UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ENG. GEOGRÁFICA,
GEOFÍSICA E ENERGIA

PREDICTING OCCURRENCE OF IBERIAN WOLF: THE ROLE OF SAMPLE SIZE AND SPATIAL SCALE

Mariana Gomes Costa Seara

Projecto

MESTRADO EM SISTEMAS DE INFORMAÇÃO GEOGRÁFICA, TECNOLOGIAS E APLICAÇÕES

2013

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ENG. GEOGRÁFICA, GEOFÍSICA E ENERGIA

PREDICTING OCCURRENCE OF IBERIAN WOLF: THE ROLE OF SAMPLE SIZE AND SPATIAL SCALE

Mariana Gomes Costa Seara

Relatório de Projecto

Orientadores: Prof. Drª Cristina Catita, Prof. Drº Francisco Fonseca

MESTRADO EM SISTEMAS DE INFORMAÇÃO GEOGRÁFICA, TECNOLOGIAS E APLICAÇÕES

2013

Agradecimentos

À entidade fundadora do projecto, Fundação da Ciência e Tecnologia, FCT

Ao Professor Francisco Fonseca, pela oportunidade que me deu de participar no projecto

À Professora Cristina Catita, pelos conselhos e palavras de motivação

À Clara Grilo e ao Gonçalo Costa, por toda a ajuda, orientação e paciência que tiveram comigo

À Isabel, por estar sempre disponível para me ajudar

À minha mãe, por me dar a oportunidade de fazer as minhas escolhas

Ao meu namorado, Ricardo, por me apoiar sempre. Nos bons e maus momentos, por me fazer seguir em frente.

Pela informação cedida para o projecto, gostaria ainda agradecer às seguintes entidades: Centro de Biologia Ambiental (CBA), Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Associação de Conservação do Habitat do Lobo Ibérico (ACHLI), Estación Biológica Doñana, Centre for Environmental and Marine Studies (CESAM), Unidade Vida Selvagem e ao Instituto Geográfico Português.

Table of Contents

Table of figures

Tables

Abstract

The Iberian Wolf (*Canis lupus signatus*) is classified in the Portuguese Red Book of Vertebrates as an Endangered Species (EN). Conservation measures for wolf habitat are necessary to prevent further declining of the number of species individuals. The studies that support these measures should, however, integrate spatial considerations, under the penalty of not having an actual positive impact on the species conservation. The main goal of this study is to evaluate the transferability of variables that influence the actual Iberian wolf distribution across three spatial scales (100x100m, 2x2km and 10x10km grids), and to identify the spatial scale that explains better the species presence.

We used data from wolf distribution in Portugal and Spain and from environmental variables to model its potential occurrence at different spatial scales: a 10x10km grid, for the Iberian Peninsula (data collected from both Portugal and Spain); a 2x2km grid and a 100x100m grid for Portugal only. Environmental variables used to assess correlation with wolf presence were divided into landscape (altitude and land use), domestic prey availability (cattle, sheep and goat density), and human disturbance (road density and human population density).

Two distinct methods were used to model potential wolf occurrence: Maxent (Maximum Entropy Model), at the finest resolution, and a Generalized Linear Model, Logistic Regression.

Our results suggest that there should be a compromise between scale and spatial resolution, since, even though all models had high AUC values, the one that was able to extrapolate with the highest correct classification was the model of Portugal at the 2x2km grid.

Regarding the environmental variables, landscape variables had the highest contribution to the models, especially mean altitude, which is supported by previous studies of several authors.

This study demonstrates the high potentialities of Geographic Information Systems for creating biogeographic models for large areas and comparing the importance of spatial parameters used in each model.

Keywords: *Canis lupus*, Iberian Wolf, spatial resolution, Geographic Information Systems, Maximum Entropy Model, Logistic Regression.

Resumo

O Lobo Ibérico (*Canis lupus signatus*) está classificado no Livro Vermelho dos Vertebrados de Portugal como uma espécie Em Perigo. São necessárias medidas de conservação do habitat do Lobo para evitar o progressivo decréscimo da população lupina. Os estudos que sirvam de base para a construção destas medidas devem, contudo, integrar considerações espaciais, sob pena de não terem um impacto positivo real na conservação da espécie.

O objectivo principal deste estudo é avaliar a capacidade de transferir variáveis entre três escalas espaciais diferentes. Mais especificamente, avaliar as variáveis que influenciam a presença de Lobo Ibérico às escalas de 100x100m, 2x2km e 10x10km e identificar qual destas melhor explica a presença da espécie.

Utilizaram-se dados de distribuição de Lobo Ibérico de Portugal e Espanha, e variáveis ambientais para modelar a potencial ocorrência de lobo às três escalas: à quadrícula de 100x100m (dados da região a Norte do rio Douro); de 2x2km (dados de Portugal continental); e 10x10km (dados de toda a Península Ibérica). As variáveis ambientais utilizadas na avaliação da sua correlação com a presença de lobo foram divididas nas categorias de paisagem (altitude e uso do solo), disponibilidade de presas domésticas (gado bovino, ovino e caprino) e perturbação humana (densidade de estradas e densidade populacional humana). Outras variáveis ambientais foram inicialmente testadas, mas descartadas por evidenciarem autocorrelação e/ou fraca correlação com a presença da espécie, como a rugosidade e o declive.

Foram utilizados dois métodos diferentes para modelar a potencial ocorrência de lobo: Maxent (Modelo da Máxima Entropia), à resolução espacial mais fina, de 100x100m; e um Modelo Linear Generalizado, a Regressão Logística, utilizado nos modelos de 2x2km e 10x10km. O primeiro modelo utiliza apenas dados de presença, evitando o problema das pseudo-ausências, ou seja, ausências não confirmadas de facto, no campo;

enquanto para a regressão logística são necessárias presenças e ausências confirmadas.

Os resultados deste estudo sugerem que deve haver um compromisso entre a escala e a resolução espacial uma vez que, apesar de todos os modelos terem tido valores de AUC elevados, aquele que extrapolou com maior classificação correcta foi o modelo de Portugal à quadrícula de 2x2km. Conclui-se também que a amostra utilizada para a modelação de uma espécie generalista como o Lobo Ibérico deverá ter uma boa representatividade de áreas, ou seja, estar completa com dados espacialmente dispersos. Relativamente às variáveis ambientais, as que tiveram maior contributo nos modelos foram as de paisagem, em particular, a altitude média, o que é suportado por outros estudos realizados.

Este estudo demonstra as grandes potencialidades dos Sistemas de Informação Geográfica na criação de modelos biogeográficos em áreas extensas e na comparação da importância dos parâmetros espaciais utilizados em cada modelo.

Palavras-Chave: *Canis lupus*, Lobo Ibérico, resolução espacial, Sistemas de Informação Geográfica, Modelo da Máxima Entropia, Regressão Logística.

Introduction

The Iberian Wolf (*Canis lupus signatus* Cabrera, 1907) is classified in the Portuguese Red Book of Vertebrates as an Endangered Species (Cabral et al., 2006), which means that its survival will be unlikely if the limiting factors continue to exist. In Portugal, the wolf population has been declining since the last century, from South to North and from West to East (Petrucci-Fonseca, 1990, Grilo *et al*., 2002). Habitat decrease and fragmentation, human persecution and decrease in wild and domestic prey are the main causes for the decline in the species population (Petrucci-Fonseca, 1990). Habitat fragmentation generally leads to smaller and more isolated populations which are more vulnerable to local extinction due to stochastic events (Grilo *et al*, 2002). Habitat and, therefore, species conservation measures have to be taken in order to avoid Iberian wolf extinction. In Portugal, in 1990, a law for full wolf protection was published, but it is necessary to build a national conservation and recovery strategy for the Portuguese wolf population (Grilo *et al.,* 2002).

However, it is important that these measures are supported by studies about the species distribution and habitat suitability that take into account factors like the models used and possible spatial scale effects in the analyses.

This study is a part of a project lead by the Geographic Portuguese Institute (IGP), named "*Wildlife corridors: Spatial modeling of human pressure and its usefulness for Iberian Wolf conservation", part of a nine month research grant* financed by the Portuguese Foundation for Science and Technology *(PTDC/AAC-AMB/097511/2008).*

In this project, ecological corridors will be defined in Portugal taking into account the model of human pressure built. The models built for the Iberian wolf will later be used to build ecological corridors that will validate (or not) the ones based on human pressure only (Fig. 1).

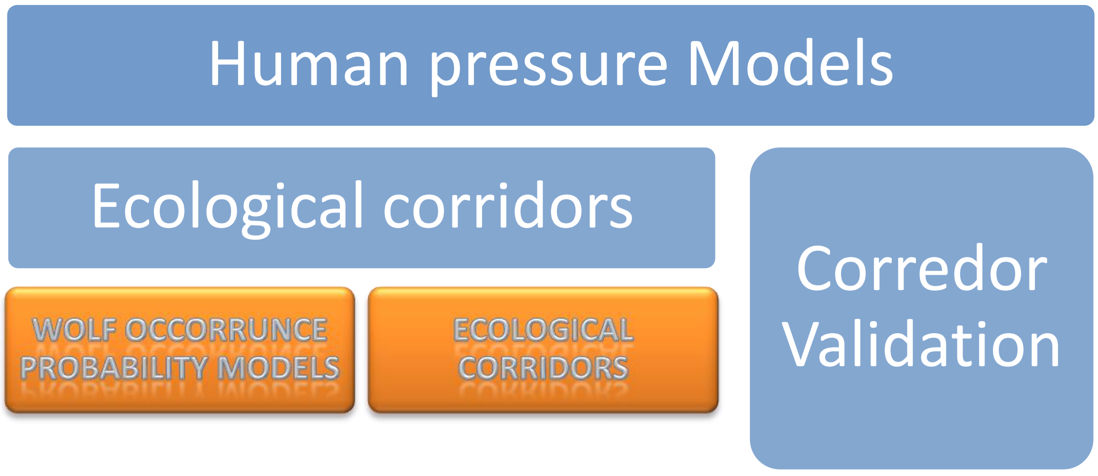

Fig. 1 – Workflow for the project "*Wildlife corridors: Spatial modelling of human pressure and its usefulness for Iberian Wolf conservation*" *(PTDC/AAC-AMB/097511/2008).*

The main goal of this study is to evaluate the transferability of variables and the behavior of models across three spatial scales and three different sample sizes.

The model with the finest spatial resolution and lower sample size was built using the Maximum Entropy Model and its corresponding software, Maxent (Phillips, 2004). This method was chosen because it is known to have good performance with low sample size species presence-data (Kumar & Stohlgren, 2009). The model was built at a 100x100m grid, using wolf presence data in the north of Portugal (Vila Real and Bragança counties), with a total of 94 squares. This model was extrapolated for the entire country.

Two other different models were built, both using logistic regression, but with different spatial resolution and sample sizes: one model was built using a 2x2km grid, with 318 squares of wolf presence data, in North and southern Douro river (Peneda-Gerês, Alvão, Arada/Trancoso); the other model was built using a 10x10km grid, using 953 wolf presence squares in the Iberian Peninsula (both Portugal and Spain). The first model was also extrapolated for the entire country.

All the models were evaluated according their AUC value (Area Under the Curve, which will be explained in further detail in chapter 2 - Study Area, Data, Methods, Models, pp. 11 to 27) and they were then extrapolated to the other two scales and the results compared with wolf presence data. This allowed a comparison of the actual wolf presence data used with the probability of wolf occurrence areas given by each model, in each spatial scale. The logistic regression models were also evaluated according to the respective correct classification.

Though this methodology has been used with many different species (Lleblond *et al*,. 2011, Martin *et al*., 2012), in the course of this project, there had been no similar work with the Iberian Wolf. Furthermore, the data used in the project weren't limited to a single country, but was provided both from Portugal and Spain, a collaboration much needed in order to better understand the true spatial movements of the species.
This document is divided in different chapters:

1 - State of the Art

Summarized presentation of previous studies about habitat suitability for other species using identical methods and data transferability across spatial scales

2 – Data: Study Area, Methods, Models

Ecology of the species and environmental data used, the different spatial resolutions, the two models and how they work.

3 - Results and 4 - Discussion

Results for the three models are shown, and compared with actual wolf presence records. Model extrapolation capacity is evaluated by applying each model to the spatial resolutions of the other models.

5 - Final Remarks:

The main results of this study, critics and suggestions of improvement for future studies.

1 - State of the Art

Several scientific papers have shown the importance of integrating spatial scale and sample size considerations in studies regarding biodiversity conservation (Turner et al., 2001, Wu et al., 1997, Guisan et Thuiller, 2005). When handling a species distribution, it is important to avoid over and underestimating its presence, which will most likely happen, when dealing with only one choice for spatial scale, because there is no such thing as the 'right' scale resolution (Turner *et al*., 2001). Erroneous conclusions may result if scale effects are not considered explicitly in spatial analysis with area-based data (Wu *et al*., 1997).

The choice of an appropriate resolution might depend on the size of the species home range and the way the species uses resources in the landscape (Guisan *et* Thuiller, 2005. The choice of the geographical extent might also depend on a prior knowledge of environmental gradients in the study area (Guisan *et* Thuiller, 2005).

Being a generalist species, the Iberian wolf can and once had a large distribution, being the prey availability the most important factor in individuals or wolf pack establishment (Mech & Boitani, 2003). This means that, even though there might be areas where there are currently no wolves, that doesn't mean those can't constitute suitable habitat for the wolves. However, the distance to these areas can be an impeditive factor for the species to expand there. In this study, we did not consider the distance as a variable, but rather environmental variables only, so as to predict the potential of habitat for the wolf in distant areas from its current habitat.

In a study to assess how local resource selection by the threatened forest-dwelling woodland caribou was influenced by both broad-scale landscape context and local resource heterogeneity in the region of Charlevoix, Canada. Lleblond *et al*. (2011) conclude that landscape context fundamentally constrains the choices available to animals, and that failing to consider landscape context, or arbitrarily choosing an inappropriate scale for measuring covariates, may provide biased inferences with respect to habitat selection patterns.

Two different studies that analyze the environmental variables that influence wolf occurrence, at two different scales, point out similar results: Grilo *et al*. (2002), working with a 2x2km grid cell, concludes that altitude and mixed forest were positively related with Iberian wolf occurrence, whereas the high livestock density was negatively related with wolf occurrence in southern Douro river (the analysis was performed with 108 wolf occurrence); and Llaneza *et al*. (2012), working at a 5x5km grid, concludes that altitude, roughness and refuge strongly determine the Iberian wolf occurrence, followed by human pressure and food availability. In this study, altitude was the main predictor that explained wolf occurrence (the analysis was in this case performed with 267 wolf occurrence).

In a different study Martin *et al*., (2012), in an attempt to make the most of scarce data of the brown bear distribution, in the Pyrenees used two spatial scales to analyze habitat suitability. At a coarse scale, logistic regression was used to develop a habitat suitability model and, at a finer scale, with presence-data only, the authors described the species ecological-niche and then both models were integrated to obtain a more integrative understanding of bear requirements. Both models were consistent: good and suitable habitats predicted at the fine scale were located within source-like habitats predicted by the coarse-scale model. The authors conclude that using local-scale preferences may facilitate the choice between the conservation strategies and management decisions and that the integration of both models at different spatial scales can be most important in making the most of scarce data regarding the species population.

The two different methods for modeling used in the present study - Maxent and Logistic Regression, a particular case of Generalized Linear Models (GLM) - are often used in species habitat suitability studies. These two methods require different data types: while Maxent works with presence-only data, GLM requires presence and absence data.

Philips *et al*. (2006) compared Maxent predictions of two Neotropical mammals (a lowland species of sloth, and a small mountain murid rodent) with those of a commonly used presence-only modeling method, the Genetic Algorithm for Rule-Set Prediction (GARP). The study showed that the area under the ROC curve (AUC) was almost always higher for Maxent, indicating better discrimination of suitable versus unsuitable areas for the species and thus showing that the Maxent modeling approach can be used in its present form for many applications with presence-only datasets.

In a study to acquire the potential spatial distribution of Asiatic black bear and Japanese serow, Doko *et al*. (2008), compared three algorithms: GARP, Maxent, and GLMs. In particular, they concluded that for bear, Maxent was the best algorithm, but GLM has good transferability.

Brotons *et al*. (2004) used breeding bird atlas data in Catalonia as a working example and attempt to analyze the relative performance of two methods: the Ecological Niche factor Analysis (ENFA) using presence data only and Generalized Linear Models (GLM) using presence/absence data. Their results support the idea that GLM predictions are more accurate than those obtained with ENFA, which was particularly true when species were using available habitats proportionally to their suitability, making absence data reliable and useful to enhance model calibration. The authors also conclude that it is difficult to predict generalist species distributions accurately and this is independent of the method used.

In the present study, we attempt to and assess the importance of sample size and spatial resolution in Maxent and Logistic Regression models and also to compare the results obtained.

2 - Data

Study Area
The study area includes both Portugal and Spain (Fig. 2).

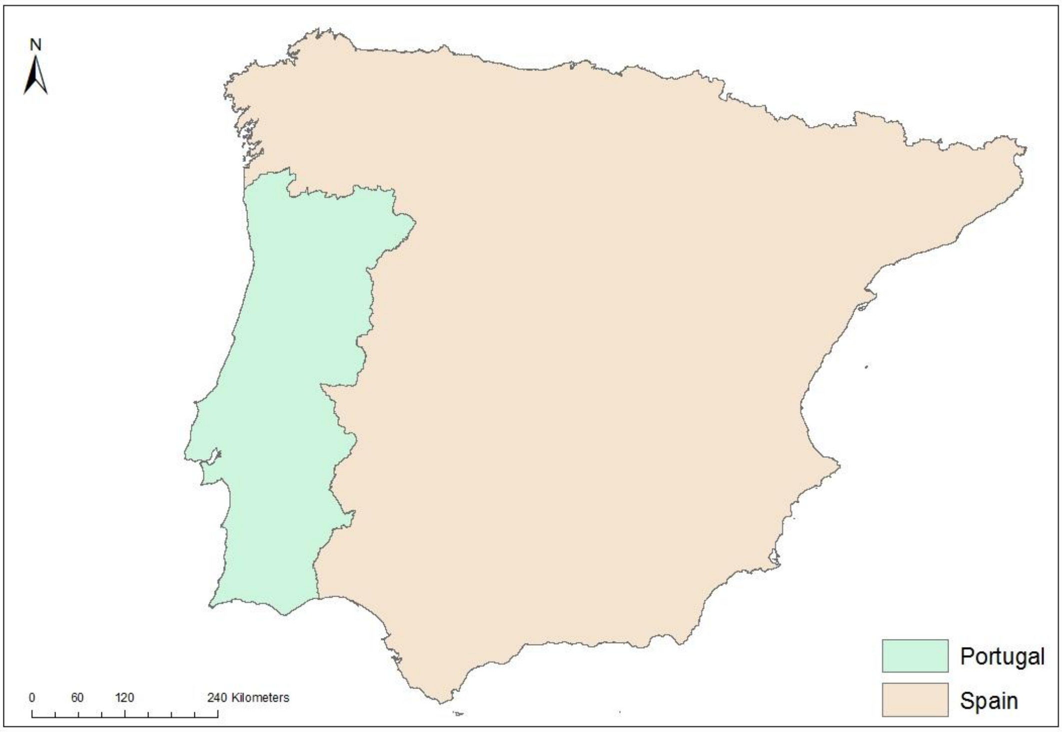

Fig. 2 - Study Area - The Iberian Peninsula.

Portugal:

Portugal is located in the southwest of Europe, and has an area of 92,212 km2. Its climate varies from north to south and east to west but can be overall described as Mediterranean.

Though no wolves have been found at the south of Tejo River in Portugal for over decades, the whole country was included in the models. Though in most areas wolves are extinct, it does not necessarily mean that these don't have potential to be re-colonized by the species (Grilo *et al.*, 2002). In this study, we refer only to continental Portugal, seeing that there is no proof of the existence, present or passed of wolves in the Madeira or the Açores archipelagos.

Spain

Spain is located in the southwestern Europe, in the Iberian Peninsula, being the mainland borderd to the northwest and west by Portugal.

Spain has a total area of 505, 992km2 and it's climate is mainly mediterranean, being the southeastern region considered semiarid. The northern region has an temperate oceanic submediterranean, which differs from the mediterranean climate because it has no drought season.
We will only refer to continental Spain as well, for the same reasons as for Portugal.

In the Iberian Peninsula, the wolf population reached its lowest level in the 1970s, with wolves surviving mainly in the north-west, and later expanding southwards and eastwards (Llaneza *et al*., 2012).

Data

A Geographic Information System (GIS) was used to compile data from wolf occurrence as well as environmental data considered relevant for wolf habitat. A database was built in Arcmap 10.0 (ESRI, 2012).

All the variables' coordinate systems were transformed from the original into different coordinate systems so as to avoid grid disruption:

PT-TM06/ETRS89 – for the 100x100m model
ED50 UTM Zone 29N – for the 2x2km model
ETRS89 UTM zone 30N – for the 10x10km model

Wolf Data

The wolf occurrence data (Table 1) was obtained from different sources: genetic analyses, dead wolves (wolves killed by car hits where not taken into account so as to avoid spatial correlation with road network), photographic traps and breeding places, from different years, since 2005 being the most recent information from 2011. At the 100x100m and 2x2km models, the data from wolf scats gathered were confirmed by genetic analyzes, but at the 10x10km model, the data used did not have this confirmation, which can be a limitation to the sample quality (and, therefore, model quality).

Table 1 – Wolf Data used for each model.

Sample Size	Resolution	Distribution Range	Characteristics	Source
94 squares	100x100m	Northern Portugal (Vila Real and Bragança counties)	Direct observations, scats confirmed with genetics, photos	Grupo Lobo
318	2x2km	North and	Direct observations,	Grupo Lobo

squares		southern Douro river (Peneda-Gerês, Alvão, Arada/Trancoso)	scats confirmed with genetics, camera trapping photos, telemetry data	
953 squares	10x10km	Iberian Peninsula	Observations, scats without confirmation with genetics, camera trapping photos	Pimenta *et al*., 2005. *Wolf's situation in Portugal: 2002/2003 National Census Results* Palomo *et al*. 2007. Atlas y Libro Rojo de los Mamíferos Terrestres de España.

<u>Environmental variables</u>

In order to characterize the study area, several environmental variables were selected according to the wolf's known ecological requirements (Mladenoff *et al*., 1995). Human density and the type of human activities carried out in a given area may be important factors determining the level and the type of human pressure on a wolf population, but landscape attributes may drive this human – wolf interaction by providing protection from humans (Llaneza *et al*., 2012).

We can separate three types of variables: landscape, human presence and domestic prey availability. All of the variables were transformed to different coordinates system in the three spatial scales, in order to prevent major deformations: at 100x100m (Portugal), the chosen coordinate system was the TM06-PT/ETRS89; at the 2x2km (Portugal), the ED50 UTM zone 29N; and at the 10x10km (Iberian Peninsula), the ETRS89/UTM zone 30N.

<u>Landscape</u>

Altitude – altitude was obtained from the Portuguese Environmental Agency, in vectorial format. The contour lines were transformed in a TIN file (using Spatial Analyst, ESRI, 2009) and after that into a raster with altitude information, with a resolution of 100 meters. Mean, maximum and altitude amplitude were calculated.

Slope – Slope was derived from the altitude raster data, using the Spatial Analyst extension from Arcmap (ESRI, 2009), with the same resolution (100 meters). Mean and maximum slope were calculated.

Roughness index – Roughness index was obtained using Jenness Entreprises' DEM tools, which allows to calculate a ratio (surface area / planimetric area) for the land area contained within that cell's boundaries (Jenness, 2009).

Land use – Land use was obtained from CORINE Land cover map (Coordination of Information on the Environment, European Environment Agency, 2006). Land use classes were reclassified in order to best represent the most significant for wolf habitat. The variables chosen after testing were Open Areas and Forest, because both had correlation with wolf presence (Pearson coefficient higher than 0.5) and had no correlation between one and another.

Open areas include pastures, natural grassland, bare rock and sparsely vegetated areas. Forest includes broad leaved, coniferous, mixed forests, moors and heathland, sclerophylous vegetation and transitional woodland/shrub.

Prey availability

Livestock – Livestock data was obtained from the National Institute of Statistics, from both Portugal and Spain, from 2011 and 2009, respectively. Livestock includes only cattle, sheep and goat. The total of individuals of each (cattle, sheep and goat) was divided by the total area (square km) of each parish, thus obtaining the livestock density.

Human pressure

Population density – Total of population was obtained from Portuguese 2011 population census. The smaller administrative Portuguese boundary used was the parish, which area was obtained in square km. Total population was divided for each parish total area, thus obtaining population density.

Road network - For each grid, we calculated the total length of the roads that crossed each square and divided that for the area of each square, obtaining the road density for each. For the xy coordinates, a map of Euclidean distances to the roads was generated, using Arcmap's Spatial Analyst extension (ESRI, 2009).

The data base used is systematized in Fig. 3 and the environmental data used is summarized in Table 2.

As said previously, the same variables were used at all three scales, but at the 2x2km and 10x10km (the logistic regression models), the correlation between each one was tested previously. This is the reason why the variables presented will differ slightly in each model.

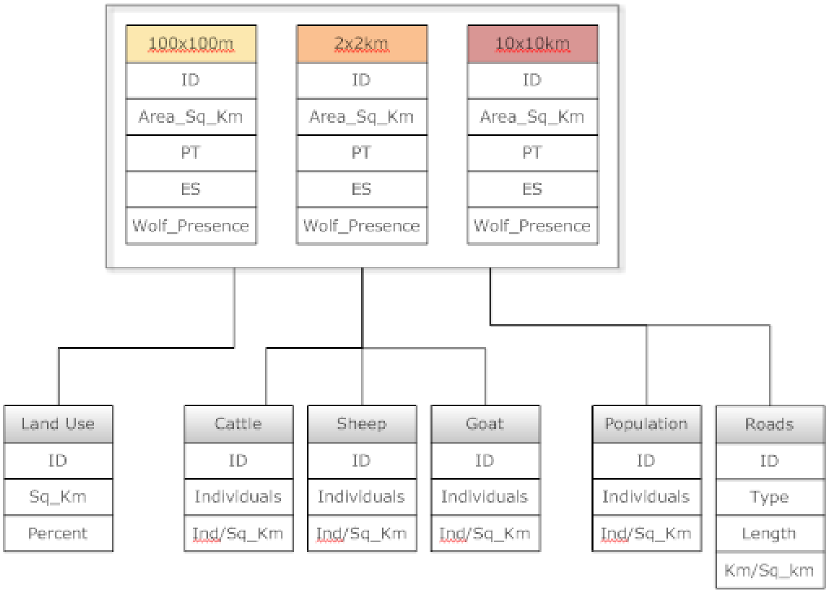

Fig. 3 - Simplified view of the database created for each model.

Table 2 - Data Sources and parameters used

Variables	Source/Year	Resolution	Parameters	Description	GPS coordinates	2x2km squares	10x10 km squares
					Mean (min-max)		
Landscape	Corine Land Cover 2006	25ha	Altitude (m)	Average altitude	0 - 1202	0 - 1784	1. 2649
			Open areas	Pastures, natural grassland, bare rock, sparsely vegetated areas		0 – 100%	0-100%
			Forest	Broad leaved, coniferous, mixed forests, moors and heathland, sclerophylous vegetation, transitional woodland/shrub		0 – 100%	0-100%
Prey	INE		Cattle	nind./km2	0 – 42.23	90 - 4022	0 – 1089
	INE		Sheep	nind./km2	0 – 278	0 - 2004	0 – 1236
	INE		Goat	nind./km2	0 - 38	0-2686	0 - 1053
Humanpressure	INE		Population density	nind./km2	5.4 - 144	0 - 13221	0 - 16243
	IGP		Road density	km/km2	---	0 - 2.92	0 – 4.5
	IGP		Distance to roads*	m	0 - 12610	---	---

Models

To prevent misreading the results of wolf occurrence probability models, we used three different spatial resolutions: 100x100m (for Portugal only),; 2km x 2km, for Portugal only, and10km x 10km in all Iberian Peninsula (data collected from both Portugal and Spain)

Two distinct methods were also used: Maxent (Maximum Entropy Model), at the finest resolution (100x100m) and a Generalized Linear Model, Logistic Regression at 2x2km and 10x10km resolutions.

The same variables were used in all three models. In a previous stage, variables were tested in order to determine which actually contributed for the model and those who didn't were excluded from the analyses. The Spearman's correlation coefficient was calculated for each pair of variables of each group (Landscape, Prey and Human Pressure). Whenever the coefficient was higher than 0.5, both variables were compared in their correlation with wolf occurrence. The variable that had higher correlation coefficient would be selected and the other one, excluded, in order to prevent variable correlation, which could be prejudicial to the model. The excluded variables were slope, roughness, and soil classification such as agriculture or water bodies.

Maximum Entropy Model

The Maximum Entropy model was applied to Portugal, using the GPS coordinates from wolf occurrence. The model was built using Maxent software, according to Steven Philips' tutorial and recommendations (Philips, 2006).

The principle of Maximum Entropy solves real, nontrivial problems in a way that cannot be approached by other statistical methods (Jaynes, 1985).

One of the great problems of species probability of occurrence modeling is that of having records of species presence, but not having confirmed data regarding the species absence. Even though, by knowing the species biology, specialists can presume that there are no individuals in a specific area, the data absence is not confirmed, so we can be dealing with "pseudo-absence" data, and not real data. The main characteristic of the maximum entropy model is that it uses only presence data, to avoid this issue, because the Maxent algorithm does not allow you to assign zero probability to any situation unless your information really rules out that situation. Any other distribution would necessarily either assume information that we do not have, or contradict information that we do have. The problem is that the information is incomplete. The only way known to set up a probability distribution that honestly represents a state of incomplete knowledge is to maximize the entropy, subject to all the information we have (Jaynes, 1985).

Since becoming available in 2004, the Maximum Entropy model has been utilized extensively for modeling species distributions. (Elith *et al*., 2011). Maxent allows making inferences from partial or incomplete information, using the probability distribution which has maximum entropy regarding what is known (Jaynes, 1957; Philips *et. al*, 2006). With Maxent we assume nothing about that which is unknown by, given a collection of facts, choosing a model consistent with all the facts, but otherwise as uniform as possible (Berger, 1996).

The idea of Maxent is to estimate the target 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).

In the maximum entropy approach, we consider the class of all hypotheses $\{H_1 \ldots H_n\}$ consistent with the one data set D_{obs} that was actually observed. Prior information I is also used and represents the knowledge of the possible ways in which Nature could have generated the various H_i. Out of the class C of hypotheses consistent with the data used, the chosen one is the one that is favored by the prior information I (Jaynes, 1985).

Each successive piece of data that is obtained is a new constraint that restricts the possibilities permitted by the previous information gathered (Jaynes, 1985).

In this study, Maxent was applied to presence-only data for the distribution modeling. In this case, the pixels of the study area correspond to the space on which the Maxent probability distribution was defined.

The set of pixels of the study area constitute the space (X) where Maxent's probability distribution ($\boldsymbol{\pi}$) is defined. Pixels with known species occurrence records (x_1, x_2, …x_n

belonging to X) constitute the sample points, and the features (used f_1, ..., f_n) are the environmental variables (Philips *et al*., 2006) and the constraints are that its values coincide with its empirical average. The distribution π assigns a non-negative probability π(x) to each point x, and these probabilities sum to 1 (Philips *et al*., 2006). The goal is to estimate the area of occurrence of a given species, considering that the distribution π coincides with the biologists' concept of the species' potential distribution. (Philips et al., 2004). The purpose is to, from a set of points x_1, x_2, …x_n, chosen independently from an unknown distribution π, build a distribution ^π that is close to π (Ferrão da Costa, 2007).

As a default Maxent randomly samples 10,000 background locations from covariate (environmental variables) grids. Using background data informs the model about the presence or not of the species, the density of covariates in the region, and provides the basis for comparison with the density of covariates occupied by the species. Constraints are imposed so that the solution is one that reflects information from the presence records (Elith *et al*., 2011).

Maxent Model

To build this model, we used a total of 94 wolf GPS localizations in the North of Portugal (Fig. 4).

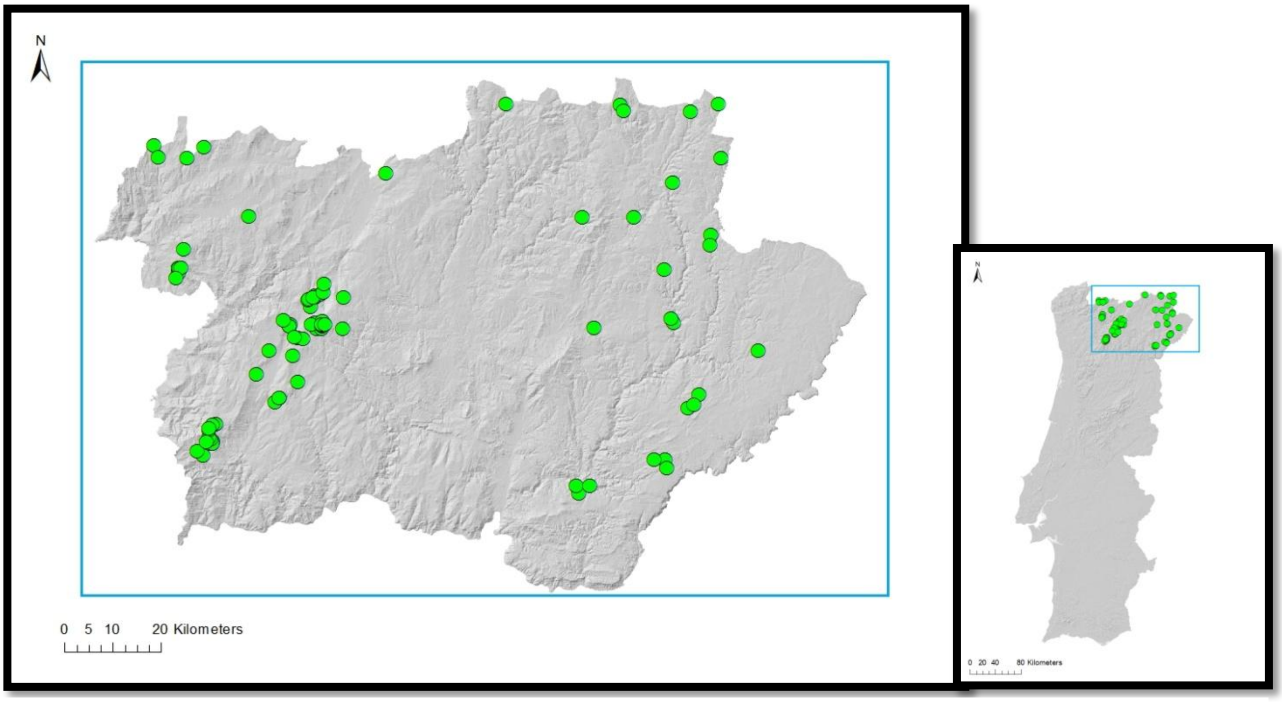

Fig. 4 - Wolf Data used to build the 100x100m model - 94 Wolf GPS coordinates in Bragança and Vila Real, Portugal.

Logistic Regression

The logistic regression is a particular case of Generalized Linear Models (GLM). GLM are often used when presence-absence reliable information is available. These allow establishing correlation between presence, absence and the environmental variables, being able to predict the probability of presence, and therefore, habitat suitability for the species. GLM are mathematical extensions of classic linear models that allow for non linearity and incorporating data with non-gaussian distributions and without constant variance (Ferrão da Costa, 2007).

Unlike classical linear models, which assume a Gaussian (i.e. normal) distribution and an identity link, the distribution of Y in a GLMs may be any of the exponential family distributions (e.g. Gaussian, Poisson or binomial).

When the response variable is binary (i.e. presence/absence), a common approach is to use a generalized linear model, a particular case of multiple regression, with binomial distribution and logistic link: the logistic regression (Hirzen *et al*., 2002).

A binomial GLM is specified with three steps (Zuur *et al*., 2009):

Step 1 - we assume that Yi is binomial distributed, and define the mean and variance of Yi;

Step 2 - the systematic part of the model (a function of the explanatory variables) is specified by the predictor function; and

Step 3 - we define the relationship between the expected value of Yi, πi, and the predictor function η.

This function will map the values of η between 0 and 1. We used the logit link, which assumes that you have approximately an equal number of zeros and ones, which was what we used in this study.

This way, the multiple logistic regression has the following form: π(x) = e gx/(eg(x) +1) where π(x) is the probability of species occurrence and g(x) is given by g(x) = β0 + β1x1 + β2x2 + …. + βpxp, where β0 is a constant and β0, β1, β2 …. βp, are the partial regression coefficients of the x1, x2… xp environmental variables.

As said previously, logistic regression was used at 2x2km and 10x10km grid modeling. We used 80% of the presence data (and the equivalent in absence data) to build the model and the 20% left were used to validate the models.

The 80% sample was selected randomly, using a random feature selection tool box “Random Features”, from the website (Fergunson, C. 2011, [online] Available at: <https://sharepoint.gru.wvu.edu/sites/digital_soils/DSM Tools/ArcGIS Models and Scripts/random_features.zip; or sharepoint.gru.wvu.edu>, [accessed 21 April 2012]).

At the 2x2km grid, the sample corresponds to 318 grid-cells (in a total of 398 cells with presence data), of the total grid-cells (22761), (Fig. 5).

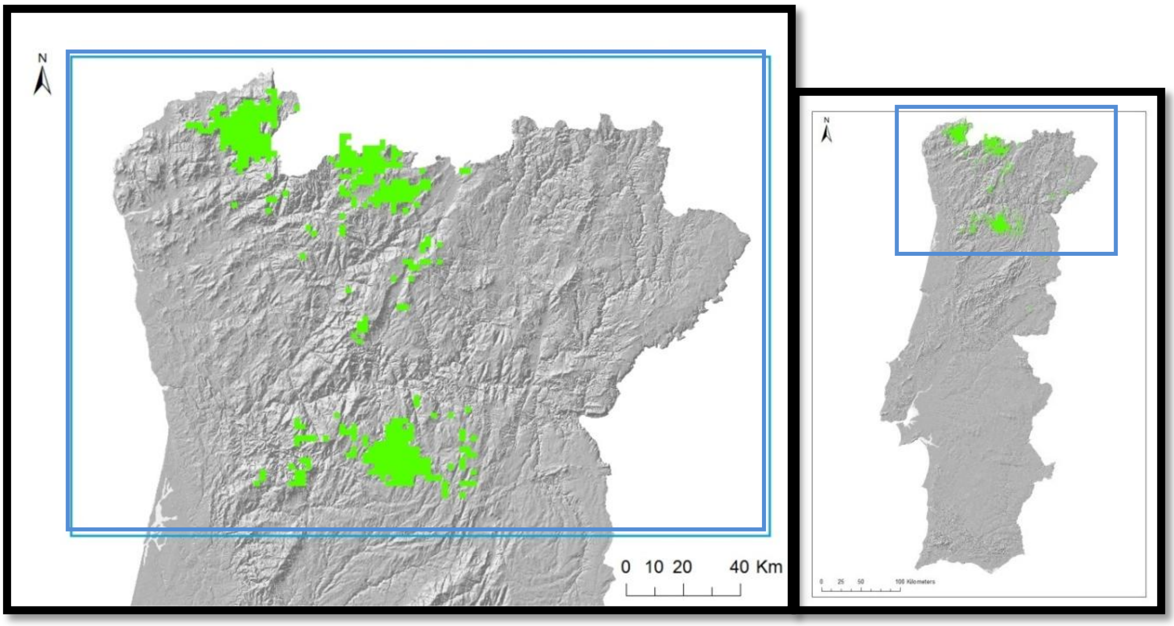

Fig. 5 - Wolf data used to build the 2x2km model - 398 grid cells of wolf presence both North and South of Douro River, Portugal.

At the 10x10km grid, the number of grid-cells was 953 (in a total of 1192 cells with presence data), of a total of 1283 grid-cells (Fig. 6). We used only the continuous distribution of the northwest of the Iberian Peninsula, because those data are more reliable than the ones from the South at Sierra Morena, in Spain.

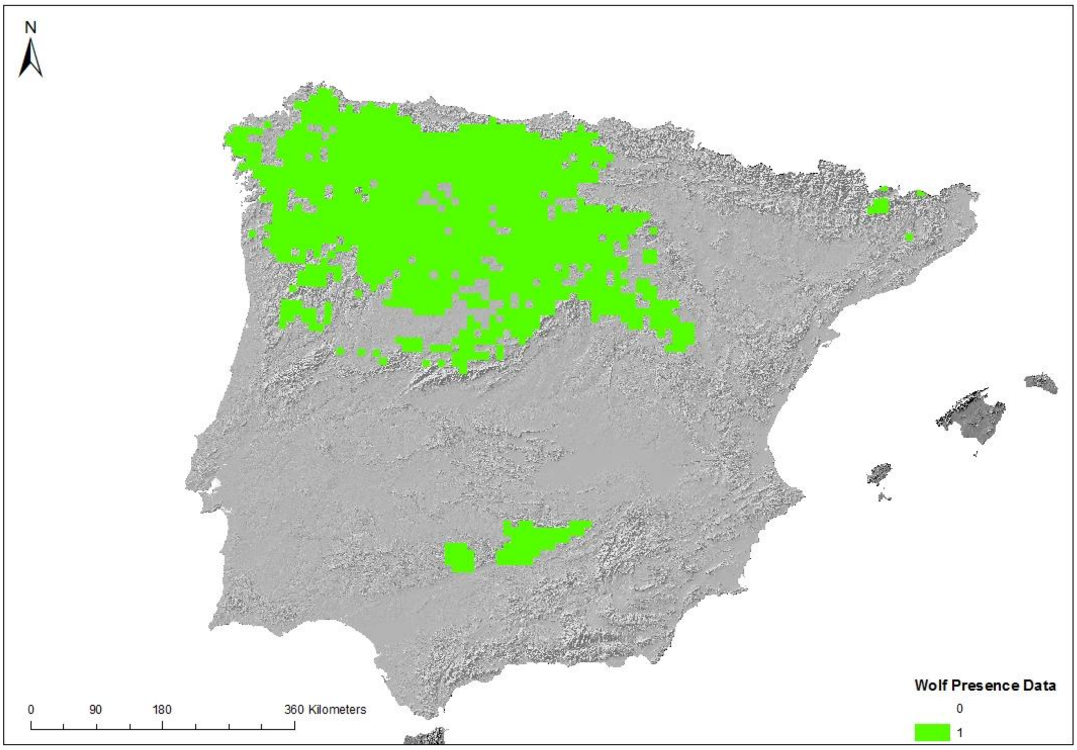

Fig. 6 - Wolf data used to build the 10x10km model - 1192 grid cells of wolf presence in the Iberian Peninsula.

In the selection of the ‘absence cells’, we took in consideration a distance from the original presence data cells. Our goal was to understand why certain areas are currently occupied by wolf and why close areas to those aren’t. We applied a 100km buffer to the known wolf distribution in each of the models and selected the absence cells randomly from that area. We did not want to include cells from distant places, which nevertheless, could also be reliable absence places, because then we would be including another variable in the model: the distance. The selected cells are shown in figs. 7 and 8.

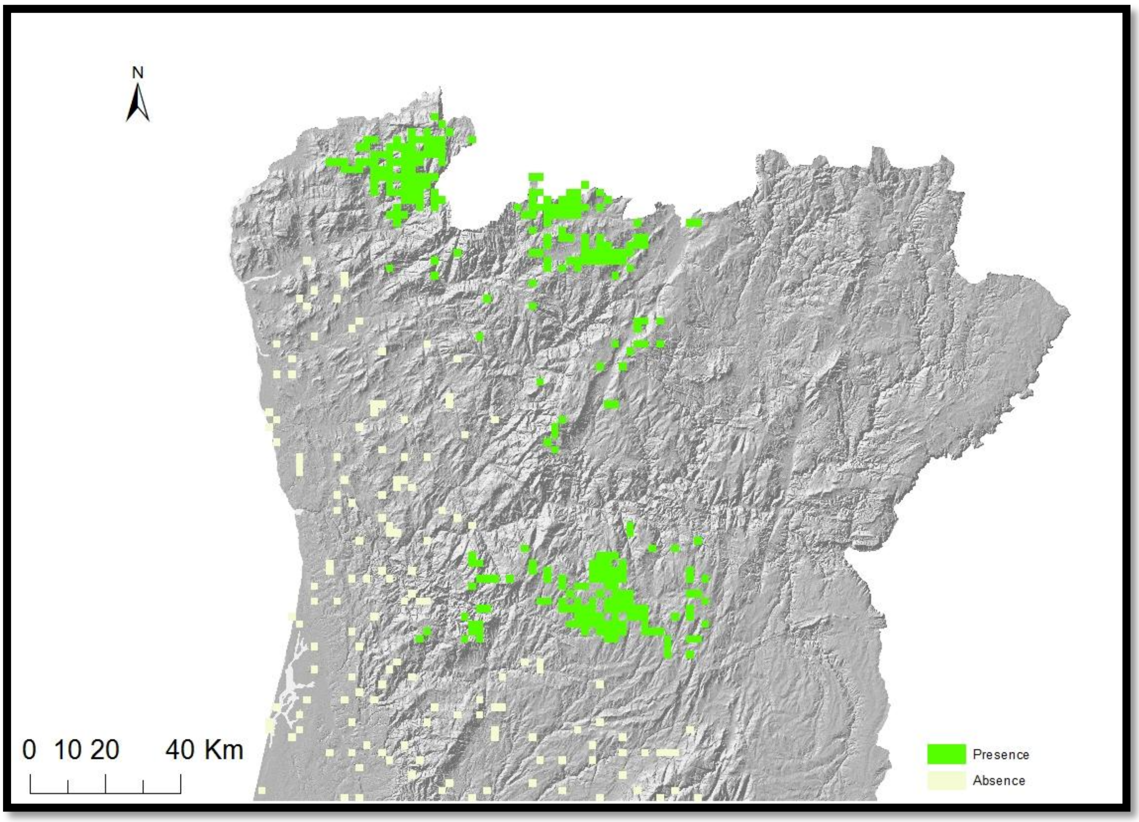

Fig. 7 - Wolf presence and absence cells used to build the 2x2km model.

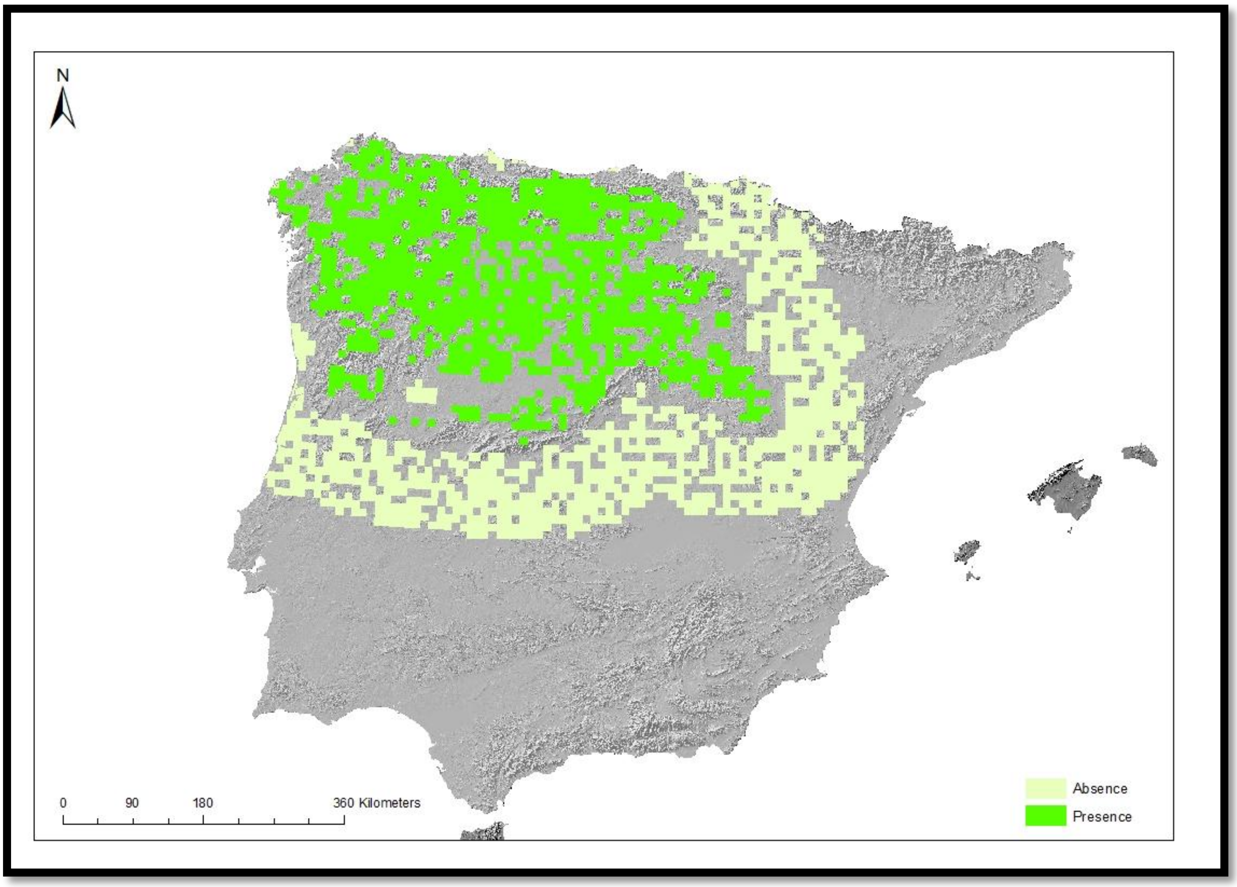

Fig. 8- Wolf presence and absence cells used in the 10x10km model.

The software used for the logistic regression analyses was R statistics 2.15.1 (R Foundation, 2004).

We first analyzed the correlation for each variable using Spearman's rank correlation coefficient. The threshold chosen for the selection of variables was +-0.5. The variables that showed high correlation were compared as to each one's correlation with wolf presence. The ones that had a higher coefficient were chosen and discarded the ones with the lowest coefficient.

Maximum altitude, altitude range, mean slope, maximum slope, slope range and the roughness index were discarded, as well as other soil classification other than open areas and forest for having lower correlation with wolf presence.

Then, the models were tested with several combinations of the variables chosen as well as the quadratic function for all variables (in tables 3 and 5 referred as "Open Areas 2", for example).

The model chosen was the one with the lowest AIC – Akaike Information Criteria (Sakamoto *et al*, 1986) – which is a measure of the relative goodness of fit of a statistical model (the lower the AIC value, the better the model).

3 - Results

In Tables 3 and 5 show all the variable’s combinations tested for each scale (2x2km and 10x10km). The combinations chosen for each scale were the ones with lower AIC value (Tables 4 and 6).

Table3 - Candidate 2x2km Models

Models	AIC	Deviance	ΔAIC	wi	AICwi	Dispersion
Landscape						
Alt_m	535.690	531.7	268.2	0.00	0.00	0.83862776
Forest	863.560	859.6	596.1	0.00	0.00	1.355772871
OpenAreas	708.150	704.2	440.7	0.00	0.00	1.110646688
Alt_m+Alt_m2	502.030	496.0	234.6	0.00	0.00	0.783617694
Forest+Forest2	828.610	822.6	561.1	0.00	0.00	1.299541864
OpenAreas+OpenAreas2	651.620	645.6	384.2	0.00	0.00	1.019936809
Alt_m+Floresta	488.050	482.1	220.6	0.00	0.00	0.761532385
Alt_m+OpenAreas	496.290	490.3	228.8	0.00	0.00	0.774549763
Floresta+OpenAreas	710.140	704.1	442.7	0.00	0.00	1.112385466
Alt_m+Floresta+OpenAreas	475.810	467.8	208.3	0.00	0.00	0.740205696
Alt_m+Alt_m2+Forest	426.680	418.7	159.2	0.00	0.00	0.662468354
Alt_m+Altm2+Forest+Forest2	420.500	410.5	153.0	0.00	0.00	
Alt_m+Alt_m2+Forest+ OpenAreas+OpenAreas2	391.390	379.4	123.9	0.00	0.00	0.602206349
Alt_m+Alt_m2+Forest+Forest2+ OpenAreas+OpenAreas2	391.730	377.7	124.3	0.00	0.00	0.600524642
Human						
Roads	853.850	849.9	586.4	0.00	0.00	1.340457413
Population_dens	**653.770**	649.8	386.3	0.00	0.00	1.024873817
Roads+Roads2	854.870	848.9	587.4	0.00	0.00	1.341026856
Population_dens+Population_dens2	649.260	643.3	381.8	0.00	0.00	1.016208531
Roads+Population_dens	655.190	649.2	387.7	0.00	0.00	1.025576619
Roads:Population_dens	713.710	709.7	446.2	0.00	0.00	1.119416404
Prey						
Cattle	883.490	879.5	616.0	0.00	0.00	1.387208202
Sheep	884.810	880.8	617.3	0.00	0.00	1.389290221
Goat	862.130	858.1	594.7	0.00	0.00	1.35351735
Cattle+Sheep	884.820	878.8	617.4	0.00	0.00	1.388341232
Cattle+Goat	863.070	857.1	595.6	0.00	0.00	1.353981043
Sheep+Goat	863.490	857.5	596.0	0.00	0.00	1.35464455
Cattle+Sheep+Goat	864.560	856.6	597.1	0.00	0.00	1.355316456
Cattle+Cattle2	756.960	751.0	489.5	0.00	0.00	1.186350711
Sheep+Sheep2	845.050	839.1	577.6	0.00	0.00	1.325513428
Goat+Goat2	861.900	855.9	594.4	0.00	0.00	1.352132701
Cattle+Cattle2+Sheep	758.940	750.9	491.5	0.00	0.00	1.188196203
Cattle+Cattle2+Sheep+Sheep2	725.300	715.3	457.8	0.00	0.00	1.133597464
Cattle+Cattle2+Goat	739.680	731.7	472.2	0.00	0.00	1.157721519
Cattle+Cattle2+Goat+Goat2	740.210	730.2	472.7	0.00	0.00	**1.157226624**
Sheep+Sheep2+Cattle	846.330	838.3	578.9	0.00	0.00	1.326471519
Sheep+Sheep2+Goat	828.010	**820.0**	560.5	0.00	0.00	1.297484177
Sheep+Sheep2+Goat+Goat2	827.310	817.3	559.8	0.00	0.00	1.29526149
Goat+Goat2+Cattle	862.580	854.58	595.1	0.00	0.00	1.352183544
Goat+Goat2+Sheep	863.100	855.1	595.6	0.00	0.00	1.353006329
Cattle+Sheep+Sheep2+Goat+Goat2	828.970	816.97	561.5	0.00	0.00	1.296777778
Sheep+Cattle+Cattle2+Goat+Goat2	742.190	730.19	474.7	0.00	0.00	1.159031746
Goat+Cattle+Cattle2+Sheep+ Sheep2	**709.64**	697.64	442.2	0.00	0.00	1.107365079
Goat+Goat2+Cattle+Cattle2+Sheep+Sheep2	710.07	696.07	442.6	0.00		1.106629571
Cattle+Sheep+Goat+Goat2	863.92	853.92	596.5	0.00	0.00	1.353280507
Cattle+Goat+Sheep+Sheep2	829.8	819.8	562.3	0.00	0.00	1.299207607
Sheep+Goat+Cattle+Cattle2	741.64	731.64	474.2	0.00	0.00	1.159492868
Landscape+ Prey						**0**
Alt_m+Alt_m2+Forest+OpenAreas+OpenAreas2+Goat+Cattle+Cattle2+Sheep+Sheep2	321.29	299.29			0.00	0.478864
Human + Prey						**0**
Population_dens+Goat+Cattle+Cattle2+Sheep+Sheep2	485.7	471.7			0.00	0.749920509
Landscape+ Human + prey						
Alt_m+Alt_m2+ Forest + OpenAreas+OpenAreas2 + Population_dens + Cattle+Cattle2+Sheep+Sheep2	**267.47**	245.47	0.0	1.00	1.00	0.392752

2x2km Best Model

The parameters for the best model chosen are showed in Table 4. The *p*-value was always lower that 0.02 and the standard error did not surpass the value of the estimate.

Table 4 - Best 2x2km Model and each variable's parameter

Variable	Estimate	Std Error	z Value	*p*
intercept	-5,72	0,988500	-5,78	<0.001
Altitude	0,02	0,002531	7,62	<0.001
Altitude2	0,00	0,000002	-5,93	<0.001
Open Areas	6,99	2,483000	2,81	0,005
Open Areas2	-7,29	3,083000	-2,36	0,018
Forest	-3,95	0,778600	-5,07	<0.001
Population density	-0,03	0,006012	-4,80	<0.001
Cattle	0,39	0,052050	7,42	<0.001
Cattle2	-0,01	0,001231	-5,92	<0.001
Sheep	0,05	0,029200	1,57	0,117
Sheep2	0,00	0,000460	-2,92	0,004

Table 5 - Candidates 10x10km Models

Models	AIC	Deviance	ΔAIC	wi	AICwi	Dispersion
Landscape						
Alt_m	2353.400	2349.4	694.5	0.00	0.00	1.233928571
Forest	2642.900	2638.9	984.0	0.00	0.00	1.385976891
OpenAreas	2606.100	2602.1	947.2	0.00	0.00	1.36664916
Alt_m+Forest	2344.200	2338.2	685.3	0.00	0.00	1.22869154
Alt_m+OpenAreas	2329.900	2323.9	671.0	0.00	0.00	1.221177089
Forest+OpenAreas	2604.200	2598.2	945.3	0.00	0.00	1.365317919
Alt_m+Forest+OpenAreas	2321.2	2313.2	662.3	0.00	0.00	1.216193481
Alt_m+Alt_m2	2303.600	2297.6	644.7	0.00	0.00	1.207356805
Forest+Forest2	2632.200	2626.2	973.3	0.00	0.00	1.380031529
OpenAreas+OpenAreas2	2581.900	2575.9	923.0	0.00	0.00	1.35359958
Alt_m+Alt_m2+Forest	2301.200	2293.2	642.3	0.00	0.00	1.205678233
Alt_m+Altm2+Forest+Forest2	2274.500	2264.5	615.6	0.00	0.00	1.19121515
Alt_m+Alt_m2+Forest+OpenAreas+OpenAreas2	2238.000	2226.0	579.1	0.00	0.00	1.171578947
Alt_m+Alt_m2+Forest+Forest2+ OpenAreas+OpenAreas2	**2230.000**	2216.0	571.1	0.00	0.00	1.166929963
Human						
Roads	2600.800	2596.8	941.9	0.00	0.00	1.363865546
PopulationDens	2568.500	2564.5	909.6	0.00	0.00	1.346901261
Roads:PopulationDens	2599.800	2595.8	940.9	0.00	0.00	1.363340336
Roads+PopulationDens	**2440.000**	**2434.0**	781.1	0.00	0.00	**1.279033106**
Prey						**0**
Cattle	2646.200	2642.2	987.3	0.00	0.00	1.387710084
Sheep	2599.300	2595.3	940.4	0.00	0.00	1.363077731
Goat	2562.800	2558.8	903.9	0.00	0.00	1.343907563
Cattle+Sheep	2589.100	2583.1	930.2	0.00	0.00	1.357383079
Cattle+Goat	2533.200	2527.2	874.3	0.00	0.00	1.328008408
Sheep+Goat	2546.500	2540.5	887.6	0.00	0.00	1.334997373
Cattle+Sheep+Goat	2519.300	2511.3	860.4	0.00	0.00	1.320347003
Cattle+Cattle2	2560.100	2554.1	901.2	0.00	0.00	1.342143983
Sheep+Sheep2	2599.100	2593.1	940.2	0.00	0.00	1.36263794
Goat+Goat2	2564.400	2558.4	905.5	0.00	0.00	1.344403573
Cattle+Cattle2+Sheep	2510.700	2502.7	851.8	0.00	0.00	1.315825447
Cattle+Cattle2+Sheep+Sheep2	2497.600	2487.6	838.7	0.00	0.00	1.308574435
Cattle+Cattle2+Goat	2453.300	2445.3	794.4	0.00	0.00	1.285646688
Cattle+Cattle2+Goat+Goat2	2452.300	2442.3	793.4	0.00	0.00	1.284744871
Sheep+Sheep2+Cattle	2591.100	2583.1	932.2	0.00	0.00	1.35809674
Sheep+Sheep2+Goat	2533.100	2525.1	874.2	0.00	0.00	1.327602524
Sheep+Sheep2+Goat+Goat2	2535.100	2525.1	876.2	0.00	0.00	1.328300894
Goat+Goat2+Cattle	2534.800	2526.8	875.9	0.00	0.00	1.32849632
Goat+Goat2+Sheep	2548.200	2540.2	889.3	0.00	0.00	1.335541535
Cattle+Sheep+Sheep2+Goat+Goat2	2511.100	2499.1	852.2	0.00	0.00	1.315315789
Sheep+Cattle+Cattle2+Goat+Goat2	2428.400	2416.4	769.5	0.00	0.00	1.271789474
Goat+Cattle+Cattle2+Sheep+Sheep2	2418.2	2406.2	759.3	0.00	0.00	1.266421053
Cattle+Sheep+Goat+Goat2	2520.9	2510.9	862.0	0.00	0.00	1.320831142
Cattle+Goat+Sheep+Sheep2	2509.2	2499.2	850.3	0.00	0.00	1.314676486
Sheep+Goat+Cattle+Cattle2	2429.5	2419.5	770.6	0.00	0.00	1.272751184
Goat+Goat2+Cattle+Cattle2+Sheep+Sheep2	**2417.4**	**2403.4**	758.5	0.00	0.00	**1.265613481**
Landscape+ Prey						**0**
Alt_m+Alt_m2+Forest+Forest2+ OpenAreas+OpenAreas2 +Goat+Goat2+Cattle+Cattle2+Sheep+Sheep2	1860.9	1834.9	202.0	0.00	0.00	0.968796199
Human + Prey						**0**
Roads+PopulationDens+Goat+Goat2+Cattle+Cattle2+Sheep+Sheep2	**2233.2**	**2.22E+03**	574.3	0.00	0.00	**1.17E+00**
Landscape+ Human + prey						**0**
Alt_m+Alt_m2+Forest+ OpenAreas+OpenAreas2 +Roads+pop_dens+ Goat+Goat2+Cattle+Cattle2+ Sheep+Sheep2	**1658.9**	**1630.9**	**0.0**	**1.00**	**1.00**	**0.861997886**

10x10 km Best Model

Table 6 shows the parameters for the best model chosen. The *p*-value was always lower that 0.002 and the standard error did not surpass the value of the estimate.

Table 6 - Best 10x10km Model and each variable's parameter

Variable	**Estimate**	**Std Error**	**z Value**	***p***
Intercept	-7,636	0,534	-14,30	< 0.001
Altitude	0,013	0,001	13,60	< 0.001
Altitude2	0,000	0,000	-9,91	< 0.001
Forest	-1,326	0,255	-5,20	< 0.001
Open Areas	5,546	1,289	4,30	< 0.001
Open Areas2	-7,550	2,205	-3,43	< 0.001
Road density	4,497	0,388	11,58	< 0.001
Population density	-0,005	0,001	-5,09	< 0.001
Goat	-0,192	0,024	-8,03	< 0.001
Goat2	0,001	0,000	8,35	< 0.001
Cattle	0,131	0,012	11,24	< 0.001
Cattle2	-0,001	0,000	-8,68	< 0.001
Sheep	-0,021	0,004	-5,21	< 0.001
Sheep2	0,000	0,000	3,15	0,002

All three models have high AUC values (Table 7). The Area Under the Curve (AUC) corresponds to the area under the receiver operating characteristic (ROC), or ROC curve, which plots the true positives (sensitivity) vs. false positives (specificity), for a binary classifier system as its discrimination threshold is varied (Tuszynski, 2004). More specifically, the AUC represents the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

This means that, in all three models, there was a good presence discrimination. In the logistic regression models, the percentage of correct classification and the validation (with the 20% sample) was also high.

Table 7 - Comparison between the three models used, with different spatial resolution and sample sizes

	AUC	Correct Classification	Validation
100x100m grid	0.93	--	--
2x2km grid	0.97	92%	93%
10x10km grid	0.88	81%	76%

Maximum Entropy Model (100x100m)

The model built with the GPS coordinates and Maxent software shows a higher concentration of areas with wolf occurrence probability in the North of Portugal. The threshold used to define the most suitable areas for the wolf corresponds to the one that correctly classifies 90% of the presence data used, which is 0.18 (Fig.).

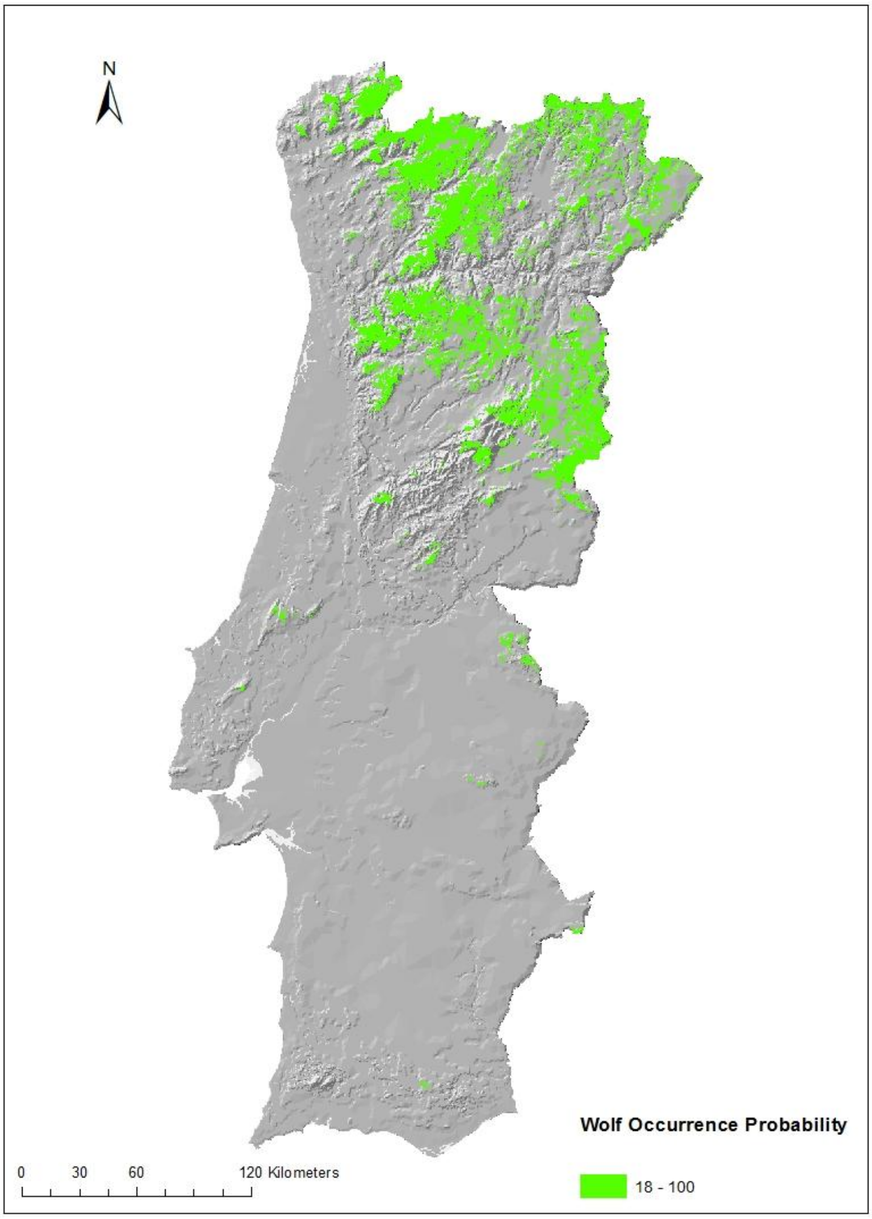

Fig. 9 - Wolf Occurrence Probability, results of the 100x100m model (Maximum Entropy).

In this model, all variables were used, and Table 8 shows the percentage of contribution of each one to the model.

Table 8 - Percentage of contribution of each variable to the 100x100m model

Variable	% Contribution
Altitude	70%
Forest	7%
Open Areas	1%
Cattle	11%
Goat	1%
Sheep	2%
Road Distance	6%
Population Density	2%

Mean altitude was the variable that has the highest contribution (73%). Cattle (11%) and forest areas (8%) are the ones with the highest values after altitude. Open areas was the variable with the smallest contribution to the model (only 1.2%).

Maxent software produces a graphic that allows to understand which of the environmental variable has the highest gain, i.e., which appears to have the most useful information by itself and also which of the variables has the most information that isn't present in the other variables (the one that, when omitted, decreases the gain) (Fig. 10).

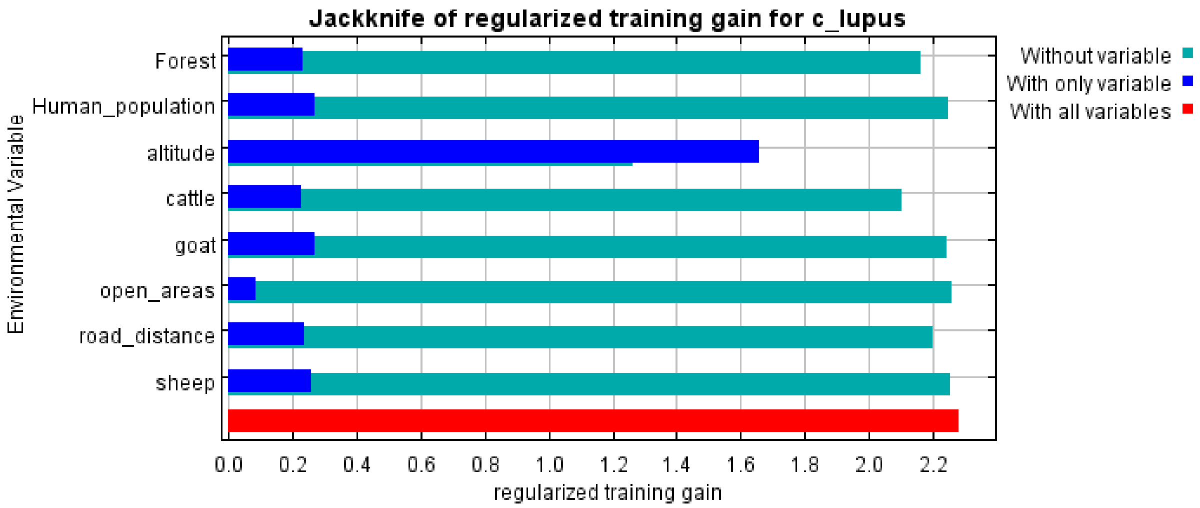

Fig. 10 - Gain of each environmental variable used in the 100x100m model.

Mean altitude is the variable that has the most gain when used isolated. Open areas have the least gain, and it is the one that less influences the model: the gain without this variable is one of the highest. The same happens when sheep, goat or population density, are not used. This means that these variables are the ones that add less information to the model, when compared to the others.

Logistic Regression Model: 2x2km grid

The model at the 2x2km scale was the one with the higher AUC value. It also had higher values of correct classification and validation than the model at the 10x10km grid**.**

The model suggests several areas along the eastern half part of Portugal which constitute suitable habitats for the Iberian wolf especially concentrated in the North of Portugal (Fig.). Nevertheless, there are also several areas with a probability above 75% of wolf occurrence in Alentejo, southern of river Tejo.

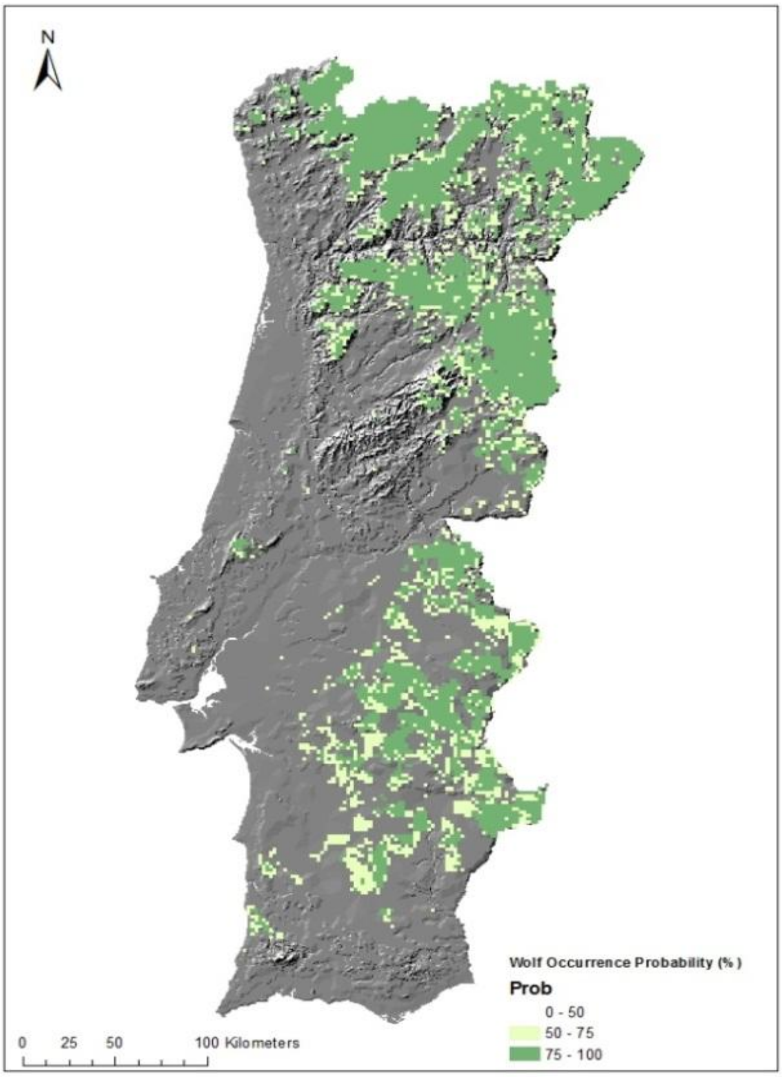

Fig. 11- Wolf occurrence probability. Results of the 2x2km model (Logistic Regression).

The coefficients (β) of the logistic regression equation are shown in Table 9.

Table 9 - - Estimate, Standard Error, Z value and P value for the environmental variables used in the 10x10km model.

Variable	**Estimate**	**Std Error**	**zValue**	***p***
Intercept	-5.72	0.99	-5.78	<0.001
Altitude	0.02	0.00	7.62	<0.001
Altitude^2	0.00	0.00	-5.93	<0.001
OpenAreas	6.99	2.48	2.81	0.005
Open Areas^2	-7.29	3.08	-2.36	0.018
Forest	-3.95	0.78	-5.07	<0.001
Populationdensity	-0.03	0.01	-4.80	<0.001
Cattle	0.39	0.05	7.42	<0.001
Cattle^2	-0.01	0.00	-5.92	<0.001
Sheep	0.05	0.03	1.57	0.117
Sheep^2	0.00	0.00	-2.92	0.004

The variables “goat” and “road density” were not included in this model because they had a higher Standard Error than the Estimate, which means we wouldn’t be able to determine if its relation with wolf occurrence is positive or negative.

From the other variables, we used the quadratic function for the Altitude and Open Areas, which were the ones that made the AIC value of the model decrease.

Logistic Regression Model: 10x10km grid

The model with the lowest spatial resolution was the model with the least AUC value and also less extrapolation accuracy.

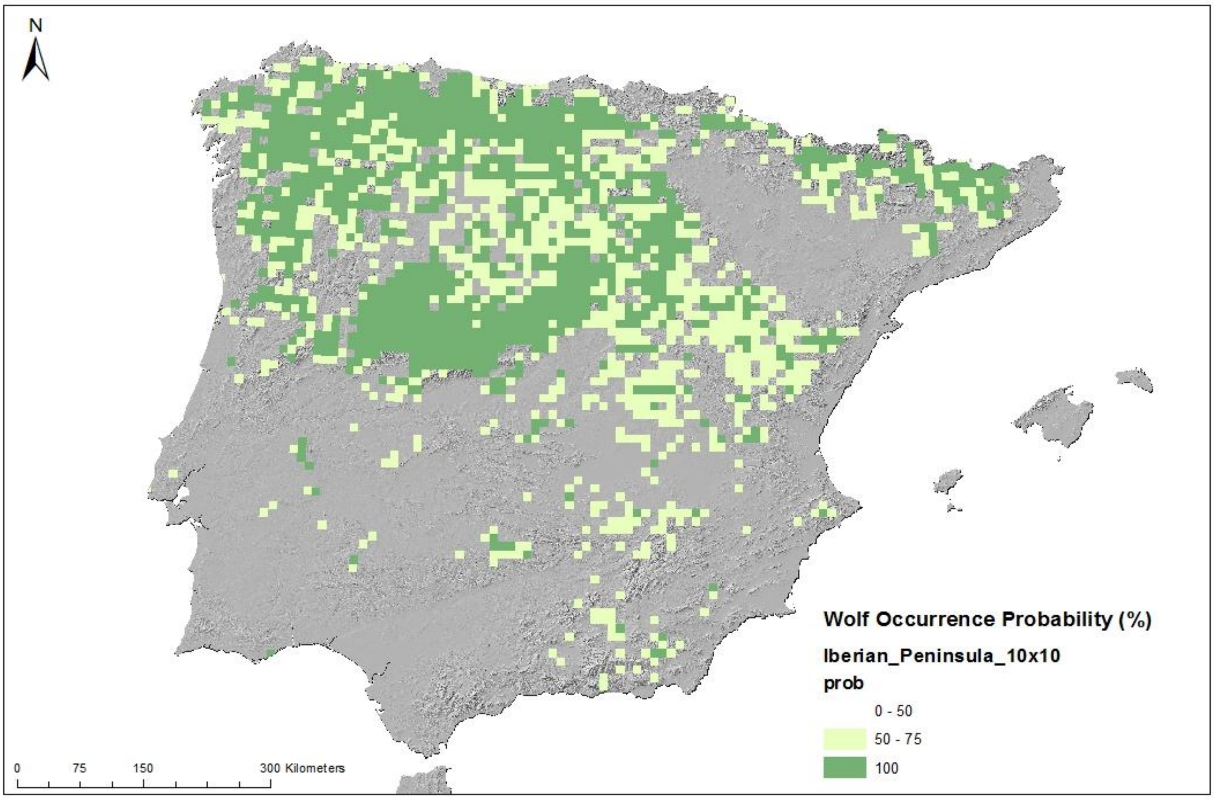

Fig. 12 - Wolf Occurrence Probability - Results of the 10x10km model (Logistic Regression).

The coefficients (β) of the logistic regression equation are shown in Table 10.

Table 10 - Estimate, Standard Error, Z value and P value for the environmental variables used in the 10x10km model

Variable	Estimate	Std Error	zValue	*p*
Intercept	-7.636	0.534	-14.30	< 0.001
Altitude	0.013	0.001	13.60	< 0.001
Altitude2	0.000	0.000	-9.91	< 0.001
Forest	-1.326	0.255	-5.20	< 0.001
OpenAreas	5.546	1.289	4.30	< 0.001
Open Areas2	-7.550	2.205	-3.43	< 0.001
Roaddensity	4.497	0.388	11.58	< 0.001
Populationdensity	-0.005	0.001	-5.09	< 0.001
Goat	-0.192	0.024	-8.03	< 0.001
Goat2	0.001	0.000	8.35	< 0.001
Cattle	0.131	0.012	11.24	< 0.001
Cattle2	-0.001	0.000	-8.68	< 0.001
Sheep	-0.021	0.004	-5.21	< 0.001
Sheep2	0.000	0.000	3.15	0.002

Surprisingly, road density has, in this model, a positive estimate, which means that it's correlation with wolf occurrence probability would be positive. This can be explained by the high road density in the area where wolf presence data was selected. In Galicia there are high levels of road density (Fig. 13). In a study of factors that can influence wolf occurrence in Galicia, Llaneza *et al*. (2012), states that wolves occur in Galicia in areas with remarkably high densities of paved roads.

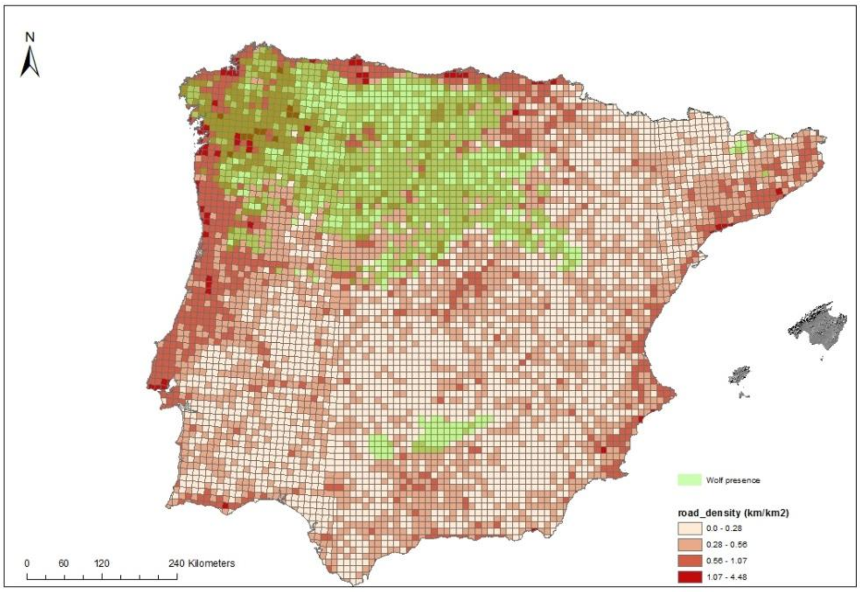

Fig. 13- Road Density in the Iberian Peninsula, 10x10km grid (km/km2).

4 - Discussion

The relation between wolf occurrence probability and each environmental variable, for all three models, are shown in Fig. 14 to 17.

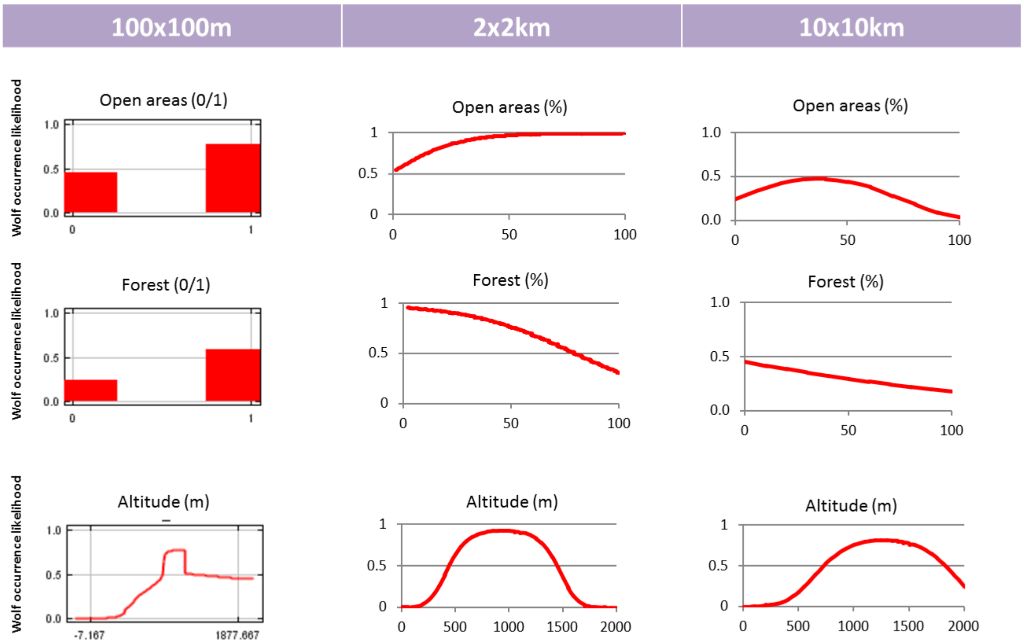

Fig.7 - Relation between wolf occurrence probability and each environmental variable of the Landscape group used, at all three scales.

In the Landscape variables, we can conclude that both open areas and forest areas are important for Wolf occurrence. Forest areas alone are not as attractive to wolf, but we can say that the complementarity between both kind of areas is important to satisfy wolves' needs for hunting and refuge.

Altitude has a similar behavior in all three scales and does not have a linear and positive correlation with wolf occurrence, possible due to wolf absence in Serra da Estrela (Portugal) and the mountain mass in midland Spain (Serra de Gredos).

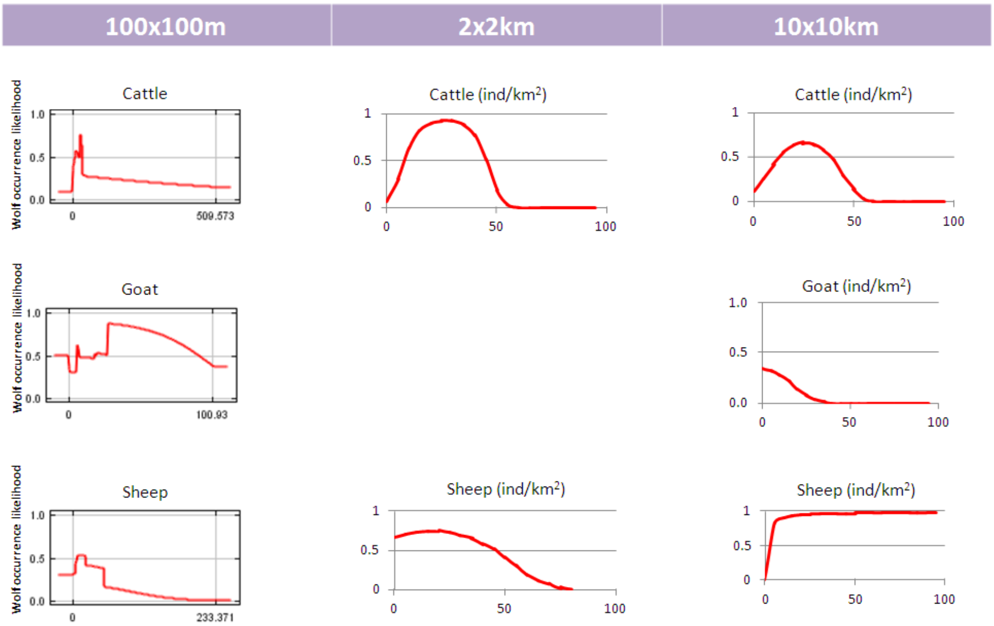

Fig. 8 - Relation between wolf occurrence probability and each environmental variable of the Prey group used, at all three scales.

In the prey group, cattle density has a similar behavior in all three scales. Surprisingly, goat density does not affect wolf occurrence at the 2x2km scale and high levels of goat density are actually negatively correlated with wolf occurrence. One possible explanation for this is the fact that the domestic prey data used in this study does not distinguish livestock kept in barns from the one kept at outdoor, or, in other words, livestock that isn't and is available to wolves, correspondingly. This means that data from high livestock density probably corresponds to large fenced livestock farms.

Sheep density has a similar response in both 100x100m and 2x2km models, but it presents a very positive relation with wolf occurrence in the 10x10km models. Again, a possible explanation is the fact that there are large livestock farms in the area of Iberian wolf distribution used to build the model. It is determining a positive relation at this scale which is not seen at models with higher spatial resolution.

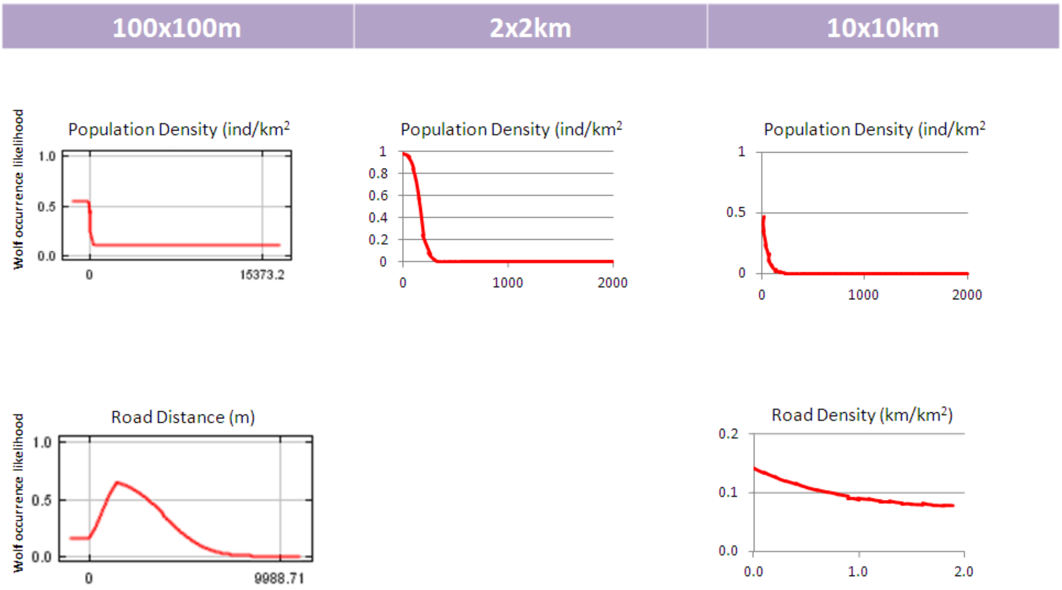

Fig. 9 - Relation between wolf occurrence probability and each environmental variable of the Human Pressure group used, at all three scales.

As expected, there is a negative relation between wolf occurrence and human pressure.

The absence of the variable road distance at the 2x2km model might be due to the similar density in the region were the sample was taken to build the model. It apparently has no influence in wolf occurrence at this scale, but it has at higher and lower spatial resolution.

It was the combination of all three types of variables (Landscape, Prey and Human pressure) that explained all models (Fig. 10). In the 100x100m model, the Landscape group was the one that had the largest contribution. In the logistic regression models (2x2km and 10x10km), the AIC weighted value of all combinations was too low, except when all three types were used together.

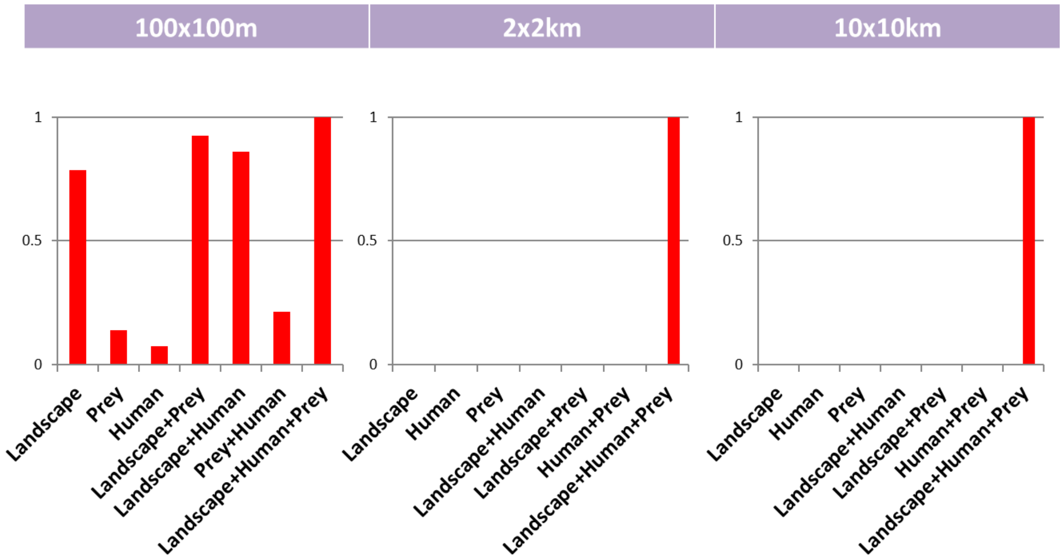

Fig. 10 - Relative importance of each environmental variables group for each model.

Validation

The model comparison had to take into account the fact that three different spatial scales were used. The only way we could test the behavior of each model was to extrapolate it to the spatial resolution of the other models. The actual wolf presence data from each region was compared with the squares that had a wolf occurrence probability higher than 50% (as a result of the model extrapolation), i.e., the models correct classification for each scale.

After this process, we concluded that it is the 2x2km model that has higher extrapolation accuracy (Table 11).

Table 11 – Comparison of the correct classification for each model, when extrapolated

Model	**Vila Real and Bragança counties, NE Portugal**	**North and southern Douro river**	**Iberianpeninsula**
100x100m	**--**	**95%**	**62%**
2x2km	**91%**	**--**	**93%**
10x10km	**66%**	**71%**	**--**

The 100x100m model, though a good model as seen previously, has low extrapolation accuracy.
The 10x10km model, with lowest spatial resolution (even if it had the highest sample size) can lead to less accurate results and extrapolation accuracy.

These results suggest that Sampling in locations that comprise all habitats used by Iberian wolf with high resolution may provide accurate habitat suitability models.

5. Final Remarks

The Iberian wolf is an endangered species and, therefore, conservation measures have to be adequate, in order to have a positive effect on the protection of the species. Results show that integrating spatial resolution and sample size considerations in the species studies, i.e., when modeling potential species occurrence, affects the results obtained. Also that, depending on the spatial scale and sample size used, models resulted in different areas of wolf occurrence probability and different relations between wolf and environmental variables, and, therefore, different AUC and validation values and extrapolation accuracy.
In this study, we conclude that Local scale (high resolution with low sample size) provided a good model with a low extrapolation accuracy; Regional scale (medium resolution and high sample size) provided the best model with the highest extrapolation accuracy; and the Iberian Peninsula scale (low resolution and the highest sample size) can lead to less accurate results and extrapolation accuracy.

One of the main technical difficulties we had was the large set of data, especially when considering high spatial resolution (100x100m), or large regions (when working at the Iberian Peninsula scale). In many cases, the software (either Maxent or Arcmap) or computer wasn't able to process the data correctly because of that.

The integration of data in the GIS environment allowed precise mapping the species distribution and the comparison of the biogeographic models in different and, in this case, large areas. However, as said previously, the software used posed some data processing problems at the Iberian Peninsula scale, especially when dealing with raster datasets. Other software can be used to overcome this problem.

Future actions, to improve the results of the analyzes made, would be to disentangle the effects of sample and scale on the wolf models accuracy, incorporate data on breeding sites and mortality to improve the wolf occurrence models and gather data with the highest spatial resolution possible. Data from domestic prey availability distinguishing the ones that are at wolf range (not in barns) and traffic values for the road network would also be important additions to this study.

We would like to add that the co-operation among wolf research groups in Portugal and Spain is crucial to evaluate the real preferences in order to perform conservation planning.

References

BERGER, A., PIETRA, S.A.D & PIETRA, V.J.D. (1996). A maximum entropy approach to natural language processing. Computational Linguistics, 22: 39-71.

BEIER, P., LOE, S. (1992). A checklist for evaluating impacts to wildlife movement corridors. Wildlife Society Bulletin 20: 434-440.

BOND, M. (2003). Principles of Wildlife Corridor Design. Center for Biological Diversity.

BROTONS, L., THUILLER, W., ARAÚJO, M. B., HIRZEL, A. (2004). Presence-absence versus presence-only modeling methods for predicting bird habitat suitability. *ECOGRAPHY* 27, pp: 437-448.

Coordination of Information on the Environment, European Environment Agency (2006), [online] Available at: <http://sia.eionet.europa.eu/CLC2006>, [accessed 20 May 2012].

ELITH, J., PHILIPS, S. J., HASTIES, T., DUDÍK, M., CHEE, Y. E., YATES, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, (2011) 17, pp: 43–57.

FERRÃO DA COSTA, G. (2007). O uso de sistemas de informação geográfica na elaboração de modelos biogeográficos da distribuição actual e histórica de espécies animais. O caso de *Acanthodactylusboskianus* (reptilia, lacertidae) no norte de África. Tese de Mestrado em Sistemas de Informação Geográfica, InstitutoSuperior Técnico.

GRILO, C., MOÇO, G., CÂNDIDO, A. T., ALEXANDRE, A. S., PETRUCCI-FONSECA, F. (2002). Challenges for the recovery of the Iberian Wolf in the Douro River South Region. *Revista de Biologia* 20: 121-133.

DOKO, T., KOOIMAN, F. A., TOXOPEUS, A. G. (2008). Modeling of species geographic distribution for assessing present needs for the ecological networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVII. Part B4. Beijing 2008, pp: 268-276.

GUISAN, A. & THUILLER, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters,* 8: 993-1009.

HIRZEN, A.H., HAUSSER, J., CHESSEL, D., PERRIN, N. (2002). *Ecological-Niche factor analysis: how to compute habitat-suitability maps without absence data?* Ecology, 83, pp: 2027-2036.

JAYNES, E.T. (1957). Information theory and statistical mechanics. *The Physycal Review,* 106 (4): 620-630.

JAYNES, E. T. (1985). Where do we go from there? C. Ray Smith and W. T. Grandy, Jr. (eds), Maximum-Entropy and Bayesian Methods in Inverse Problems, pp: 21-58.

KOOL, T. J., MOILANEN, A., TREML, E. A. (2012). Population connectivity: recent advances and new perspectives. Landscape Ecology, December 2012.

KUMAR, S., STOHLGREN, T. J. (2009). Maxent modeling for predicting suitable habitat for threatened and endangered tree Canacomyrica monticola in New Caledonia. *Journal of Ecology and Natural Environment* Vol. 1(4), pp. 094-098.

LLANEZA, L., LÓPEZ-BAU, J. V., SAZATORNIL, V. (2012). Insights into wolf presence in human dominated landscapes: the relative role of food availability, humans and landscape attributes. *Diversity and Distributions*, 18, pp: 459-469.

LLEBLOND, M., FRAIR, J., FORTIN, D., DUSSAULT, C., OUELLET, J., COURTOIS, R. (2011). Assessing the influence of resource covariates at multiple spatial scales: an application to forest-dwelling caribou faced with intensive human activity. *Landscape Ecol (2011)* 26, pp: 1433–1446.

MARTIN, J., REVILLA, E., QUENETTE, P., NAVES, J., ALLAINÉ, D., SWENSON, J. E. (2012). Brown bear habitat suitability in the Pyrenees: transferability across sites and linking scales to make the most of scarce data. *Journal of Applied Ecology*. Volume 49, Issue 3, pp: 621–631.

MECH, L. D., BOITANI, L. (Editors). 2003. Wolves: Behavior, Ecology, and Conservation. University of Chicago Press. 428 pp.

MLADENOFF, D.J., SICLEY, T. A., HAIGHT, R. G., WYDEVEN, A.P. (1995). A Regional Landscape Analysis and Prediction of Favorable Gray Wolf Habitat in the Northern Great Lakes Region. *Conservation Biology*, Vol. 9, No. 2. Pp: 279-294.

PETRUCCI-FONSECA, F. (1990). O lobo (*Canis lupus signatus* Cabrera, 1907) em Portugal. Problemática da sua conservação. Dissertação apresentada à Faculdade de Ciências da Universidade de Lisboa para a obtenção do grau de Doutor, Lisboa.

PHILLIPS, S.J., DUDIK, M. & SCHAPIRE, R.E. (2004). A Maximum Entropy Approach to Species Distribution Modeling. In: *Proceedings of the 21st International Conference on Machine Learning*, ACM Press, New York. Pp: 655-662.

PHILLIPS, S.J., [online] Available at: <http://www.cs.princeton.edu/~schapire/maxent/> [accessed 1 May 2012].

PHILLIPS, S.J., ANDERSON, R.P. & SCHAPIRE, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modeling*, 190: 231-259.

SAKAMOTO, Y., ISHIGURO, M., KITAGAWA G. (1986), D. Reidel Publishing Company, [online] Available at: <http://stat.ethz.ch/R-manual/R-patched/library/stats/html/AIC.html> [accessed 1 June 2012].

TURNER, M.G., GARDNER, R.H. O'NEILL, R.V. (2001). Landscape Ecology in Theory and Practice: Pattern and Process. Springer-Verlag, New York.

TUSZYNSKI, J. (2004), [online] Available at: <http://svitsrv25.epfl.ch/R-doc/library/caTools/html/colAUC.html>, [accessed 4 June 2012].

WU, J., GAO, J., TUELLER, P. T. (1997). Effects of changing spatial scale on the results of statistical analysis with landscape data: A case study. *Geographic Information Sciences* 3: 30-41.

ZUUR, A. F., IENO, E. N., WALKER, N. J., SAVELIEV, A. A., SMITH, G. M. (2009). Mixed Effects Models and Extensions in Ecology with R. Statistics for Biology and Health.