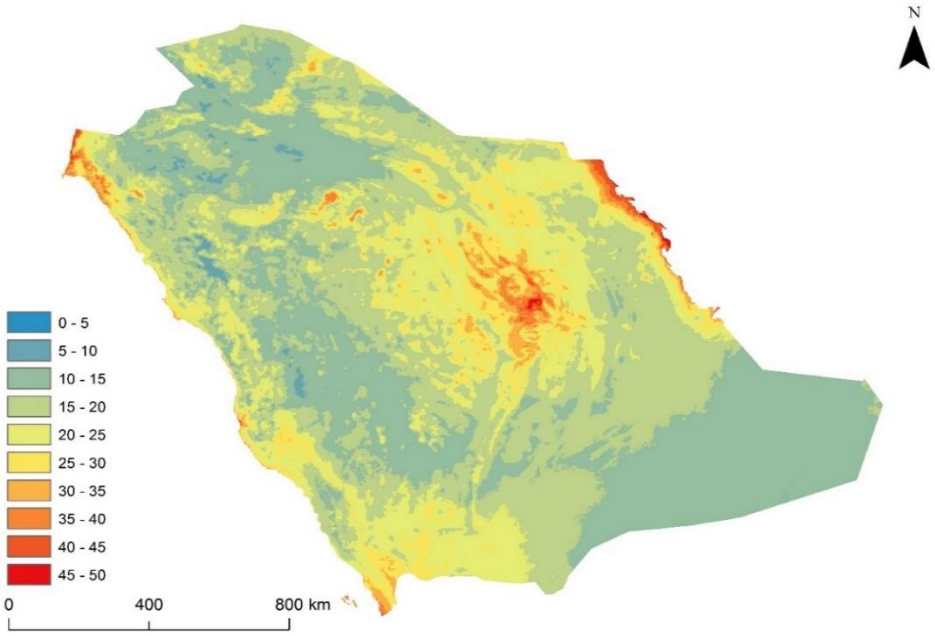


Figure 2.4: Predicted terrestrial reptile species richness for Saudi Arabia, based on summed habitat suitabilities from individual distribution models for 62 species. The map was generated by summing all averaged ascii maps for each species, and then colours rescaled to match the range of Figure 2.5.

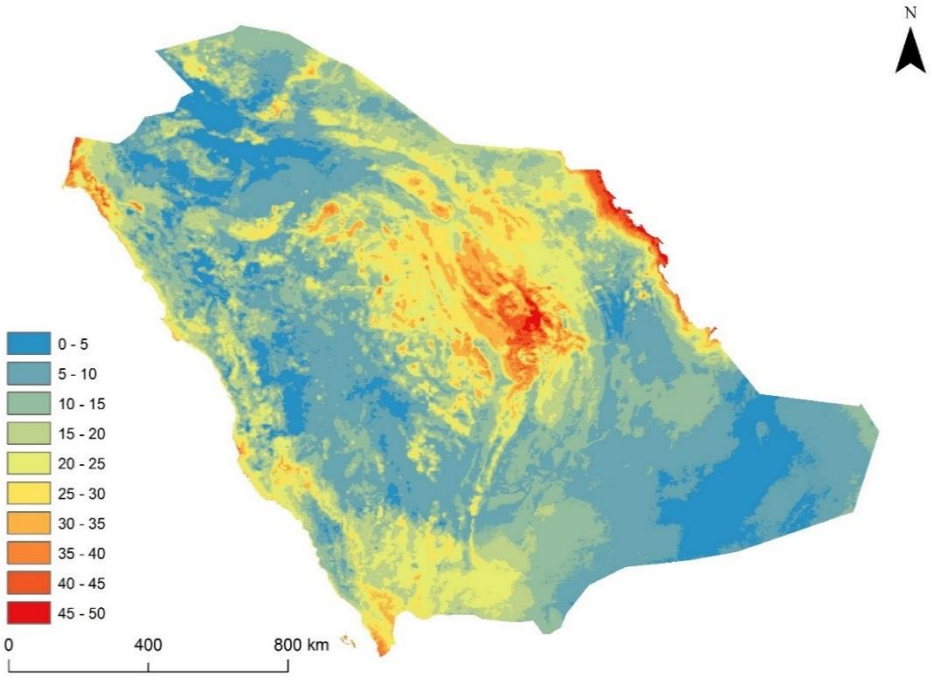


Figure 2.5: Predicted terrestrial reptile species richness for Saudi Arabia, based on summed binary predictions of suitable habitats from individual distribution models for 62 species. The map was generated by averaging the threshold maps for each reptile species, and then summing the number of species predicted to be present in each pixel.

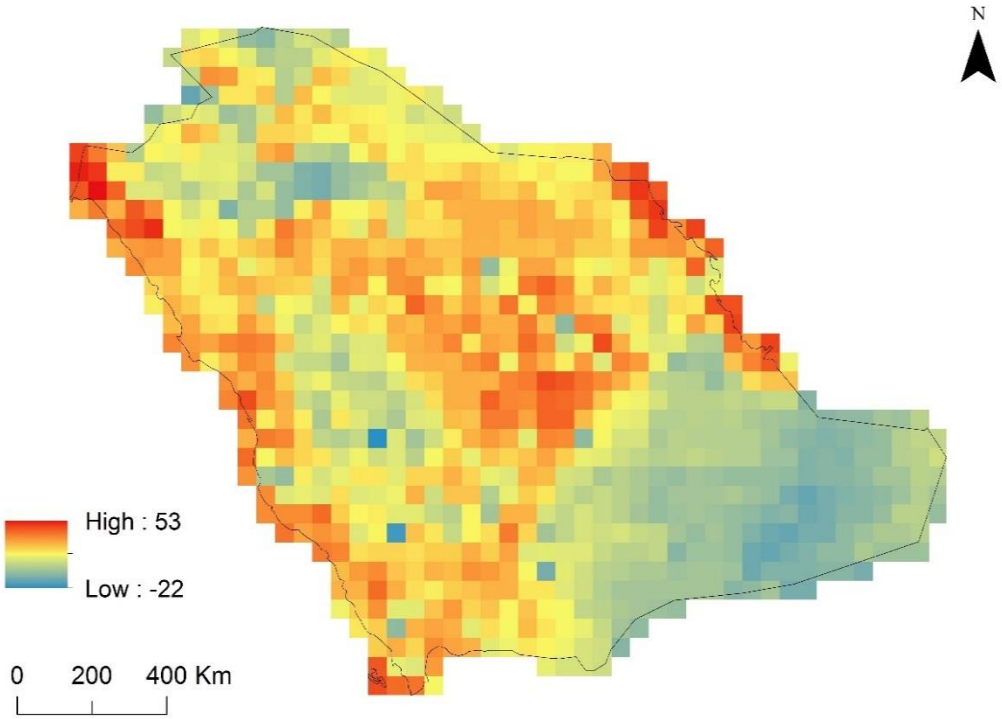


Figure 2.6: Map of Saudi Arabia (50 x 50 km grid cells) representing the difference between the predicted thresholded species richness and the actual observations from the records. Negative values mean that the observed species richness was higher than the predicted. High values mean that the predicted number of species richness at these cells was larger than the observed.

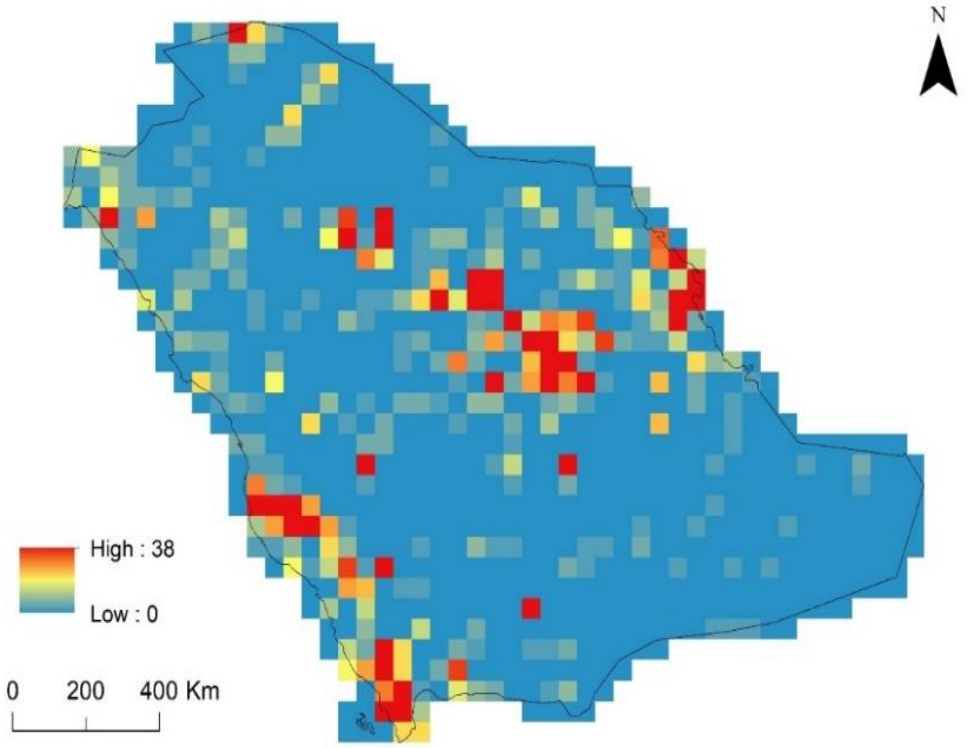


Figure 2.7: Map of Saudi Arabia (50 × 50 km grid cells) representing the observed species richness from the records.

2.7 Supplementary material

Table S2.2: Correlation matrix of all 32 bioclimatic variables considered for use in modelling. Variables are; Bio1, Annual Mean Temperature; Bio2, Mean Diurnal Range (Mean of monthly (max temp - min temp)); Bio3, Isothermality (BIO2/BIO7) (×100); Bio4, Temperature Seasonality (standard deviation ×100); Bio5, Max Temperature of Warmest Month; Bio6, Min Temperature of Coldest Month; Bio7, Temperature Annual Range (BIO5-BIO6); Bio8, Mean Temperature of Wettest Quarter; Bio9, Mean Temperature of Driest Quarter; Bio10, Mean Temperature of Warmest Quarter; Bio11, Mean Temperature of Coldest Quarter; Bio12, Annual Precipitation; Bio13, Precipitation of Wettest Month; Bio14, Precipitation of Driest Month; Bio15, Precipitation Seasonality (Coefficient of Variation); Bio16, Precipitation of Wettest Quarter; Bio17, Precipitation of Driest Quarter; Bio18, Precipitation of Warmest Quarter; Bio19, Precipitation of Coldest Quarter; Solar radiation 1 is January and 12 is December (Fick and Hijmans, 2017; https://worldclim.org/data/worldclim21.html).

	alt	bio_01	bio_02	bio_03	bio_04	bio_05	bio_06	bio_07	bio_08	bio_09	bio_10	bio_11	bio_12	bio_13	bio_14	bio_15
alt	1.00															
bio_01	-0.58	1.00														
bio_02	-0.19	0.13	1.00													
bio_03	0.20	0.40	-0.27	1.00												
bio_04	-0.21	-0.37	0.53	-0.94	1.00											
bio_05	-0.69	0.84	0.53	-0.08	0.18	1.00										
bio_06	-0.28	0.84	-0.30	0.75	-0.79	0.43	1.00									
bio_07	-0.25	-0.23	0.73	-0.85	0.96	0.33	-0.71	1.00								
bio_08	-0.12	0.81	-0.03	0.57	-0.57	0.53	0.83	-0.45	1.00							
bio_09	-0.70	0.55	0.43	-0.33	0.38	0.78	0.14	0.46	0.25	1.00						
bio_10	-0.74	0.86	0.42	-0.09	0.15	0.99	0.47	0.27	0.54	0.78	1.00					
bio_11	-0.35	0.92	-0.13	0.70	-0.71	0.56	0.98	-0.59	0.86	0.25	0.59	1.00				
bio_12	0.38	-0.13	-0.12	-0.01	0.04	-0.11	-0.08	0.00	0.03	-0.11	-0.11	-0.11	1.00			
bio_13	0.29	0.13	-0.01	0.01	0.03	0.16	0.10	0.02	0.28	0.03	0.16	0.10	0.77	1.00		
bio_14	0.41	0.01	-0.43	0.56	-0.58	-0.30	0.35	-0.60	0.25	-0.44	-0.29	0.26	0.52	0.20	1.00	
bio_15	-0.32	0.45	0.24	-0.06	0.10	0.52	0.25	0.15	0.31	0.28	0.54	0.30	-0.26	0.35	-0.51	1.00
bio_16	0.27	0.10	-0.05	-0.01	0.04	0.13	0.08	0.02	0.23	0.02	0.14	0.07	0.87	0.95	0.31	0.19
bio_17	0.36	0.12	-0.41	0.60	-0.63	-0.22	0.43	-0.62	0.35	-0.37	-0.20	0.36	0.51	0.20	0.97	-0.48
bio_18	0.44	0.05	-0.45	0.63	-0.65	-0.30	0.41	-0.66	0.30	-0.51	-0.28	0.32	0.51	0.24	0.93	-0.40
bio_19	-0.14	-0.11	0.09	-0.44	0.47	0.13	-0.31	0.42	-0.29	0.28	0.14	-0.28	0.74	0.44	0.06	-0.24
srad_01	-0.16	0.78	-0.07	0.64	-0.64	0.46	0.84	-0.52	0.77	0.05	0.48	0.86	-0.11	0.14	0.28	0.37

	alt	bio_01	bio_02	bio_03	bio_04	bio_05	bio_06	bio_07	bio_08	bio_09	bio_10	bio_11	bio_12	bio_13	bio_14	bio_15
srad_02	-0.08	0.76	-0.07	0.69	-0.68	0.43	0.84	-0.55	0.79	0.01	0.44	0.87	-0.06	0.17	0.30	0.34
srad_03	-0.03	0.66	-0.15	0.75	-0.75	0.29	0.82	-0.63	0.69	-0.13	0.31	0.82	-0.06	0.07	0.41	0.19
srad_04	-0.09	0.33	-0.15	0.58	-0.58	0.03	0.50	-0.50	0.25	-0.24	0.05	0.49	-0.27	-0.35	0.28	-0.09
srad_05	-0.27	-0.31	0.09	-0.32	0.33	-0.17	-0.41	0.30	-0.56	0.06	-0.16	-0.39	-0.15	-0.54	-0.22	-0.45
srad_06	0.04	-0.60	0.10	-0.51	0.53	-0.35	-0.70	0.46	-0.66	0.02	-0.37	-0.69	0.14	-0.21	-0.24	-0.46
srad_07	0.15	-0.71	0.06	-0.51	0.52	-0.47	-0.76	0.43	-0.72	-0.06	-0.49	-0.76	0.13	-0.22	-0.22	-0.50
srad_08	0.11	-0.70	0.04	-0.51	0.51	-0.47	-0.74	0.41	-0.75	-0.08	-0.49	-0.75	0.08	-0.30	-0.24	-0.51
srad_09	-0.19	-0.21	-0.14	-0.09	0.07	-0.21	-0.18	0.03	-0.40	-0.09	-0.18	-0.19	0.06	-0.36	0.01	-0.46
srad_10	-0.23	0.75	-0.15	0.64	-0.66	0.42	0.84	-0.55	0.69	0.05	0.45	0.85	-0.05	0.05	0.33	0.20
srad_11	-0.11	0.74	-0.08	0.68	-0.68	0.41	0.83	-0.55	0.75	-0.01	0.43	0.85	-0.08	0.12	0.32	0.33
srad_12	-0.21	0.77	-0.09	0.64	-0.64	0.46	0.84	-0.52	0.74	0.05	0.49	0.86	-0.14	0.09	0.28	0.36
	bio_16	bio_17	bio_18	bio_19	srad_01	srad_02	srad_03	srad_04	srad_05	srad_06	srad_07	srad_08	srad_09	srad_10	srad_11	srad_12
bio_16	1.00															
bio_17	0.31	1.00														
bio_18	0.36	0.94	1.00													
bio_19	0.59	0.06	0.00	1.00												
srad_01	0.12	0.36	0.41	-0.36	1.00											
srad_02	0.14	0.39	0.43	-0.36	0.99	1.00										
srad_03	0.08	0.48	0.54	-0.35	0.94	0.95	1.00									
srad_04	-0.29	0.32	0.38	-0.33	0.64	0.65	0.79	1.00								
srad_05	-0.41	-0.27	-0.27	0.32	-0.41	-0.41	-0.30	0.24	1.00							
srad_06	-0.14	-0.31	-0.35	0.46	-0.82	-0.78	-0.74	-0.35	0.75	1.00						
srad_07	-0.18	-0.30	-0.34	0.40	-0.90	-0.85	-0.81	-0.42	0.67	0.97	1.00					
srad_08	-0.24	-0.32	-0.35	0.38	-0.87	-0.84	-0.78	-0.35	0.72	0.96	0.99	1.00				
srad_09	-0.20	-0.03	0.01	0.39	-0.24	-0.23	-0.11	0.32	0.87	0.64	0.54	0.61	1.00			
srad_10	0.10	0.40	0.45	-0.20	0.94	0.93	0.92	0.71	-0.18	-0.66	-0.77	-0.72	0.05	1.00		
srad_11	0.11	0.40	0.45	-0.36	0.99	0.99	0.96	0.69	-0.37	-0.78	-0.86	-0.84	-0.19	0.94	1.00	
srad_12	0.08	0.36	0.41	-0.36	0.99	0.97	0.94	0.66	-0.36	-0.80	-0.89	-0.86	-0.20	0.95	0.98	1.00

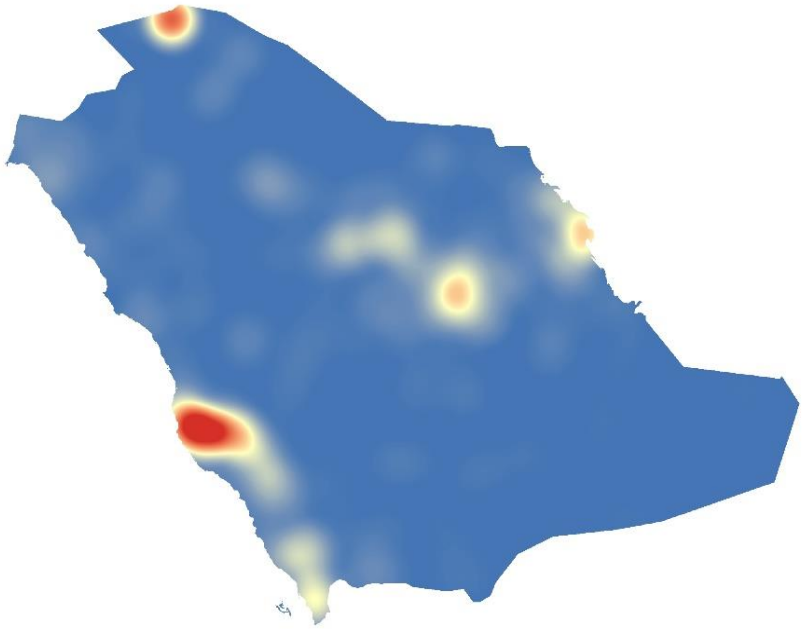
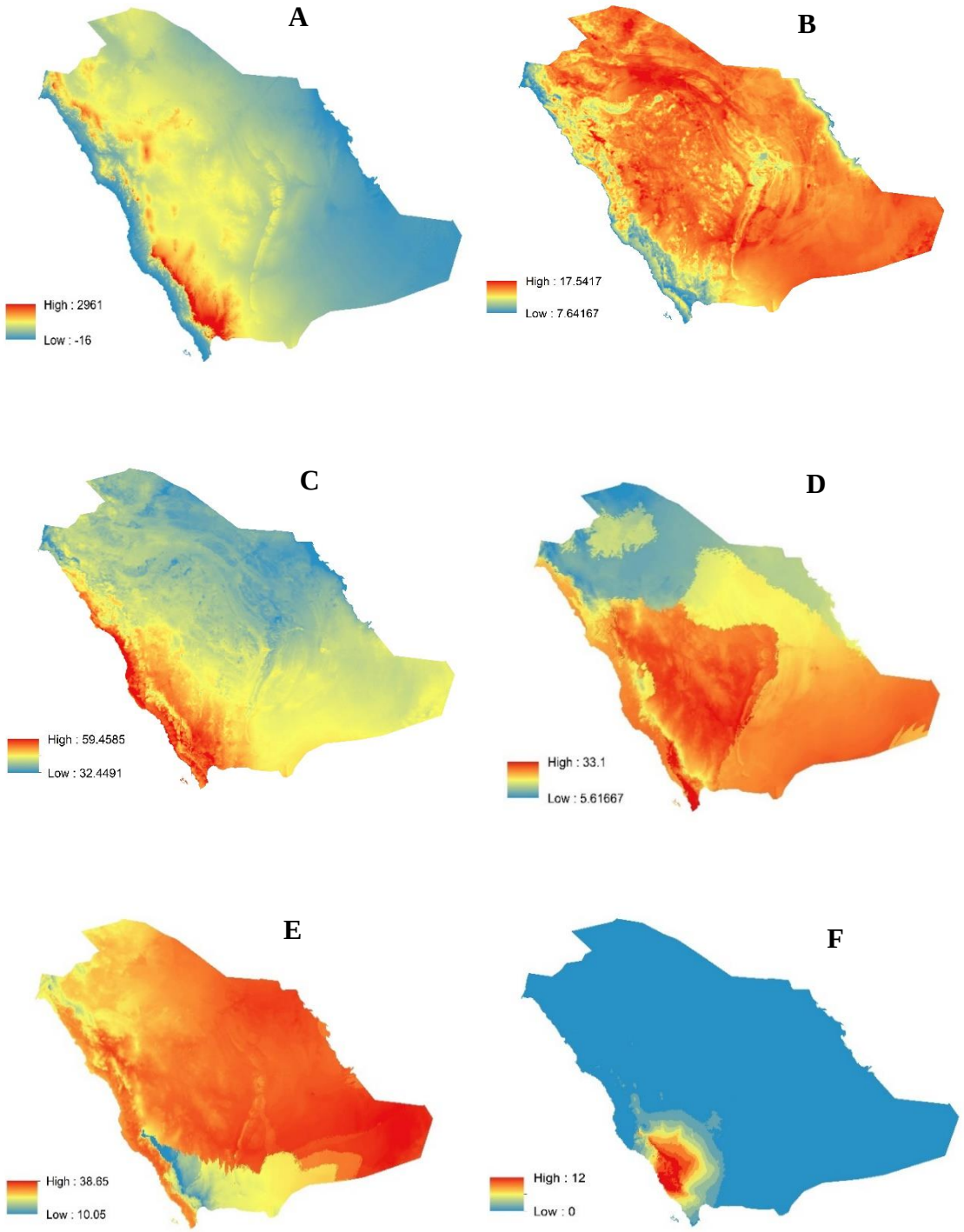


Figure S2.8: The bias file that was generated using SDMtoolbox 2.0 which then is used in Maxent to minimize the potential impact of the bias in species records. Red areas represent clusters of presence points where sampling has been most extensive.


A
High : 2961
Low : -16
B
High : 17.5417
Low : 7.64167
C
High : 59.4585
Low : 32.4491
D
High : 33.1
Low : 5.61667
E
High : 38.65
Low : 10.05
F
High : 12
Low : 0

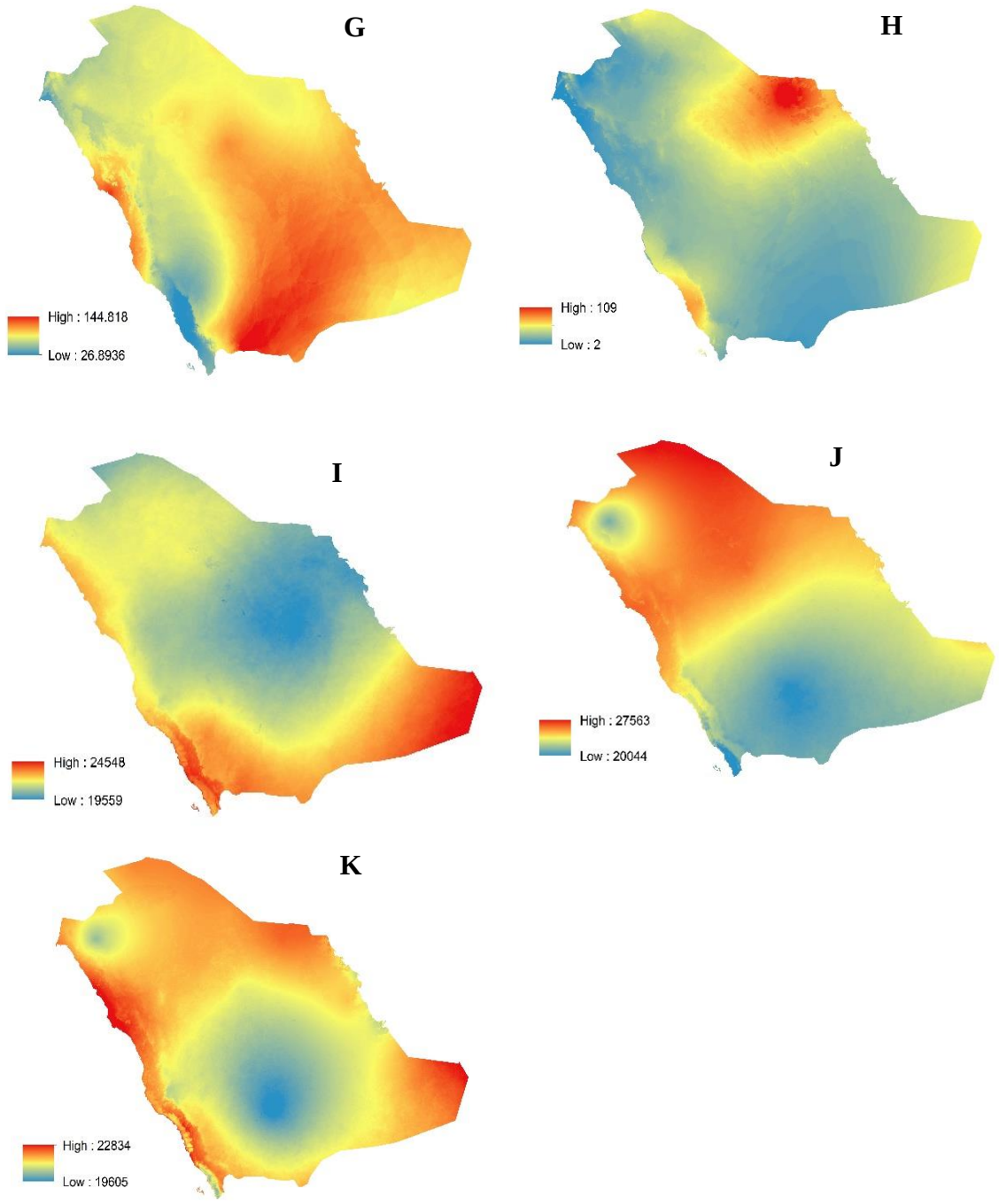


Figure S2.9: the 11 environmental variables that were used to run the model in Maxent. A (altitude), B (bio-02, mean diurnal range); C (bio-03, Isothermality); D (bio-08, Mean temperature of wettest quarter); E (bio-09, mean temperature of driest Quarter); F (bio-14, precipitation of driest month); G (bio-15, precipitation seasonality); H (bio-19, precipitation of coldest quarter); I (srad-04; solar radiation in April), J (srad-06, solar radiation in June); K (srad-09, solar radiation in September).

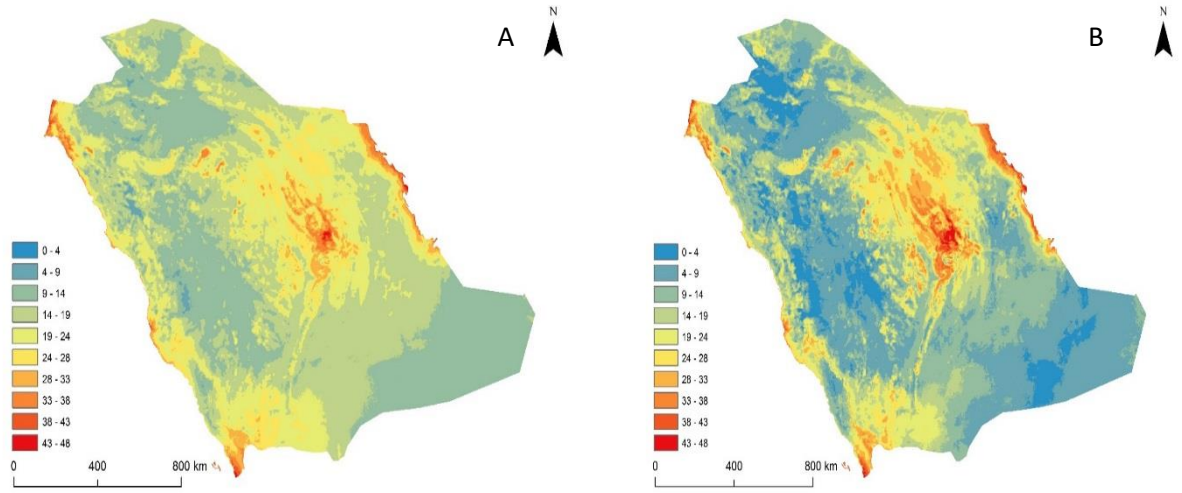


Figure S2.10: A; Predicted terrestrial reptile species richness for Saudi Arabia, based on summed habitat suitabilities from individual distribution models for 62 species. The map was generated by summing all averaged ascii maps for each species, and then colours rescaled to match the range of map S2.10B. B; Predicted terrestrial reptile species richness for Saudi Arabia, based on summed binary predictions of suitable habitats from individual distribution models for 62 species. The map was generated by averaging the threshold maps for each reptile species, and then summing the number of species predicted to be present in each pixel. Both maps were generated using 10 replications of cross-validation method.

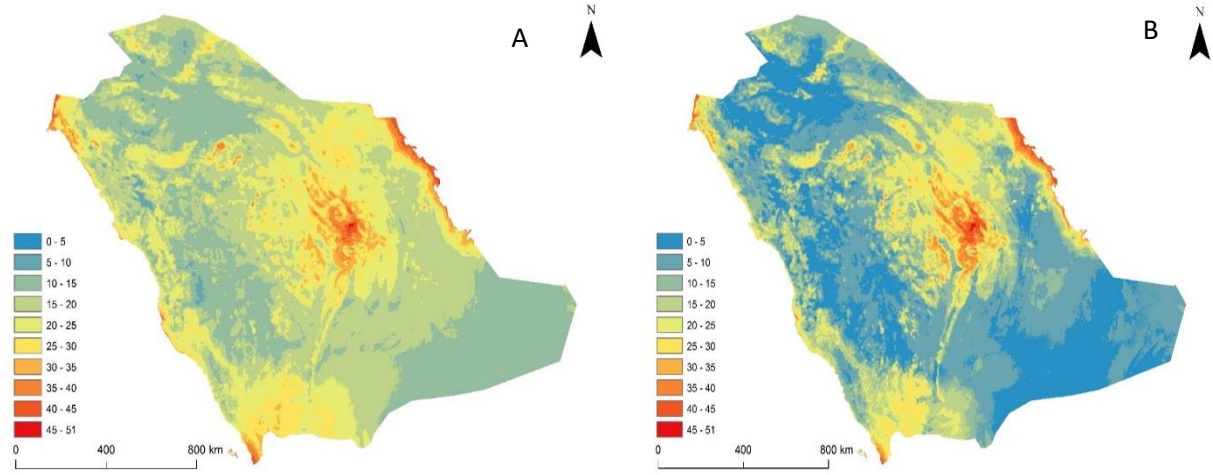
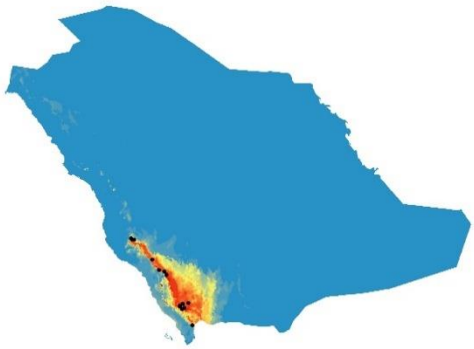


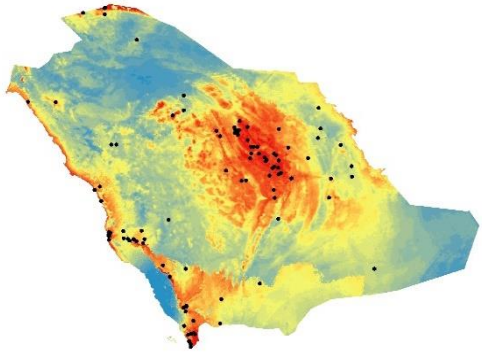
Figure S2.11: A; predicted terrestrial reptile species richness for Saudi Arabia, based on summed habitat suitabilities from individual distribution models for 62 species. The map was generated by summing all averaged ascii maps for each species, and then colours rescaled to match the range of Figure S2.11B. B; predicted terrestrial reptile species richness for Saudi Arabia, based on summed binary predictions of suitable habitats from individual distribution models for 62 species. The map was generated by averaging the threshold maps for each reptile species, and then summing the number of species predicted to be present in each pixel. Both maps were generated using 70% of data as a training data while 30% of data used as testing data.



Acanthocercus adramitanus

Records = 34

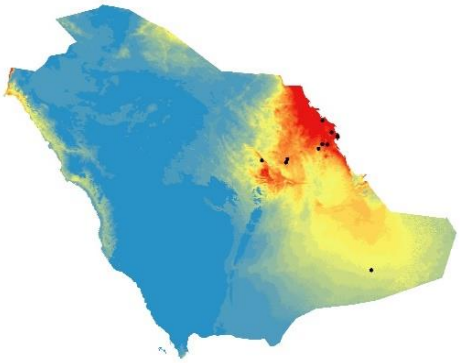
AUC = 0.99



Acanthodactylus boskianus

Records = 222

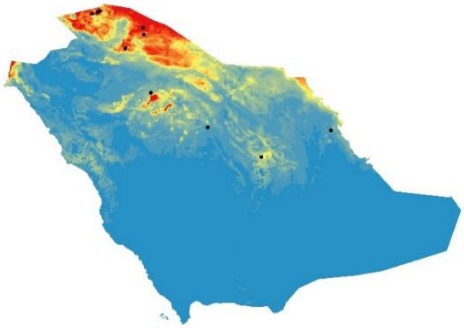
AUC = 0.75



Acanthodactylus gongrorhynchatus

Records = 20

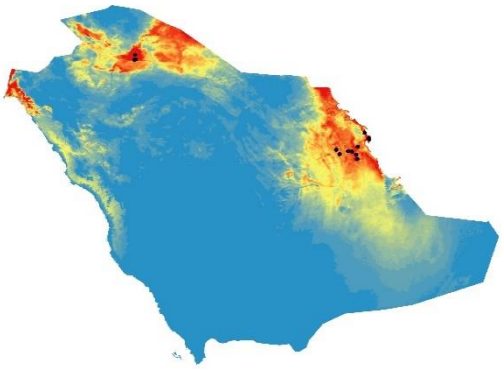
AUC = 0.92



Acanthodactylus grandis

Records = 17

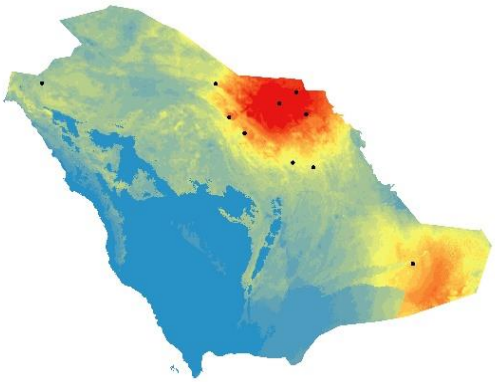
AUC = 0.93



Acanthodactylus haasi

Records = 24

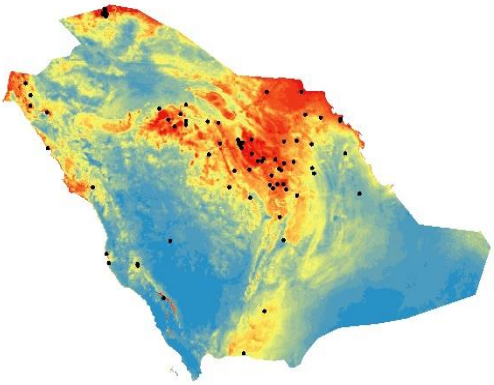
AUC = 0.96



Acanthodactylus hardyi

Records = 13

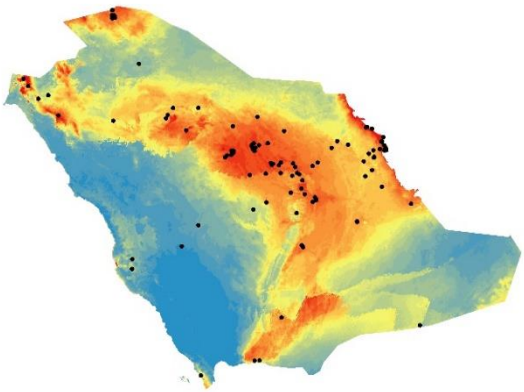
AUC = 0.82



Acanthodactylus opheodurus

Records = 90

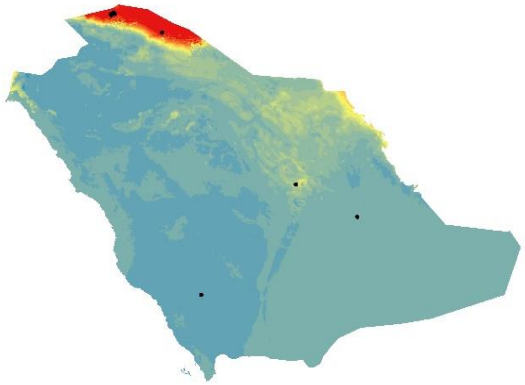
AUC = 0.75



Acanthodactylus schmidti

Records = 148

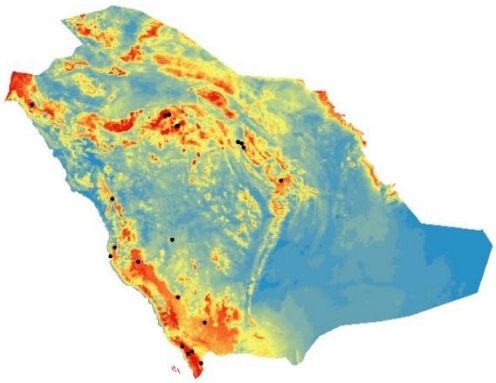
AUC = 0.81



Acanthodactylus scutellatus

Records = 12

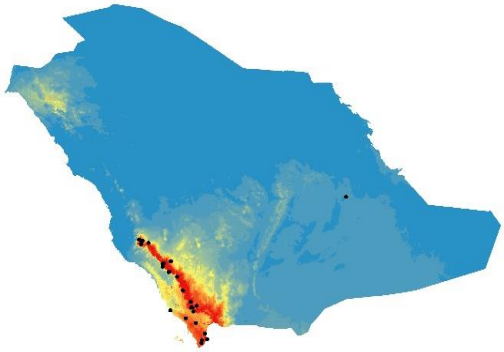
AUC = 0.8



Atractaspis engaddensis

Records = 21

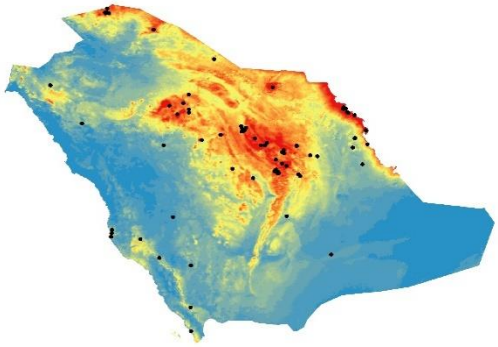
AUC = 0.84



Bitis arietans

Records = 34

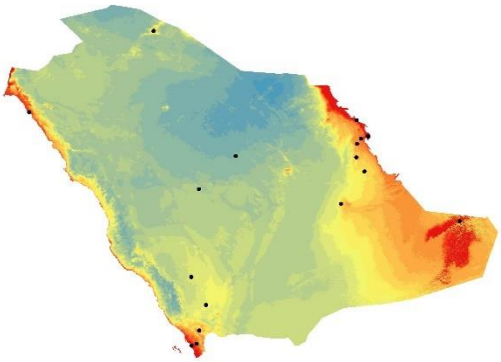
AUC = 0.91



Bunopus tuberculatus

Records = 99

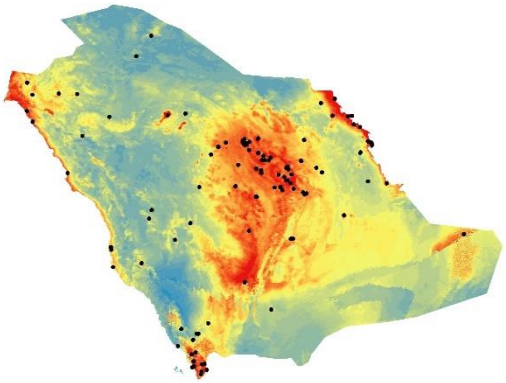
AUC = 0.83



Cerastes cerastes

Records = 30

AUC = 0.68



Cerastes gasperettii

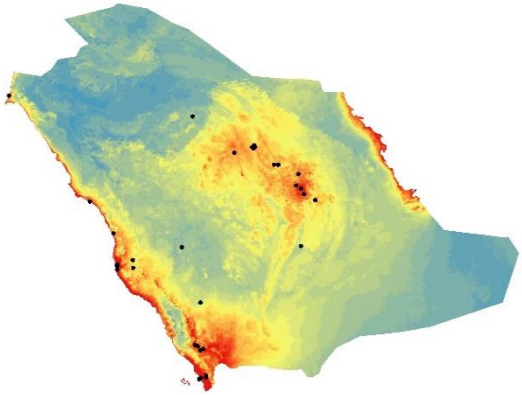
Records = 155

AUC = 0.78

Cerastes vipera

Records = 10

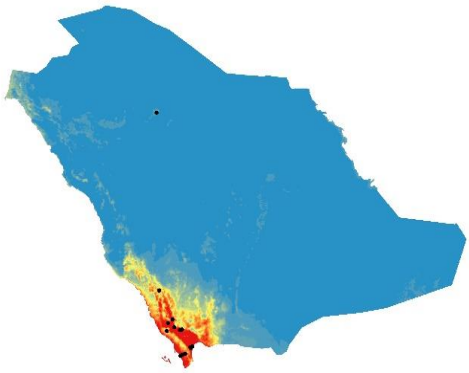
AUC = 0.99



Chalcides ocellatus

Records = 50

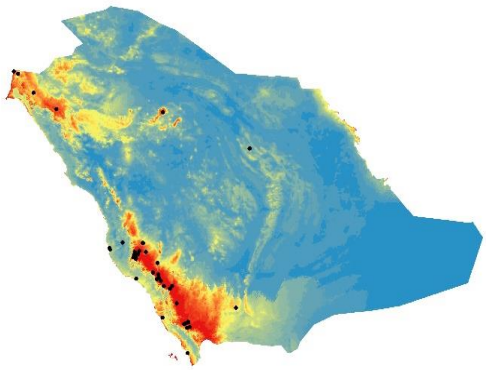
AUC = 0.86



Chamaeleo calyptratus

Records = 37

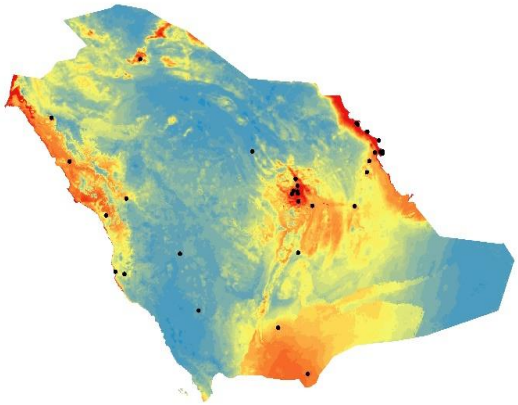
AUC = 0.98



Chamaeleo chamaeleon

Records = 67

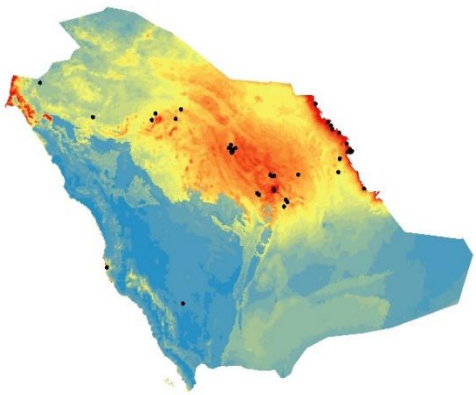
AUC = 0.88



Cyrtopodion scabrum

Records = 55

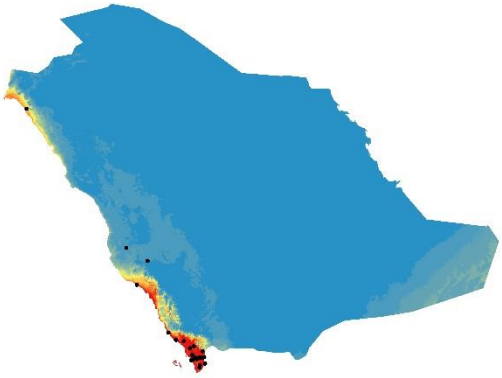
AUC = 0.86



Diplometopon zarudnyi

Records = 79

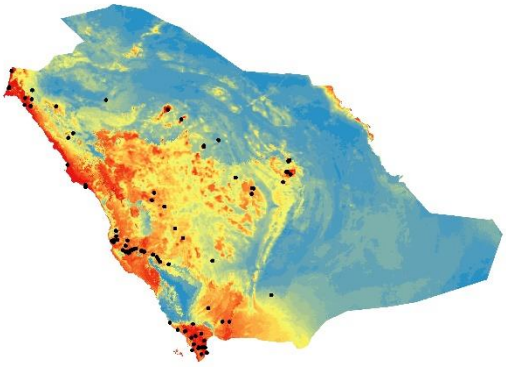
AUC = 0.92



Echis carinatus

Records = 51

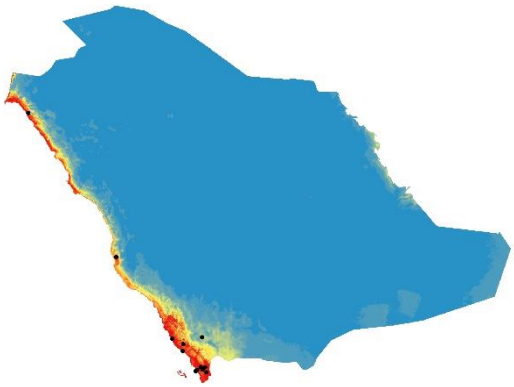
AUC = 0.88



Echis coloratus

Records = 125

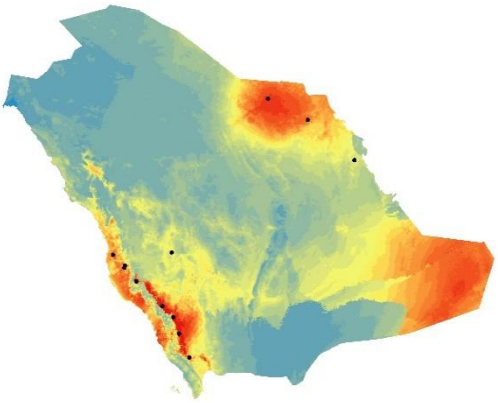
AUC = 0.86



Echis pyramidum

Records = 15

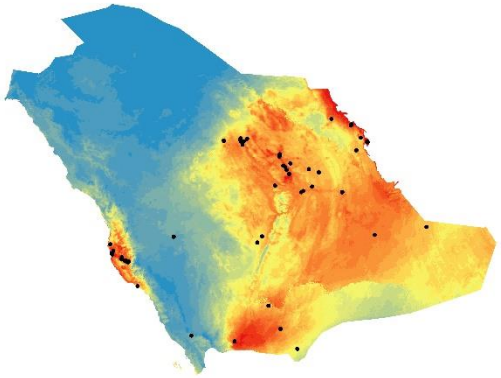
AUC = 0.98



Eirenis coronella

Records = 25

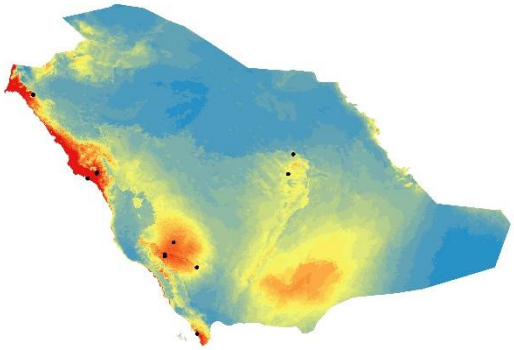
AUC = 0.84



Eryx jayakari

Records = 96

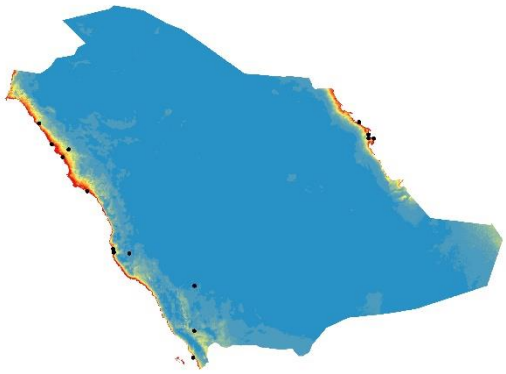
AUC = 0.86



Eumeces schneideri

Records = 14

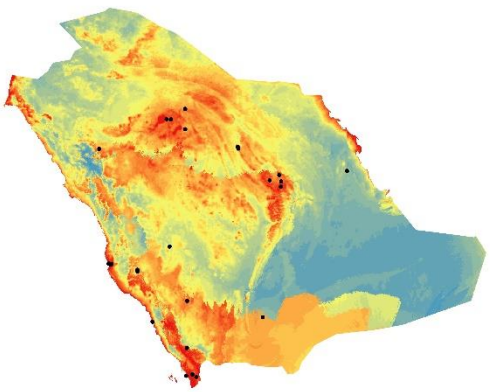
AUC = 0.94



Hemidactylus flaviviridis

Records = 31

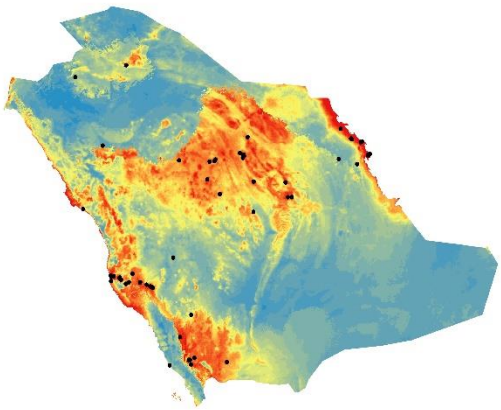
AUC = 0.94



Hemidactylus turcicus

Records = 31

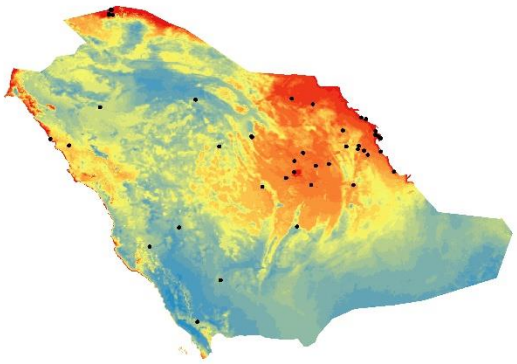
AUC = 0.82



Lytorhynchus diadema

Records = 94

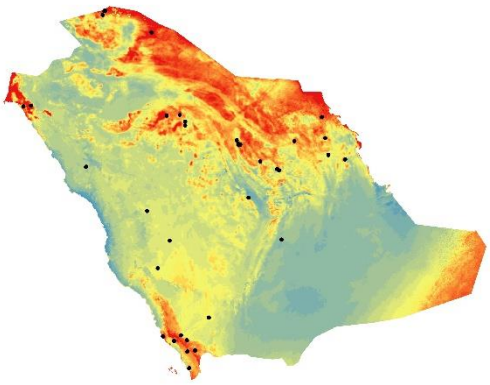
AUC = 0.81



Mesalina brevirostris

Records = 77

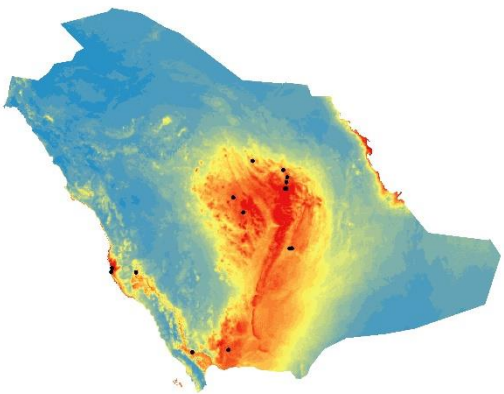
AUC = 0.89



Mesalina guttulate

Records = 42

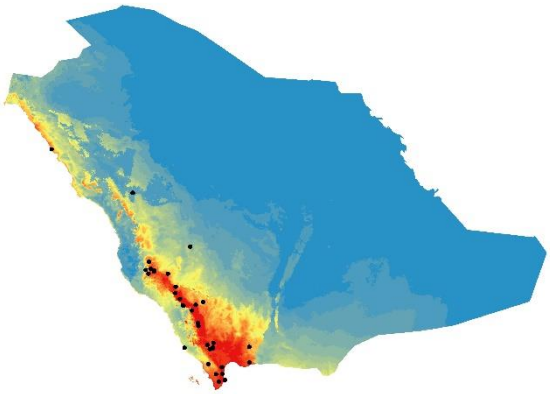
AUC = 0.70



Myriopholis macrorhyncha

Records = 16

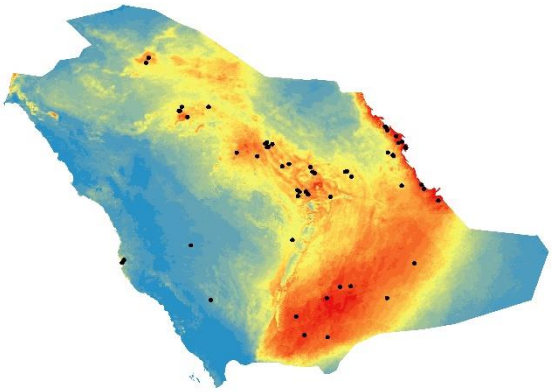
AUC = 0.88



Naja arabica

Records = 49

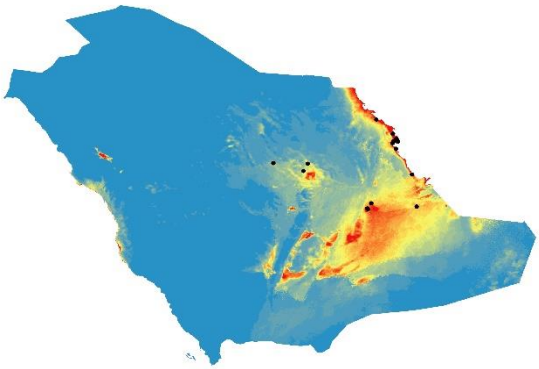
AUC = 0.92



Phrynocephalus arabicus

Records = 132

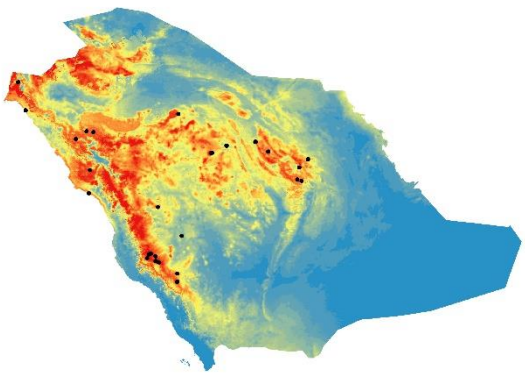
AUC= 0.76



Phrynocephalus maculatus

Records = 34

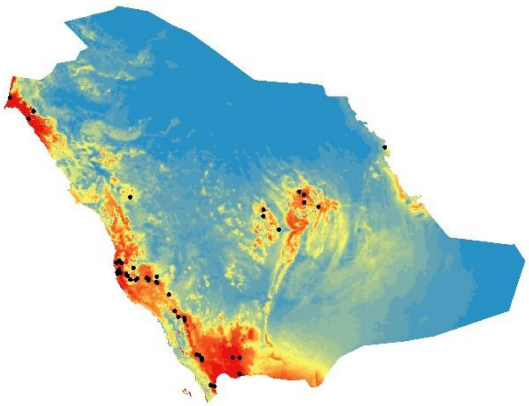
AUC = 0.97



Platyceps elegantissimus

Records = 47

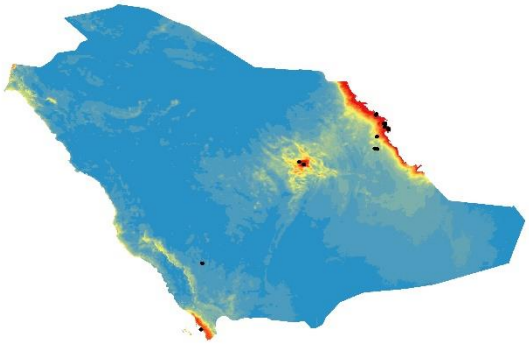
AUC = 0.80



Platyceps rhodorachis

Records = 79

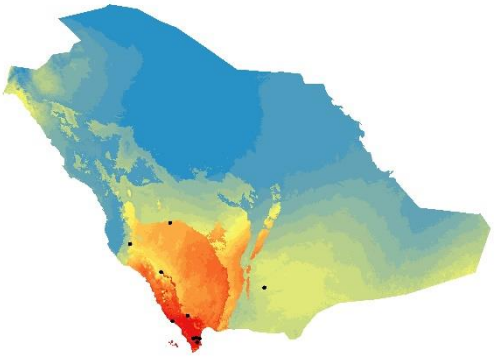
AUC = 0.92



Platyceps ventromaculatus

Records = 39

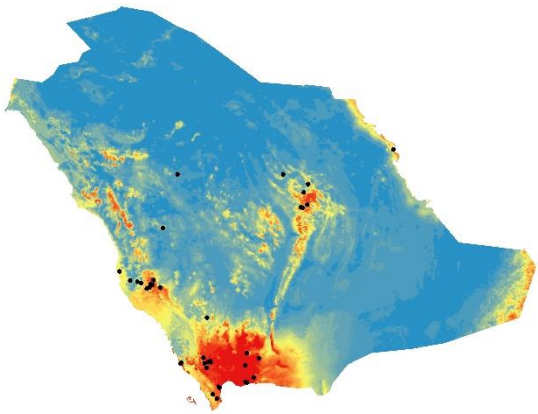
AUC = 0.99



Pristurus flavipunctatus

Records = 21

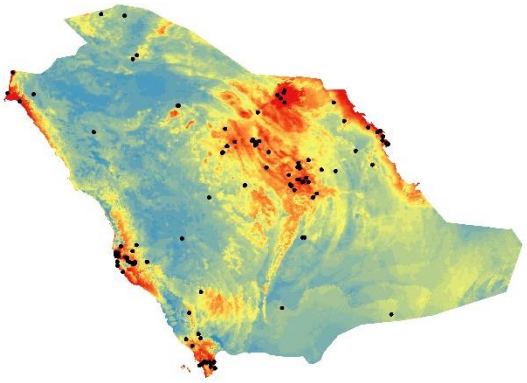
AUC = 0.74



Pristurus rupestris

Records = 61

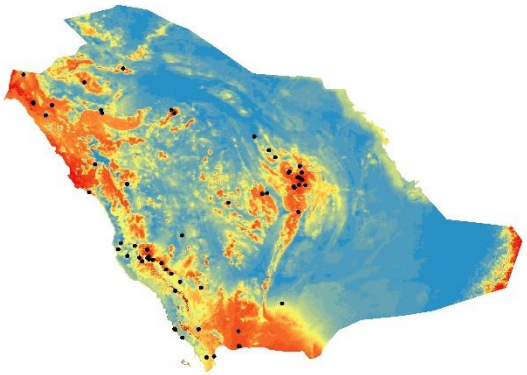
AUC = 0.91



Psammophis schokari

Records = 190

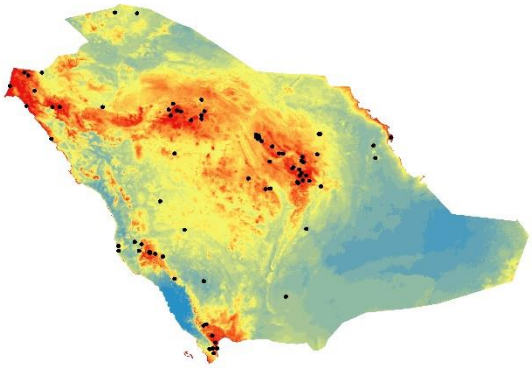
AUC = 0.83



Pseudotrapelus sinaitus

Records = 78

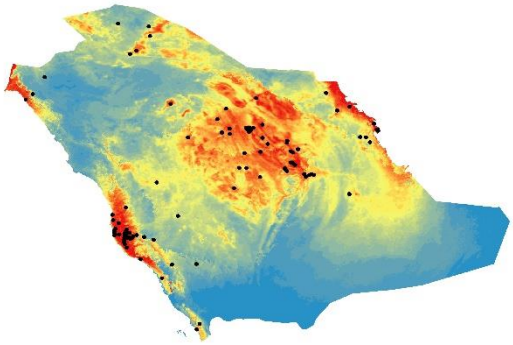
AUC = 0.82



Ptyodactylus hasselquistii

Records = 100

AUC = 0.79



Rhagerhis moilensis

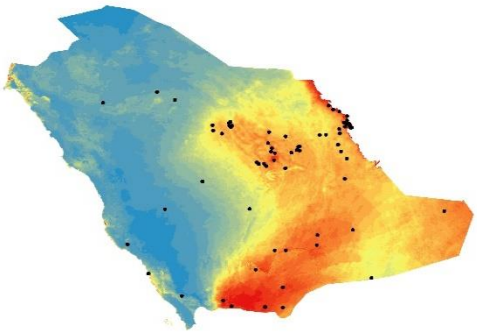
Records = 181

AUC= 0.86

Scincus hemprichii

Records = 22

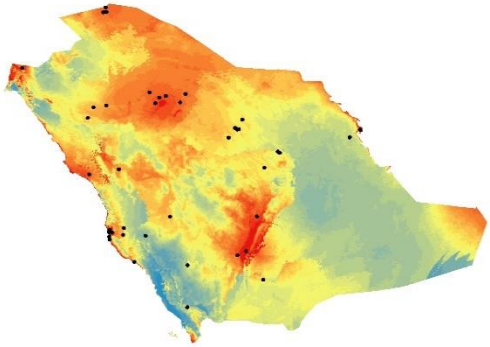
AUC = 0.99



Scincus mitranus

Records = 172

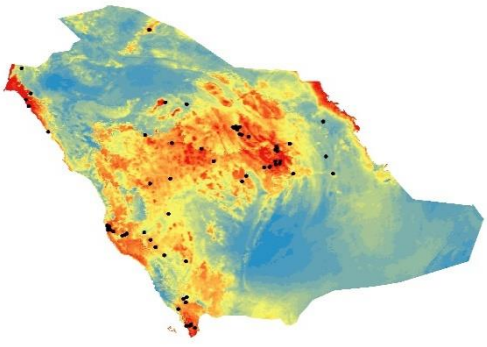
AUC = 0.71



Scincus scincus

Records = 62

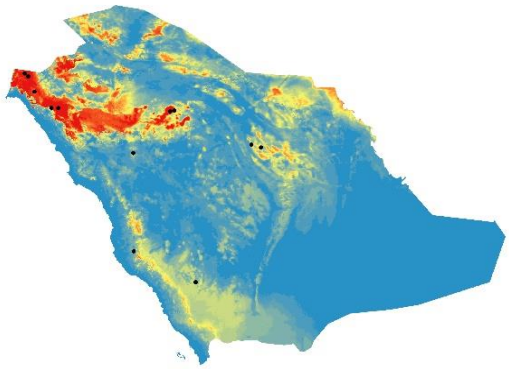
AUC = 0.68



Spalerosophis diadema

Records = 125

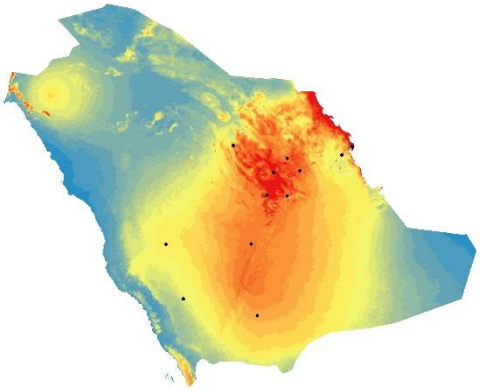
AUC = 0.84



Stellagama stellio

Records = 14

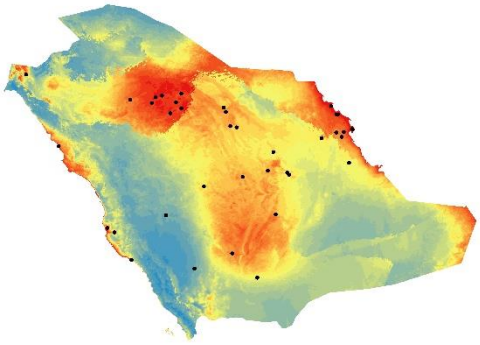
AUC = 0.90



Stenodactylus arabicus

Records = 20

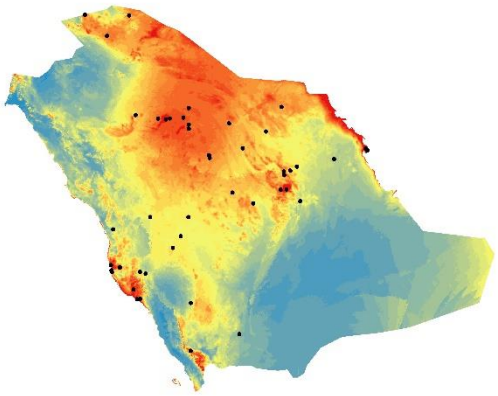
AUC = 0.70



Stenodactylus doriae

Records = 70

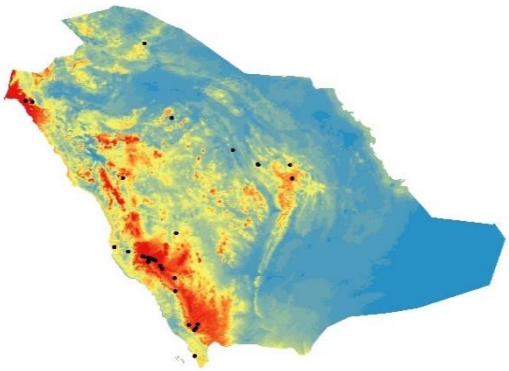
AUC = 0.67



Stenodactylus slevini

Records = 69

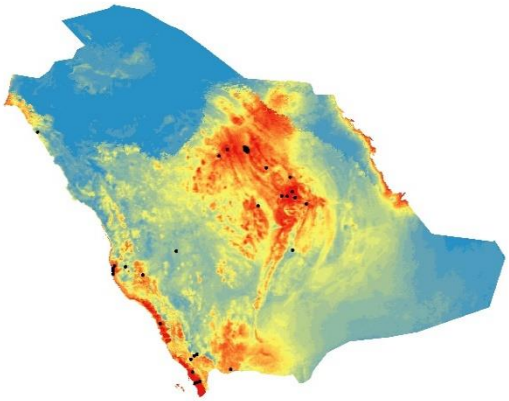
AUC = 0.81



Telescopus dhara

Records = 49

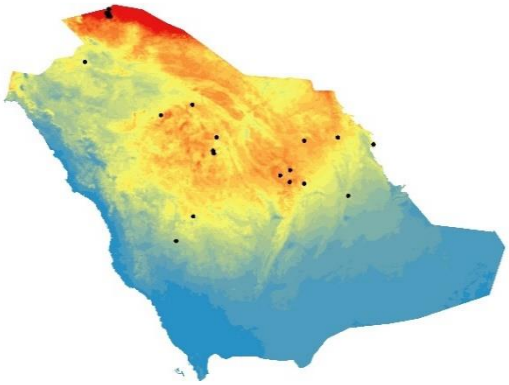
AUC = 0.84



Trachylepis brevicollis

Records = 73

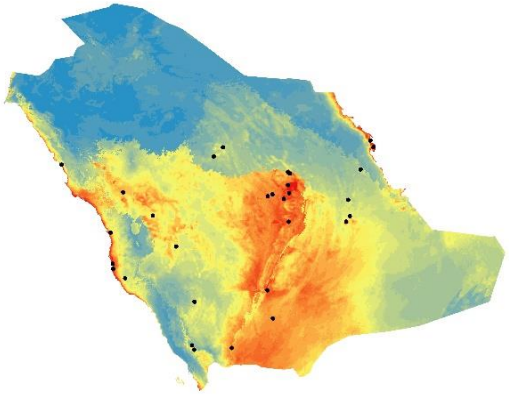
AUC = 0.81



Trapelus agnetae

Records = 35

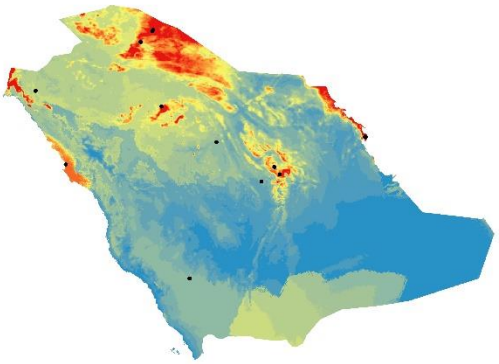
AUC = 0.84



Trapelus flavimaculatus

Records = 36

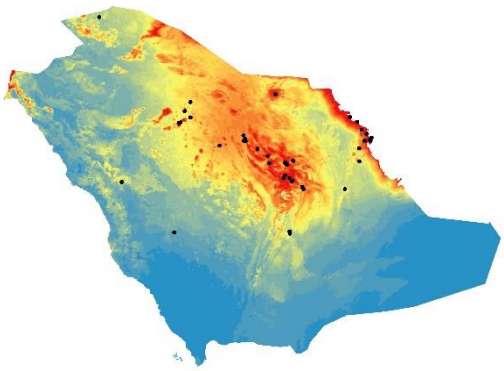
AUC = 0.72



Trapelus mutabilis

Records = 15

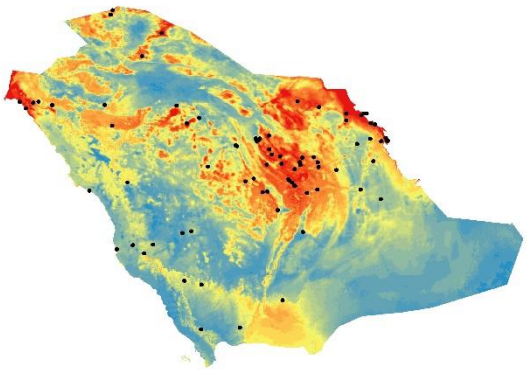
AUC = 0.72



Trapelus ruderatus

Records = 49

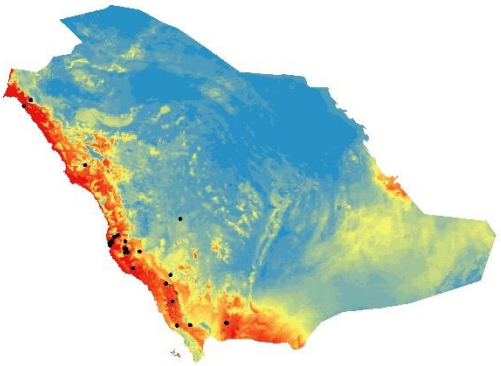
AUC = 0.93



Uromastyx aegyptia

Records = 122

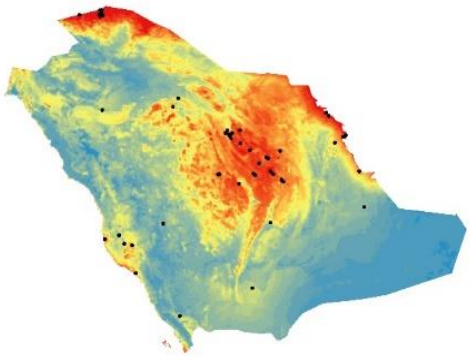
AUC = 0.73



Uromastyx ornate

Records = 52

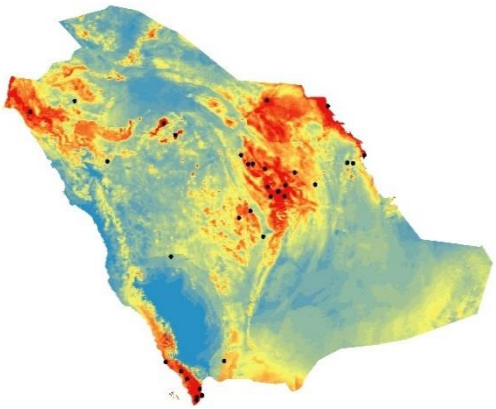
AUC = 0.94



Varanus griseus

Records = 67

AUC = 0.89



Walterinnesia aegyptia

Records = 46

AUC = 0.88

Figure S2.12: The predicted habitat suitability of each of the 62 species (41 lizards, 21 snakes) of reptile in Saudi Arabia. The final consensus probability map for each species is the mean output from ten replicate distribution models. All maps are coloured to match the same legend.

3 Chapter Three: The transferability of species distribution models: comparing the performance of local and Egyptian models for Saudi Arabian reptiles

Abstract

Efforts to study patterns of species richness are not evenly distributed around the globe, resulting in poor knowledge of species distributions. The ability to transfer the predictions of species distribution models (SDMs) from one area to another has the potential to inform conservation decision making, and to direct fieldwork, in poorly studied geographic regions. However, SDM transference is relatively poorly studied, and rarely tested using independent data from the region of interest. In this study, I took advantage of the well-studied terrestrial reptiles of Egypt to test the performance of SDMs built for Egypt spatially transferred into Saudi conditions, considering both species richness and the distributions of individual species. I compared transferred model performance with the performance of alternative models built with newly collated Saudi data, with a method of controlling for environmental differences. I used Maxent software to transfer Egyptian models into Saudi conditions. Multivariate Environmental Similarity Surface analysis (MESS) investigated and controlled for differences in environmental conditions between the two regions. For reptile species in common between Egypt and Saudi Arabia, I compared the transferred model with independent Saudi models built using Saudi data. The Egyptian models yielded very effective predictions, broadly analogous to those generated by local independent models based on Saudi data. Indeed, for some species, the Egyptian model performed better than the equivalent Saudi model. Stacked transferred Egyptian models showed similar patterns of species richness to Saudi models. The transference of SDM predictions appears to be a very effective tool that can be used to explore spatial patterns of diversity, especially when there is

a paucity of data, giving potentially valuable insights into fundamental ecological questions as well as applied conservation problems.

Keywords: Egypt, Maxent, model transferability, Saudi Arabia, species distribution model, species richness

3.1 Introduction

Species distribution modelling is one of the most rapidly growing techniques in ecology, widely applied in the last two decades (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005; Elith and Leathwick, 2009). Species distribution models (SDMs) usually work by combining known species occurrence records with spatial layers of digital environmental data, using statistical techniques to estimate the likely geographical distribution over an area of interest (Pearson, 2008; Elith and Leathwick, 2009). The main objective of species distribution modelling is to produce maps of the predicted distribution that explain and capture species-environment associations (Guisan and Zimmermann, 2000; Aguirre-Gutierrez *et al.*, 2013). The predicted maps demonstrate how habitat suitability varies through space and time, by identifying sites in the study area that have similar environmental variables to those where the species has been detected (Pearson, 2008). However, despite considerable progress in the development of these techniques, predicting species distributions is still a complex and challenging process (Merow *et al.*, 2014).

To predict the distribution of any particular species, a distribution model can be used in three ways (Elith and Leathwick, 2009; Peterson *et al.*, 2011; Duque-Lazo *et al.*, 2016). First, *interpolation* predicts the habitat suitability of new unvisited sites within the environment extent/range used to calibrate the model. Second, *extrapolation* can be used to predict into areas experiencing environmental conditions outside the range of those for which

the original model was calibrated (such as new climatic conditions generated by climate change). Third, *transference* uses the model to predict outside the geographical area where it was calibrated (Elith and Leathwick, 2009; Peterson *et al.*, 2011).

One of the most interesting recent applications of species distribution models is transference, predicting the potential distribution of species in landscapes different from those in which the data were gathered (Randin *et al.*, 2006; Elith and Leathwick, 2009; Peterson *et al.*, 2011; Werkowska *et al.*, 2017; Yates *et al.*, 2018). Transference can be used to answer fundamental questions in ecology and biogeography about the movement of species in space (Randin *et al.*, 2006; Peterson *et al.*, 2011; Heikkinen *et al.*, 2012; Duque-Lazo *et al.*, 2016; Werkowska *et al.*, 2017; Yates *et al.*, 2018). It has already been used to test the predictive accuracy (transferability) of distribution models between contiguous and disjunct areas: Randin *et al*. (2006) tested the transferability success between Switzerland and Austria for plants, and Duque-Lazo *et al.* (2016) looked at the effect of the number of predictor variables on transferability between south-western Spain and Australia for oomycete fungi.

Transference of SDM predictions is thought to work best when using predictors which have a direct (physiological) effect on species, rather than indirect predictors, since the latter may fail to encompass the true species-habitat association (Austin, 2002; Randin *et al.*, 2006; Franklin, 2009). As always with SDMs, transference requires the assumption that a species is in equilibrium with its environment (Elith and Leathwick, 2009; Varela *et al.*, 2009). If this assumption is correct, accurate transferability is still not guaranteed even when the model shows a good level of accuracy within the area in which it has been calibrated (Varela *et al.*, 2009; Heikkinen *et al.*, 2012; Moreno-Amat *et al.*, 2015; Duque-Lazo *et al.*, 2016). Transference into new landscapes is expected to perfom better when similar environmental conditions are present (Werkowska *et al.*, 2017). Overall, the transference of SDM predictions is a challenging process: caution is needed because the validity of predictions in

new areas is generally unknown (Zanini *et al.*, 2009), and constraints need to be carefully considered (Werkowska *et al.*, 2017; Sequeira *et al.*, 2018; Yates *et al.*, 2018).

There are many factors that may affect transferability (Werkowska *et al.*, 2017; Sequeira *et al.*, 2018; Yates *et al.*, 2018), such as species characteristics (e.g., dispersal ability), the complexity of the model, environmental ranges in the two landscapes, and the sources and types of occurrence data (Randin *et al.*, 2006; Werkowska *et al.*, 2017; Yates *et al.*, 2018). Choosing appropriate strategies to maximise transferability is thus important in determining the value of the model predictions (Randin *et al.*, 2006; Phillips, 2008; Moreno-Amat *et al.*, 2015; Werkowska *et al.*, 2017; Yates *et al.*, 2018). Partly because of these difficulties, spatial transferability is still poorly understood and rarely applied (Yates *et al.*, 2018), although it is gaining more attention, especially with the general growth in the popularity of species distribution modelling (Wenger and Olden, 2012). Modellers are trying to come up with comprehensive guidelines applicable in different landscapes with different species groups to enhance transferability in space and time (Werkowska *et al.*, 2017; Sequeira *et al.*, 2018; Yates *et al.*, 2018).

Here I study the transferability of SDMs for reptiles between Egypt and Saudi Arabia, comparing transferred models to models generated from Saudi data. Reptiles in Egypt and Saudi Arabia present a good opportunity for a case study of distribution model transference because: i) reptiles are widely distributed and diverse in each country; ii) they are the dominant vertebrates in their ecosystems; iii) they cannot move much, and hence are more tied to the local environment than other more vagile taxa; and iv) they are well-studied in Egypt (Baha El Din, 2006), but not in Saudi Arabia, where there have been few reptile surveys. Transference of predictions from models built using Egyptian data has the potential to provide insights into Saudi reptile communities, and identify areas where fieldwork and

conservation actions should be prioritised. New datasets for Saudi reptiles provide an opportunity to test the validity of the transference approach.

My main objective in this paper is therefore to use the most comprehensive dataset for the herpetofauna of Egypt, complied by BioMap project, to create models for terrestrial reptile species in Egypt and transfer their predictions into the conditions of Saudi Arabia. I then compare the transferred predictive power of the Egyptian models with those of independent models built from interpolating Saudi data, and hence assess transferability accuracy between Saudi Arabia and Egypt.

3.2 Methods & Materials

3.2.1 Study areas

Saudi Arabia is a large hyper-arid country that covers most of the Arabian Peninsula (Figure 3.1) and contains a wide range of topographical and climate variation. Due to its tropical and subtropical location, the climate in Saudi Arabia is generally hot during the summer, reaching 48 °C in some locations, with scarce rainfall (Vincent, 2008; Gosling *et al.*, 2011). Egypt is in the extreme northeast of the African continent (Figure 3.1), and is officially the driest country on Earth (El-Gabbas *et al.*, 2016), having slightly less average rainfall than Saudi Arabia. The climate in Egypt is also generally hot during the summer. The presence of the Nile River and its surrounding valley and delta have a positive impact on Egyptian wildlife, including its herpetofauna (El-Gabbas *et al.*, 2016). To date, at least 118 reptiles have been recorded from Egypt (Baha El Din, 2006; El-Gabbas *et al.*, 2016). The dominant habitat in both countries is arid open desert, with very limited vegetation occurring only in particular locations. Egypt and Saudi Arabia are geographically separated by the Red Sea with the Gulf of Aqaba in the north. The similarity in the environments of the two countries make them highly suitable for a study of model transferability, whilst their

relatively slight geographical disjunction is not thought to have a strong influence on transferability (see Yates *et al.*, 2018).

3.2.2 Species occurrence data

Presence-only reptile occurrences in Egypt were obtained from the data of the BioMap project (see El-Gabbas *et al.*, 2016). This is the most comprehensive dataset for the herpetofauna of Egypt, and it has been extensively checked for taxonomy and geo-referencing. I limited modelling to terrestrial lizards and snakes with at least five occurrences, leaving a total of 71 species (49 lizards and 22 snakes) with 11,103 records to be modelled using Maxent (Figure 3.1). In Saudi Arabia, there was no collated dataset on terrestrial reptiles, and thus I created one using all available resources from museums and relevant literature in presence-only format. This involved extensive geo-referencing of locations. The result was a dataset for 62 reptile species (41 lizards and 21 snakes) with 3,943 records (Figure 3.1), the most comprehensive for the herpetofauna of Saudi Arabia to date.

When modelling the distribution of reptile species richness in both countries, and testing transferability, I used: 1) all species in Egypt; and 2) only species in common between the two countries. In my dataset, there are 25 terrestrial reptiles common to both Egypt (6,203 data points) and Saudi Arabia (1,791 data points) (Figure S3.6).

3.2.3 Environmental predictors

Environmental variables were downloaded from WorldClim Version 2.0, and included 19 bio-layers and 12 indices of solar radiation; altitude was downloaded from Version 1.4 (see www.worldclim.org) (Hijmans *et al.*, 2005; Fick and Hijmans, 2017). To reduce multicollinearity, I employed the Variance Inflation Factor (VIF) using the library

'*usdm*' in R (Naimi *et al.*, 2014; R Developments Core Team, 2018). Since my aim was to test transferability of Egyptian models into Saudi Arabia, I used Saudi Arabian environmental variables for the VIF analysis. Restricting the predictors to variables with a VIF < 10 retained 11 variables for use in the modelling (Table 3.1). Variables of vegetation cover (such as NDVI) were not used because the vegetation in hyper-arid environments is extremely sparse and has been shown to be a poor predictor of reptile distributions in Egypt (see El-Gabbas *et al.*, 2016). All the environmental variables were clipped to the size of the study area and converted to ascii files using ArcGIS (Version 10.3.1).

3.2.4 Distribution modelling and similarity analysis

I modelled reptile distributions using Maxent Version 3.4.1, downloaded from https://biodiversityinformatics.amnh.org/open_source/maxent/ (Phillips *et al.*, 2006). Maxent is a powerful tool based on machine-learning methods that is much used in species distribution studies. I chose Maxent because: i) it has the capability of projecting to different scenarios; ii) it is very robust, producing reliable models from presence-only data; iii) it works with low numbers of records; and iv) it appears to be relatively insensitive to sample bias (Baldwin, 2009; Franklin, 2009; Merow *et al.*, 2013). Finally, and most importantly, Maxent has been used to transfer models spatially and temporally (Tuanmu *et al.*, 2011; Verbruggen *et al.*, 2013; Moreno-Amat *et al.*, 2015; Manzoor *et al.*, 2018), and showed good transferability in comparison with regression techniques (see Heikkinen *et al.*, 2012; Duque-Lazo *et al.*, 2016).

I predicted the distributions of individual reptile species, and total species richness, using three types of model. First, I modelled the distribution of reptiles in Egypt, using Egyptian occurrence data (hereafter "Egypt (inter) models"). Second, I transferred these

models to create predictions for reptiles in Saudi Arabia, based on Egyptian data ("Egypt (trans) models". Finally, I used Saudi occurrence data to create independent predictions for reptiles in Saudi Arabia ("Saudi models"). For clarity, a flow-chart (Figure S3.7) shows the process of generating the Egyptian models.

When transferring models from Egypt to Saudi Arabia, I used the capability of Maxent to measure and allow for environmental dissimilarity through Multivariate Environmental Similarity Surface Analysis (MESS), using the commands from the appendix of Elith *et al.* (2010), and the tutorial of Phillips (2017). First, I created models which included the full set of 11 environmental variables as predictors, hereafter referred to as Egyptian Model Set 1 (all species included), and Egyptian Model Set 2 (only species in common between Egypt and Saudi Arabia included) (Figure S3.8). I then ran MESS using these models and removed successively variables that contributed most to any dissimilar predicted environments between Egypt and Saudi Arabia until no further reductions in dissimilarity were evident, and I had a mostly positive surface for Saudi Arabia (Figure S3.9). A total of six variables were retained (highlighted grey in Table 3.1) and used to run and transfer the model again in Maxent for all species (Egyptian Model Set 3) and only species in common between Egypt and Saudi Arabia (Egyptian Model Set 4). Comparing between sets of models with the full set of predictors and the reduced set allowed us to test the possible influence of extrapolating outside the environmental data of Egypt on the transferability of model predictions.

In creating all models, I used the same features and settings in Maxent to generate the transferred predictions: product and hinge features; ten subsampled replicates; 1000 iterations; and the new default *cloglog* output (for more details, see Phillips *et al.*, 2017). I partitioned the data into 90% training and 10% test datasets; and I used the '10% training presence' as the threshold rule to convert continuous habitat suitability into dichotomous

suitable/unsuitable values (Liu *et al.*, 2005). The permutation importance table was used to determine the importance of each variable and the degree it contributed to the model.

3.2.5 Bias and spatial autocorrelation

Overall, the geographical coverage of the sample points in Egypt and Saudi are satisfactory (Figure 3.1). However, as is expected with museum data and unsystematic surveys (Graham *et al.*, 2004; Newbold, 2010), the data show an unavoidable bias toward cities, road networks, and easily accessible areas. To minimize the impact of bias and spatial autocorrelation in my model predictions, I employed a target-group bias file (Phillips *et al.*, 2009) for all Egyptian and Saudi model sets. This uses the sum of all the records of the taxon of interest as an estimate of sampling effort, smoothed using a Gaussian kernel estimation function (using the SDMtoolbox 2.0 in ArcGIS: Brown *et al.*, 2017). The bias file is used in Maxent to promote the selection of more background points from sampled (i.e. biased) locations. It is a common approach used in distribution modelling to correct for sampling bias, and is known to improve model accuracy (Phillips *et al.*, 2009), and hence transferability.

3.2.6 Model evaluation

To evaluate the accuracy of the model, I used the Area Under the Curve ("AUC") calculated from the "Receiver Operating Curve" (ROC) (Fielding and Bell, 1997; Pearce and Ferrier, 2000). The ROC curve plots sensitivity against (1 – specificity) for all possible values of the threshold habitat suitability above which the habitat is assumed to be suitable (Fielding and Bell, 1997). The AUC can be interpreted as the probability that a randomly chosen

presence site will be more highly ranked than a randomly chosen absence (Pearce and Ferrier, 2000).

3.2.7 Assessing transferability

Assessing the accuracy of the predictions derived from the transferred model is an important goal of this study. Following Duque-Lazo *et al.* (2016), and Heikkinen *et al.* (2012), the transferability index was used to measure the accuracy of the Egyptian model transferred to Saudi Arabia (AUC_{trans}), using the '*dismo*' package (R Developments Core Team, 2018) compared with its accuracy when predicting distributions was calibrated in Egypt (AUC_{inter}):

$$\text{Transferability index } T = \frac{AUC\_trans(Egypt)}{AUC\_inter(Egypt)}$$

An index value < 1 is an indication that the model performed better in its original site, and a value > 1 is an indication that the model performed better in the new landscape (Heikkinen *et al.*, 2012; Duque-Lazo *et al.*, 2016).

Additionally, I wanted to compare the predictions from the Egyptian transferred model with the accuracy of the Saudi model created using Saudi data. I created a novel index of relative transferability (to my knowledge), using a similar equation but with AUC_{inter} for the Saudi model in the denominator:

$$\text{Relative transferability index } U = \frac{AUC\_trans(Egypt)}{AUC\_inter(Saudi)}$$

A value of this index < 1 is an indication that the Saudi model performed better than the transferred Egyptian model for a species, while a value > 1 shows that the transferred Egyptian model performed better than the Saudi model.

3.2.8 Species richness maps

The ten replicate maps from Egyptian (transferring projection) and Saudi models (interpolation) for each species were reduced to a single average map by taking the arithmetic mean of the habitat suitability values for each cell, or by creating a consensus binary map (see below). I then combined species distribution predictions for each modelled species, creating two maps summarising the distribution of terrestrial reptile species richness: i) an unthresholded species-richness map, which shows the sum of the habitat suitabilities for all species in each grid cell; and ii) a thresholded species-richness map, which shows the predicted species richness based on the conversion of the continuous surface of habitat suitability for each species into a dichotomous surface (suitable = 1 and not suitable = 0). Both maps were created in ArcGIS (10.3.1) using the Raster Calculator and the Reclassify Tool. To create the unthresholded species-richness map, all the averaged ascii maps for each species were added together to produce a final map of predicted species richness. To create the consensus thresholded species-richness map from the 10 replicate maps for each species, the ten ascii thresholded replicates for each species were added together. A species was considered to be present in a pixel if more than five replicates of the model predicted its presence. All these consensus maps were then added together to create a final thresholded species-richness map.

3.3 Results

3.3.1 Model performance and variable contributions

The overall performance of the Maxent models of the 71 terrestrial reptiles in Egypt was excellent: 80% of species had mean test AUCs greater than 0.90. Altitude was the greatest contributor to the model and was the most significant predictor for 30 species, with

two temperature and two precipitation variables the most significant for some species. Solar radiation variables also contributed to explaining some species distributions.

3.3.2 Model transferability and accuracy

In general, the transferability and relative transferability indices suggest that transferred models performed well (Table 3.2). The transferability index showed that Egyptian models performed quite well when transferred, but significantly less well on average than they did for Egypt. The mean transferability index (T) was 0.71, significantly lower than 1 (one-sample t-test: $t_{24} = 11.1$, $p < 0.001$). For only one species (*Echis pyramidum*) (Table 3.2) did the Egyptian model perform better when transferred to Saudi Arabia than when predicting the Egyptian distribution.

The mean relative transferability index (U) was 0.79, again significantly lower than 1 ($t_{24}= 6.9$, $p < 0.001$). The individual values show that the Egyptian model performs better than the Saudi model for two species (*Cerastes cerastes* and *Pristurus flavipunctatus*) (Table 3.2). The transferability index was significantly lower than the relative transferability index (paired t-test, $t_{24} = 4.48$, $p < 0.001$).

There was no indication that the relative performance of the transferred model predictions was affected by the availability of occurrence data for Saudi reptiles. There was no significant correlation between the number of reptile occurrences in Saudi Arabia and the transferability index ($r = -0.287$, $n = 25$, $p > 0.05$; Figure S3.10) or the relative transferability index ($r = - 0.186$, $n = 25$, $p > 0.05$; Figure S3.11).

Additionally, there was no indication that the relative performance of the transferred model predictions was affected by the availability of occurrence data for the models upon which they depended. Thus there was no significant correlation between the transferability

index and the number of reptile occurrences in Egypt ($r = -0.002$, $n = 25$, $p > 0.05$; Figure S3.12), and also no significant correlation between the relative transferability index and the number of occurrences in Egypt ($r = 0.04$, $n = 25$, $p > 0.05$; Figure S3.13).

3.3.3 Predicted patterns of transferred species richness using Egyptian models

The unthresholded and thresholded species-richness maps generated for Saudi Arabia using all environmental predictors (Egyptian Model Set 1) show broadly similar distribution patterns (Figure 3.2) (for transferred prediction maps for individual species, see Figure S3.14). Areas of Saudi Arabia predicted to have suitable habitat for relatively many species (high species richness), including coastal areas, the Eastern Province, the Southwest of Jizan Province, and the West of Tabuk Province. Areas predicted to have relatively low richess include fragmented locations in the west of Saudi Arabia. Overall, when the model is transferred using the MESS-corrected subset of variables (Egyptian Model Set 3) (Figure 3.3), the spatial pattern of predicted species richness is lower across Saudi Arabia, but again highlights the same locations for relatively high diversity as with the full set of variables.

Considering only those species found in both countries, the unthresholded and thresholded species-richness maps (Egyptian Model Set 2) show similar patterns to those described above, again emphasizing that the central plateau and coastlines have high to medium predicted richness (Figure 3.4). When only the MESS-corrected subset of variables was used (Egyptian Model Set 4), the unthresholded and thresholded species richness predicts lower species richness, especially in the north, the west and the area between the central plateau and western coast, but the overall pattern is still broadly similar (Figure 3.5).

3.3.4 Egyptian models (transference) vs Saudi model (interpolation)

An important result from this study is that all transferred Egyptian models generated similar predictions for the spatial patterns in species richness to those produced by the independent Saudi Models 1 & 2 (Figure S3.15). The same locations are predicted to have high species richness: the coastlines, the central plateau, the southern and north western parts of Saudi Arabia, and a few other fragmented locations. The transferred predictions followed a similar pattern of to the observed species richness across Saudi Arabia, especially emphasizing the areas mentioned above, but these similarities were again less when the subset of variables was used.

3.4 Discussion

Transferring the predictions of species distribution models into a new poorly-studied habitat is a more complicated and challenging process than gap-filling within the existing range of a model (Peterson *et al.*, 2007; Barbosa *et al.*, 2009; Heikkinen *et al.*, 2012). However, the results presented here suggest that the process can yield effective models whose predictions are similar to those generated by models that use data from the target locality. In particular, the patterns of species richness predicted from transferred models were very similar to those from spatially independent models built using Saudi data.

3.4.1 Transference versus interpolation predictions

Most models for Egyptian reptiles showed excellent performance within their native territory. However, good performance in the training area does not necessarily guarantee good transferability to the new area. It is promising, therefore, that model transference using Egyptian models predicted similar distribution patterns to those predicted using an interpolation approach with a completely independent dataset (Saudi models).

These predicted patterns suggest that proximity to the sea and to major cities positively influence habitat suitability and hence reptile species richness. Cox *et al.* (2012) reported similar patterns based on analysis of range maps for reptiles over the Arabian Peninsula. They found greater terrestrial reptile richness around the coastline, and at a few scattered locations in the centre of Saudi Arabia. My finding is in-line with El-Gabbas *et al.* (2016) who modelled reptile richness and their habitat suitability in Egypt, and reported that areas around major cities and along some coasts tend to have higher predicted species richness and habitat suitability, especially around the Sinai Peninsula and along the Mediterranean Sea. Carranza *et al.* (2018) in Oman found areas with high reptile richness around the capital Muscat and its coast; having sampled thoroughly, they concluded that reptiles are not negatively affected by the presence of human settlements and may actually benefit.

These similar associations in spatial distribution from different reptile studies in similar habitat conditions may reflect the fact that human activity is associated with more food or otherwise suitable environmental condition, either because humans choose to live in areas which are productive or benign, or because some humans activities influence reptile distributions positively by acting as a source of suitable habitat or resources (e.g. farmland) (Carranza *et al.*, 2018). The combination of suitable temperatures most of the year and low to moderate altitude along the coasts (AbuZinada *et al.*, 2004; Vincent, 2008) - and perhaps close to human settlements - may provide perfect conditions for terrestrial reptiles in Saudi Arabia, including both specialist and generalist species: this may explain these interpolative and transference predicted patterns.

These recent findings from similar habitat conditions certainly imply that my transferred Egyptian models reliably captured important locations at the species-richness level across Saudi Arabia. Thus transference of the Egyptian model yielded very promising

results that could be applicable under Saudi conditions. This overall pattern may be an artefact of sampling bias, but I attempted to control for this using a bias file in both situations. Further validation (especially via ground-truthing) is needed to establish definitively if the predicted richness from the Egyptian models really are useful at Saudi locations.

3.4.2 Transference & accuracy of distribution modelling

Although these results suggest that transference is a useful method of predicting patterns of species richness, transferred models were on average less accurate than interpolated models. There are various factors suggested in the literature that could be responsible for limiting the transferability index to be < 1, along with proposed approaches to improve the accuracy of transferred models (Randin *et al.*, 2006; Duque-Lazo *et al.*, 2016; Werkowska *et al.*, 2017; Sequeira *et al.*, 2018; Yates *et al.*, 2018). Here, I discuss possible constraints that may limit the transferability of the Egyptian models.

Environmental dissimilarity and geographical distance between the two areas may limit model transferability. Environmental dissimilarity is considered one of the biggest constraints that needs to be checked prior to transference and extrapolation of an SDM (Elith *et al.*, 2010; Werkowska *et al.*, 2017; Yates *et al.*, 2018). Transferred species distribution studies by Randin *et al.* (2006), Barbosa *et al.* (2009), and Duque-Lazo *et al.* (2016) found that environmental dissimilarity and variation in the environmental range could affect transferability into new spaces. Here the reptiles in each landscape experience both direct and indirect influences of environmental variables (e.g. day-length, wind patterns, and solar radiation) on their ecology and physiology which could affect transferability of the model. The environmental variables and ranges used to transfer the Egyptian models are listed in Table 3.1. Generally the two landscapes have similar environmental and habitat conditions

that make them highly suitable for transferring such models. Furthermore, the magnitude of geographical disjunction seen in my study is not thought to have a strong influence on transferability success (see Yates *et al.*, 2018), and since the two areas used here show similar environmental ranges, it is possible to have an accurate transference into the new space (Werkowska *et al.*, 2017).

Another potential limitation on model transferability could be the quality and the quantity of the biological data (Yates *et al.*, 2018). Occurrence records that are not accurate enough (e.g. low in quality and quantity) are known to affect SDMs negatively because they may not represent the reality of the species-environment associations (Wisz *et al.*, 2008), and hence may lead to lowered accuracy in transference (Yates *et al.*, 2018). In this study, occurrences records from Egypt and Saudi Arabia may suffer from such problems since most of them come from museum records and opportunistic observations. However, having high-quality data to transfer and extrapolate an SDM does not always guarantee successful transference in the new area (see Barbosa *et al.* 2009). Having a comprehensive coverage of occurrence records is typically rare for many regions (Tsoar *et al.*, 2007; Sequeira *et al.*, 2018; Yates *et al*., 2018). This is one of the reasons why transference of SDMs is needed under some circumstances (Barbosa *et al.*, 2009; Yates *et al.*, 2018), and why it has become an important scientific application for use in many ecological and biogeography disciplines.

A possible factor that could limit transferability is the spatial scale/extent (Randin *et al.*, 2006) of study areas, which then lead each study area to have different and distinctive geographical features. Saudi Arabia is twice the size of Egypt and probably encompasses wider environmental and geographical variations than Egypt because of the difference in latitudinal and longitudinal range. Even though the dominant habitats in both are open arid deserts, each country still has geographically distinctive features. The south-eastern part of Saudi Arabia is dominated by the huge sand dunes of the Empty Quarter. South-western

Saudi Arabia has long, richly vegetated mountain chains (up to 2900 m) (AbuZinada *et al.*, 2004; Vincent, 2008), with some areas receiving approximately 200 mm of rainfall (Gosling *et al.*, 2011): such conditions are not present in Egypt. In contrast, the unique presence in Egypt of the Nile River and its surrounding delta and valley influences the spatial distribution of reptiles (El-Gabbas *et al.*, 2016). These geophysical and geographical features in each country probably influence local reptiles in terms of adaptation to new environmental conditions, which in turn alters their distribution and habitat selection.

Recently, grain size has been suggested to have an influence on transferability across regions. Manzoor *et al.* (2018) tested three different grain sizes and their effect on transferability of plant SDMs in two separate national parks in Wales (UK). They found that transferability was improved by using medium grain size (300 m), while models using the grain size normally used for bioclimatic variables (1 km^2) had the worst transferability. However, Guisan *et al.* (2007) showed that changing grain size does not greatly influence the results of species distribution models, where sample size appears to have more impact on model accuracy. Terrestrial reptiles are not highly mobile and hence should be well connected to their local environment, making the chosen grain size suitable for our purposes. However, it is certainly possible that the transference procedure had such grain-size issues, and the predicted models may have had better performance for some species if smaller, or perhaps even larger grain sizes had been used.

Model complexity (e.g. the number and the type of environmental variables) could also have an influence on transferability. The effect of model complexity on SDM and its transference are still an active area of research (Verbruggen *et al.*, 2013; Moreno-Amat *et al.*, 2015; Duque-Lazo *et al.*, 2016; Werkowska *et al.*, 2017; Yates *et al.*, 2018). Model complexity should be considered along with species characteristics and their response to environmental predictors, because using resource and 'direct' variables is known to improve

SDMs (Austin, 2002) and hence transferability (Randin *et al.*, 2006; Franklin, 2009). Moreno-Amat *et al.* (2015) tested model complexity and its effect on transferability using Maxent, and found that intermediate complexity gave better performance in temporal transference, while more complex models had a negative impact on transferability. Here, I used two sets of predictors to produce the transferred species richness maps across Saudi Arabia, which were filtered via similarity analyses to ensure their validity in transferring the model between the two studied areas. The results of models using the two predictor sets were broadly similar, but absolute species richness was typically higher when using the full set.

There are other constraints that could implicitly affect the overall transferability of the Egyptian models, such as biotic/abiotic interactions, environmental adaptation, anthropogenic pressures, niche shifts and species traits (Randin *et al.*, 2006; Yates *et al.*, 2018). So far, these have not been studied much in research on transference in SDMs (Werkowska *et al.*, 2017). Frankly, it is impossibly complicated to account for all possible constraints that influence transferability (Randin *et al.*, 2006; Barbosa *et al.*, 2009), and it is difficult to know whether there might have been such hidden impacts on my Egyptian models without extensive and systematic field surveys conducted specifically to answer such issues.

It consumes both time and resources to have 'perfect' data (Tsoar *et al.*, 2007; Sequeira *et al.*, 2018) that satisfy all required conditions for transferring and extrapolating an SDM between two disjunctive areas or between different times. More detailed data and investigations are needed specifically on how to overcome difficulties known to limit transferability (Werkowska *et al.*, 2017; Yates *et al.*, 2018).

3.4.3 Transferability and relative transferability indices vary among studied species

Indices of transferability varied among the 25 species in common (Table 3.2). Whilst transferred models performed less well for most species, models for the Egyptian Saw-scaled

Viper (*E. pyramidum*), the Desert Horned Viper (*C. cerastes*), and Middle Eastern Rock Gecko (*P. flavipunctatus*) performed better in Saudi Arabia using the transferred Egyptian model than in at least one of the other scenarios considered (interpolation in Egypt or Saudi Arabia). Most of the occurrence records in Saudi Arabia and Egypt come from museums and relevant literature in presence-only format, which may partly explain this variation. In other words, some species have received more attention because of either researcher preferences or because they are easy to detect, while other species are only opportunistically detected (Phillips *et al.*, 2009; Newbold, 2010). Hence species can be recorded in habitats that are actually not suitable, or recorded as absent from suitable habitats and conditions (Hirzel *et al.*, 2002; Guisan and Thuiller, 2005).

Variation among modelled species is expected due to the various ecological and environmental obstacles mentioned above. In correlative SDM studies, models of narrow-range species tend to be more accurate than those for wider-range species (Tsoar *et al.*, 2007; Newbold *et al.*, 2010). Dispersal can also influence transferability (Yates *et al.*, 2018). Heikkinen *et al.* (2012) found that SDMs for birds have higher transferability than butterflies and plants, while Roach *et al.* (2017) reported poor spatial transferability for widespread bird species. Here, *C. cerastes* is a species with a wide distribution range (Baha El Din, 2006), and *P. flavipunctatus* and *E. pyramidum* are habitat specialists with relatively narrow ranges (Baha El Din, 2006). That being said, my transferred Egyptian model was able to capture good predictions regardless of species dispersal ability and that may be considered as a sign of the strength of the transferred Egyptian models. Detailed ecological knowledge about reptiles in the Middle East is still developing, and modellers need such knowledge in order to interpret the variation in transferability among species.

There is no obvious relationship between sample size and either of the transferability indices. Species with relatively low to medium sample sizes tended to have the highest

indices, although there was no statistical association. Poorly sampled species may be more likely to have overfitted models, but overfitting is typically expected to reduce transferability (Randin *et al.,* 2006). I believe overfitting is not a concern with my transferred models, since I reduced spatial autocorrelation using the bias file, deletion of unreliable spatial records and MESS analysis. In addition, Maxent is less prone to overfitting than other methods due to its regularization procedure, especially when modelling species with small sample sizes (Baldwin, 2009; Phillips *et al.*, 2006; Wisz *et al.*, 2008; Aguirre-Gutierrez *et al.*, 2013; Merow *et al.,* 2013).

3.5 Conclusion

Effort is unevenly distributed in the study of species richness and distribution in most regions (Tsoar *et al*., 2007; Sequeira *et al.*, 2018), and some locations probably have greater species richness and uniqueness than is currently known (e.g. the southwest of the Arabian Peninsula: Ficetola *et al.,* 2013). There is an obvious deficiency of data for the reptile fauna in this part of the world, a clear indication for the need for more fieldwork, even in places that have been surveyed before (Cox *et al.,* 2012). Species distribution modelling and its transference (either in space and time) represent a solution for situations where biological data are very scarce and fewer studies have been conducted, especially areas that are remote and inaccessible (Tsoar *et al.,* 2007; Barbosa *et al.*, 2009; Ficetola *et al.*, 2013; Yates *et al.*, 2018).

The transferred Egyptian models give useful ecological insights about the distribution of species richness of terrestrial reptiles in Saudi Arabia. The results facilitate new fieldwork by indicating locations where species richness is expected to be high but records are scarce, and provide evidence for the accuracy of SDM predictions transferred into new spatial conditions. My study contributes to the developing science of spatial and temporal

transference; more accurate applications of this approach have great promise in helping us to understand species distributions in understudied regions in the future.

3.6 References

AbuZinada, A., Robinson, E. R., Nader, I. & Al Wetaid, Y. 2004. *First Saudi Arabian National Report on the Convention on Biological Diversity*, The National Commission for Wildlife Conservation and Development.

Aguirre-Gutierrez, J., Carvalheiro, L. G., Polce, C., van Loon, E. E., Raes, N., Reemer, M. & Biesmeijer, J. 2013. Fit-for-purpose: species distribution model performance depends on evaluation criteria - Dutch Hoverflies as a case study. *PLoS ONE*, 8**,** e63708.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157**,** 101-118.

Baha El Din, S. 2006. *A Guide to the Reptiles and Amphibians of Egypt*, The American University in Cairo Press.

Baldwin, R. 2009. Use of maximum entropy modeling in wildlife research. *Entropy*, 11**,** 854-866.

Barbosa, A. M., Real, R. & Vargas, J. M. 2009. Transferability of environmental favourability models in geographic space: The case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. *Ecological Modelling*, 220**,** 747-754.

Brown, J. L., Bennett, J. R. & French, C. M. 2017. SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ*, 5**,** e4095.

Carranza, S., Xipell, M., Tarroso, P., Gardner, A., Arnold, E. N., Robinson, M. D., et al. 2018. Diversity, distribution and conservation of the terrestrial reptiles of Oman (Sauropsida, Squamata). *PLoS ONE*, 13**,** e0190389.

Cox, N. A., Mallon, D., Bowles, P., Els, J. & Tognelli, M. F. 2012. *The Conservation Status and Distribution of Reptiles of the Arabian Peninsula*. Cambridge, UK and Gland, Switzerland: IUCN, and Sharjah, UAE: Environment and Protected Areas Authority.

Duque-Lazo, J., van Gils, H., Groen, T. A. & Navarro-Cerrillo, R. M. 2016. Transferability of species distribution models: The case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. *Ecological Modelling,* 320**,** 62-70.

El-Gabbas, A., Baha El Din, S., Zalat, S. & Gilbert, F. 2016. Conserving Egypt's reptiles under climate change. *Journal of Arid Environments,* 127**,** 211-221.

Elith, J., Kearney, M. & Phillips, S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution,* 1**,** 330-342.

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.

Ficetola, G. F., Bonardi, A., Sindaco, R. & Padoa-Schioppa, E. 2013. Estimating patterns of reptile biodiversity in remote regions. *Journal of Biogeography,* 40**,** 1202-1211.

Fick, S. E. & Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology,* 37**,** 4302-4315.

Fielding, A. H. & Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation,* 24**,** 38-49.

Franklin, J. 2009. *Mapping Species Distributions: Spatial Inference and Prediction*, Cambridge University Press, Cambridge, UK.

Gosling, S. N., Dunn, R., Carrol, F., Christidis, N., Fullwood, J., et al. 2011. *Climate: Observations, projections and impacts: Saudi Arabia*, The Met Office, UK.

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 & Evolution,* 19**,** 497-503.

Guisan, A., Graham, C. H., Elith, J., Huettmann, F. & the NCEAS Species Distribution Modelling Group. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions*, 13**,** 332-340.

Guisan, A. & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8**,** 993-1009.

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

Heikkinen, R. K., Marmion, M. & Luoto, M. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography*, 35**,** 276-288.

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.

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

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

Manzoor, S. A., Griffiths, G. & Lukac, M. 2018. Species distribution model transferability and model grain size – finer may not always be better. *Scientific Reports*, 8**,** 7168.

Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., McMahon, S. M., et al. 2014. What do we gain from simplicity versus complexity in species distribution models? *Ecography*, 37**,** 1267-1281.

Merow, C., Smith, M. J. & Silander, J. A. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36**,** 1058-1069.

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.

Naimi, B., Hamm, N., Groen, T. A., Skidmore, A. K. & Toxopeus, A. G. 2014. Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37**,** 191-203.

Newbold, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, 34**,** 3-22.

Newbold, T., Reader, T., El-Gabbas, A., Berg, W., Shohdi, W. M., Zalat, S., Baha El Din, S. & Gilbert, F. 2010. Testing the accuracy of species distribution models using species records from a new field survey. *Oikos*, 119**,** 1326-1334.

Pearce, J. & Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133**,** 225-245.

Pearson, R. G. 2008. Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*. Available at http://ncep.amnh.org.

Peterson, A. T., Papeş, M. & Eaton, M. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, 30**,** 550-560.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M. & Araújo, M. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press.

Phillips, S. J. 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). *Ecography*, 31**,** 272-278.

Phillips, S. J. 2017. A Brief Tutorial on Maxent. Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. 2017. Opening the black box: an open-source release of Maxent. *Ecography*, 40**,** 887-893.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190**,** 231-259.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. & Ferrier, S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19**,** 181-197.

R Development Core Team, 2018. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.

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.

Roach, N. S., Hunter, E. A., Nibbelink, N. P. & Barrett, K. 2017. Poor transferability of a distribution model for a widespread coastal marsh bird in the southeastern United States. *Ecosphere*, 8**,** e01715.

Sequeira, A. 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.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, 13**,** 397-405.

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.

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

Verbruggen, H., Tyberghein, L., Belton, G. S., Mineur, F., Jueterbock, A., Hoarau, G., Gurgel, F. D. & Clerck, O. 2013. Improving transferability of introduced species' distribution models: New tools to forecast the spread of a highly invasive Seaweed. *PLoS ONE*, 8**,** e68337.

Vincent, P. 2008. *Saudi Arabia: An Environmental Overview*. Taylor & Francis Group.

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. 2017. A practical overview of transferability in species distribution modeling. *Environmental Reviews*, 25**,** 127-133.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A. & NCEAS Predicting Species Distributions Working Group. 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., et al. 2018. Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33**,** 790-802.

Zanini, F., Pellet, J. & Schmidt, B. R. 2009. The transferability of distribution models across regions: an amphibian case study. *Diversity and Distributions*, 15**,** 469-480.

Table 3.1: Bioclimatic variables with a Variance Inflation Factor (VIF) of less than 10 that were used to build the reptile distribution models using Maxent. The third and the fourth column represent the range of enviromental variables used in the analysis. Variables which are higlighted in grey are those which were used to rerun the model to minimize the effect of environmental conditions that do not occurr in Egypt.

		Range of values	
Abbrev	Name	Egypt	Saudi Arabia
Alt	Altitude	-131 − +2494 m	-16 − +2961 m
Bio_02	Mean Diurnal Range (Mean of monthly (max temp - min temp))	7.00 − 19.19 °C	7.64 − 17.54 °C
Bio_03	Isothermality (BIO2/BIO7) (* 100)	33.70 − 57.05 %	32.44 − 59.45 %
Bio_08	Mean Temperature of Wettest Quarter	4.48 − 34.36 °C	5.61 − 33.10 °C
Bio_09	Mean Temperature of Driest Quarter	10.85 − 34.18 °C	10.05 − 38.65 °C
Bio_14	Precipitation of Driest Month	0 – 2 mm	0 – 12 mm
Bio_15	Precipitation Seasonality (Coefficient of Variation)	0 − 154.3 %	26.9 − 144.8 %
Bio_19	Precipitation of Coldest Quarter	0 – 156 mm	2 – 109 mm
Srad_04	Solar radiation in April	21596−25137 kJ m^{-2} day^{-1}	19559−24548 kJ m^{-2} day^{-1}
Srad_06	Solar radiation in June	25064−28299 kJ m^{-2} day^{-1}	20044−27563 kJ m^{-2} day^{-1}
Srad_09	Solar radiation in Sept	20540−23675 kJ m^{-2} day^{-1}	19605−22834 kJ m^{-2} day^{-1}

Table 3.2: The value of the transferability index generated by the Egyptian model and the relative transferability index generated by the Saudi model for the 25 species in common. AUC-inter(Egypt) accuracy calibrates the model in Egypt; AUC-inter(Saudi) accuracy calibrates for Saudi; AUC-trans calibrates the transfer from Egypt to Saudi. Calculation of both transferability indices are explained in the main text.

Species	AUC-inter (Egypt)	AUC-inter (Saudi)	AUC-trans	Trans-index	Relative trans-index
Acanthodactylus boskianus	0.878	0.753	0.674182	0.767861	0.895328
Acanthodactylus scutellatus	0.848	0.8	0.583922	0.688587	0.729903
Cerastes cerastes	0.82	0.684	0.71572	0.872829	1.046374
Cerastes vipera	0.946	0.991	0.676971	0.715614	0.683119
Chalcides ocellatus	0.937	0.865	0.651173	0.694955	0.752801
Chamaeleo chamaeleon	0.977	0.881	0.877613	0.898274	0.996156
Cyrtopodion scabrum	0.978	0.864	0.758142	0.775196	0.877479
Echis coloratus	0.91	0.865	0.70127	0.770626	0.810717
Echis pyramidum	0.843	0.989	0.90262	1.070724	0.912659
Eumeces schneiderii	0.939	0.941	0.677562	0.721578	0.720045
Hemidactylus turcicus	0.935	0.829	0.607173	0.649383	0.732417
Laudakia stellio	0.979	0.908	0.669306	0.683663	0.737121
Lytorhynchus diadema	0.898	0.818	0.618622	0.688889	0.756262
Malpolon moilensis	0.945	0.866	0.660825	0.699286	0.763077
Mesalina guttulata	0.787	0.707	0.531138	0.674889	0.751256
Pristurus flavipunctatus	0.991	0.742	0.946733	0.955331	1.275921
Psammophis schokari	0.96	0.834	0.553069	0.576113	0.663152
Pseudotrapelus sinaitus	0.901	0.82	0.570593	0.633289	0.695845
Ptyodactylus hasselquistii	0.912	0.798	0.582414	0.638612	0.729842
Scincus scincus	0.847	0.686	0.553788	0.653823	0.807271
Spalerosophis diadema	0.923	0.84	0.619236	0.670895	0.737186
Telescopus dhara	0.896	0.847	0.522106	0.582707	0.616417
Trapelus mutabilis	0.933	0.726	0.63446	0.680021	0.873912
Uromastyx aegyptia	0.948	0.737	0.422311	0.445476	0.573013
Varanus griseus	0.903	0.895	0.539749	0.597729	0.603072

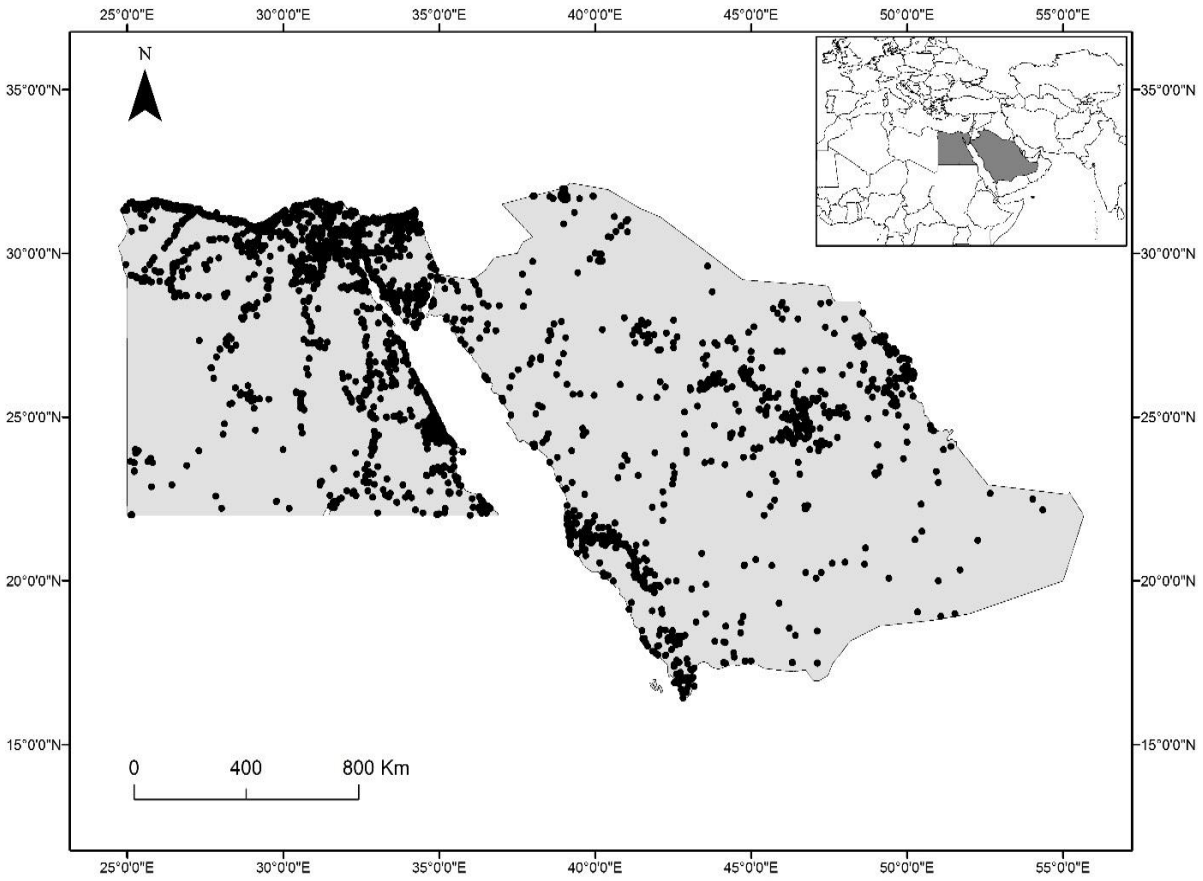


Figure 3.1: The study areas of Egypt and Saudi Arabia, along with 11,103 (71 species) and 3,943 (62 species) recorded occurrences of terrestrial reptiles in Egypt and in Saudi, respectively. The occurrence records in Egypt were used to develop distribution models that were then transferred into Saudi Arabia, while the occurrence records in Saudi Arabia were used to build independent Saudi models.

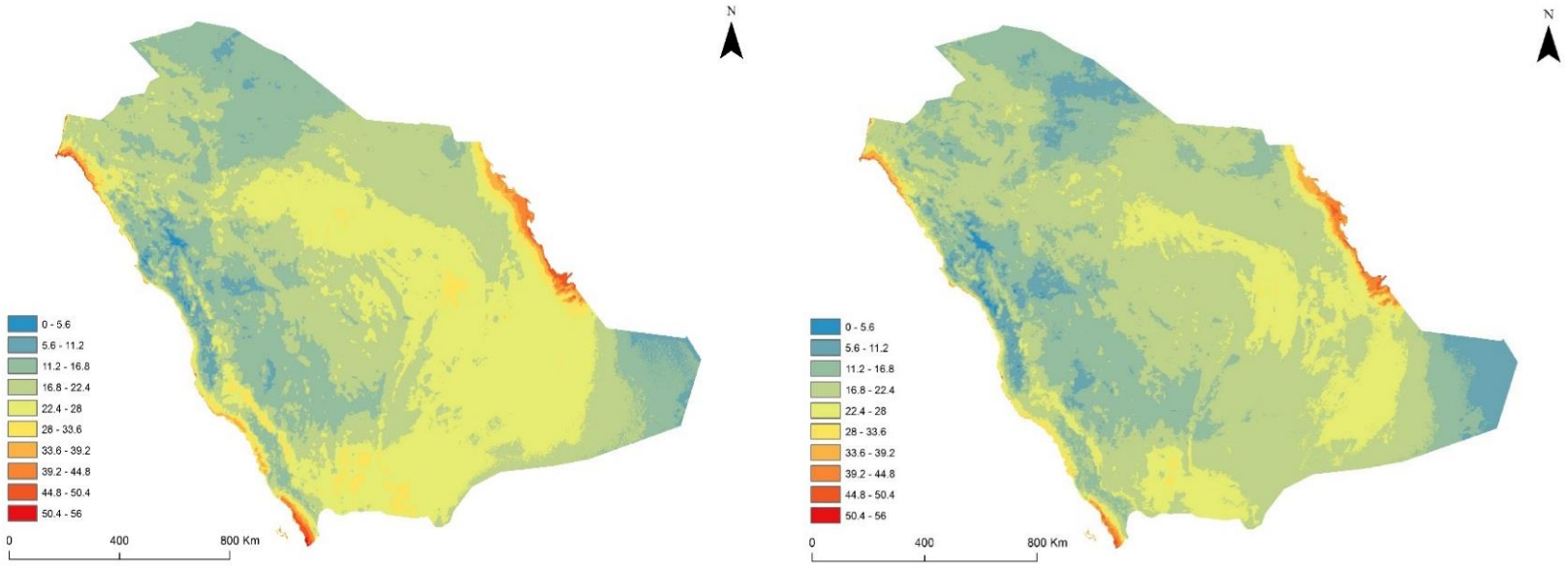


Figure 3.2: On the left, an unthresholded species richness map generated using the full set of environmental predictor variables for 71 reptile species (Egyptian Model Set 1). The predicted species richness was the sum of the predicted habitat suitability for each species in a given grid cell. On the right, a thresholded species richness map generated using the full set of environmental predictor variables for 71 species. The predicted species richness in each grid cell was the number of species for which suitability was predicted by more than five out of ten replicate model runs. The two maps were rescaled to match the thresholded Egyptian Model Set 3 for comparison purposes. Dark red indicates high predicted species richness.

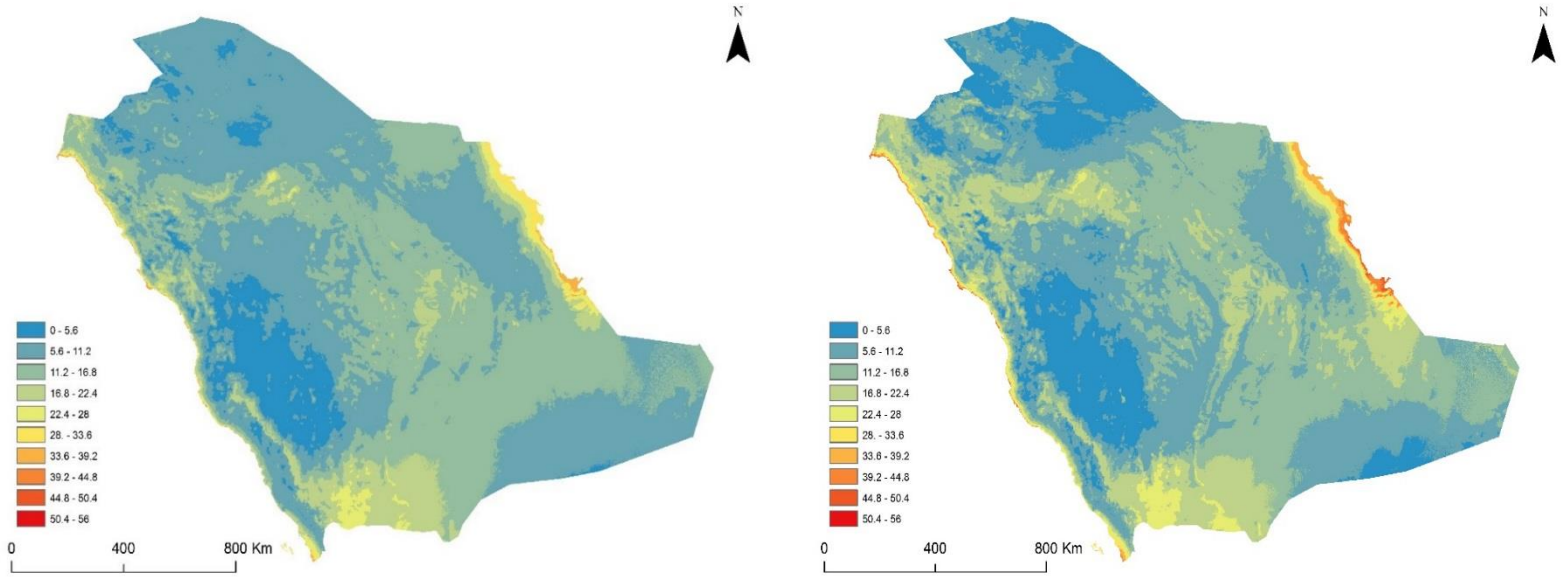


Figure 3.3: On the left, an unthresholded species richness map generated using a sub-set of variables for 71 species (Egyptian Model Set 3). In this case, predicted species richness was the sum of the predicted habitat suitability for each species in a given grid cell. The map was rescaled to match the thresholded version for comparison purposes. On the right, a thresholded species richness map generated using a sub-set of environmental predictor variables for 71 species. In this case, predicted species richness in each grid cell was the number of species for which suitability was predicted by more than five out of ten replicate model runs. Dark red indicates high predicted species richness.

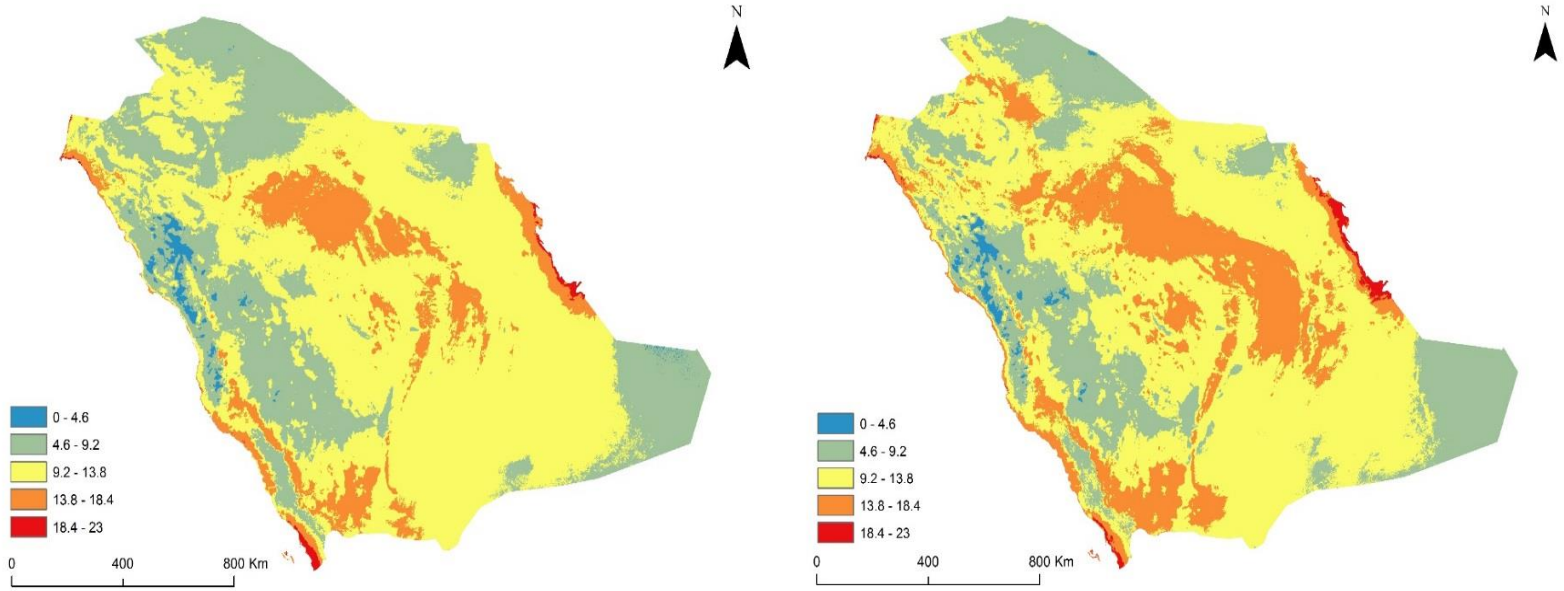


Figure 3.4: On the left, an unthreshold species richness map generated using the full set of environmental predictor variables for the 25 reptile species in common between Egypt and Saudi Arabia (Egyptian Model Set 2). In this case, predicted species richness was the sum of the predicted habitat suitability for each species in a given grid cell. The map was rescaled to match the thresholded version for comparison purposes. On the left, a thresholded species richness generated in the same way. In this case, predicted species richness in each grid cell was the number of species for which suitability was predicted by more than five out of ten replicate model runs. Dark red indicates high predicted species richness.

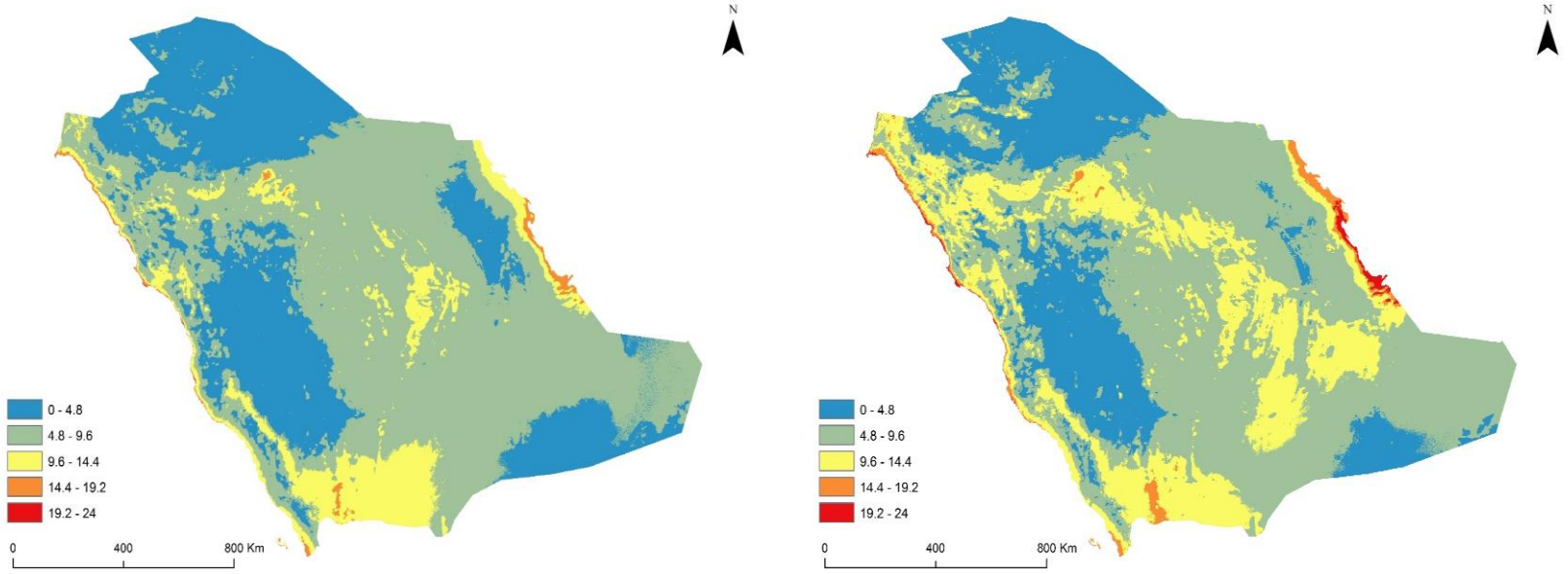


Figure 3.5: On the left, an unthresholded species richness map generated using a sub-set of environmental predictor variables for the 25 species in common between Egypt and Saudi Arabia (Egyptian Model Set 4). In this case, predicted species richness was the sum of the predicted habitat suitability for each species in a given grid cell. The map was rescaled to match the thresholded version for comparison purposes. On the right, a thresholded species richness map generated similarly. In this case, predicted species richness in each grid cell was the number of species for which suitability was predicted by more than five out of ten replicate model runs. Dark red indicates high predicted species richness.

3.7 Supplementary materials

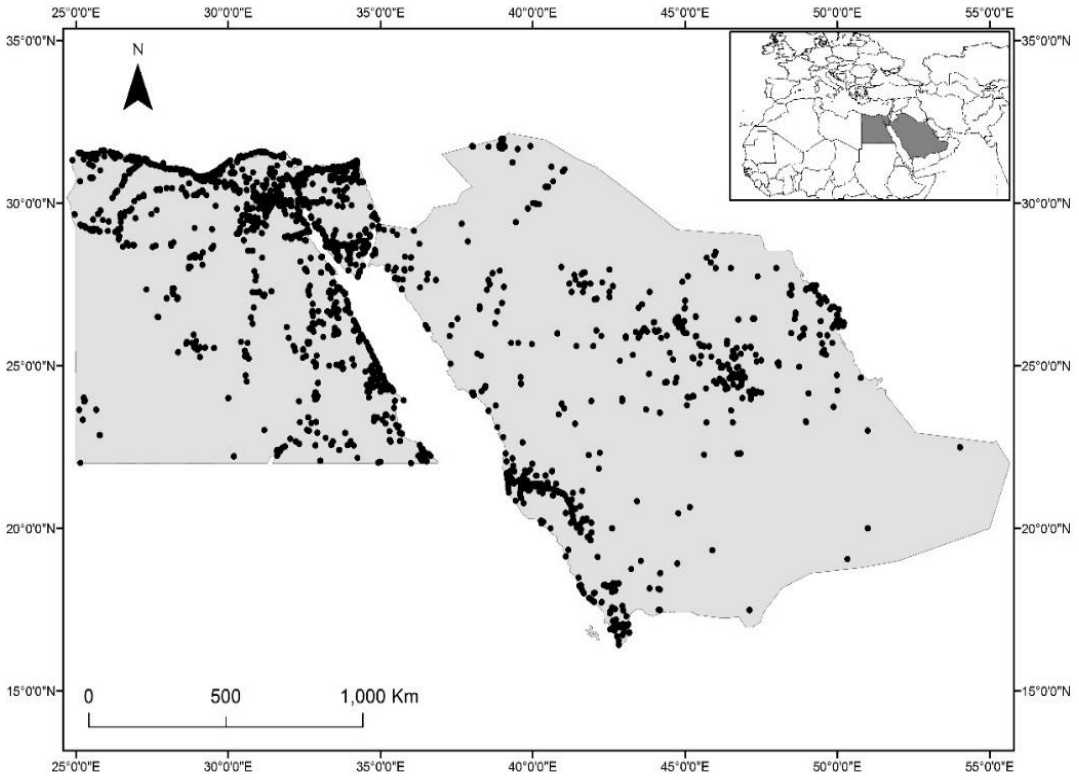


Figure S3.6: The study areas of Egypt and Saudi Arabia, along with 6,203 and 1,791 black dots representing the occurrence data for the 25 terrestrial reptile species in common between Egypt and Saudi Arabia, respectively. The occurrence records in Egypt were used to develop distribution models that were then transferred into Saudi Arabia, while the occurrences records in Saudi Arabia were used to build independent Saudi models.

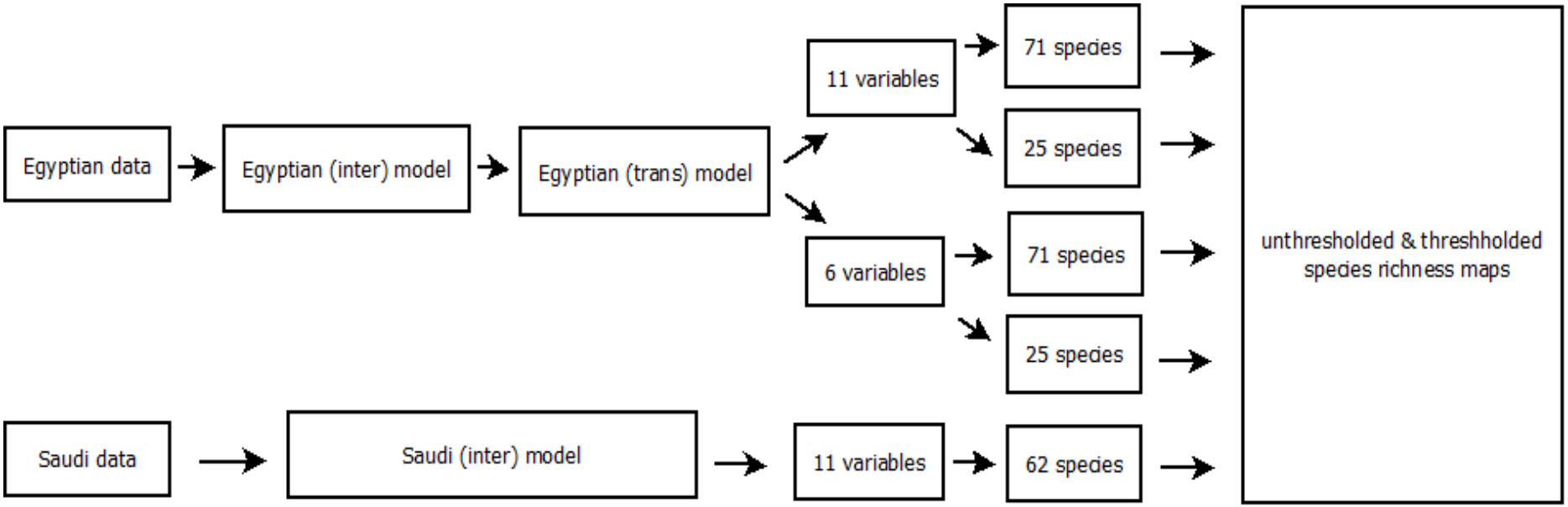


Figure S3.7: Flow-chart representing the process by which models of reptile distributions were created for Saudi Arabia, based on either Egyptian or Saudi data. Each Egyptian model uses either the full set of 11 environmental variables, or a reduced set of 6 that minimize the extrapolation to environmental conditions that do not occur in Egypt. Each model has two output maps i) probability species richness (i.e. un-thresholded richness), produced by adding together the raw output of Maxent across all species; and ii) binary species richness (thresholded richness), produced by converting the raw output of Maxent for each species into a binary map (each pixel is either suitable or unsuitable) using a threshold, and then summing over all species. All the final transferred maps from the Egyptian models are compared with the Saudi models (i.e. unthresholded & thresholded richness maps) to test the accuracy of transferability predictions of the Egyptian models into Saudi Arabia.

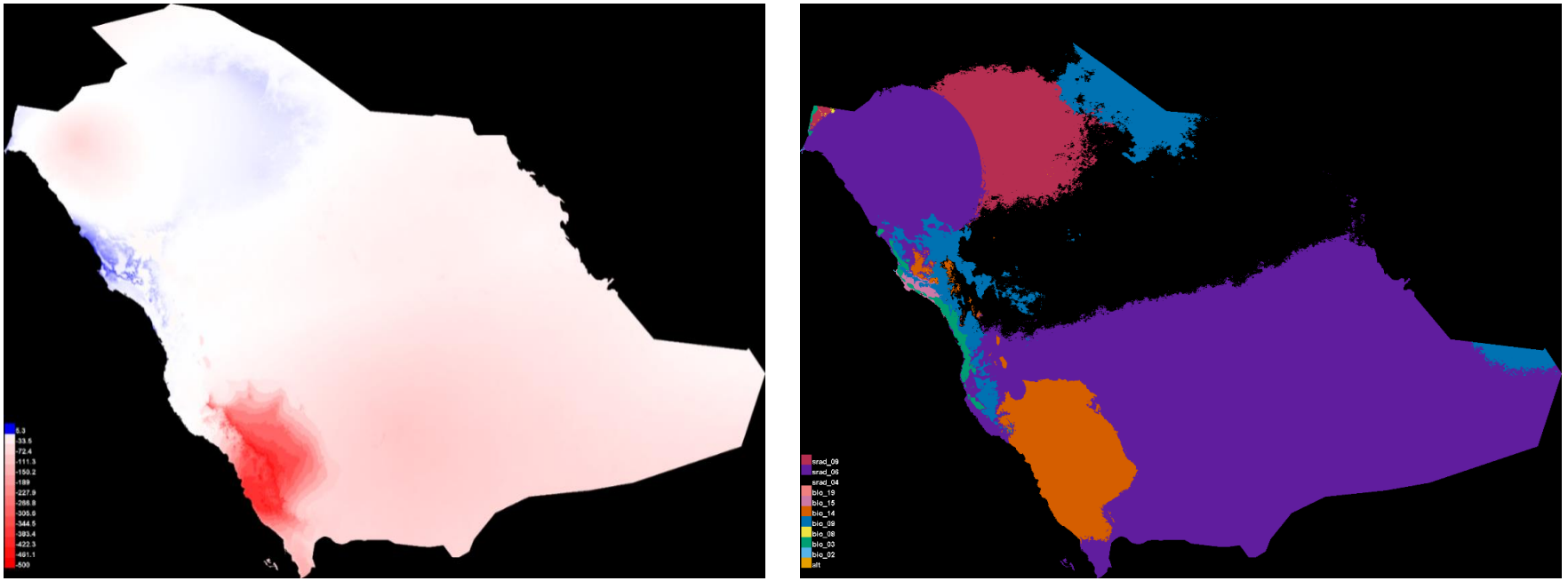


Figure S3.8: On the left, multivariate environmental similarity surface analysis (MESS) generated using Maxent. The red areas indicate novel climate conditions in Saudi Arabia that are not present in the data from Egypt used to develop the model, whilst blue areas indicate areas with similar conditions and white areas indicate neutral areas. This model used the full set of variables (Egyptian Models Sets 1 & 2). On the right, the 'most dissimilar variable' (MoD) map generated using Maxent, which shows environmental variables that are novel in the new landscape at each point.

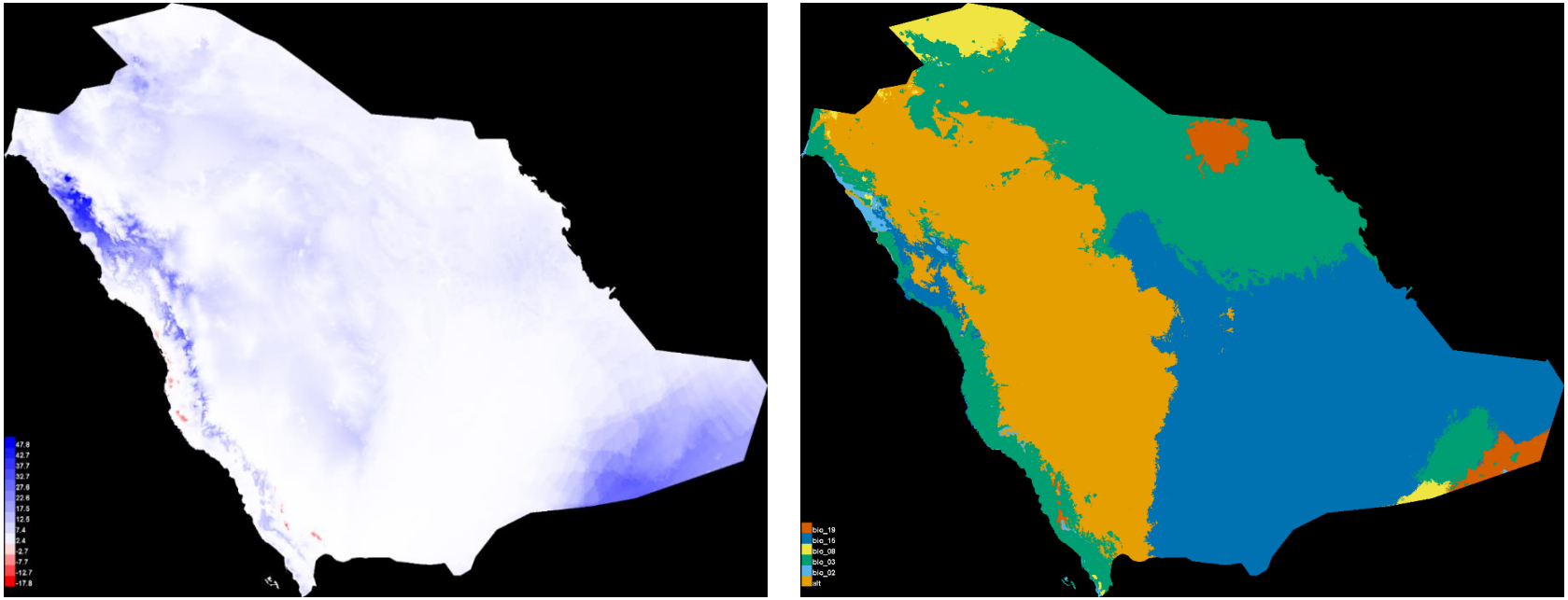


Figure S3.9: On the left, multivariate environmental similarity surface analysis (MESS) generated using Maxent after reducing the predictor environmental variables (Egyptian Model Sets 3 & 4) to minimize the occurrence of environmental conditions that do not occur in Egypt (i.e. extrapolating beyond the data). The red areas indicate novel climate conditions in Saudi Arabia that are not present in the data from Egypt used to develop the model, whilst blue areas indicate areas with similar conditions and white areas indicate neutral areas. On the right, the 'most dis-similar variable' (MoD) map generating using Maxent, which shows environmental variables that are novel in the new landscape at each point.

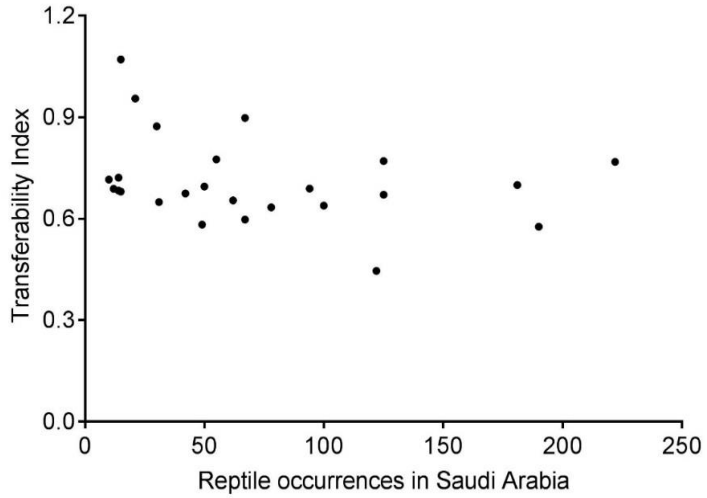


Figure S3.10: Correlation between the number of reptile records in Saudi Arabia and the transferability index ($r = -0.287$, $p > 0.05$) for the 25 species common to both countries.

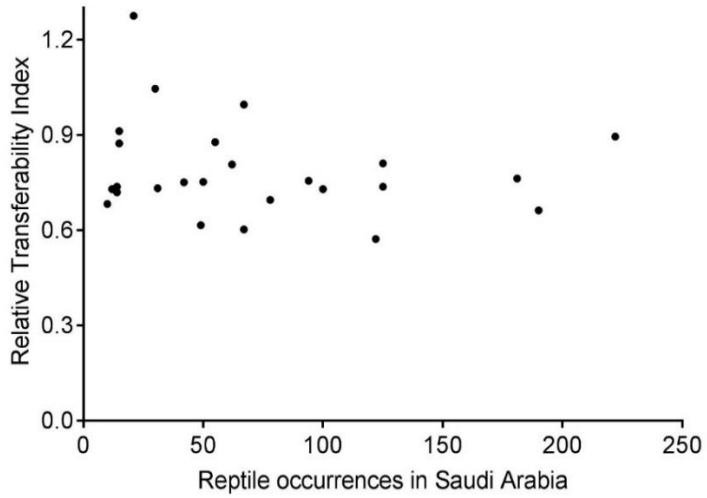


Figure S3.11: Correlation between the number of reptile records in Saudi Arabia and the relative transferability index ($r = -0.186$, $p > 0.05$) for the 25 species common to both countries.

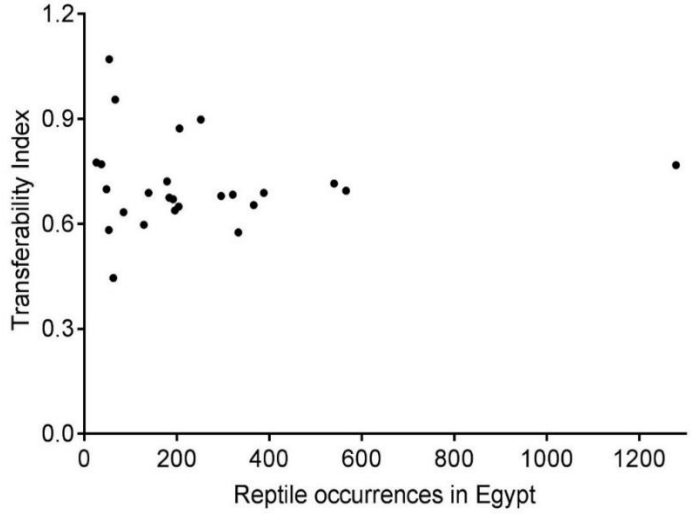


Figure S3.12: Correlation between the number of reptile records in Egypt and the transferability index ($r = -0.002$, $p > 0.05$) for the 25 species common to both countries.

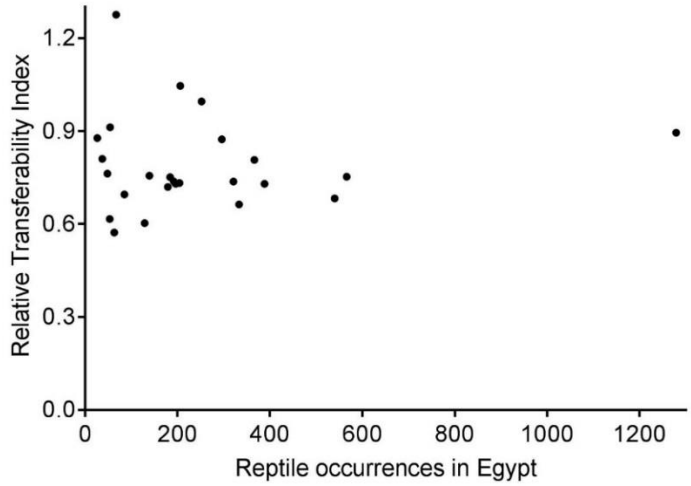
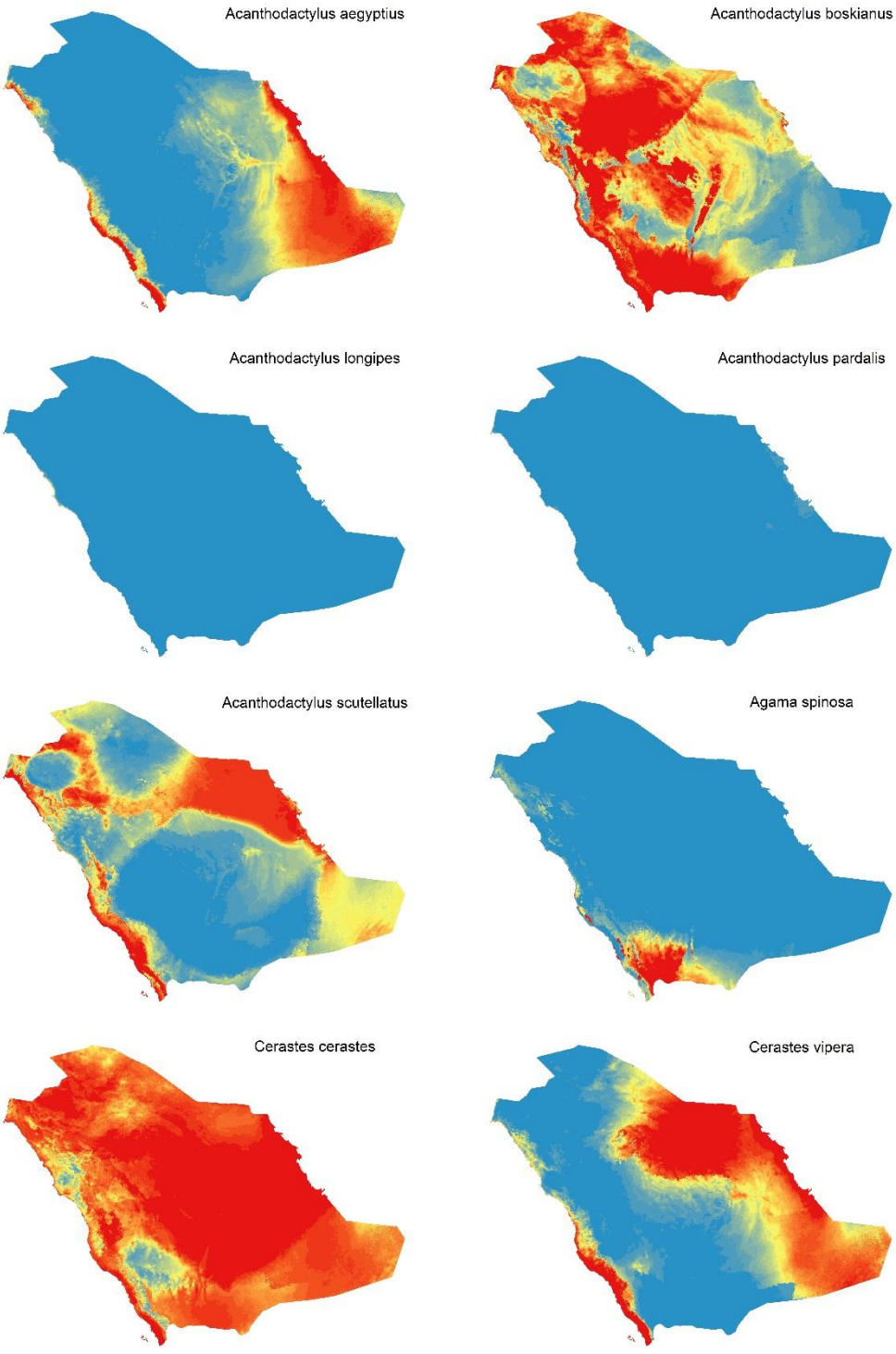
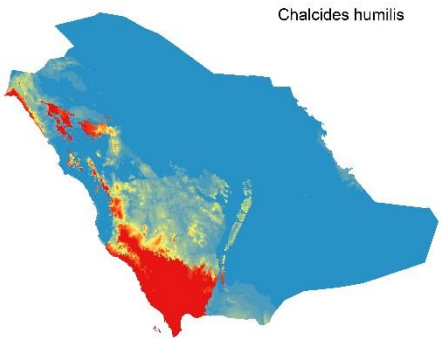
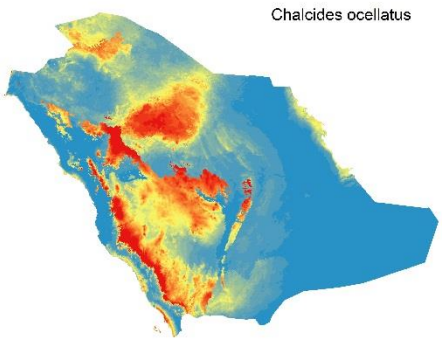
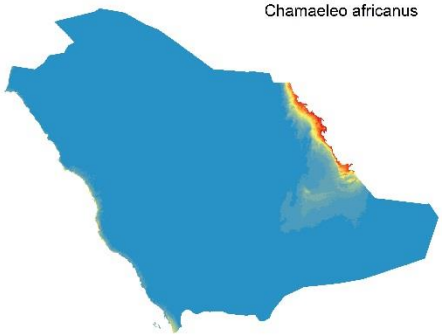


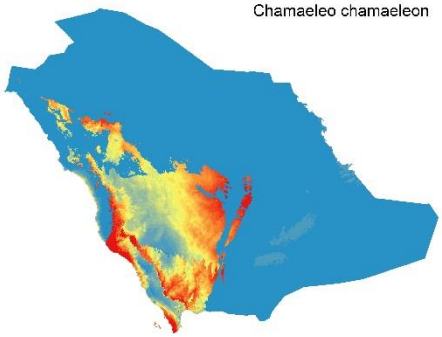
Figure S3.13: Correlation between the number of reptile records in Egypt and the relative transferability index ($r = 0.04$, $p > 0.05$) for the 25 species common to both countries.


Acanthodactylus aegyptius
Acanthodactylus boskianus
Acanthodactylus longipes
Acanthodactylus pardalis
Acanthodactylus scutellatus
Agama spinosa
Cerastes cerastes
Cerastes vipera

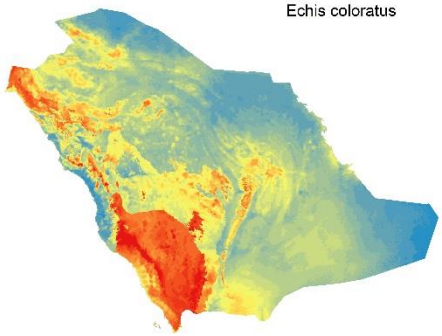

Chalcides humilis

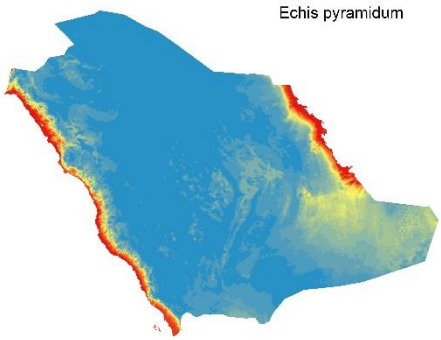

Chalcides ocellatus

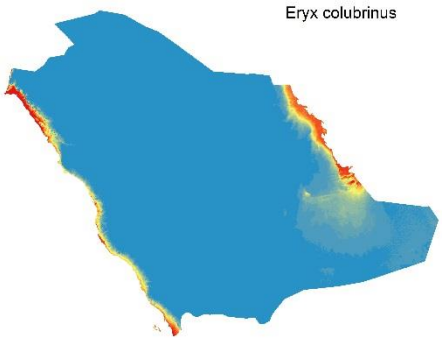

Chamaeleo africanus

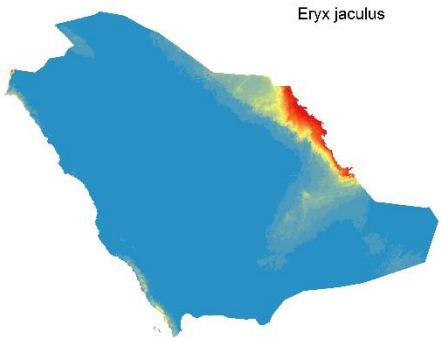

Chamaeleo chamaeleon

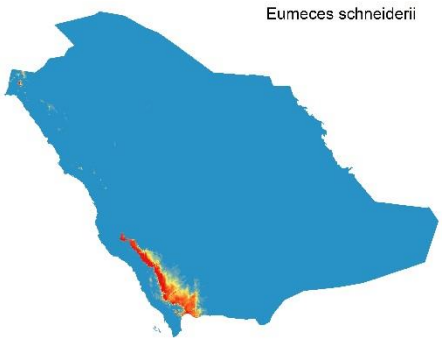

Cyrtopodion scabrum

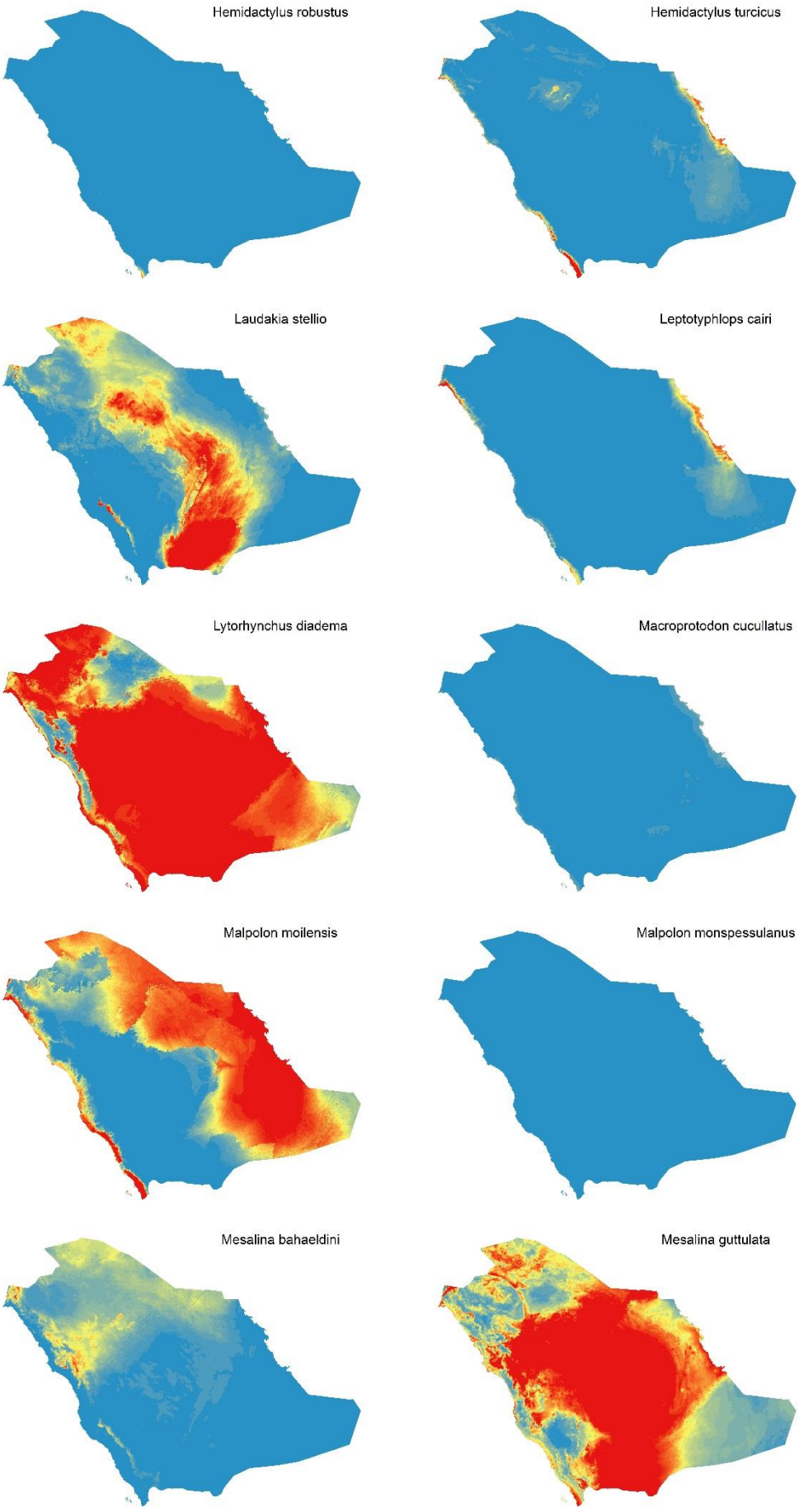

Echis coloratus

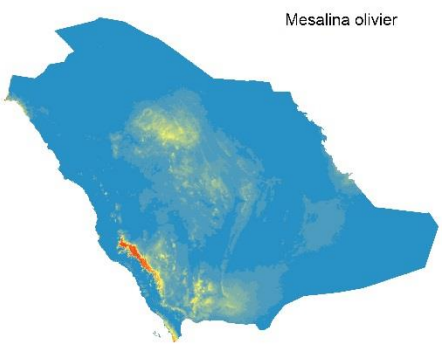

Echis pyramidum

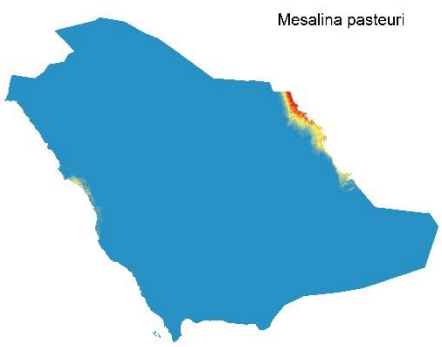

Eryx colubrinus

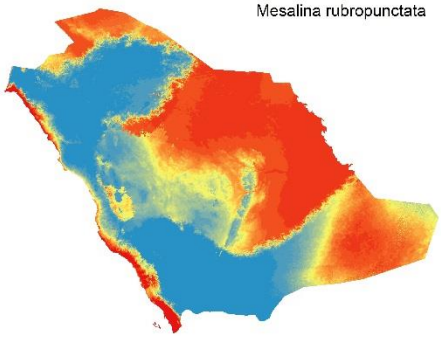

Eryx jaculus


Eumeces schneiderii


Hemidactylus robustus
Hemidactylus turcicus
Laudakia stellio
Leptotyphlops cairi
Lytorhynchus diadema
Macroprotodon cucullatus
Malpolon moilensis
Malpolon monspessulanus
Mesalina bahaeldini
Mesalina guttulata

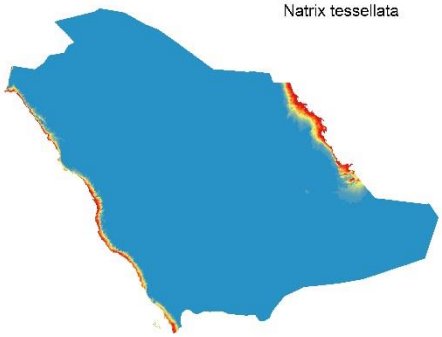

Mesalina olivier

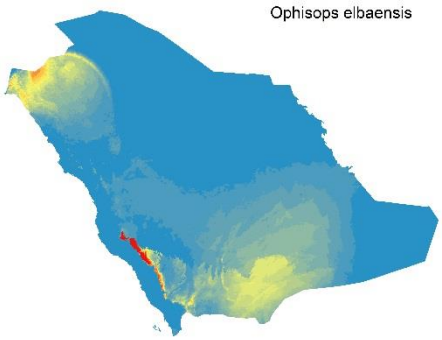

Mesalina pasteuri

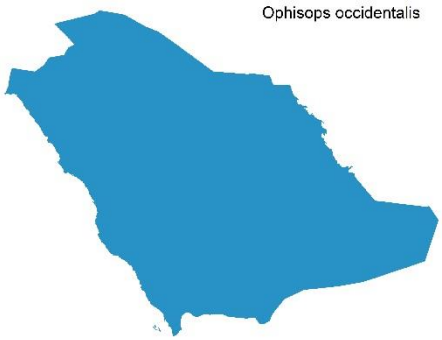

Mesalina rubropunctata

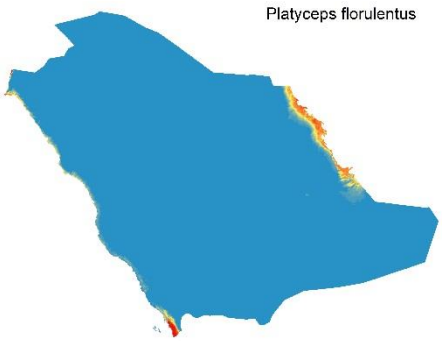

Naja haje

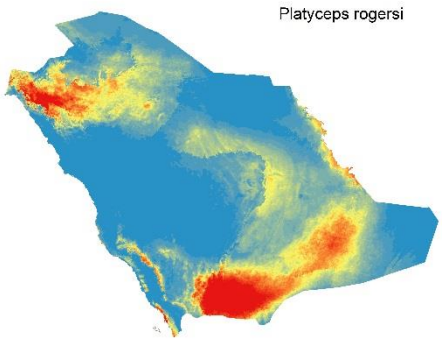

Naja nubiae

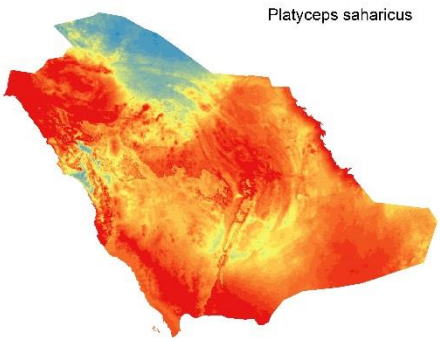

Natrix tessellata

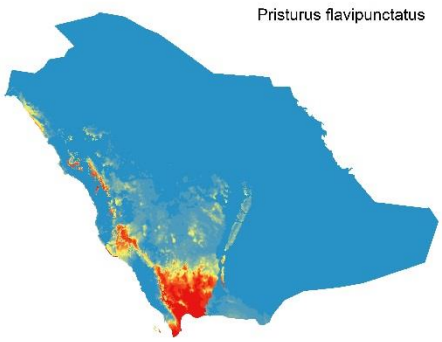

Ophisops elbaensis


Ophisops occidentalis

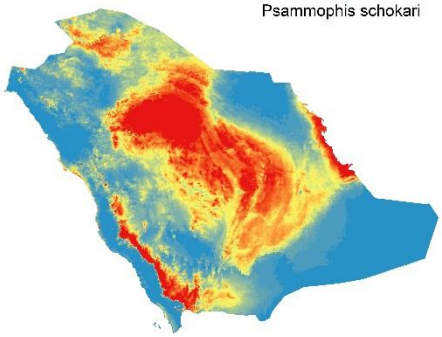

Platyceps florulentus


Platyceps rogersi

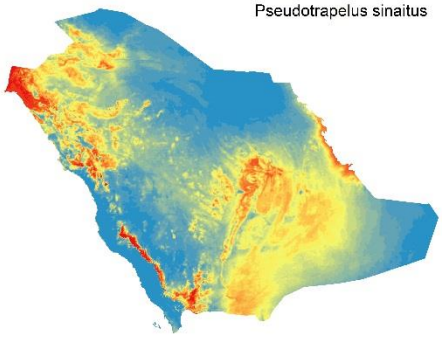

Platyceps saharicus

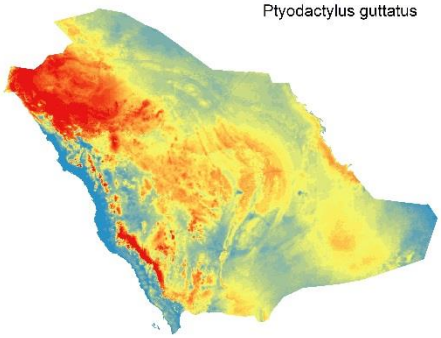

Pristurus flavipunctatus

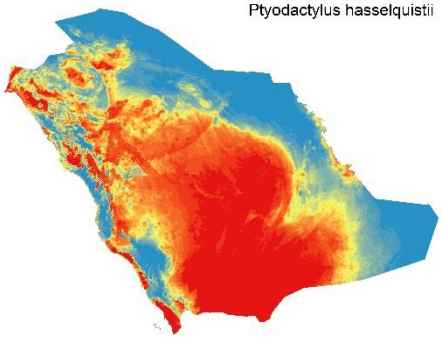

Psammophis aegyptius

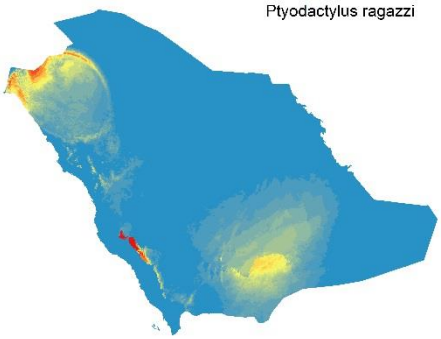

Psammophis schokari

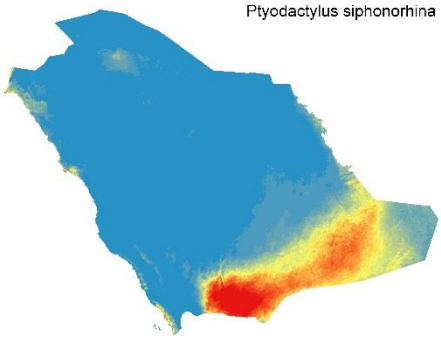

Psammophis sibilans

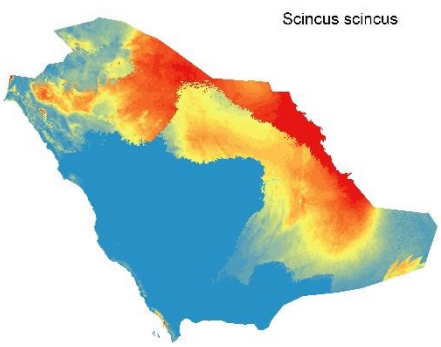

Pseudotrapelus sinaitus

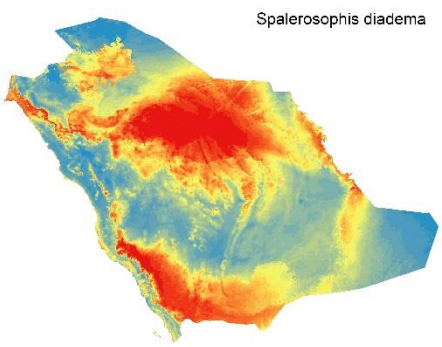

Ptyodactylus guttatus

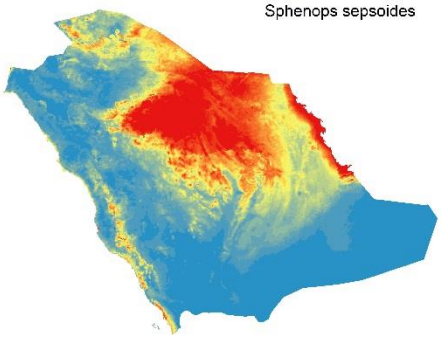

Ptyodactylus hasselquistii

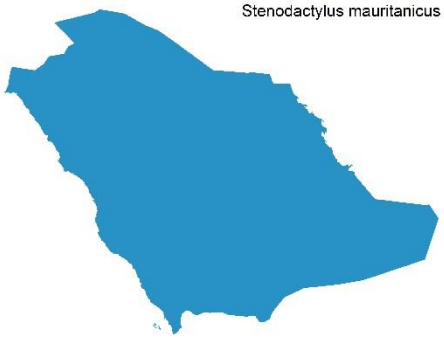

Ptyodactylus ragazzi

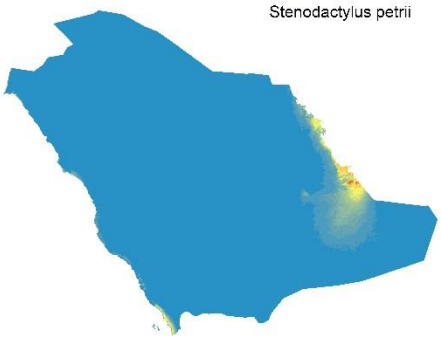

Ptyodactylus siphonorhina

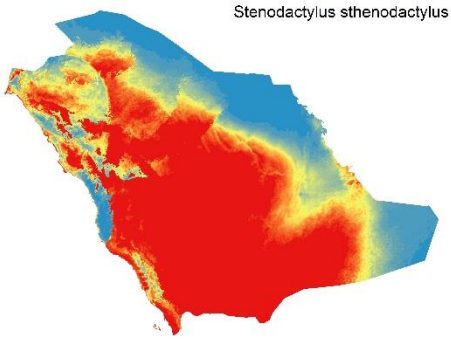

Scincus scincus

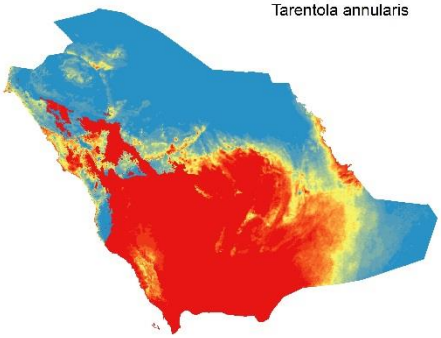

Spalerosophis diadema


Sphenops sepsoides

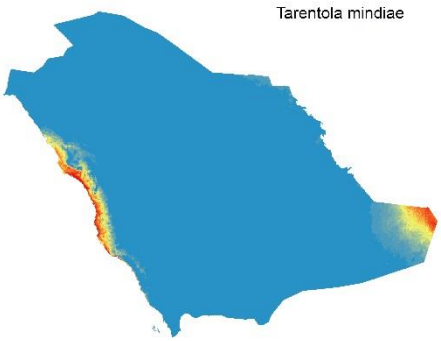

Stenodactylus mauritanicus

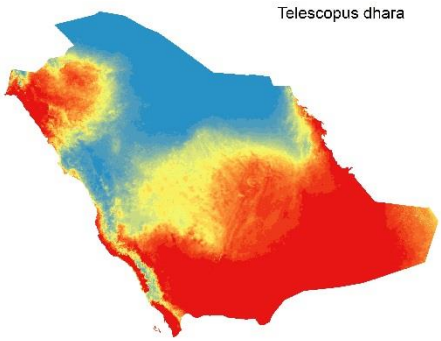

Stenodactylus petrii

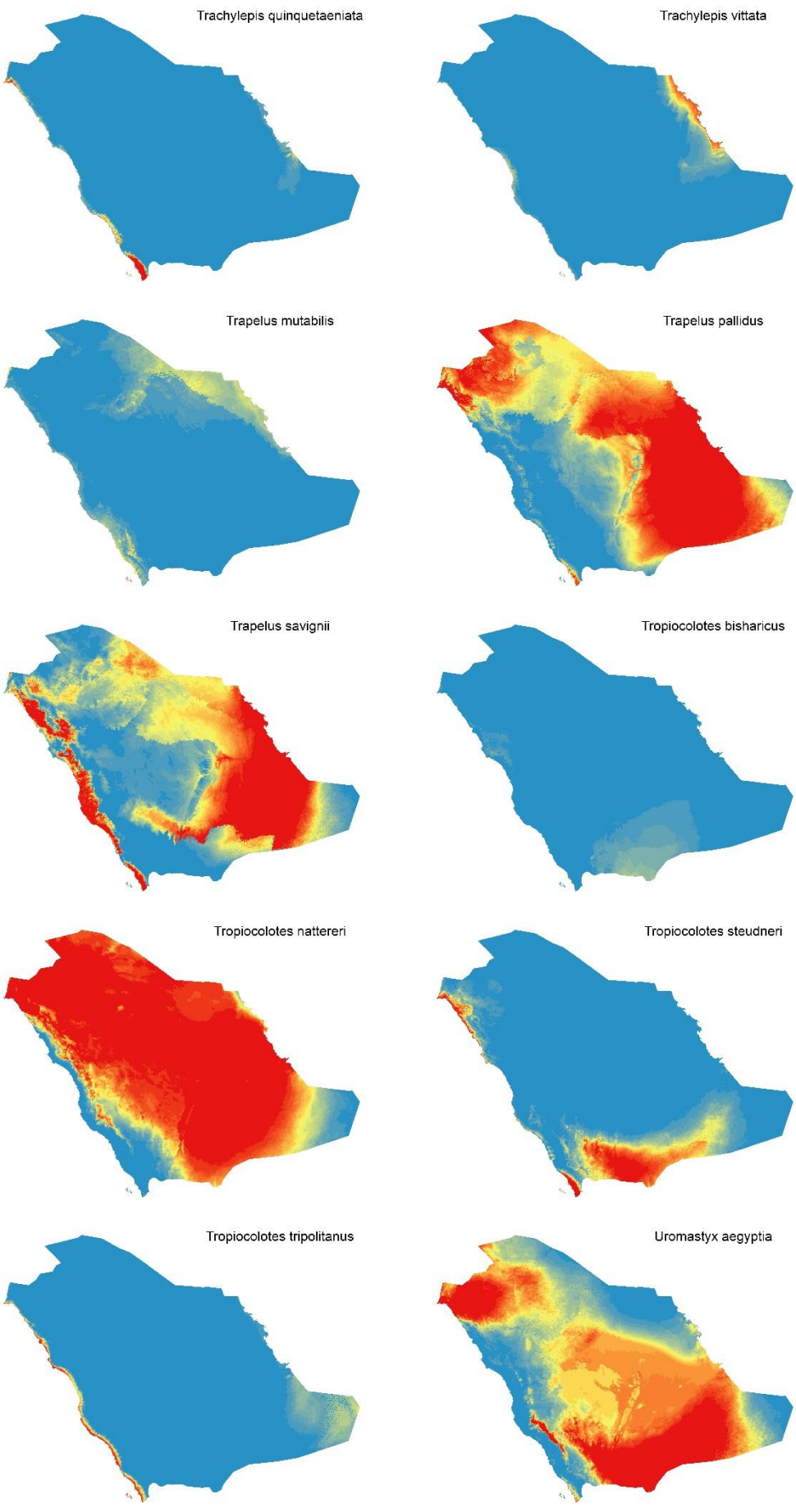

Stenodactylus sthenodactylus


Tarentola annularis


Tarentola mauritanica


Tarentola mindiae


Telescopus dhara


Trachylepis quinquetaeniata
Trachylepis vittata
Trapelus mutabilis
Trapelus pallidus
Trapelus savignii
Tropiocolotes bisharicus
Tropiocolotes nattereri
Tropiocolotes steudneri
Tropiocolotes tripolitanus
Uromastyx aegyptia

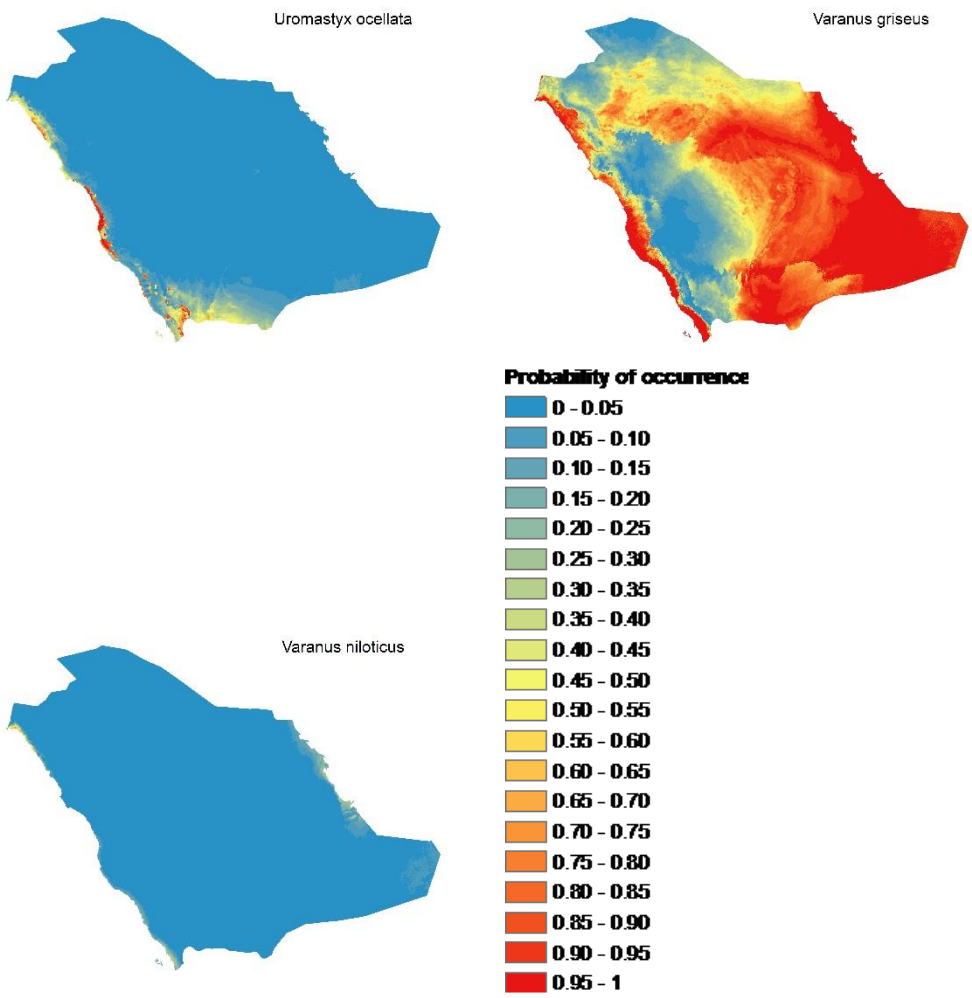


Figure S3.14: The transferred habitat suitability in Saudi Arabia for each reptile included in the Maxent modelling using the full set of 11 variables. The final consensus probability map for each species is an average of ten transferred replicates. The symbology of each map has been edited to match the Saudi map for comparision. All maps were generated by Maxent (Version 3.4.1) and manipulated by ArcGIS (Version 10.3.1).

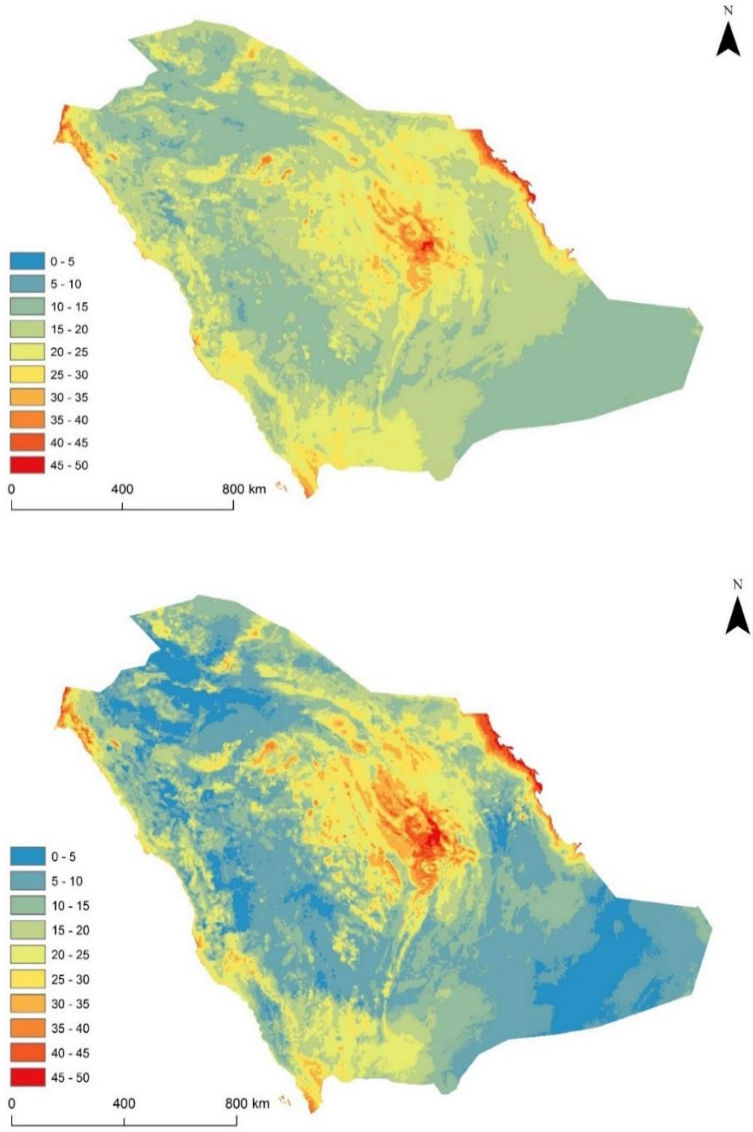


Figure S3.15: Upper: unthresholded predicted terrestrial reptile species richness for Saudi Arabia (Saudi model 1), based on summed habitat suitabilities from individual distribution models for 62 species. The map was generated by summing all averaged ascii maps for each species, and then rescaling to match Saudi model 2. Lower: thresholded predicted terrestrial reptile species richness for Saudi Arabia (Saudi model 2), based on summed binary predictions of suitable habitat from individual distribution models for 62 species. The map was generated by averaging the thresholded maps for each reptile species, and then summing the number of species in each pixel. Dark red indicates high predicted species richness.

4 Chapter Four: Field-testing the accuracy of predictions of species distribution models for reptiles in Saudi Arabia

Abstract

Validation of species distribution models is essential if their utility in conservation planning is to be fully realised. Due to the nature of the records, validation is typically achieved by partitioning available species occurrence data into training (used to build the distribution model) and testing sets (used to test model accuracy). However, this technique is problematic in the presence of sampling bias, and can overestimate model accuracy. Using newly collected, independent field data appears to be the best method to test the predictive power of a model. In this study, I used Maxent models built with both Egyptian (71 species) and Saudi Arabian (62 species) data to predict reptile distributions and species richness across Saudi Arabia. I then conducted directed field surveys based on these spatial predictions to test model accuracy, using both threshold-independent and threshold-dependent evaluation. Overlaying the maps of predicted species richness from Egyptian and Saudi models identified where they agreed and disagreed across Tabuk Province, and these were targeted in field sampling. My field validation was designed to facilitate occupancy modelling, to estimate species detectability and occupancy. The predicted richness was approximately three times greater than the richness recorded in the field. The ground-truthing validation of predictions showed a much lower estimation of model accuracy using field data compared to partitioning predictions ($r = 0.5$). Despite this, the transferred Egyptian model was able to predict presences better than the Saudi model built from local data, but the Saudi prediction was better at capturing distributions than the transferred Egyptian model. Detectability was weakly positively correlated with model accuracy calculated from the new independent field data. My field data included species that had not been recorded within

Saudi Arabia before. I conclude that field data are important in validating SDMs, to show the true predictability of the models, but that both transferred and interpolated models predict the spatial differences reasonably well.

Keywords

Detection probability, Maxent, occupancy modelling, reptile species richness, Saudi Arabia, transferability,

4.1 Introduction

Understanding the factors determining the spatial and temporal distribution of species is fundamental in ecology and biogeography (Rushton *et al.*, 2004; Guisan and Thuiller, 2005). Typically, species distributions are influenced by a variety of biological and geo-physical factors (Soberón and Peterson, 2005), some of which are being modified significantly by human activities (Chapin *et al.*, 2000). Determining the importance of these factors is required to inform specific conservation actions (Pardo *et al.*, 2013), and also to understand better the ecological and biological processes that shape communities (Pineda and Lobo, 2009). However, because of unequal sampling effort (Pardo *et al.*, 2013), some geographic regions are predicted to have higher species richness than is currently recorded (Ficetola *et al.*, 2013), which indicates that knowledge of species distributions is still far from complete (Mora *et al.*, 2011).

In the last two decades, considerable attention has been given to the use of species distribution modelling (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005; Pearson, 2008) and occupancy modelling (MacKenzie *et al.*, 2002, 2006) to fill the gaps in our knowledge of species distributions (Peterman *et al.*, 2013). Such models seek to identify

empirical relationships between observed species presence/absence and environmental variables, with the aim of predicting distributions across wide geographic areas. Ideally, these models are built with data from comprehensive and systematic field sampling. However, comprehensive sampling of species occurrences is rare for various reasons, including cost, time constraints, and logistics (Mora *et al.*, 2011; Pardo *et al.*, 2013). Scientists often therefore take advantage of the large number of occurrence records in museum and herbaria databases (Graham *et al.*, 2004; Newbold, 2010; Pardo *et al.*, 2013), as an alternative source of data with which to build distribution models. However, extra attention is required when using museum data to build species distribution models because there is always a degree of uncertainty associated with the records caused by spatial and temporal sampling bias, and issues with geographic and taxonomic accuracy (more details in Graham *et al.*, 2004; Newbold, 2010; Newbold *et al.*, 2010; Pardo *et al.*, 2013 and the references therein). Furthermore, since museum records rarely indicate locations in which species have been recorded as absent, it is hard to use such records with traditional distribution modelling approaches which rely on presence/absence data. Despite the known problems associated with museum data, in many cases there is no alternative, making them too valuable to ignore (Newbold *et al.*, 2010; Pardo *et al.*, 2013), especially in case of relatively understudied habitats such as the Middle East. Modelling approaches such as 'Maxent' have been developed in recent decades which only require presence information for the species of interest; these have expanded the potential for museum records to be used in distribution modelling studies (see Phillips *et al.*, 2006; Baldwin, 2009).

Testing the performance of distribution models (also known as validation) is a challenging task, but is an important step in attempts to understand and predict species distributions (Pearce and Ferrier, 2000). Testing model predictions allows us to compare different modelling techniques, provides insights into the suitability of a model for specific

tasks, and can lead to the improvement of models that perform poorly (Pearce and Ferrier, 2000). The testing approach taken is typically dependent on the goal of the study and the algorithm being used (Araújo *et al.*, 2005, Pearson, 2008). Measures of the validity of a distribution model normally fall under discrimination capacity (whether a model can differentiate between presence and absence at a site) (see Pearce and Ferrier, 2000).

Testing model predictions can be done using independent data sets, or via data partitioning methods (Verbyla and Litvaitis, 1989; Fielding and Bell, 1997; Pearce and Ferrier, 2000; Araújo *et al.*, 2005). Due to the nature of the records, the partitioning approach is most commonly used, where the original data are partitioned into training (used to build the distribution model) and testing sets (used to test model accuracy) (Verbyla and Litvaitis, 1989; Fielding and Bell, 1997; Pearson, 2008). However, this technique can yield uncertain predictions, especially if there is bias in the original data, as is typically expected when using museum data, for example (Newbold *et al.*, 2010). Various techniques have been developed to correct the effects of sampling bias and minimize its impact on predicted distributions (e.g., target-group background; Phillips *et al.*, 2009). However, using newly collected, independent field data appears to be the best method to test the predictive power of a model, because it avoids some of the drawbacks of methods involved in the splitting of datasets (Verbyla and Litvaitis, 1989; Pearce and Ferrier, 2000, Newbold *et al.*, 2010). This is still not a common practice in testing distribution models (Newbold *et al.*, 2010) because obtaining new data is typically costly, time-consuming and/or logistically difficult (Verbyla and Litvaitis, 1989; Newbold *et al.*, 2010).

As well as testing the performance of a distribution model in the geographic region for which it was developed, we may seek to validate model predictions in landscapes different from those in which the data were collected, i.e. transference (Randin *et al.*, 2006; Peterson *et al.*, 2011; Werkowska *et al.*, 2017; Yates *et al.*, 2018). If a distribution model can

provide valid predictions for species in other geographic areas, this could justify the use of models built in areas with extensive data on species occurrences to predict distributions in poorly studied areas, where otherwise we know very little about patterns of biodiversity (Werkowska *et al.*, 2017; Yates *et al.*, 2018). Model transference has already been used to make predictions for disjunct (Duque-Lazo *et al.*, 2016) and contiguous areas (Randin *et al.*, 2006; Barbosa *et al.*, 2009).

Even with systematic independent field surveys, obtaining new data for model testing is difficult. One problem is that the failure to record a species on a visit to a site can result from a failure to detect a species which is actually present (MacKenzie *et al.*, 2002, 2006). A solution to this problem is provided by occupancy modelling, a powerful technique used to estimate detectability and the factors influencing it, and hence to estimate true occupancy when the probability of detection is less than 1 (MacKenzie *et al.*, 2002, 2006). In occupancy modelling, occupancy (ψ) can be defined as the probability of a randomly selected site being occupied by the species of interest, while detectability (p) is the probability of detecting the species at site given that it is present (MacKenzie *et al.*, 2002, 2006; Donovan and Hines, 2007). If the detectability of a target species is likely to be below 1, estimating both occupancy and detectability is highly recommended when trying to understand the factors influencing the distribution; otherwise estimated parameters and inferences about causal relationships may be biased (MacKenzie *et al.*, 2002, 2006; MacKenzie and Royle, 2005). More generally, occupancy modelling has important advantages when surveying species, including savings in time and expense (if exhaustive sampling can be avoided as a result), and a reduction in the impact of unequal sampling effort (MacKenzie *et al.*, 2002, 2006; MacKenzie and Nichols, 2004).

A variety of recent studies have used predictions generated from species distribution models to direct new field surveys (Peterman *et al.*, 2013; West *et al.*, 2016; Rhoden *et al.*,

2017; Tronstad *et al.*, 2018). Some of these studies have used the newly collected presence/absence data to build a new distribution model to compare with the initial model (Newbold *et al.*, 2010; West *et al.*, 2016). I am only aware of a few studies that have used occupancy modelling when testing the validity of species distribution models (Newbold *et al.*, 2010; Gormley *et al.*, 2011; Peterman *et al.*, 2013). Considering the value of these two forms of modelling in isolation, the potential benefits of combining them are considerable. Robust occupancy modelling can help produce more reliable tests of distribution model accuracy, and hence facilitate better predictions about the distributions of important species, especially in areas of the world where new data are challenging to collect.

In this study, I used the most comprehensive datasets to date of terrestrial reptile species in Saudi Arabia and Egypt to test the accuracy of the predictions of species distribution models for Saudi Arabia. Terrestrial reptiles are an important component of biodiversity in Saudi desert ecosystems, but their distributions are poorly understood because of a shortage of suitable ecological studies. Hence, I aimed to 1) build species distribution models for Saudi terrestrial reptiles using presence-only data collected from museum records and published sources, 2) compare these with transferred distribution model predictions from a different (better-studied) geographic area (Egypt), 3) conduct intensive fieldwork, directed by the spatial predictions from the Saudi model and Egyptian model, 4) estimate occupancy and detectability from the new field data, and 5) using directed field surveys data, assess the broad-scale accuracy of the two sets of predictions. To my knowledge, this is the first time for any taxon that new field data and occupancy modelling have been combined to evaluate predictions of distribution models built both using local data and using data from a separate geographic area.

4.2 Methods & Materials

4.2.1 Study areas and data

Saudi Arabia is a large hyper-arid country [2,149,690 km^2] (Vincent, 2008), and Egypt is in the extreme northeast of the African continent [1,019,600 km^2] (Figure 4.1) (Baha El Din, 2006). Both countries contain a wide range of topographical and climate variation with large areas of arid open desert; the very limited vegetation occurs only in particular locations. Saudi Arabia has a list of 103 recorded reptile species (AbuZinada *et al.*, 2004) while at least 118 have been recorded from Egypt (Baha El Din, 2006).

Presence-only occurrence records were obtained for Saudi terrestrial reptiles from museums and relevant literature. Extensive georeferencing of locations was done to ensure as much precision as possible. Results are presented for a total of 62 reptile species (41 lizards and 21 snakes; hereafter called the original Saudi data), with 3,943 records in total (Figure 4.1). The accepted names of the reptile species were checked using The Reptiles Database (www.reptile-database.org), and Catalogue of Life (ITIS; www.catalogueoflife.org).

Presence-only occurrence records were obtained for Egypt from the database of the BioMap project (see El-Gabbas *et al.*, 2016). This is the most comprehensive dataset for the herpetofauna of Egypt, which has been extensively checked for taxonomy and georeferencing. In this study, I used the occurrences of 71 species (49 lizards and 22 snakes) with 11,103 records in total (Figure 4.1).

Environmental variables for Saudi Arabia were downloaded from WorldClim Version 2.0 (19 bio-layers and 12 indices of solar radiation), and Version 1.4 (altitude) (www.worldclim.org) (Hijmans *et al.*, 2005, Fick and Hijmans, 2017), at 30-sec resolution (~1 km^2). To reduce multicollinearity, stepwise selection of environmental variables was performed by calculating the Variance Inflation Factor (VIF) using the library *usdm* in R (Naimi *et al.*, 2014; R Development Core Team, 2019). Restricting the predictors to variables

with a VIF < 10 retained 11 variables for use in the modelling (Table S4.2). Since my aim was to test model performance in Saudi Arabia, I used the same 11 Saudi Arabian environmental variables to transfer the model from Egypt. Variables of vegetation cover (such as NDVI) were not used because the vegetation in hyper-arid environments is extremely sparse and has been shown to be a poor predictor of reptile distributions in Egypt (see El-Gabbas *et al.*, 2016). All the environmental variables were prepared by clipping them to the size of the study area and converting them to ascii files using ArcGIS software (Version 10.3.1).

4.2.2 Species richness distribution modelling

I created two types of distribution models, one from the Saudi data ("Saudi models") and one from Egyptian data ("Egyptian models"), using Maxent Version 3.4.1, downloaded from https://biodiversityinformatics.amnh.org/open_source/maxent/ (Phillips *et al.*, 2006). Maxent identifies areas that have similar values on environmental variables to those at the locations of recorded occurrences. By default, Maxent generates 10000 background points chosen randomly from the study area. I chose Maxent because: i) it is very robust, producing reliable models from presence-only data; ii) it works with low numbers of records; and iii) it appears to be relatively insensitive to sample bias (see Baldwin, 2009). Maxent has been shown to perform very well in comparisons with other modelling approaches (Baldwin, 2009). For both Saudi and Egyptian models, ten maps were created for each species using randomly selected subsamples (90%) of the records.

Overall, the geographical coverage of the sample points in Egypt and Saudi Arabia is fairly extensive (Figure 4.1). However, as is expected with museum data and unsystematic surveys (Graham *et al.*, 2004; Newbold, 2010), the data show an unavoidable bias toward cities, road networks and easily accessible areas. To minimize the impact of bias and spatial

autocorrelation in my model predictions, I employed a target-group bias file (Phillips *et al.*, 2009). This uses the sum of all the records of the taxon of interest as an estimate of sampling effort, smoothed using a Gaussian kernel estimation function (using the SDMtoolbox 2.0 in ArcGIS: Brown *et al.*, 2017). The bias file is used in Maxent to promote the selection of more background points from biased locations. It is a common approach to correcting sampling bias used in distribution modelling and is known to improve model accuracy (Phillips *et al.*, 2009).

For each of the two model types (Saudi and Egyptian), I created two types of map predicting species richness across Saudi Arabia. First, an "unthresholded" species-richness map was created by adding together all averaged (across 10 replicates) ascii probability maps for each species. Second, a consensus "thresholded" species-richness map was created by adding together ten replicate ascii thresholded presence-absence maps for each species. A species was considered to be present in a pixel if more than five replicates of the model predicted its presence. All these consensus maps were then added together to create a final thresholded species-richness map for each model.

4.2.3 Model evaluation

For a threshold-independent evaluation, I calculated the Area Under the Curve (AUC) calculated from the "Receiver Operating Curve" (ROC) (Fielding and Bell, 1997; Pearce and Ferrier, 2000) to evaluate the predictive performance of both Saudi and Egyptian models using a data partitioning approach. I also calculated AUC to evaluate the performance of the Saudi model using new independent field test data, with the entire original Saudi dataset being used for model calibration only.

For a threshold-dependent evaluation, I used confusion matrices to evaluate the predictions of the Egyptian and Saudi models against observed species records. The confusion matrix can be used in this way on the assumption that predictions in the form of continuous habitat suitabilities can be converted to binary presence/absence (actually suitable/unsuitable) data using a threshold above which a species is predicted to be present (Liu *et al.*, 2005). The confusion matrix is used to calculate "sensitivity" (the model's ability to predict presence correctly), "specificity" (the model's ability to predict absence correctly), and the correct classification rate (which uses both sensitivity and specificity). Additionally, the confusion matrix is used to calculate the "commission" error rate (i.e. proportion of sites in which a species is predicted to be present but is actually recorded as absent), and the "omission" error rate (the proportion of sites in which a species is predicted to be absent but is actually recorded as present) (Fielding and Bell, 1997; Pearce and Ferrier, 2000; Allouche *et al.*, 2006; Pearson, 2008).

4.2.4 Fieldwork to test predictive accuracy

Field validation of species distribution models was carried out in Tabuk Province in the extreme north-west of Saudi Arabia during summer 2017 & 2018 (Figure 4.2). Tabuk has an area of approximately 139,000 km^2 (about 7% of Saudi Arabia) (Al-Harbi, 2010), with distinct combinations of topographical and environmental characteristics (e.g. coastal areas, Red Sea escarpment, sand dunes, the Hisma plateau, lava desert) (Vincent, 2008). The climate is generally hot during the summer (June – August) ranging from 30 – 40 °C, while the winter (December – February) is exceptional in comparison to the rest of Saudi Arabia in getting very cold. Temperatures can sometimes drop below 0 °C (Vincent, 2008) leading to

localized snowfall. The annual rainfall in Tabuk is generally very low, averaging about 35 mm (Gosling *et al.*, 2011).

Fieldwork was undertaken in June, July and August of 2017 and 2018, partially guided by the spatial predictions of reptile species richness from the Saudi and Egyptian models. In 2017, before modelling based on existing records had been completed, I visited 10 sites, minimizing bias by selecting them to be in different habitats widely separated across Tabuk Province, and not visited before by other local studies of reptiles. In 2018, by which time modelling was complete, I classified the predicted species richness maps generated from each model into five categories using a quantile classification. I then focussed on areas which fell into extreme categories, representing the lowest (< 14), and highest (> 19) predicted potential species richness, ignoring areas predicted to have intermediate species richness. I overlaid the two maps of categorized species richness from the Egyptian and Saudi models to identify where they agreed and disagreed across Tabuk Province (Figure 4.2, Table S4.3). Three replicate sites were then chosen at random for field sampling within areas with each possible combination of extreme model predictions (i.e. high-high, high-low, low-high and low-low), yielding a total of a further 12 sites. Seven of the sites visited in 2017 turned out to have intermediate predicted species richness from at least one model; these seven sites were not used in the statistical analysis involving the extreme sites (see below) (Figure 4.2: Table S4.3).

Each site was visited on three separate occasions for one day each, separated by minimum of at least 12 days, and each time I searched 1 km² of area to match the resolution of the environmental variables. During each visit, searches were carried out during the active period for reptiles in Saudi Arabia (early to late morning, and early to late evening). At each site, two trained observers, including one Bedouin guide expert at collecting reptiles, haphazardly searched the area looking for signs of reptile activity. When an animal was

spotted, it was caught by hand if it was a lizard, and a pair of tongs if it was a snake. All individuals were photographed for subsequent expert verification of identity (by Dr. Sherif Baha El Din). Any captured individual was then released in the same location where it was captured. The exact coordinates of each individual were recorded using Garmin eTrex 10 device. Although I recorded the abundances of each species observed, here I focus purely on the presence/absence of each species on each sampling occasion.

4.2.5 Occupancy modelling

The field validation was designed to facilitate occupancy modelling, to account for imperfect detection. All analyses of occupancy modelling were performed with PRESENCE software (Hines, 2006). I ran PRESENCE without any predictor covariates to avoid over-fitting and biased estimation, given the small sample size (three repeat visits). A detection (1) and non-detection (0) history was assigned to each species recorded during the fieldwork, following MacKenzie *et al.* (2002). There were therefore eight potential detection/ non-detection sequences for each species at each site (111, 000, 100, 101, 110, 010, 011, 001). Note that 000 does not indicate species absence: the species may be there but not detected (MacKenzie *et al.* 2002). The detection histories from each site were entered into PRESENCE's matrix interface. The software uses maximum likelihood to estimate occupancy (ψ) and detectability (p) using a logit function/link (MacKenzie *et al.*, 2002, 2006; Donovan and Hines, 2007).

Pearson and Spearman correlations, and t-tests, were implemented using R version 3.5.1 (R Developments Core Team, 2019). A specific non-parametric two-way Anova (Barnard *et al.*, 2017) was conducted to compare observed reptile species richness among sites with high and low predicted species richness from the Saudi and Egyptian models.

4.3 Results

My field survey revealed populations of the Sinai Fan-fingered Gecko (*Ptyodactylus guttatus*) (three individuals observed at one site) and *Ptyodactylus ananjevae* (> 20 individuals observed from one site) that have not been reported from Saudi Arabia before. Another notable result was that I was able to add seven new site records for Natterer's Pigmy gecko (*Tropiocolotes nattereri*), extending its known range considerably. According to my database and other published work (Wilms *et al.*, 2010), this species was described from Tabuk Province around 40 years ago from only two localities. I also recorded Schneider's Skink (*Eumeces schneideri*) from a new site, a species rarely reported from Saudi Arabia to date (14 records).

Among 22 sites visited, only one had no species detected, whilst the most diverse site had nine species detected. Occupancy modelling showed that in the field the estimated probabilities of detection and occupancy varied considerably among species (Table 4.1). The detection probability (*p*) ranged from 0.02 to 1, while occupancy ranged from 0.05 to 1. Overall, lizards had a significantly higher chance of being detected in the field than snakes (independent t-test: $t_{8.3} = 2.4$, $p < 0.05$).

The observed species richness was often quite different from that predicted by the Saudi and Egyptian models. For instance, both Al-Beer and Sheqrei were predicted to have low potential species richness and habitat suitability by both models (Table S4.3), but I found a substantial difference in observed species richness, with the former site having the lowest observed richness (0), whilst the latter had seven observed species. Additionally, a site near the coast (South of Haqeel) was predicted to have high potential species richness by both models, but only two were recorded.

Model accuracy estimated from partitioning data when building the Saudi model ranged from 0.68 to 0.94, with models for all species performing better than would be expected at random (> 0.5; one-sample t-test: $t_{20} = 20.6$, $p < 0.001$). Accuracy calculated from the new independent field data was significantly lower than when estimated from partitioned data (Figure S4.8; Paired t-test, $t_{20} = 5.5$, $p < 0.001$) ranging from 0.16 to 0.99, and was not significantly better than random (t-test: $t_{20} = 0.93$, $p > 0.05$). There was, however, a significant positive correlation between the two measures of accuracy across species (Figure 4.3; Pearson's correlation: $r = 0.5$, $df = 19$, $p < 0.05$). There was a marginally non-significant positive correlation between species detectability, estimated from occupancy modelling, and distribution model accuracy (Figure 4.4; $r = 0.4$, $n = 21$, $p = 0.07$).

There was a significant difference in observed reptile species richness among extreme categories of field site (high or low predicted richness) based on predictions from the Saudi distribution models: observed species richness was significantly greater in sites which were predicted by the Saudi model to have high diversity (specific non-parametric 2-way ANOVA: $z = 1.66$, $p < 0.05$). However, there was no significant difference among categories based on the Egyptian model ($z = - 0.65$, $p > 0.05$), and no interaction between the effects of the two types of category ($z = - 0.54$, $p > 0.05$) (the negative z-values show that the means are actually opposite to the expected directions).

The predicted species richness in each grid square cell generated by the Saudi model was positively but not significantly correlated with observed species richness in the field using unthresholded (Spearman: $r_s = 0.20$, $n = 22$, $p > 0.05$) or thresholded ($r_s = 0.21$, $n = 22$, $p > 0.05$) predictions of species richness (Figure 4.5).

For the Egyptian models, the relationships between the predicted and observed species richness using the unthresholded and thresholded species richness habitat suitability

maps were both weakly negative and not significant (unthresholded: r_s = -0.21, n = 22, p > 0.05; thresholded: r_s = -0.22, n = 22, p > 0.05; Figure 4.6).

Egyptian and Saudi model generated different accuracy outcomes when testing on the ground in terms of discrimination capacity. Using presences/absence data collected from the field, I found that the transferred Egyptian model had a higher average of correct predictions of presences (sensitivity) while the Saudi model was better in terms of correct predictions of absences (specificity) (Figure 4.7).

4.4 Discussion

Validation of species distribution models is a critical part of model calibration. Using new independent test samples is considered best practice to test model accuracy, because it avoids some of the issues involved in the splitting of datasets (Pearce and Ferrier, 2000, Newbold *et al.,* 2010). Here I used independent field surveys to validate two types of distribution model (interpolated and transferred) to predict at least the gross features of reptile species richness in the study area. I found that, across species, field validation gave much lower estimates of accuracy than the normal partitioning method, suggesting that the data partitioning approach may overestimate accuracy, and hence impute too much confidence in the models. Field validation of the predictions led me to record species hitherto unknown in Saudi Arabia. Comparing different approaches to test model accuracies can overall enhance our understanding about model's weakness and strengthens (Pearce and Ferrier, 2000).

4.4.1 Model accuracy

The suitability of a model is determined by its accuracy, but this can be determined in various ways. In this study, accuracy calculated using partitioning of the original Saudi data

was significantly higher than that estimated from independent field data. This suggests that the data partitioning approach may overestimate accuracy, giving us too much confidence in my models. This is potentially problematic if a conservation decision is based solely on these predictions. Araújo *et al.* (2005) temporally and Newbold *et al.* (2010) spatially validated model accuracy using independent data, and also showed that model accuracy dropped when using independent data. The difference in accuracy estimates between partitioning and field testing may result from the fact that partitioned data are likely to share unavoidable sampling biases, which can lead to inflated estimates of accuracy (Araújo *et al.*, 2005; Newbold *et al.*, 2010). Another possible reason could be that the field data may not be reliable because of inadequate sampling, identification issues, sampling at the wrong time, or problems with detectability and the sampling protocol. The lower accuracy associated with independent data may be also explained by the low number of records (Newbold *et al.*, 2010), and also models may lack some of the potential factors that may influence the prediction (biotic factors, dispersal limitation, etc.,), all of which may result in an unsuccessful validation (Araújo *et al.*, 2005).

Model accuracy can also be measured through its ability to differentiate between presence and absence at a site (Pearce and Ferrier, 2000). Interestingly, the ground-truthing validation showed that transferred Egyptian models were able to predict presences better than the Saudi model built from local data, demonstrating the potential utility of species distribution modelling and its spatial transference from well-studied to less-studied areas. My Egyptian data form an extensive well-established dataset covering most ecological aspects of the Egyptian landscape, which might capture species-environment associations better when modelled, and hence might be better at prediction even when spatially transferred. In this study, my survey covered only Tabuk Province, a small area relative to the rest of the study area (Saudi Arabia). The accuracy and discrimination capacity of the Egyptian and Saudi

models for the rest of the study area is not known, and hard to predict for certain until specific ground-truthing is conducted.

Even when models are built with data from surveys that are not systematic, and which contain spatial and temporal biases, partitioning may still be useful in comparing models. Here, species with apparently high accuracies from the partitioning approach also were judged to be relatively accurate using the independent data approach, resulting in the positive correlation. Of course, models should be tested using independent data whenever possible (Verbyla and Litvaitis, 1989; Newbold *et al.*, 2010), but although providing useful insights into model predictive performance, such independent new field data require a lot of effort to obtain. Here, the new field data gave much lower accuracy in comparison with the normal software-based partitioning technique. Therefore, the limitations of such data in species distribution models must be acknowledged (mentioned above) as it may lead to bias in the estimation of accuracy. Overall, comparing different testing approaches to investigate species distribution patterns has some important benefits, including comparing different model techniques, viability of models for specific tasks, and identifying model drawbacks that need attention (Pearce and Ferrier, 2000). To conclude, models are not perfect and typically there are many inaccuracies (Araújo *et al.*, 2005; Pineda and Lobo, 2009), and therefore a successful model validation does not always suggest that a model is accurate enough for a specific task, and unsuccessful model validation does not typically indicate that a model is wrong (see Araújo *et al.*, 2005).

4.4.2 Species detection variation in the field

It is still an uncommon practice to use occupancy modelling in conjunction with species distribution modelling (but see Newbold *et al.*, 2010; Gormley *et al.*, 2011; Peterman

et al., 2013). The species I surveyed in the field show considerable variation in occupancy and detectability probabilities (Table 4.1), and species with high detectability tend to produce models with greater accuracy. This implies that distribution models built with museum data may be better at predicting distributions of easily detected species in the field, or possibly that field surveys are less effective at assessing the distributions of hard-to-detect species on the ground. Species detectability at a particular site typically depends on various factors, including local density, size, behaviour, environmental conditions and sampling effort (see Bailey *et al.*, 2004). Local density is probably the most important reason for variation in detectability (Bailey *et al.*, 2004), which may well be the cause of my finding that lizards had a higher detectability than snakes. Most lizards are typically more conspicuous and more abundant than snakes. Behaviour is also important in explaining the low detectability of elusive species. For instance, the Sinai agama (*Pseudotrapelus sinaitus*) is an active conspicuous species during the hottest parts of the day in dry rocky areas (Baha El Din, 2006), making it relatively easy to detect (Norfolk *et al.*, 2010), while a snake such as Burton's Carpet Viper (*Echis coloratus*) is crepuscular on rocky steep slopes (Baha El Din, 2006), making detection difficult. Lastly, the estimated detection probabilities might also be influenced by sampling effort (Bailey *et al.*, 2004; MacKenzie and Royle, 2005), although this was consistent here across all sites.

Newbold *et al.* (2010) conducted similar fieldwork in Egypt to investigate the occupancy and detectability of terrestrial reptiles, and also found that conspicuous abundant species such as lizards had higher detectabilities than elusive species such as snakes. Both studies demonstrate the advantages of using occupancy modelling when testing SDMs. However, using more sites across a larger scale would be advisable because this would enable statistical testing with greater power.

4.4.3 Predicted and observed species richness

A better understanding of species richness patterns involves comparing observed and predicted species richness (Pineda and Lobo, 2009). I found that the observed species richness was significantly greater in sites which were predicted by the Saudi model to have high diversity, and there was a positive although not significant relationship between predicted and observed species richness. This somewhat contrasts with the weak negative non-significant correlation for the transferred Egyptian model. This suggests that the Saudi prediction may have been better at capturing distributions than the transferred Egyptian model, but the evidence is for this difference is not strong. In my case, most areas predicted to have high potential species richness were coastal areas and their surroundings. However, during field work it was evident that some large divergences between predicted and observed species richness may have been the result of variation in habitat complexity. Anecdotally at least, I found habitats with more diverse habitats (vegetation, mountains, and sandy areas) had more species, but this was not captured by the models. For instance, Wadi Rumaihiah and South of Haqeel are coastal sites predicted to have high potential species richness by both models, but I observed 7 and 2 species, respectively, and the former had more diverse habitats than the latter. In 2018, I opportunistically searched a farm close to the Al-Beer site (which had 0 species detected), and three species were recorded. This suggests that the models may not capture important features of the environment which are good predictors of diversity, such as those associated with human activity. However, this is a subjective impression from field work which would need further study to confirm.

The pattern of moderate to high species richness next to the coast maybe an artefact of model bias, or just simply because suitable environmental areas have not yet been occupied, or (more likely) not adequately surveyed yet. This richness pattern predicted by both models could also have arisen because using stacked species distribution models, and using only

environmental data to relate them with species occurrences, can both overestimate species richness (Algar *et al.*, 2009; Pineda and Lobo, 2009). However, stacked models are able to produce predictions of species richness validated by macroecological modelling (Distler *et al.*, 2015). Similar patterns of species richness have been seen in other studies of terrestrial reptiles in similar arid environments (e.g., in Arabian Peninsula by Cox *et al.*, 2012, Egypt by El-Gabbas *et al.*, 2016, and in Oman by Carranza *et al.*, 2018). Even if the predictions have captured the patterns well, more rigorous sampling efforts could reveal different outcomes contrary to these predictions. Overall, a lot of detailed information and extensive analysis is required to predict accurately a species richness pattern (Pineda and Lobo, 2009), which unfortunately does not exist for many areas.

4.4.4 Guided field surveys to study unknown species richness

Regardless of what model accuracies tell us about model performance, the newly collected field data tell us important things about the very poorly studied reptile communities in Saudi Arabia. It is clear that a lot of work is still needed in terms of more systematic surveys, because large numbers of species new to science are still being described from Saudi Arabia and from the whole Arabian Peninsula (see Wilms and Schmitz, 2007; Wilms *et al.*, 2010; Busais and Joger, 2011; Carranza and Arnold, 2012; Nazarov *et al.*, 2013; Šmíd *et al.*, 2016) and adjacent areas (e.g., Jordan; Melnikov *et al.*, 2012; Nazarov *et al.*, 2013). My guided field surveys were successful in leading me to record unknown populations of *P. guttatus* and *P. ananjevae*. These two species have been recorded before in Jordan (Al-Quran, 2009; Nazarov *et al.*, 2013) and *P. guttatus* in Egypt (Baha El Din, 2006). I also considerably extended the range of some species, including *T. nattereri*. According to my dataset of records and the map in Wilms *et al.* (2010), the distribution of *T. nattereri* was only known from two records (1978, 1979) from Tabuk Province, while my fieldwork added seven new

sites to its distribution. I consider this finding regarding *T. nattereri* important from a conservation perspective, giving that all records for the species occur outside the current network of Saudi's protected areas (see Cox *et al.*, 2012).

A recent study by Aloufi and Amr (2015) also conducted a reptile survey in Tabuk Province (2011-2013), recording three new populations for the first time in Saudi Arabia (different species from those I recorded). It is striking to see that within four years, two different field surveys in the same study area have added five new species to the herpetofauna of Saudi Arabia. Based on my findings and those of others, I conclude that the accepted national list in AbuZinada *et al.* (2004) is far from complete taxonomically and biogeographically.

I can say with confidence that the Maxent predictions used to direct the field surveys were useful in these isolated less-studied habitats, as they have been in other contexts (see Raxworthy *et al.*, 2003; Peterman *et al.*, 2013; West *et al.*, 2016; Rhoden *et al.*, 2017; Tronstad *et al.*, 2018). Rhoden *et al.* (2017) aimed to validate Maxent predictions, and this led them to record new populations of endemic crayfish, whilst Raxworthy *et al.* (2003) in Madagascar were guided by SDMs produced by Genetic Algorithm for Rule-set Prediction (GARP), and ended up with descriptions of seven species new to science. Species distribution modelling has thus emerged as an important conservation tool that can be used to explore distributions in poorly known habitats (Raxworthy *et al.*, 2003).

4.5 Conclusion

Species distribution modelling has a lot of potential uses, including in conservation planning and directing new field surveys. Therefore, it is an important to validate models correctly; currently, the use of independent data is considered the best practice (Verbyla and Litvaitis, 1989; Newbold *et al.*, 2010). My results indicate that model accuracy calculated

from independent field data of terrestrial reptiles is lower than accuracy estimated from partitioned data, suggesting that the partitioning approach may over-estimate model accuracy. Incorporating occupancy modelling in conjunction with species distribution modelling, and its spatial transference, to direct field surveys and validate predictions is promising approach that should improve the utility of species distribution modelling overall. However, the weak relationships revealed here between my new observations and model predictions, despite intensive and time-consuming field work, may illustrate not just the potential limitations of the models, but also the difficulty of collecting high quality independent data for model validation.

4.6 References

AbuZinada, A., Robinson, E. R., Nader, I. & Al Wetaid, Y. 2004. *First Saudi Arabian National Report on the Convention on Biological Diversity*, The National Commission for Wildlife Conservation and Development.

Algar, A. C., Kharouba, H. M., Young, E. R. & Kerr, J. T. 2009. Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography*, 32**,** 22-33.

Al-Harbi, K. M. 2010. Monitoring of agricultural area trend in Tabuk region – Saudi Arabia using Landsat TM and SPOT data. *The Egyptian Journal of Remote Sensing and Space Science*, 13**,** 37-42.

Allouche, O., Tsoar, A. & Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43**,** 1223-1232.

Aloufi, A. A. & Amr, Z. S. 2015. On the herpetofauna of the Province of Tabuk, northwest Saudi Arabia (Amphibia, Reptilia). *Herpetozoa*, 27**,** 147-158.

Al-Quran, S. 2009. The Herpetofauna of the Southern Jordan. *American-Eurasian Journal of Agricultural and Environmental Sciences*, 6**,** 385-391.

Araújo, M. B., Pearson, R. G., Thuiller, W. & Erhard, M. 2005. Validation of species–climate impact models under climate change. *Global Change Biology*, 11**,** 1504-1513.

Baha El Din, S. 2006. *A Guide to the Reptiles and Amphibians of Egypt*, The American University in Cairo Press.

Bailey, L. L., Simons, T. R. & Pollock, K. H. 2004. Estimating site occupancy and species detection probability parameters for terrestrial salamanders. *Ecological Applications*, 14**,** 962-702.

Baldwin, R. 2009. Use of maximum entropy modeling in wildlife research. *Entropy*, 11**,** 854-866.

Barbosa, A. M., Real, R. & Vargas, J. M. 2009. Transferability of environmental favourability models in geographic space: The case of the Iberian desman (*Galemys pyrenaicus*) in Portugal and Spain. *Ecological Modelling*, 220**,** 747-754.

Barnard, C. J., Gilbert, F. & McGregor, P. K. 2017. *Asking Questions in Biology*. Pearson, UK.

Brown, J. L., Bennett, J. R. & French, C. M. 2017. SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ*, 5**,** e4095.

Busais, S. M. & Joger, U. 2011. Three new species and one new subspecies of *Hemidactylus* OKEN, 1817 from Yemen (Squamata, Gekkonidae). *Vertebrate Zoology*, 61**,** 267-280.

Carranza, S. & Arnold, E. N. 2012. A review of the geckos of the genus *Hemidactylus* (Squamata: Gekkonidae) from Oman based on morphology, mitochondrial and nuclear data, with descriptions of eight new species. *Zootaxa*, 3378**,** 1-95.

Carranza, S., Xipell, M., Tarroso, P., Gardner, A., Arnold, E. N., Robinson, M. D., et al. 2018. Diversity, distribution and conservation of the terrestrial reptiles of Oman (Sauropsida, Squamata). *PLoS ONE*, 13**,** e0190389.

Catalogue of Life. 2018. http://www.catalogueoflife.org/ (accessed on 15 March 2018).

Chapin, F. S., III, Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., Hooper, D. U., Lavorel, S., Sala, O. E., Hobbie, S. E., Mack, M. C. & Díaz, S. 2000. Consequences of changing biodiversity. *Nature*, 405**,** 234-242.

Cox, N. A., Mallon, D., Bowles, P., Els, J. & Tognelli, M. F. 2012. *The Conservation Status and Distribution of Reptiles of the Arabian Peninsula*. Cambridge, UK and Gland, Switzerland: IUCN, and Sharjah, UAE: Environment and Protected Areas Authority.

Distler, T., Schuetz, J. G., Velásquez-Tibatá, J. & Langham, G. M. 2015. Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. *Journal of Biogeography*, 42**,** 976-988.

Donovan, T. M. & Hines, J. 2007. Exercises in occupancy modeling and estimation. Available online at: http://www.uvm.edu/rsenr/vtcfwru/spreadsheets/?Page=occupancy/occupancy.htm

Duque-Lazo, J., van Gils, H., Groen, T. A. & Navarro-Cerrillo, R. M. 2016. Transferability of species distribution models: The case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. *Ecological Modelling*, 320**,** 62-70.

El-Gabbas, A., Baha El Din, S., Zalat, S. & Gilbert, F. 2016. Conserving Egypt's reptiles under climate change. *Journal of Arid Environments*, 127**,** 211-221.

Ficetola, G. F., Bonardi, A., Sindaco, R. & Padoa-Schioppa, E. 2013. Estimating patterns of reptile biodiversity in remote regions. *Journal of Biogeography*, 40**,** 1202-1211.

Fick, S. E. & Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37**,** 4302-4315.

Fielding, A. H. & Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24**,** 38-49.

Gormley, A. M., Forsyth, D. M., Griffioen, P., Lindeman, M., Ramsey, D. S., Scroggie, M. P. & Woodford, L. 2011. Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology*, 48**,** 25-34.

Gosling, S. N., Dunn, R., Carrol, F., Christidis, N., Fullwood, J., Gusmao, D., Golding, N., Good, L., Hall, T., Kendon, L., Kennedy, J., et al. 2011. *Climate: Observations, Projections and Impacts: Saudi Arabia*. The Met Office, UK.

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 & Evolution*, 19**,** 497-503.

Guisan, A. & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8**,** 993-1009.

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

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.

Hines, J. E. 2006. PRESENCE- Software to estimate patch occupancy and related parameters. USGS-PWRC. http://www.mbr-pwrc.usgs.gov/software/presence.html

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

MacKenzie, D. I. & Nichols, J. D. 2004. Occupancy as a surrogate for abundance estimation. *Animal Biodiversity and Conservation*, 27**,** 461-467.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A. & Langtimm, C. A. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83**,** 2248-2255.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L. & Hines, J. E. 2006. *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*, Elsevier.

MacKenzie, D. I. & Royle, J. A. 2005. Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology,* 42**,** 1105-1114.

Melnikov, D., Nazarov, R., Ananjeva, N. & Disi, A. 2012. A new species of *Pseudotrapelus* (Agamidae, Sauria) from Aqaba, southern Jordan. *Russian Journal of Herpetology,* 19**,** 143-154.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. & Worm, B. 2011. How many species are there on earth and in the ocean? *PLoS Biol,* 9**,** e1001127.

Naimi, B., Hamm, N., Groen, T. A., Skidmore, A. K. & Toxopeus, A. G. 2014. Where is positional uncertainty a problem for species distribution modelling? *Ecography,* 37**,** 191-203.

Nazarov, R., Melnikov, D. & Melnikova, E. 2013. Three new species of *Ptyodactylus* (Reptilia; Squamata; Phyllodactylidae) from the Middle East. *Russian Journal of Herpetology,* 20**,** 147-162.

Newbold, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography,* 34**,** 3-22.

Newbold, T., Reader, T., El-Gabbas, A., Berg, W., Shohdi, W. M., Zalat, S., Baha El Din, S. & Gilbert, F. 2010. Testing the accuracy of species distribution models using species records from a new field survey. *Oikos,* 119**,** 1326-1334.

Norfolk, O., Melotte, A., Gilbert, F., Zalat, S. & Reader, T. 2010. A comparative study of two agamid lizards, *Laudakia stellio* and *Pseudotrapelus sinaitus,* in southern Sinai. *Egyptian Journal of Biology,* 12**,** 27-43.

Pardo, I., Pata, M. P., Gomez, D. & Garcia, M. B. 2013. A novel method to handle the effect of uneven sampling effort in biodiversity databases. *PLoS ONE,* 8**,** e52786.

Pearce, J. & Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 133**,** 225-245.

Pearson, R. G. 2008. Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*. Available at http://ncep.amnh.org.

Peterman, W. E., Crawford, J. A. & Kuhns, A. R. 2013. Using species distribution and occupancy modeling to guide survey efforts and assess species status. *Journal for Nature Conservation*, 21**,** 114-121.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M. & Araújo, M. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190**,** 231-259.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. & Ferrier, S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19**,** 181-197.

Pineda, E. & Lobo, J. M. 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, 78**,** 182-190.

R Development Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.

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.

Raxworthy, C. J., Martínez-Meyer, E., Horning, N., Nussbaum, R. A., Schneider, G. E., Ortega-Huerta, M. A. & Peterson, A. T. 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, 426**,** 837-841.

Reptiles Data-Base. 2017. http://www.reptile-database.org/ (accessed on 20 October 2017).

Rhoden, C. M., Peterman, W. E. & Taylor, C. A. 2017. Maxent-directed field surveys identify new populations of narrowly endemic habitat specialists. *PeerJ*, 5**,** e3632.

Rushton, S. P., Ormerod, S. J. & Kerby, G. 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology*, 41**,** 193-200.

Šmíd, J., Shobrak M., Wilms, T., Joger, U. & Carranza, S. 2016. Endemic diversification in the mountains: genetic, morphological and geographical differentiation of the *Hemidactylus* geckos in southwestern Arabia. *Organisms Diversity and Evolution*, 17**,** 267-285.

Soberón, J. & Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, 2**,** 1-10.

Tronstad, L. M., Brown, K. M. & Andersen, M. D. 2018. Using species distribution models to guide field surveys for an apparently rare aquatic beetle. *Journal of Fish and Wildlife Management*, 9**,** 330-339.

Verbyla, D. L. & Litvaitis, J. A. 1989. Resampling methods for evaluating classification accuracy of wildlife habitat models. *Environmental Management*, 13**,** 783-787.

Vincent, P. 2008. *Saudi Arabia: An Environmental Overview*. Taylor & Francis Group.

Werkowska, W., Márquez, A. L., Real, R. & Acevedo, P. 2017. A practical overview of transferability in species distribution modeling. *Environmental Reviews*, 25**,** 127-133.

West, A. M., Kumar, S., Brown, C. S., Stohlgren, T. J. & Bromberg, J. 2016. Field validation of an invasive species Maxent model. *Ecological Informatics*, 36**,** 126-134.

Wilms, T. & Schmitz, A. 2007: A new polytypic species of the genus *Uromastyx* MERREM 1820 (Reptilia: Squamata: Agamidae: Leiolepidinae) from southwestern Arabia. *Zootaxa*, 1394**,** 1-23.

Wilms, T., Shobrak, M. & Wagner, P. 2010. A new species of the genus *Tropiocolotes* from Central Saudi Arabia (Reptilia: Sauria: Gekkonidae). *Bonn zoological Bulletin*, 57**,** 275-280.

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., et al. 2018. Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33**,** 790-802.

Table 4.1: Probability of occupancy (ψ) and detection (*p)* of species that were found at 22 sites in Tabuk Province during field work summer 2017 & 2018. The last two columns show the predictions accuracy for each species. There are a few species detected in the field, but they do not have accuracy value because they were not part of the model process of independent Saudi model.

species name	ψ	SE (ψ)	*p*	SE (*p*)	AUC-partitioning	AUC-field data
Acanthodactylus boskianus	0.4651	0.1091	0.7167	0.0896	0.75	0.45
Acanthodactylus opheodurus	0.3206	0.1001	0.8034	0.0908	0.76	0.49
Acanthodactylus schmidti	0.1364	0.0732	1	0	0.82	0.91
Bunopus tuberculatus	0.4705	0.1108	0.6762	0.0951	0.83	0.46
Cerastes gasperettii	0.4404	0.2286	0.2753	0.1513	0.79	0.69
Chamaeleo chamaeleon	1	0.0002	0.0152	0.015	0.88	0.86
Cyrtopodion scabrum	0.1434	0.0777	0.634	0.183	0.86	0.79
Echis coloratus	0.0478	0.0471	0.634	0.317	0.87	0.51
Eumeces schneideri	1	0	0.0152	0.015	0.94	0.61
Laudakia stellio	0.2047	0.0968	0.518	0.1764	0.91	0.99
Lytorhynchus diadema	0.9275	0.7728	0.1307	0.1169	0.82	0.31
Mesalina guttulate	1	0	0.0455	0.0256	0.71	0.32
Phrynocephalus arabicus	0.238	0.1268	0.382	0.1876	0.76	0.37
Pristurus rupestris	0.1852	0.0839	0.7362	0.1372	0.92	0.67
Psammophis schokari	1	0	0.0606	0.0294	0.83	0.25
Pseudotrapelus sinaitus	0.119	0.0932	0.382	0.2653	0.82	0.62
Ptyodactylus ananjevae	0.0455	0.0444	1	0	-	-
Ptyodactylus guttatus	0.0478	0.0471	0.634	0.317	-	-
Ptyodactylus hasselquistii	0.0455	0.0444	1	0	0.80	0.95
Scincus mitranus	1	0	0.1061	0.0379	0.71	0.25
Spalerosophis diadema	1	0	0.0455	0.0256	0.84	0.47
Stenodactylus doriae	0.5258	0.1141	0.634	0.0956	0.68	0.43
Stenodactylus slevini	0.2047	0.0968	0.518	0.1764	0.81	0.16
Stenodactylus sthenodactylus	1	0.0001	0.0152	0.015	-	-
Trapelus pallidus	0.0455	0.0444	1	0	-	-
Tropiocolotes nattereri	0.4095	0.1223	0.518	0.1248	-	-

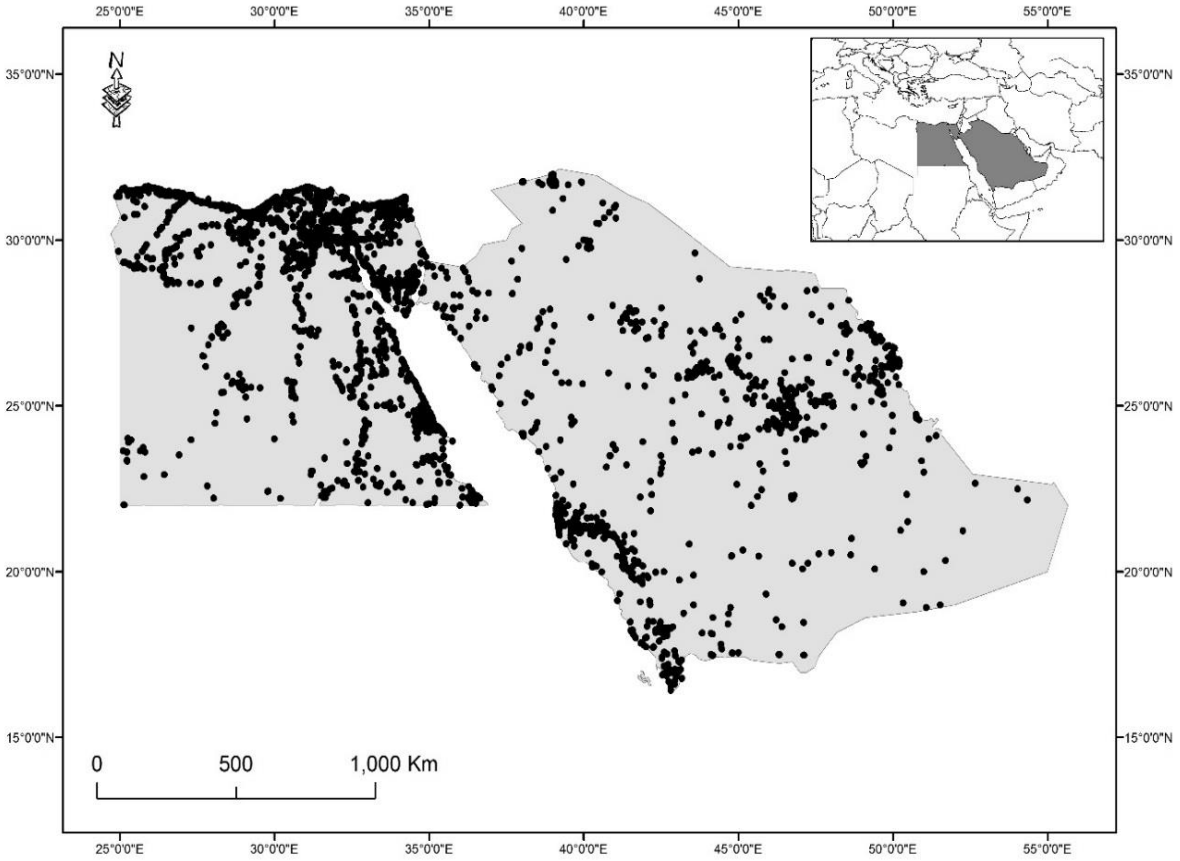


Figure 4.1: Study areas of Saudi Arabia and Egypt, along with 3,943 presence records (for 62 species) in Saudi Arabia used to calibrate the Saudi models, and 11,103 presence records (for 71 species) in Egypt used to calibrate the Egyptian models.

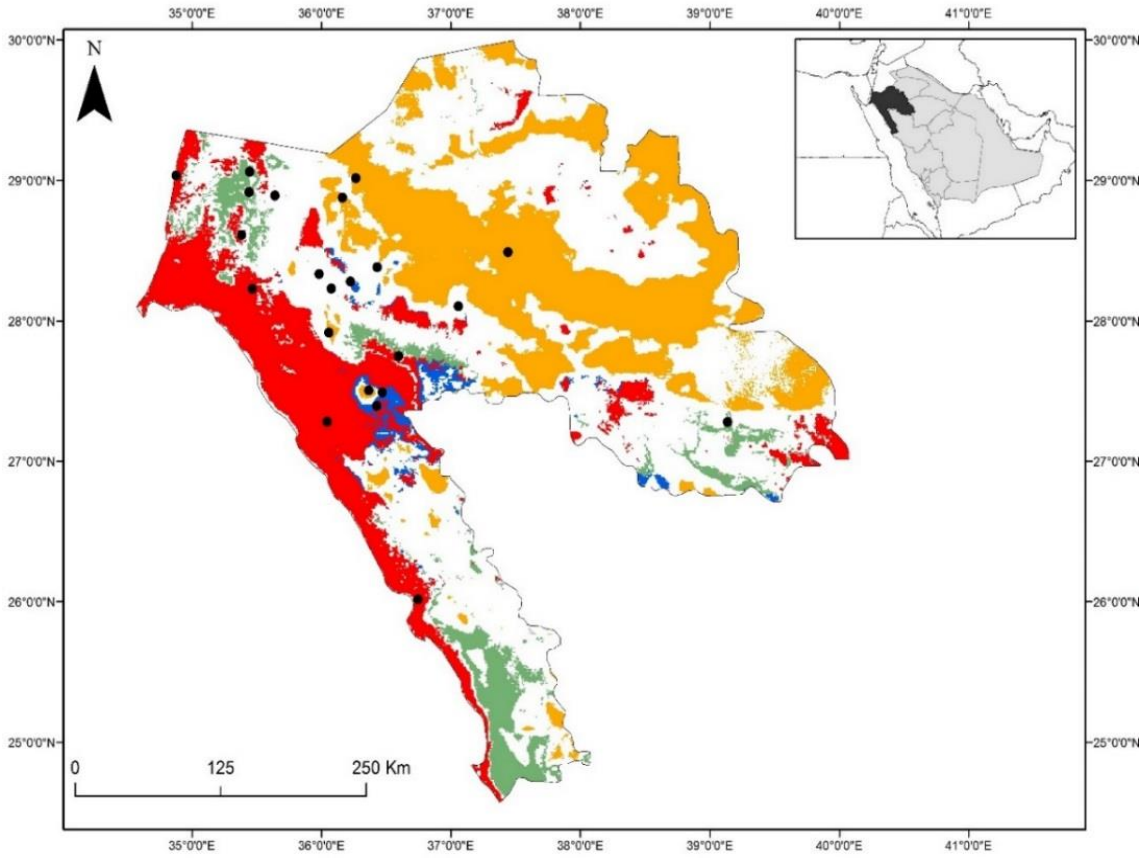


Figure 4.2: Maxent directed field validation localities (black dots) to test two types of species distribution models in Tabuk province, Saudi Arabia. For both Egyptian (transferred) and Saudi (interpolated) species richness predictions, I identified areas with predictions in the highest and lowest quintiles. Then, I combined both predictions to identify areas of agreement and disagreement between the models. Red grid cells represent areas in the highest quintile for both models, while orange cells represent areas predicted to be in the lowest quintile for both models. Green cells represent areas in the highest quintile for the Saudi model and the lowest quintile for the Egyptian model, and blue cells represent areas with opposite mismatching predictions. Areas coloured white are where one or both models predicted moderate species richness.

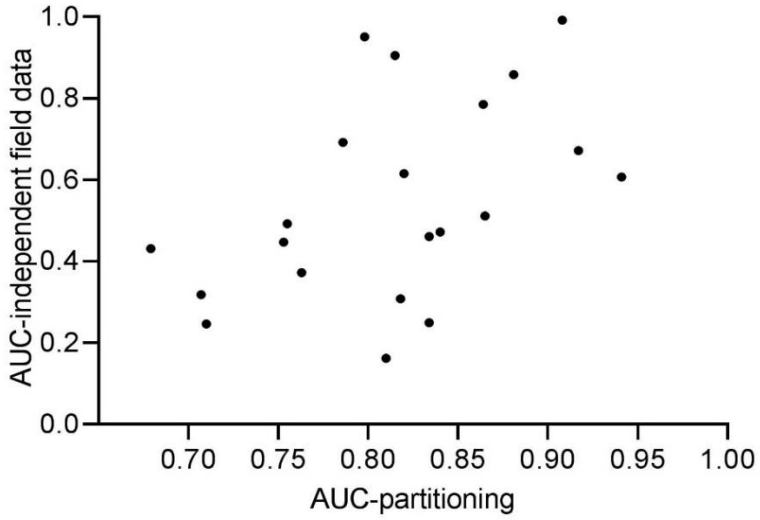


Figure 4.3: The relationship between two measures of model accuracy for Saudi models of reptile distributions ($r = 0.5$, $p < 0.05$), one derived from data-partitioning approach and one derived from testing against new field data, for 21 terrestrial reptile species across 22 visited sites summer 2017 & 2018.

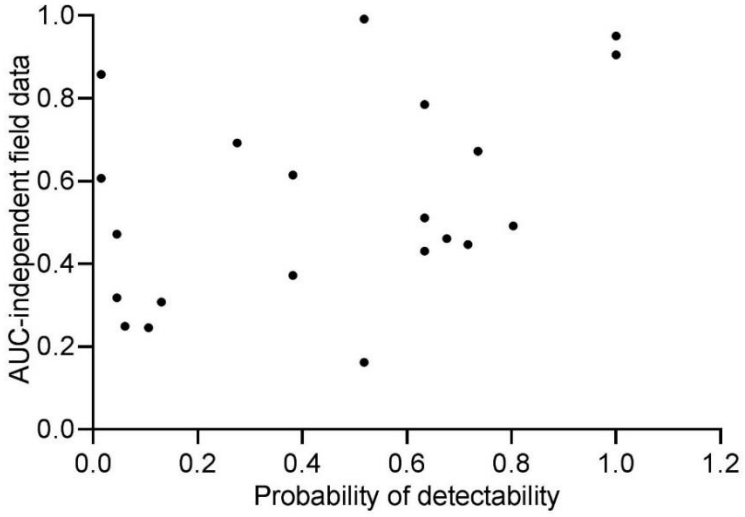


Figure 4.4: The relationship between the estimated probability of species detection at a site ($r = 0.4$, $p = 0.07$), given that the species is present, and accuracy for Saudi distribution models, as estimated from new field data, for 21 species of reptile across 22 visited sites during summer 2017 & 2018.

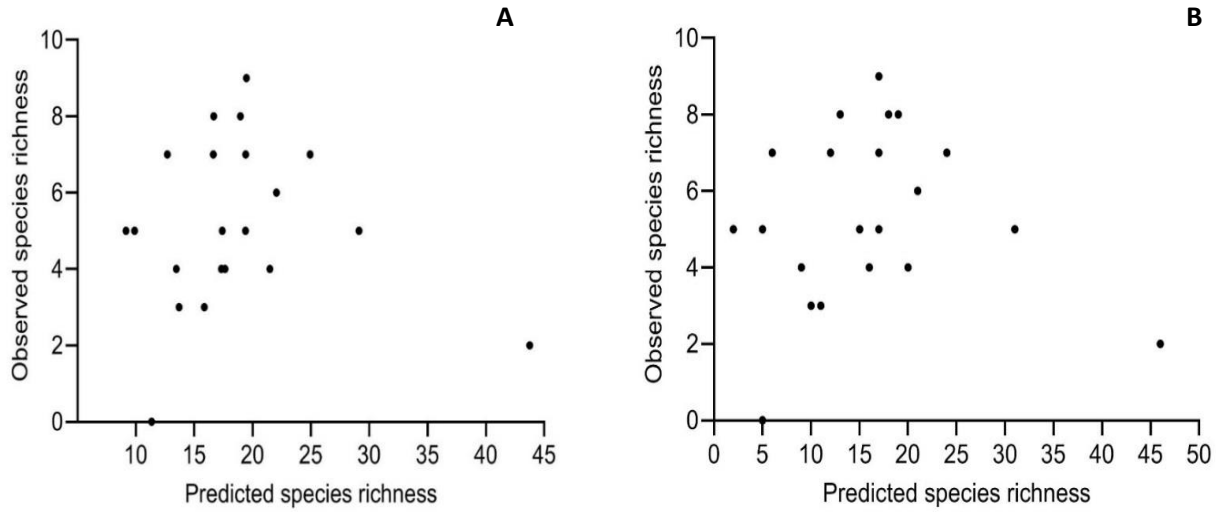


Figure 4.5: The correlation between the unthresholded (A) (r_s = 0.20, p > 0.05) and the thresholded (B) (r_s = 0.21, p > 0.05) predicted species richness from Saudi model and actual species richness estimated from field work conducted in Tabuk Province in summer 2017 and 2018 across 22 visited sites.

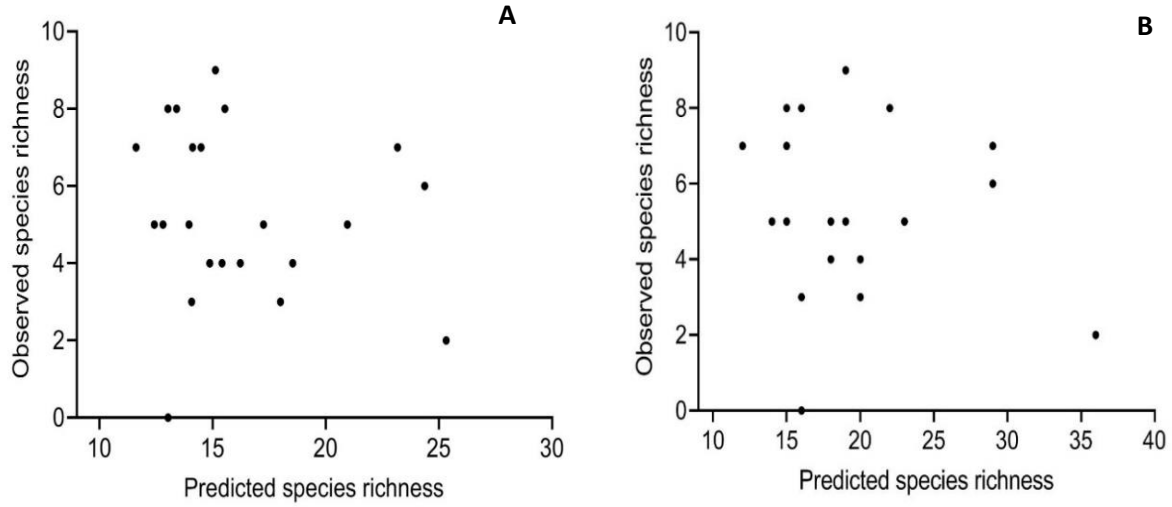


Figure 4.6: The correlation between the unthresholded (A) (r_s = -0.21, p > 0.05) and the thresholded (B) (r_s = -0.22, p > 0.05) predicted species richness from the transferred Egyptian model and actual species richness estimated from field work conducted in Tabuk Province in summer 2017 and 2018 across 22 visited sites.

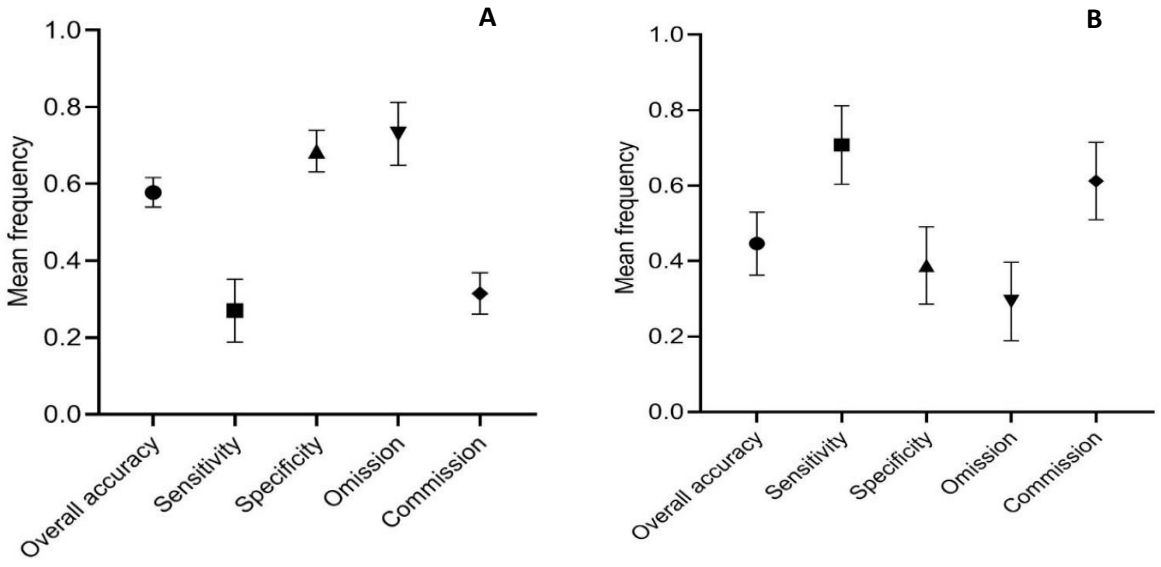


Figure 4.7: The median, range and quartiles of different statistics obtained from the confusion matrix reflecting the accuracy of predictions of distribution models built with Saudi museum data (A), and with transferred Egyptian museum data (B) when compared with estimated species occurrence from field surveys in Tabuk Province in the summer of 2017 and 2018.

4.7 Supplementary materials

Table S4.2: Bioclimatic variables (downloaded at 30 sec resolution) with a Variance Inflation Factor (VIF) of less than 10 which were used to build and transfer the reptile distribution models using Maxent. For the full the list of Bioclimatic variables see (https://www.worldclim.org/).

Abbreviation	Name	Range of values and units
Alt	Altitude	-16 − +2961 m
Bio_02	Mean Diurnal Range (Mean of monthly (max temp - min temp))	7.64 − 17.54 °C
Bio_03	Isothermality (BIO2/BIO7) (* 100)	32.44 − 59.45 %
Bio_08	Mean Temperature of Wettest Quarter	5.61 − 33.10 °C
Bio_09	Mean Temperature of Driest Quarter	10.05 − 38.65 °C
Bio_14	Precipitation of Driest Month	0 – 12 mm
Bio_15	Precipitation Seasonality (Coefficient of Variation)	26.9 − 144.8 %
Bio_19	Precipitation of Coldest Quarter	2 – 109 mm
Srad_04	Solar radiation in April	19559−24548 kJ m^{-2} day^{-1}
Srad_06	Solar radiation in June	20044−27563 kJ m^{-2} day^{-1}
Srad_09	Solar radiation in September	19605−22834 kJ m^{-2} day^{-1}

Table S4.3: The fifteen locations that were visited during field work summer 2017 & 2018 to validate the spatial predictions of two types of species distribution models (i.e., Saudi and Egyptian models). Sites were randomly selected from areas in which both model predictions for species richness were either in the highest or lowest quintile.

Site name	Saudi prediction category	Egyptian prediction category	Observed species richness
Wadi Rumaihiah	High	High	7
Near Al-Amoud Village	High	High	6
Road to Sharma	High	High	5
South of Haqeel	High	High	2
Near Alagaan village	High	Low	8
Al-harra	High	Low	7
Al-Koutib sand dunes	High	Low	5
Road to Deba	Low	High	5
Abo Al-Ajaj village	Low	High	3
Near Ashwaq village	Low	High	4
Road to Al-Qaleebah	Low	Low	5
Al-Beer	Low	Low	0
Al-Oyaynah	Low	Low	3
Sheqrei	Low	Low	7
Road to Al-Disa	Low	Low	5

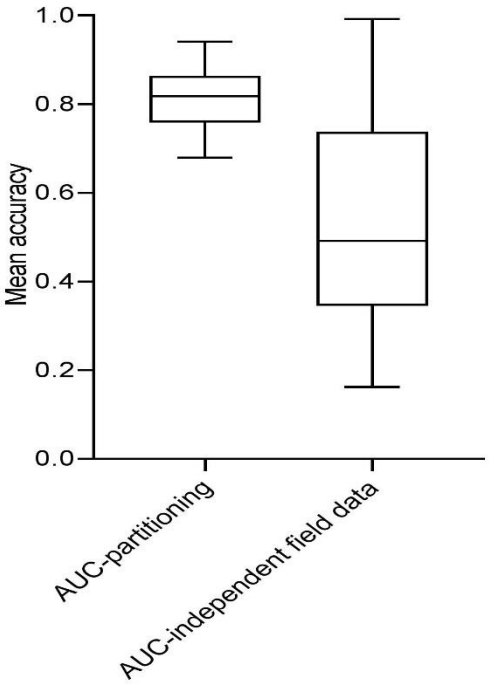


Figure S4.8: Mean module accuracy as estimated using partitioning of Saudi data and new independent field data.

5 Chapter Five: Measuring the effectiveness of Saudi Arabian protected area network in conserving terrestrial reptile species richness

Abstract

In response to the increased global awareness of conserving and maintaining biodiversity, a remarkable expansion of protected areas has been noticed. It is critically important that spatial prioritization planning of protected areas (PAs) be assessed to ensure effectiveness to meet their expected objectives, especially in fragile habitats where ecological information is scarce. For the first time in Saudi Arabia (to my knowledge), I applied two techniques to evaluate and measure the effectiveness and conservation value of current and proposed Saudi's protected areas using terrestrial reptile species richness as case study by i) calibrating species distribution models using Maxent to create habitat suitability maps, and then ii) prioritizing important conservation areas using Zonation. I compared several ways of prioritising sites using Zonation, incorporating a human influence index (HII) and uncertainty analysis, in order to identify robust conservation solutions. I found that there is substantial variation in the location of the areas predicted to be most important for reptile conservation, depending on the basic assumptions of the selection algorithm used in Zonation. However, across the various models there is a good agreement in many of the areas identified as being important. Importantly, I found that the conservation value of both the current and recently proposed PAs was significantly better than the areas outside them. I conclude that examining ongoing and proposed conservation actions using distribution models combined with socioeconomic data and uncertainty analysis is a promising approach to conservation prioritization planning, especially in situations where the detailed data are in short supply.

Keywords

Maxent, Saudi Arabia, socioeconomic data, spatial conservation prioritization, species distribution modelling, uncertainty analysis, Zonation

5.1 Introduction

Biodiversity continues to decline globally at an unprecedented rate (Butchart *et al.*, 2010) negatively affecting ecosystem function (Chapin *et al.*, 2000), and triggering regional, national, international efforts to halt this decline (Di Minin *et al.*, 2017). A good understanding of species distributions is critically important to conserve biodiversity features effectively (Tsoar *et al.*, 2007; Pardo *et al.*, 2013), but much remains to be done to achieve this (Mora *et al.*, 2011; Ficetola *et al.*, 2013). Species distribution modelling has been successfully applied to fill gaps in our knowledge regarding species distributions in different parts of the globe (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005; Pearson, 2008; Elith and Leathwick, 2009) and effectively utilized in spatial conservation prioritization planning (Leach *et al.*, 2013; El-Gabbas *et al.*, 2016; Kaky and Gilbert, 2019).

Spatial prioritization planning is a core concept in biological conservation science (Margules and Pressey, 2000; Moilanen *et al.*, 2005). The concept involves using biogeographic and economic analyses to identify important suitable areas to which resources can be allocated to conserve biodiversity features in such a way as to achieve high conservation value at minimal cost (Moilanen *et al.*, 2005; Kukkala and Moilanen, 2013; Di Minin *et al.*, 2014; Moilanen *et al.*, 2014). Because the drivers of conservation action vary depending on circumstances, each conservation planning exercise has different objectives. Moilanen *et al.*, (2006a) listed four possibilities in the reserve-selection algorithms used to identify the top targets for important areas with the least expensive solutions. For example, one objective of reserve selection is protecting the largest number of different biodiversity

features for the least resource. The result in each case is a grid of cells ranked by conservation importance that, along with literature and expert knowledge, can guide systematic conservation planning to prioritize action (Kukkala and Moilanen, 2013). In this process, conservationists and stakeholders can work closely together using ecological and socio-political data to plan actions which ensure long-term biodiversity protection (Margules and Pressey, 2000; Kukkala and Moilanen, 2013; Moilanen *et al.*, 2014; Sinclair *et al.*, 2018).

In most real cases, biodiversity protection forces involved parties to work under constraints. Pursuing a *robust* optimal reserve is typically preferred because it can minimize the impact of errors in the input data (Moilanen *et al.*, 2006b). Indeed, a robust sub-optimal solution is preferred over an optimal solution with no robustness (Moilanen *et al.*, 2006b). The robustness of conservation prioritization planning can be increased by including socio-economic data in prioritisation algorithms (Kaky and Gilbert, 2019) and reducing uncertainty in the inputs (Moilanen *et al.*, 2006a, b; Moilanen *et al.*, 2014; Dunn *et al.*, 2016). The use of socio-economic data can enable the selection of suitable areas with least human impact on ecosystem services (Kaky and Gilbert, 2019), while accounting for uncertainty can generate conservation decisions which have a stronger chance of success (Moilanen *et al.*, 2006a, b; Moilanen *et al.*, 2014; Dunn *et al.*, 2016). Using both is ideal when robust planning is needed.

One of the most effective ways to maintain biodiversity over space and time is by establishing a protected area network (hereafter PAs) (Margules and Pressey, 2000; Chape *et al.*, 2005; Jenkins and Joppa, 2009; Woodley *et al.*, 2012; Saura *et al.*, 2017; UNEP-WCMC *et al.*, 2018). Well established and managed PAs are effective in conserving species diversity, ecosystem function, particular endangered species, genetic diversity and the sustainability of habitats (Bruner *et al.* 2001; Saura *et al.*, 2017). Thus, PAs have received global political attention and their establishment is a key target of the Convention on Biological Diversity (CBD). In the last two decades, as a result of this target there has been a remarkable

expansion of PAs (Jenkins and Joppa, 2009; UNEP-WCMC *et al.*, 2018) to cover approximately 14.9% and 7.3% of land surface and earth's oceans, respectively (UNEP-WCMC *et al.*, 2018). This figure is subject to continuous change as more areas become protected; there is also the opposite process as some PAs are regrettably downgraded, downsized or degazetted (known as PADDD: Mascia and Pailler, 2011; Woodley *et al.*, 2012; Watson *et al.*, 2014; UNEP-WCMC *et al.*, 2018). In 2010 the parties to the CBD adopted a new Strategic Plan for Biodiversity (2011-2020), including 20 ambitious targets known as the Aichi Biodiversity Targets (CBD COP, 2010). For conservationists, special attention is given to Aichi Target 11 because it explicitly deals with the target that PAs cover 17% of terrestrial and inland water surfaces and 10% of coastal and marine surfaces by 2020 (CBD COP, 2010; UNEP-WCMC *et al.*, 2018); individual countries plan to cover 3% to 50% of their landscapes (Butchart *et al.*, 2015). Evaluating the effectiveness of PAs nationally and internationally in achieving this target is complicated and needs a lot of theoretical and practical information (Butchart *et al.*, 2015; Dunn *et al.*, 2016).

Designing and placing PAs should ideally be based on science to ensure ecological quality and effectiveness (Tang *et al.*, 2011; Woodley *et al.*, 2012) in the face of unequal spatial distribution and expansion of PAs, partly dependent on location characteristics (realms, ecoregions, biomes: Jenkins and Joppa 2009). Uncertainty in the spatial distribution of species might affect the establishment of PAs in the correct places (Dunn *et al.* 2016). Butchart *et al.* (2015) suggested that the recent expansion of PAs might not have been by design, and hence the expected conservation goals may not have been achieved.

Using species distribution modelling in conjunction with a spatial conservation prioritization algorithm forms a robust technique that allows conservationists to assess current and/or proposed networks of PAs, and to identify important areas for conservation planning. Various studies in different geographical areas using various taxa have applied this

combined technique: examples have involved Egyptian biodiversity (Leach *et al.*, 2013; El-Gabbas *et al.*, 2016; Kaky and Gilbert, 2019), Himalayan Galliformes (Dunn *et al.*, 2016), plants in China (Zhang *et al.*, 2012; Wan *et al.*, 2014), birds in central and southern America (Prieto-Torres *et al.*, 2018), and fish and invertebrates in part of the United States (Wenger *et al.*, 2009). One manifest advantage is that the approach can be used for remote and/or little-studied regions, such as developing countries where the cost in time and resources required to obtain many more data is large.

The present study aims to use distribution models in combination with spatial conservation prioritization to plan and evaluate PAs for terrestrial reptiles in Saudi Arabia. This is the first country-wide study of its kind in Saudi Arabia. In 2001, Saudi Arabia became a signatory to the Convention of Biological Diversity, thus promising to ensure the protection of its biodiversity (AbuZinada *et al.*, 2004). The majority of PAs in Saudi Arabia were established specifically to conserve particular endangered species, such as the Arabian oryx (*Oryx leucoryx*) and Houbara bustard (*Chlamydotis undulata*). Their more general effectiveness in conserving other aspects of biodiversity, such as terrestrial reptiles, is not yet clear. Here, I build species distribution models using Maxent (Phillips *et al.*, 2006) and implement conservation prioritization using Zonation (Moilanen *et al.*, 2014) to assess the effectiveness of current and proposed PAs managed by the Saudi Wildlife Authority in protecting terrestrial Saudi reptile species. Overall, I question whether the current and proposed PAs in Saudi Arabia are an effective way to conserve terrestrial reptile species richness.

5.2 Methods & Materials

5.2.1 Study area and presence records

Covering an area of 2,149,690 km^2, Saudi Arabia constitutes three quarters of the Arabian Peninsula (Figure 5.1) (Vincent, 2008), with a list of 103 recorded reptiles (at least 85 of them terrestrial lizards and snakes) (AbuZinada *et al.*, 2004). Being both tropical and sub-tropical, there is a wide variation of topographical and climate conditions (AbuZinada *et al.*, 2004; Vincent, 2008). The predominant habitat in Saudi Arabia is hyper-arid open desert, with very limited vegetation occurring only in particular locations.

I used historical data from museums and relevant literature to collate a dataset of records of terrestrial reptiles. I then carried out an extensive geo-referencing of the locations, deleting any records of uncertain locality. As a result, a total of 62 reptile species (41 lizards and 21 snakes) with 3,943 records were modelled.

5.2.2 Environmental variables

Environmental variables were downloaded from WorldClim Version 2.0, and included 19 bio-layers (describing temperature and precipitation) and 12 indices of solar radiation; altitude was downloaded from WorldClim Version 1.4 at a resolution of 30 sec (~ 1 km^2) (www.worldclim.org) (Hijmans *et al.*, 2005; Fick and Hijmans, 2017). A stepwise selection of environmental variables to reduce multicollinearity among variables was performed by calculating the Variance Inflation Factor (VIF) using the library *usdm* in R (Naimi *et al.*, 2014; R Development Core Team, 2019). Among the 32 original variables, only 11 with a VIF less than 10 were retained for modelling (Table S5.1). Vegetation cover was ignored because this is exceptionally sparse in Saudi Arabia, and has been shown to be a poor predictor of reptile distributions in similar studies (e.g. Egypt: El Gabbas *et al.*, 2016).

5.2.3 Modelling reptile distribution and their habitat suitability

I created distribution models using MaxEnt Version 3.4.1 (Phillips *et al.*, 2006). Maxent identifies areas that have similar values for environmental variables to those at the locations of recorded occurrences. It is one of the most popular software packages to model species distributions with presence-only data (Merow *et al.*, 2013). I chose features and settings in Maxent that maximized the area under curve (AUC), on the basis of extensive testing. I therefore used: product and hinge features; ten subsampled replicates; 1000 iterations; and the new default *cloglog* output (see Phillips *et al.*, 2017). I portioned the data into 90% training and 10% testing subsets. For each species, Maxent created a final habitat suitability map by taking the arithmetic mean of the ten subsampled replicate maps for each species.

5.2.4 Bias file

Overall, the geographical cover of the records is satisfactory (Figure S5.7). However, as is expected with museum data and unsystematic surveys (Newbold, 2010), the data show an unavoidable bias toward cities, road networks, and easily accessible areas. To minimize the impact of bias and spatial autocorrelation in my model predictions, I employed a target-group bias file (Phillips *et al.*, 2009), created using a Gaussian kernel estimation function of the combined records of all species (in the SDMtoolbox 2.0 in ArcGIS: Brown *et al.*, 2017). The bias file is used in Maxent to promote the selection of more background points from biased locations. It is a common approach to correct sampling bias used in distribution modelling, and is known to improve model accuracy (Phillips *et al.*, 2009).

5.2.5 Spatial conservation prioritisation planning and evaluation

Once the habitat suitability maps had been created, Zonation v.4 (Moilanen *et al.*, 2014) was used to prioritize important areas for conservation planning, evaluating the current and proposed Saudi PA network. Zonation is a reserve-selection algorithm that ranks the entire landscape based on the calculated conservation values of the grid cells (Moilanen *et al.*, 2014). Zonation iteratively removes grid cells (or sets of grid cells) one by one, starting with the least valuable cells, via cell removal rules (Moilanen *et al.*, 2005; Moilanen *et al.*, 2014). Some grid cells are given more weight during the ranking process according to the assigned conservation value of the different species (Moilanen *et al.*, 2014). The top-ranking grid cells represent areas that are important for any potential conservation prioritization planning of biodiversity (Moilanen *et al.*, 2005; Moilanen *et al.*, 2014).

Ranking and assessing the grid cells across the landscape of interest proceeded via two cell removal rules (Moilanen *et al.*, 2014; Di Minin *et al.*, 2014). 'Core-area Zonation' (*caz*) identifies important parts of the landscape where rare species occur. As an alternative, I used the 'additive benefit function' (*abf*), which identifies species-rich areas and gives them more weight. According to Moilanen *et al.* (2014), comparing the results of these two alternative analyses is useful because it helps demonstrate the impact of prioritising diversity or rarity in the conservation planning process. The number of cells removed at a time in each iteration is called the 'warp factor' (Moilanen *et al.*, 2014): I set this to 100 in all my analyses. Even though a lower warp factor may be preferred, it has little influence on the solution when compared with a warp factor of 100 (Moilanen *et al.*, 2014). Setting the warp factor at 100 can help greatly in minimizing run-times with no great difference in the solution (Moilanen *et al.*, 2014).

To ensure landscape connectivity, I applied 'distribution smoothing' as a cell aggregation method. This requires a typical dispersal distance to be specified (Moilanen *et al.*, 2014), which is not known for reptiles in Saudi Arabia. A similar study on Egyptian

reptiles applied a uniform dispersal distance of 1 km to all species (El-Gabbas et al 2016), and I adopted this strategy. Using it, I calculated the dispersal parameter α from the equation provided in Moilanen *et al.* (2014: 97) - the value was 207.

$$\alpha = \frac{2}{Dispersal\ distance\ in\ km\ (1) * \frac{0.00833}{0.860}}$$

In Zonation, each species is given a conservation value to reflect the fact that it is more important to conserve some species than others. I used the global and regional status of species using the IUCN assessments, the regional Red List assessments for the Arabian Peninsula compiled by Cox *et al.* (2012) and endemicity to Arabia. The way these were combined for my species is given in Cox *et al.* (2012: 21). The same weights were used in the different Zonation models because it was not the purpose of this work to explore the impact of different weightings.

For each cell removal rule, I started first with a Zonation "basic model", i.e. without incorporating uncertainty or human impacts (see below).

5.2.6 Uncertainty analysis

Uncertainty in species distributions is expected due to the nature of the existing records, and incorporating this uncertainty into conservation prioritization planning is important (see Moilanen *et al.*, 2006a, b; Moilanen *et al.*, 2014). I used the standard deviation map generated by Maxent from the ten replicates for each species to measure uncertainty. I then applied species discounting analysis by subtracting the standard deviation map from the average habitat suitability map: this creates a map where positive values imply areas with

high certainty of high suitability, used by Zonation in a final run (Moilanen *et al.*, 2014; Dunn *et al.*, 2016).

5.2.7 Socio-economic data

To prioritize areas with low human impacts, I used the Human Influence Index (HII) v2 (years 1995 – 2004) as a 'cost layer' in Zonation, downloaded from the Socioeconomic Data and Applications Center (SEDAC), hosted by Center for International Earth Science Information Network (CIESIN), Columbia University. This index is generated by combining nine different global layers believed to influence terrestrial ecosystem services, including human population density, built-up areas, roads, railroads, rivers, coastline, land use/land cover, night-time light (WCS, 2005). The cost layer gives low weight to grid cells where high human influence occurs during the cell ranking/removal process (Figure 5.2) (Kaky and Gilbert, 2019); it was only used with the *abf* removal rule rather than with *caz* because it may jeopardize the overall performance of Zonation using *caz*, producing illogical solutions (Moilanen *et al.* 2014).

5.2.8 Identifying important conservation areas

Zonation generates a map that shows low (0) to high (1) conservation value for each combination of settings. I had eight combinations of settings, each of which produces a map (flow-chart in Figure S5.8). I considered the top 20% of grid cells as the most important parts of the landscape for conservation, and identified areas on the eight maps where they agreed on the top 20%. I also used the 'performance curve' for each map, which plots the average proportion of the distribution of each species which is protected, as increasing percentages of the landscape are protected, weighted by conservation importance (Moilanen *et al.*, 2014). In

one of the scenarios in each technique (*abf* and *caz*), the current PAs were forced to be in the top fraction of important conservation areas. This is the recommended practice to allow for existing PAs when the aim is to identify new important areas for conservation that complement the existing list (Moilanen *et al.*, 2014). The strength of this technique is that it can be used to direct where the next efforts should be focussed (Moilanen *et al.*, 2014).

I considered several Zonation models in this assessment (Figure S5.8). The first model was the basic model (no layers included) which a conservation planning study would normally start with. Such a model can be used as a reference with which to compare other scenarios/models (Moilanen *et al.*, 2014). The second model accounting for current PAs is a recommended conservation practice, widely used when the aim is to identify potential new PAs (Moilanen *et al.*, 2014). The third model incorporates the Human Influence Index in identifying the top priority conservation areas, a new technique for Saudi Arabia. It allows Zonation to avoid areas that are undesirable because they are already in use by humans, and focuses purely on areas with high potential for conservation with the least possible human impact. Such a scenario can also be used to direct new development and urbanization schemes away from identified targeted areas. The fourth model uses uncertainty analysis to increase the robustness of solutions generated by Zonation (Moilanen *et al.*, 2014). This is a very useful technique, especially given the uncertainty of habitat suitability maps. Finally, combining the human index and uncertainty layers takes into consideration the possible influence of these layers simultaneously, again trying to identify a robust solution.

5.2.9 Assessing the Saudi Protected Area network

The shapefile of Saudi PAs was downloaded from the World Database on Protected Areas (UNEP-WCMC, 2019), which was given to the database by the Saudi Arabian National Commission for Wildlife Conservation and Development. In this study, I evaluated

the effectiveness of the 14 current (about 4% of land area of Saudi Arabia) and 21 proposed (about 4.6%) PAs managed by the Saudi Wildlife Authority (SWA) (Table S5.2), excluding two tiny island PAs (one current and one proposed) where nothing is known of their terrestrial reptiles. I created a 50-km buffer around each PA to allow a paired comparison of conservation value inside the PA vs. outside (in the buffer). The mean difference (inside - outside) in value for each combination of Zonation parameters allowed us to compare performance. I compared each PA with its immediate surroundings because of the massive extent of the study area. To compare with the entire country, with its different bioregions and habitat conditions (Al-Saadon and SWA, 2012), and in view of the very different sizes of the PAs (e.g., Raydah 9.33 km^2; Al-Khunfah 19,339 km^2; and Al-'Uruq al-Mu'taridah 50,414.15 km^2), might lead to conclusions which reflect only very large scale geographic factors, and underplay the importance of locally/regionally significant parts of the fauna. Comparing each PA with a surrounding buffer is believed to give a better reflection of the true effect of the network.

5.3 Results

5.3.1 Conservation prioritization using additive benefit function

Prioritizing areas using *abf* as the cell removal rule in the basic model generally emphasized the central plateau, the coastline and its surroundings, and a few other scattered locations, as the best areas for conserving terrestrial reptile (Figure 5.3a). About 36% of terrestrial reptile conservation value (i.e., the scaled ranked Zonation values) is predicted to be conserved when 20% of the landscape is protected (Figure S5.9a). The forcible inclusion of the current PAs into the top fraction slightly altered the locations of top-ranking areas across the landscape,

but the overall pattern remained similar (Figure 5.3b). Thus it had no great impact on the conservation value remaining conserved within the top 20% areas (Figure S5.9b).

Using the Human Influence Index (HII) as a cost layer identified substantially different areas for conservation prioritization when compared with the *abf* basic model. Under this model, large parts of the north of Saudi Arabia were highly ranked, while the conservation value of the coastline and central plateau were noticeably reduced (Figure 5.3c). About 31% of terrestrial reptile conservation value is predicted to be conserved when protecting 20% of the landscape (Figure S5.9c).

Applying *abf* whilst allowing for uncertainty in distributions removed most of the central plateau from the best areas for conservation, while still emphasizing the coastline and surroundings (Figure 5.3d). Furthermore, the right-hand bottom corner of the Empty Quarter desert was newly identified as an important area for conservation. The prediction was that about 32% of terrestrial reptile conservation value is predicted to be conserved when protecting 20% of the landscape (Figure S5.9d).

Incorporating uncertainty and the HII simultaneously dramatically changed the location of the 20% of the landscape with the highest conservation priority, highlighting the southern parts of the Empty Quarter desert, and western and eastern areas of the north, especially Tabuk Province (Figure 5.3e). Most of the coastline except that in Tabuk Province was removed from the top 20%. Approximately 25% of terrestrial reptile conservation value is predicted to be conserved when protecting 20% of landscape (Figure S5.9e).

There are several areas of Saudi Arabia which all five *abf* Zonation models agreed were in the top 20% of landscape ranked by conservation priority (see Figure 5.3f). These areas are located next to the eastern and extreme north-western coastal areas, and in the south-western mountain chain and adjacent interior desert. There are also few fragmented locations in the central north.

5.3.2 Conservation prioritization using core-area Zonation

Prioritizing areas using *caz* as a cell removal rule with the basic model shrunk the areas around the coastline included in the top 10% priority, and increased central and northern areas in the top 20%, relative to the results from using the *abf* rule (Figure 5.4a). About 33% of terrestrial reptile conservation value is predicted to be conserved when protecting the top 20% of the landscape (Figure S5.10a). The forcible inclusion of the current PAs into the top fraction slightly altered the location of the top-ranking areas across landscape, but again a similar pattern was still evident (Figure 5.4b). Thus it had no great impact on the conservation value remaining conserved within the top 20% areas (Figure S5.10b).

When uncertainty in distributions was included with the *caz* rule, areas in the top 20% were changed substantially in comparison with the basic model with *caz* (Figure 5.4c). The conservation value of the coastal areas was greatly reduced, with more inland areas identified, starting from north-western down to south-western parts of the landscape. This technique did not suggest sand dune deserts as important areas, with the exception of a few scattered small areas in the extreme south of the Empty Quarter. Protecting 20% of the landscape is predicted to preserve about 29% of the terrestrial reptile conservation value (Figure S5.10c).

The overall areas of conservation prioritization where the three *caz* Zonation models agreed about the top 20% are shown in Figure 5.4d. Briefly, these small areas were located near the coasts of the extreme north-west, the extreme south-west, and eastern areas. There were also more scattered areas across the interior of central and northern Saudi Arabia that were identified by all of the *caz* models than by the *abf* models.

5.3.3 Current & proposed PAs network

The relative performance of the current PAs overall is shown in Figure S5.11. Their size was positively correlated but not significant with their relative performance (Spearman's rank correlation: $r_s = 0.4$, n = 14, $p > 0.05$). The averaged mean conservation value across all Zonation models inside the current PAs was significantly higher than outside them in the adjacent buffer zone (paired t-test, $t_7 = 2.6$, $p < 0.05$). The differences inside and outside the current PAs were positive for all Zonation runs (Figure 5.5).

The overall relative performance of the proposed PAs is shown in Figure S5.12. Their size was negatively but not significantly correlated with their relative performance ($r_s = -0.2$, n = 21, $p > 0.05$). The mean conservation value across all Zonation runs for the proposed PAs was significantly higher inside than in adjacent buffer zones ($t_7 = 15.8$, $p < 0.001$), and the differences (inside - outside) were positive for all Zonation runs (Figure 5.6).

5.4 Discussion

The role of Protected Areas in conserving biodiversity in vulnerable habitats cannot be ignored. In 1987 the first Saudi PA was established (Table S5.2), and in 2010 a new revised Protected Area System Plan was announced that aims to protect up to 10.42% of the landscape, covering most bioregions within Saudi Arabia (Al-Saadon and SWA, 2012). To date, there has been no evaluation of these current and proposed PAs for protecting any aspect of biodiversity using species distribution modelling and reserve-selection algorithms, and therefore assessing their effectiveness to meet their expected objectives is critically important. In this study, I found that there is substantial variation in the location of the areas predicted to be most important for reptile conservation, depending on the basic assumptions

of the selection algorithm used in Zonation. However, across the various models there is good agreement in many of the areas identified as being in the top 20% of importance. Importantly, I found that the conservation value of both the current and the proposed PAs was significantly better than the areas outside them, suggesting that the PAs are situated in suitable places. In the absence of reliable quantitative distribution modelling and spatial prioritization analysis for other taxonomic groups in Saudi Arabia, I consider predictions of conservation value based on reptile data to be very valuable. In Saudi Arabia, reptiles are the second-largest vertebrate group, and due to their abundance, diversity, adaptation to local environment, and wide distribution, they are likely to be ecologically important in many ways.

5.4.1 Robust conservation prioritization planning

This is the first conservation planning study for Saudi Arabia to take into account uncertainty and human impacts to produce a robust conservation prioritization solution. Using socio-economic data, which include land use and other layers, to evaluate and orient conservation actions is not a new conservation practice (Faleiro and Loyola, 2013; Di Minin *et al.*, 2016; Di Minin *et al.*, 2017; Kaky and Gilbert, 2019, and the references therein). Ideally, targeting areas with the least human impact is preferable whenever possible, because this maximizes the likelihood that protected areas will incorporate land where the conservation measures are likely to be effective and feasible. Saudi Arabia is a large country that is not densely populated, and thus large areas with low to very low human impact can be identified (Figure 5.2). However, the spatial distribution of habitat suitability for terrestrial reptiles in Saudi Arabia is concentrated around centres of human populations (Alatawi *et al.*, 2020), and it is worrying that such habitats might be irreplaceable if damaged by human

activities: there is thus potential for human-wildlife conflict. Indeed, Cox *et al.* (2012) reported that human resource use in the form of grazing, logging and hunting is considered one of the greatest threat to terrestrial reptiles in the Arabian Peninsula.

My results clearly indicate that the predicted conservation value of grid cells is greatly changed after taking human impacts into account, which strongly suggests that local and national wildlife authorities should consider these impacts when planning long-term conservation of terrestrial reptiles, and biodiversity in general. Kaky and Gilbert (2019) reported similar effects on conservation prioritization planning using Egyptian medicinal plants, concluding that the use of human impacts as a cost layer is especially important for little-studied habitats. Di Minin *et al.* (2016) emphasized that accounting for land use can be used by planners at all scales, from local to national, to avoid the risk of potential human-wildlife conflict. It is true that I have found that terrestrial reptiles might be doing better near human settlements (Alatawi *et al.*, 2020), but for conservation planning PAs are more likely to be successful away from human activities to avoid possible conflict, especially for species that are consumed by humans (e.g., *Uromastyx aegyptia*).

Uncertainty in species data may occur because not all areas have been thoroughly surveyed (Naimi *et al.*, 2014), but conservation planning can still be scientific even with uncertainty in the species data (Moilanen *et al.*, 2006b, 2014). In my case, accounting for uncertainty changed the predicted top 20% of areas in both *abf* and *caz* models compared with the basic model. I consider this important because some of these areas are now proposed to be new PAs (e.g., southern parts of the Empty Quarter desert). Dunn *et al.* (2016) reported that even though the inclusion of uncertainty did not change their spatial prioritization much, their results were robust because of its use. Moilanen *et al.* (2006b) recommended using uncertainty in reserve selection because of its efficiency in producing a robust solution in the face of inevitable uncertainties about the input data. Overall, I can say that my maps of the

top-ranked areas represent the current best solution, especially in a data-deficient country such as Saudi Arabia.

It is a recommended practice to apply different techniques (e.g., both *abf* emphasizing species richness, and *caz* emphasizing rarity) to identify areas of high conservation value in common across the landscape (Moilanen *et al.*, 2014). In this study, the different Zonation settings emphasized different prioritized areas, some of which are currently unprotected, or protected but not managed by the Saudi Wildlife Authority (SWA). Lack of strict wildlife regulations (e.g. relating to public accessibility and resource use) may jeopardize the conservation value of these areas. Identified areas in common between the different model outputs are clearly important (Figures 5.3e, 5.4d), and I argue that my robust prioritization has created maps that are useful to planners, stakeholders and conservationists for allocating resources and taking the next steps in conservation.

5.4.2 Conservation effectiveness of Saudi's PAs in conserving terrestrial reptiles

Currently, great attention is focussed by local and national authorities on fulfilling their obligations in conserving biodiversity (Al-Saadon and SWA, 2015; 5th National Report of the Kingdom of Saudi Arabia to the Convention on Biological Diversity, 2014). In this study, my conservation assessment shows that under most models the Saudi PAs succeed in capturing valuable areas for conservation for terrestrial reptiles, which explicitly means that these PAs are in suitable places. However, my model predictions allow us to make three main conservation suggestions for the SWA and other involved partners.

First, I propose the extension of some current (e.g. Ibex Reserve; Nafud al-Urayq; Raydah) and some proposed (e.g. Al-Wajh Bank; Wadi Lajb/Jabal al-Qahar; Wadi Taraj/Jabal Jandaf; Ra's Suwayhil/Ra's al-Qasbah) PAs, and the establishment of others, to encompass more of the highest value areas, and hence the effectiveness of the PA network in

conserving reptiles. I propose the establishment of new PAs to conserve areas the north-western coast and its surroundings: all the models I ran agreed on these areas. If an extension is logistically difficult to establish, a corridor to the next identified suitable important habitat might suffice. The robustness of these recommendations is underscored by the fact that a separate study by Cox *et al.* (2012), which used different non-quantitative methods, broadly identified that similar areas that need to be conserved.

Second, I do not recommend any form of PA downgrading, downsizing, and degazettement (PADDD) at the moment because of the likely negative effects of such actions (Mascia and Pailler, 2010; UNEP-WCMC *et al.*, 2018). Watson *et al.* (2014) reported that all forms of PADDD are regrettably increasing across the globe in both developing (e.g. Arabian Oryx Sanctuary in Oman) and developed countries (e.g. Sites of Special Scientific Interest in the UK) (more details in Frey and Steiner, 2011; Watkins *et al.*, 2014).

Third, my maps of the top-ranked sites identified some important PA areas not currently managed by the SWA, but managed instead by various local and national partners. The level of ongoing conservation action is not yet clear under the management of these partners. I suggest potentially transferring some of these areas, such as the Asir National Park and Jubail Marine Wildlife Sanctuary, to management by the SWA. This might constitute a better strategy to ensure long-term conservation of terrestrial reptiles and other biodiversity. If this were not possible, then at least some kind of agreement is needed to allow wildlife regulations to be applied and regular checks to be made by wildlife conservationists and rangers. Whether managed by SWA or not, to implement all the Aichi targets (especially target 11), PAs should consider all aspects of habitat suitability for terrestrial reptiles and other biodiversity, taking into account the needs of both specialist and generalist species.

5.4.3 Efforts to implement Aichi target 11

Achievement of Aichi target 11 will be difficult, requiring strong commitment from governments and key stakeholders. The target number itself is not based on science (see Woodley *et al.*, 2012), and deciding what percentage of land should be protected typically depends on the taxon and the region being considered, because each has special attributes not applicable when studying other taxa or regions. However, some taxa might need to be used as surrogates for others, especially in the absence of good data. About of 4% of the Saudi landscape is currently protected, managed by SWA, and an extra 4.6% will be protected if planned PAs are created: these figures are designed to cover the various ecological conditions of Saudi Arabia (Al-Saadon and SWA, 2012). There is a promising proposal to join together three large current PAs (Al-Khunfah, At-Tubayq, and Harrat al-Harrah) in the north into a single PA. Al-Saadon and SWA (2012) reported 17 new strategies in the form of a National Biodiversity Strategy and Action Plan to implement a Programme of Work for Protected Areas. It is clear that local and national Saudi wildlife authorities are moving forward to contribute to international conservation efforts, and indeed they have already made significant contributions in protecting endangered species such as the Arabian oryx, Mountain gazelle and Houbara bustard (AbuZinada *et al.*, 2004; 5th National Report of the Kingdom of Saudi Arabia to the Convention on Biological Diversity, 2014).

My conservation priority maps enhance our limited knowledge and fill gaps about the role of PAs in conserving terrestrial reptiles. Of course, the requirements of other taxa with larger home ranges (Arabian oryx) may be different, but until the relevant analyses have been done, it is hard to know. No-one can judge with confidence how large a PA is required to conserve a species effectively until extensive appropriate fieldwork has been carried out both inside and outside the PA (see Al-Saadon and SWA, 2012).

5.4.4 Caveats

I have carefully applied the best possible model combinations to produce robust predictions of the value of land in Saudi Arabia for reptile conservation, yet there are some caveats that might implicitly limit the usefulness of my prioritizations, especially regarding species distribution modelling. First, the records used to build my models come in presence-only format: species can be recorded in habitats that are actually not suitable, or fail to be recorded in suitable habitats. These deficiencies may limit the performance of distribution models (Zaniewski *et al.*, 2002; Guisan and Thuiller, 2005), which in turn might influence the conservation prioritization process. Second, using only environmental variables to model habitat suitability can be an issue. It is possible that where a species is found is affected by other factors such as biotic interactions - we know very little about these for Saudi reptiles. Ultimately, all ecological modelling is limited by the assumptions made, and the quality of the data used to generate predictions. Only further empirical and theoretical study of a wide range of taxa will improve our understanding of the best strategies to conserve Saudi biodiversity in the long-term.

5.5 Conclusion

Much of the world's biodiversity is likely to be lost without active conservation (Bruner *et al.*, 2001). Where and how to conserve species is an important question (Moilanen *et al.*, 2014) which should be based on sound science to ensure its effectiveness. There are obvious deficiencies in data on the terrestrial reptiles of Saudi Arabia regarding their conservation status and distributions that need to be remedied. It is encouraging that my quantitative findings here are supported by the qualitative conclusions of a previous study (Cox *et al.*, 2012). I have demonstrated the utility of combining species distribution modelling with reserve-selection algorithms to assess ongoing and proposed conservation

actions. My conservation prioritization is unique for Saudi Arabia, and I hope it will move conservation science in the country forward.

5.6 References

AbuZinada, A., Robinson, E. R., Nader, I. & Al Wetaid, Y. 2004. *First Saudi Arabian National Report on the Convention on Biological Diversity*. The National Commission for Wildlife Conservation and Development.

Alatawi, A. S., Gilbert, F. & Reader, T. 2020. Modelling terrestrial reptile species richness, distributions and habitat suitability in Saudi Arabia. *Journal of Arid Environments*, 178**,** 104153. https://doi.org/10.1016/j.jaridenv.2020.104153

Al-Saadon, K. H. & Saudi Wildlife Authority. 2012. *Action Plan for Implementing the Convention on Biological Diversity's Programme of Work on Protected Areas*.

Brown, J. L., Bennett, J. R. & French, C. M. 2017. SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ*, 5**,** e4095.

Bruner, A. G., Gullison, R. E., Rice, R. E. & Fonseca, G. 2001. Effectiveness of parks in protecting tropical biodiversity. *Science*, 291**,** 125-128.

Butchart, S. H. M., Clarke, M., Smith, R. J., Sykes, R. E., Scharlemann, J. P. W., et al. 2015. Shortfalls and solutions for meeting national and global conservation area targets. *Conservation Letters*, 8**,** 329-337.

Butchart, S. H. M., Walpole, M., Collen, B., Strien, A., Scharlemann, J. P. W., et al. 2010. Global biodiversity: Indicators of recent declines. *Science*, 328**,** 1164-1168.

CBD COP. 2010. Decision adopted by the Conference of the Parties to the Convention on Biological Diversity at its Tenth Meeting X/2. The Strategic Plan for Biodiversity 2011-2020 and the Aichi Biodiversity Targets. Convention on Biological Diversity Conference of the Parties. Available from https://www.cbd.int/doc/decisions/cop-10/cop-10-dec-02-en.pdf. Accessed 20-08-2019.

Chape, S., Harrison, J., Spalding, M. & Lysenko, I. 2005. Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets. *Philos. Trans. R. Soc. B,* 360**,** 443-455.

Chapin, F. S., III, Zavaleta, E. S., Eviner, V. T., Naylor, R. L., Vitousek, P. M., Reynolds, H. L., Hooper, D. U., Lavorel, S., Sala, O. E., Hobbie, S. E., Mack, M. C. & Díaz, S. 2000. Consequences of changing biodiversity. *Nature*, 405**,** 234-242.

Cox, N. A., Mallon, D., Bowles, P., Els, J. & Tognelli, M. F. 2012. *The Conservation Status and Distribution of Reptiles of the Arabian Peninsula*. Cambridge, UK and Gland, Switzerland: IUCN, and Sharjah, UAE: Environment and Protected Areas Authority.

Di Minin, E., Slotow, R., Hunter, L. T., Pouzols, F. M., Toivonen, T., Verburg, P. H., Leader-Williams, N., Petracca, L. & Moilanen, A. 2016. Global priorities for national carnivore conservation under land use change. *Scientific Reports*, 6**,** 23814.

Di Minin, E., Soutullo, A., Bartesaghi, L., Rios, M., Szephegyi, M. N. & Moilanen, A. 2017. Integrating biodiversity, ecosystem services and socio-economic data to identify priority areas and landowners for conservation actions at the national scale. *Biological Conservation*, 206**,** 56-64.

Di Minin, E., Veach, V., Lehtomäki, J., Pouzols, F. & Moilanen, A. 2014. *A Quick Introduction to Zonation*. C-BIG, the Conservation Biology Informatics Group, at the University of Helsinki University of Helsinki.

Dunn, J. C., Buchanan, G. M., Stein, R. W., Whittingham, M. J. & McGowan, P. 2016. Optimising different types of biodiversity coverage of protected areas with a case study using Himalayan Galliformes. *Biological Conservation*, 196**,** 22-30.

El-Gabbas, A., Baha El Din, S., Zalat, S. & Gilbert, F. 2016. Conserving Egypt's reptiles under climate change. *Journal of Arid Environments*, 127**,** 211-221.

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.

Faleiro, F. V. & Loyola, R. D. 2013. Socioeconomic and political trade-offs in biodiversity conservation: a case study of the Cerrado Biodiversity Hotspot, Brazil. *Diversity and Distributions*, 19**,** 977-987.

Ficetola, G. F., Bonardi, A., Sindaco, R. & Padoa-Schioppa, E. 2013. Estimating patterns of reptile biodiversity in remote regions. *Journal of Biogeography*, 40**,** 1202-1211.

Fick, S. E. & Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37**,** 4302-4315.

Fifth National Report of the Kingdom of Saudi Arabia to the Convention on Biological Diversity. 2014. Riyadh, Saudi Wildlife Authority.

Frey, B. S. & Steiner, L. 2011. World Heritage List: does it make sense? *International Journal of Cultural Policy*, 17**,** 555-573.

Guisan, A. & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8**,** 993-1009.

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

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.

Jenkins, C. N. & Joppa, L. 2009. Expansion of the global terrestrial protected area system. *Biological Conservation*, 142**,** 2166-2174.

Kaky, E. & Gilbert, F. 2019. Allowing for human socioeconomic impacts in the conservation of plants under climate change. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology***,** 1-11.

Kukkala, A. S. & Moilanen, A. 2013. Core concepts of spatial prioritisation in systematic conservation planning. *Biological Reviews*, 88**,** 443-464.

Leach, K., Zalat, S. & Gilbert, F. 2013. Egypt's Protected Area network under future climate change. *Biological Conservation*, 159**,** 490-500.

Margules, C. R. & Pressey, R. L. 2000. Systematic conservation planning. *Nature*, 405**,** 243-253.

Mascia, M. B. & Pailler, S. 2011. Protected area downgrading, downsizing, and degazettement (PADDD) and its conservation implications. *Conservation Letters*, 4**,** 9-20.

Merow, C., Smith, M. J. & Silander, J. A. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36**,** 1058-1069.

Moilanen, A., Franco, A. M., Early, R. I., Fox, R., Wintle, B. & Thomas, C. D. 2005. Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. *Proc R. Soc. B*, 272**,** 1885-1891.

Moilanen, A., Pouzols, F. M., Meller, L., Veach, V., Arponen, A., Leppanen, J. & Kujala, H. 2014. *Zonation – Spatial Conservation Planning Methods and Software*. Version 4. User Manual. C-BIG Conservation Biology Informatics Group Department of Biosciences University of Helsinki, Finland.

Moilanen, A., Runge, M. C., Elith, J., Tyre, A., Carmel, Y., Fegraus, E., Wintle, B. A., Burgman, M. & Ben-Haim, Y. 2006b. Planning for robust reserve networks using uncertainty analysis. *Ecological Modelling*, 199**,** 115-124.

Moilanen, A., Wintle, B. A., Elith, J. & Burgman, M. 2006a. Uncertainty analysis for regional-scale reserve selection. *Conservation Biology*, 20**,** 1688-1697.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. & Worm, B. 2011. How many species are there on earth and in the ocean? *PLoS Biol*, 9**,** e1001127.

Naimi, B., Hamm, N., Groen, T. A., Skidmore, A. K. & Toxopeus, A. G. 2014. Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37**,** 191-203.

Newbold, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography*, 34**,** 3-22.

Pardo, I., Pata, M. P., Gomez, D. & Garcia, M. B. 2013. A novel method to handle the effect of uneven sampling effort in biodiversity databases. *PLoS ONE*, 8**,** e52786.

Pearson, R. G. 2008. Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*. Available at http://ncep.amnh.org.

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E. & Blair, M. E. 2017. Opening the black box: an open-source release of Maxent. *Ecography*, 40**,** 887-893.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190**,** 231-259.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. & Ferrier, S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19**,** 181-197.

Prieto-Torres, D. A., Nori, J. & Rojas-Soto, O. R. 2018. Identifying priority conservation areas for birds associated to endangered Neotropical dry forests. *Biological Conservation*, 228**,** 205-214.

Saura, S., Bastin, L., Battistella, L., Mandrici, A. & Dubois, G. 2017. Protected areas in the world's ecoregions: How well connected are they? *Ecological Indicators*, 76**,** 144-158.

Sinclair, S. P., Milner-Gulland, E. J., Smith, R. J., McIntosh, E. J., Possingham, H. P., Vercammen, A. & Knight, A. T. 2018. The use, and usefulness, of spatial conservation prioritizations. *Conservation Letters*, 11**,** e12459.

Tang, Z., Fang, J., Sun, J. & Gaston, K. J. 2011. Effectiveness of protected areas in maintaining plant production. *PLoS ONE*, 6**,** e19116.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, 13**,** 397-405.

UNEP-WCMC. 2019. Protected Area Profile for Saudi Arabia from the World Database of Protected Areas, August 2019. Accessed at https://www.protectedplanet.net/ on 16-08-2019.

UNEP-WCMC, IUCN & NGS. 2018. Protected Planet Report 2018. UNEP-WCMC, IUCN and NGS: Cambridge UK; Gland, Switzerland; and Washington, D.C., USA.

Vincent, P. 2008. *Saudi Arabia: An Environmental Overview*. Taylor & Francis Group.

Wan, J., Wang, C., Yu, J., Nie, S., Han, S., Zu, Y., Chen, C., Yuan, S. & Wang, Q. 2014. Model-based conservation planning of the genetic diversity of *Phellodendron amurense* Rupr due to climate change. *Ecology and Evolution*, 4**,** 2884-2900.

Watkins, H., Hawkins, K. & Cormack, A. 2014. *HS2: The Case for a Greener Vision*. The Wildlife Trusts.

Watson, J. E., Dudley, N., Segan, D. B. & Hockings, M. 2014. The performance and potential of protected areas. *Nature*, 515**,** 67-73.

WCS. 2005. Wildlife Conservation Society – WCS, and Center for International Earth Science Information Network - CIESIN - Columbia University. 2005. Last of the Wild Project, Version 2, 2005 (LWP-2): Global Human Influence Index (HII) Dataset (Geographic). Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4BP00QC. Accessed 01-08-2019.

Wenger S. J., Hagler M. M. & Freeman B. J. 2009. Prioritizing areas of the Conasauga River sub-basin in Georgia and Tennessee for preservation and restoration. *Southeastern Fishes Council Proceedings*, 51**,** 31-38.

Woodley, S., Bertzky, B., Crawhall, N., Dudley, N., Londoño, J. M., MacKinnon, K., Redford, K. & Sandwith, T. 2012. Meeting Aichi Target 11: what does success look like for protected area systems? *Parks*, 18**,** 23-36.

Zaniewski, A. E., Lehmann, A. & Overton, J. M. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, 157**,** 261-280.

Zhang, M-G., Zhou, Z-K., Chen, W-Y., Slik, J. W. F., Cannon, C. H. & Raes, N. 2012. Using species distribution modeling to improve conservation and land use planning of Yunnan, China. *Biological Conservation*, 153**,** 257-264.

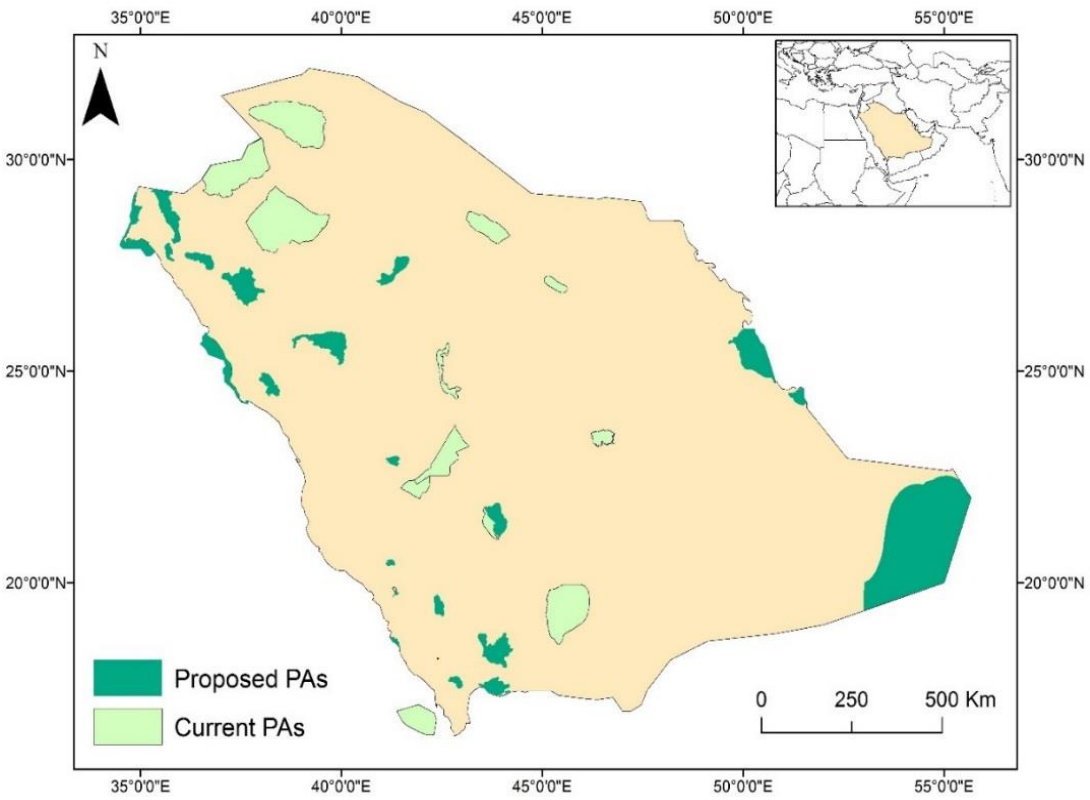


Figure 5.1: Map showing the study area along with proposed (21 dark green) and current (14 light green) Saudi Arabian Protected Areas (PAs) which all are/will be managed by the Saudi Wildlife Authority. For more information about each PA please see Table S5.2 in supplementary materials (source: The World Database on Protected Areas website (UNEP-WCMC, 2019).

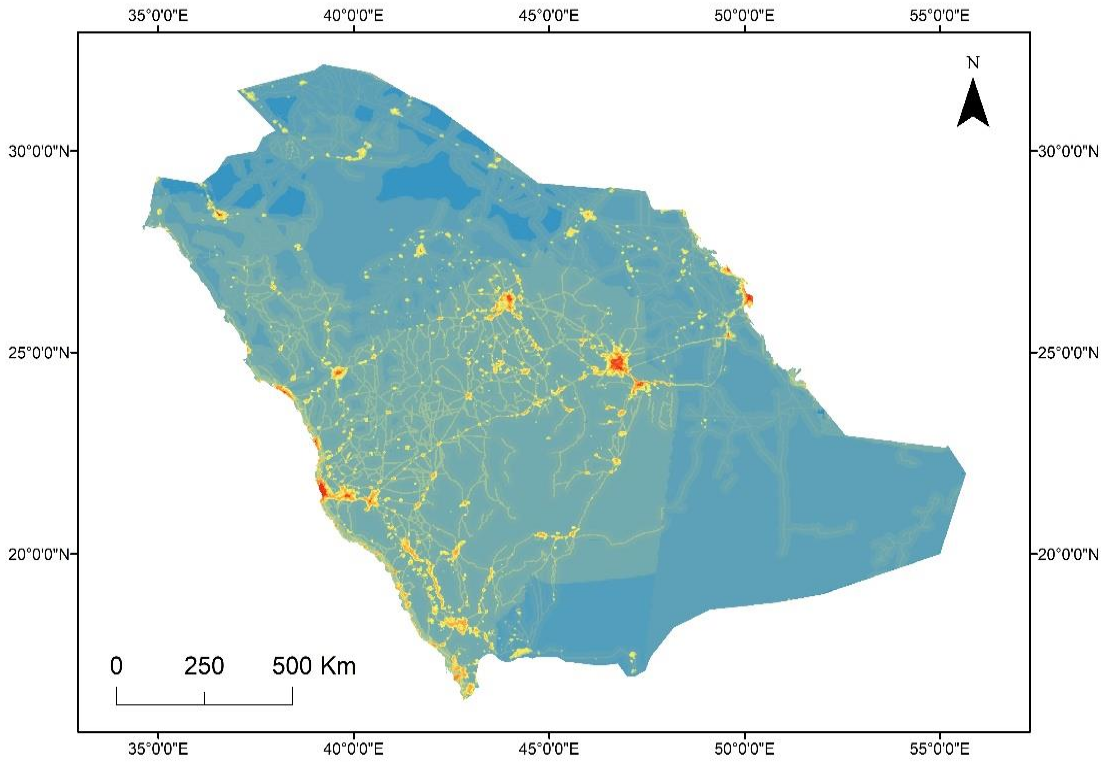


Figure 5.2: The Human Influence Index (HII) across the Saudi landscape. HII was used as a cost layer in Zonation to identify areas with least human impact on ecosystem services by giving these grid cells high weight during the cell ranking process to prioritize these areas for conservation planning. HII layer is created by stacking different layers (population density, built-up areas, roads, railroads, rivers, coastline, land use/land cover, night-time lights). Dark red colour indicates high HII, yellow medium while blue indicates low HII.

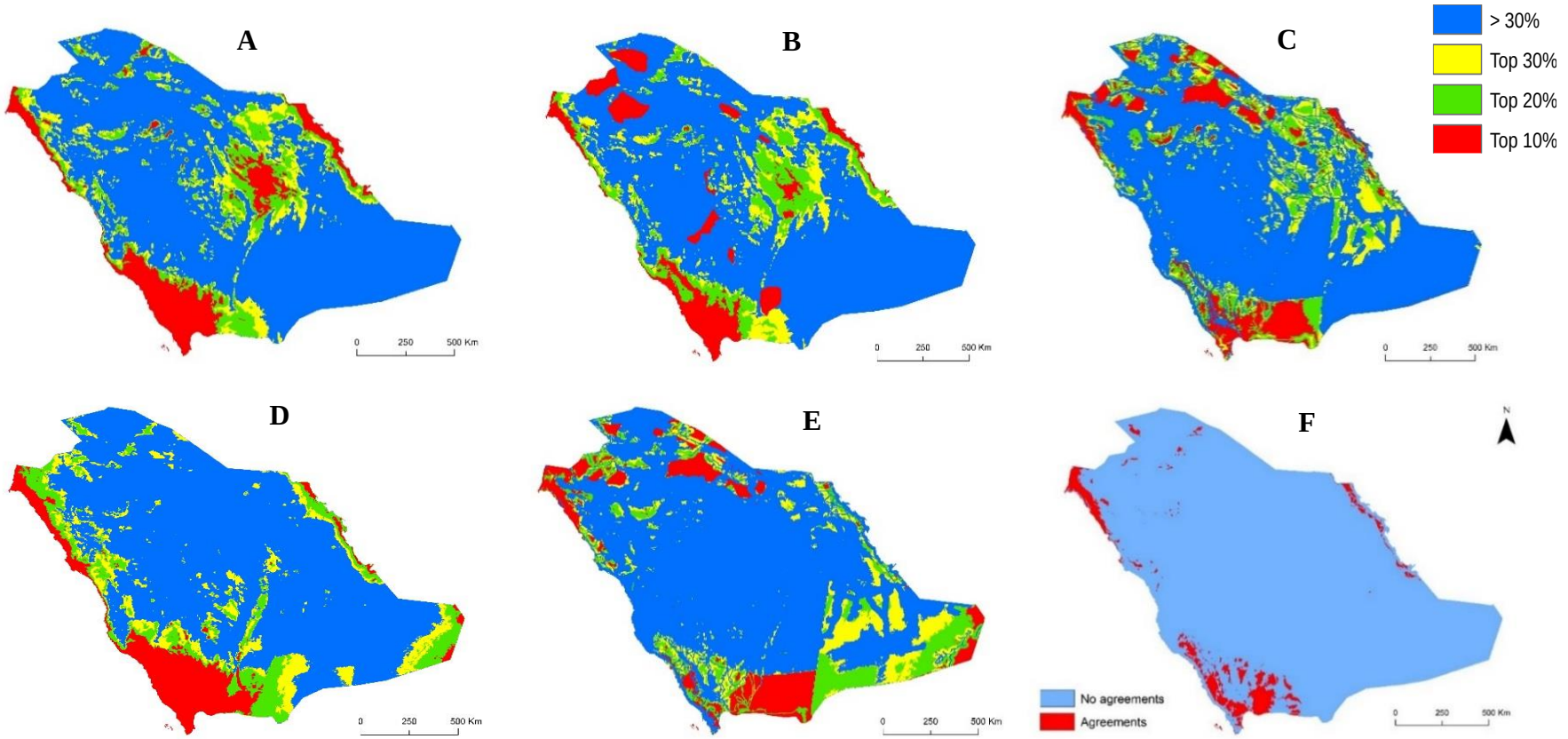


Figure 5.3: Different outputs from the Zonation model run using the *abf* cell removal rule to identify areas for conservation prioritization. On the first five maps, colour indicates the areas prioritized, based on ranked conservation value; i.e. red areas are those which the model identified as being in the top 10% of all areas according to conservation value, and so on. A, *abf* with basic model; B, *abf* with current PAs forced to be the top fraction when ranking grid cells; C, *abf* with a 'cost' layer based on the Human Influence Index; D, *abf* with uncertainty analysis used in form of distribution discounting; E, *abf* with both a cost layer and uncertainty analysis used simultaneously; F shows the agreement (red) and disagreement (blue) between the five different Zonation models on the top 20% of grid cells.

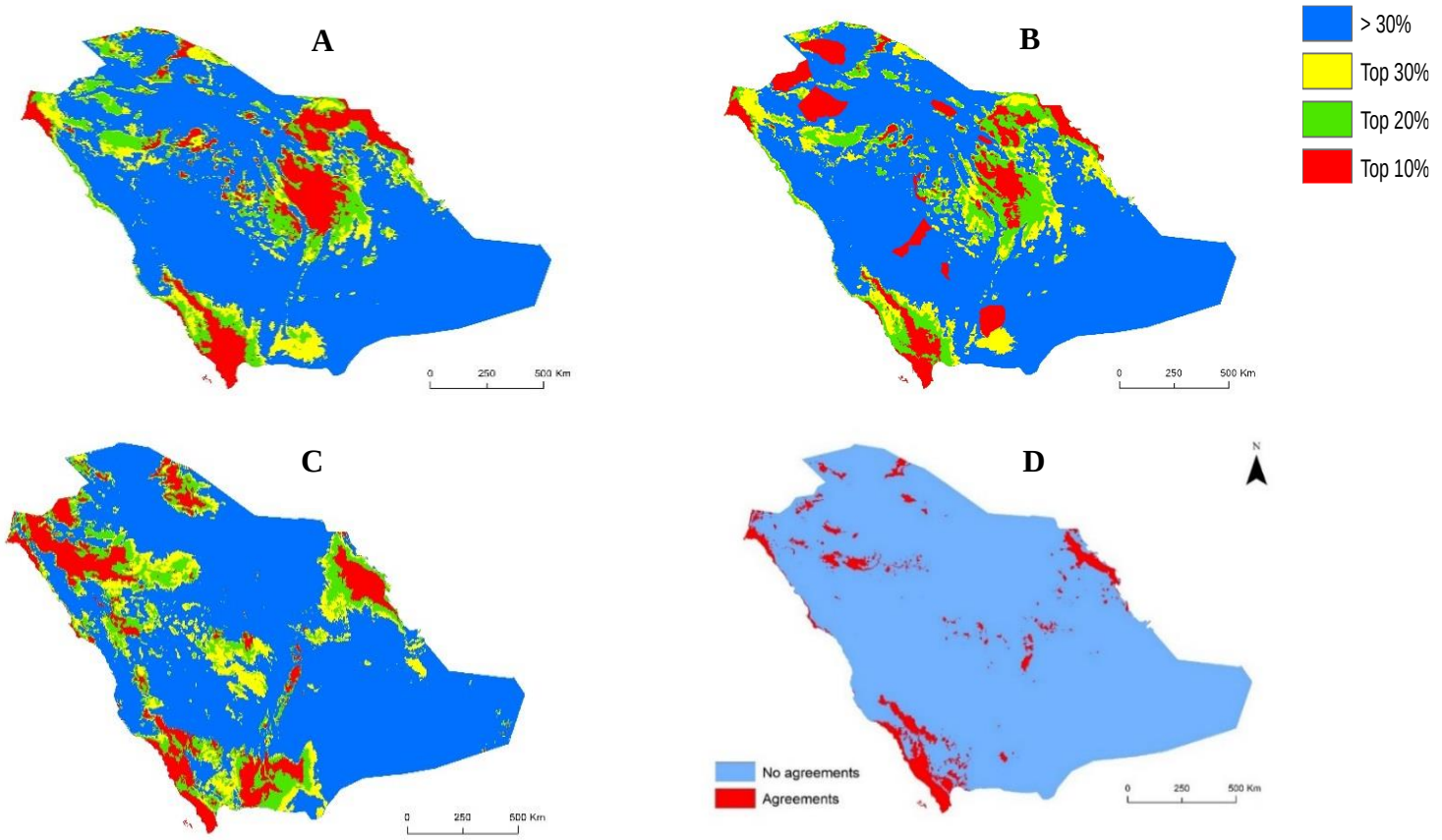


Figure 5.4: Different outputs from the Zonation model run using the *caz* cell removal rule to identify areas for conservation prioritization. On the first five maps, colour indicates the areas prioritized, based on ranked conservation value; i.e. red areas are those which the model identified as being in the top 10% of all areas according to conservation value, and so on. A, *caz* with basic model; B, *caz* with current PAs forced to be the top fraction when ranking grid cells; C, *caz* with uncertainty analysis used in form of distribution discounting; D shows the agreement (red) and disagreement (blue) between the three different Zonation models on the top 20% of grid cells.

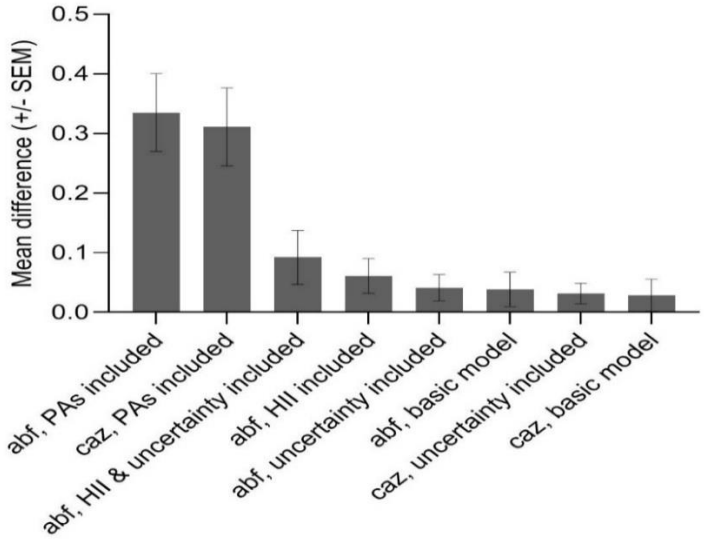


Figure 5.5: Mean differences in conservation value of grid cells inside and outside current PAs using one of two cell removal rules (*abf* or *caz*). *abf* and *caz* cell removal rules were used with basic model, forced inclusion of PAs in the top priority fraction of landscape, and accounting for uncertainty in species distribution. *abf* was also used with HII as a cost layer, and HII and uncertainty analysis simultaneously. Further details are given in the main text.

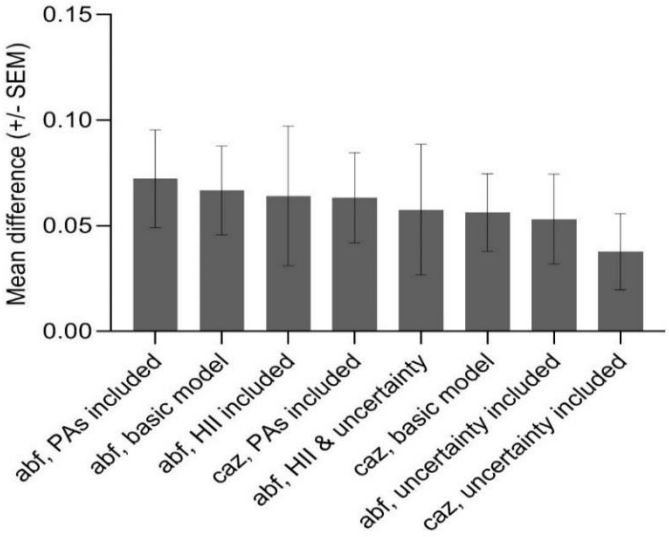


Figure 5.6: Mean differences in conservation value of grid cells inside and outside proposed PAs using one of two cell removal rules (*abf* or *caz*). *abf* and *caz* cell removal rules were used with basic model, forced inclusion of PAs in the top priority fraction of landscape, and accounting for uncertainty in species distribution. *abf* was also used with HII as a cost layer, and HII and uncertainty analysis simultaneously. Further details are given in the main text.

5.7 Supplementary materials

Table S5.1: Bioclimatic variables (downloaded at 30 sec resolution) with a Variance Inflation Factor (VIF) of less than 10 which were used to build reptile distribution models for Saudi Arabia using Maxent. For the full the list of Bioclimatic variables see (https://www.worldclim.org/).

Abbreviation	Name	Range of values and units
Alt	Altitude	-16 − +2961 m
Bio_02	Mean Diurnal Range (Mean of monthly (max temp - min temp))	7.64 − 17.54 °C
Bio_03	Isothermality (BIO2/BIO7) (*100)	32.44 − 59.45 %
Bio_08	Mean Temperature of Wettest Quarter	5.61 − 33.10 °C
Bio_09	Mean Temperature of Driest Quarter	10.05 − 38.65 °C
Bio_14	Precipitation of Driest Month	0 – 12 mm
Bio_15	Precipitation Seasonality (Coefficient of Variation)	26.9 − 144.8 %
Bio_19	Precipitation of Coldest Quarter	2 – 109 mm
Srad_04	Solar radiation in April	19559−24548 kJ m^{-2} day^{-1}
Srad_06	Solar radiation in June	20044−27563 kJ m^{-2} day^{-1}
Srad_09	Solar radiation in September	19605−22834 kJ m^{-2} day^{-1}

Table S5.2: The current and proposed list of protected areas in Saudi Arabia that are/will be managed by Saudi Wildlife Authority. The relevant data and shapefile are downloaded from World Database on Protected Areas (accessed at https://www.protectedplanet.net/ on 16-08-2019) based on information provided by National Commission for Wildlife Conservation and Development, Saudi Arabia.

Name	Area (km^2)	Status	year
Jabal Aja	2202.91	Proposed	1987
Jabal Batharah / Wadi Turabah	316.56	Proposed	1990
Maqla‘ Timiyah / Harrat Kishb	599.24	Proposed	1990
Hisma	3699.29	Proposed	1990
Jabal ad-Dubbagh	628.91	Proposed	1990
Jabal Radwa	1673.02	Proposed	1990
Jibal Qaraqir	1678.26	Proposed	1990
Harrat Khaybar / Wadi Hadiyah	4895.66	Proposed	1990
Harrat ‘Uwayrid	5165.17	Proposed	1990
Jazirat al-Huwaysat / Dawhat Duwayhin*	1223.88	Proposed	1990
Ra’s Suwayhil / Ra’s al-Qasbah*	3705	Proposed	1990
Al-Wajh Bank*	3857.6	Proposed	1990
Makhshush*	267.63	Proposed	1990
Murtafa‘at Najran	2241.69	Proposed	1990
Wadi Tarj / Jabal Jandaf	1119.98	Proposed	1990
Wadi al-Hizmah / Wadi Tathlith	4437.77	Proposed	1990
Wadi Lajb / Jabal al-Qahar	719.85	Proposed	1990
Raydah	9.33	Designated	1989
Al-‘Uruq al-Mu‘taridah	50414.15	Proposed	1990
Extension to Majami‘ al-Hadb	2703.78	Proposed	1987
Khalij Salwa*	7802.48	Proposed	1990
Majami'al-Hadb	3400	Designated	1993
Ibex Reserve	2369	Designated	1988
Jabal Shada al-A‘la	68.62	Designated	2002
Nafud al-‘Urayq	2036.1	Designated	1995
Saja / Umm ar-Rimth	6528.2	Designated	1995
Mahazat as-Sayd	2553	Designated	1988
Al-Jandaliyah	1188.9	Designated	1995
At-Tubayq	12105	Designated	1989
Harrat al-Harrah	13775	Designated	1987
Al-Khunfah	19339	Designated	1987
Extension to Jabal Shada	76.21	Proposed	1990
Farasan Islands*	5408	Designated	1988
At-Taysiyah	4272.2	Designated	1995
‘Uruq Bani Ma‘arid	12787	Designated	1993

* Island and coastal protected areas

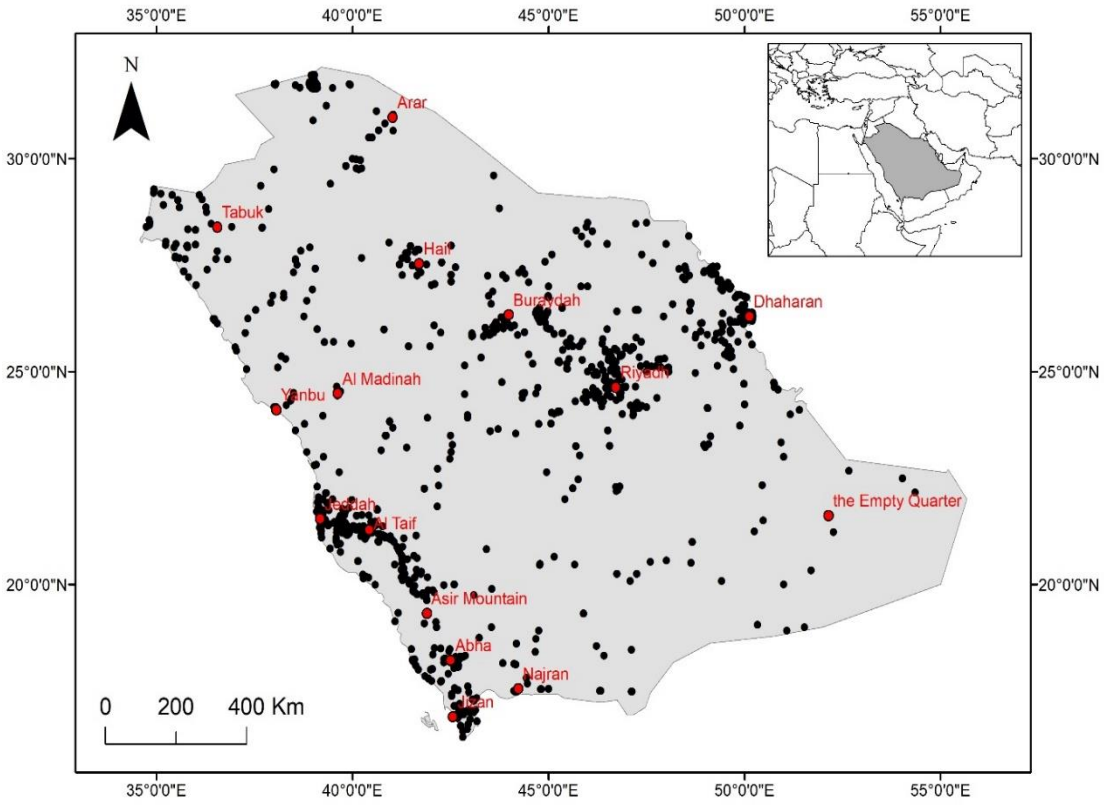


Figure S5.7: The locations in Saudi Arabia of the observations of 62 terrestrial reptile species (n = 3,943 observations in total) that were used to build distribution models.

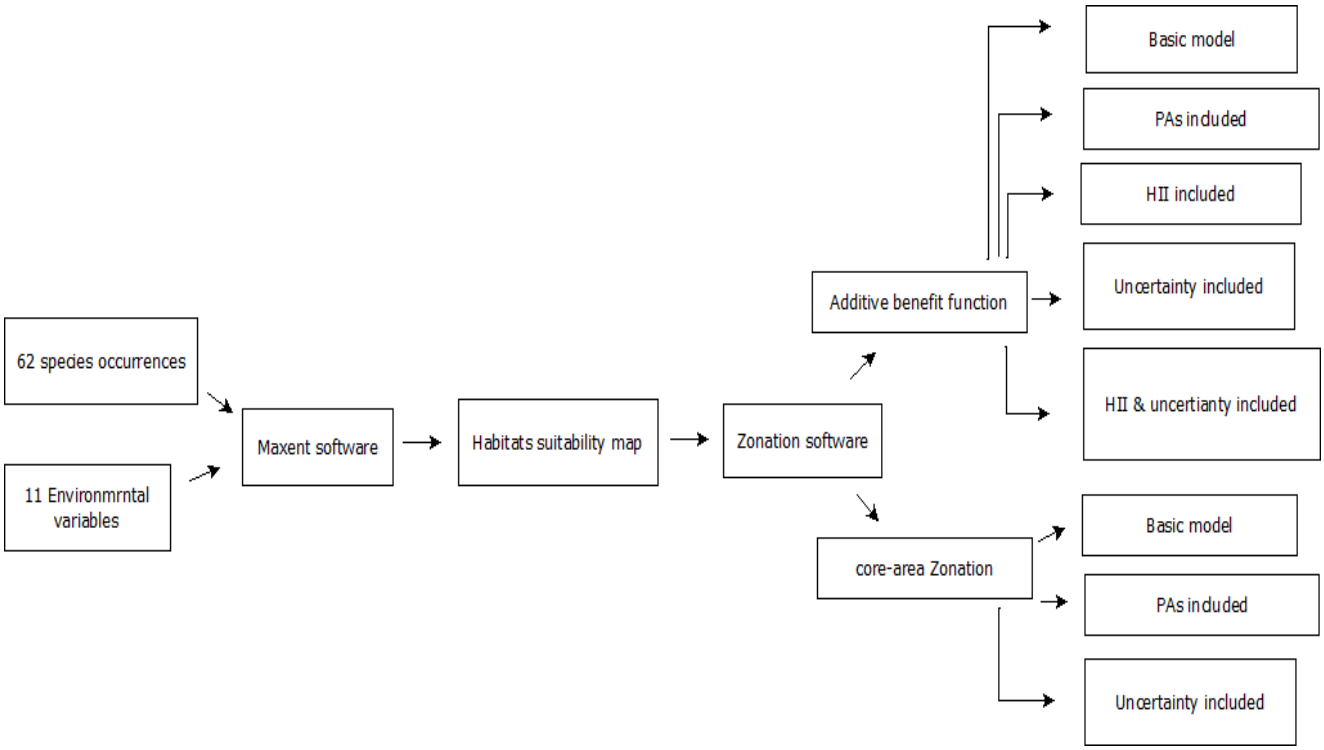


Figure S5.8: A flow-chart shows the process of evaluating Saudi's PAs network and the identification of top 20% of landscape. I first calibrated species distribution modelling (e.g., Maxent) and then optimized each species habitat suitability map to be used in reserve-selection algorithm, (e.g., Zonation). Two cell removal rules, *abf* (5 models) and *caz* (3 models), have been used to investigate different conservation value of landscape. I did not consider the usage of HII layer with *caz* because it may jeopardize the overall performance of Zonation, and hence producing illogical solutions (see Moilanen et al., 2014: 32).

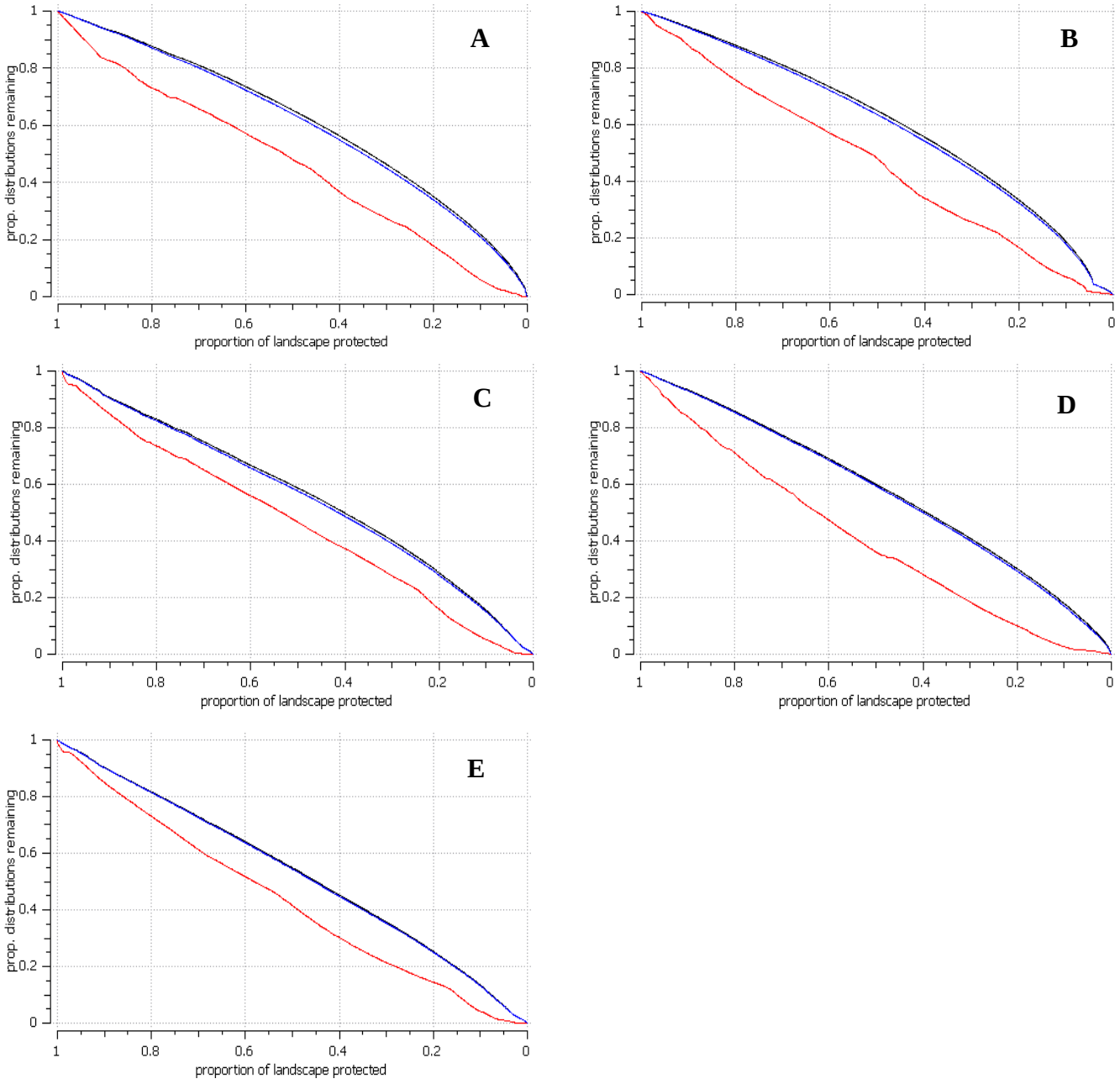


Figure S5.9: Performance curves showing the proportion of species distributions predicted to be conserved as a function of the proportion of the landscape which is protected when *abf* is used as a cell removal rule in Zonation. A; basic model using *abf* which means that no additional layers have been used, B; *abf* with current PAs forced to be in the top fraction when grid cells are ranked, C; *abf* with a 'cost' layer based on the Human Impact Index, D; *abf* with uncertainty analysis used in form of distribution discounting analysis, E; *abf* with both cost layer and uncertainty analysis used simultaneously. The red line represents species with the lowest fraction of distribution remaining among all species, while blue line (unweighted) and black (weighted) represent the average proportion of each species distribution remaining.

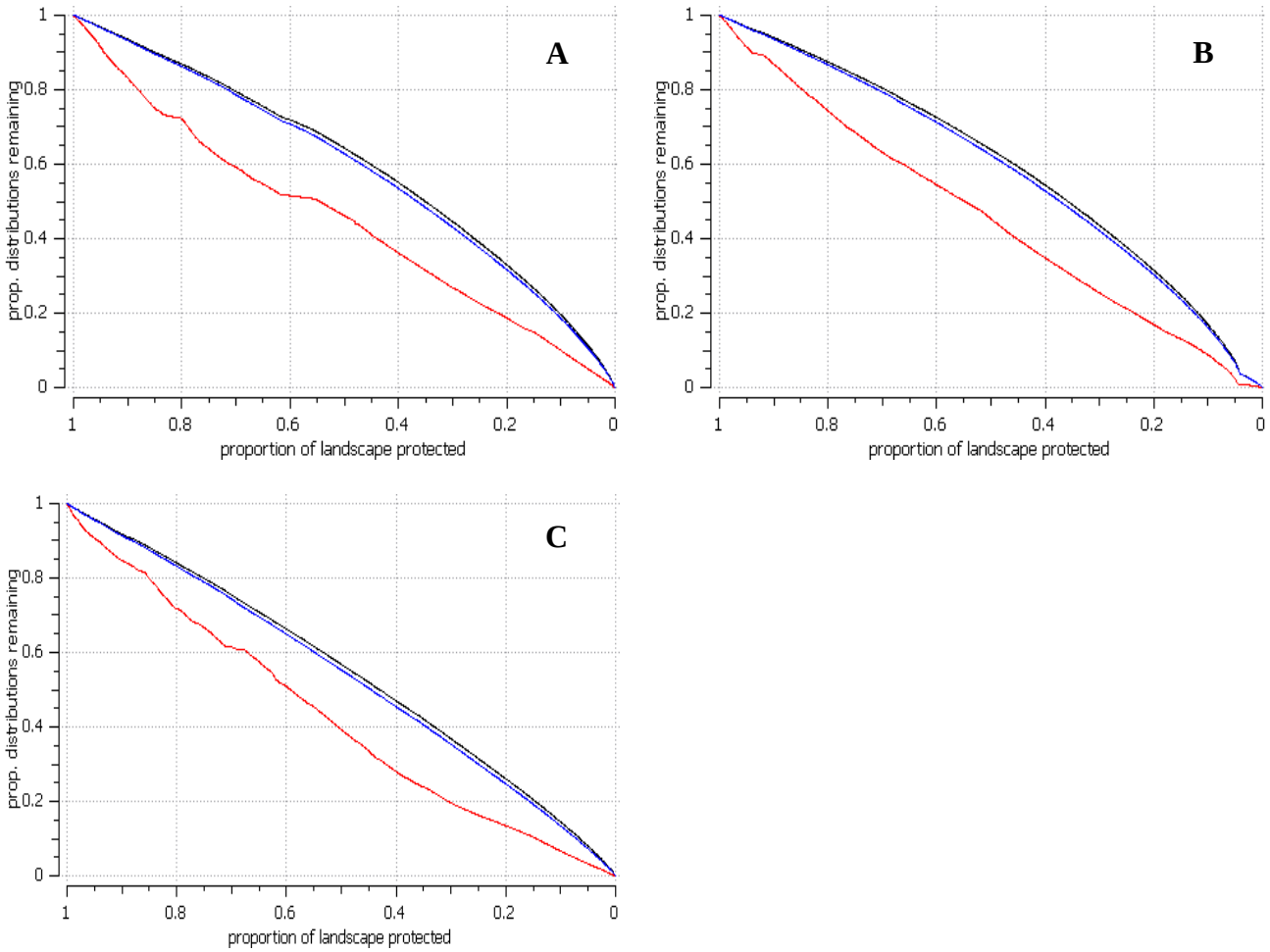


Figure S5.10: Performance curves showing the proportion of species distributions predicted to be conserved as a function of the proportion of the landscape which is protected when *caz* is used as a cell removal rule in Zonation. A; basic model using *caz* which means that no additional layers have been used, B; *caz* with current PAs forced to be in the top fraction when grid cells are ranked, C; *caz* with uncertainty analysis used in form of distribution discounting analysis. The red line represents species with the lowest fraction of distribution remaining among all species, while blue line (unweighted) and black (weighted) represent the averaged proportion of each species distribution remaining.

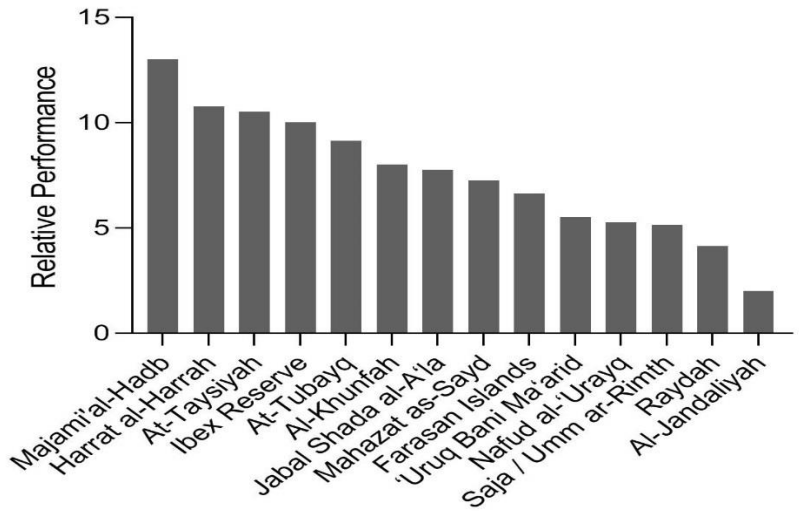


Figure S5.11: The relative performance of current PAs in terms of protecting terrestrial reptile species in Saudi Arabia from the best to the lowest. The relative performance of PAs was ranked according to their conservation performance indexed by the difference between inside and outside across all different Zonation models run.

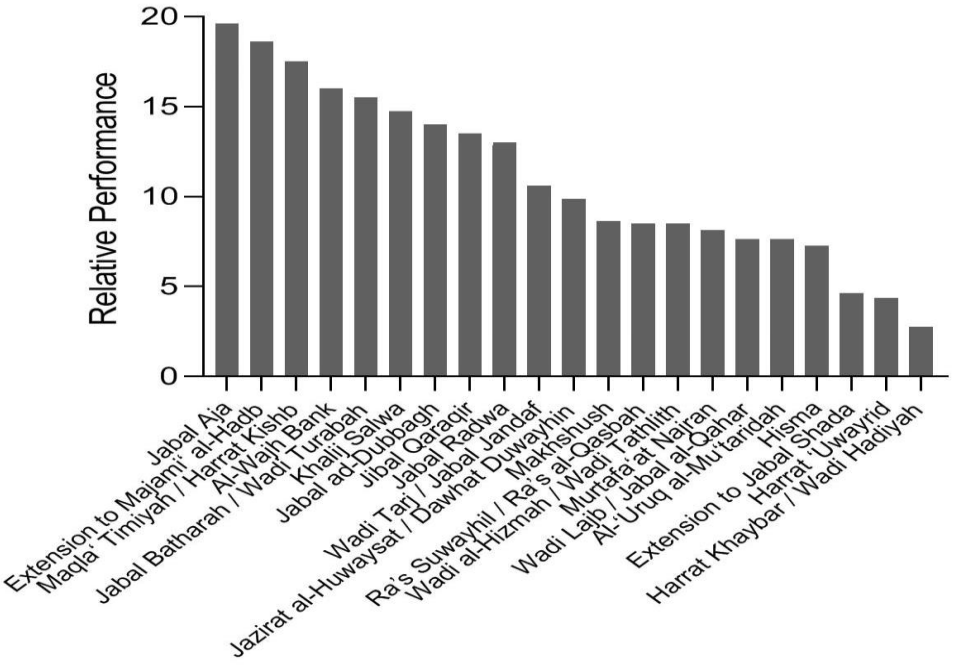


Figure S5.12: The relative performance of proposed PAs in terms of protecting terrestrial reptile species in Saudi Arabia from the best to the lowest. The relative performance of PAs was ranked according to their conservation performance indexed by the difference between inside and outside across all different Zonation models run.

6 Chapter Six: Final Conclusion

As the second largest vertebrate group, reptiles are an important part of Saudi biological diversity and ecosystems. The conservation status for some species has been classified as Vulnerable, Near Threatened, and Data Deficient, with major threats caused by habitat loss and hunting (Cox *et al.*, 2012). From a conservation perspective, this sounds worrying as most of these habitats are fragile, and any potential damage can be irreversible. The majority of ongoing conservation actions were established to protect other taxa (e.g., mammals) the conservation challenges for which can be quite different. However, an encouraging message from this thesis is that I found that the current and proposed Protected Areas of Saudi Arabia do have more conservation value for lizards and snakes than the areas outside them. This result also supports the ambitious plan of local authorities to increase the protection of biodiversity across the country covering different ecoregions (Al-Saadon and SWA, 2012). However, one concern is the paucity of published information so far about reptile conservation and distribution in these PAs and also outside them (but see Wilms *et al.*, 2011). Therefore, planning for reptile conservation based on reliable analyses and science should be a priority.

Through this research I found two main issues/gaps regarding Saudi terrestrial reptile species that need to be urgently addressed in order to understand their distribution, and hence their conservation requirements. These are i) large parts of Saudi Arabia lack systemic survey efforts for reptiles; and ii) most records, even in newly published work, are opportunistically/ haphazardly collected without taking in consideration any underlying ecological processes. This indicates that the current species richness distribution patterns may not represent the reality taxonomically or biogeographically. I found that the degree of sampling effort was very low across the country, and efforts must at least be doubled in the form of well-planned systematic surveys across the country to achieve a satisfactory level of knowledge of reptile

distributions. Such new sampling efforts must also include areas that have been surveyed before (Cox *et al.*, 2012).

Before the work presented in this thesis, there was very little contemporary knowledge about Saudi terrestrial species richness distribution patterns, habitat suitability, and potential important conservation areas based on quantitative analysis. This thesis significantly improves our overall understanding, and delivers valuable and important ecological and biogeographical information. Its findings are concordant with the majority of the conclusions of various other studies in different and similar habitat conditions (Newbold *et al.*, 2010; Cox *et al.*, 2012; El-Gabbas *et al.*, 2016; Carranza *et al.*, 2018).

Conservation in Saudi Arabia is still a relatively new practice (Barichievy *et al.*, 2018). Initiating proper conservation actions requires having a good level of knowledge of species distributions (Pardo *et al.*, 2013), but in reality collecting new data is difficult (Tsoar *et al.*, 2007; Pardo *et al.*, 2013). Therefore, the demand for new techniques that can study species distributions remotely is very high. Many studies have used species distribution modelling (SDM) techniques to answer important ecological and biogeographical questions. Its applications include studying species richness distribution patterns, transferring models between regions, guiding field surveys efforts, directing conservation effort, and prioritizing conservation planning (Guisan and Zimmermann, 2000; Raxworthy *et al.*, 2003; Guisan and Thuiller, 2005; Randin *et al.*, 2006; Pearson, 2008; Peterson *et al.*, 2011; Werkowska *et al.*, 2017; Yates *et al.*, 2018). Interpolation and transference of species distribution models represent a promising solution for situations where biological data are very scarce, and few studies have been conducted (see Chapters 2 & 3). Such predictions are very important because they can show the suitability/reliability of the model for specific tasks in this region when properly evaluated (see Chapter 4) and lead to an applied quantitative assessment of the effectiveness of ongoing and proposed conservation actions (see Chapter 5). Species

distribution modelling also has the potential to contribute to other goals in Saudi Arabia, which may be vital in future studies, such as seeking to ameliorate the impact of global environmental change, and the national assessment of ecological status (i.e., IUCN).

This project suggests a number of research priorities which hopefully can be addressed in the future. First, it was surprising to realize that all the accessible records of reptiles in Saudi Arabia are held in international museums, with nothing locally as far as I am aware. Two main concerns can be highlighted: i) we need a reliable national Natural History Museum where local specimens and data can be preserved and shared with others; and ii) from my reading I strongly believe that there is a large collection of specimens in existence (especially at King Saud University) but the data are not yet online and available. Second, we need to establish a good network among local ecologists to facilitate information sharing and to be up to date with new data about Saudi Arabia. Third, more well-designed field surveys across Saudi Arabia are badly needed. It is very clear that some areas and species have received more effort than others: such bias need to be minimized in the future. Ensuring good survey coverage can help to understand reptile biogeography properly, especially for vulnerable and threatened species that need extra attention from conservationists. Fourth, we should endeavour to increase local community involvement and awareness about the importance of preserving wild fauna and flora in a desert ecosystem. Fifth, conservation action planning is typically based on political-economic reasons, but proper scientific evaluation and planning can also help to satisfy all stakeholders.

In the absence of systematic surveys of the reptile fauna across the country, the Saudi data that I collated represent a very valuable information resource. The presence-only distribution modelling I applied here provides the best current option for describing and understanding of the spatial distribution of reptile species richness in this understudied part of the world. Although based only on reptile data, this conservation assessment may be valuable

in understanding biodiversity more generally, by illuminating their abundance, diversity, adaptation to local environment and wide distribution. I hope that my work will motivate more studies to follow a similar approach for other local taxa, leading to more systematic attempts to understand species richness patterns and support conservation actions.

6.1 References

Al-Saadon, K. H. & Saudi Wildlife Authority. 2012. *Action Plan for Implementing the Convention on Biological Diversity's Programme of Work on Protected Areas*.

Barichievy, C., Sheldon, R. Wacher, T., Llewellyn, O., Al-Mutairy, M. & Alagaili, A. 2018. Conservation in Saudi Arabia; moving from strategy to practice. *Saudi Journal of Biological Sciences*, 25**,** 290-292.

Carranza, S., Xipell, M., Tarroso, P., Gardner, A., Arnold, E. N., Robinson, M. D., et al. 2018. Diversity, distribution and conservation of the terrestrial reptiles of Oman (Sauropsida, Squamata). *PLoS ONE*, 13**,** e0190389.

Cox, N. A., Mallon, D., Bowles, P., Els, J. & Tognelli, M. F. 2012. *The Conservation Status and Distribution of Reptiles of the Arabian Peninsula*. Cambridge, UK and Gland, Switzerland: IUCN, and Sharjah, UAE: Environment and Protected Areas Authority.

El-Gabbas, A., Baha El Din, S., Zalat, S. & Gilbert, F. 2016. Conserving Egypt's reptiles under climate change. *Journal of Arid Environments*, 127**,** 211-221.

Guisan, A. & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8**,** 993-1009.

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

Newbold, T., Reader, T., El-Gabbas, A., Berg, W., Shohdi, W. M., Zalat, S., Baha El Din, S. & Gilbert, F. 2010. Testing the accuracy of species distribution models using species records from a new field survey. *Oikos*, 119**,** 1326-1334.

Pardo, I., Pata, M. P., Gomez, D. & Garcia, M. B. 2013. A novel method to handle the effect of uneven sampling effort in biodiversity databases. *PloS ONE*, 8**,** e52786.

Pearson, R. G. 2008. Species' distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*. Available at http://ncep.amnh.org.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M. & Araújo, M. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press.

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.

Raxworthy, C. J., Martínez-Meyer, E., Horning, N., Nussbaum, R. A., Schneider, G. E., Ortega-Huerta, M. A. & Peterson, A. T. 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, 426**,** 837-841.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D. & Kadmon, R. 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions*, 13**,** 397-405.

Werkowska, W., Márquez, A. L., Real, R. & Acevedo, P. 2017. A practical overview of transferability in species distribution modeling. *Environmental Reviews*, 25**,** 127-133.

Wilms, T. M., Wagner, P., Shobrak, M., Rödder, D. & Böhme, W. 2011. Living on the edge? - On the thermobiology and activity pattern of the large herbivorous desert lizard *Uromastyx aegyptia microlepis* Blanford, 1875 at Mahazat as-Sayd Protected Area, Saudi Arabia. *Journal of Arid Environments*, 75**,** 636-647.

Yates, K. L., Bouchet, P. J., Caley, M. J., Mengersen, K., Randin, C. F., et al. 2018. Outstanding challenges in the transferability of ecological models. *Trends in Ecology & Evolution*, 33**,** 790-802.