Influence of Human Activities on The Distribution of Amphibians and Reptiles in The Salamanca Province, Spain

Yan Sun

March, 2007

Course Title:	Geo-Information Science and Earth Observation for Environmental Modelling and Management
Level:	Master of Science (MSc)
Course Duration:	September 2005 - March 2007
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:	2005-07

Influence of Human Activities on the Distribution of Amphibians and Reptiles in the Salamanca Province, Spain

By

Yan Sun

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.

Thesis Assessment Board

Chairman: Prof. Dr. A.K. Skidmore (ITC)
External Examiner: Prof. Dr hab. K. Dąbrowska-Zielińska (University of Warsaw)
Primary Supervisor: Mr. Bas Wesselman (ITC)
Secondary Supervisor: Dr. A.G. (Bert) Toxopeus (ITC)

ITC International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands

Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

This thesis is an outcome of the BIOFRAG–ITC internal research project in collaboration among:
International Institute for Geo-Information Science and Earth Observation (ITC), The Netherlands
Dr. Neftalí Sillero University of Porto, Portugal

To my parents

献给我的父母

Special thanks to Dr. Neftalí Sillero
from CICGE (Investigation Centre on Geo-spatial Sciences),
University of Porto

Gracias especiales a Dr. Neftalí Sillero
de CICGE (Investigation Center on Geo-spatial Sciences),
Universidade do Porto

Abstract

The population decline of amphibians and reptiles has aroused worldwide attentions during the past decades. Human activities were considered as one of the most important potential factors leading to this problem. For studying the relationship between species distribution and human activities, 4 amphibians (*Pelobates cultripes*, *Bufo calamita*, *Hyla arborea*, and *Pleurodeles waltl*) and 4 reptiles (*Psammodromus algirus*, *Psammodromus hispanicus*, *Blanus cinereus*, and *Timon lepida*) were selected, and 2780 species presence points were obtained in the Salamanca Province of Spain from 2000 to 2002. Statistical frequency method and Bayes' theorem were applied for finding out the relationship between species presence and human activities influence (represented by human population density). Bayes' probability theorem could avoid the bias caused by sampling method and showed more reasonable results than frequency method. The results showed that both the amphibians and reptiles prefer comparatively weak human influence, which was most probably because strong human influence could cause the loss of species habitat and the death of the fragile species by accident. However, different species reacted differently along with the change of human influence. For those reptiles, the results showed the species presence decreased with the increase of human influence, while for those amphibians, the moderate human influence would favour the species presence to some extend. The reason could be in the moderate human influence area, human activities like building ponds and irrigating farmland closed to the villages provided more essential habitats for amphibians, especially for their larva. On the contrary, the growth of reptiles does not rely on water body very much.

For the purpose of better studying the species habitat, MAXENT (Maximum Entropy model) was applied for modelling the species distribution, especially considering human activities parameter. The modelling was constructed firstly with only environmental parameters and then with both environmental and human activities parameter for the eight species. The results showed that adding human activities parameter could slightly improve the modelling accuracy for most of species. In addition, the output distribution maps from two models had significant difference. The model added human activities parameter displayed a distribution map with lower habitat probability in human resident places, and more fragmental species habitat than the model with only environmental parameters, which means the former map was better indicating that species avoiding strong human influence. Studying the influence of human activities on species distribution and creating the species distribution maps can not only give a better understanding of human influence to the nature, but also help decision-maker to take proper methods protecting the species.

Acknowledgements

This thesis could not have come into being without the help and support of many people as well as institutions.

First of all, many thanks to Erasmus Mundus—GEM MSc course, and its organizers—the staff from Southampton University, Lund University, Warsaw University and ITC, as well as Prof. Liu Xuehua who introduced this course to me. During the past 18 months, I have not only learned knowledge from teachers, but also made good friends and learned lots of things from them which could not be found from any textbook. This experience will be remembered in my heart forever.

I would like to express my sincere thanks to Mr. Bas Wesselman. I appreciate the time you spent on numerous occasions to discuss with me about research issues, study problems as well as many other topics during my study. I am also very grateful for the discussions with Dr. A.G. (Bert) Toxopeus and Prof. Andrew Skidmore. Thank all of you for the valuable suggestions to the thesis and encouragement to me when I was depressed by the work.

Specially thanks to Dr. Sillero Neftalí and your colleagues from CIBIO, University of Porto. Without you cooperation and willingness to share data and assist in the field this research would not have been possible. Thank you also for your selfless and great help with research ideas and comments for my thesis. That was you who brought me to the world of amphibians and reptiles for the first time, and helped me really get close to those lovely species and taught me how to identify them and protect them. Your attitude to fieldwork and research impressed me and I will never forget in my future research.

I would like to thank all of my GEM classmates who were the most important part of my 18 months' fantastic life. From UK to Sweden, from Poland to the Netherlands, we met each other, knew each other, got used to each other, and it has been more and more difficult to leave each other. I've imaged millions of ways to say goodbye while none of them worked without tears. Sarika and Mila, I've never thought I could really use English to talk about everything with my best friends but you girls made it come true. And Shawn, thank you for being my live translator and for everything we've been through together. Girls, I love you all and hope we are best friends forever. Many thanks also to Chen since June of 2005 when I was disappointed by IELTS.

Last but not least, I would like to thank my parents and Shi Jianbo, for your love and support forever.

Table of contents

Abstract i

Acknowledgements ii

List of figures v

List of tables vii

Acronyms viii

1. GENERAL INTRODUCTION 1

1.1. Introduction 1

1.2. Research objectives 3

1.2.1. Overall objective *3*

1.2.2. Specific objectives *3*

1.3. Research questions 3

1.4. Research hypothesis 3

2. MATERIALS AND GENERAL METHODOLOGY 5

2.1. Study area 5

2.2. Field survey and species data 6

2.3. Parameters dataset 10

2.4. General introduction of methodology 12

3. HUMAN ACTIVITIES INFLUENCING SPECIES DISTRIBUTION 15

3.1. Introduction 15

3.2. Statistical frequency analysis 15

3.2.1. Method *15*

3.2.2. Results *15*

3.2.3. Bias of frequency method *17*

3.3. Bayes' probability theorem 18

3.3.1. What is Bayes' theorem *18*

3.3.2. Application of Bayes' theorem *19*

3.3.3. Results *20*

3.4. Statistical testing 22

3.5. Discussion 23

3.6. Conclusion 25

4. HUMAN ACTIVITIES PARAMETER INFLUENCING DISTRIBUTION MODELLING (MAXENT Model) 27

4.1. Introduction 27

4.2. MAXENT modelling 27

4.3. Results 28

4.3.1. Overall modelling accuracies for eight species 28

4.3.2. Statistical testing 29

4.3.3. Results for Pelobates cultripes *30*

4.3.4. Results of Psammodromus algirus *33*

4.4. Discussion 36

4.5. Conclusion 40

5. OVERALL CONCLUSION AND RECOMMENDATION 41

5.1. Human activities influence the distribution of species 41

5.2. Human activities parameter influence the distribution modelling 41

5.3. Comparison of frequency information and probability information 42

5.4. Limitation of the research 42

5.5. Recommendations 43

REFERENCE 45

APPENDIX 51

List of figures

Figure 1-1: Scheme displaying the impact on the distribution of an animal species of three broad categories of environmental factors. 2

Figure 2-1: Position of the Salamanca Province, Spain 5

Figure 2-2: Species field survey.. 7

Figure 2-3: Photographs of the eight species studied in the thesis 7

Figure 2-4: Species presence points from field survey for 8 species in Salamanca Province 9

Figure 2-5: General methodology of this thesis 13

Figure 3-1: Presence frequency of the 8 amphibians and reptiles corresponding with different human population density levels 16

Figure 3-2: Sampling equal areas in different classes 17

Figure 3-3: Comparison of the occupied area for every population density level within the whole study area 18

Figure 3-4: Presence probability of eight amphibians and reptiles. 21

Figure 3-5: Killed species on the roads of the Salamanca Province, at night 24

Figure 4-1: Output potential habitat maps of *Pelobates cultripes* from MAXENT: 31

Figure 4-2: Comparison of the two potential habitat maps of *Pelobates cultripes* from MAXENT, with threshold value 30. 32

Figure 4-3: The Jack-knife test of variable importance for modelling distribution of *Pelobates cultripes* 33

Figure 4-4: Output potential habitat maps of *Psammodromus algirus* from MAXENT 33

Figure 4-5: Comparison of the two potential habitat maps of *Psammodromus algirus* from MAXENT, with threshold value 16. 34

Figure 4-6: The Jack-knife test of variable importance for modelling distribution of *Psammodromus algirus* 35

Figure 4-7: Comparison of the frequency and probability trend lines for *Pelobates cultripes* and *Psammodromus algirus* 37

Figure 4-8: Response curves of human activities parameter (human population density) in the MAXENT models 38

Figure 4-9: The curve of exponential function $y=10^x$ 38

App. Figure 1: The random points generated from the study area 51

App. Figure 2: Correlation test of parameters about temperature 52

App. Figure 3: Correlation test of parameters about precipitation 53

App. Figure 4: Correlation test of parameters about terrain 54

List of tables

Table 2-1: Introduction of 8 species studied in the thesis 8

Table 2-2: List of available dataset and resources 11

Table 3-1: Correlation coefficients between species present frequency/probability and human population density 22

Table 4-1: MAXENT modelling AUC accuracy of the eight species 28

Table 4-2: T test for the difference between distribution maps with and without human activities parameters 30

Acronyms

AUC	The Area under ROC Curve
FAO	The Food and Agriculture Organization
GAM	Generalized Additive Models
GARP	Genetic Algorithm for Rule-Set Prediction
GIS	Geographical Information System
GLM	General Linear Models
GPS	Global Position System
GTOPO 30	Global 30 Arc Second Elevation Data
IUCN	International Union for Nature and Natural resource
MAXENT	The Maximum Entropy Method
NDVI	Normalized Difference Vegetation Index
ROC	The Receiver Operating Characteristic
RS	Remote Sensing
SPOT	Satellite Pour l'Observation de la Terre, satellite program of France
USGS	U.S. Geological Survey

1. GENERAL INTRODUCTION

1.1. Introduction

Amphibians and Reptiles are active species in the world. Amphibians (*Amphibia*) are composed by three groups of vertebrates: frogs, salamanders and caecilians, which have smooth, scaleless skin that is permeable to water (Duellman and Trueb 1994; Pough *et al.* 1998). Reptiles (*Reptilia*), which are amniotes, mainly include turtles, crocodilians, lizards, snakes and tuatara (Kimball 1986; Myers 2001).

Many amphibians and reptiles live in both aquatic and terrestrial habitats and are particularly sensitive to the natural ecosystems (Duellman and Trueb 1994). During their lifetime, those species interact with a large range of other species in the local environment (Gardner 2001). Therefore, amphibians and reptiles could be very helpful for indicating global environmental health. Additionally, with high collective biomass and high digestion and production efficiencies, amphibians and reptiles also contribute a lot to the ecosystem energetic and carbon flow (Pearman 1997).

Globally declines of amphibians and reptiles have aroused world-wide concern (Blaustein and Wake 1990; Wake 1991; Blaustein *et al.* 1994; Alford and Richards 1999; Christopher *et al.* 2003). For protecting the endangered species, many researches have been concentrated on studying the environmental and human factors that influence species habitats, as well as modelling their potential geographic distribution (Manel *et al.* 1999; Guisan and Zimmermann 2000; Carey *et al.* 2001; Parris 2001; Christopher *et al.* 2003; Parra-Olea *et al.* 2005; Elith *et al.* 2006; Latimer *et al.* 2006).

Various techniques like GPS (Global Position System), GIS (Geographical Information System) and RS (Remote Sensing) have been widely used for species presence-absence data collection, environmental information extraction and spatial analysis of the geo-information (Lassueur *et al.* 2006). Different models, such as MAXENT (The Maximum Entropy Method) (Phillips *et al.* 2004; Phillips *et al.* 2006), GARP (Genetic Algorithm for Rule-Set Prediction) (Stockwell and Peters 1999; Stockman *et al.* 2006; Stockwell *et al.* 2006), BIOMAPPER (Hirzel *et al.* 2002; Chefaoui *et al.* 2005), GLM (General Linear Models) (Brotons et al. 2004) and GAM (Generalized Additive Models) (Nicolas *et al.* 2002), have been

developed for modelling the species distribution, and also applied widely on amphibians and reptiles. In addition, species distribution information can also be used for conservation area planning and urban planning (Lofvenhaft *et al.* 2004; Maret *et al.* 2006).

The species distribution models are mainly built according to three types of information: a) environmental parameters (such as temperature, precipitation and elevation); b) human activities influence (like incautious killing, preying for economic profit); and c) the species resource database (species presence-absence information obtained primarily from field survey). The relationships of those factors were showed in Figure1-1. There are also many other parameters influencing the species distribution, such as historical factors (like topological and climate upheaval), and ecological factors (like species competition).

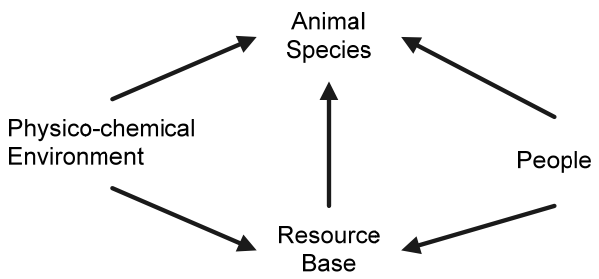


Figure 1-1: Scheme displaying the impact on the distribution of an animal species of three broad categories of environmental factors. (Skidmore 2002)

Current distribution models were normally constructed with only environmental variables and the variables derived from human activities were often ignored (Omullo 1996; Guisan and Zimmermann 2000). Generally speaking, environmental condition is the most important deciding factor for species habitat. However, with the development of human society, human activities, which have both direct and indirect influence to the species distribution (see Figure 1-1), have become one of the main reasons leading to the decline of some amphibians and reptiles (Goudie 1993; Alford and Richards 1999; Carey *et al.* 2001). For example, in many places of the world, the process of urbanization is occupying the species habitat, building of roads is leading to the habitat fragment, and some species are killed by people as food resource or even splurge of luxury. On the contrary, some species may like live with human because of increasing of farmland or ponds. Therefore, for better protecting the species, the relationship between species and people should be

studied, and the influence of human activities should be considered when mapping species distribution.

1.2. Research objectives

1.2.1. Overall objective

The overall objective of this thesis is to find out how human activities influence the distribution of amphibians and reptiles and the process of modelling in the Salamanca province, Spain.

1.2.2. Specific objectives

(a) To study the influence of human activities on species present frequency and probability;

(b) To compare and assess the outputs of distribution models (i) based on only environmental parameters, and (ii) based on both environmental and human activities parameter, applying MAXENT model.

1.3. Research questions

(a) Do human activities significantly influence the distribution of species, with reference to the available datasets?

(b) Does human activities parameter significantly influence the species distribution modelling, using MAXENT model?

1.4. Research hypothesis

H_1: Human activities significantly influence the distribution of species.
H_0: Human activities do not significantly influence the distribution of species.

H_1: Adding human activities parameter significantly influences the species distribution modelling.
H_0: Adding human activities parameter does not significantly influences the species distribution modelling.

2. MATERIALS AND GENERAL METHODOLOGY

2.1. Study area

Salamanca is a province of Spain, in the western part of the autonomous community of Castilla y León (as showed in Figure 2-1). To the north, Salamanca shares a border with Zamora and Valladolid; to the south, with Cáceres; to the east with Ávila; and to the west with Portugal. Salamanca occupies 12 336 Km^2 (Oviedo 2007) , which makes it the third-largest province of Castilla y León.

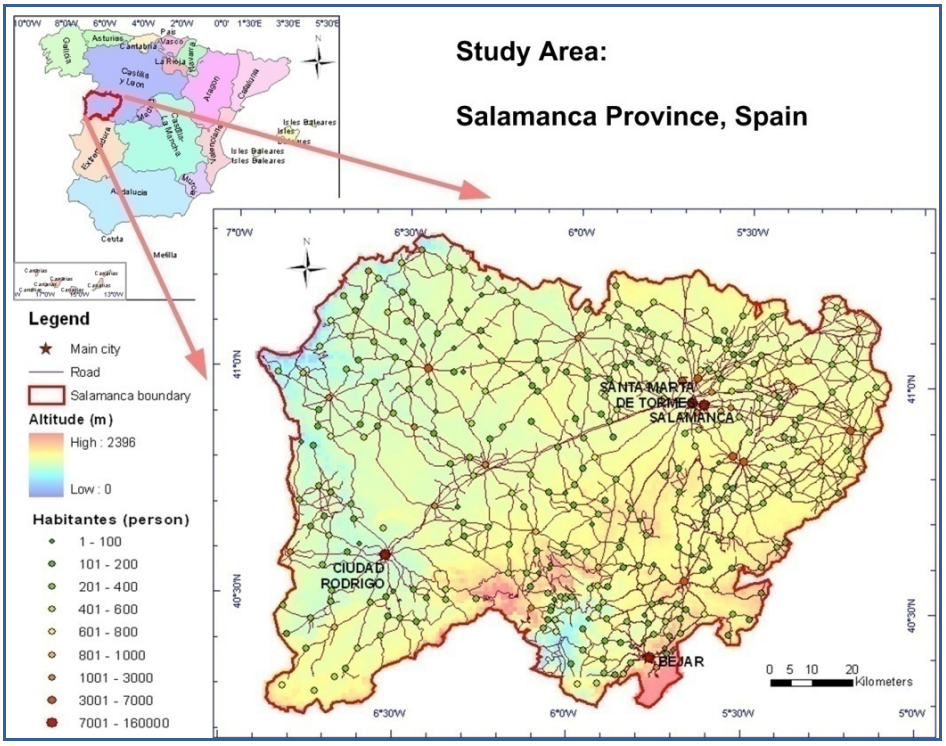


Figure 2-1: Position of the Salamanca Province, Spain

There are 347 120 people (2002) (Wikipedia 2007) live in Salamanca province, around 29 inhabitants per square kilometre, and the population growth is almost

stalled or even negative sometime. More than 45% of the population live in the capital, Salamanca city. The four biggest cities (Salamanca, Béjar, Ciudad Rodrigo, and Santa Marta de tormes), as showed in Figure 2-1, all have more than 10 000 people, and totally own around 60% population of the Salamanca Province. On the other hand, there are 362 municipalities in Salamanca province, of which more than half are villages with fewer than 300 people.

The Salamanca Province lies on Spain's Northern Plateau. It is mainly a flat area forming part of the Duero river system where Paleozonic soils, poorly suited to cultivation, have resulted in a landscape of pastures, dotted with trees and bushy undergrowth. The vast pastures are very suitable for cattle and pig farming, and thus the province is famous for its meat industry. The average altitude of the province is 830m above sea level, while the mountainous area, locating in south of the province, has an average of 900-950m (Sillero 2006).

The continental climate of Salamanca Province implies cold, dry winters and hot summers. The average temperature in winter is 4℃, and can plummet below zero on snowy days. However, summer temperatures average 25℃ and can rise as high as 35-40℃. The annual rainfall of Salamanca is about 637mm (SpanishTown 2007).

According to the introduction above, Salamanca Province has a comparatively concentrated population distribution, vast grassland and bushy area, and moderate climate. Many amphibian and reptile species were found in the Salamanca Province (Sillero 2006).

2.2. Field survey and species data

Most of the species present points were existing data collected from 2000 to 2002 (Sillero 2006). Some field surveys were carried through on 2006, mainly for investigating the habitats of different species and their potential relation with human influence.

Several methods were applied for field work, and the surveyed area has almost covered the whole Salamanca Province. Amphibians were surveyed using dip-netting, egg search on water vegetation and margins, refuge search, night search for calls and raining-night search by car for migrating individual on roads (see Figure 2-2a). Reptiles were searched using visual encounter surveys and refuge search (see Figure 2-2b) (Sillero 2006).

(a) (b)

Figure 2-2: Species field survey. (a) Raining night search by car for migrating individuals on roads; (b) Visual encounter survey for species during the day time.

Within all of the field survey data, eight species (four amphibians and four reptiles) are selected for this research. The selection process was mainly considered two criteria: firstly, the representability of species present points for possible distribution (that is, the species present points should better represent all possible habitats according to the expert knowledge of this species), and secondly, the diversity of species (mainly considering various family and different endangered levels). The photos of eight species were showed in Figure 2-3, and their general information was listed in Table 2-1.

Figure 2-3: Photographs of the eight species studied in the thesis

Among all of the eight species, *Psammodromus algirus* (Miras et al. 2005), *Psammodromus hispanicus* (Pleguezuelos *et al.* 2005b), *Blanus cinereus* (Pleguezuelos *et al.* 2005a), and *Timon lepida* (also named *Lacerta lepida* before) (Pleguezuelos *et al.* 2005c) are reptiles. *Pelobates cultripes* (Beja *et al.* 2006b), *Bufo calamita* (Beja et al. 2004), *Hyla arborea* (Böhme et al. 2006), and *Pleurodeles waltl* (Beja *et al.* 2006a) are amphibians. 3 of the species are listed as near threatened species in the IUCN Red List, and the others are less concerned.

Table 2-1: Introduction of 8 species studied in the thesis (IUCN 2007)

Name	*Class*	*Order*	*Family*	*English Name*	*Red List*
Psammodromus algirus	*Reptilia*	*Squamata*	*Lacertidae*	Large psammodromus	LC*
Psammodromus hispanicus	*Reptilia*	*Squamata*	*Lacertidae*	Spanish psammodromus	LC*
Blanus cinereus	*Reptilia*	*Squamata*	*Amephisbaenidae*	Iberian worm lizard	LC*
Timon lepida	*Reptilia*	*Squamata*	*Lacertidae*	Ocellated lizard	NT**
Pelobates cultripes	*Amphibia*	*Anura*	*Pelobatidae*	Western spadefoot	NT**
Bufo calamita	*Amphibia*	*Anura*	*Bufonidae*	Natterjack toad	LC*
Hyla arborea	*Amphibia*	*Anura*	*Hylidae*	European tree frog	LC*
Pleurodeles waltl	*Amphibia*	*Caudata*	*Salamandridae*	Spanish ribbed newt	NT**

* Listed as Least Concern in view of its wide distribution, presumed large population, and because it is unlikely to be declining fast enough to qualify for listing in a more threatened category.
** Listed as Near Threatened because this species is probably in significant decline (but probably at a rate of less than 30% over ten years) because of widespread habitat loss through much of its range and the impacts of invasive predators, thus making the species close to qualifying for Vulnerable.

The distribution of species presence points from field survey were showed in Figure 2-4. Most of the species have around 100 points and further more, *Psammodromus algirus*, *Pelobates cultripes*, and *Bufo calamita* have more than 300 points each.

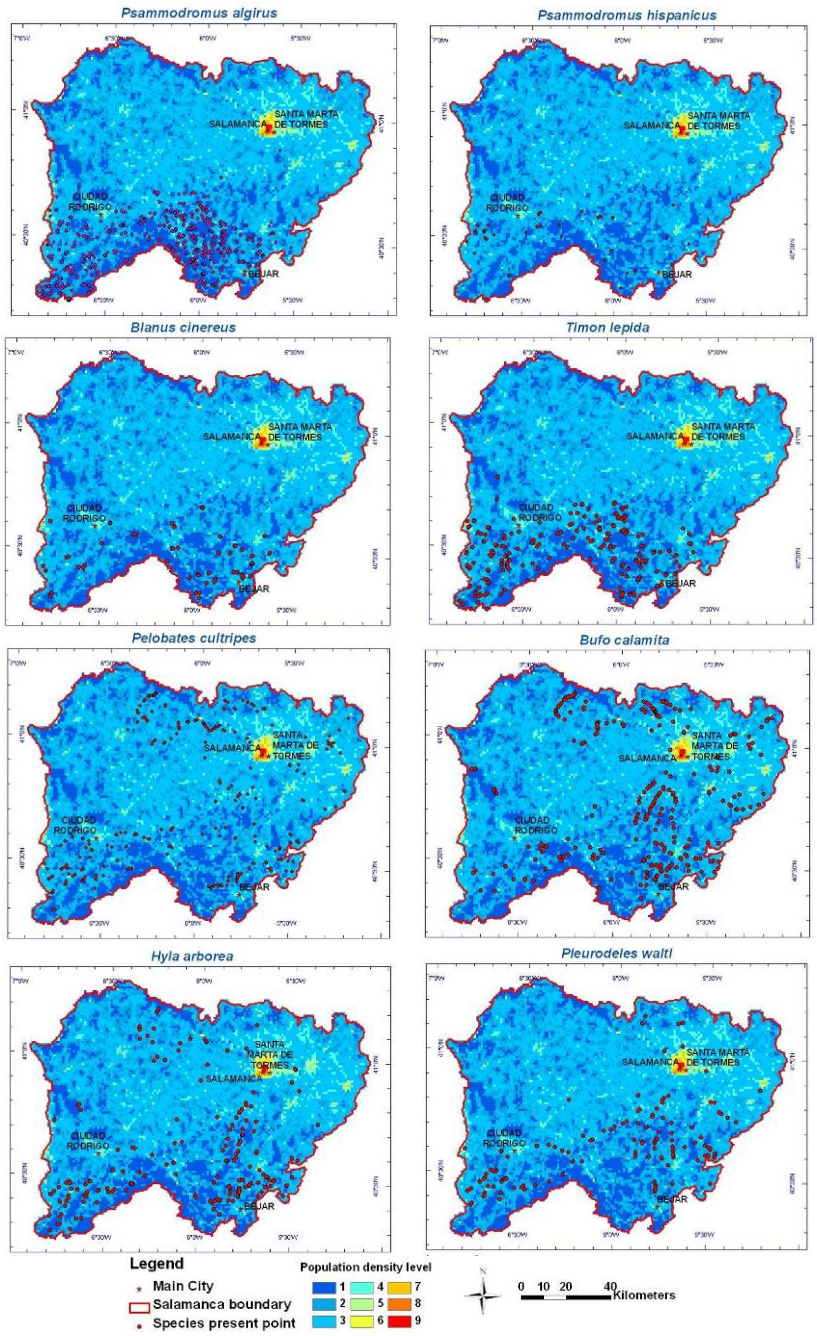


Figure 2-4: Species presence points from field survey for 8 species in Salamanca Province (NT: Population density level from 1 to 9 shows the density from the lowest to the highest)

2.3. Parameters dataset

The parameter data used for analysis and modelling were listed in Table 2-2.

Five kinds of environmental parameters were chosen for modelling:

- 2 sets of parameters from climate conditions (temperature and precipitation), including: annual mean temperature and precipitation, temperature and precipitation seasonality (coefficient of variation), mean temperature of warmest and coldest quarter, and precipitation of wettest and driest quarter;
- 3 parameters standing for terrain conditions (altitude, slope and aspect);
- 1 parameter for the soil condition (soil type);
- 1 parameter for distance to the water body;
- 1 parameter of NDVI (Normalized Difference Vegetation Index).

The selection of environmental parameters considered two criteria:

(1) Biologically, the parameter which has potential influence to species distribution will be considered;

(2) Statistically, for similar parameters, independent test will be applied to find the most representative parameter for saving time during the modelling.

In terms of the criteria above, statistical test for parameters about temperature, precipitation and terrain were carried out. The testing strategy and results were showed in the Appendix. Finally, eight environmental parameters were selected for distribution modelling, which are anural mean temperature, anural mean precipitation, altitude, slope, aspect, soil type, distance to water body, and NDVI.

Human population density was chosen as the human activities parameter, standing for the strength of human influence within a certain area. There are many parameters could represent the human influence, like landuse and the distance to the road. However, none of them is proper for this research. For landuse, it is difficult to simply classify the human influence according to the strength, because different landuse types could have totally different influence to species. While for the distance to the road, since one of our sampling methods is survey along the roads, this parameter is already biased and could not be used. Compared with landuse and distance to road, human population density avoided their shortages and thus suitable for this study. Therefore, human population density was divided to 9 levels, from 1 to 9 representing the density from the lowest where there is hardly people, to the highest where is the Salamanca City with around 4150 person/km² (see Figure 2-4).

All of those parameters were prepared to be raster files with the resolution of 1×1km. The spatial reference used for all of those data was WGS 1984 UTM ZONE 30N. In addition, the software used for data processing is ArcGIS 9.1.

Table 2-2: List of available dataset and resources

Category	*Material*	*Parameter used*	*Vector/Raster**	*Resource*
Environmental parameters	Climate	Annual mean temperature Annual mean precipitation	Raster	BIOCLIM (Houlder et al. 2000)
	Terrain	Altitude Slope Aspect	Raster	USGS GTOPO30
	Soil	Soil type	Vector (polygon)	FAO Soil (FAO 1995)
	Main water body	Distance to water body	Vector (line)	Local database, 2002
	Vegetation	NDVI	Raster	SPOT Vegetation May, 2006
Human activities parameter	Human population density		Raster	USGS
	Resident points with population		Vector (point)	Local database, 2002

* The resolution of original raster data is 1×1km.

2.4. General introduction of methodology

The overall aim of this study is to find out the influence of human activities on the species distribution and its modelling. The general study approach of this thesis was illustrated in Figure 2-5.

First of all, the collected data from field work and other resourced were grouped into three types (species data, human activities data, and environmental parameters data), and then formatted properly for the following research (in Chapter 2).

Secondly, the relation between human activities and species distribution was studied in chapter 3. Statistical frequency method and Bayes' probability theorem were applied. Further more, correlation test was applied to check out the significance of influence.

After that, in chapter 4, MAXENT modelling was run (i) with only environmental parameters and (ii) both environmental parameters and human activities parameter. Comparisons of models were applied on modelling accuracy assessment, output distribution maps, and the parameter sensitivity.

Finally, Chapter 5 provides an overall conclusion and recommendation for the research.

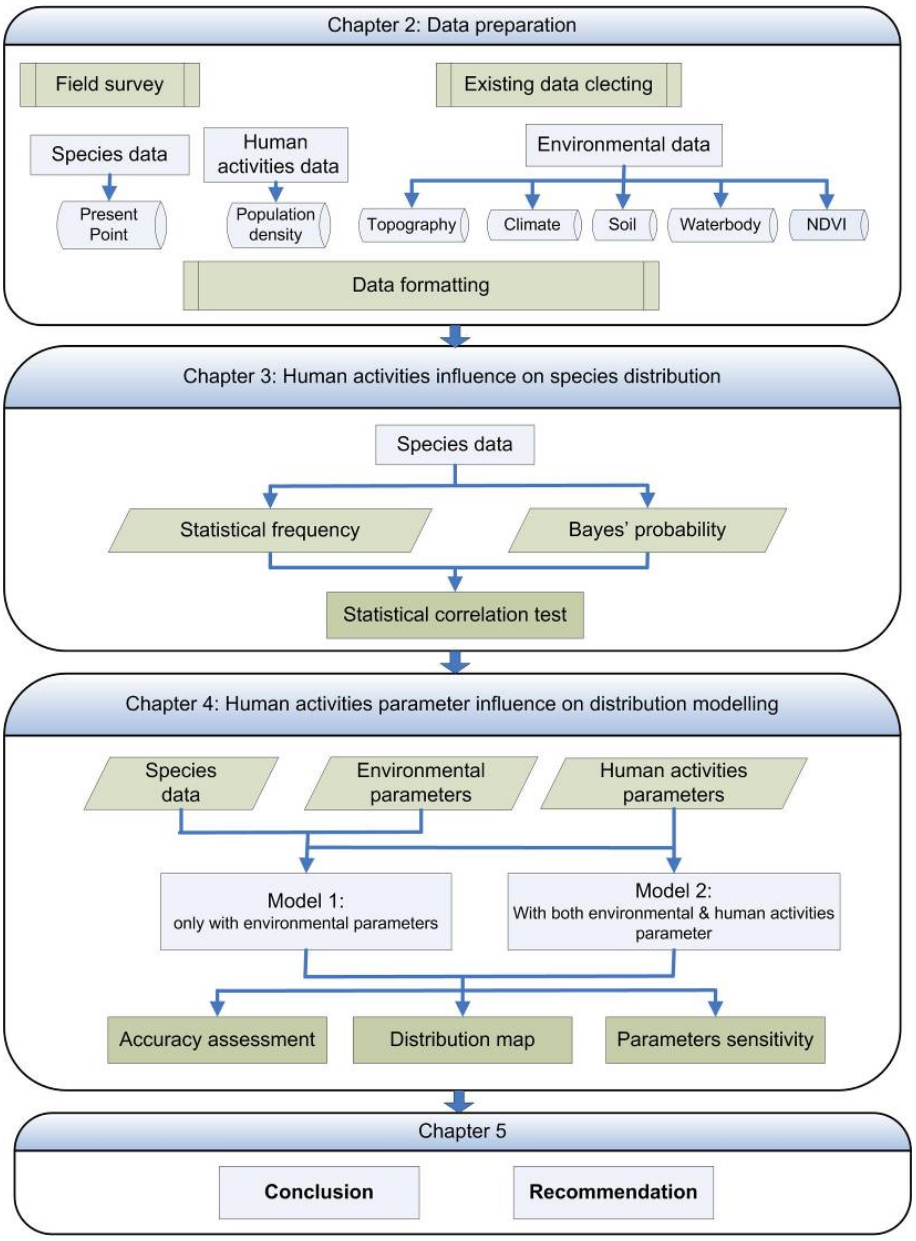


Figure 2-5: General methodology of this thesis

3. HUMAN ACTIVITIES INFLUENCING SPECIES DISTRIBUTION

3.1. Introduction

In this chapter, the main purpose is to study if human activities influence the distribution of eight amphibian and reptile species in the Salamanca Province. In another word, our objective is to find out if the species prefer certain strength of human influence (which means they prefer to live with a certain level of human population density). For achieving this objective, two methods were applied.

Firstly, through statistical analysis method, the frequency of the species presence was displayed against human population density levels, in a form of a histogram. However, this statistically counting method may cause bias when the areas for human population density levels significantly differ from each other (detailed discussion in chapter 2.2.3). To avoid this bias, on the other hand, Bayes' probability theorem was applied. The results of these two methods were compared and discussed for the purpose of gaining an insight of the relationship between human activities and species distribution. Finally, statistical testing was applied to check that with what kind of confidence, the species presence frequency/probability and human population density are correlated.

3.2. Statistical frequency analysis

3.2.1. Method

In this section, the research objective is to identify the human activities influence on species presence. Therefore, the preference of these species to live (or to cope) with a certain population density, have been addressed. To achieve this, the frequencies of species presence were counted, grouped by different human population density levels and displayed in the form of frequency histograms. Furthermore, the cumulative frequency of the species presence was calculated and the tendency and speed of this frequency to reach 100 percent, is noted.

3.2.2. Results

The results of the statistical frequency analysis were displayed in Figure 3-1. The figure shows that the presence points of all the eight species are mainly concentrated in lower population density area (normally lower than human population density level 4). However, the output histograms showed two different styles. In group 1,

there are four species (*Psammodromus algirus*, *Psammodromus hispanicus*, *Blanus cinereus*, and *Timon lepida*), which are all reptiles, presented almost the same high frequencies within the lowest 3 levels, and very low frequencies in the higher density levels. On the contrary, in group 2, another four species (*Pelobates cultripes*, *Bufo calamita*, *Hyla arborea*, and *Pleurodeles waltl*), which are all amphibians showed their preference within the first 3 levels, while with an obvious peak in area with human population density level 3.

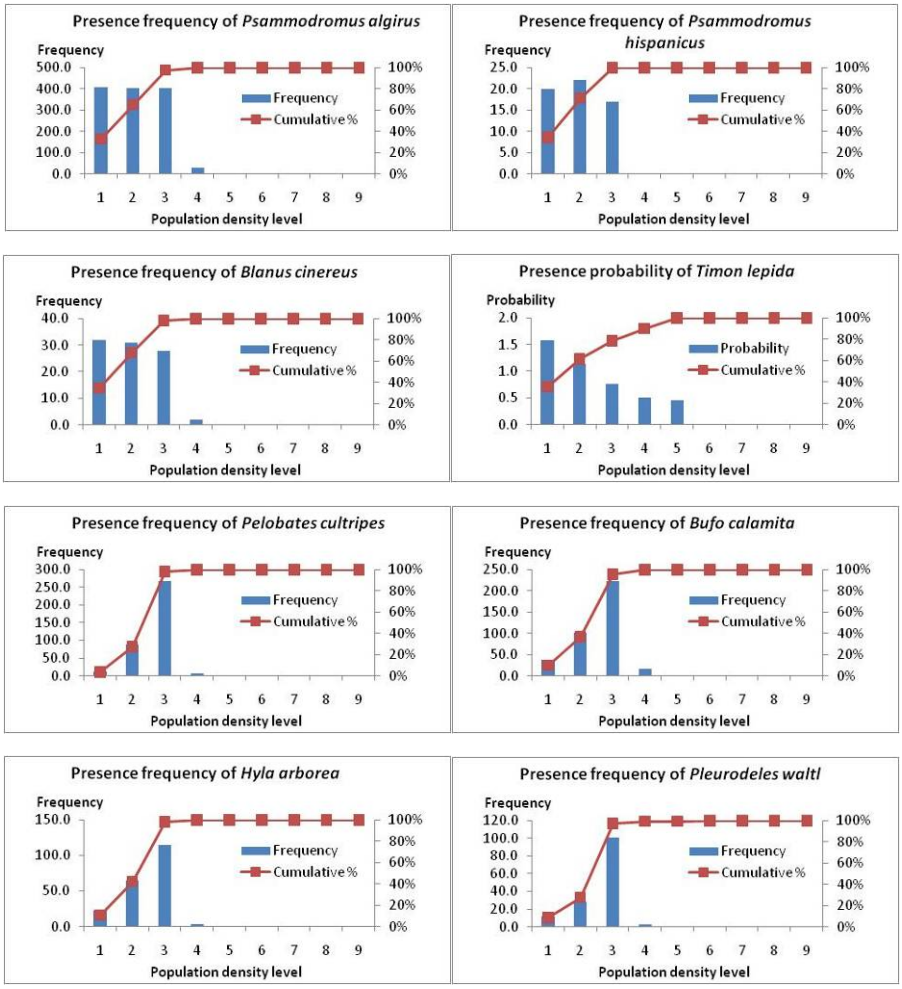
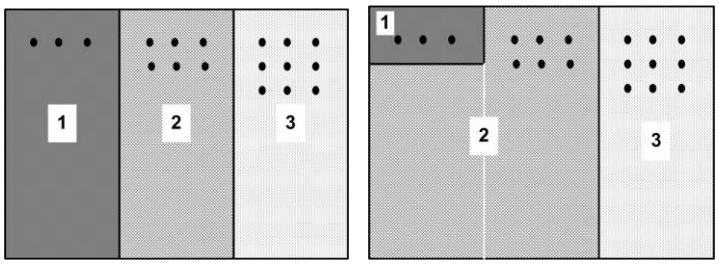


Figure 3-1: Presence frequency of the 8 amphibians and reptiles corresponding with different human population density levels (from 1 to 9 shows density from lowest to highest)

3.2.3. Bias of frequency method

The frequency method has a bias when displaying the preference of species, and the bias is derived from the method for obtaining the species data, the sampling method during field survey.



(a) (b)

Figure 3-2: Sampling equal areas in different classes (a) and unequal areas in different classes (b) (dots are species present points)

As we know, if the sampling area is equal for different background classes (as showed in Figure 3-2a), the more presence points, the more preference for the class. In that case, the frequencies can be used for comparing the different preferences in different classes. However, if the sampling area is different in different classes (see Figure 3-2b), that is when class 2 has much larger area than class 1, even though there are more presence points in class 2 (6 points) than class 1 (3 points), the frequencies can no longer accurately show the real preference of species because of the disturbance of sampling area.

In our case, animals are always moving and hiding, and therefore, it is not very easy to see them in a certain area at a certain time. The sampling method used for amphibians and reptiles (as explained in Chapter 2.2) was mainly random linear survey trying to cover the whole study area (Salamanca Province). Therefore, the visited area could not be the same for every class of human population density. In fact, the survey was applied within the whole study area and the comparison of the area for each human population density level was showed in Figure 3.3. Therefore, the bias was aroused when we counting the frequency of species presence for each level of human population density. In other words, the counts of points were not proper to show the real possibility of species presenting in a certain human density area. For getting the real preference of species, the disturbance of different sampling area must be removed, and thus, Bayes' probability theorem was applied and introduced in section 3.3.

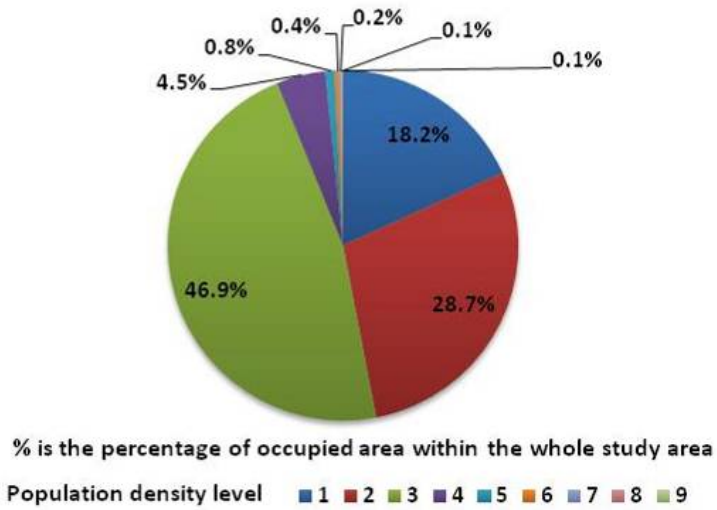


Figure 3-3: Comparison of the occupied area for every population density level within the whole study area

3.3. Bayes' probability theorem

3.3.1. What is Bayes' theorem

Probability is the extent to which something is likely to happen or be the case (Oxford 1948). Here, probability was used to describe how possible the species could present in an area with a certain level of human population density. Bayes' probability rule was used in this analysis, which relates the conditional and marginal probability distributions of random variables.

The function of Bayes' probability rule was showed as follows.

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \qquad (3.1)$$

In which,

A and B are two stochastic events;
$P(A)$ is the prior or marginal probability of A;
$P(B)$ is the prior or marginal probability of B;
$P(A|B)$ is the conditional probability of A, given B;
$P(B|A)$ is the conditional probability of B, given A.

3.3.2. Application of Bayes' theorem

In this research, what we want to know, is the probability of a certain species present in a certain level of population density, therefore, two events were also constructed:

A – Presence of a certain species;
Bi – Presence of the *i* level of population density.

Therefore, function 3.1 could be changed as follows,

$$\boldsymbol{P(A|Bi) = \frac{P(Bi|A) \times P(A)}{P(Bi)}} \qquad (3.2)$$

In which,

P(*A*) is the probability of a certain species present in the whole study area;
P(*Bi*) is the probability of *i* level population density present in the whole study area;
P(*A*|*Bi*) is the probability of a certain species present, under the condition of the population density level is *i* (that's what we want to know);
P (*Bi*|*A*) is the probability of *i* level population density, under the condition of a certain species present.

Here, *Bi* became a conditional influence of *A*, and when calculating *P*(*A*|*Bi*), the influence of *Bi* was removed, left only the probability of *A* (species presence) under the condition of *Bi* (the i level of human population density). According to the function above, three values were needed to calculate *P*(*A*|*Bi*).

(1) ***P*(*Bi*)**. It is the percentage of population density level *i* within the whole area, which can be calculated by the following function:

$$\boldsymbol{P(Bi) = \frac{Bi}{C}} \qquad (3.3)$$

In which,

Bi- number of pixels whose population density level is *i* in the whole study area;
C- number of pixels in the whole study area.

(2) ***P(Bi|A)***. It means to calculate the number of pixels of level *i* within the area where species is presence. Because the whole presence event *A* are not available, this value was replaced by *a*, a sample of *A*, which is the presence points obtained from field survey. Here, we assumed the sample can represent all of the species within the study area. And the calculation function is:

$$\boldsymbol{P(Bi|A) \approx P(bi|a) = \frac{bi}{a}} \qquad (\mathbf{3.4})$$

In which,

bi- number of pixels whose population density level is *i* in the sample;

a - number of pixels in the sample of present points.

(3) ***P(A)***. It is the percentage of species presence in the whole area. This value is also not available for us. However, since it is the same when calculating for every population density level, it could be ignored and will not influence the comparison of different human population density levels.

Therefore, inputting function 3.3 and 3.4 to 3.2, the final function was changed to:

$$\boldsymbol{P(A|Bi) \sim \frac{P(bi|a)}{P(Bi)} = \frac{bi/a}{Bi/C}} \qquad (\mathbf{3.5})$$

3.3.3. Results

The results of the presence probabilities for eight species are shown in Figure 3-4. In this figure, the results of frequency analysis were also showed as lines for the comparison of those two methods.

Using Bayes' probability method, the graph behaviours of the four species in group 1 (*Psammodromus algirus*, *Psammodromus hispanicus*, *Blanus cinereus*, and *Timon lepida*, which are all reptiles) were significantly changed. As we can see in Figure 3-4, instead of showing almost equal frequencies in the lowest 3 levels, all of those species displayed a peak in the level 1 and the probabilities gradually decreased with the increase of population density. The decrease of *Psammodromus hispanicus* was much sharper than the other three species within the group 1. However, for the four species in group 2 (*Pelobates cultripes*, *Bufo calamita*, *Hyla arborea*, and *Pleurodeles waltl*, which are all amphibians), the graph style of probabilities were

more or less the same with frequency method, maintaining the peaks in human population density level 3.

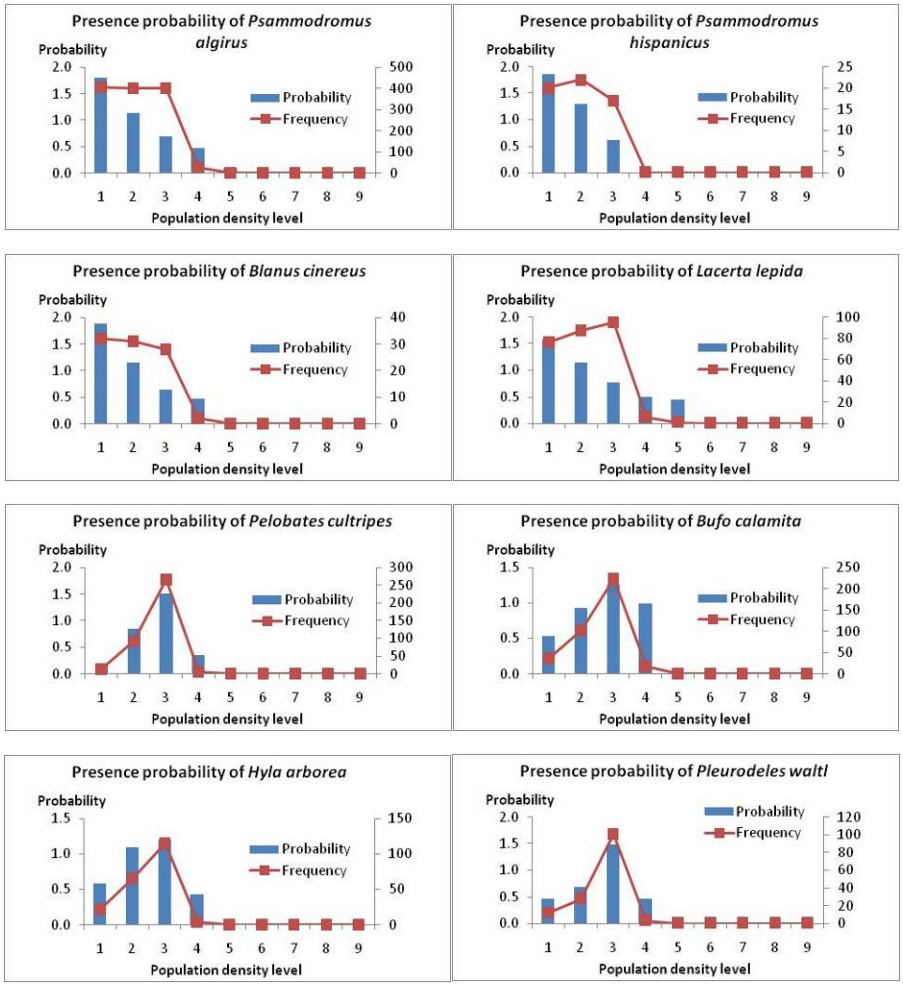


Figure 3-4: Presence probability of eight amphibians and reptiles.
NT: The probability is corresponded with population density (from 1 to 9 shows density from lowest to highest), and compared with frequency counting.

3.4. Statistical testing

To test if species present frequency and probability are significantly influenced by human activities, correlation test was applied in this section. The results were showed in Table 3-1. According to Table 3-1, there are obvious negative correlation between human population density and species present frequency/probability (R<0). With 95% confidence (P<0.05), the presence of all the reptile species (*Psammodromus algirus*, *Psammodromus hispanicus*, *Blanus cinereus*, and *Timon lepida*) have significantly negative correlation with human population density. On the other hand, 85% (P<0.15) confidence is needed for *Bufo calamita* and *Hyla arborea* to show the negative correlation. Furthermore, only 73% (P<0.27) confidence was showed for *Pelobates cultripes* and *Pleurodeles waltl*. The lower confidence of amphibians are because the influence of human population density on species is not absolutely negative but displayed positive on the level of moderate human population density (around level 3).

Table 3-1: Correlation coefficients between species present frequency/probability and human population density

Species name	***Human population density & species present frequency***		***Human population density & species present probability***	
	R*	***P****	***R****	***P****
Psammodromus algirus	-0.837	<= 0.008	-0.888	<= 0.004
Psammodromus hispanicus	-0.829	<= 0.009	-0.817	<= 0.010
Blanus cinereus	-0.846	<= 0.007	-0.868	<= 0.005
Timon lepida	-0.813	<= 0.011	-0.934	<= 0.002
Pelobates cultripes	-0.440	<= 0.238	-0.578	<= 0.106
Pleurodeles waltl	-0.462	<= 0.212	-0.414	<= 0.269
Bufo calamita	-0.552	<= 0.126	-0.737	<= 0.027
Hyla arborea	-0.580	<= 0.105	-0.784	<= 0.017

* R is correlation efficient, P is level of significant.

3.5. Discussion

There are two methods (statistical frequency analysis and Bayes' probability analysis) were used to study the relationship between species presence and human activities influence in this chapter.

First of all, the results of both methods showed that the presence of the eight species were mostly concentrated within the area with the lowest four levels of human population density. That means generally speaking, human activities have negative influence on the species presence of amphibians and reptiles in the Salamanca Province, which is also a worldwide problem causing species decline and habitat loss (Gibbon *et al.* 2000; Brooks *et al.* 2002). The specific reasons of the negative influence within our study area could be:

a. Human activities occupied the habitat of species. In this research, those amphibians and reptiles normally used to live in the places covered by some vegetation or closed to the waterbody. However, human activities like building human facilities (houses, roads) occupied the habitat of species, and thus reduced the species presence in the area. In the Salamanca Province, the expanding of urban area is slow compared with developing countries. Nevertheless, the capital Salamanca City was concentrated more and more people and more roads were built for linking the main cities in the past decades. All of those activities could occupy the habitat of species or cause fragmentation of habitat.

b. Amphibians and reptiles are easy to be killed by people on purpose or by accident. Sometimes, some special kinds of species are valuable for people, and could be used as food or medicine. While sometimes, the animals are killed because they are disturbing people's life. In our case, most probably some species are very easy to be killed by car especially at night (see Figure 3-5) when they are migrating, since these species are small and not easy to be noticed on the ground. Therefore, the higher population density, the heavier traffic on the road. Especially when the road was built on the migration route of species, lots of animals could be killed because of people's incaution.

Figure 3-5: Killed species on the roads of the Salamanca Province, at night

Secondly, different species have different reactions to the human influence. Bayes' probability analysis performed more clearly for this result (see Figure 3-4).

On one hand, the reptile species (*Psammodromus algirus*, *Psammodromus hispanicus*, *Blanus cinereus*, and *Timon lepida*) has decreased presence probabilities with the increase of human influence (human population density), which means human activities has negative influence to those species. Further more, within those four species, *Psammodromus hispanicus*, which usually lives in a very particular habitat, decreased the most sharply with the increase of human population density, and it presented only within the area with human population density level lower than 3. Compared with *Psammodromus hispanicus*, the other three species decreased more slowly and showed their stronger tolerance to human influence.

On the other hand, for those amphibian species (*Pelobates cultripes*, *Bufo calamita*, *Hyla arborea*, and *Pleurodeles waltl*), a proper extent of human activities (with human population density level around 3) is better than without any human influence. For this group of species, moderate human influence could improve the development of species to some extent.

The different behaviours between amphibians and reptiles could be caused by the demand of water. As we know, the growth of amphibians needs water body as a proper habitat, especially for their larva. Therefore, the human influenced area like villages with more ponds and ditches built by people are more suitable than the dry, desolate area without any human disturbance, for the surviving of amphibians. Compared with amphibians, reptile species do not have the special demand of water, and thus they do not have the special preference to live with a certain level of human population density.

Finally, comparing those two methods, Bayes' probability analysis removed the bias caused by sampling method, and showed clearer results than statistical frequency analysis. However, the disadvantage of probability method lies in the process of function 3.4 (Chapter 3.3.2), where the value of all the presence points within the study area were replaced by the points obtained from field survey. At that time, the sample must have the ability to represent the whole group (all the presence of species within the study area). To achieve that, what can be done is to make a random sample and increase the sample size. In our case, the whole group is not available for the research, and thus random points can not be selected from the whole group. Therefore, the solution is to enlarge the sampling size, which means more survey is needed. The larger the sampling size, the better of the sample representing the whole area. In terms of that, the probability method could be rectified by increasing of sampling size. On the contrary, frequency method will be more biased with the increase of sampling size, because sampling area will be more closed to the area of the background classes which is not equal in most of the times.

3.6. Conclusion

In this chapter, we can conclude as follows:

(1) The results showed human activities significantly influence the presence of species. And all of the species prefer comparatively lower human population density.

(2) However, those species do not react the same under different strengthens of human influence. Those reptile species have decreased presence with the increase of human influence. On the contrary, for those amphibians, moderate human activities have positive influence to the development of species.

(3) Bayes' probability analysis is better than statistically frequency analysis for studying the relation between species presence and human influence, while large sampling size is needed for getting more accurate results.

4. HUMAN ACTIVITIES PARAMETER INFLUENCING DISTRIBUTION MODELLING (MAXENT Model)

4.1. Introduction

In last chapter, we found out that human activities have significant influence to the species distribution. Therefore, in this chapter, MAXENT Model was used to study if adding human activities parameter will influence the distribution modelling of the eight species in the Salamanca Province. Within the eight species, one reptile (*Psammodromus algirus*) and one amphibian (*Pelobates cultripes*) were selected for the further processing, and more detailed comparison of the modelling outputs were given.

4.2. MAXENT modelling

Maximum entropy theorem was derived from statistical mechanics (Jaynes 1957; Good 1963). MAXENT Model for species geographic distribution modelling were developed by Phillips, et al (2004) and compared with other often used distribution model GARP (Phillips, et al 2006). The idea of MAXENT is to estimate the species distribution by finding the distribution of maximum entropy (i.e., that is closest to uniform), subject to the constraint that the expected value of each feature under this estimated distribution matches its empirical average (Phillips *et al.* 2004).

As displayed in section 2.3, eight environmental parameters (annual mean temperature, annual mean precipitation, altitude, slope and aspect, soil type, distance to the waterbody, and NDVI) and one human activities parameter (human population density) were prepared for the modelling.

After the data preparing, firstly, MAXENT was run with only eight environmental parameters, and then with both environmental parameters and human activities parameter. 75% of the data points were used for training the model and 25% were used for testing. The results were compared between different runs and different species in section 4.3.

4.3. Results

4.3.1. Overall modelling accuracies for eight species

The outputs accuracies of all models were showed in Table 4-1. From the table, we can see the difference of the accuracies between models with only environmental parameters and with both environmental and human activities parameters. The accuracy assessment method used here is ROC (the receiver operating characteristic) and AUC (the area under ROC curve).

Table 4-1: MAXENT modelling AUC accuracy of the eight species

Species	***Model with only environmental parameters***		***Model with environmental & human activities parameters***		***Accuracy improvement***	
	Training	***Test***	***Training***	***Test***	***training***	***test***
Psammodromus algirus	94.0%	92.6%	94.2%	92.7%	0.2%	0.1%
Psammodromus hispanicus	97.7%	90.4%	97.7%	90.5%	0.0%	0.1%
Blanus cinereus	94.4%	90.5%	94.3%	89.9%	0.0%	-0.6%
Timon lepida	94.3%	93.6%	94.3%	93.4%	0.0%	-0.2%
Pelobates cultripes	88.2%	83.6%	89.2%	85.0%	1.0%	1.4%
Bufo calamita	87.4%	87.2%	87.9%	89.0%	0.5%	1.8%
Hyla arborea	90.6%	86.4%	90.7%	86.7%	0.1%	0.3%
Pleurodeles waltl	93.3%	88.3%	94.3%	90.2%	1.0%	1.9%

Generally speaking, the accuracies of distribution models using MAXENT were very high (all higher than 85%). However, for most species, adding human activities parameter caused positive influence (improvement) to the distribution modelling. Especially for *Pelobates cultripes*, *Pleurodeles waltl* and *Bufo calamita*, the testing data accuracies were improved by 1.4%, 1.8% and 1.9%. However, for some species, like *Timon lepida* and *Blanus cinereus*, adding human activities parameter even caused negative influence to the modelling accuracy in terms of testing data.

Comparing reptile species with amphibian species, all the amphibians had positive improvement to the modelling when adding human activities parameter into the model, while reptile species had comparative low positive improvement and even negative improvement for some species. This showed human activities parameter played more important roles when modelling the distribution of those amphibian species in this study. Comparison of other outputs like distribution maps and importance of parameters will be discussed later.

4.3.2. Statistical testing

To test if human activities parameter significantly influenced the species distribution modelling, two tailed t test was applied to find out if the two output distribution maps made with and without human activities parameter are significantly different. To achieve this, 347 random points were selected representing the whole distribution map. The positions of the random points were showed in App. Figure 1, and the testing results were showed in Table 4-2.

According to Table 4-2, with 95% confidence ($P < 0.05$) we can say the differences between distribution maps with and without human activities parameter are significant for *Psammodromus algirus*, *Timon lepida*, *Pelobates cultripes*, *Hyla arborea*, and *Pleurodeles waltl*. In another word, human activities parameter has significant influence to the species distribution modelling for those species. On the contrary, the influence of human activities parameter was not significant for *Psammodromus hispanicus, Blanus cinereus*, and *Bufo calamita*.

Therefore, within the species whose distribution modelling were significantly influenced by adding human activities parameter, *Psammodromus algirus* and *Pelobates cultripes* were selected for the further analysis of the modelling outputs.

Table 4-2: T test for the difference between distribution maps with and without human activities parameters

Species	***Testing results (P*)***
Psammodromus algirus	0.00671
Psammodromus hispanicus	0.36706
Blanus cinereus	0.87254
Timon lepida	0.00002
Pelobates cultripes	0.00206
Bufo calamita	0.25092
Hyla arborea	0.00018
Pleurodeles waltl	0.04282

* P is level of significant.

4.3.3. Results for *Pelobates cultripes*

The output distribution maps of *Pelobates cultripes* from MAXENT Models were showed in Figure 4-1. In this figure, 4-1a showed the map from model using only environmental parameters, and 4-1b displayed the result from model using both environmental parameters and human activities parameter.

Comparing Figure 4-1a and Figure 4-1b, generally speaking, probability of potential habitat is lower in the map considered human activities parameter than the one without human parameter in many places. The decrease is more obvious within the area around Salamanca City, which can be seen in Figure 4-2c (the biggest city in the Salamanca Province which concentrated around 45% population of the whole province). The probability in Salamanca City area is significantly reduced from red or yellow to blue (the difference of probabilities between those two colours is around 10).

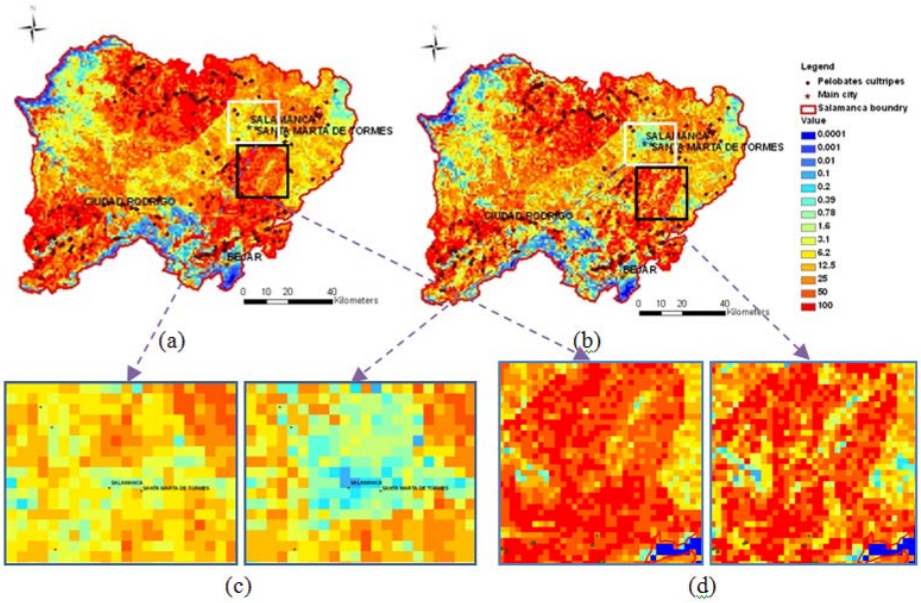


Figure 4-1: Output potential habitat maps of *Pelobates cultripes* from MAXENT: (a) model with only environmental parameters; (b) model with both environmental parameters and human activities parameter; (c) enlarged area around Salamanca City; (d) enlarged area in the southeast of Salamanca Province.

In addition, from Figure 4-1a to Figure 4-1b, the large area of high probability habitat (red area) were separated by small pieces of low probability area (yellow or blue area). This change is very obvious in the north, southeast (see Figure 4-2d), and southwest of the Salamanca Province. The phenomenon of separating large habitat is called fragmentation.

Those two distribution maps were converted to present-absent maps and the results were compared in Figure 4-2. In this figure, the blue pixels represented the pixels predicted to be presence by the model with only environmental parameters, which means those pixels were changed to be absence when adding human activities parameter. As we can see from the figure, those blue pixels normally distributed within large species habitats, and divides them into smaller pieces. There are also some pixels predicted to be presence by the model adding human activities parameter with changed to be absence when only using environmental parameters (green pixels), while that kind of pixels were more distributed between large habitats.

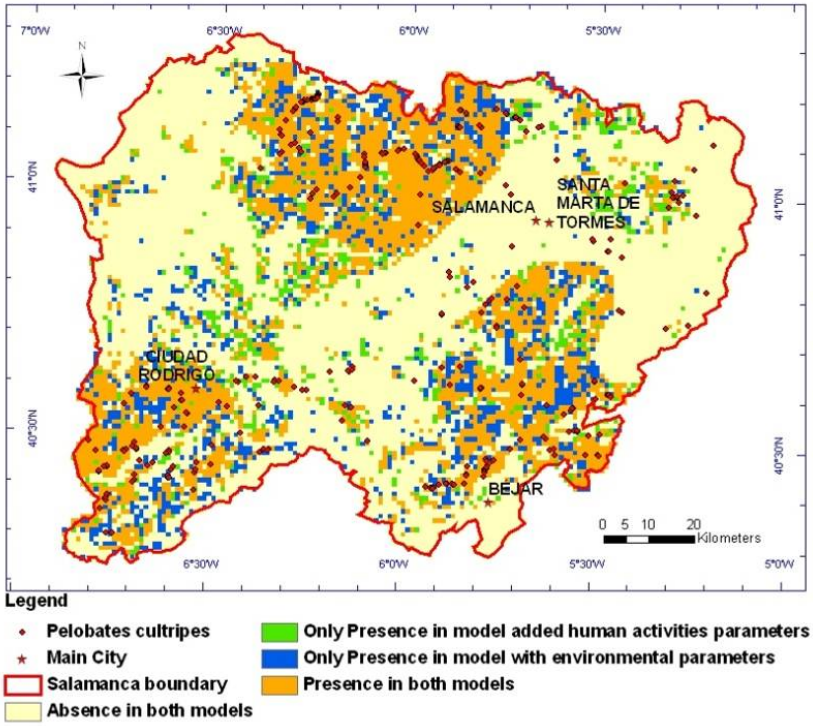


Figure 4-2: Comparison of the two potential habitat maps of *Pelobates cultripes* from MAXENT, with threshold value 30.

The comparison of variable importance in the model with both environmental and human parameters was showed in Figure 4-3, by means of Jack-knife test. From this figure, we can see the variable with highest gain when used in isolation is population density, which therefore appears to have the most useful information by itself. Furthermore, the variable that decreases the gain the most when it is omitted is also population density, which therefore appears to have the most information that isn't present in the other variables. Therefore, human population density was showed to be the most important parameter when modelling the distribution of *Pelobates cultripes*. The secondary important parameters were annual mean temperature, annual mean precipitation and slope.

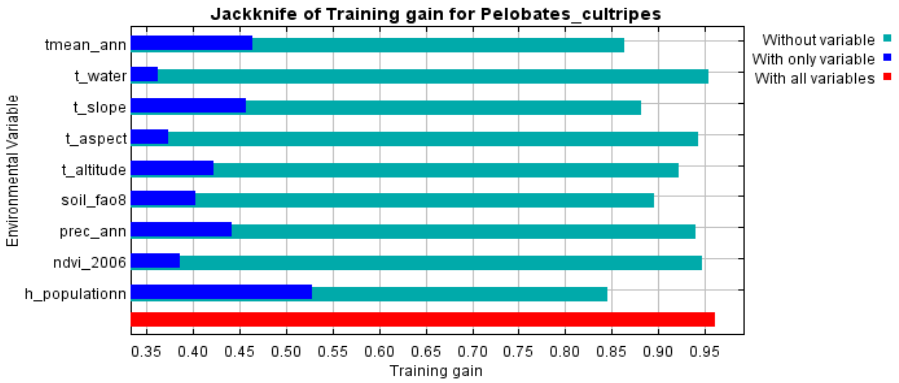


Figure 4-3: The Jack-knife test of variable importance for modelling distribution of *Pelobates cultripes*

4.3.4. Results of *Psammodromus algirus*

The output distribution maps of *Psammodromus algirus* from MAXENT Models were showed in Figure 4-4. In this figure, 4-4a showed the map from model using only environmental parameters, and 4-4b displayed the result from model using both environmental parameters and human activities parameter.

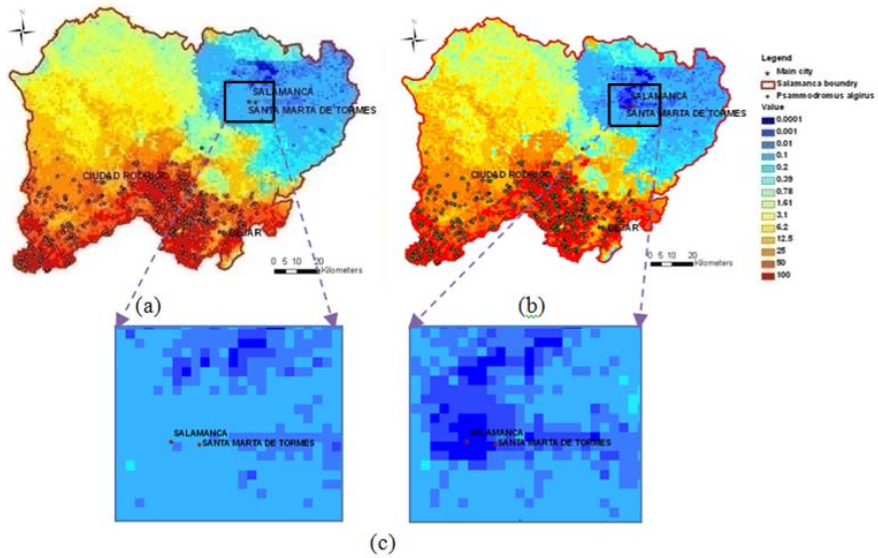


Figure 4-4: Output potential habitat maps of *Psammodromus algirus* from MAXENT: (a) model with only environmental parameters; (b) model with both environmental parameters and human activities parameter; (c) enlarged area around Salamanca City.

Comparing Figure 4-4a and Figure 4-4b, the general decrease of the probability for potential habitat is not as obvious as the results of *Pelobates cultripes*. The main high probability area, which is concentrated in the south part was remained almost the same. However, within the area near Salamanca city, the probability is still reduced a little bit, from light blue to dark blue (the difference of probabilities is only around 0.1).

There is no significantly fragmentation could be observed when adding human activities parameters to the distribution model. The influence of human activities to the modelling process mainly happened within the area with highest population density (the area around Salamanca City).

Those two distribution maps were converted to present-absent maps and the results were compared in Figure 4-5.

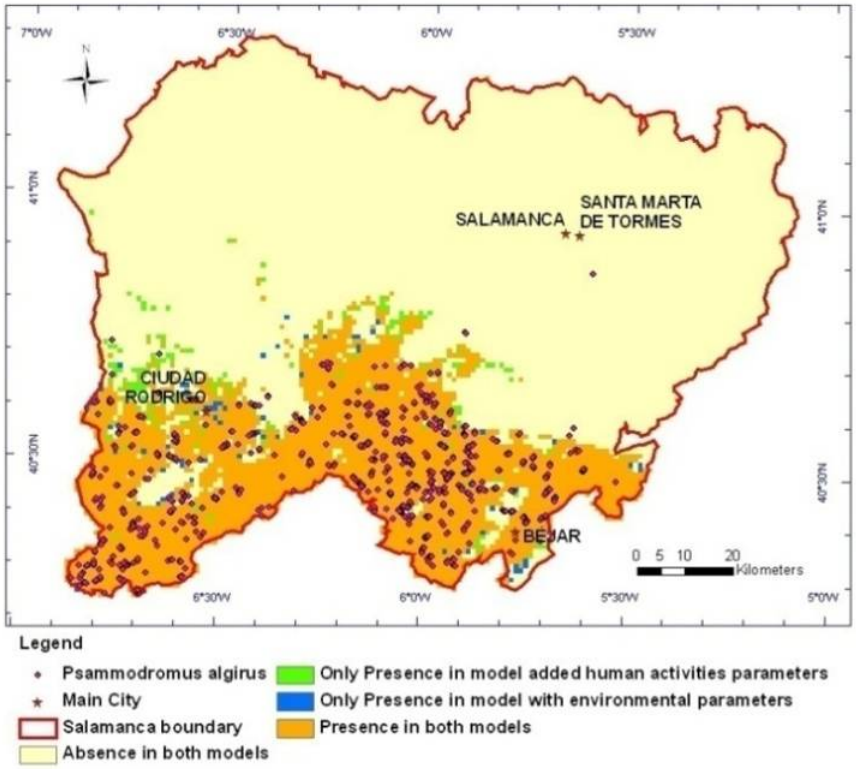


Figure 4-5: Comparison of the two potential habitat maps of *Psammodromus algirus* from MAXENT, with threshold value 16.

In Figure 4-5, most of the area predicted to be presence by the model with only environmental parameters was kept to be presence when adding human activities parameter (dark orange pixels). There are a few pixels used to be presence in the model with only environmental parameters and were changed to be absence when adding human activities (blue pixels). And there are a little bit more pixels used to be absence in the model with only environmental parameters and were changed to be presence when adding human activities (green pixels). The influence of adding human activities parameter didn't showed very clear in the distribution map.

The results of Jack-knife test for comparing variable importance in the model with both environmental and human parameters was showed in Figure 4-6. The variable with highest gain when used in isolation is soil type, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is annual mean temperature, which therefore appears to have the most information that isn't present in the other variables. The secondary important parameters were annual precipitation, slope and altitude. Human activities parameter has very low gain when used in isolation which means it did not contribute much to the modelling.

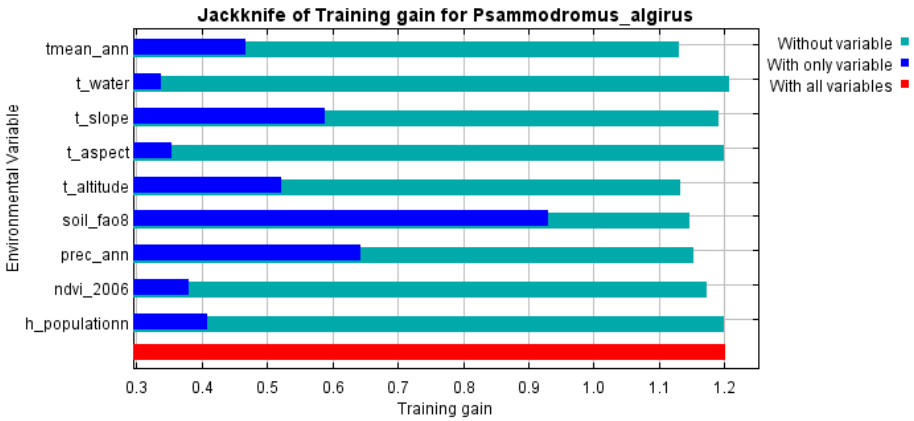


Figure 4-6: The Jack-knife test of variable importance for modelling distribution of *Psammodromus algirus*

4.4. Discussion

To study if human activities parameter influences the species distribution modelling, MAXENT model was applied in this chapter. The model was run once with only environmental parameters and once with both environmental and human activities parameter for every species.

The modelling results showed the accuracies of the output distribution maps are all higher than 85%, which indicated that MAXENT is very good at making use of the existing presence data for species distribution modelling. On the other hand, the influences of human parameter to the distribution modelling are different for different species. According to the comparison of modelling accuracies and influence significant test, most of the species were significantly influenced by human activities parameter while the human influence showed less significant for *Psammodromus hispanicus* and *Timon lepida*. The reason could be those two species have special preference for habitat and some other parameter played the main role.

In addition, the further comparison was applied in three aspects for *Psammodromus algirus* and *Pelobates cultripes*.

First of all, the output accuracies improvements were different for *Psammodromus algirus* and *Pelobates cultripes*. When adding human activities parameter into the model, more than 1% accuracy improvement were found for *Pelobates cultripes*, while only 0.1% improvement for *Psammodromus algirus*.

Secondly, comparing the output distribution maps, the influences of adding human activities parameters are different. For *Pelobates cultripes*, the distribution map from the model added human activities parameter showed not only reduction of habitat probabilities (especially area around Salamanca City reduce around 8%) in many places but also more fragmentation of species habitat (see Figure 4-2). Nevertheless, the results of *Psammodromus algirus* showed slightly decrease of probabilities even around Salamanca City (about 0.1), and there was no obvious fragmentation appeared comparing the two distribution maps (see Figure 4-5).

Finally, the sensitivities of human activities parameter are different during the modelling of *Psammodromus algirus* and *Pelobates cultripes*. Human activities parameter showed the largest sensitivity when modelling *Pelobates cultripes* while comparatively less sensitive than other parameters when modelling *Psammodromus algirus* (see Figure 4-3 and 4-6).

As a result, according to the three aspects discussed above, we can generally discover that human activities parameter played more important roles when modelling *Pelobates cultripes* than modelling *Psammodromus algirus*. Actually, *Psammodromus algirus* is a kind of wide distributed animal and not quite afraid of people, while *Pelobates cultripes* is an endangered species which could be influenced more buy human activities. Therefore, the reason for the different influence could be the other parameters are in fact more sensitive than human activities parameter for *Psammodromus algirus* than *Pelobates cultripes*. On the other hand, as we knew, both of these two species are influenced by human activities (Chapter 3), but in different ways. Therefore, the reason could also be the influence of human activities parameter was not expressed properly for *Psammodromus algirus*, which means that the sensitivity of parameter was ignored or reduced during the modelling process.

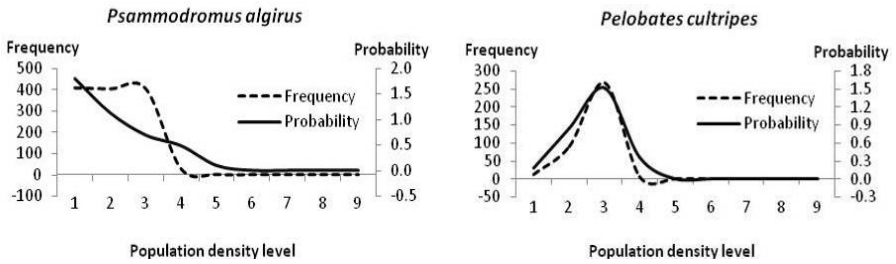


Figure 4-7: Comparison of the frequency and probability trend lines for *Pelobates cultripes* and *Psammodromus algirus*

To find out if the influence of human activities parameter was properly expressed, we should go back to Chapter 3. Comparing the two methods used in Chapter 3, it can be seen that the presence of both *Pelobates cultripes* and *Psammodromus algirus* are significantly influenced by human population density. However, the results of two methods showed different behaviours. For *Pelobates cultripes*, the sensitivities of human influence are almost the same for both of the methods (see Figure 4-7). While on the contrary, comparing the results from frequency method and probability method for *Psammodromus algirus*, it can be seen that the behaviour of frequency is less sensitive (the change is mainly between level 3 and 4) than the probability (the change is a gradually decrease from 1 to 6). Therefore, if MAXENT makes use of the information of frequency, the human activities parameter could not be sensitive and the influence of modelling will not be significant. Here, a question is aroused about what kind of information MAXENT is using, frequency or probability.

Now, we should have a look at the response curves for parameters with presence only data in MAXENT model. The response curves of human activities parameter are showed in Figure 4-8. These curves are used to show how each variable affects the MAXENT prediction and how the exponent changes as each variable is varied, keeping all other variables at their average sample value. Here, the response curves are not from the original models (with 25% points for testing models) but from the model run for all of the available point data. These new runs are better for the comparison with frequency and probability curves, because no testing data were missing and the curves better showed the theorem of models.

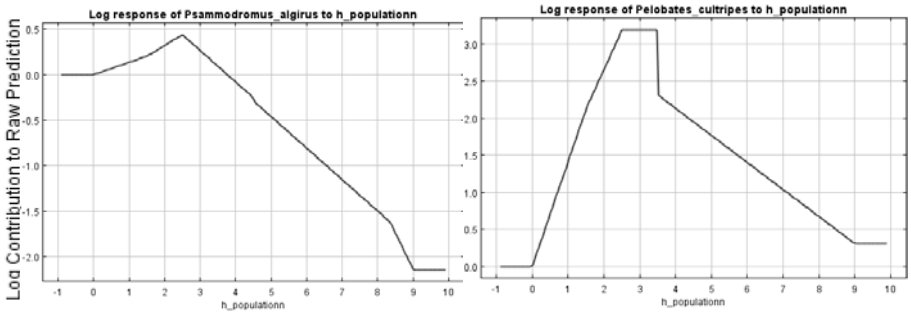


Figure 4-8: Response curves of human activities parameter (human population density) in the MAXENT models

In Figure 4-8, the vertical axis showed the log contributions to the raw prediction (could be simply expressed as X→log(Y), where X is the response value, and Y is the prediction). While, since the transposed function of log is $Y \rightarrow 10^X$, which is an increasing function (as showed in Figure 4-9), the function conversion did not change the trend of the curve. That is, the response curves displayed in Figure 4-8 have the same trend with the response curves of prediction in MAXENT model.

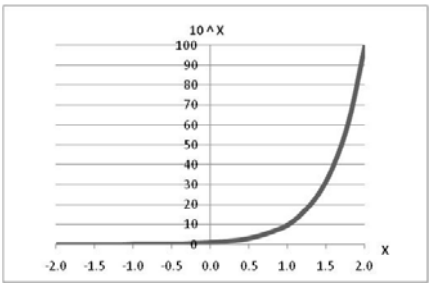


Figure 4-9: The curve of exponential function $y=10^x$

Comparing the curves of *Psammodromus algirus*, it can be seen that the trend of response curve is similar with the frequency curve. Therefore, another question

appeared: if the output response curves means the model is derived from frequency information of the available species presence data? To answer this, we should look at the inside of distribution modelling.

During the past decades, the focuses of distribution modelling are mainly on two aspects: one is how to deal with presence/absence data; and another one is how to combine different background environmental parameters and find out the most proper expression of distribution probabilities. For the models using presence only data, information about parameters from where species were found is used for training the distribution model. Some models, like MAXENT or some regression models, use this information trying to find the balance between different parameters. And thus, naturally, the more species presence points appeared in a certain section of parameter value range, the balance will incline to that section. That is, the frequency information was used to train the model according to the information from the area where species appears.

Frequency information will not cause bias if strict sampling method was taken. As discussed in Chapter 3.2.3, equal sampling area for each category (for categorized parameter) or for each section within the range (for continuous parameter) should be surveyed. At this time, the frequency information is the simplest and most direct method for showing the species' real preferred environmental conditions and will not cause bias.

Unfortunately, the equal area field survey is difficult to carry out. The reason is firstly, values of a parameter do not concentrate within an area but are mixed in the whole study area (like population density). For the purpose of equal area sampling, some area should be picked while others should be ignored. It is a waste of time especially when study area is big. Secondly, for continuous parameters, there is hardly some area with equal values, and thus either large numbers of sections have to be defined within the range or big sections should be chosen. Either of those will cause bias for the original data. Last but not least, several parameters are usually used altogether for the modelling, which means equal areas have to be chosen for each parameter. This is almost impossible when there are lots of parameters.

In terms of the discussion above, we can conclude that on one hand, improper sampling method causes the bias of frequency method, which leads to the bias of related distribution model. While on the other hand, most of the time perfect sampling is impossible for the species.

Compared with frequency information, probability information can better representing the preference of species (as discussed in Chapter 3.3). However, probability is not omnipotent, but also has many limitations. First of all, the parameter had better be category data. That is because if grouping continuous data for calculating the probability, some information will be lost and accuracy will be reduced for the parameter. Secondly, the field survey should be widely spread or randomly displayed within the whole study area, for the purpose of calculating function 3.4 in chapter 3.3.2. Finally, probability information could not be applied for the current distribution models (like MAXENT, GARP, GLM, GAM, etc), and thus new models are expected to be developed.

In our research, the human population density parameter is category data and our sampling was almost everywhere of the study area, which could be assumed covering the whole area. And thus it's proper to utilize Bayes' theorem to extract the probability information of human population density parameter and applied to the species distribution modelling.

4.5. Conclusion

In this chapter, we can conclude as follows:

(1) The output distribution maps have high accuracies (higher than 85% according to AUC) using MAXENT model.

(2) Human activities parameters significantly influenced the species distribution modelling for most of the species.

(3) The influences of adding human activities parameter to the species distribution modelling lie in 3 aspects:
   a. The accuracies are positively improved for most species.
   b. Species distribution probabilities were reduced especial within the area with higher population density, and more habitat fragmentation was displayed in the output distribution maps for *Pelobates cultripes*.
   c. Human activities parameter showed to be the most sensitive parameter during modelling distribution of *Pelobates cultripes*.

(4) Using frequency information in distribution modelling will cause bias when sampling is not properly designed. Probability method should be considered when its criteria are met.

5. OVERALL CONCLUSION AND RECOMMENDATION

5.1. Human activities influence the distribution of species

- Human activities significantly influence the distribution of species. Generally speaking, amphibians and reptiles prefer comparatively lower human population density.

- The reason for this phenomenon could be the decrease of proper habitat of species by occupation of human facilities and environmental pollution. And it could also be the killing of species by people on purpose or by accident.

- However, different species do not react the same under different strengthens of human influence. For those reptile species, human activities have significantly negative influence to the presence of species. However, those amphibians prefer moderate human influence. In another word, moderate disturbance is good for the development of some amphibian species.

- The reason for the different reaction could be amphibians preferred to be closed to some water bodies during their lifetime. And thus building of ponds and irrigation in the villages or countryside provide proper habitat for amphibians to survive.

5.2. Human activities parameter influence the distribution modelling

- Using MAXENT to model the distribution of amphibians and reptiles in this research, the accuracies of the output maps (assessed by AUC) are all higher than 85%.

- Human activities parameters significantly influenced the species distribution modelling for most of the species.

- Adding human activities parameter into MAXENT Model influenced the modelling more for *Pelobates cultripes* than *Psammodromus algirus*. This can be concluded as follows:

- The accuracies of modelling were improved more for *Pelobates cultripes* (around 1%) than *Psammodromus algirus* (around 0.1%);
- More habitat fragmentation and probability reduction were displayed in the output distribution maps for *Pelobates cultripes*;
- Human activities parameter showed to be the most sensitive parameter during modelling of *Pelobates cultripes* while comparatively less sensitive for *Psammodromus algirus*.

- The reason of the different influence between *Psammodromus algirus* and *Pelobates cultripes* could be biologically the former distributed widely and not afraid of people, while the latter is an endangered species which could be influenced more buy human activities.

5.3. Comparison of frequency information and probability information

- Statistically frequency method is biased when studying the influence of human activities on the species distribution, and the bias is caused by the species sampling method. On the contrary, Bayes' probability theorem considered the bias, and is more suitable to show the relation between species and human activities.

- Frequency information is widely used in species distribution modelling (like MAXENT), and thus the bias could be passed to the model.

- Using frequency information in distribution modelling will cause bias when sampling is not properly designed. Probability method should be considered when criteria are met.

- There are also many limits when applying Bayes' probability method, such as: categorized parameter is preferred; widely or random sampling is demanded, and information could be lost when grouping values.

5.4. Limitation of the research

- More methods are required during field survey, for the purpose of getting more accurate data.

- Larger area and wider distributed surveyed is need for avoiding the error when calculating probability.

- The output distribution maps based on the available presence points, even though have high accuracies while still have unavoidable errors resource, such as unvisited area and unconsidered parameters.

5.5. Recommendations

- The decline of distribution of amphibians and reptiles should arouse worldwide attention, and more researches should be applied for the species protection.

- Different protecting methods should be adopted for different species. For example, about those species who prefer to live with moderate human activities, pollution should be avoided near the residential area; while for those who prefer total natural habitat, any kind of human disturbance should be avoided within their habitat. And also, protecting large species habitat should be considered when building roads or other facilities.

- Probability information should be utilized in the distribution modelling. For example, method like Neural Network could be used for defining the relationship between species present probability and background parameters. Proper weight could be obtained and a suitable function could be applied for creating the distribution models.

- For more detailed research, small area could be picked (such as an area around a village) and finer resolution should be considered in the further research, which is better to understanding the relationship between species and human activities.

REFERENCE

Alford, R. A., and S. J. Richards. 1999. Global amphibian declines: a problem in applied ecology. *Annual Review of Ecology and Systematics* 30 (133-165).

Beja, P. et al. 2004. Bufo calamita: In: IUCN 2006. IUCN Red List of Threatened Species. <www .iucnredlist.org>. Downloaded on 22 January 2007.

Beja, P., J. Bosch, M. Tejedo, P. Edgar, D. Donaire-Barroso, M. Lizana, I. Martínez-Solano, A. Salvador, M. García-París, E. Recuero Gil, S. Tahar, E.H. El Mouden, and P. Geniez. 2006a. Pleurodeles waltl: In: IUCN 2006. IUCN Red List of Threatened Species. <www .iucnredlist.org>. Downloaded on 22 January 2007.

Beja, P., J. Bosch, M. Tejedo, M. Lizana, I. Martínez-Solano, A. Salvador, M. García-París, E. Recuero Gil, V. Perez-Mellado, C. Diaz Paniagua, M. Cheylan, R. Marquez, and P. Geniez. 2006b. Pelobates cultripes: In: IUCN 2006. IUCN Red List of Threatened Species. <www .iucnredlist.org>. Downloaded on 22 January 2007.

Blaustein, A. R., and D. B. Wake. 1990. Declining amphibian populations: A global phenomenon? . *Trends Ecology* 5:203-204.

Blaustein, A. R., D. B. Wake, and W. P Sousa. 1994. Amphibian declines: judging stability, persistence and susceptibility of populations to local and global extinctions. *Cons. Biol.* 8 (60-71).

Böhme, W., S. Kuzmin, D. Tarkhnishvili, V. Ishchenko, B. Tuniyev, T. Papenfuss, S. Anderson, M. Sparreboom, I. Ugurtas, P. Beja, F. Andreone, P. Nyström, B. Schmidt, B. Anthony, A. Ogrodowczyk, M. Ogielska, J. Bosch, M. Vogrin, M. Tejedo, D. Cogalniceanu, T. Kovács, I. Kiss, M. Puky, J. Vörös, M. Lizana, I. Martínez-Solano, A. Salvador, M. García-París, E. Recuero Gil, R. Marquez, M. Cheylan, and I. Haxhiu. 2006. Hyla arborea: In: IUCN 2006. IUCN Red List of Threatened Species. <www .iucnredlist.org>. Downloaded on 22 January 2007.

Brooks, Thomas M., Russell A. Mittermeier, Cristina G. Mittermeier, Gustavo A. B. da Fonseca, Anthony B. Rylands, William R. Konstant, Penny Flick, John Pilgrim, Sara Oldfield, Georgina Magin, and Craig Hilton-Taylor. 2002. Habitat Loss and Extinction in the Hotspots of Biodiversity. *Conservation Biology* 16 (4):909.

Brotons, Lluı´s, Wilfried Thuiller, Miguel B. Arau´jo, and Alexandre H. Hirzel. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27:437-448.

Carey, C., R. W. Heyer, J. Wilkinson, R. A. Alford, J. W. Arntzen, T. Halliday, L. Hungerford, K. R. Lips, E. M. Middleton, S. A. Orchard, and A. S. Rand. 2001. Amphibian declines and environmental change: use of remote sensing data to identify environmental correlates. *Conservation Biology* 15:903-913.

Chefaoui, Rosa M., Joaquı´n Hortal, and Jorge M. Lobo. 2005. Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian Copris species. *Biological Conservation* 122:327–338.

Christopher, J. Raxworthy, Martinez-Meyer Enrique, Horning Ned, A. Nussbaum Ronald, E. Schneider Gregory, A. Ortega-Huerta Miguel, and A. Peterson, Townsend. 2003. Predicting distributions of known and unknown reptile species in Madagascar. *NATURE* 426 (18/25):837-840.

Duellman, W. E. , and L. Trueb. 1994. *Biology of Amphibians*: John Hopkins University Press.

Elith, Jane, Catherine H. Graham, Robert P. Anderson, Miroslav Dudik, Simon Ferrier, Antoine Guisan, Robert J. Hijmans, Falk Huettmann, John R. Leathwick, Anthony Lehmann, Jin Li, Lucia G. Lohmann, Bette A. Loiselle, Glenn Manion, Craig Moritz, Miguel Nakamura, Yoshinori Nakazawa, Jacob McC. M. Overton, A. Townsend Peterson, Steven J. Phillips, Karen Richardson, Ricardo Scachetti-Pereira, Robert E. Schapire, Jorge Soberon, Stephen Williams, Mary S. Wisz, and Niklaus E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29 (2):129-151.

FAO. 1995. The digital soil map of the world: FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. <http://www.fao.org/AG/agl/agll/prtsoil.stm>.

Gardner, T. 2001. Declining amphibian populations: a global phenomenon in conservation biology. *Animal Biodiversity and Conservation* 22 (4):25-44.

Gibbon, J. Whitfield, David E. Scott, Travis J. Ryan, Kurt A. Buhlmann, Tracey D. Tuberville, Brian S. Metts, Judith L. Greene, Tony Mills, Yale Leiden, Sean Poppy, and Christopher T. Winne. 2000. The Global Decline of Reptiles, Déjà Vu Amphibians. *BioScience* 5- (8):653–666.

Good, I.J. 1963. Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *The Annals of Mathematical Statistics* 34:911-934.

Goudie, Andrew. 1993. Human influence in geomorphology. *Geomorphology* 7 (1-3):37-59.

Guisan, Antoine, and Niklaus E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-186.

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

Houlder, David, Michael Hutchinson, Henry Nix, and June McMahon. 2000. BIOCLIM: the Centre for Resource and Environmental Studies at the Australian National University. <http://cres.anu.edu.au/outputs/anuclim/doc/bioclim.html>.

IUCN. 2007. www.iucnredlist.org: Accessed at Jan 23, 2007.

Jaynes, E.T. 1957. Information theory and statistical mechanics. *Physical review* 106:620-630.

Kimball, J.W. . 1986. *Biology*. 5th ed ed. Don Mills: The Benjamin/Cummings Publishing Company, Inc. 972 p.

Lassueur, Thierry, Stephane Joost, and Christophe F. Randin. 2006. Very high resolution digital elevation models: Do they improve models of plant species distribution? *Ecological Modelling* In Press, Corrected Proof.

Latimer, A. M., S. Wu, A. E. Gelfand, and J. A. Silander Jr. 2006. Building statistical models to analyze species distributions. *Ecological Applications* 16 (1):33-50.

Lofvenhaft, Katarina, Siv Runborg, and Per Sjogren-Gulve. 2004. Biotope patterns and amphibian distribution as assessment tools in urban landscape planning. *Landscape and Urban Planning* 68 (4):403-427.

Manel, Stephanie, Jean-Marie Dias, and Steve J. Ormerod. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling* 120 (2-3):337-347.

Maret, Timothy J., Jonathan D. Snyder, and James P. Collins. 2006. Altered drying regime controls distribution of endangered salamanders and introduced predators. *Biological Conservation* 127 (2):129-138.

Miras, J.A.M., M. Cheylan, M.S. Nouira, U. Joger, P. Sá-Sousa, and V. Pérez-Mellado. 2005. Psammodromus algirus: In: IUCN 2006. IUCN Red List of Threatened Species. <www .iucnredlist.org>. Downloaded on 22 January 2007.

Myers, P. . 2001. "Reptilia" (On-line): Animal Diversity Web. Accessed January 23, 2007 at http://animaldiversity.ummz.umich.edu/site/accounts/information/Reptilia.html.

Nicolas, Ray, Lehmann Anthony, and Joly Pierre. 2002. Modeling spatial distribution of amphibian populations: a GIS approach based on habitat matrix permeability. *Biodiversity and Conservation* (11):2143–2165.

Omullo, L. 1996. Modelling the distribution of wildlife in relation to environmental gradients : a case study of the Tsavo national park, Kenya, ITC, Enschede.

Oviedo, Cuesta de. 2007. Salamanca, City and Province http://www.palaciocongresossalamanca.com/esp/ver.asp?ver=salamanca&idioma=ing: Salamanca Convention Centre. Accessed at Jan 23, 2007.

Oxford. 1948. *Oxford Dictionary of English, Second Edition*. New York: Oxford University Press, Inc.

Parra-Olea, G., E. Martinez-Meyer, and G. Perez-Ponce De Leon. 2005. Forecasting climate change effects on salamander distribution in the highlands of central Mexico. *Biotropica* 37 (2):202-208.

Parris, Kirsten M. 2001. Distribution, habitat requirements and conservation of the cascade treefrog (Litoria pearsoniana, Anura: Hylidae). *Biological Conservation* 99 (3):285-292.

Pearman, N, P. B. 1997. Correlates of amphibian diversity in an altered landscape of Amazonian Ecuador. *Conservation Biology* 11:1211-1225.

Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190 (3-4):231-259.

Phillips, Steven J., Dud´ık Miroslav, and Robert E. Schapire. 2004. A Maximum Entropy Approach to Species Distribution Modeling. *In Proceedings of the Twenty-First International Conference on Machine Learning*:655-662.

Pleguezuelos, J., P. Sá-Sousa, V. Pérez-Mellado, and R. Marquez. 2005a. Blanus cinereus: In: IUCN 2006. IUCN Red List of Threatened Species. <www .iucnredlist.org>. Downloaded on 22 January 2007.

Pleguezuelos, J., P. Sá-Sousa, V. Pérez-Mellado, R. Marquez, and M. Cheylan. 2005b. Psammodromus hispanicus: In: IUCN 2006. IUCN Red List of Threatened Species. <www .iucnredlist.org>. Downloaded on 22 January 2007.

Pleguezuelos, J., P. Sá-Sousa, V. Pérez-Mellado, R. Marquez, M. Cheylan, and C Corti. 2005c. Timon lepidus: In: IUCN 2006. IUCN Red List of Threatened Species. <www .iucnredlist.org>. Downloaded on 22 January 2007.

Pough, F.H., R.M. Andrews, J.E. Cadle, M.L. Crump, A.H. Savitzky, and K.D. Wells. 1998. *Herpetology*. Toronto: Prentice Hall of Canada, Inc, 577p.

Sillero, Neftali. 2006. Application of remote sensing and geographical information systems in the biogeography of Iberian amphibians and reptiles, University of Leon.

Skidmore, Andrew. 2002. *Environmental Modelling with GIS and Remote Sensing*. London: Taylor & Francis.

SpanishTown. 2007. Facts About Salamanca http://www.spanish-town-guides.com/Salamanca_Facts.htm: Accessed at Jan 23, 2007.

Stockman, A. K., D. A. Beamer, and J. E. Bond. 2006. An evaluation of a GARP model as an approach to predicting the spatial distribution of non-vagile invertebrate species. *Diversity and Distributions* 12 (1):81-89.

Stockwell, D., and D. Peters. 1999. The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Systems* 13 (2):143-158.

Stockwell, David R. B., James H. Beach, Aimee Stewart, Gregory Vorontsov, David Vieglais, and Ricardo Scachetti Pereira. 2006. The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas of species biodiversity. *Ecological Modelling* 195 (1-2):139-145.

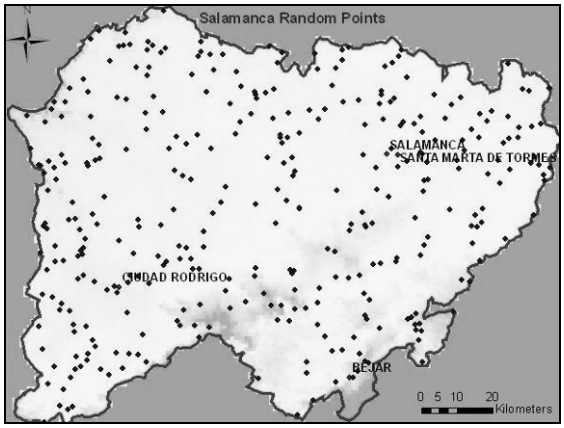
Wake, D. B. 1991. Declining amphibian populations. *Science* 203:860.

Wikipedia. 2007. http://en.wikipedia.org/wiki/Salamanca_%28province%29: Accessed at Jan 23, 2007.

APPENDIX

The correlation test of grouped parameters

There are 347 random points were selected from the whole study area and the values of each parameter in those points were extracted for the correlation test (see App. Figure 1).



App. Figure 1: The random points generated from the study area

Three groups of parameters were tested using JMP software (http://www.jmp.com/) and spearman's test, which are:

(1) Parameters about temperature: annual mean temperature, temperature seasonality, and the mean temperature of warmest and coldest quarter;
(2) Parameters about precipitation: annual mean precipitation, precipitation seasonality, and the precipitation of wettest and driest quarter;
(3) Parameters about terrain conditions: altitude, slope and aspect.

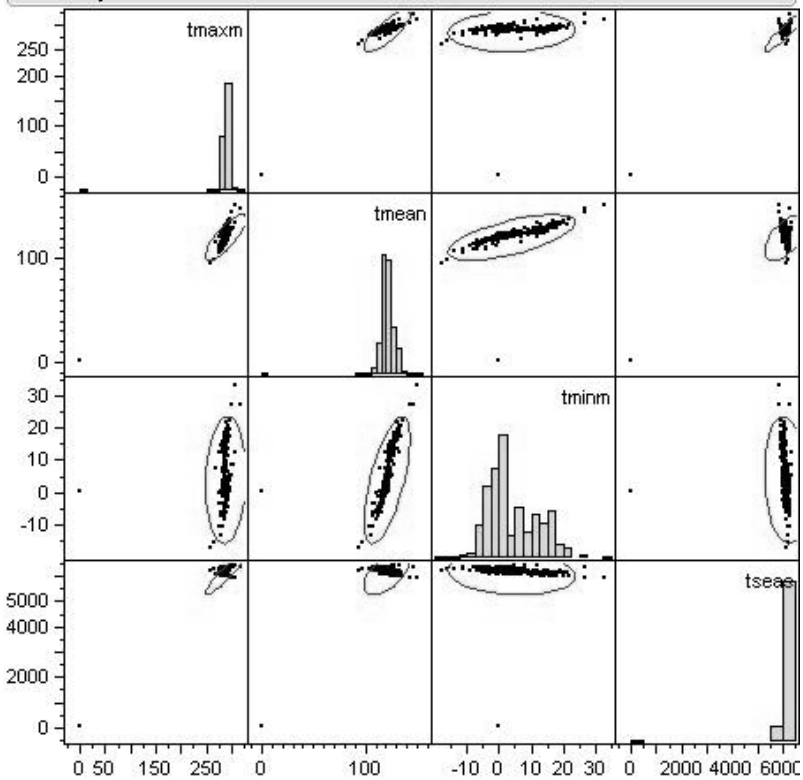
The results were showed in App. Figure 2, 3 and 4. According to the results, annual mean temperature and annual mean precipitation are highly correlated to other parameters within their groups, and thus could be used representing the groups. However, for terrain parameters, they are not correlated to each other, and therefore should be all used in the models.

Multivariate

Correlations

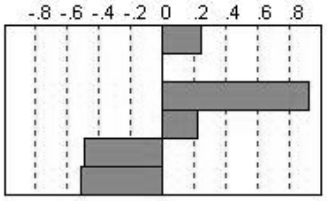
	tmaxm	tmean	tminm	tseas
tmaxm	1.0000	0.7992	0.0808	0.9282
tmean	0.7992	1.0000	0.6559	0.6013
tminm	0.0808	0.6559	1.0000	-0.1287
tseas	0.9282	0.6013	-0.1287	1.0000

Scatterplot Matrix



Nonparametric: Spearman's ?

Variable	by Variable	Spearman ?	Prob>|?|
tmean	tmaxm	0.2413	<.0001*
tminm	tmaxm	0.0000	0.9996
tminm	tmean	0.9324	<.0001*
tseas	tmaxm	0.2286	<.0001*
tseas	tmean	-0.4892	<.0001*
tseas	tminm	-0.5092	<.0001*



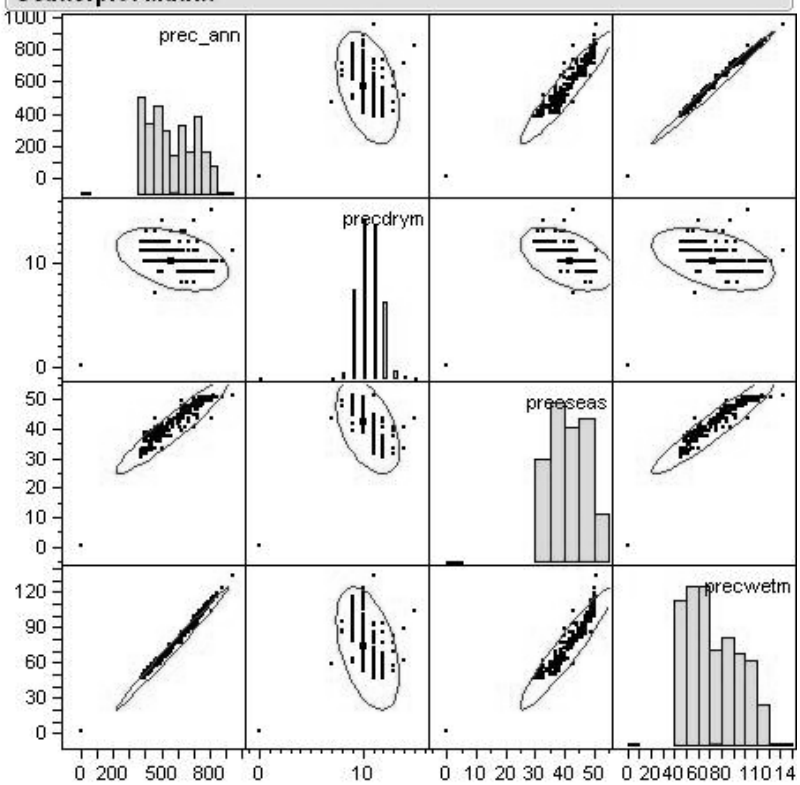
App. Figure 2: Correlation test of parameters about temperature

Multivariate

Correlations

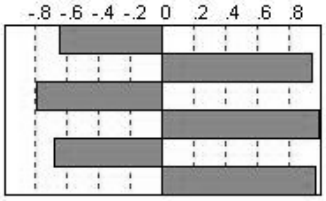
	prec_ann	precdrym	precseas	precwetm
prec_ann	1.0000	-0.4076	0.9445	0.9951
precdrym	-0.4076	1.0000	-0.4750	-0.4456
precseas	0.9445	-0.4750	1.0000	0.9461
precwetm	0.9951	-0.4456	0.9461	1.0000

Scatterplot Matrix



Nonparametric: Spearman's ?

Variable	by Variable	Spearman ?	Prob>|?|
precdrym	prec_ann	-0.6530	<.0001*
precseas	prec_ann	0.9546	<.0001*
precseas	precdrym	-0.7898	<.0001*
precwetm	prec_ann	0.9950	0.0000*
precwetm	precdrym	-0.6781	<.0001*
precwetm	precseas	0.9703	<.0001*



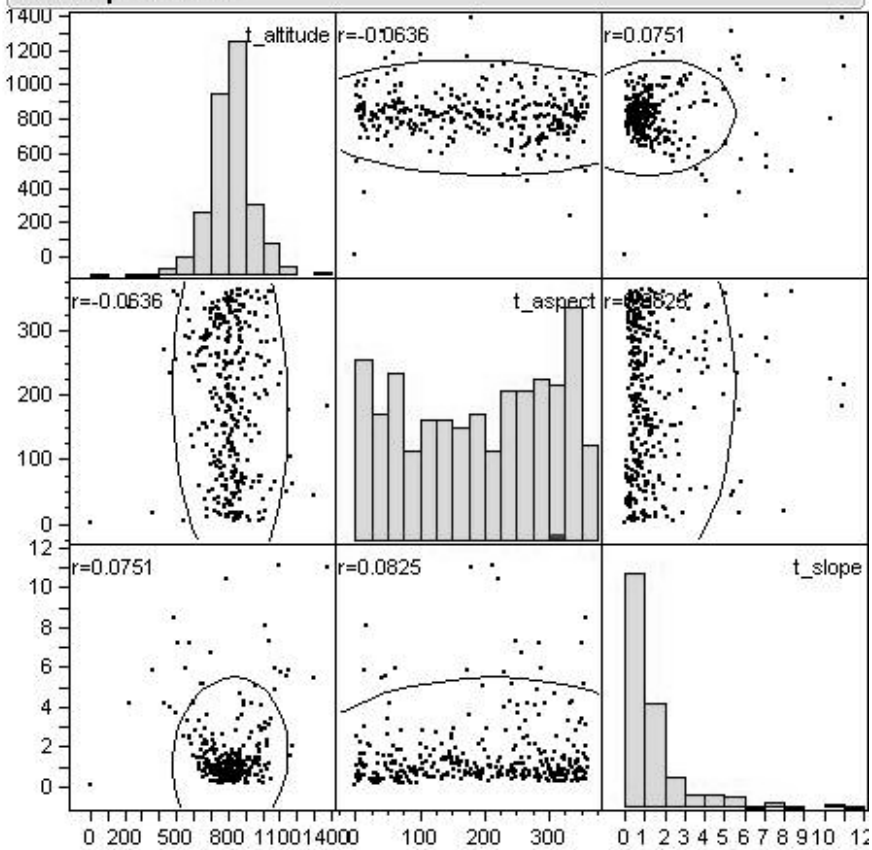
App. Figure 3: Correlation test of parameters about precipitation

Multivariate

Correlations

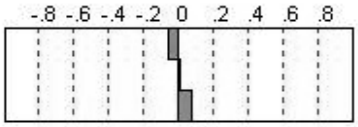
	t_altitude	t_aspect	t_slope
t_altitude	1.0000	-0.0636	0.0751
t_aspect	-0.0636	1.0000	0.0825
t_slope	0.0751	0.0825	1.0000

Scatterplot Matrix



Nonparametric: Spearman's ?

Variable	by Variable	Spearman ?	Prob>|?|
t_aspect	t_altitude	-0.0529	0.3258
t_slope	t_altitude	0.0167	0.7571
t_slope	t_aspect	0.0805	0.1347



App. Figure 4: Correlation test of parameters about terrain