Predictive Distribution Modelling of *Timon lepida* in Spain

Ednah Kgosiesele
March, 2010

Course Title:	Geo-Information Science and Earth Observation for Environmental Modelling and Management
Level:	Master of Science (Msc)
Course Duration:	September 2008 - March 2010
Consortium partners:	University of Southampton (UK) Lund University (Sweden) University of Warsaw (Poland) International Institute for Geo-Information Science and Earth Observation (ITC) (The Netherlands)
GEM thesis number:	2010-30

Predictive Distribution Modelling of *Timon lepida* in Spain

by

Ednah Kgosiesele

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation for Environmental Modelling and Management

Thesis Assessment Board

Chairman: Prof. Dr. A.K. (Andrew) Skidmore, ITC
External Examiner: Prof. T.P. (Terry) Dawson
First Supervisor: Dr. A.G. (Bert) Toxopeus, ITC
Second Supervisor: Msc. (V) Valetijn Venus

ITC UNIVERSITY OF TWENTE.
FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

Disclaimer

Abstract

Ecological factors determining the geographic distribution of *Timon lepida* are poorly known. This work modelled the potential geographic distribution of *Timon lepida* at two spatial scales: 1. Landscape (Andalucía) 2. Regional (Spain) and analyzed the degree to which this distribution is associated with different predictor variables (e.g. temperature, solar radiation, topography, vegetation etc). The objectives of this study are to: (1) determine the most important predictor variables influencing the spatial distribution of *Timon lepida*; (2) generate potential geographic distribution maps for this species and (3) compare the predictive powers of environmental variables and hyper temporal NDVI to predict this distribution. Maxent, a presence-only distribution modelling technique was used to model predicted ranges for *Timon lepida*, using a large dataset of 10*10 km UTM presence only records collected between 1998-2002 period over Europe and a set of biophysical variable of 1*1 km resolution. To test the average behavior of the algorithm, 10 iterative models were produced by dividing all the presence records into 70% for training and 30% for testing. Three sets of model scenarios were generated: (i) models including environmental variables, (ii) models including environmental variables and vegetation and (iii) models including vegetation indices.

Model accuracy was measured with binomial tests of omission rates and the area under the curve (AUC). All models were significantly better than random by the binomial test and AUC measure. The AUC score for models built using environmental variables was always higher indicating better discrimination of suitable and unsuitable areas for the species. For the two spatial scales, environmental variables models had a superior predictive ability than vegetation models. These findings did not support our hypothesis. The results indicate that at a landscape level, topographic variables (aspect and slope) are the most important whereas at a regional scale, climatic variables (temperature seasonality, solar radiation, altitude) and hyper temporal NDVI appear to have a significant effect on this distribution pattern. The results of the present study can be an important contribution to a better understanding of the ecological requirements of the species. The conclusions would be more precise if the adequate precise high resolution environmental data is included in the future application and reliable datasets of current conditions are identified to improve results.

Keywords: Maxent, AUC, *Timon lepida*, NDVI, environmental variables

Acknowledgements

My sincere thanks to the GEM MSc course, and its organizers and directors – Professors Andrew Skidmore from ITC, Peter Atkinson from University of Southampton, Petter Pilesjõ from Lund University and Katarzyna Dabrowska from University of Warsaw for granting me the opportunity to be part of this scholarship programme, the experience has been truly great.

I would like to express my special gratitude to my supervisors Dr Bert Toxopeus and Mr Valentijn Venus for the support they provided me throughout my research. Dr Bert, the very first day I came to you with confusions you opened the door to me and have been supportive throughout my work. You introduced me to a totally new field and have helped me throughout with your guidance. Above all, thank you for your good sense of humour which made me relax during discussions.

The same token of gratitude goes to Dr Kees de Bie for his excellent GPS instructions and for providing the NDVI imagery used for this study. I also am grateful to my GEM course coordinator, Mr Andre Kooiman for the moral support he provided even before I met him. I would also like to thank Dr David Rossiter for his series of lectures "from proposal writing to thesis phase". I am truly grateful to Dr Tiejun Wang for his help on the statistical analysis and helpful ideas about my thesis.

To all my fellow GEM classmates: thank you for your friendship and the nice birthday surprises. You have been a family to me in absence of my maternal family and have made the past 18 months fantastic. A special thank you goes to Francis Kamau Muthoni, Diriba Kumssa and Ciro Patricio Romero Romero for their patience in sharing valuable ideas during my research and helping me with software problems. Eric Bissila Buedi, I am grateful for your friendship during the course.

Finally, let me express countless thanks to my family for their great encouragement, support and loving care. I made it up to this stage because you stood by my side through it all. Above all I give thanks and praise to God who gave me the strength to accomplish this work.

Table of contents

List of figures

List of tables

1. Introduction

1.1. Background and Significance

Habitat fragmentation caused by clearing and other human-related disturbances is one of the most serious ecological threats confronting the long-term survival of flora and fauna in the world today (Goode *et al.* 1998; Myers *et al.* 2000; Fahrig 2003; Bascompte and Ricard 2009). One of the most serious outcomes of environmental degradation with regard to ecosystem transformation is the current loss of biodiversity, which is occurring at a faster rate than at any time in human history (Bascompte and Ricard 2009). The main factors that trigger biodiversity loss at the global level are over-harvesting, alien species introduction, pollution, habitat fragmentation, and habitat destruction (Barnard and Thuiller 2008; Phoebe and Thuiller 2008). Local diversity is constrained proximally by resource abundance, competition and predation, but it is also influenced by larger temporal and spatial scale processes and events such as emigration and large scale disturbances (Barnard and Thuiller 2008). Moreover, climate change is having serious repercussions on biodiversity. As climate change intensifies, the negative effects produced are expected to worsen, causing habitat alteration, population reduction and species extinctions (Myers *et al.* 2000).

At the continental level, Europe's biodiversity is under threat because of urban development, industrialization and tourism increase. Of prime concern is the rapid decline of European herpetofauna species (amphibians and reptiles) (Hamilton 2005). For example; the Mediterranean region is recognized as a biodiversity hotspot (Myers *et al.* 2000). Its flora diversity is outstanding with 15,000 to 25,000 species, 60% of which are unique to the region (Baillie *et al.* 2004; Stuart *et al.* 2004). Spain, the country with the highest level of biodiversity in Europe is experiencing most of these negative effects (IUCN 2001). In Spain, reptiles account for about 15% of the 136 species that are under some degree of threat (IUCN 2001). Understanding the effects of environmental covariates on reptiles is therefore potentially useful and can lead to new insights into their distribution patterns.

Predictive Distribution Modelling

Species distribution modelling is becoming increasingly important in the face of accelerated rate of biodiversity loss and limited datasets on the species, since it informs conservation strategies and biodiversity management (Barry and Elith 2006; Peterson *et al.* 2007). Species distribution models relate the occurrence of species to environmental predictors in order to facilitate the mapping of predictions across the landscape even into areas that have not been surveyed before (Guisan and Zimmermann, 2003). A variety of predictive models are widely used to simulate the spatial distribution of plant and animal species. The most commonly used are the Generalized Linear Model (Guisan *et al.* 2002; Lehmann *et al.* 2002; Guisan and Zimmermann 2003; Engler *et al.* 2004), Generalized Additive Models (GAM), Classification and Regression Trees (CART's), Principal Components Analysis (PCA) and Artificial Neural Networks (ANNs). Recently, presence only models have become a powerful tool due to the huge datasets available in national museums and Herbaria and the questionable value of absence data. Examples of presence only distribution models include Ecological Niche Factor (ENFA), Genetic Algorithm for Rule-set Production (GARP) (Stolkwell 1999; Peterson *et al.* 2007) and Maximum Entropy (Maxent) (Phillips *et al.* 2004; Phillips *et al.* 2006; Phillips and Dudik 2008; Phillips *et al.* 2009). Species distribution models are especially useful in ecosystem management for the identification and habitat suitability mapping of areas containing high species occurrences and those species requiring attention (Graham and Hijmans 2006). Predictive distribution of species is thus an important tool for conservation, monitoring and assessing the possible impacts of environmental changes on this distribution (Hernandez *et al.* 2006). The need for accurate information on the distribution of species further illustrates the need for such kind of research.

Remote Sensing and Species Distribution Modelling

Concern over the future of biodiversity has compelled conservationists to come up with ways of determining its current status in order to predict its distribution in response to global environmental change. This is evidenced by the recent initiative taken by the country of Spain to devise strategies to "halt biodiversity loss before 2010" (Agency 2009). Knowing which areas are under threat is often a challenge. With the advent of species distribution modelling and Remote Sensing tools, it is easier to determine the factors driving the distribution patterns of species and therefore generate the knowledge required to better conserve them (Araujo and Williams 2000; Polasky and Solow 2001). Normalized Difference Vegetation Index (NDVI) is the oldest and most widely used index (Sellers 1989). While it has seen extensive use in mapping plant and animal species, its application to herpetofauna species is relatively new (Oindo and Skidmore 2002; Leyequien *et al.* 2007).

Species distributions are not only affected by climatic factors such as temperature and precipitation, but are also a result of abiotic and non-climatic factors such as topography, geology and landuse. Since NDVI is an integration of both climatic and biophysical variables, it may provide an index of ecosystem processes and productivity compared to climate based models as it is spatially-explicit (de Bie *et al.* 2006; Skidmore *et al.* 2006). Hyper temporal NDVI data can provide temporal quantitative information on vegetation reflectance that can be used to estimate relevant environmental factors influencing patterns of occurrence and abundance of various kinds of species. It is believed that using NDVI may improve the accuracy of the results since Remote Sensing provides direct measurements of vegetation variability (Oindo and Skidmore 2002). Hyper temporal NDVI appears to be of fine enough temporal resolution to capture the fluctuations of vegetation response in changing environmental conditions (de Bie *et al.* 2006). This study used hyper temporal classified NDVI, Corine Land Cover and a suite of other environmental variables in order to predict the distribution of *Timon lepida* in Andalucía and Spain for a better understanding of the current range of this species.

1.2. Research Problem

Accurate estimates of the spatial distribution of species assist conservation practitioners in predicting how a species will respond to landscape alteration and climate change. Very little is known about the critical environmental habitat requirements of *Timon lepida* (ocellated lizard). This lack of knowledge on the distribution and ecological niche requirements limits the ability to develop conservation strategies for this species. Although previous research has focused on modelling the distribution of many plant and animal species; For mammal species: (Toxopeus *et al.* 1994; Corsi *et al.* 1999; Corsi *et al.* 2000; Oindo and Skidmore 2002; Guisan and Hofer 2003; Guisan and Zimmermann 2003; Said *et al.* 2003), few papers in literature have predicted the distribution of reptiles (Owen 1989; Guisan and Hofer 2003); (Leyequien *et al.* 2007). Even amongst the few papers that have modelled the distribution of reptiles, there is still lack of literature on any case studies on reptiles (Maurer 1994; Guisan and Hofer 2003). Therefore, lack of adequate data on the distribution patterns of *Timon lepida* raise the question of what factors potentially restrict its present day range extensions.

The population of *Timon lepida,* the target species of this research, is generally threatened because of ongoing habitat loss, pesticide pollution and poisoning. Although reasonable populations are present in Spain, human defined vegetation patches and widespread habitat loss have led to a substantial decline of the species in many areas (IUCN 2001). Predators might also be eating this species more due to the decline of rabbits (Baillie *et al.* 2004). In the IUCN red list of threatened species, *Timon lepida* is classified as "Near Threatened" (IUCN 2004). There is lack of literature on case studies relating the distribution of *Timon lepida* to environmental variables and vegetation (classified NDVI and Corine Land Cover). Further studies are needed for understanding and predicting the distribution of this species in relation to both biophysical and climatic factors in order to understand which predictor variables are restricting its distribution.

Remote Sensing is becoming an indispensable tool in species distribution modelling. While remotely sensed data for animal diversity assessment using habitat characteristics is increasingly used, few studies have incorporated it in reptile distribution modelling (Leyequien *et al.* 2007). It is hypothesized that species abundance increase with ecosystem productivity and therefore NDVI can be used as a surrogate measure of productivity (Box *et al.* 1989; Oindo and Skidmore 2002). There is lack of literature on case studies relating environmental variables and classified hyper temporal NDVI in predicting the distribution of *Timon lepida* in Spain. The aim of this study was to incorporate both the conventional approach to species distribution modelling that focuses on the use of environmental predictors (such as precipitation, temperature) and the use of remotely sensed data (hyper temporal classified NDVI) as a means of improving predictions of the distribution of *Timon lepida* at the two spatial scales of Andalucía (landscape) and Spain (regional). According to the scale domain of (Pearson and Dawson 2003), landscape level characterises areas between 10-200 km and regional level characterises areas corresponding to 200-2000 km. In the end, the results of this research may aid in determining whether vegetation indices will be better predictors than environmental variables, thereby acting as a guide for future research and contributing to existing knowledge on the factors influencing the distribution of *Timon lepida* in Spain.

1.3. Overall Objective

The overall aim of this study was to explore the predictive power of vegetation indices (hyper temporal classified NDVI and Corine Land Cover) and environmental variables to predict the distribution and observed patterns of *Timon lepida* in Andalucía and Spain using Maximum Entropy modelling approach (Maxent).

1.3.1. Specific Objectives

Specifically, the study sought to:

1. Generate potential geographic distribution maps for *Timon lepida* based on the Maxent model output.

2. Determine the most important predictor variable (s) potentially responsible for the distribution of *Timon lepida* at two different spatial scales: (i) Andalucía and (ii) Spain.

3. Evaluate whether environmental variables or vegetation indices (classified NDVI and Corine Land Cover) would better predict the distribution of *Timon lepida* at the two spatial modelling scales.

1.4. Research Questions

1. Which set of predictor variables have the strongest predictive power to determine the potential distribution of *Timon lepida* in (i) Andalucía and (ii) Spain?

2. Which set of predictor variables are more important at each spatial scale to model the potential distribution of the target species?

3. Does classified NDVI and land cover variables significantly predict the distribution of *Timon lepida* better than environmental variables?

1.5. Research Hypotheses

Hypothesis 1: Testing the concept that the potential distribution models for *Timon lepida* generated by Maxent algorithm will perform significantly better than random.

1-H_0: The potential distribution models of *Timon lepida* will not predict the distribution of *Timon lepida* significantly better than random;

1-H_1: The potential distribution models will predict the distribution of *Timon lepida* significantly better than random.

Hypothesis 2: Testing the concept that vegetation indices (Classified NDVI and Corine Land Cover) would predict the distribution of *Timon lepida* significantly better than environmental variables (mean temperature of driest month, solar radiation, etc).

2-H_0: Models that include classified NDVI and land cover variables would not predict the distribution of *Timon lepida* significantly better than models that do not include classified NDVI and land cover variables as additional variables;

2-H_1: Models that include classified NDVI and land cover variables would predict the distribution of *Timon lepida* significantly better than models that include classified NDVI and land cover variables as additional variables.

1.6. Research Approach

Modelling was divided into 3 stages: 1. Data acquisition and preparation stage, 2. Modelling stage, and 3. Model validation stage.

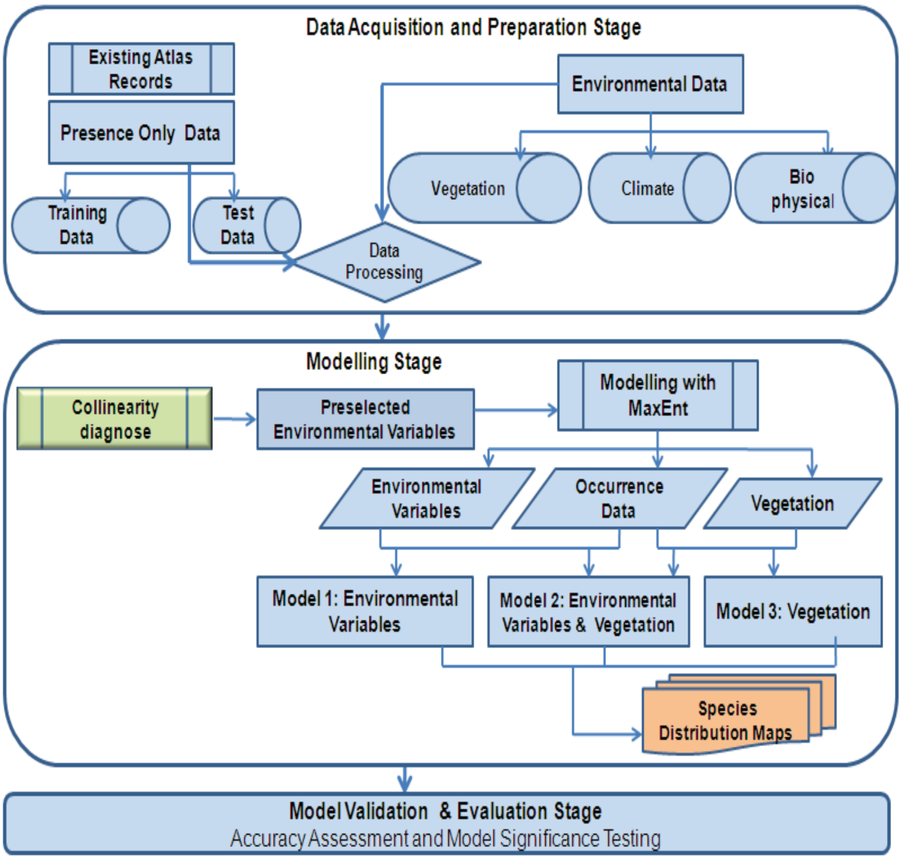


Figure 1-1: Conceptual Framework of the Study

2. Materials and Methods

2.1. Study Area

Andalucía is the southern-most region of Spain, the bridge between two continents (Europe and Africa) and it is located where the Mediterranean meets the Atlantic. It is the second largest region in Spain occupying an area of about 87,300 km^2 and also the largest one of the autonomous regions (Giannakopoulos *et al.* 2005). Andalucía can be divided into two main areas: high Andalucía made up of the mountain ranges and low Andalucía which is the huge depression created by the Guadalquivir River and its numerous tributaries. These two regions are very different in terms of climate. The lower Andalucía is comprised of huge flatlands up to 300 km wide. The flatlands are important for their rich variety of plant and bird species, particularly the marshlands of the Coto de Doñana. The flatlands, being low and in the sheltered south west, have a warmer climate than the mountainous areas. They are mild and pleasant in winter but experience high temperatures and humidity in midsummer (Giannakopoulos *et al.* 2005). Half of the Andalucía surface is mountainous, one third on a level of more than 600 meter altitude and about 46 peaks are higher than 1000 meters. The highest mountains of the Spanish peninsula are situated in the Sierra Nevada Mountains: the Mulhacén (3.481 meters) and Veleta (3.398 meters) (Giannakopoulos *et al.* 2005).

The Andalucía landscape is varied. Besides the expanded valley of the Guadalquivir River and its tributaries there are large forests of deciduous and cork trees in the low mountain ranges, snow-covered alpine high mountains, over 500 miles of coastline, the volcanic landscape of the Coto de Gata and even a half desert close to Tabernas Almeria .Topography, elevation and soils are the most influencing factors for the variety of biodiversity in this region. Andalucía supports a wide range of biodiversity such as plant communities including woodlands, shrub lands, broad-leafed forests and about 5000 different species of plants out of which 150 are unique to the area (Bario 2006). The climate also supports a number of numerous animal species. The dominant landuse type is agriculture (Bario 2006).

Andalucía's weather is Mediterranean. Mainly it is mild all-year-round, with short winters and long summer season. Rainfall is irregular; concentrating mainly in the fall and spring, while in summertime it is dry. Average maximum temperature is 23°C and minimum temperatures range as low as 12°C (Roberts 1986). The dominant rock type in Andalucía is limestone though crystalline rocks (granites, schist and gneiss) are also evident. The three major soil types in Andalucía are peridotite soil, limestone soil, and a spectrally quenched soil. These soils influence the vegetation and crop production in the province (Giannakopoulos *et al.* 2005).

Figure 2-1: Map of the study area

Top: Spain, Bottom: Andalucia

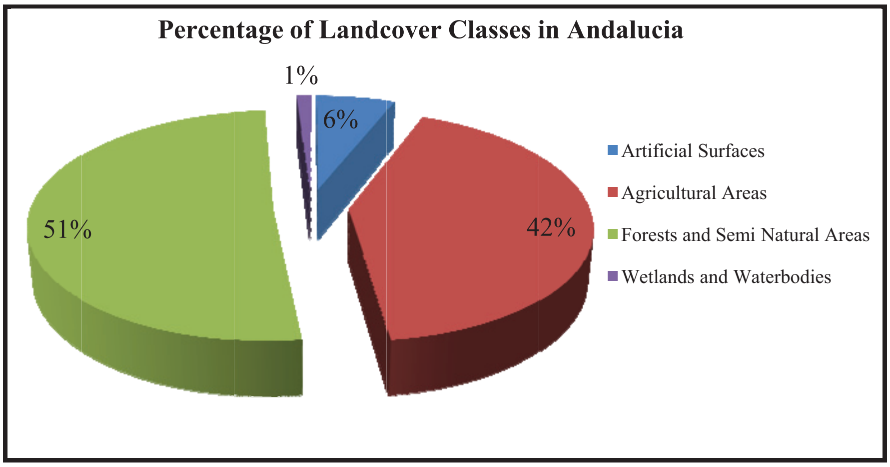


Figure 2-2: Percentage of land cover classes in Andalucía

Spain covers a total surface area of 505,370 sq km including the Balearic and Canary Islands. It occupies 85% of the Iberian Peninsula which it shares with Portugal in the south-west. It is bordered to the south and east by the Mediterranean sea, to the north by France, Andorra and the Bay of Biscay, and to the northwest and west by the Atlantic Ocean (Bario 2006). Peninsular Spain experiences three major climatic types: Continental, Oceanic, and Mediterranean (Bario 2006). In terms of agriculture, Spain is a large exporter of olives which has been boosted by agricultural subsidies since the incorporation of the country in the European Union (Bario 2006).

2.2. Target Species

The main focus of this study is the ocellated lizard, also known as *Timon lepida*. Formally known as *Lacerta lepida*, the ocellated lizard is the largest legged lizard in Europe. It is found in Spain, Portugal, France and north-western Italy (Diaz and Carrascal 1991). Large individuals can grow up to 80 cm long including the tail. They are characterized by their large size, huge heads (for males) and blue spots on the flanks (mainly in adult males). Their habitat consists of typical Mediterranean scrub land, dry and often densely vegetated areas (Diaz and Carrascal 1991). In southern Spain they can live up to 2100 m in altitude. *Timon lepida* prefers dry bushy areas, such as open woodland and scrub, old olive groves and vineyards, and is sometimes found on more open rocky or sandy areas. The lizard usually stays on the ground, but climbs well on rocks and on trees. It can dig holes and sometimes uses abandoned rabbit burrows. Although it can tackle prey as large as a small rabbit, its normal diet is insects, snails, rodents and some sweet fruits (Diaz and Carrascal 1991).

Timon lepida is preyed upon by eagles such as Short toed (*Circaetus gallicus)* (Vlachos and Papageorgiou 1994). The young ones and the juveniles are preyed upon by large snakes such as Montpellier snake. There has been a substantial decline of this species in many areas due to widespread habitat loss and persecution by hunters that fear the lizard eats all of the partridge eggs and young rabbits. In the past larger lizards were hunted and eaten as well and this might have led to the current reduction in numbers. The subspecies *Timon.lepida. oteroi* is endemic to Salvora Island in North-western Spain while the subspecies *Timon.lepida. nevadensis* occupy the South-eastern region (IUCN 2001). *Timon lepida* prefers temperature ranges between 24-27°C but can occasionally be seen basking at temperatures between 30-35°C.

Figure 2-3: *Timon lepida*

2.3. Data Used

2.3.1. *Timon lepida* Presence Data

The amphibians and reptiles National Atlas of Spain (*Atlas y Libro Rojo de los Anfibios Reptiles de Espana*) (Pleguezuelos *et al.* 2004) provided the basis for the Maxent models used in this study. The database contains presence and absence data for *Timon lepida* collected between 1998 to 2000 for the whole of Spain. Presence only data was used for modelling. There are 5901 grids in the form of 10 km by 10 km UTM Zone 30N recorded as “presence and absence” in Spain. 718 grids fall within Andalucía while Spain encompasses 3442 grids as “presence” of *Timon lepida*. Using the Hawths sampling tools of ArcGIS 9.3, all presence records were extracted based on the central point of each grid cell and used to model the distribution of *Timon lepida* using Maxent models.

Table 2-1: Presence and absence records for *Timon lepida* in Spain

	Andalucía	Spain
Presence	718	3442
Absence	158	2458
% Occupancy	82	58.3

2.4. Environmental Data Layers

Species distributions are limited to a certain time and space due to certain environmental conditions (Barnard and Thuiller 2008). The choice of environmental variables greatly influences the outcome of species distribution models and the careful selection of predictor variables is therefore a central step in modelling. Expert knowledge was used for selecting the suite of environmental variables that was used as input data for the models. There appears to be a significant correlation between precipitation, temperature, solar radiation and the abundance and the daily activities of reptiles (Nicholson *et al.* 2005). Net Primary Production (NPP) is related to the ecosystems and long term NDVI thus can indicate the overall productivity of an ecosystem (Oindo and Skidmore 2002). The environmental variables used to fit the models are known to have a major eco-physiological impact on *Timon lepida* and may therefore influence its distribution (Jellinek *et al.* 2004). Many previous applications of species distribution models have used similar indices as predictors of distribution patterns (Skidmore *et al.* 2006).

A total of 16 predictor variables were pre-selected for modelling. Volume of stones was however omitted from the final modelling process since its contribution to the overall gain of the models was very minimal.

Table 2-2: Pre-selected Environmental Variables

Category	Variable	Data Format	Resolution	Data Source
Climate	Temperature Seasonality Mean Temperature of Driest Quarter Mean Temperature of Wettest Quarter Precipitation of Driest Quarter Precipitation of warmest Quarter	Raster	1km	WORLDCLIM
	Potential Evapotranspiration	Raster	1km	CGIAR_CSI
	Direct Annual Radiation Aridity Cloud Cover	Raster	1km	USGS/NIEHS
Soil	Soil Type	Raster	1 km	ESDB
Terrain	Altitude Slope Aspect (Southness and Westness)	Raster	1 km	USGS/SRTM
Vegetation	Classified NDVI	Raster	1 km	SPOT Vegetation
	Corine Land Cover	Raster	100 m	CLC 2000

2.4.1. Climatological Variables

Previous studies have indicated that there might be a causal relationship between climatic variables and species abundance (Badgley and Fox 2000; Lennon *et al.* 2000). Climate data layers for this study were obtained from Worldclim bioclimatic database (http://www.worldclim.org/). This database contains 19 climatic variables of precipitation and temperature for the period 1950-2000 (Hijmans *et al.* 2005). The climate data layers were generated through interpolation of average monthly data from weather stations using thin plate smoothing splines. Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. The climatic variables were derived in ESRI Grid Format at a resolution of 1km^2 (30 arc seconds).

2.4.2. Potential Evapotranspiration Data

Potential Evapo-Transpiration (PET) is a measure of the ability of the atmosphere to remove water through Evapo-Transpiration (ET) processes. The Global-PET was modelled using the data available from the WorldClim Global Climate Data (Hijmans *et al.* 2005) as input parameters. Annual PET (mm) layer was obtained from (http://csi.cgiar.org) at a spatial resolution of 30 arc-seconds (~ 1 km at tropics) for the 1950-2000 period. PET is calculated using the Hargreaves method (described below) with available layers of monthly average temperature parameters, available from WorldClim database, and extra-terrestrial radiation, calculated for specific months using a methodology presented by (Hargreaves *et al.* 1985; Allen *et al.* 1998; Hargreaves and Allen 2003). Temperature range (TD) is an effective proxy to describe the effect of cloud cover on the quantity of extra-terrestrial radiation reaching the land surface and, as such, it describes more complex physical processes with easily available climate data at high resolution.

PET = 0.0023 * RA * (Tmean + 17.8) * TD 0.5 (mm / day) eqn. (2-1)

Where:
RA stands for mean monthly extra-terrestrial radiation (RA, radiation on top of atmosphere)
Tmean: mean monthly temperature
TD: mean monthly temperature range

2.4.3. Annual Aridity Data

Aridity is usually expressed as a generalized function of precipitation, temperature, and potential Evapo-Transpiration (PET). An Aridity Index (UNEP, 1997) can be used to quantify precipitation availability over atmospheric water demand. These datasets have been downloaded and are available from the CGIAR-CSI GeoPortal (http://www.csi.cgiar.org). Global mapping of mean Aridity Index from the 1950-2000 period at 30 arc second ($1km^2$) spatial resolution is calculated as:

$$\text{Aridity Index (AI)} = \frac{\text{MAP}}{\text{MAE}} \quad \text{eqn. (2-2)}$$

Where:

MAP = Mean Annual Precipitation

MAE = Mean Annual Potential Evapo-Transpiration (PET)

Aridity Index values reported within the Global-Aridity geo dataset have been multiplied by a factor of 10,000 to derive and distribute the data as integers (with 4 decimal accuracy). This multiplier has been used to increase the precision of the variable values without using decimals (real or floating values are less efficient in terms of computing time and space compared to integer values). Global-Aridity values need to be multiplied by 0.0001 to retrieve the values in the correct units. For a full documentation of this dataset refer to (Trubacco *et al.* 2008).

2.4.4. Cloud Cover Data

Cloud Cover Data was downloaded from USGS/NIEHS at a spatial resolution of 1km. The data layer was re-projected from the initial ETRS 50 into UTM Zone 30N and clipped using the study area boundary for conformity with other variables used in this study.

2.4.5. Radiation Data

Radiation data was acquired from ESRA (European Solar Radiation Atlas, 2000) through ITC. Solar radiation data are ten-year (1981-1990) averages of monthly means of daily radiation in watt hour per square meter (Wh/m^2) on flat plane. Annual data for horizontal direct irradiation were used for this study. The data was resampled to 0.00833 degree (roughly 1 km^2 grid size for mid-latitudes)**.** For a full description of the dataset see ESRA publication, also available in the ITC-library (ISBN: 2-911762-22-3).

2.4.6. Digital Elevation Model (DEM)

Altitude, slope and topographic exposure layers were derived from the Shuttle Radar Topographic Mission dataset (SRTM; available at http://srtm.csi.cgiar.org/). The DEM used for this study is the 1km STRM DEM which has been resampled and significantly improved from earlier versions using new interpolation algorithms and auxiliary DEMS. The data was derived in ASCII Info Format in geographic coordinate system - WGS84 datum. The CGIAR-CSI version 4 provides the best global coverage full resolution SRTM dataset which has been processed to fill in no-data voids (Reuter *et al.* 2007). After mosaic, the resulting DEM was clipped with the study area boundary. Altitude, Slope and aspect were derived using the following pre-processing steps:

2.4.6.1. Altitude and Slope

1. All negative values were re-classified into zeros
2. The DEM was re-projected to WGS_1984_UTM_Zone_30N (to change the measurement units from decimal degrees to meters to reduce errors and allow easy calculation of slope). Altitude layer was derived from the clipped DEM above.

3. Slope and aspect were calculated using *Surface Analysis tool* of ArcGIS 9.3
4. The layers were converted to ASCII format coincident with other environmental variables.

2.4.6.2. Aspect

Aspect can be defined as the direction a slope faces with respect to the sun. Aspect identifies the down slope direction of the maximum rate of change in a value from each cell to its neighbors. Chang *et al.* (2006) defined southness as the degree to which an aspect is south, and westness as the degree to which it is west. Aspect was transformed to linear measures of southness and westness in order to avoid identical aspects (e.g., 0 and 360 degrees) and create two data layers with unique numerical representation. Southness and westness are the one of the best ways to handle aspect from an ecological perspective. Aspect was transformed into westness using the sine function while southness was derived by transforming aspect using the cosine function as supported by (Roberts 1986; Mollenbeck *et al.* 2009; Schaller *et al.* 2010). All the transformations were done in ArcGIS 9.3.

2.4.7. Soil Variables

Soil data layers were all obtained from the version 2 raster library of the European Soil Database (ESDB) at a resolution of 1:1000 000 ($1km^2$). The database contains a list of Soil Typological Units (STU) for the period 1998-2006 (Database 2004). Besides the soil names they represent, these units are described by variables (attributes) specifying the nature and properties of the soils: for instance, dominant land use, soil type, volume of stones, and dominant parent material of the full soil code of the Soil Typological Unit (STU) of the World Reference Base (WRB) for Soil Resources. Only 1 variable: soil type was used for this study. The layers were re-projected from the initial ETRS 1989 to WGS 84 UTM Zone 30 N coincident with other environmental variables used for this study.

Table 2-3: Description of Soil Types

Code	Soil Type	Code	Soil Type	Code	Soil Type
1	Rock outcrops	10	Podzol	19	Fluvisol
2	Water body	11	Planosol	20	Cryosol
3	Soil disturbed by man	12	Phaeozem	21	Cambisol
4	Town	13	Luvisol	22	Calcisol
5	Vertisol	14	Leptosol	23	Chernozem
6	Umbrisol	15	Kastanozem	24	Arenosol
7	Solonetz	16	Histosol	25	Andosol
8	Solonchak	17	Gypsisol	26	Acrisol
9	Regosol	18	Gleysol	27	Albeluvisol

2.4.8. Corine Land Cover 2000

Corine Land Cover (CLC) 2000 version 4 was obtained from the European Environment Agency (EEA) (http://dataservice.eea.europa.eu/dataservice/). This is the latest available version of the dataset. The new CORINE2000 represent land cover and land cover changes for the period 1999-2001. The nomenclature comprises of 44 land cover classes which have been created by on-screen interpretation and digitizing of Landsat images in a GIS environment (Neumann *et al.* 2007). The data was retrieved at a resolution of 100 m^2. The accuracy of CLC has been reported to be over 85% (Martin de Santa Olalla Manas *et.al* 2003) and version 4 possesses the following geographic information quality label:

- Completeness: Good
- Logical Consistency: Excellent
- Position Accuracy: Excellent
- Temporal Accuracy: Excellent
- Thematic Accuracy: Excellent

The raster layer was re-projected from the initial ETRS Projection to the target projection and resampled to 1 km^2 resolution using bilinear interpolation. This interpolation technique is more realistic than simply using nearest neighbour interpolation though it does not increase the resolution of the data (Phillips *et al.* 2006).

Table 2-4: Description of CLC classes

Corine code	Corine land cover
111	Continuous Urban Fabric
112	Discontinuous Urban Fabric
121	Industrial or Commercial Units
123	Port Areas
124	Airports
131	Mineral extraction sites
133	construction sites
142	sport and leisure facilities
211	Non-irrigated arable land
212	Permanently irrigated land
221	Vineyards
222	Fruit trees and berry plantations
223	Olive groves
241	Annual crop associated with permanent crop
242	Complex cultivation pattern
243	Land principally occupied by agriculture
244	Agro-forestry area
311	Broad-leaved forest
312	Coniferous forest
313	Mixed forest
321	Natural grassland
323	Sclerophyllous vegetation
324	Transitional woodland scrub
332	Bare rock
334	Burnt areas
511	Water courses
512	Water bodies

All the above predictor variables were clipped by ArcGIS 9.3 *Spatial Analyst* tool using the study area boundary, re-projected to WGS 1984 UTM Zone 30 N and converted to ASCII raster grid format to meet the data requirements of Maxent.

2.4.9. Hyper Temporal Normalized Difference Vegetation Index

The Normalized Difference Vegetation Index (NDVI) is a greenness index derived by dividing the divergence between near-infrared and red reflectance measurements by their sum (Sellers 1989; Oindo and Skidmore 2002). The spatial and temporal heterogeneity in primary productivity and photosynthetically active biomass is an explanatory variable to assess species occurrence (Skidmore *et al.* 2006). The NDVI data used for this study is the ten day composite SPOT images (S10 product) obtained via ITC from www.VGT.vito.be at a resolution of 1 km^2. The 393 NDVI images were pre-processed in order to generate a vegetation variable that would distinguish between the different vegetation types in the study area. The acquired images were geo-referenced and de-clouded. De-clouded means: using by image and pixel the supplied quality record, only pixels with a 'good' radiometric quality for bands 2 (red; 0.61-0.68 µm) and 3 (near IR; 0.78-0.89 µm), and not having 'shadow', 'cloud' or 'uncertain', but 'clear' as general quality, were kept (removed pixels were labeled as 'missing') (de Bie *et al.* 2006).

2.4.9.1. Iterative Self-Organizing Data Analysis Technique (ISODATA) Classification

Using the ISODATA clustering algorithm of ERDAS-Imagine software, the 10 year (1998-2008) stack image layers were classified using unsupervised classification. ISODATA calculates the spectral distance and iteratively classifies the pixels until a minimum spectral distance is achieved (Tou and Gonzalez, 1974). The maximum number of iterations was set to 50 and the convergence threshold was set to 1.0. The pre-defined number of classes for Andalucía ranged from 10-100 while for Spain it ranged from 10-180 classes. SPOT hyper temporal NDVI was chosen because the images possess a fine temporal resolution and they are commonly decadal (every ten days) ; which acts as an effective source to capture the fluctuations of vegetation in response to changing environmental conditions (Storms and Etes 1993). To determine the optimum number of classes, signature separability (in Divergence distance) was calculated for the stacked image in ERDAS' Signature Editor and plotted in Excel. A clear, evident peak in the separability classes signifies the optimum number of classes and these classes were chosen by visual inspection. The optimum number of classes for Andalucía is 45 and for Spain is 104 (Figure 2-4, 2-5). Each NDVI class represents an NDVI-profile showing changes in vegetation greenness overtime which is assumed to relate to the types of land cover and landuse present in the study area (de Bie *et al.* 2006). The final classified NDVI maps were converted to ASCII format to ensure compatibility with other datasets used for further analysis.

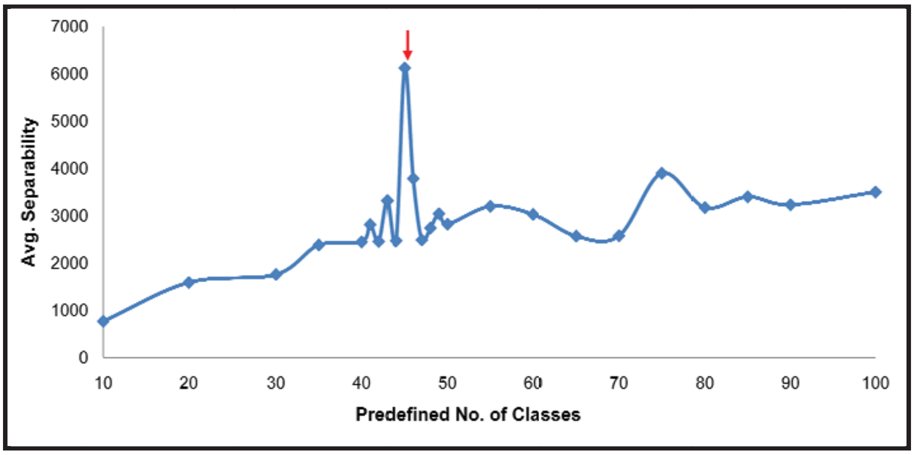


Figure 2-4: Average Separability between classes in Andalucía

de Bie *et al*. (2009)

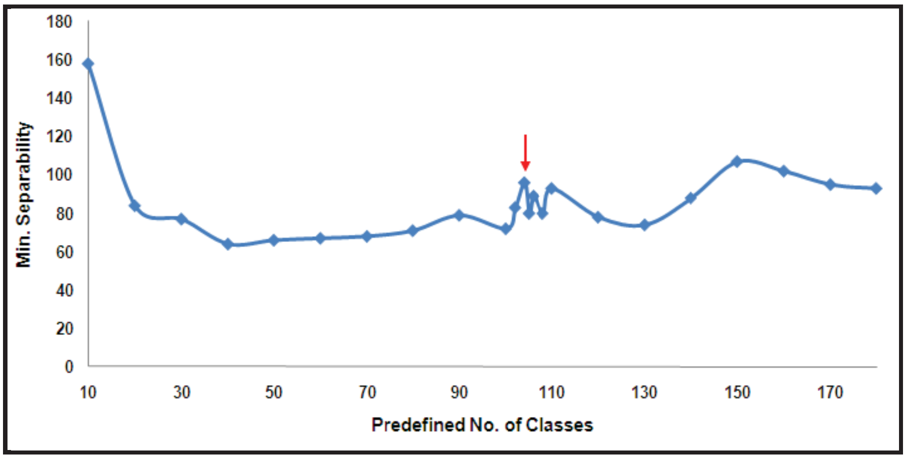


Figure 2-5: Minimum Separability between classes in Spain

2.5. Multi Collinearity Analysis

Multi-collinearity refers to a situation in which the independent or predictor variables are highly correlated (Fox 1997). When independent variables are multi-collinear, there is overlap or sharing of predictive power. This may lead to the paradoxical effect whereby the regression model fits the data well, but none of the predictor variables has a significant impact in predicting the dependent variable.

This is because when the predictor variables are highly correlated, they share essentially the same information. Thus together they may explain a great deal of the dependent variable, but may not individually contribute significantly to the model. Highly correlated variables may cause some serious problems in validation, interpretation, and analysis of the model, such as unstable estimates and high-standard errors (Fox 1997).

The Variance Inflation Factor (VIF) measures the impact of collinearity among the variables in a regression model (Montgomery and Peck 1982). The VIF is 1/Tolerance, it is always greater than or equal to 1. VIF measures the impact of collinearity among variables in a regression model on the precision of estimation. It expresses the degree to which collinearity among the predictors degrades the precision of an estimate. Typically, a VIF of greater than 10 is of concern. The absence of collinearity is as such essential for a regression model!

Mathematically speaking: $VIF = \frac{1}{1-R^2}$ eqn. (2-3)

Where R^2 is the multiple correlation coefficient.

Checking for multi-collinearity (Montgomery and Peck 1982) was accomplished using SPSS 16.0 by requesting the display of linear regression “Tolerance” and “VIF” values for each predictor variable. All the categorical variables: classified NDVI, CLC and soil type were excluded from the collinearity diagnose as they cannot be tested for collinearity.

During collinearity tests, parameters with the highest values were removed from the list. This was followed by running the collinearity test again and again, until all the remaining variables had a VIF of less than 10. The variables with high values were eliminated one by one, step by step, as the removal of that variable could cause the values of all the other variables to change drastically when that one variable was removed (Bert, personal communication). The ultimate purpose was to keep as many parameters with different meaning in the list of final predictor variables as possible.

Slope, altitude and precipitation of the warmest quarter were included in the model even though they had high VIF values. Based on expert knowledge, these variables are potentially important for explaining the distribution of *Timon lepida* and may impose distributional limits on its range presently and in the future.

Table 2-5: Multicollinearity Diagnose results for predictor variables

Predictor	Environmental Variable Description	VIF
1	Temperature seasonality (standard deviation * 100)	8.406
2	Cloud cover	1.710
3	Volume of stones	7.097
4	Altitude	8.851
5	Southness	4.879
6	Westness	4.196
7	Aridity Index	9.613
8	Precipitation of wettest quarter	6.348
9	Global Annual Radiation	3.632
10	Potential Evapotranspiration	5.397

2.6. Predictive Distribution Modelling with Maxent

Maxent version 3.3.1 available from http://www.cs.princeton.edu/_schapire/maxent/ was used for this study. Maxent was developed by the machine learning community and uses a statistical technique called maximum entropy which makes prediction from incomplete information (Phillips *et al.* 2004; Phillips *et al.* 2006). Maxent models the geographic distributions of species using geo-referenced occurrence records and environmental variables. This algorithm estimates species distributions based on presence-only data by finding the distribution of maximum entropy (i.e. most spread out, closest to uniform), subject to the constraint that the expected value of each environmental variable under this estimated distribution should match its empirical average (Phillips *et al.* 2006; Cardona and Loyola 2008). The probabilities in Maxent should sum up to 1. Maxent model output reveals the relative probability of a species distribution over all grid cells in the defined geographical space, in which a high probability value associated with a particular grid cell indicates the likelihood of this cell having suitable environmental conditions for the modelled species. Maxent probability of distribution, under convex duality, is equivalent to Gibbs Distribution and can thus be represented mathematically as:

$$q\lambda^{(x)} = \frac{e^{\lambda . f^{(x)}}}{Z_\lambda} \qquad \text{eqn. (2-4)}$$

Where λ is a vector of n real valued coefficients or feature weights
f denotes the vector of all n features and
Z_λ is a normalising constant that ensures that $q\lambda$ sums to 1

Previous studies have demonstrated Maxent's ability to predict geographical and ecological distributions of species in a wide range of ecological and geographical regions (Phillips *et al.* 2006). Maxent was chosen because it is one of the few methods available that does not require absence data. Contrary to GARP or linear regression methods which require large sample sizes, Maxent can produce useful results with sample sizes as small as 5, 10 and 25 occurrences (Phillips *et al.* 2004; Barry and Elith 2006; Phillips *et al.* 2006; Hernandez *et al.* 2008). Maxent can therefore be deemed substantially superior to the standard method since it has outperformed most other modelling methods. An added advantage of Maxent is that it also performs the ROC statistical analysis. Most importantly, it can be easily interpreted by human experts (Phillips *et al.* 2004; Phillips *et al.* 2006; Phillips and Dudik 2008; Phillips *et al.* 2009).

Major features of Maxent models for Species Distribution Modelling:

1. Pixels with known species records are called sample points, environmental variables (e.g. precipitation, elevation, slope, maximum temperature etc) constitute features and the geographical region of interest is the space on which this distribution is defined.

2. Randomly selected pixels from the study area are treated as "background pixels" or "pseudo absences" used in place of absences during modelling and the pixels in which the presence data fall are treated as positives.

3. Maxent output formats are of 3 types: Logistic, cumulative and raw. First, the raw output is just the Maxent exponential model itself. Second, the cumulative value corresponding to a raw value is the percentage of the Maxent distribution with raw value. Third, the logistic output gives an estimate between 0 and 1 of probability of presence. Of the 3, Logistic output is the easiest to conceptualize and this justifies why it has been adopted for this study.

4. Maxent can be applied to both presence only data as well as presence/absence data by using a conditional model. Both presence and absence data are needed to train a conditional model, which is why an un-conditional model was used in this study due to the uncertainty and the questionable value of absence data.

2.6.1. Model Building with Maxent

As already mentioned in Section 2.3.1, central points were utilized for modelling the distribution of *Timon lepida* both in Andalucía and Spain. Adopting a methodology from Phillips *et al.*(2006) and Yost *et al.*(2008), 10 iterative models were created with all the presence records in order to test the predictive performance of Maxent. The 10 iterative models were deemed sufficient to generate stable model performance. In each partition, Maxent was configured such that 70% of the total presence records (n=503 for Andalucía, n=2410 for Spain) were used for training the models and 30% (n=215 for Andalucía and n= 1033 for Spain) were reserved for testing the resulting models, again following the steps of (Phillips *et al.* 2006). Ten random subsets were created rather than only one in order to provide the best estimates of the species potential distribution by assessing the average behaviour of the algorithm in the different model runs. This allowed for testing the differences in performance using Wilcoxon signed rank test as well as reduce the probability of biased or uncertain predictions caused by variability within individual models (Araujo and New 2007).

It is therefore worth pointing that predictive model performance and comparisons were based on the average of the 10 subsets rather than individual model runs. Examples of studies using the average can be found in (McNees 1987; Araujo and New 2007; Yost *et al.* 2008). The same training and testing data were used in the consecutive model runs to facilitate comparability of model outputs. Logistic output format was chosen in order to represent a probability of presence. It is scale independent, calibrated such that typical presence points yields a value of 0.5 on a scale of 1.0. Logistic output format gives an estimate of probability of presence between 0 and 1. Logistic output format was also selected since it is robust to unknown prevalence being also easier to interpret as the estimated species probability of presence given the constraints imposed by environmental variables (Phillips and Dudik 2008).

All the species occurrence points were converted to .csv format as required by Maxent. The .csv files contained the species name, latitude and longitude in WGS_1984_UTM_Zone_30N. The environmental data were all harmonized to 1*1 km resolution, projected in the same system and converted to ASCII format.

Fine tuning of regularization parameters optimize the predictive accuracy of Maxent models and prevents over fitting (matching the input data too closely) which negatively affect model performance (Phillips *et al.* 2004). Taken in the language of Phillips *et al.* (2006), regularization forces Maxent to focus on the most important features. Choosing the best regularization parameters is a topic of ongoing research; refer to (Elith *et al.* 2006). For this study, regularization parameters for all model runs were user specified as:

Table 2-6: Parameterization of Maxent Model

Regularization Parameter	Tuning
Feature type	Auto
Regularization multiplier	1
Max. number of background points	10000
Maximum Iterations	1000
Convergence Threshold	0.00001

Maxent was run using "Auto features" as suggested by (Phillips *et al.* 2004). The use of default settings was reasonable since they have been used and validated over a wide range of species, environmental conditions and diverse ecological regions (Phillips *et al.* 2004).

2.6.1.1. Model Scenarios

In order to test the hypothesis that vegetation indices (classified NDVI and CLC) were better predictors in modelling the distribution of *Timon lepida*, Maxent algorithm was fitted with three sets of predictor variables:

1. Models including Environmental Variables

In this set of experiment, Maxent was fitted with all climatologic, soil and topographic variables.

2. Models including Environmental Variables and Vegetation

In the 2nd set of tests, Maxent was fitted with all the environmental variables as mentioned in (1) and in addition, classified NDVI and CLC were tested for significance.

3. Models including Vegetation Indices

In the 3rd set of experiments, all the climatologic variables (e.g. temperature, precipitation) were excluded from the model and Maxent was fitted with only classified NDVI, CLC, soil and topographical variables. For a full list of the environmental variables for the 3 sets of models, see appendix A). The aim of this experiment was to investigate whether classified NDVI and land cover variables would predict better in the absence of the climatological variables.

Table 2-7: Model Scenarios

Model Scenario	Included Model Variables
1.Environmental Variables	Climatological, Topographic and Soil
2. Environmental & Vegetation	Climatological, Topographic, Soil and Vegetation
3. Vegetation Indices	Topographic, Soil and Vegetation

2.6.2. Relative contribution of Environmental Variables

An important application of species distribution model is to answer the question of which environmental variables mostly affect the species being modelled. Maxent can address this question in two ways:

1. Variable Percentage Contribution

While Maxent is being trained, it is possible for the modeler to keep track of which environmental variables are making the greatest contribution to the model. Maxent model assigns the increase in gain to the environmental variable(s) that the feature depends on. At the end of the training process, the gain is converted to percentage.
The relative contributions of the environmental variables are only heuristic estimates and caution must be used when employing this method as strong collinearity can influence results by indicating more importance for one of two or more highly correlated variables.

2. Jackknife Test

Jackknife test can also be performed in order to show relative importance of environmental variable (Phillips *et al.* 2004; Phillips *et al.* 2006; Phillips and Dudik 2008). During Jackknife test, each variable is excluded in turn, and a model is created with the remaining variables. Then a model is created using each variable in isolation. In addition, a model is created using all variables as before. The environmental variable with the highest gain when used in isolation will make the greatest influence in the species being modelled. The environmental variable that decreases the gain the most when it is omitted appears to have the most information that is not present in other variables therefore it is likely to be highly influential. Variable importance in this study was determined using Jackknife test since it is a more reliable measure compared with the percentage contribution values mentioned in (1). In so doing, Jackknife provides information on the performance of each variable in the model in terms of how important each variable is at explaining the species distribution and how much unique information each variable provides. Jackknife test takes the following steps:

1. Run the model with all the predictor variables
2. Check Jackknife test result of the model
3. Remove the variable with the most negative effect on the total gain/variable with lowest decrease in the training gain when omitted. In other words, omission of this variable will result in a higher training gain that when included.
4. Run the Jackknife test again with the remaining variables in order to evaluate their predictive capabilities
5. Omit another variable that has the most negative effect to the total gain
6. Repeat step 4 and 5 until all the variables have a positive effect to the total gain

The final models excluded volume of stones. This variable was dropped due to its low overall contributions to the model performance in Jackknife tests. Liu *et al.* (2005) noted that conclusions about model performance requires serious statistical testing; as such the model partitions trained without volume of stones were statistically compared with the full predictor variables models in order to determine which model partitions were more significant. The model subsets with the full set of predictor variables were found to be more significant using the Wilcoxon sign rank test (*p*-value=0.002) and hence were used for modelling in this study.

2.7. Model Output Evaluation, Stability and Comparison

Assessing the predictive power of a model is of paramount importance, both for theoretical and applied issues. Araujo and Guisan (2006) describe model evaluation as "the testing process required to justify the acceptance of a model for its intended purpose". Many indices can be used in the assessment of species distribution models performance including specificity, sensitivity and Cohen's Kappa (Cohen 1960; Liu *et al.* 2005; Allouche *et al.* 2006). This study employed both threshold dependent and independent evaluation methods.

2.7.1. Threshold-dependent Evaluation

2.7.1.1. Binomial Test

In order to investigate whether the models performed significantly better than random, a threshold dependent binomial test was used based on omission rates and predicted area. Maxent models automatically calculate the statistical significance of the prediction using test omission rates and fractional predicted area and provide the corresponding *p*-values, which can be used to directly evaluate the performance. A one tailed binomial test was used to determine whether the models predicted the test localities significantly better than random at the "cumulative threshold of 1, 5 and 10" following (Anderson 2003; Phillips *et al.* 2006; Yost *et al.* 2008). The binomial test requires that threshold be set in order to convert predictions into binaries showing suitable and unsuitable areas for the species. After applying a threshold, the model performance can be investigated using the *extrinsic omission rate*, which is the fraction of all test localities that fall into pixels not predicted as suitable for the species and the *proportional predicted area*, which is the fraction of all the pixels that are predicted as suitable for the species. As mentioned by (Anderson 2003), a low omission rate is a necessary condition for potentially predicting the species distribution ranges. Determining optimal thresholds for Maxent models is an area of ongoing research and there is no thresholding rule that has been developed for the algorithm yet (Phillips *et al.* 2006).

2.7.1.2. Kappa Statistics

The spatial accuracy of the predictions was also determined using the threshold dependent (omission and predicted area) by calculating the "equal test sensitivity and specificity" (Araujo and New 2007; Freedman *et al.* 2008; Martinez-Freiria *et al.* 2008; Sillero 2009). This measure minimises over and under-prediction associated with other threshold measures (Phillips *et al.* 2004; Lobo *et al.* 2008). The equal test sensitivity and specificity measure was used for selecting the threshold for kappa calculation. Kappa statistics is an index of inter-rater reliability that is commonly used to measure the level of agreement between two sets of dichotomous ratings or scores (Cohen 1960; Manel *et al.* 2001; Guisan *et al.* 2002). Kappa statistics can be used to objectively assess the level of agreement between observed and predicted data.

In species distribution modeling, Kappa statistic can be computed by error matrix

Table 2-8: Error matrix for Kappa calculation

		Observation		
		Recorded Presence	**Recorded Absence**	
Prediction	**Predicted Presence**	a (true positive)	b (false positive)	a+b
	Predicted Absence	c (false negative)	d (true negative)	c+d
		a+c	b+d	n

Where:
a is the number of correctly predicted presences
b is the number of incorrectly predicted presences
c is the number of correctly predicted absences
d is the number of incorrectly predicted absences

The confusion matrix records the frequencies of each of the four possible types of prediction from analysis of test data:

1. True positive (the model predicts that the species is present and test data confirms this to be true),

2. False positive (the model predicts presence but test data show absence),

3. False negative (the model predicts absence but test data show presence),

4. True negative (the model predicts and the test data show absence).

To calculate Kappa, you first have to calculate the Observed level of agreement:

$$Po = \frac{a+d}{n} \qquad \text{eqn. (2-5)}$$

Expected level of agreement:

$$Pe = P_{\cdot 1}P_{1\cdot} + P_{\cdot 2}\,P_{2\cdot} \qquad \text{eqn. (2-6)}$$

Where $p_{\cdot 1} = \frac{a+c}{n}$ eqn. (2-7)

$$P_{1\cdot} = \frac{a+b}{n} \qquad \text{eqn. (2-8)}$$

$$p_{\cdot 2} = \frac{b+d}{n} \qquad \text{eqn. (2-9)}$$

$$p_{2\cdot} = \frac{c+d}{n} \qquad \text{eqn. (2-10)}$$

Therefore Kappa is calculated using the equation:

$$K = \frac{Po-Pe}{1-Pe} \qquad \text{eqn. (2-11)}$$

Where o is the observed level of agreement
e is the expected level of agreement

(Monserud and Leemans 1992) suggested the following ranges to describe the levels of agreement for the Kappa statistic.

Table 2-9: Kappa Range and Interpretation

Kappa Range	Interpretation
<0.05	No agreement
0.05-0.20	Very poor
0.20-0.40	Poor
0.40-0.55	Fair
0.55-0.70	Good
0.70-0.85	Very good
0.85-0.99	Excellent
0.99-1.00	Perfect

Negative values indicate extremely poor agreement.

Kappa indicates correctly classified presences or absences after accounting for chance effects (Moisen and Frescino 2002). A value of 1 will indicate perfect agreement, >0.75 indicates good model; > 0.80 will indicate excellent agreement. 0.50 indicates a performance not better than chance. Kappa statistics requires that a threshold be set in order to convert continuous predictions into presence/absence or suitable and unsuitable areas for *Timon lepida*. While various threshold determining approaches exist, subjective approaches should be avoided, for example, choosing a fixed value of 0.5, as these approaches are very arbitrary and lack any ecological basis (Osborne *et al.* 2001). Although there are so many approaches to determining thresholds, there is no comparative study on their relative performance. Following Cantor *et al.*(1999) and Liu *et al.*(2005), the point at which sensitivity and specificity are equal was used to determine the threshold for computing the confusion matrix. Kappa was computed using SPSS 16.0 by running: Analysis>Descriptive statistics> Crosstabs.

2.8. Threshold Independent Evaluation

2.8.1. Area Under the Curve (AUC)

The Area Under the Curve of the Receiver Operating Characteristic (ROC) measures the quality of a ranking of sites (Guisan and Hofer 2003). Practical difficulties in evaluating predictions from presence only data models have been noted by (Pearce and Ferrier 2000) since absence data is missing and therefore cannot be used to assess model predictions. A common method is to use background points or pseudo absences (Hirzel *et al.* 2001). Pseudo-absences are sites, randomly selected across the geographical area of interest, at localities where no species presence was recorded and for which species occurrence is set as absent (Anderson 2003; Elith *et al.* 2006; Phillips *et al.* 2006). Maxent uses the randomly selected background pixels as negative instances and the pixels in which the presence data falls as positive instances (Yost *et al.* 2008).

The Area Under the Curve is the probability that a randomly chosen presence site will be ranked above a random background site. The ROC AUC is determined by plotting sensitivity (true positives) on the y-axis against specificity (false positives) on the x-axis for all possible thresholds as described by Wilson *et al.* (2005). ROC predicts using values between 0 and 1. The main advantage of ROC is that that area under the ROC curve (AUC) provides a single measure of model performance, independent of any particular choice of threshold. The AUC has an intuitive interpretation, namely the probability that a random positive instance and a random negative instance are correctly ordered by the classifier. This interpretation indicates that the AUC is not sensitive to the relative numbers of positive and negative instances in the test data set. An AUC value can be interpreted as indicating the probability that, when a presence site (site where a species is recorded as present) and an absence site (site where a species is recorded as absent) are drawn at random from the population, the presence site has a higher predicted value than the absence site (Elith *et al.* 2006; Phillips *et al.* 2006).

When AUC is applied to presence only data and pseudo absences it would appear that ROC curves are inapplicable, since without absences, there seems to be no source of negative instances with which to measure specificity, the maximum achievable AUC value is no longer 1 but 1- a/2 (where a is the fraction of the geographical area of interest covered by a species' true distribution, which typically is not known (Phillips *et al.* 2006) and is even smaller for wider ranging species.

Table 2-10: AUC Range and Interpretation

AUC Range	Interpretation
0.90-1	Excellent
0.80-0.90	Good
0.70-0.80	Fair
0.60-0.70	Poor
0.50-0.60	Fail

The stability of the models was evaluated by comparing the test dataset with the training dataset. The stability values were calculated using the following equations:

$$Stability_{AUC} = \frac{AUCtest}{AUCtraining} \quad \text{eqn. (2-12)}$$

$$Stability_{Gain} = \frac{\text{Regularized test gain}}{\text{Regularized training gain}} \quad \text{eqn. (2-13)}$$

In the threshold independent test, the AUC for the dataset was analyzed to see how significantly each model prediction differed from random using the Wilcoxon signed rank test (Phillips *et al.* 2006). Two tailed Wilcoxon signed rank test was used to evaluate whether there was a statistical difference in the performance of the 3 model scenarios. The average AUC values of the models were compared using a two tailed Wilcoxon sign rank test which estimates the significance of the model using a *p*-value. The Wilcoxon signed ranked test, which is almost as powerful as the t-test, works by ranking two predictions better than another (Elith *et al.* 2006). The *p*-values were tested on the AUC values for all the random subsets, using significance of 95% (alpha 0.05) for each set of model. Tests that allowed us to determine significant differences were run e.g. to determine if the models built using environmental variables were better than models built using vegetation indices or better than a combination of environmental variables and vegetation. The *p*-values were compared with the alpha (0.05), if the *p*- value is less than or equal to alpha, then the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid. Finally, by using Friedman's analysis of variance (ANOVA) for multiple dependent samples, we determined if there was a significant difference among all the 3 model scenarios. Non parametric tests were performed since normality within the datasets was violated.

2.9. Habitat Suitability Maps

A normal practice in species distribution modelling is to select one of the best performing models to be used for mapping the distribution of a species in the study area among a random set of partitions. “Best” is one of the consensus methods used in species distribution modelling (Marmion *et al.* 2009) and it is based on picking up the best of the separate models in all the partitions based on some predefined criteria. Following Yost *et al.* (2008) and Seoane *et al.*(2005) the best model was the one with the highest regularized training gain given the higher sensitivity of this measure relative to the AUC. Continuous suitability maps were chosen to represent the potential geographic distribution of *Timon lepida* since they reveal a whole gradient in habitat suitability and avoid the subjective task of choosing a threshold. The output maps were also evaluated visually to see how well they fit with the presence records. According to (Yost *et al.* 2008), a good model will produce regions that correspond well to the points of the presence records.

2.10. Employed Software

ERDAS Imagine 9.3 was used for NDVI image pre-processing and for performing unsupervised classification. Maxent 3.3.1 was used for the modelling process. ArcGIS 9.3 was used for data preparation and post modelling map generation. SPSS 16.0 and Statistica 7.0 were used for non parametric tests of the model results. Microsoft word and Excel 2007 were used for analysis and reporting. EndNote X2 was used for writing references.

3. Modelling Results for the potential distribution of *Timon lepida* in Andalucia and Spain

3.1. Introduction

This section describes the model prediction results for *Timon lepida* in Andalucía and Spain under the following sub headings:

- Potential geographic distribution of *Timon lepida*
- Model Comparison and Stability Evaluation
- Predictor variables determining the potential geographic distribution of *Timon lepida*

3.2. Potential geographic distribution of *Timon lepida* in Andalucia

Maxent output maps of the predicted geographic distribution in Andalucía are shown in Figure 3-1. The potential distribution maps revealed that *Timon lepida* thrives in a wide range of habitats. Suitable areas and unsuitable areas are widely distributed throughout the whole Andalucía province which made it difficult to distinguish suitable versus unsuitable habitats for the species. Model outputs were visually very similar which also made comparison between three model outputs difficult. From all the prediction maps, it is however clear that low suitability was predicted in the central parts in areas around Seville and Cordoba and the south eastern extreme corner around Almeria. This north to south spatial constriction corresponds to the presence of the Guadalquivir river basin in the central and the dominance of arid areas of Almeria to the south east. All the model scenarios also indicated unsuitable conditions in the south-eastern parts in Sierra Nevada Mountains.

The potential distribution map based on environmental variables (Figure 3-1b) pointed high suitability areas to most parts of the southern western extreme corner of Andalucía in areas including Malaga. These regions correspond indeed to the areas with known presences of *Timon lepida* (Figure 3-1a). However, if we compare the 2 maps of potential distribution that included vegetation (Figure 3-1c, d) with the output map that included environmental variables, small differences exist. These 2 maps identified areas of low suitability in some parts of the western Andalucía traversing Huelva possibly because of the presence of marshy areas of the river Tinto and Odiel.

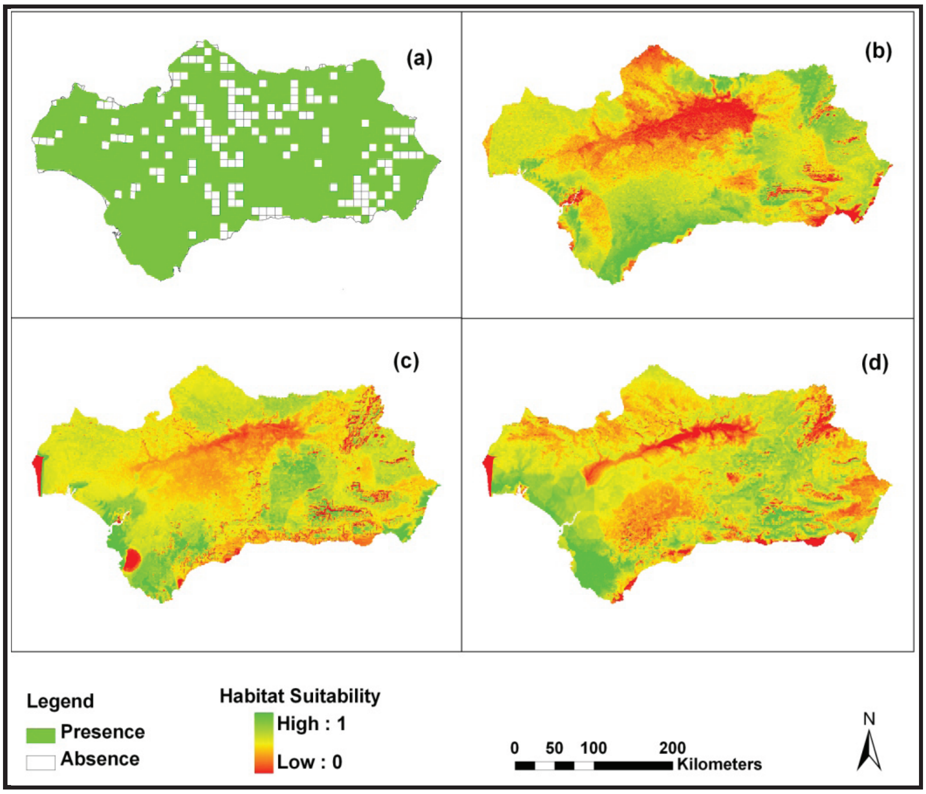


Figure 3-1: Distribution of *Timon lepida* in Andalucía

(a) Documented distribution; green represents presence and white represents absence (b) Suitability based on Environmental Variables (c) Environmental Variables and Vegetation, (d) Vegetation Indices

3.2.1. Model Evaluation and Stability Comparisons

3.2.1.1. Binomial Test and Kappa Results

Results for the threshold dependent tests are presented in Table 3-1. Except for the threshold level of 1, *p*-values from the binomial test revealed that all the model scenarios performed significantly better than random because they attained ($p<0.005$) at the thresholds of 5 and 10 (Table 3-1). An interesting part of this analysis is that as the threshold changed from 1 to 10, the binomial *p*-values from the models decreased considerably to less than 0.005 ($p<0.005$) hence signifying highly significant predictions. This indicated that as the threshold changed from 1 to 10, there was a higher probability of rejecting the null hypothesis for all the models. Reliability of the model predictions using Kappa statistics demonstrated that models generated using environmental variables had the highest Kappa value followed by models that combined both environmental variables and vegetation. Comparatively, low Kappa values were obtained when evaluating models based on vegetation indices (see Table 3-1).

Table 3-1: *p*- values from the Binomial test for Andalucía

	Binomial Test *p*-value for threshold			Kappa-Value
	1	5	10	
Spatial Model Scenario				
Environmental Variables Only	0.0379	0.00113	0.00049	0.512
Environmental Variables & Vegetation	0.0543	0.00106	0.000007	0.481
Vegetation Indices	0.0179	0.00071	0.000490	0.467
Kappa was calculated using equal training sensitivity and specificity threshold				

3.2.1.2. ROC/AUC Analysis

The predictive performance of the 3 sets of models trained in Andalucía is summarized in Table 3-2.

Table 3-2: Species distribution-model summary statistics for Andalucía

	Environmental Variables	Environmental &Vegetation	Vegetation
Training AUC	0.917	0.920	0.864
Test AUC	**0.753**	0.739	<u>0.690</u>
Stability_{AUC}	**0.821**	0.803	<u>0.798</u>
Model Gain			
Regularized training Gain	0.705	0.735	0.529
Test gain	**0.525**	0.503	<u>0.327</u>
Stability_{Gain}	**0.745**	0.684	<u>0.618</u>

The results represent mean values of random partitions (n=10) with all occurrence localities included. The bold indicates the most stable and accurate model based on the test dataset and the underlined indicates the model with the lowest performance considering the above criteria.

The accuracy of the modelled distributions based on test localities was significantly better than random for all the models (test AUC>0.5) (Elith *et al.* 2006). Model scenarios generated including vegetation indices were statistically compared to correspondent distribution models generated without vegetation indices. The AUC values (calculated on the extrinsic test data) of the models built with environmental variables was higher than that of the models that combined environmental variables and vegetation and those that were built using vegetation indices. The test AUC showed a fair performance (shown in Table 3-2).

Models generated using both environmental variables and vegetation were not statistically superior among all the models. This outcome defies logical expectation, where we anticipated a strong response to the combination of the two suites of variables both in terms of higher AUC and training gain. This is unlike the observation made by (Phillips *et al.* 2006) that the inclusion of vegetation should increase the AUC since there is more information available to the classifier.

3.2.2. Predictor Variables determining the distribution of *Timon lepida* in Andalucía

3.2.2.1. Models including Environmental Variables

When modelling using environmental variables, the analysis of variable importance provided by Maxent Jackknife test ranked slope as the most important predictor variable, south exposure as the 2nd and west exposure as the 3rd (Figure 3-2) determining the occurrence of *Timon lepida* in Andalucía. The environmental variable with the highest gain when used in isolation was south exposure (0.176) which therefore appeared to have the most useful information by itself vis-a-vis other variables. The average regularized training gain for this model was 0.705 and this was the highest compared to the average gains of the models built using environmental variables and vegetation and models built using vegetation indices.

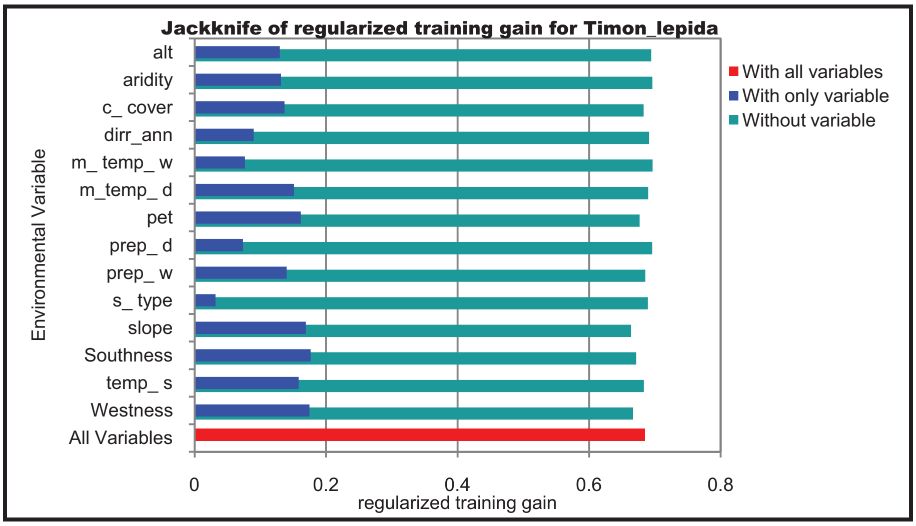


Figure 3-2: Jackknife test results for Environmental Variables model in Andalucía

3.2.2.2. Models including Environmental Variables and Vegetation

Models trained in Andalucía using environmental variables and vegetation depicted south exposure, slope and west exposure as the 3rd most important variables looking at the drop in gain when these variables were omitted (Figure 3-3). Looking at the training gain with only variable, south facing slope emerged as the top most contender with the highest gain when used in isolation thus suggesting this variable alone contained the most useful information by itself. The regularized training gain of this model was higher (0.735) than that of the models fitted with environmental variables.

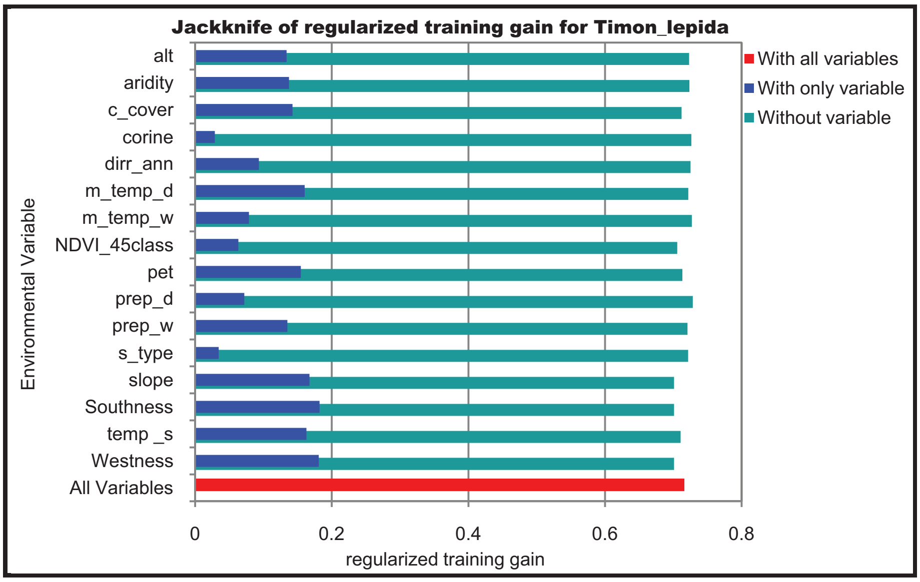


Figure 3-3: Jackknife test results for Environmental Variables and Vegetation model in Andalucía

3.2.2.3. Models including Vegetation Indices

Models trained in Andalucía using vegetation indices depicted south exposure as the 1st, slope as the 2nd followed by west exposure as the 3rd most important predictor variables respectively. South exposure decreased the gain the most when it is omitted from the model suggesting that it harboured the most information that was not present in other variables and it had the most useful information by itself (see Figure 3-4). The regularized training gain of the model fitted with vegetation was 0.529 which was the lowest among the 3 models at the landscape level. A comparative analysis of the three models showed that in all instances and for all partitions, topographical variables had the greatest contribution in defining areas of *Timon lepida* occurrence in Andalucía.

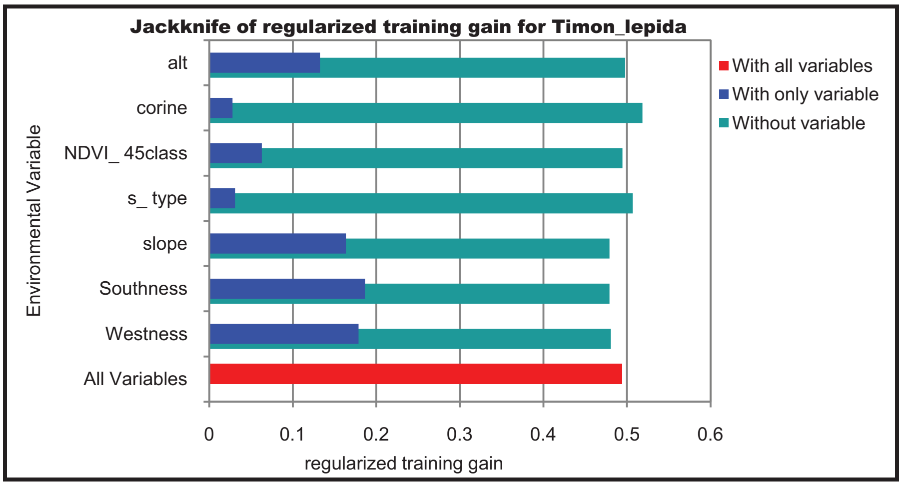


Figure 3-4: Jackknife test results for Vegetation Indices model in Andalucía

3.3. Potential Geographic Distribution of *Timon lepida* in Spain

The potential geographic distribution maps of *Timon lepida* in Spain are shown in Figure 3-5. All the model scenarios predicted a relatively lower distribution in the northern part of Spain and increasing probability of distribution moving towards the central and southern parts of the country. Maxent produced impressive predictions in the northern part corresponding well to the Atlas data (Figure 3-5a). The potential distribution maps based on environmental variables and those based on environmental variables and vegetation indicated a spatial constriction in the southern part in areas coinciding with the Guadalquivir river basin as well as the south eastern corner in the half desert of Almeria. In addition, models based on environmental variables indicated unsuitable conditions in the highland areas of Sierra Nevada Mountains but the models based on environmental variables and vegetation failed to capture this spatial constriction. On the other hand, predictions based on vegetation indices (Figure 3-5c) agreed well with the other models to a spatial constriction in the half desert of Almeria as well as the Sierra Nevada Mountains, but on the contrary, showed areas of suitability even in areas around the Guadalquivir depression.

By comparing the documented distribution ranges of *Timon lepida* with the predicted distribution maps: an important difference in the species geographical distribution can be clearly visualized. There is an increase in the western part in Balearic Islands which could be due to the fact that (1) there exists suitable habitats that the species has never occupied or it could be due to (2) over prediction of the potential distribution by all the models. In general, all the models produced broadly similar and widespread geographic distribution which made comparison between the different model scenarios difficult.

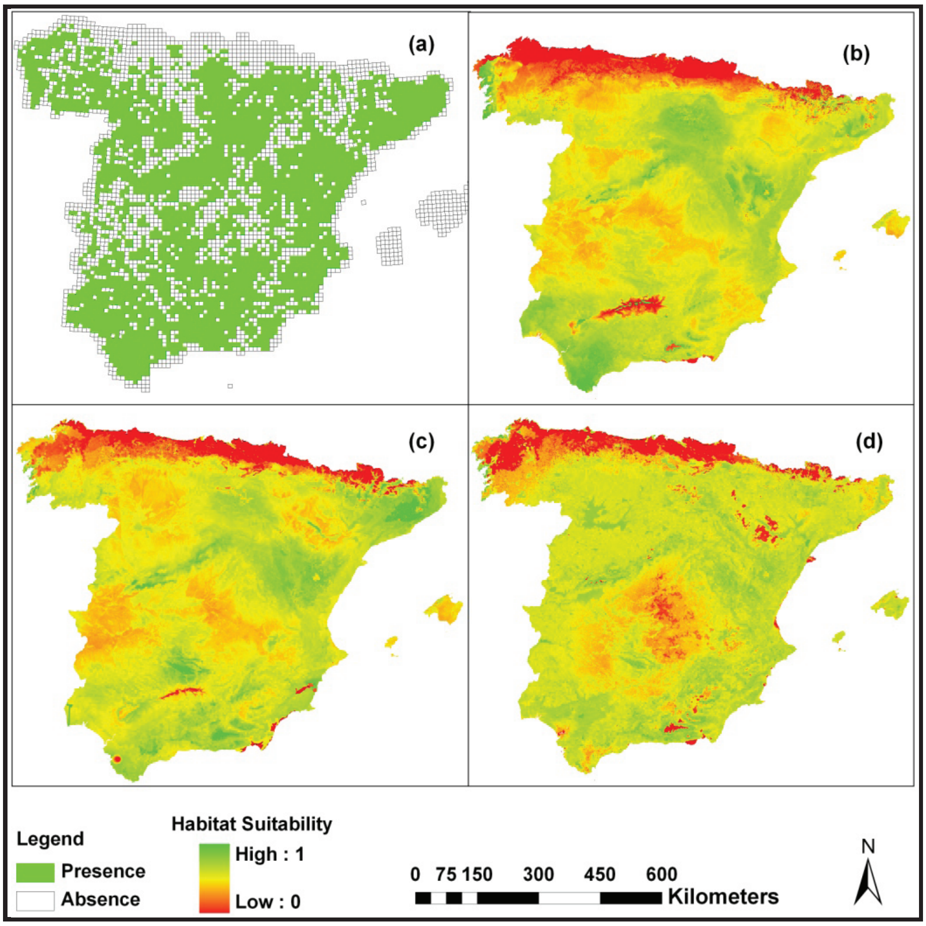


Figure 3-5: Distribution of *Timon lepida* in Spain

(a) Documented distribution; green represents presence, white represents absence (b) Suitability based on Environmental Variables (c) Environmental Variables and Vegetation, (d) Vegetation Indices

3.3.1. Model Evaluation and Stability Comparisons

3.3.1.1. Binomial Test and Kappa Results

The first step in evaluating model performance was to determine if the models predicted the test localities significantly better than random using the binomial test. Results for the threshold dependent tests are presented in Table 3-3:

Table 3-3: *p*-values from the Binomial test for Spain

	Binomial Test *p*-value for threshold			Kappa-Value
	1	5	10	
Spatial Model Scenario				
Environmental Variables	0.03700	0.00305	0.00039	0.503
Environmental Variables & Vegetation	0.07400	0.00253	0.00376	0.458
Vegetation Indices	0.01202	0.07427	0.00415	0.421

Kappa was calculated using equal training sensitivity and specificity threshold

All the models yielded predictions that were not better than random at the threshold of 1 (*p*-value>0.005). The models generated using environmental variables and those that combined environmental variables and vegetation yielded predictions that were significantly better than random (*p*-value<0.005) at the thresholds of 5 and 10 respectively. The model scenario that was generated including vegetation indices improved performance at the threshold level of 10 (see Table 3-3). In general, models including environmental variables always seemed to give a better prediction. Looking at the Kappa statistics, all the models indicated a fair prediction. The models generated using environmental variables had a slightly higher Kappa value among all the model scenarios (Table 3-3).

3.3.1.2. ROC/AUC Analysis

The predictive performance of the 3 sets of models trained in Spain is summarized in Table 3-4. In terms of the AUC, models built with environmental variables had the highest predictive performance (AUC training score=0.914, test score=0.668). Models built with both environmental variables and vegetation had slightly weaker predictive strength (AUC training score=0.906, test score=0.640) while models built with vegetation indices obtained the lowest performance (AUC training score =0.858, AUC test score=0.612). The test results indicate that models built using environmental variables were the most powerful in discriminating suitable from unsuitable habitats. Based on the average training gain and AUC criteria, this model was more stable compared to the other models (Stability_{AUC}=0.737; stability_{Gain}= 0.315).

Table 3-4: Species distribution-model summary Statistics for Spain

	Environmental Variables	Environmental &Vegetation	Vegetation
Training AUC	0.914	0.906	0.858
Test AUC	**0.668**	0.640	0.612
$Stability_{AUC}$	**0.737**	0.700	0.713
Model Gain			
Regularized training Gain	0.614	0.700	0.511
Test gain	**0.194**	0.116	0.057
$Stability_{Gain}$	0.315	0.166	0.111

The results represent mean values of random partitions (n=10) with all occurrence localities included. The bold indicates the most stable and accurate model based on the test dataset and the underlined indicates the model with the lowest performance considering the above criteria.

Though the environmental and vegetation model had a higher test AUC than the vegetation indices model, models built using vegetation indices appeared to be more stable than it (0.713) compared to (0.700) for environmental and vegetation. Addition of vegetation to the other environmental variables did not increase the AUC. Indeed the AUC generally decreased (0.640). $Stability_{AUC}$ (0.713) of the models built with vegetation was higher than that of the models that combined environmental variables and vegetation but the $stability_{Gain}$ (0.057) was the lowest among the 3 model scenarios. An interpretation of all the 3 models is that they all gave a poor prediction of the test localities but gave very good prediction of the training localities.

3.3.2. Predictor Variables determining the distribution of *Timon lepida* in Spain

3.3.2.1. Models including Environmental Variables

At a regional scale of Spain, Maxent Jackknife test results for models generated with environmental variables ranked temperature seasonality as the most important predictor variable determining the occurrence of *Timon lepida*. This variable decreased the training gain the most when omitted from the model suggesting that it contained the most useful information that was not present in other variables. Altitude and solar radiation were also top predictors occupying the 2^{nd} and 3^{rd} positions respectively. These variables also contained the most useful information by themselves (Figure 3-6). The regularized training gain of the models fitted with environmental variables was 0.614 which was lower than that of the models fitted with both environmental variables and vegetation (0.700). This model scored the highest stability gain (0.315).

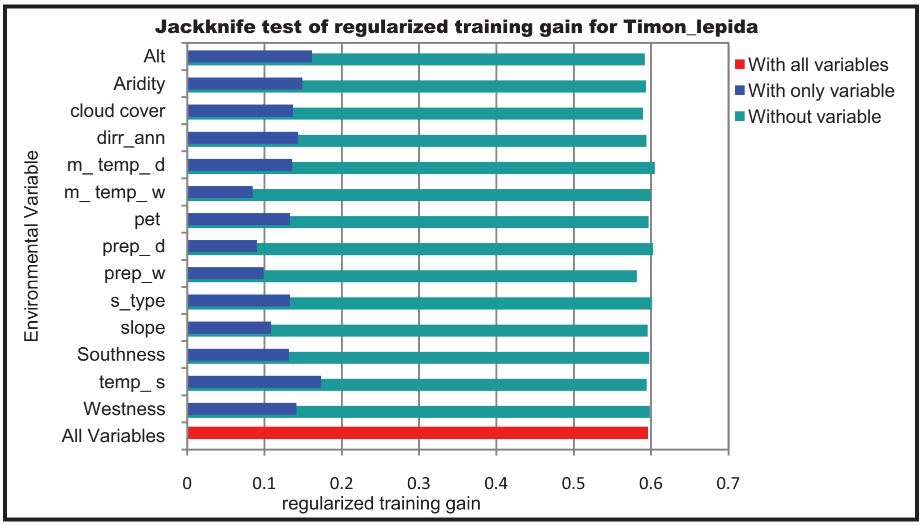


Figure 3-6: Jackknife test results for Environmental Variables model in Spain

3.3.2.2. Models including Environmental Variables and Vegetation

In contrast to Andalucía, models generated with both environmental variables and vegetation at a regional scale depicted classified NDVI as the single most important variable that decreased the gain the most when omitted from the model, which suggests that this variable contained most information that was not present in other variables. Potential Evapotranspiration (PET) and temperature seasonality emerged as the 2nd and 3rd contenders respectively (Figure 3-7). Additionally, aridity, solar radiation and altitude played a major role in defining the occurrence of the target species. When used in isolation these environmental variables still contained most environmental information not accounted for by any other predictor variables. Models fitted with both environmental variables and vegetation achieved the highest regularized training gain (0.700). In terms of stability, this model was less stable than models built using environmental variables with a stability gain of (0.166).

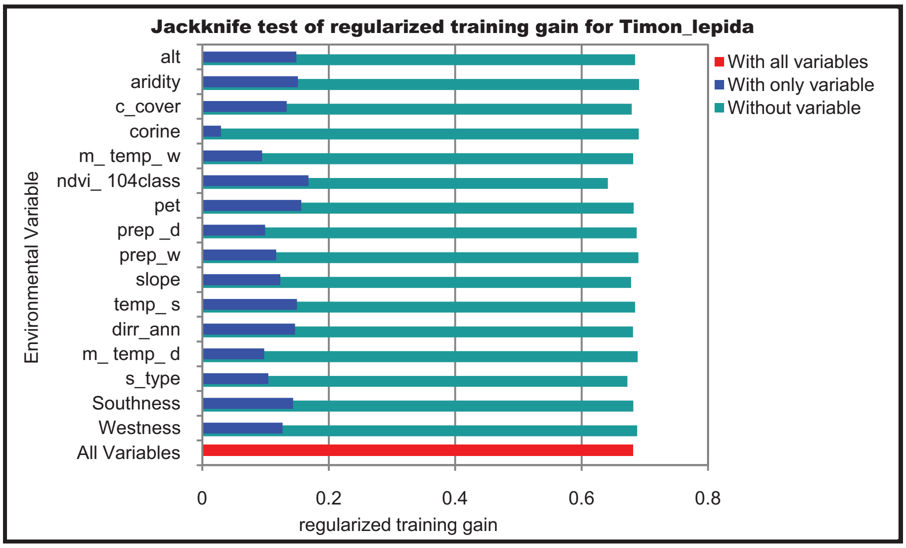


Figure 3-7: Jackknife test results for Environmental Variables and Vegetation model in Spain

3.3.2.3. Models including Vegetation Indices

Examining experiments using vegetation indices at a regional scale singled out classified NDVI, altitude and south exposure as equally the most important variables that decreased the gain the most when omitted from the model. The variable that increased the gain the most when used in isolation was classified NDVI, followed by altitude and south exposure (see Figure 3-8). The regularized training gain and the stability gain of the model fitted with vegetation indices were the lowest (0.111) among the 3 sets of models trained in Spain. This analysis revealed that at a regional scale, climatic variables and vegetation were the most important explanatory variables defining the distribution ranges of *Timon lepida*. While topographic variables were of great importance at a landscape level, they were not the best predictors at a regional scale.

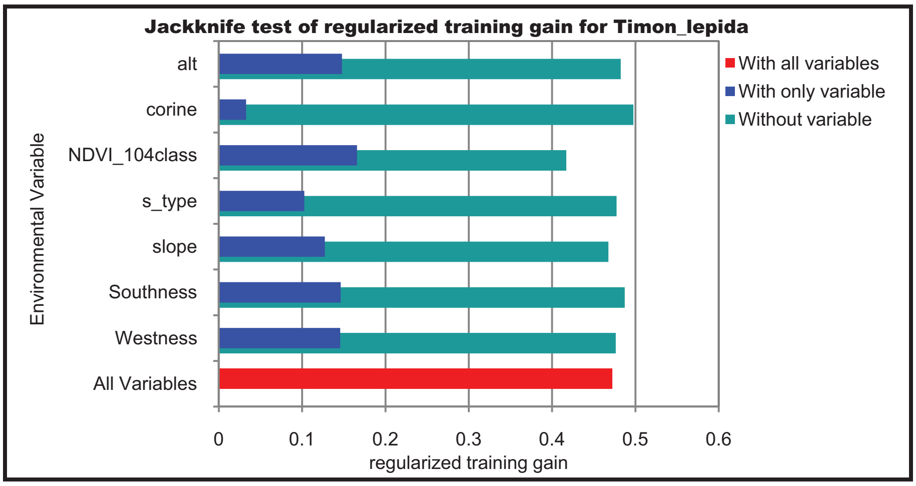


Figure 3-8: Jackknife test results for Vegetation Indices model in Spain

3.4. Model Evaluation and Stability Comparisons at different Scales

3.4.1. Binomial Test (Hypothesis 1 Testing)

A true statistical test for any model prediction is how significantly and consistently it performs better than a random prediction (Peterson 2001). As the results indicated in sections 3.2.1 for Andalucía and 3.3.1 for Spain, all the models at the two spatial scales performed significantly better than random (0.5) when using the threshold binomial test. This implies small omission rates associated with reasonable fractions of predicted area, again suggesting meaningful model predictions (Buermann *et al.* 2008). These results were further supported by the outcome of Wilcoxon signed rank test (discussed below). Therefore with 95% confidence; we reject the null hypothesis.

3.4.2. Wilcoxon signed rank Test and Friedman's ANOVA (Hypothesis 2 Testing)

Further significance tests using the two tailed Wilcoxon signed rank test indicated that all model scenarios performed significantly better than a random prediction (p-value < 0.05). A pair wise evaluation of the test AUC values using the Wilcoxon signed rank test indicated that model predictions were significantly higher when using environmental variables (Table 3-5). Statistically speaking, models generated using environmental variables proved to be superior (p-value<0.05; smaller p-values means better models) to models generated using both environmental variables and vegetation as well as those generated with vegetation indices in terms of predictive accuracy; two tailed non parametric test of (Delong *et al.* 1988).

Table 3-5: Wilcoxon signed rank test results for Model Scenarios

	Andalucía	Spain
Environmental Variables	p= 0.002	p= 0.012
Environmental Variables and vegetation	p= 0.024	p=0.039
Vegetation Indices	p= 0.043	p=0.046

***p*-values were calculated at alpha (0.05)**

A further comparison using Friedman test confirmed this trend. The (chi-square) x^2 for Andalucía was 67.55 and *p*-value = 0.0338 whereas the x^2 value for Spain was 77.59 and *p*- value= 0.0430. These results did not support our hypothesis (section 1.5) that models that included vegetation will perform significantly better than models that included environmental variables since the models that included vegetation performed poorly.

3.5. Model Comparisons at Different Scales

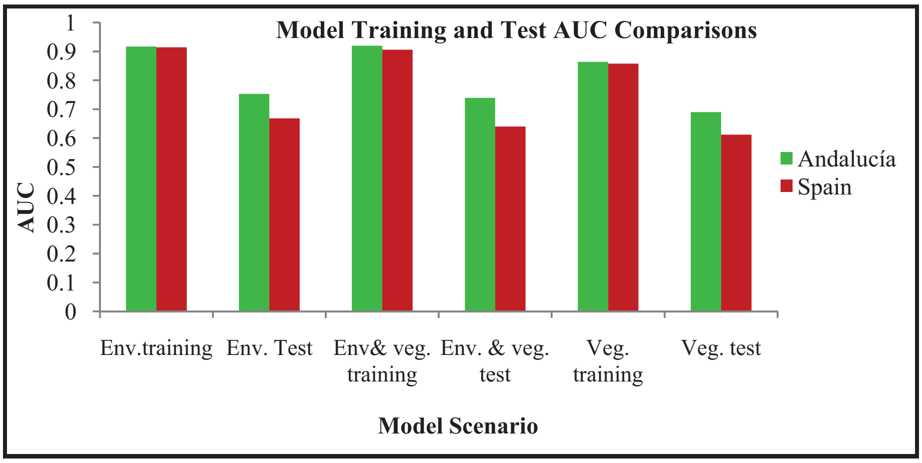


Figure 3-9: Full Variable Model AUC comparison between Andalucía and Spain

Env: Environmental variables model scenario; **Env& Veg**: Environmental variables and vegetation model scenario; **Veg:** Vegetation indices model scenario.

3.5.1. Best fit model Comparisons between Andalucía and Spain

Models generated using environmental variables have proved to be the most accurate compared to the other model scenarios both in Andalucía and Spain. The accuracy of the "best" fit models was therefore compared at these 2 spatial scales based on the AUC and training gain. Andalucía had a higher accuracy than Spain.

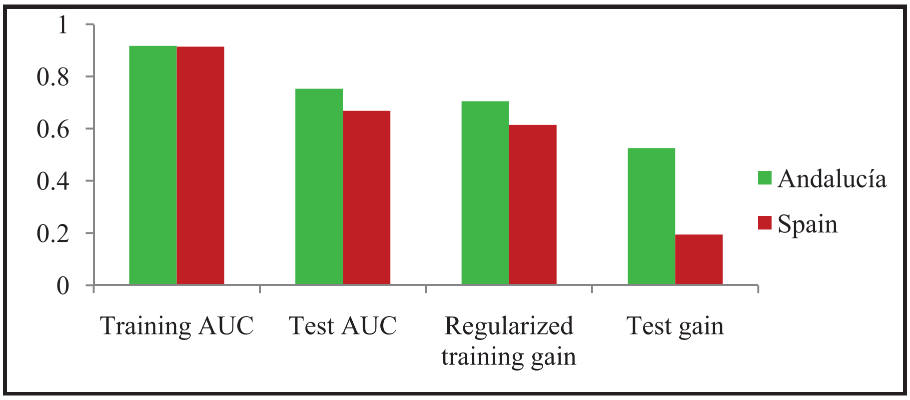


Figure 3-10: Best fit model AUC and model Gain

Table 3-6: Average AUC values for Models without Top Predictors

	Environmental Variables	Environmental &Vegetation	Vegetation
Andalucía			
Training AUC	0.847	0.756	0.864
Test AUC	0.698	0.567	0.690
Spain			
Training AUC	0.829	0.798	0.703
Test AUC	0.603	0.507	0.473

The average AUC values are based on the 10 random partitions. The results indicate that when the most important variables are omitted from the models, the AUC values drop drastically. This shows that these variables contained much more useful information than the rest of the variables. These findings are suggestive of a generally decreased discrimination of suitable vs. unsuitable habitats when top predictors were omitted.

3.5.2. Variable Percentage Contribution for Andalucía and Spain

Table 3-7: Variable importance comparison between Andalucía and Spain

Spatial Model	Andalucía	Spain
Environmental Variables	Southness, Westness,slope, pet, temp_s, m_temp_d, prep, w, c_cover, aridity, alt dirr_ann, m_temp_w, prep_d, s_type	temp_s, alt, aridity, dirr_ann, Westness c_cover, m_temp_d, s_type, pet, Southness, slope, prep_w, prep_d, m_temp_w
Environmental Variables and Vegetation	Southness, Westness,slope, temp_s, m_temp_d, pet, c_cover, aridity, prep_w, alt, dirr_ann, m_temp_w, prep_d NDVI_45class, s_type, corine	NDVI_104class, pet, aridity, temp_s, alt, dirr_ann, Southness, c_cover, Westness, slope, prep_w, s_type, prep_d, m_temp_d, m_temp_w, corine
Vegetation Indices	Southness, Westness, slope, alt, NDVI_45class, s_type, corine	NDVI_104class, alt, Southness, Westness, slope, s_type, corine

Variable descriptions match the ones used in the model scenarios. This table shows variable contribution in terms of the gain they add to the model and they have been ranked from the most contributing to the least contributing. Full variable names are listed in Appendix A.

Table 3-8: Environmental Variables percentage Contribution

Variable	% Contribution Andalucía	Variable	% Contribution Spain
southness	14.1	temp_s	13
Westness	13.1	Alt	12.2
slope	12.7	dirr_ann	11
pet	9.4	aridity	9.3
temp_s	8.4	c_cover	8.8
c_cover	7.9	slope	7
m_temp_w	6.7	southness	6.2
dirr_ann	6	m_temp_d	5.9
prep_w	5.8	westness	5.9
aridity	5.5	pet	5.8
prep_d	3.3	prep_d	5.6
m_temp_d	2.8	s_type	5.3
s_type	2.2	m_temp_w	2
Alt	2.1	prep_w	2

Table: 3-9: Vegetation Indices percentage Contribution

Variable	% Contribution Andalucía	Variable	% Contribution Spain
Southness	27.9	NDVI_104class	28
Westness	26.1	alt	18.7
Slope	26	southness	17.5
alt	9.4	westness	16
NDVI_45class	4.6	slope	8
s_type	4	s_type	6.5
corine	2	corine	5.3

Table: 3-10: Environmental Variables and Vegetation percentage Contribution

Variable	% Contribution Andalucía	Variable	% Contribution Spain
southness	13.4	NDVI_104class	13.5
slope	11.9	temp_s	12
westness	11.3	pet	10
pet	8.8	dirr_ann	6.8
temp_s	8.1	aridity	6.3
c_cover	6.5	Alt	6
m_temp_w	5.7	slope	6
dirr_ann	5.2	southness	6
prep_w	5	c_cover	5.8
aridity	4.9	westness	5.7
prep_d	4.6	prep_d	5.1
NDVI_45class	3.8	m_temp_d	5
m_temp_d	3.2	s_type	4
corine	3	prep_w	2.9
s_type	2.4	m_temp_w	2.5
Alt	2.2	corine	2.4

These results represent the average percentage contributions based on the 10 random partitions (n=10). The percentage contributions are heuristic estimates of relative contributions of the predictor variables to the Maxent Model. Variable contributions should be interpreted with caution due to the likelihood of correlations between predictor variables

4. Discussion

4.1. Potential geographic distribution of *Timon lepida* in Andalucia and Spain

This study provided insights into the spatial distribution of *Timon lepida* at the two spatial scales: (1) landscape (2) regional, it identified the environmental variables that were most influential in determining this distribution and it further made a comparative analysis of the models generated with and without vegetation indices.

The results revealed that *Timon lepida* is widely distributed throughout Andalucía and the whole of Spain and persists in a wide range of habitat types and environmental conditions that are not easily defined by the data, independent variables or Maxent model design. An overlay analysis of the distribution datasets with explanatory variables revealed that *Timon lepida* inhibits a high ecological tolerance as could be seen from the distribution maps (Figure 3-1, 3-5). Low suitability areas in the south eastern Andalucía e.g. Almeria correspond to some of the warmest and most arid areas in Europe with temperatures as high as 34^{0}C, while the presence of marshy lowlands created by the river Guadalquivir and its tributaries together with high temperatures could be responsible for the spatial constriction in the central part in areas like Seville and Cordoba. The northern part of Spain is characterised by high altitude areas traversing Basque and Pyrenees and lies on a high altitude zone with cooler temperatures. These areas are quite humid with annual precipitation records of over 1000 hence the resultant low predicted suitability (Bario 2006). The south-eastern parts of Andalucía correspond to Sierra Nevada Mountains which are quite cold for the species survival.

The outcome of this study is consistent with the findings made by (Busack and Visnaw 1989) in their study about the habitat preferences of *Timon lepida* in Cadiz, Spain that the species is found in artificial shelter sites as well as natural sites with equal frequency and does not appear to respond to any clear environmental gradients. These results corroborate other published work (Busack and Visnaw 1989; Mateo and Castanet 1994) who found out that generalist species like *Timon lepida* can exhibit good adaptability to diverse microclimates including altered habitats. Therefore it was not surprising to see that models tested in this study had difficulty in defining *Timon lepida's* ecological niche or identifying specific landscape characteristics that may inhibit the species distribution.

This study confirmed the results of other modelling efforts, e.g. (Dettmers *et al.* 2002; Hepinstall *et al.* 2002; Alley *et al.* 2004; Seoane *et al.* 2005; Crall *et al.* 2006) who had similar results when modelling widely distributed species and concluded that generalist species are unlikely to be modelled with great accuracy. Hernandez *et al.*(2006) discovered that ecological characteristics affect model accuracy where species that are widespread in both geographic and environmental space are difficult to model than species with specialized habitat, as is the case with *Timon lepida*. This was presumably because it was difficult to discriminate suitable from unsuitable habitat for the species, and not as an artefact of the methodology or the set of predictor variables used (Franklin *et al.* 2009).

Visual Interpretation

Most strikingly, Maxent models in general, showed a much wider distribution than is presently known which might indicate potential suitable habitat for the species into newer areas such as the Balearic Islands, which it has not been able to colonise due to geographic barriers or this could simply indicate model over-prediction. This scenario is not uncommon in species distribution modelling and matches closely the hypothesis by (Fielding and Bell 1997; Evangelista *et al.* 2008) that two possible errors may occur in species distribution models: false-negatives (under prediction) and false positives (over-prediction). Problems of over-prediction in species distribution modelling, according to Graham and Hijmans (2006) can be improved by restricting the potential ranges using expert drawn range maps and bio-geographical regions information. In principle, if environmental variables included in the model are appropriate, over-prediction can indicate that other factors not included in the model may be determining the geographical distributions such as historical connectivity, geographical barriers or competitive exclusion by closely related or ecologically similar taxa (Parra *et al.* 2004). This scenario, according to Mateo and Castanet (1994) and Hodar *et al.* (1996) may be understood by incorporating into modelling processes other ecological factors which may be of over-riding importance in driving the distribution such as availability and daily activities of potential prey.

Habitat suitability does not guarantee occupancy because other factors may influence the distribution other than environmental variables e.g. dispersal success, geographic barriers etc (Peterson 2003). In spite of the above mentioned possibilities of over-prediction, it is believed that these models generally represent reasonable approximations of the current potential geographical distribution. We did not observe any clamping effects which indicate projection of future climate outside of the current climate space did not occur. Visually speaking, the derived potential distribution maps from all the models trained in Andalucía appeared less smooth due to topography varying over shorter distances and providing more contrasted values between adjacent cells. This concurs with the results obtained by (Guisan and Hofer 2003) when predicting the distribution of reptiles in relation to climate and topography.

These results concur with other published work. For instance Said *et al.* (2003) found out that vegetation indices were less accurate predictors when modelling mammalian species richness in East Africa and Parra *et al.* (2004) yielded similar results when modelling the distribution of birds in the Ecuadorian Andes. The high accuracy of the predictions by models based on environmental variables could be due to the fact that the environmental variables used were of primary importance to the species.

4.1.1. ROC/AUC Comparisons

Threshold-independent ROC analysis showed significantly better than random model performance for all partitions in all the models. All of the generated models are above random (>0.5). The models had a fair to poor accuracy assessment by the AUC measurement (Manel *et al.* 2001). For instance, those models with an AUC of <0.7 are thought to be poor whereas those with AUC>0.7 are deemed potentially useful (Phillips *et al.* 2006). Models built with environmental variables had the highest AUC score (0.753 and 0.668) for the landscape and regional modelling scales respectively. It is clear that all model scenarios trained in Andalucía were more accurate than those trained in Spain especially when looking at both the training and test AUC.

The test AUC values for all the models were not very high and consistently declined from the landscape to the regional scale. If a species is widespread and the probability of presence increases with predictor values, an accurate model will have low AUC values which will only denote the true generalist nature of the species distributions (Lobo *et al.* 2008). This realization should help guide the appropriate interpretation of predictive habitat maps for generalist species such as *Timon lepida*.

4.1.2. Environmental Variables vs. Environmental Variables and Remotely Sensed Vegetation Indices

Models built using environmental variables and vegetation had a relatively lower predictive ability despite the large number of predictor variables included in this set. We expected this combination to provide the best estimates of the distribution for the species range. Frankly speaking, classified hyper temporal NDVI and CLC did not add much improvement in the modelling process besides increasing the regularized training gain. The low predictive ability could be due to correlation between the input variables such that any addition of other variables did not improve model performance due to redundancy. As documented by (Buermann *et al.* 2008), adding more variables with no significant new information does not lead to improved model performance. If datasets are not correlated, predictions should be different and possibly complementary, resulting in an overall improvement of model performance (Parra *et al.* 2004). This could be further linked to some of the limitations of using remotely sensed data as described in (section 4.6) possibly as a shortcoming of remote sensing not capturing smaller scale vegetation patterns. Alternatively, it can be argued that relative to climate, vegetation was of secondary importance to the target species.

Several researchers have credited the strong performance of the model with the identification of key variables associated with habitat suitability for the analysis (Parra *et al.* 2004). On the other hand, poor model performance is often blamed on lack of significant variables to predict suitable habitats e.g. choosing variables that are not biologically meaningful for the study scales (Parra *et al.* (2004). First, we suggest that poor predictive performance of the model may not always be a shortcoming in methodology or the predictor variables used. Secondly, it is worth noting that high AUC values do not imply high suitability accuracy. An alternative interpretation might consider poor model performance as indicative of species traits e.g. generalist species. As such, the predictive accuracy of the models tested in this study lies in their low AUC values!

4.2. Drivers of the potential distribution at a Landscape level

According to our results obtained for Andalucía, topography, (aspect and slope), has proved to be the top most predictor for the potential distribution of *Timon lepida* for all the model scenarios trained at this spatial scale. As stated by (Austin 1980, 2002), and supported by (Wang *et al.* 2009), topography is an important variable affecting the spatial variability of micro climate, soil properties and species distributions. This is not surprising given the physiological needs of the target species. South and west facing slopes receive more sunshine necessary for basking of the lizards, a statement supported by (Rouag *et al.* 2006) who found out that the daily activities of lizards are intense during the early morning hours. This result also proved reasonable since maximum temperatures are higher on south and west facing slopes associated with their greater insolation. This is especially important when considering the findings of Chiaraviglio and Bertona (2007) who reported that most reptiles prefer sunny and warm habitats to increase their body temperature and decrease it with shaded forest in order to facilitate thermoregulation.

Most lizards tend to be found at relatively lower warm slope gradients than steep cooler inclines (Huey and Hertz 1984). Lizards are extreme solar ectotherms with complex physiological and behavioural mechanisms for maintaining their body temperatures (Heatwole, 1982). The linear decline of temperature along with elevation explains the decline of suitability in the northern parts of Spain. In short, our results suggest that at a landscape level of Andalucía, the distributional range of *Timon lepida* is associated with topographical variables and any conservation measures may be implemented within these distributional limits (Mackey and Lindenmayer 2001).

4.3. Drivers of the potential distribution at a Regional Level

At a regional scale of Spain, the distribution of *Timon lepida* seemed to be driven by various predictor variables of which the most important were NDVI, PET, solar insolation and altitude. A great deal of accuracy was achieved in the predictions generated at the two spatial scale since topography was more relevant at a landscape level while in contrast climate and vegetation were more limiting at a regional scale. Our results corroborate previous studies showing the importance of these variables at the appropriate spatial scale (Peterson 2001; Guisan and Hofer 2003; Hulbert and Haskell 2003). Vegetation indices i.e. NDVI is a variable that integrates the effects of temperature, precipitation, soils and land cover thereby making primary productivity the most informative predictor at this spatial scale.

Nonetheless, from a predictive standpoint vegetation indices did not significantly improve the predictive performance of the models. Several hypotheses can be attributed to explain this outcome: (1) Vegetation as an indirect variable has little direct physiological relevance for the target species making it difficult to establish the causal relationship between this variable and species distributions. (2) Vegetation is usually driven by climatic gradients, but in each cell there is an underlying heterogeneity of land cover information that is lost given the coarse resolution of the vegetation indices (1km^2). This heterogeneity is a crucial factor in predicting the distribution of species (Crowling and Lombard 2002; Benton *et al.* 2003).

4.4. Difference in Variable importance from Landscape to Regional Scale

This study discovered that topographic variables were significant at the landscape level, while climatic variables and vegetation indices were important at the regional scale thus confirming the established theory. These findings were supported by Pathey (2003) who indicated that modelling at a large extent reveals environmental variables that best characterise the overall species range whereas modelling at a smaller extent disclose those environmental variables that best characterise habitat at home-range or population level. While climatic parameters can be relevant at all spatial scales (Guisan and Hofer 2003), they dominate more at regional, continental and global scales. NDVI and climate data are the two most commonly used predictors of species distributions at a regional scale because of their presumed importance as limiting factors at this spatial scale (Peterson 2001; Guisan and Zimmermann 2003).

4.5. Limitations of using Remotely sensed NDVI for species distribution modelling

The low performance of Maxent models when vegetation indices were included warrants some discussion. Remotely sensed NDVI has been used with immense success for vegetation mapping since it provides measurements directly related to forest structure and the overall health of the ecosystem that can collectively improve our understanding of suitable habitats for the species (Ji and Peters 2007). However, it presents some limitations that impact on the utility of its use for global vegetation studies and species distribution modelling.

Wang *et al.* (2005)and Ji and Peters (2007) stated that vegetation indices (VI's) are greatly affected by external factors such as atmospheric interference, solar zenith angle and viewing angle. VI's are highly prone to atmospheric noise caused by scattering, cloud and aerosol contamination (smoke and biomass burning). For instance, the atmosphere reduces the contrast between the red and NIR leading to decreased VI signal which ultimately culminates in the underestimation of the ground surface vegetation. Uncertainties caused by sun angles and variations in sensor viewing geometry often reduce the computed VI values resulting in degraded accuracy. We however controlled for cloud interference by including de-clouded images (discussed in section 2.4.9).

In semi-arid regions like Spain where vegetation is thin and widely spaced, VI's are affected by background interference caused by bare soil reflecting off the ground (Huete *et al.* 2001; Houborg *et al.* 2007). Bare soil reflectance resulting from soil properties such as iron amount, soil colour brightness and organic matter content may cause large variations in NDVI and further lead to a reduction in accuracy when using these indices for prediction (Huete and Tucker 1991) e.g. NDVI values are shown to decrease with the increasing radiance reflected from the soil especially due to rough terrain (Santos and Negri 1997).

Another limitation of NDVI is the 1km grain size. It is likely that sub-pixel landscape features exist which influence the distribution of the species, but could not be captured using this coarse resolution imagery (Pettorelli *et al.* 2005). This makes it even more complicated given the mobile nature of species such as *Timon lepida* since NDVI does not directly quantify species but species habitats (Leyequien *et al.* 2007). There is no consensus as to which scale results in the greatest accuracy (Gillespie *et al.* 2008).

All the aforementioned reasons could possibly explain why vegetation indices predicted the distribution of *Timon lepida* less accurately than other environmental factors. Some studies yielded similar results when using NDVI for predicting mammal species (Oindo 2001; Oindo and Skidmore 2002; Oindo and Skidmore 2002; Said *et al.* 2003; Parra *et al.* 2004) compared to those that used climatic parameters (Justice *et al.* 1991; Egbert *et al.* 2000).

4.6. Maximum Entropy Modelling Approach: Usefulness and Future Advancements

Maximum Entropy modelling approach has proved to be a simple, robust and effective method for predicting the distribution and ecological requirements of *Timon lepida,* enabling us to identify the variables that best predict areas of apparent suitability. While its usefulness has been widely documented and has been proven by this study, there are a few improvements that may strengthen its efficacy for species distribution modelling such as: 1. developing protocol for selection of appropriate threshold values. A more consistent or more clearly defined selection approach when the need is to convert continuous probability into binary data would provide greater consistency in model output and is worthy of further investigation, 2. developing methodology for selecting the best approximation method. One of the biggest obstacles to be overcome in Maxent pertains to model evaluation and subsequent model selection. Many models that are often included in full models have little influence on the species distribution patterns, therefore their elimination may increase the accuracy of the model and avoid overfittting. Consequently, the development of information-theoretic (AIC) (Marmion *et al.* 2009) may provide the greatest opportunity for model testing.

5. Conclusions and Recommendations

In this study, we systematically demonstrated the usefulness of a novel approach, Maximum Entropy, for modelling the potential geographic distribution of *Timon lepida*. In order to test the hypothesis that remotely sensed vegetation indices are better predictors of species distribution of species compared to environmental variables ,(climatic, topography, etc) 3 main objectives were set:

The first objective aimed at mapping the distribution of *Timon lepida* and the second objective was to determine which set of independent explanatory variables would explain well this distribution. The findings of this study revealed that *Timon lepida* is widely distributed in Spain, and this distribution may be explained by topography at a landscape level and climate and vegetation at a regional scale. Objective 3 sought to investigate whether vegetation indices would outperform environmental variables in terms of predictive accuracy. Overall, for Andalucía and Spain, environmental variables proved to be superior predictors compared to remotely sensed vegetation indices. This was consistently demonstrated by high AUC score, high kappa value, *p*-values from the test omission rates, Wilcoxon signed rank and Friedman test results. In summary, the results did not support our hypothesis that vegetation indices would significantly predict better than environmental variables.

From the standpoint of conserving biodiversity, our results have significant implications. The development of geographic distribution maps for *Timon lepida* with the assistance of Maxent models will enhance the ability to develop conservation strategies for the species (Svenning and Condit 2008). Maxent also provided detailed information about the variables along with their importance in relation to the contribution to the model, which may have important implications for the conservation of the species. Finally, we provided insights into model performances and their relations to species traits and we hope our results will prompt an alternative interpretation of model accuracy based on the AUC by incorporating species characteristics and traits. Finally, practitioners should remember that models are simply an estimate of a species potential distribution. Species distribution modeling cannot replace fieldwork intended to collect more distributional data but can be a useful tool for data exploration to help identify potential knowledge gaps. Interpretation and use of our model results for conservation purposes should therefore be done cautiously.

5.1. Specific Conclusions

1. The accuracy of the species distribution predicted by Maxent models is significantly different than random. All the models achieved an AUC score of more than 0.5 and a *p*-value (<0.05) using Wilcoxon signed rank and Friedman tests.

2. In Andalucía and Spain, vegetation indices were less accurate predictors of the distribution of *Timon lepida* than environmental variables. The relatively low predictive accuracy of vegetation indices could be attributed to various factors such as bare ground reflectance or coarse grain size which may have degraded the quality of vegetation indices and hampered its accuracy.

3. The study results indicated that variables changed in importance at different spatial scales: topographic variables were more important at a landscape level while vegetation and climatic variables were more important at a regional scale. These results are due to the performance of variables at different spatial resolutions and this supports established ecological theory.

4. High AUC score does not always imply high predictive accuracy of the model. The ecological characteristics and traits of the species being modelled must guide the appropriate interpretation of predictive habitat maps for generalist species such as *Timon lepida*.

5.2. Recommendations

1. Future research may consider other variables that may be driving the distribution such as availability and daily activities of prey.

2. It is important to note that while our results illustrate the current extent of suitable habitat throughout Spain, some of the variables may not accurately depict current conditions (e.g., the bioclimatic variables represent mean values for the time period 1950 - 2000). If reliable datasets of current conditions can be identified, these data can be incorporated into Maxent to improve results.

3. Kaliontzopoulou *et al.* (2008) proved that variables with higher spatial resolutions produce better predictive models. The conclusions would be more precise if the adequate precise environmental resolution data are included in the future application. To re-test this hypothesis, the current low resolution data e.g., NDVI used in these models could be replaced with high resolution data that capture fine ecological details.

6. References

Agency, E. E. (2009). "Halting the loss of biodiversity: managing Europe's ecosystems and wild species." www.eea.eu.int: Accessed 18-08-09.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, J. M. B. (1998). "Crop evapotranspiration: Guidelines for computing crop requirements. Irrigation and Drainage Paper No, 56, FAO, Rome,Italy.".

Alley, N., Stolhgren, T. J., Evangelista, P. H. and Guenther, D. (2004). "Iterative model development for natural resource managers: a case study in Utah's grand Staircase-Enscalate national Monument." Geographic Information Sciences **10**: 1-9.

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

Anderson, R. P. (2003). "Real vs. artefactual absences in species distributions for Oryzomys albigularies (Rodentia: Muridae) in Venezuela." J. Biogeogr **30**: 591-605.

Araujo, M. B. and Guisan, A. (2006). "Five or so challenges for species distribution modelling." Journal of Biogeography **33**: 1677-1688.

Araujo, M. B. and New, M. (2007). "Ensemble forecasting of species distributions." Trends in Ecology and Evolution **22**: 42-47.

Araujo, M. B. and Williams, P. H. (2000). "Selecting areas for species persisitence using occurance data." Biological Conservation **96**(331-345).

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

Austin, M. P. (2002). "Spatial predictions of species distributions: an interface between ecological theory and statistical modeling." Ecological Modelling **157**: 101-118.

Badgley, C. and Fox, D. L. (2000). "Ecological biogeography of North American mammals:species density and ecological structure in relation to environmental gradients." Journal of Biogeography **27**: 1437-1467.

Baillie, J. E. M., Hilton-Taylor, C. and Stuart, S. N. (2004). "2004 IUCN Red List of Threatened Species: A Global Species Assessment." IUCN, Gland, Switzerland and Cambridge, UK. ISBN:2831708265.

Bario, J. F. (2006). "Si, Spain." http://www.sispain.org/english/index.html: Accessed 16-August 2009.

Barnard, P. and Thuiller, W. (2008). "Global change and biodiversity: future challenges." Biol. Lett **4**: 553-555.

Barry, S. and Elith, J. (2006). "Error and Uncertainty in habitat models." Journal of Applied Ecology **43**: 413-423.

Bascompte, J. and Ricard, V. (2009). "Effects of Habitat Destruction in a Prey Predator Metapopulation." Journal of Animal Ecology **78**: 182-190.

Benton, T. G., Vickery, J. A. and Wilson, J. D. (2003). "Farmland biodiversity: is habitat heterogeneity the key?" TREE **18**: 279-289.

Box, E., O., Holben, B. N. and Kalb, V. (1989). "Accuracy of AVHRR Vegetation Index as a predictor of biomass, primary productivity and net CO2 flux." Vegetatio **80**: 71-80.

Buermann, W., Saachi, S., Smith, T. B., Zutta, B. R., Chaves, J. A., Mila, B. and Graham, C. H. (2008). "Predicting species distributions across the Amazomian and Andes regions using remote sensing data." Journal of Biogeography **35**: 1160-1176.

Busack, S. D. and Visnaw, J. A. (1989). "Observation on the natural history of *larceta lepida* in Cadiz Province, Spain." Amphabia-Reptilia **10**(3): 201-213.

Cantor, S. B., Sun, C. C., Tortolero-Luna, G., Richards-Kortum, R. and Follen, M. (1999). "A comparison of C/B ratios from studies using Receiver Operating Characteristic Curve Analysis." Journal of Clinical Epidemiology **52**(9): 885-892.

Cardona, J. N. and Loyola, R. D. (2008). "Applying niche-based models to predict endangered-hylid potetial dsitributions: are neotropical protected areas effective enough?" TropicalConversation Science **1**(4): 417-445.

Chang, C., Lee, P., Bai, M. and Lin, T. (2006). "Identifying the Scale Thresholds for Field-Data Extrapolation via Spatial Analysis of Landscape Gradiants." Ecosystems **9**: 200-214.

Chiaraviglio, M. and Bertona, M. (2007). "Reproduction and thermoregulation as main factors influencing habitat choice in the Argentine Boa constrictor." Biology of Boas and pythons, p478-488. Henderson, R.W, Powel, R. Eds, Utah, Eale Mountain Publishing.

Cohen, J. (1960). "A coefficient of agreement for nominal scales." Education Psychology Measurement **20**: 37-46.

Corsi, F., Dupre, E. and Boitani, L. (1999). " A large-scale model of wolf distribution in Italy for Conservation Planning." Conservation Biology **7**: 150-159.

Corsi, F., Skidmore, A. K. and de Leeuw, A. J. (2000). "Modelling Species Distributions with GIS. In L. Boitani & T.K. Fuller (Eds)." Research Techniques in animal ecology: 389-434.

Crall, A. W., Newman, G. J., Stohlgren, T. J., Jarvenich, C. S., Evangelista, P. H. and Guenther, D. (2006). "Evaluating dominance as a component of non-native species invasions." Diversity and Distributions **12**: 195-204.

Crowling, R. M. and Lombard, A. T. (2002). "Heteregoneity, speciation/extinction history and climate: explaning regional plant diversity patterns in the Cape Floristic Region." Diversity and Distributions **8**: 163-179.

Database, E. (2004). "European Soil Bureau Network and the European Commission." EUR 199945 EN.

de Bie, C. A., Khan Mobushir, R., Toxopeus, A. G., Venus, V. and Skidmore, A. K. (2006). "Hypertemporal Image Analysis and Crop Mappin and change detection." ITC, Enschede, the Netherlands.

Delong, E. R., Delong, D. M. and Clarke-Pearson, D. L. (1988). "Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach." Biometrics **44**: 837-845.

Dettmers, R., Buehler, D. and Barlett, J. (2002). "A test and comparison of wildlife-habitat modeling techniques for predicting bird occurrence at a regional scale." Predicting species occurrence . Issues of scale and Accuracy (ed by Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. Island Press, Washington, D.C.: 607-615.

Diaz, J. A. and Carrascal, L., M. (1991). "Regional Distribution of a Mediterranean lizard: influence of habitat cues and prey abundance." Journal of Biogeography **18**: 291-297.

Egbert, S. L., Martinez-Meyer, E., Orgetta-Huerta, M. and Peterson, A. T. (2000). "Use of Datasets Derived from Time-series AVHRR Imagery as surrogates for Landcover Maps in predicting species distributions." International Journal of Remote Sensing **10**(2337-2339).

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R., J., Huettmann, F., Leathwick, J., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K., Schacheti-Pereira, R., Schapire, R. E., Soberon, J., Williams, S., Wisz, M. S. and Zimmermann, N. E. (2006). "Novel methods improve predictions of species distributions from occurence data." Ecography **29**: 129-151.

Engler, R., Guisan, A. and Rechsteiner, L. (2004). "An improved approach for predicting the distribution of rare and endangered species from occurance and pseudo-absence data." Journal of Applied Ecology **41**: 263-274.

Evangelista, P. H., Kumar, S., Stohlgren, T. J., Jarvenich, C. S., Crall, A. W., Norman III, J. B. and Barnett, B. T. (2008). "Modelling invasion for a habitat generalist and a specialist plant species." Diversity and Distributions **14**: 808-817.

Fahrig, L. (2003). "Effects of habitat fragmentation on Biodiversity." Ecography **34**: 487-515.

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

Fox, J. (1997). "Applied regression analysis, linear models and related methods. California:Sage Publications."

Franklin, J., Wejnert, K. E., Hathaway, S. A., Carlton, J. R. and Fisher, R. N. (2009). "Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California." Divers. Distrib. **15**: 167-177.

Freedman, A. H., Buermann, W., Lebreton, M., Chirio, L. and Smith, T. B. (2008). "Modeling the Effects of Anthropogenic Habitat Change on Savanna Snake Invasions into African Rainforest." Conservation Biology **23**(1): 81-92.

Giannakopoulos, C., Bindi, M., Moriondo, M., Lesager, P. and Tin, T. (2005). "Climate change impacts in the Mediterranean resultin from a 2^{o}C global temperature rise." Gland, Switzerland, WWF.

Gillespie, T. W., Foody, G. M., Rocchini, D., Giorgi, A. P. and Saachi, S. (2008). "Measuring and Modelling Biodiversity from Space." Progress in Physical Geography **32**(2): 203-221.

Goode, M., J., Swanna, D. E., Shwalbe, C., R. and Mannan, W. (1998). "The effects of micro habitat destruction on reptile abundance."

Graham, C. H. and Hijmans, R., J. (2006). "A comparison of methods for mapping species ranges and species richness." Global Ecology and Biogeography.

Guisan, A., Edwards, T. C. and Hastie, t. (2002). "Generalized Linear and generalized additive models in studies of species distributions: setting the scene." Ecological Modelling **157**: 89-100.

Guisan, A. and Hofer, U. (2003). "Predicting reptile distribution at the mesoscale :relation to cimate and topography." Biogeography **30**: 1233-1243.

Guisan, A. and Hofer, U. (2003). "Predicting reptile distributions at the mesoscale:relation to climate and topography." Biogeography **30**: 1233-1243.

Guisan, A. and Zimmermann, N. E. (2003). "Predictive Habitat Distribution models in Ecology." Ecological Modelling **135**: 147-186.

Hamilton, A. J. (2005). "Species Diversity or biodiversity?" Journal of Environmental Management **75**(89-92).

Hargreaves, G. H. and Allen, R. G. (2003). "History and Evaluation of Hargreaves Evapotranspiration Equation." J. Irrig. Drain. Eng. ASCE **129**(1): 53-63.

Hargreaves, G. L., Hargreaves, G. H. and Riley, J. P. (1985). "Irrigation Water Requirements for Senegal River Basin." J. Irrig. Drain. Eng. ASCE **111**(3): 265-275.

Hepinstall, J., Krohn, W. and Sadler, S. (2002). "Effects of niche width on the performance and agreement of avian habitat models." Predicting species occurrence . Issues of scale and Accuracy (ed by Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. Island Press, Washington, D.C.: 593-606.

Hernandez, P. A., Franke, E. I., E.S.K, H., Pacheco, E. V., Paniagua, E. L., Quitana, E. H. L., Soto, E. A., Swenson, E. J. J., Tovar, E. C., Valqui, E. T. H., Vargus, E. J. and Young, E. B. E. (2008). "Prediting species distributions in poorly-studied landscapes." Biodiversity Conservation **17**: 1353-1366.

Hernandez, P. A., Graham, C. H., Master, L. L. and D.L., A. (2006). "The effect of sample size and species characteristics on perfomance of different species distribution modelling methods." Ecography **29**: 773-785.

Hijmans, R., J., Cameron, S., E., Parra, J., L., Jones, P., G. and Jarvis, A. (2005). "Very high resolution interpolated climate surfaces for global land areas." International journal of Climatology **25**: 1965-1978.

Hirzel, A. H., Helfer, V. and Metral, F. (2001). "Assesing habitat-suitability models with a virtual species " Ecological Modelling **145**: 111-121.

Hodar, J. A., Campos, F. and Rosales, B. A. (1996). "Trophic ecology of the Ocellated Lizard *Larceta lepida* in an arid zone of Southern Spain: relationships with availability and daily activity of prey." Journal of Arid Environments **33**: 95-107.

Houborg, R., Soegaard, H. and Boegh, E. (2007). "Combining vegetation index and model inversion methods for the extraction of key vegetation biophysical parameters using Terra Aqua MODIS reflectance data." Remote Sensing of Environment **106**: 39-58.

Huete, A. R., Didan, K., Miura, T., Rodriguez, E. P., Gao, X. and Ferreira, L. G. (2001). "Overview of the radiometric and biophysical performance of the MODIS vegetation indices." Remote Sensing of Environment **83**: 195-213.

Huete, A. R. and Tucker, C. J. (1991). "Investigation of soil influences in AVHRR vegetation imagery." International Journal of Remote Sensing **12**: 1223-1242.

Huey, R. B. and Hertz, P. E. (1984). "Effects of body size and slope on acceleration of a lizard (Stellio Stellio)." J.exp. Biol. **110**: 113-123.

Hulbert, A. H. and Haskell, J. P. (2003). "The effect of energy and seasonality in on avian species richness and community composition." Am.Nat **161**: 83-97.

IUCN (2001). "IUCN Red List Categories and Criteria: Version 3.1." IUCN, Gland, Switzerland and Cambridge, UK. (Accessed 18- August 2009) ISBN:2831706335.

IUCN (2004). "2004 IUCN Red List of Threatened Species." IUCN, Gland, Switzerland and Cambridge, UK. (Accessed 24- August 2009).

Jellinek, S., Driscoll, D. A. and Kirkpatrick, J. B. (2004). "Environment and vegetation variables have a greater influence than habitat fragmentation in structuring lizard communities in remnant urban bushland." Austral Ecology **29**: 294-304.

Ji, L. and Peters, A. J. (2007). "Performance evaluation of spectral vegetation indices using a statistical sensitivity function." Remote Sensing of the Environment **106**: 59-65.

Justice, C. O., Eck, T. F., Tanre, D. and Holben, B. N. (1991). "The effect of water vapour on the NDVI derived for the Sahelian region from NOAA AVHRR data." International Journal of Remote Sensing **12**: 1165-1967.

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

Lehmann, A., Overton, J. M. and Austin, M. P. (2002). "Regression models for spatial prediction: their role for biodiversity and conservation." Biodiversity and Conservation **11**: 2085-2092.

Lennon, J. J., Greenwood, J. J. B. and Turner, J. R. G. (2000). "Bird Diversity and environmental gradients in Britian: A test of the species-energy hypothesis." Journal of Animal Ecology **69**: 581-598.

Leyequien, E., Verrelst, J., Slot, M., Schaepman-Strub, G., Heitkonig, I., M.A and Skidmore, A. K. (2007). "Applying remote sensing to terrestrial animal distribution and diversity." International Journal of Applied Earth Observation and Geoinfomation **9**: 1-20.
Liu, C., Berry, P. M., P., D. and Pearson, R. G. (2005). "Selecting thresholds of occurence in the prediction of species distributions." Ecography **28**: 385-393.
Lobo, J. M., Jimenez-Valverde, A. and Real, R. (2008). "AUC: a misleading measure of the performance of predictive distribution models." Global Ecology and Biogeography **17**: 145-151.
Mackey, B. G. and Lindenmayer, D. B. (2001). "Towards a hierachical framework for modelling the spatial distrbutions of animals." J. Biogeogr **28**: 1147-1166.
Manel, S., Williams, H. C. and Ormerod, S. J. (2001). "Evaluating presence-absence models in ecology: the need to account for prevalence." Journal of Applied Ecology **38**: 921-931.
Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K. and Thuiller, W. (2009). "Evaluation of consensus methods in predictive species distribution modelling." Diversity and Distributions **15**: 59-69.
Martinez-Freiria, F., Sillero, N., Lizana, M. and Brito, J. C. (2008). "GIS-based niche models identify environmental correlates sustaining a contact zone between three species of Europen vipers." Divers. Distrib. **14**: 452-461.
Mateo, J. A. and Castanet, J. (1994). "Reproductive strategies of three Spanish populations of the Ocellated Lizard Larceta lepida (Sauria, Larcetidae)." Acta Oecologica **15**: 215-229.
Maurer, B. A. (1994). "Geographical population analysis: Tools for the analysis of of biodiversity. Methods in ecology." Oxford:Blackwell Scientific Publications.
McNees, S. K. (1987). "Consensus forecasts:tyranny of the majority?" New England Economic Review **Nov/Dec**: 15-21.
Moisen, G. G. and Frescino, T. S. (2002). "Comparing five modelling techniques for predicting forest characteristics." Ecological Modelling **157**(2-3): 209-225.
Mollenbeck, V., Hermann, G. and Fartman, T. (2009). "Does prescribed burning mean a threat to the rare satyrine butterfly *Hipparchia fagi?* Larval-habitat preferences give the answer." J. Insect Conserv. **13**: 77-87.
Monserud, R. A. and Leemans, R. (1992). "Comparing global vegetation maps with Kappa Statistic tool." Ecol. Model **62**: 275-293.
Montgomery, D. C. and Peck, E. A. (1982). "Introduction to linear regression analysis.New York: John Wiley and Sons."
Myers, N., Mittermeier, R. A., Mittermeier, C. G., Fonseca, D. and Kent, J. (2000). "Biodiversity hotspots for conservation areas." Nature **403**: 853-858.
Neumann, K., Herold, M., Hartley, A. and Schmullius, C. (2007). "Comparative assessment of CORINE2000 and GLC2000:Spatial analysis of landcover data for Europe." International Journal of Applied Earth Observation and Geoinformation **9**: 425-437.

Nicholson, K. L., Torrence, S., M., Ghioca Dana, M., Bhattacharjee, J., Andrei, A., E., Owen, J., Radke, N. J. A. and Perry, G. (2005). "The influence of temperature and humidity on activity patterns of the lizards *Anolis stratulus* and *Ameiva exsul* in the British Virgin Islands." Caribbean Journal of Science **41**(4): 870-873.

Oindo, B. (2001). "Spatial patterns of species diversity in Kenya." Unpublished PHD, ITC/ Wageningen.

Oindo, B. and Skidmore, A. K. (2002). "Interannual variability of NDVI and species richness in Kenya." International Journal of Remote Sensing **23**: 285-298.

Oindo, B. and Skidmore, A. K. (2002). "Interannual variability of NDVI and species richness in Kenya." International Journal of Remote Sensing **23**: 285-298.

Osborne, P. E., Alonso, J. C. and Bryant, R. G. (2001). "Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards." J.Appl. Ecol **38**: 458-471.

Owen, J. G. (1989). "On productivity as a predictor of rodent and canivore diversity." Ecology **69**: 1161-1165.

Parra, J., L., Graham, C. H. and Freile, J. F. (2004). "Evaluating alternative datasets for ecological niche models of birds in the Andes." Ecography **27**: 350-360.

Pathey, P. (2003). "Habitat and corrridor selection of an expanding red deer." Unpublished PHD, University of Luasanne, Luasanne, Switzerland.

Pearce, J. and Ferrier, S. (2000). "Evaluating the predictive performance of habitat models developed using logistic regression " Ecol. Model **133**: 225-245.

Pearson, R. G. and 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**: 361-71.

Peterson, A. T. (2001). "Predicting species geographic distributions based on ecological niche modeling." Condor **103**: 599-605.

Peterson, A. T. (2003). "Predicting the geography of species' invasions via ecological niche modeling." Quarterly Review of Biology **78**: 419-433.

Peterson, A. T., Papes, M. and Eaton, M. (2007). "Transferability and model evaluation in ecological modelling: A comparison of GARP and Maxent." Ecography **30**: 550-560.

Pettorelli, N., Vik, O. J., Mysterud, A., Gaillard, J., Tucker, C. J. and Stenseth, N. C. (2005). "Using the satellite-derived NDVI to assess ecological responses to environmental change." Trends in Ecology and Evolution **20**(9): 503-510.

Phillips, S. J., Anderson, R. P. and Schapire, R. E. (2006). "Maximum entropy modelling of species geographic distributions." Ecological Modelling **190**: 231-259.

Phillips, S. J. and Dudik, M. (2008). "Modelling of species distributions with Maxent: new extensions and a comprehensive evaluation." Ecography **31**: 161-175.

Phillips, S. J., Dudik, M., Elith, J., C.H., G., Lehmann, A., Lehmann, A., Leathwick, J. and S., F. (2009). "Sample Selection bias and prescence-only distribution modelling: implications for background and pseudo-abscence data." Ecological Applications **19**(1): 181-197.

Phillips, S. J., Dudik, M. and Schapire, R. E. (2004). "A maximum entropy approach to species distribution to species distribution modelling: Proceedings of the Twenty-first century International Conference on Machine Learning."

Phoebe, B. and Thuiller, W. (2008). "Global change and biodiversity: Future challenges." Biol. Lett **4**: 553-555.

Pleguezuelos, J. M., Marquez, R. and Lizana, M. e. (2004). "Atlas Y Libro Rojo De Los Anfibios Y Reptiles De Espana. Madrid:Ministerio De Medio Ambiente.".

Polasky, S. and Solow, A. R. (2001). "The value of information in reserve site selection." Biodiversity and Conservation **10**: 1051-1058.

Reuter, H. I., Nelson, A. and Jarvis, A. (2007). "An evaluation of void filling interpolation methods for SRTM data." International Journal of Geographic Information System **21**(9): 983-1008.

Roberts, D. W. (1986). "Ordination on the basis of fuzzy set theory." Vegetatio **66**: 123-31.

Rouag, R., Berrahma, I. and Luiselli, L. (2006). "Food habits and daily activity patterns of the North Africab ocellated lizard Timon pater from northeastern Algeria " Journal of Natural History **40**: 1369-1279.

Said, M. Y., Skidmore, A. K., de Leeuw, A. J., Aligula, H. M., Kumar, L. and Prins, J. H. T. (2003). "Analysis of relation between ungulate species richness in East Africa and climatic and remotely sensed productivity indices." Journal of African Ecology.

Santos, P. and Negri, A. J. (1997). "A comparison of the Normalized Difference Vegetation Index and rainfall for the Amazon and Northeastern Brazil." Journal of Applied Meteorology **36**: 958-965.

Schaller, M., Blum, J. D., Hamburg, S. P. and Vadeboncoerur, M. A. (2010). "Spatial variability of long-term chemical weathering rates in the white mountains, New Hampshire, USA." Geordema **154**: 294-301.

Sellers, P. J. (1989). "Canopy reflectance, photosynthesis and transpiration." International Journal of Remote Sensing **6**: 1335.

Seoane, J., Carrascal, L., M. and Palomino, D. (2005). "Species specific traits associated to prediction errors in bird habitat suitabiility modeling." Ecological Modelling **185**: 299-308.

Sillero, N. (2009). "Modelling a species in expansion at local scale:is *Hyla meridionalis* colonising new areas in Salamanca, Spain?" Acta Herpetologica **4**(1): 37-46.

Skidmore, A. K., Toxopeus, A. G., de Bie, K., Corsi, F., Venus, V., Omolo, D. P., Marquez, J. and Giminez, R., Real (2006). "Herpetological Species Mapping for the Meditteranean."

Stolkwell, D. a. P., D. (1999). "The GARP modelling system: problems and solutions to automated spatial prediction." International Journal of Geographic Information System **13**(2): 143-158.

Storms, D. M. and Etes, J. E. (1993). "A remote sensing agenda for mapping and monitoring biodiversity." International Journal of Remote Sensing **14**: 1839-1860.

Stuart, S. N., Chanson, J. S., Cox.N.A, Young, E. B. E., Rodrigues, A. S. L., Fischman, D. L. and Waller, R. W. (2004). "Status and trends of reptiles declines and extinctions worldiwide." Science **306**: 1783-1786.

Svenning, J. C. and Condit, R. (2008). "Biodiversity in a warmer world." Science **322**: 206-207.

Toxopeus, A. G., Bakker, X. and Kariuki, A. (1994). "Interactive spatial modellin ISM system for the management of the Amboseli biosphere reserve, Southeast Kenya." ITC Journal **4**: 392-400.

Trubacco, A., Zomer, R. J., Bossio, D. A., van Straaten, O. and Verchot, L. V. (2008). "Climate change Mitigation through Afforestation/ Reforestation: A global analysis of hydrologic impacts with four case studies " Agric.Ecosystems and Environment **126**: 81-97.

Vlachos, C. G. and Papageorgiou, N. K. (1994). "Diet, Breeding Success and the nest site selection of the short toed eagle (Circaetusg allucus) in Northeastern Greece." Raptor, Res **28**(1): 39-42.

Wang, Q., Aduku, S., Tenghunen, J. and Granier, A. (2005). "On the relationship of NDVI with leaf area index in a deciduous forest site." Remote Sensing of Environment **94**: 244-255.

Wang, Z., Ye, W., Cao, H., Huang, Z., Lian, J., Li, L., Wei, S. and Sun, I. (2009). "Species-topography association in a species-rich subtropical forest of china " Basic and Applied Ecology **10**: 648-655.

Yost, A. C., Peterson, S. L., Gregg, M. and Miller, R. (2008). "Predictive modeling and mapping of sage grouse(*Centrocercus urophasianus*) nesting habitat using Maximum Entropy and a long-term dataset from Southern Oregon." Ecological Informatics **3**: 375-386.

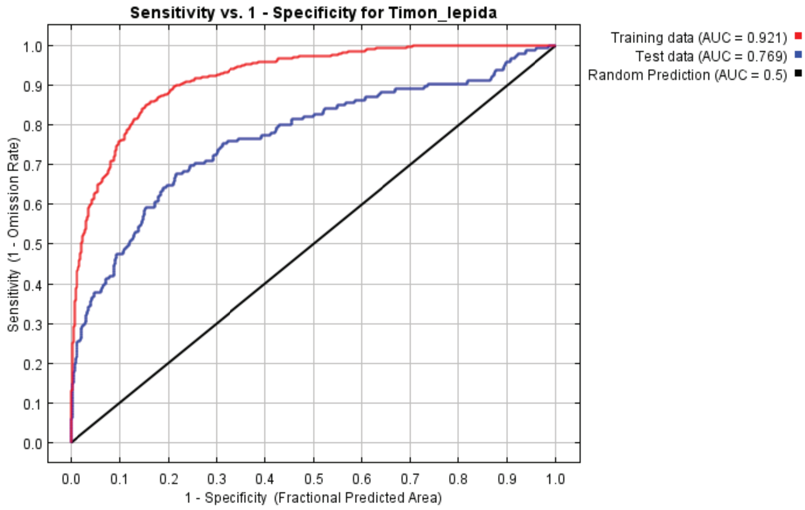
7. Appendices

7.1. Appendix A: Description of Predictor Variables

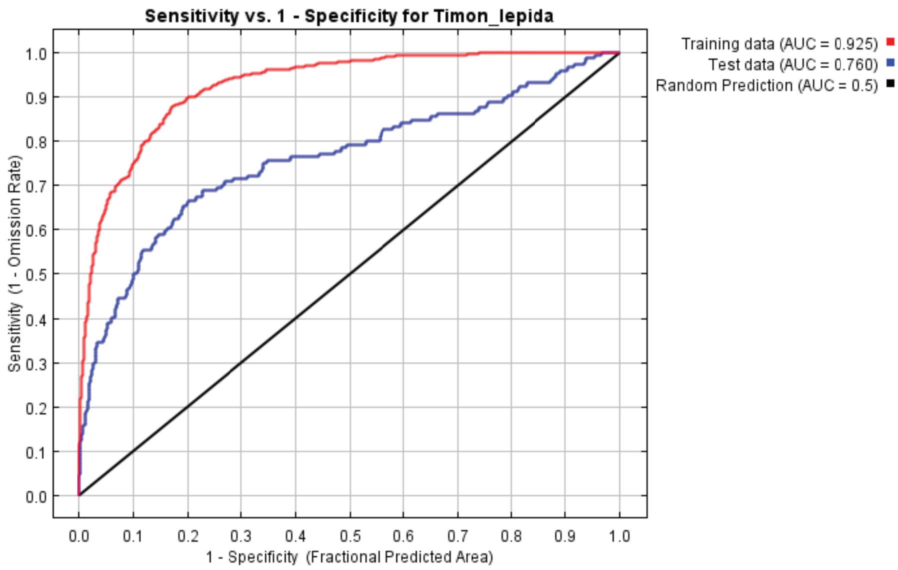
Abbreviation	Variable Full Name
Alt	Altitude
Aridity	Aridity
southness	southness
westness	westness
slope	slope
c_cover	Cloud Cover
dirr_ann	Annual Radiation
m_temp_d	Mean temperature of the driest quarter
m_temp_w	Mean Temperature of the wettest quarter
PET	Potential Evapotranspiration
prep_d	Precipitation of the driest quarter
prep_w	Precipitation of the wettest quarter
s_type	Soil type
Temp_s	Temperature seasonality
corine	Corine Land Cover
NDVI	Normalized Difference Vegetation Index

7.2. Appendix B: Best Model ROC Curves

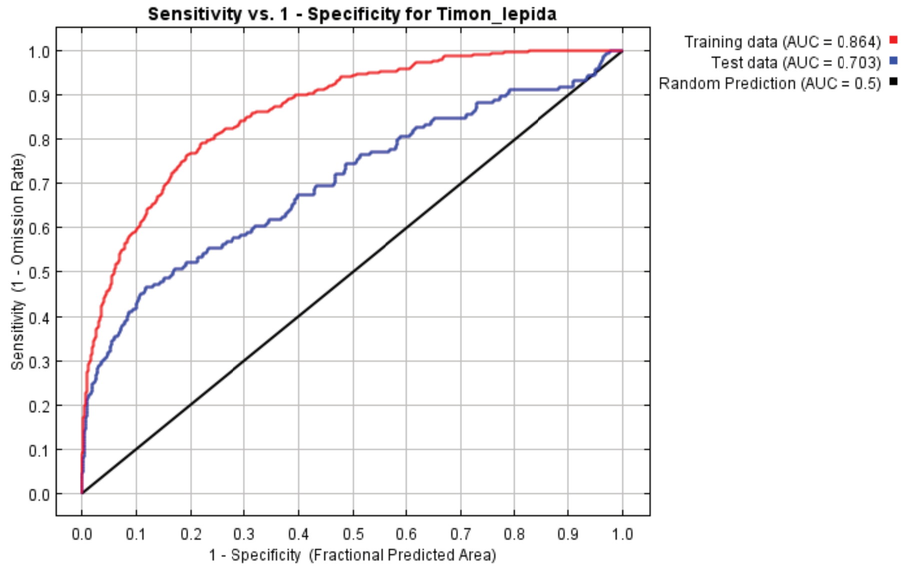
1. Andalucía



(a) Environmental Variables model

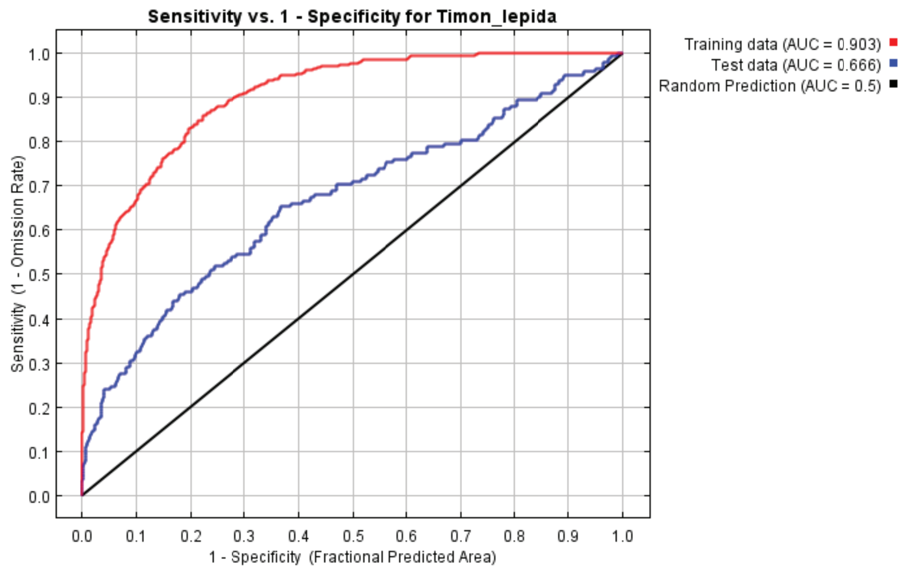


(b) Environmental Variables and vegetation model

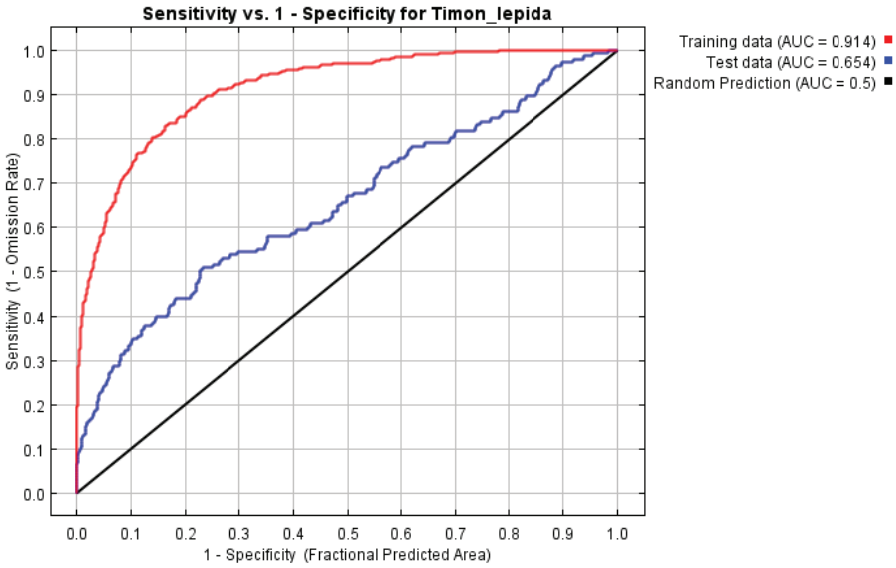


(c) Vegetation model

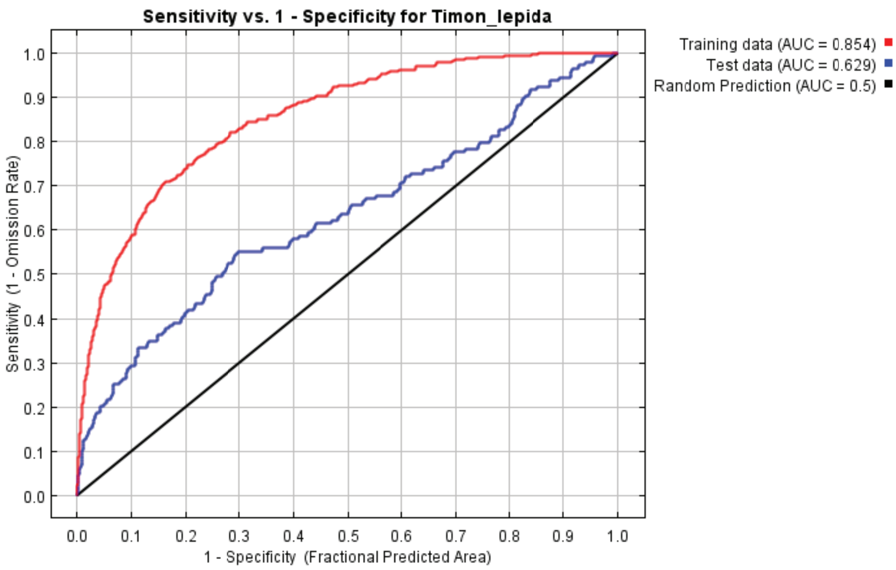
2. Spain



(a) Environmental Variables model

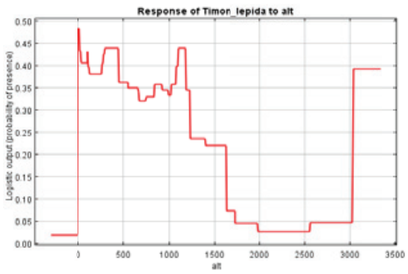


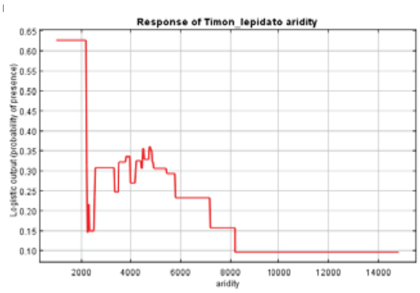
(b) Environmental variables and vegetation model

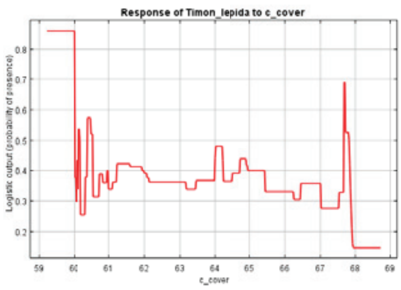


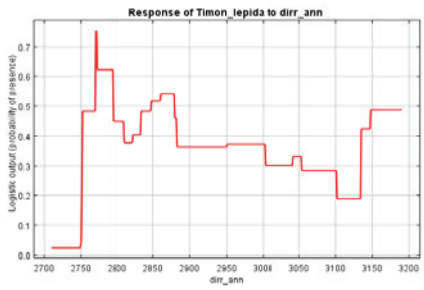
(c) Vegetation model

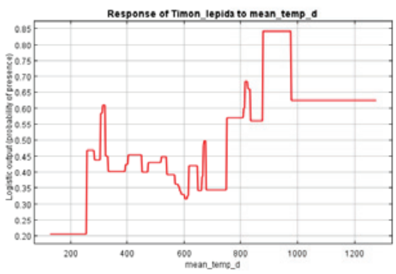
7.3. Appendix C: Response Curves of Predictor Variables

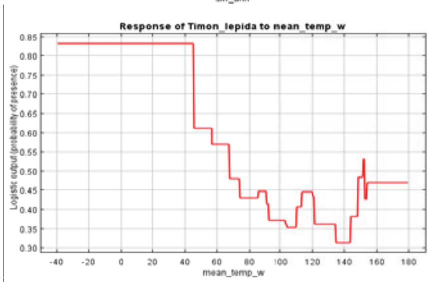


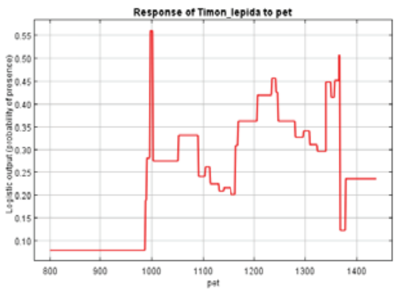


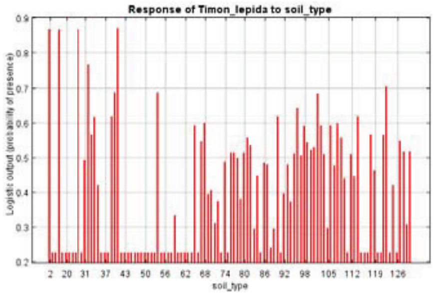


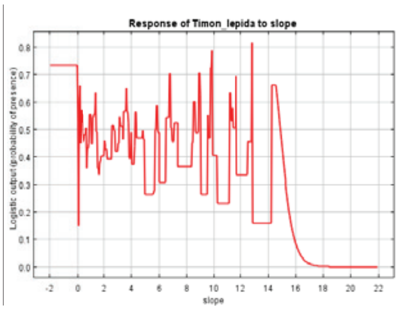


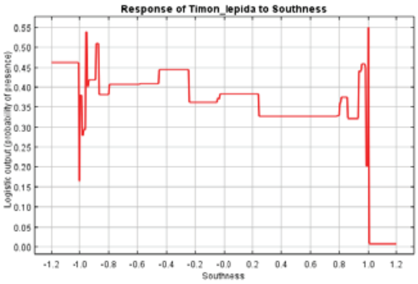


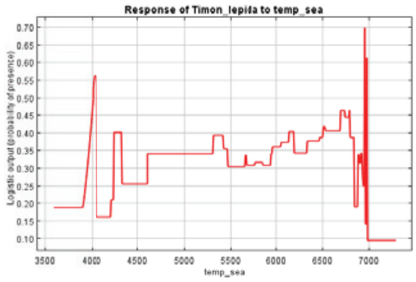


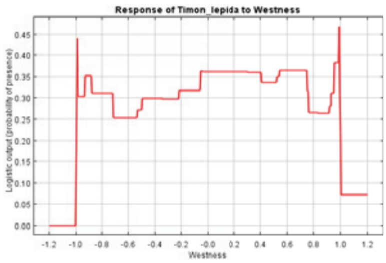


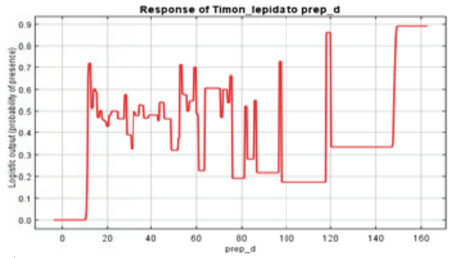

Response of Timon_lepida to soil_type
Logistic output (probability of presence)
soil_type

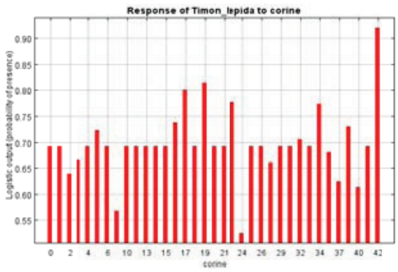

Response of Timon_lepida to slope
Logistic output (probability of presence)
slope

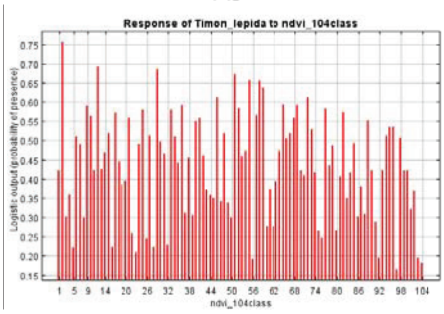

Response of Timon_lepida to Southness
Logistic output (probability of presence)
Southness


Response of Timon_lepida to temp_sea
Logistic output (probability of presence)
temp_sea


Response of Timon_lepida to Westness
Logistic output (probability of presence)
Westness


Response of Timon_lepida to prep_d
Logistic output (probability of presence)
prep_d


Response of Timon_lepida to corine
Logistic output (probability of presence)
corine


Response of Timon_lepida to ndvi_104class
Logistic output (probability of presence)
ndvi_104class