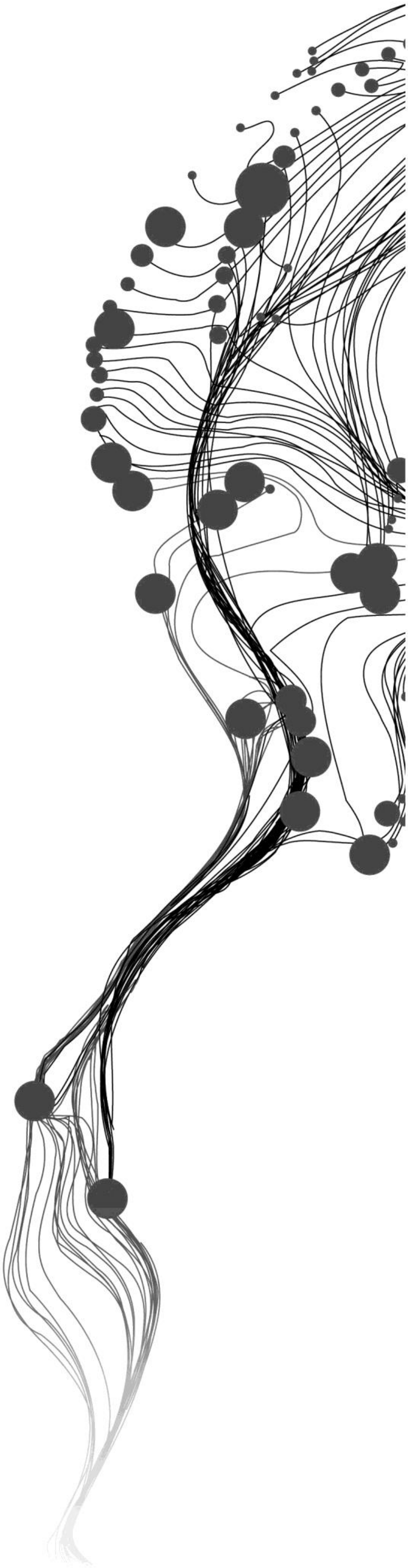
Decrypting Range Limit: The Curious Case of *Podarcis cretensis* in Crete

SURAJ BARAL
March 2019

SUPERVISORS:
Dr A. G. Toxopeus
Dr. T. A. Groen



Decrypting Range Limit: The curious case of *Podarcis cretensis* in Crete

SURAJ BARAL
Enschede, The Netherlands, [03,2019]

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: [Name course (e.g. Applied Earth Sciences)]

SUPERVISORS:
Dr. A. G. Toxopeus, NRS
Dr. T. A. Groen, NRS

THESIS ASSESSMENT BOARD:
Dr T. Wang (Chair)
Dr Petros Lymberakis (External Examiner, University of Crete)

DISCLAIMER
This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author and do not necessarily represent those of the Faculty.

ABSTRACT

A species meets its range limit if the climatic condition favourable for the species exceeds its physiological limits or if it is controlled by biotic interaction of competitor or limited by dispersal and demographic stochasticity or gene flow. Not all species can fully exploit the potentially suitable habitat available to it. Only a few may be capable of ranging throughout its suitable habitat while most of the species give up their distribution far before its potential range. Biotic mode of range limitation is very rare. Restricted gene flow contributes to limit the distribution of a species in most cases. A set of environmental variables or a single variable may induce restriction of gene flow into a population thereby limiting the range. With the advent of advanced molecular method and integration of population genetics and landscape study, it has been possible to check various hypothesis of range limit due to gene flow. With this concept, a candidate set of environmental variables to predict the distribution of *P. cretensis,* a Cretan endemic reptile with no known apparent reason for range limit. The working hypotheses are to know if the species meets its range limit in response to abiotic variable or if the restricted gene flow is contributing towards range limit. An ensemble species distribution approach was used to compare the potential distribution and the realized distribution. The environmental data was converted into appropriate cost raster to determine an accumulated cost distance and resistance distance. The cost distances were correlated with genetic distance calculated using a set multi-loci nuclear microsatellite gene to establish the relationship between environmental cost and gene flow. I found that the present distribution is below the potential range and abiotic variable is not the cause of the range limit. Also, the study did not find any significant relationship between selected variables and gene flow. The environmental variables used was found to be too coarse to have an impact upon the species. Use of micro-habitat scale environmental predictors and introduction of biological interaction and mechanistic models into SDM can help to solve the even curious case of *P. cretensis.*

Key Words: Range limits, genetic differentiation, Environmental Niche Model, multi-loci microsatellite, landscape genetics.

ACKNOWLEDGEMENT

I would like to express my gratitude to University of Twente, Faculty of Geoinformation Science and Earth Observation for providing me an opportunity to obtain the knowledge of application of geoinformation science and remote sensing in natural resource management. I would also like to thank ITC Excellence Scholarship Program for financial support for my studies here.

My sincere thanks to my first supervisor Dr A. G. Toxopeus without whom the project would not have been possible. Thanks to his comments, suggestion and constructive feedback during the entire research period that motivated to bring the most of myself. I could not have thought of completing the thesis without the help of Dr T. A. Groen as my second supervisor. His invaluable comments and suggestions on the technical implementation of species distribution model, cost parameterization, interpretation and manuscript enabled me to successfully complete the thesis. Also, sincere word of thanks goes to Dr C.A.J.M. de Bie for his constructive comments during proposal presentation and mid-term defence.

The thesis would not have its form that it has now without the data support from Natural History Museum of Crete, University of Crete. I particularly want to thank Dr Petros Lymberakis for his warm hospitality during my stay in Crete and expert opinion on the species biology that made the thesis possible. I do not have enough words to thank Ms Loukia Spilani for her unwearied support and conceptual guidance on genetics and molecular lab technology without which I could not think of completing the genetics part of the thesis. I would like to thank Dr P. Nyktas, Drs R. G. Nijmeijer and Dr B. Mishra for their continuos moral assurance during thesis writing. Also, thanks to Amar Kunwar and Sabina Koirala for their comments and suggestion upon the manuscript. Sincere word of thanks goes to Dr Marie-Josee Fortin, Dr Andrew Storfer, and Dr Niko Balkenhol for their support and critical suggestion on genetic distance analysis.

Thank you my dear wife Babita Paudyal for your continuous moral assurance and support throughout my study period. Also thank you my friends here and all family members back home for supporting me and making my stay here an incredible one.

Thank you very much.

Suraj Baral
Enschede, the Netherlands
February 2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AUC: Area Under Curve
BRT: Boosted Regression Trees
CORINE: Coordination of Information on the Environment
DEM: Digital Elevation Model
DNA: Deoxyribonucleic Acid
ESA: European Space Agency
ESBN: European Soil Bureau Network
GLM: Generalized Linear Model
m: meter
Maxent: Maximum Entropy
mm: millimetre
MYA: Million Years Ago
NDVI: Normalized Difference Vegetation Index
NIEHS: National Institute of Environmental Health Sciences
Pers. comm.: Personal Communication
RNA: Ribonucleic Acid
ROC: Receiver Operating Curve
SDM: Species Distribution Model
SPOT: Satellite for Observation of Earth
TSS: True Skill Statistics
USGS: United States Geological Survey
UTM: Universal Transverse Mercator
VIF: Variation Inflation Factor
WGS: World Geodetic System

1. INTRODUCTION

1.1. Background

Hutchinson (1957) was a pioneering ecologist to mathematically explain an indefinite species existence exclusively in an n-dimensional hypervolume of environmental variables given the absence of competition. However, the quest for the environmental variables that limit the range of any species has interested biologists long before Hutchinson (1957). Darwin (1859) was one of the foremost biologists to recognize the importance of competition and climate to limit the range of a species in time and space. The concept was later extended by MacArthur (1972) to explain the various mode of species range limitation by biotic, abiotic factor and their interactions. MacArthur (1972) also implicitly inferred the role of genes in adapting against unfavourable abiotic factors before finally losing on the distribution limits. Thus, the factors that limit the distribution of a species has been the fundamental issue in ecology, evolution or other biological fields (Gaston, 2009) since long. Although variously explained by different authors (Louthan, Doak, & Angert, 2015; MacArthur, 1972; Sexton, McIntyre, Angert, & Rice, 2009), it can be agreed that a species limits its range by any one of the three possible modes of species range limit: biotic factors (inter-specific interaction), abiotic factors (environmental factors) or the interaction between biotic and abiotic factors.

1.1.1. Species Range Limits

The difference in environmental factors in space and time has been mostly considered as the major driver of range limit. Both theoretical (Gaylord & Gaines, 2000; Holt, 2003; Pulliam, 2000) and empirical evidence (Araújo & Pearson, 2005; Arundel, 2005; Cumming, 2002) have considered spatial and temporal heterogeneity in environmental variables as a major driver for limiting the range of species. Temporal and spatial heterogeneity in climatic condition affect dispersal (Araújo & Pearson, 2005; Gaylord & Gaines, 2000; Holt, 2003; Holt, Keitt, Lewis, Maurer, & Taper, 2005) or decrease carrying capacity strengthening density-dependence and high demographic stochasticity (Holt et al., 2005) thereby limiting the range. Towards the range extremes, population growth rate decreases due to a less favourable environmental condition (Cumming, 2002; Gutiérrez & Defeo, 2005; Mehlman, 1997) and decreased survival rates (Gutiérrez & Defeo, 2005). Unfavourable climate condition towards the range end also leads to reproductive failure (Arundel, 2005; Gaston, 2009; Pulliam, 2000). All these factors singly or in synergy lead to limiting distribution range for plants and animals.

Biotic mode of species distribution limit is uncommon in nature (Gaston, 2009). Nonetheless empirical evidences have shown that inter-specific interaction like parasitism (Briers, 2003), predation (DeRivera, Ruiz, Hines, & Jivoff, 2005) and herbivory (Bruelheide & Scheidel, 1999) or competition (Bridle & Vines, 2007; Goldberg & Lande, 2006; Louthan et al., 2015; MacArthur, 1972) limit the distribution range of native or introduced species. Resource (Bridle & Vines, 2007; Case, Holt, McPeek, & Keitt, 2005), carrying capacity and local population dynamics are important factors that determine the competitive advantage on competing species and thus to the extent of distribution for the interacting species (Case et al., 2005). Predation limits the range of prey by direct mortality (Case et al., 2005; DeRivera et al., 2005), unstable local population dynamics of obligate predator and prey (Case et al., 2005), reduced fitness for recruitment, and dispersal (Briers, 2003; Case et al., 2005) are the factors that restrict range limit due to parasitism.

On an evolutionary perspective, environmental variables exert selection pressure to local population enabling them to adapt to local condition (MacArthur, 1972; Manel, Schwartz, Luikart, & Taberlet, 2003). These adaptive mutation and selection accumulate along the changing environmental gradient causing a new species to evolve thus creating a sharp limit for the distribution without known barriers (Dieckmann & Doebeli, 1999; Doebeli & Dieckmann, 2003; Polechová & Barton, 2015). More commonly, reduced gene flow from the central population prevents the population from local adaptation to limit the range

(Gaston, 2009; Kirkpatrick & Barton, 1997). Gene flow from centre population increases the genetic diversity of the marginal population and thus increases adaptive potential due to increased genetic diversity (Micheletti & Storfer, 2017) when gene flow is reduced due to high resistance of a some or one of the environmental factor, the adaptive fitness of peripheral population against extreme climatic conditions decreases which limits the distribution range of a species (Eckert, Samis, & Lougheed, 2008; Gaston, 2009; Kirkpatrick & Barton, 1997; Micheletti & Storfer, 2017).

1.1.2. Incorporating Gene Flow on Landscape Study

Genes, in simplified terms, can be defined as a sequence of DNA strands which are the basis for building a protein that is responsible for a particular function when transcribed by messenger RNA (H. Pearson, 2006). The particular function may represent traits like eye colour, the colour of hair, blood type, the risk to a specific disease or biochemical process undergoing in an individual. When the copies of the gene differ from each other for a diploid set of DNA, the variant is called alleles which are the basis of visible phenotypic expression like hair colour due to dominant allele while the other remains recessive. Sometimes multiple alleles are codominant thereby producing a phenotypic trait determined by all alleles, eg. ABO blood group system in human (Clark, 2005). In the specific position in DNA, tandemly repeated units ranging from 2 to 10 base pairs of nucleotides occur, which are commonly called as microsatellites (Bruford et al., 1996; Waits & Storfer, 2016). The microsatellites are characterized by high polymorphism; the genes in these loci differ within individuals (Bruford et al., 1996). This property makes the microsatellites an appropriate choice to study individual genetic variation (Storfer, Murphy, Spear, Holderegger, & Waits, 2010). The polymorphic alleles are transferred from individual to the next generation after successful mating and reproduction and to space as a result of dispersion of the individuals. When an allele of a particular individual is thus transferred to generation and space, it is termed as gene-flow which helps the organism to prevent reproductive isolation and promote local adaptation and adaptive evolution on complex landscapes (Whitlock & McCauley, 1999). These factors highlight the importance of gene-flow within the landscape.

Marking and following individual organism as a measure of gene flow was the classical method used to determine the gene flow but it was time-consuming, expensive and difficult technically (Waits & Storfer, 2016; Whitlock & McCauley, 1999). The advent of molecular techniques allowed the molecular estimate of gene flow is replacing the classical method (Waits & Storfer, 2016). The direct method of gene flow estimation includes assignment test and parentage analysis. The indirect approach consists of the coalescent approach and various population and individual-based distance metrics (Waits & Storfer, 2016). In assignment test of direct method of gene-flow estimation, multilocus genotype are assigned to its putative population based on the expected probabilities of that genotype in each potential source of population under Hardy-Weinberg and Linkage Equilibrium (Paetkau, Calvert, Stirling, & Strobeck, 1995) or in form of ancestry coefficients for the inferred population (Waits & Storfer, 2016). Alternatively, the direct estimation of gene flow can be done using parentage analysis where multilocus genotypes are used to determine the parents from a set of potential parents sampled using parental exclusion or statistically based parentage assignments approaches (Waits & Storfer, 2016). Sometimes population across environmental gradient are mostly continuous and panmictic; then it becomes difficult to differentiate between allele frequency and estimate gene flow using the direct approach. In the case of panmixia the indirect method of gene flow more useful for analysing the impact of the environmental variable on genetic structure (Waits & Storfer, 2016). These metrics also known as measures of genetic differentiation have a predictable relationship to the rate of migration (Holsinger & Weir, 2009; Waits & Storfer, 2016; Whitlock & McCauley, 1999) and thus can be estimated as the indirect measure of gene flow (Waits & Storfer, 2016). Recent advancement in molecular techniques combined with available statistical tools like geo-statistics and Bayesian approaches enabled the combination of population genetics and landscape ecology as an interdisciplinary field of landscape genetics (Manel et al., 2003; Storfer et al., 2010). This integrated approach made way to check various hypothesis about barrier to dispersal, effect of landcover change (see Storfer et al., 2010 for review), gene flow hypothesis of range limit (Micheletti & Storfer, 2015) and identification of unknown variable for range limit (Micheletti & Storfer, 2017). Wright (1943) conceived a positive correlation between geographic distance and genetic distance between population. This "Isolation by Distance" relation is now considered as working null hypothesis for establishing a connection between genetic distance and isolation by landscape (Waits & Storfer, 2016), isolation by resistance (McRae, 2006) or isolation by the environment (Wang, Glor, & Losos, 2013).

The alternative hypotheses of isolation by the landscape, resistance or environment are analysed by connectivity analysis of resistance surface (Spear, Balkenhol, Fortin, McRae, & Scribner, 2010). A resistance surface is a gridded representation of an environmental variable or a combination of them (Spear et al., 2010) the value of which represents the cost of movement or reduction of fitness to flow of the gene of the individual (Zeller, McGarigal, & Whiteley, 2012). The resistance surfaces are parameterized either by expert opinion (Spear, Cushman, & McRae, 2016) or empirical evidence like resource selection function or species distribution model (Hagerty, Nussear, Esque, & Tracy, 2011; Wang, Yang, Bridgman, & Lin, 2008).

The measure of gene flow between individuals or population is compared with the environmental resistance to infer the impact of the variable for the gene flow. Connectivity measures like Least Cost Model (Adriaensen et al., 2003) or circuit path model (Chandra, Raghavan, Ruzzo, Smolensky, & Tiwari, 1997; McRae, 2006) transform the characteristics of the intervening landscape to a measure of resistance distance by the variable (McRae, 2006). The statistical validation can be performed between the estimate of gene flow with the estimate of resistance (Spear et al., 2016) using correlative approaches like correlation or linear models (Wagner & Fortin, 2016). The model with significant statistical support for limiting the gene flow is considered as the unknown environmental variable to determine the range limit of a species (Micheletti & Storfer, 2017).

1.1.3. Study Species

Podarcis cretensis (Wettstein, 1952) (Squamata; Lacertidae) is a small lizard (Figure 1-1) with a slender body and tail almost twice of the body length (Lymberakis, Poulakakis, Kaliontzopoulou, Valakos, & Mylonas, 2008). It is an endemic in Crete and a few islets around Crete (Lymberakis, 2009; Lymberakis et al., 2008).

Figure 1-1: *P. cretensis* in the wild (Source: Natural History Museum of Crete)

The species inhabits shrublands, rocky area and dry river beds up to an altitude of 2000 m (Lymberakis, 2009). Restricted only in the western part of the island (Lymberakis, 2009; Lymberakis et al., 2008, Figure 1-2), the species is threatened due to the impacts of urbanization and tourism industry and is enlisted as "Endangered" species in IUCN (Lymberakis, 2009) with immediate threats of extinction if the causal agents do not cease soon (IUCN, 2012).

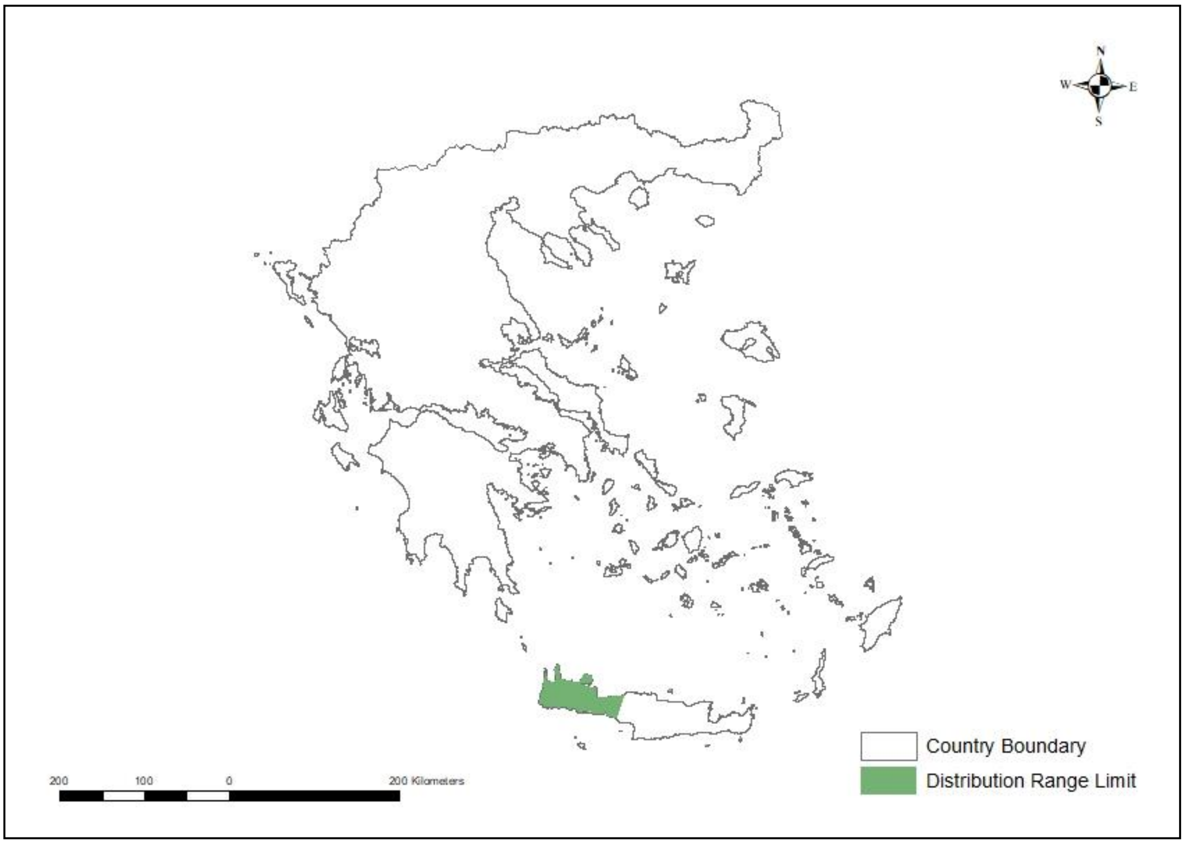


Figure 1-2: Realized Distribution Limit of *P. cretensis* in Greece. Source: IUCN Red List of Species (Lymberakis, 2009)

Phylogenetically, the species belongs to *P. erhardii* (Bedriaga 1882) group that constitute *P. erhardii, P. peloponnesiaca, P. cretensis* and *P. levendis* (Lymberakis et al., 2008; Poulakakis et al., 2003) that are similar morphologically (Lymberakis et al., 2008). The splitting of the monophyletic group to new species has been contributed to the geological event during Tortonian Age of Miocene Epoch (circa 9 MYA)- opening of Mid Aegean Trench and then later during the Messinian Crisis (circa 5 MYA) also of Miocene Epoch- (Lymberakis & Poulakakis, 2010, Figure 1-3). The common ancestor of present-day *P. cretensis* population in Crete and its islets originated during the Messinian Crisis (Lymberakis & Poulakakis, 2010; Lymberakis et al., 2008; Poulakakis et al., 2003). The distribution of *Podarcis cretensis* has been stated as curious by various authors (Herkt, 2007; Zabalaga, 2008). The present-day population of *P. cretensis* is distributed only on the western part of the island and some satellite islets on the eastern part of Crete (Lymberakis, 2009). The molecular evidences show that the population of eastern islets share a recent ancestry and are linked closely to the western population (Lymberakis et al., 2008; Poulakakis et al., 2003) and that the distribution might have been extended to the east Crete historically and have disappeared from the eastern part, surprisingly with no paleontological records (Lymberakis 2018 *pers. comm.*). The population distribution in the western region is also unusual. The species consists of two subpopulations in the western part as indicated by mitochondrial DNA analysis. A relict clade that split approximately 2.9 MYA, is localized only in higher altitude of the White Mountains and the young clade that originated circa 2.3 MYA, is found in the lower altitude of the White Mountains and the rest of the island (Lymberakis et al., 2008; Poulakakis et al., 2003; Poulakakis, Lymberakis, Valakos, Zouros, & Mylonas, 2005, Figure 1-4). The two subpopulations in the White Mountain have been known to interbreed (Lymberakis 2018*, pers. comm.*).

1.1.4. Statement of the problem

Owing to the curious distribution, its distribution has been tried to solve spatially (Herkt, 2007; Zabalaga, 2008) or phylogenetically (Poulakakis et al., 2003). However, methods are unable to explain possible mechanism of range limit as it is evident that space has an impact on phylogeny (MacArthur, 1972) and vice versa (Pometti et al., 2018).

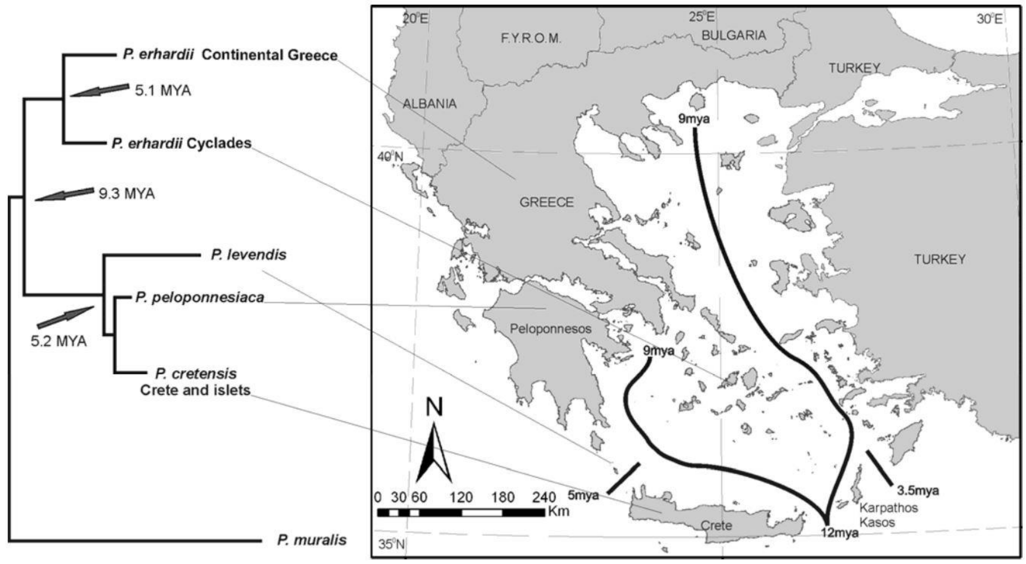


Figure 1-3: Evolution tree for *P. erhardii* group (Image adapted from Lymberakis, 2010)

Niche Models (Maxent etc.) used to define the distributional limits of *P. cretensis* (Herkt, 2007; Zabalaga, 2008) use the correlation between the observed pattern of presence (and absence) to the bio-climatic variables to predict the distribution limit of the species. However, these models might misrepresent the distribution limit because the range dynamics of the species like dispersal, migration or demographic stochasticity are not considered in these models (Schurr et al., 2012). On the other hand, Poulakakis et al. (2003) used molecular phylogeny to infer the present-day extent of *P. cretensis* from paleogeography of the Mediterranean. This model also, however, failed to incorporate the genetic variability as a result of environmental interaction which may limit the range of the species. Thus, the result from any one of the study methods might be incomplete without the other. This study tries to fill the research gap not realized by two different means to answer the curious case of the range limit of *P. cretensis.* Considering the impact the environmental variables have on gene flow might help to solve the crypic distribution.

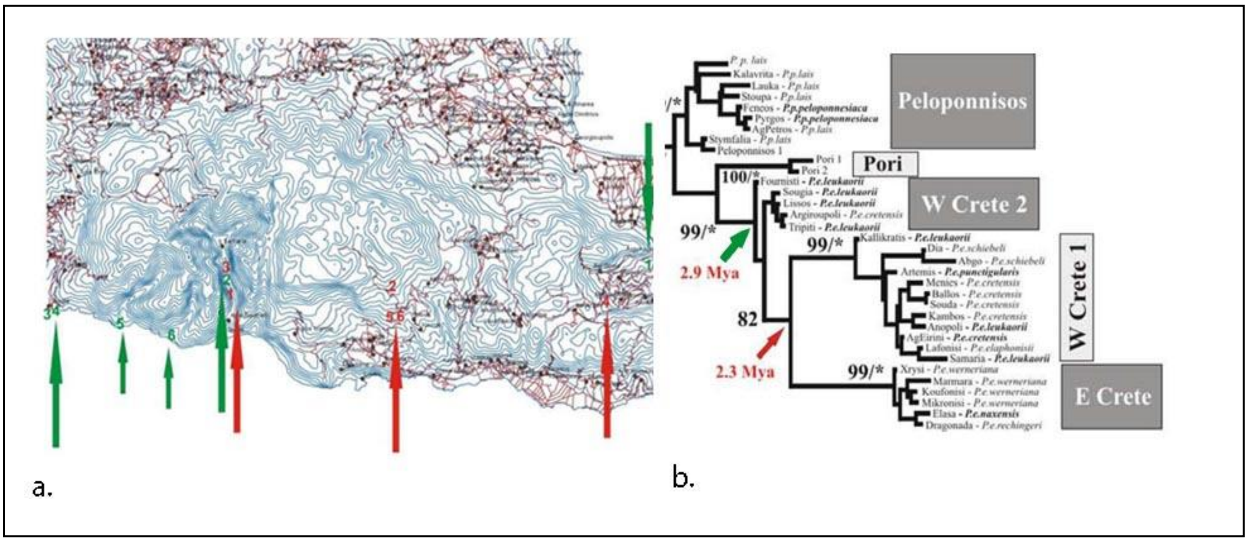


Figure 1-4: Geographical location of two subpopulations (a) and phylogeny (b) for the western population of *P. cretensis* from White Mountain (Green Arrow represents relict Population and Red young population). Image adapted from Zabalaga, 2008 (a) and Poulalakis et al., 2005 (b).

1.1.5. Research Objective

The broad objective of the study is to determine the landscape variables that restrict the gene flow to limit the range of *P. cretensis.*

Specific objectives of the study are as follows:

- To determine the suitable habitat inside the current range and potentially suitable habitat outside the range.
- To determine the variable with a high cost for gene flow.

1.1.6. Research Question

1. Does the potentially suitable habitat for the species only fall within the current range?
2. What are the variables that prevent the gene flow to limit the range of *P. cretensis*?
3. Does an SDM predict the gene flow better than any candidate variable set?

1.1.7. Research Hypotheses

Hypothesis 1

The potential distribution of *P. cretensis* is larger than the current distribution.

Hypothesis 2

Isothermality, altitude or landcover restricts the gene flow to limit the distribution to the current range.

Hypothesis 3:

The probability of presence as predicted by species distribution model restricts the gene flow to limit the distribution of *P. cretensis*.

2. MATERIALS AND METHODS

2.1. Study Area

The island of Crete (Figure 2-1) is in the Eastern Mediterranean Sea (35°20'27" N, 25°07'46" E) covering an area of approximately 8,336 sq. Km (Panagos, Christos, Cristiano, & Ioannis, 2014). The climate here is a Mediterranean type (Rakham & Moody, 1997; Vogiatzakis & Rackham, 2008) with hot, dry summer and cold, humid winters (Zabalaga, 2008). The rainfall ranges from 2000 mm at the White Mountains to 240 mm in south-east Crete (Rakham & Moody, 1997; Vogiatzakis & Rackham, 2008). About 6-degree fall in temperature per 1000 m can be felt with snow above 1600 m from October to May and occasional frost throughout the island (Vogiatzakis & Rackham, 2008). The island consists of 15 mountain ranges, three of which are above 2000 m high (Vogiatzakis & Rackham, 2008) which were created during the late Cretaceous period as a result of tectonic movement of African and Aegean Plate (Rakham & Moody, 1997). These mountains are the main geographical features that create variation in aridity, temperature (Vogiatzakis & Rackham, 2008) and precipitation (Grove & Rackham, 1993). The landscape is dominated by Mesozoic and Tertiary limestones in high, middle and low elevations with phyllite-quartzite and Neogene deposits in lower elevations (Vogiatzakis & Rackham, 2008).

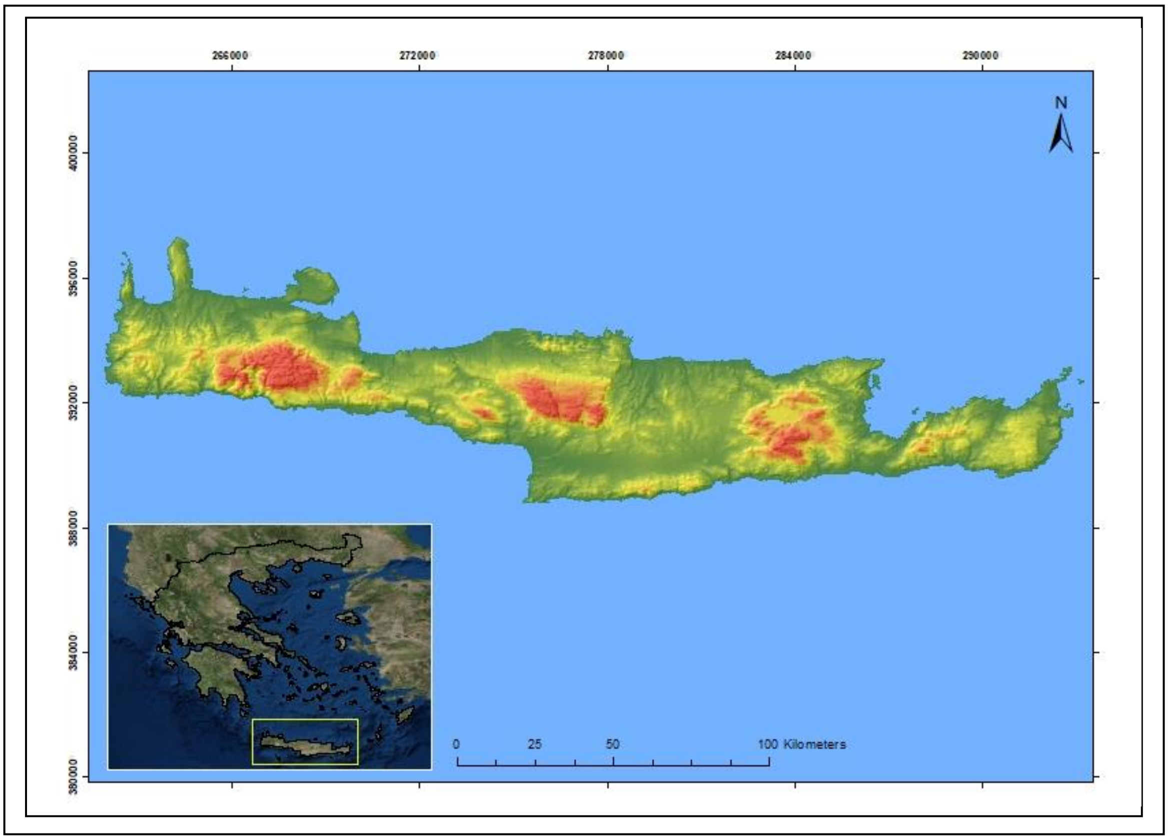


Figure 2-1: Map of Crete laid over SRTM-DEM and Hill shade (Warmer colour represent higher altitude), inset: Position of Crete in Greece (Map Source for inset: World Imagery ArcMap)

The floral element of the island is contributed by European, Asian and African affinities- largest being European Open Phrygana while the least, African floral elements like *Viola scorpiuroides, Aristidia ascensionis,* etc (Rakham & Moody, 1997). With around 200 endemic flora, Crete in one of the world's biodiversity hotspot (Vogiatzakis & Rackham, 2008). Also with high endemism in wild fauna, the species richness of fauna is also considered to be rich (Lymberakis & Poulakakis, 2010). However, the rich biodiversity of the island, the natural and semi-natural vegetation are now heavily impacted by human influence (Grove & Rackham, 1993) primarily due to habitat loss and degradation attributed to urbanization and tourism (Lymberakis, 2009).

2.2. Data Collection

This section describes the material and methods used in this study. The method consists of three parts: Species Distribution Model, Cost Raster Parameterization and Gene Flow Modelling. Collection of observation data and environmental variable to produce an ensemble species distribution model (Figure 2-1, Section 2.3.2) to predict the potential range of *P. cretensis* contributed to the first part of the study. The empirical cost parameterization of the produced the SDM and expert opinion parameterization of rest of the environmental variable (Section 2.3.6) and estimation of landscape connectivity (Section 2.3.7) consisted of the next part of the process (Figure 2-2). The next step (Figure 2-2) consisted the collection of genetic data (Section 2.2.3), and estimation of gene flow (Section 2.3.8) to test the importance of a variable to limit the gene flow (Section 2.3.9).

2.2.1. Species Observation Data and Pseudo-absence Data

The species observation data were collected as a part of a molecular study conducted by the Natural History Museum of Crete. A total of 203 geo-referenced individuals were collected on the study. I removed duplicated presence points from the database. Finally, 126 geo-referenced points (Appendix 1) were available for the analysis.

For models using presence-absence points (GLM and BRT), 1000 pseudo-absence point were generated randomly in ArcMap 10.6. The pseudo-absence points were made sure not to overlap the true presence point by buffering 1 km distance from presence points and masking it off for a pseudo-points (Appendix 1) generation.

2.2.2. Environmental Data

Species Distribution Model makes use of environmental predictors to define the probability of habitat suitability (Austin, 2002; Phillips & Dudík, 2008). Spatial data like a bio-climate map, digital terrain map, soil factors and geology maps, vegetation related maps or maps relating the anthropogenic impact can be used for the creation of habitat suitability models (Franklin & Miller, 2009). I used the environmental variables that may show some ecological relevance to the species ecology (Table 1) based on literature reviews or expert opinion. I resampled, converted from vector to raster or converted raster to ascii data format of these variables in ArcMap 10.6. The final environmental variables were projected to WGS84 UTM zone 35N reference system. Because no information on average dispersion or daily movement of *P. cretensis* was available, I used a cell size of 250 m to model the distribution of the species. The original cell size of most of the variables used, i.e. 1 km is too coarse for modelling the distribution of *P. cretensis* because of its size, and finer scale would have added the risk of spatial autocorrelation for SDMs.

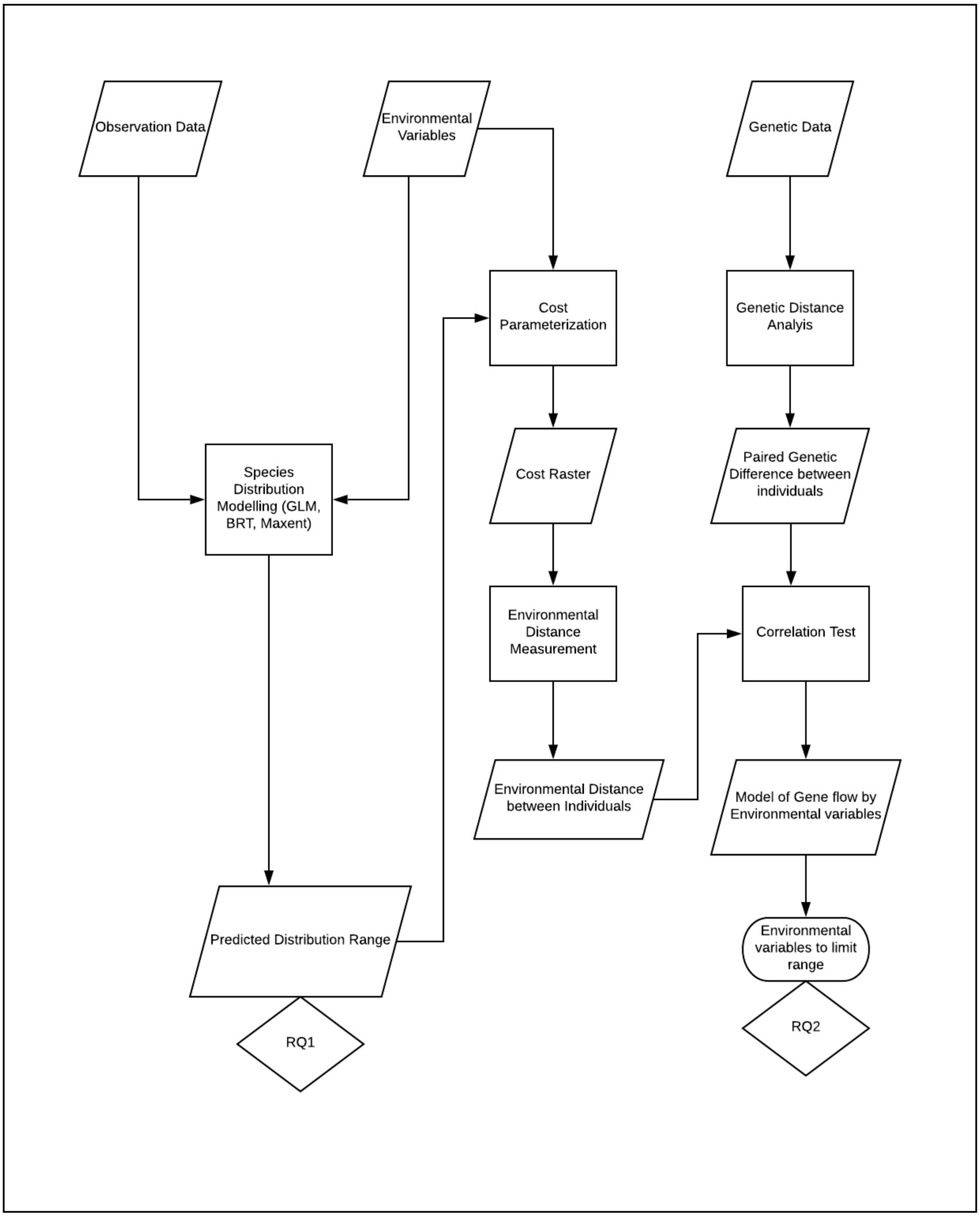
Table 2-1: Ecologically Meaningful Candidate Set of Environmental Variables

Variable	Data Type	Unit	Original Resolution	Source
Actual Evapotranspiration	Continuous	mm	0.5 degrees	USGS/NIEHS
Annual Mean Temperature (Bio1)	Continuous	°C	30 arc seconds	Hijmans et al., (2005)
Mean Diurnal Range (Bio2)	Continuous	°C	30 arc seconds	Hijmans et al., (2005)
Iso-thermality (Bio3)	Continuous	Percent	30 arc seconds	Hijmans et al., (2005)
Temperature Seasonality (Bio4)	Continuous	Percent	30 arc seconds	Hijmans et al., (2005)
Annual Precipitation (Bio12)	Continuous	mm	30 arc seconds	Hijmans et al., (2005)
Precipitation of Driest Quarter (Bio17)	Continuous	mm	30 arc seconds	Hijmans et al., (2005)
NDVI	Continuous	-	1 km	SPOT
Altitude	Continuous	m	90 m	SRTM
Slope	Continuous	degree	90 m	Calculated in ArcMap
North	Continuous	degree	90 m	Calculated in ArcMap
East	Continuous	degree	90 m	Calculated in Arc Map
Distance to road	Continuous	m	250 m	Open Street Map
Distance to river	Continuous	m	250 m	Open Street Map
Landcover	Categorical	-	100 m	ESA
Soil	Categorical	-	250 m	ESBN
Geology	Categorical	-	250 m	ESBN

Bio-climatic variables are the interpolated derivatives of basic climate parameters like precipitation, maximum and minimum temperature collected throughout the world from climatic stations (Hijmans et al., 2005). The mean temperature throughout the year has some impact on the distribution of a cold-blooded animal. Also, the Mean Diurnal Range, temperature fluctuation within a month can be related to monthly stress of temperature change which might drive the adaptation of the species. Similarly, the isothermality as an index that captures the day-night temperature oscillation relative to season oscillation has been known to be particularly important for insular species (Nix, 1986) as the physiological extreme of the species might not occur in the island or the species might be fully adapted to the extremes occurring in the island. Therefore, the variation between the daily and seasonal temperature that might cause stress for the species may contribute towards species distribution limits. The deviation from the yearly mean temperatures might also present the impact of climatic extremes in the physiology of *P. cretensis*. As seasonal means and extremes of these climatic parameters are known to relate strongly with empirical and theoretical species distribution (Stockwell, 2006).

Topographical map obtained from DEM were second classes of the environmental variable used for generating resistance raster (Table 1). These variables like altitude, slope, aspect, Eastness or Northness directly or indirectly affect temperature, soil characteristics or solar radiation and thus to the species biology (Franklin & Miller, 2009). As the species prefers a dry and open area with low water requirements (Lymberakis, 2009), availability of water and the amount transpired from the land might be related to the presence of *P. cretensis*. Therefore, I included actual evapotranspiration and distance to the river in the model. NDVI and CORINE Landcover data are the vegetation related variable used for the modelling (Table 1). These data-sets are used extensively for SDM because the gradients vegetation cover can be expressed as the habitat structure for wildlife (Franklin & Miller, 2009). Rather than using CORINE land cover data as such, ecologically meaningful habitat categories for *P. cretensis* was used by reclassifying all the available landcover type in Crete in ArcMap 10.6 (Table 2-2). Because *P. cretensis* is expected to make use of soil type for burrowing, basking, movement or camouflage, the soil type was included in the model. Habitat disturbance may be the factor to deter the species away or cause mortality in reptiles (Brehme, Hathaway, & Fisher, 2018), Euclidean distance to the road was used as disturbance maps. The Euclidean distance was calculated in ArcMap 10.6 from the available vector layer (Table 2-1) of the road.

Figure 2-2: Methodology Flow Chart



2.2.3. Genetic Data

The study includes genetic data from 203 individuals of *P. cretensis* collected by Natural History Museum of Crete throughout the island of Crete. Data were collected as a long-term survey conducted throughout Crete. The individuals were captured, or tissue sample from the tip of the tail was collected and preserved in the Natural History Museum of Crete.

A standard ammonium acetate protocol (Bruford, Hanotte, & Burke, 1998) or DNeasy Blood & Tissue Extraction kit (Qiagen®, Hilden, Germany) was used to extract DNA for 72 specimens. DNA genotype of the remaining individuals was obtained from the previous studies (Lymberakis et al., 2008; Poulakakis et al., 2003, 2005). The genomic data were then genotyped for 17 microsatellite loci (Appendix 2) for the estimation of gene flow. I used the genotype data of the individuals from the western population only if they had genotyped data for more than ten loci.

Single PCR products were mixed with an internal size standard (GeneScan 500 LIZ, Applied Biosystems) and the amplified allele sizes were visualized on an automated sequencer, type ABI3730 (Applied Biosystems). The program STRand v.2.4.109 (Toonen & Hughes, 2001) was used for genotyping. The microsatellite allele binning was conducted using the program FlexiBin v.2 (Amos et al., 2007).

All the laboratory work and genotyping were done by the colleagues from the Natural History Museum of Crete. The present study used already available data for the analysis.

Table 2-2: Reclassified Landcover type from CORINE

Corine Legend	Corine Landcover type	Classified Landcover
1	Artificial Surface	Buildup area
2.1	Arable Land	Highly Disturbed Agricultural Area
2.2 2.4	Permanent Crop Heterogenous agricultural areas	Low Disturb Agriculture
3.1	Forest	Forest
2.3 3.2	Pasture Grassland, Heath, and Moor	Grassland
3.3	Open space with little/no vegetation	Open/ Bare area
5	Water Bodies	Water Bodies

2.3. Data Analysis

2.3.1. Multicollinearity Test

Multicollinearity appears when two or more variable in a statistical model are related linearly (Dormann et al., 2012). The collinearity makes the parameter estimate unstable, standard errors are inflated and results in biased inference (Dormann et al., 2012; Graham, 2003). Therefore, it is routine to remove these collinear variables in multiple regression models. I used a Variation Inflation Factor (VIF; Marquardt, 1970) test to eliminate the correlated variable beforehand. I used a threshold of VIF= 10 (Kutner, Nachtsheim, Neter, & Li, 2005) to remove highly correlated variables. “Sample” tool in ArcMap 10.6 was used to extract the variable values at the presence and pseudo-absence points the and R-Studio was used to perform the test.

2.3.2. Habitat Modelling

To determine the probability of distribution of the species, I used an ensemble of 3 different species distribution models *viz* Generalized Linear Model (GLM), Boosted Regression Trees (BRT) and Maximum Entropy Model (MAXENT).

Generalized Linear Model (McCullagh & Nelder, 1989) is a classical regression technique that allows the response of a set of environmental variables to be modelled simultaneously (Austin, Nicholls, & Margules, 1990). A link function is used in GLM to combine the predictors with the response variable which allows flexibility to model any exponential family distribution (Guisan, Edwards, & Hastie, 2002; McCullagh & Nelder, 1989). The link function also ensures the linearity of the predictors and restricts the prediction within the range of appropriate values (Guisan & Zimmermann, 2000). GLM is a widely used species' distribution modelling approach because of its ability to model ecological relationship realistically with strong statistical foundation (Austin, 2002). I used a binomial distribution with logistic link function to model the presence and pseudo-absence data for *P. cretensis*. I ran ten replicates of the data with 70 per cent of observation used for model building. I used R studio for running the model using the "sdm" package (Naimi & Araújo, 2016).

Boosted Regression Trees considered as an advanced form of regression that combines the performance of many weak classifiers to produce a powerful classification (Friedman, Hastie, & Tibshirani, 2000) but also draws insights from Machine Learning techniques (Elith, Leathwick, & Hastie, 2008). BRT uses two algorithms to build up a model: decision trees and boosting (De'Ath & Fabricius, 2000; Friedman et al., 2000). Decision trees use a series of rules for homogenous partition of data into subgroups, growing the tree and prune when the threshold is reached (De'Ath & Fabricius, 2000; Franklin & Miller, 2009). Boosting then minimizes the deviance by adding a new tree at each step (Elith et al., 2008). The model can be fitted to various distribution similarly as GLM using link and specific distribution function. I ran ten replicates with a maximum of 10000 trees to be built with 70 per cent of observation data used for model building because a larger number of trees is preferable (Elith et al., 2008). I used R studio to run the model using the "sdm" package (Naimi & Araújo, 2016).

Maximum Entropy (Phillips, Anderson, & Schapire, 2006) belongs to the Machine Learning technique or algorithm modelling which assumes that the data is drawn from an unknown multivariate distribution, and the solution is to fit the unknown algorithm given the predictors and response data set (Breiman, 2001; Elith et al., 2008). Although the modelling technique was less common in ecological question previously (Elith et al., 2008), the use of machine learning in species distribution model has found a rapid growth (Yackulic et al., 2013) after introduction of Maxent (Phillips et al., 2006; Phillips & Dudík, 2008) in field of ecology. High predictive accuracy in comparison to many other methods (Elith et al., 2006) even with a small sample size (Hernandez, Graham, Master, & Albert, 2006; R.G. Pearson, Raxworthy, Nakamura, & Peterson, 2007) has made the model prevalent in the field of SDM. I ran ten replicates with 10000 background points and 70 per cent of data set aside for model building. The default setting for auto features and regularization multiplier was used for the analysis. For all the analysis in Maxent, I used a freely available software Maxent (Phillips et al., 2006) to run the model.

Information of a variable considered by one model is not considered by other models, therefore, it is advantageous to take a consensus of many such similar models to improve the predictive accuracy of models (Araújo & New, 2006; Bates & Granger, 1969). When making such an ensemble of the models, either a weighted or unweighted averages of candidate models can be done (Araújo & New, 2006). In this study, I used an unweighted mean of three models to calculate the ensemble model.

2.3.3. Model Evaluation

The predictive accuracy of a model for the intended use is tested using model evaluation (Allouche, Tsoar, & Kadmon, 2006; Araújo & Guisan, 2006). The assessment is important because it determines the reliability of model for the intended use (Guisan et al., 2006), compares alternative modelling techniques or effect of environmental or species data configuration on the model (Segurado & Araújo, 2004). I used the True Skill Statistics (TSS, Allouche et al., 2006) and Area Under Curve of the Receiver Operating Curve (Deleo, 1993; Fielding & Bell, 1997; Jiménez-Valverde, 2012) to assess the predictive accuracy of the models. To calculate these parameters for the ensemble model, I used the unweighted mean of the candidate models.

Estimate of model evaluation derived from data independent from the training data set is robust (Chatfield, 1995; Fielding & Bell, 1997). However, it is not always possible to collect new data (Chatfield, 1995). Therefore the collected data is split into training and validation data set, and model validation performed out of the "hold-out" data set (Fielding & Bell, 1997). I set aside 30% of observation dataset to validate the model using TSS and AUC under ROC.

TSS is a threshold dependent measure that accounts omission and commission error, unaffected by prevalence and size of validation data (Allouche et al., 2006). TSS ranges from -1 to +1 where +1 indicate perfect agreement and value of zero indicate a performance no better than random (Allouche et al., 2006).

AUC under ROC is a threshold independent evaluation measure that is obtained by plotting true positive fraction (sensitivity) value against false positive fraction (1-sensitivity) for all available threshold (Fielding & Bell, 1997). Because it does not depend on a threshold, the area under ROC can be considered an important index with a value between 0.5 and 1; a value of 0.5 represents no better than random discrimination, and 1 represents perfect discrimination (Deleo, 1993).

2.3.4. Variable Importance

I used the correlation test available in "sdm" package (Naimi & Araújo, 2016) to determine the importance of each variable in BRT and GLM. In this test, the variable under investigation is randomly permuted, and the correlation between the predicted value and permuted value is done. If the contribution of a variable is high, the correlation becomes lower as the prediction is affected more due to permutation (Naimi & Araújo, 2016; Thuiller, Lafourcade, Engler, & Araújo, 2009). The measure of variable importance is then expressed as "1-correlation" (Thuiller et al., 2009).

Also, I used the jack-knife procedure implemented in Maxent to determine the variable important for Maxent model. Maxent excludes one variable at a time to create different models, several univariate models using an individual variable and a model with all variables to compare the gain for each variable (Torres et al., 2010) to determine variable importance for each variable.

I used non-parametric scoring to find the variable importance in the ensemble model. For the purpose I scored the top-ranked variable a score of 11- the total number of variables used, descending to one with the least important variable for each model. The variable which was not used for a model was scored zero. I calculated the percentage contribution for each variable to determine the variable importance for ensemble model.

2.3.5. Population Genetics Analysis

I tested Hardy-Weinberg Equilibrium and linkage equilibrium for the assumption of population parameters for genetic study. Hardy-Weinberg Equilibrium states that the allele frequencies of a locus are constant unless gene flow, natural selection, or mutation occurs in the population. Although not conclusive the failure to meet the assumption conclude at least one of these forces to be acting on the population. Similarly, linkage equilibrium test confirms that the genes are linked to each other and assorted independently to the next generation. A failure to meet the assumption might conclude that the loci are not independent, and the inferences of the parameter cannot be conclusive. I used GENEPOP version 4.7.0 (Rousset, 2008) to test these assumptions. Also, I used GenAlEx 6.4 (Peakall & Smouse, 2006)- an MS Excel Addon for population genetics data analysis to calculate the number of alleles, expected heterozygosity and observed heterozygosity in the population as basic parameters of population genetics. All the tests were performed assuming a single population of *P. cretensis,* i.e. global test.

2.3.6. Cost Raster and Landscape Hypotheses

I created a set of hypotheses for variables that are important for the distribution of the species as indicated by different SDMs. For every variable to be analysed, I hypothesized the cost to be:

a. Inversely related for isothermality as greater isothermality means lesser relative oscillation and more favourable for gene flow.
b. Linearly related for altitude as greater altitude means the greater cost for gene flow due to terrain and inverse temperature condition.
c. *P. cretensis* is expected to have cost in increasing order from open space vegetation, grassland, forest, heterogenous agricultural area, agricultural land, urban and water bodies for dispersal and thus gene flow (Table 2-3).

Table 2-3 Cost Allocation for Candidate Environmental Variables

Allocated Cost	Isothermality	Altitude	Landcover	SDM
1	33-34	<450	Open/Bare Area	0.8-1
2	32-33	450-950	Grassland	0.6-0.8
3	31-32	950-1450	Forest Low Disturbed Agricultural Area	0.4-0.6
4	30-31	1450-1920	Highly Disturbed Agricultural Area	0.2-0.4
5	29-30	1920-2400	Built-up Water body	0-0.2

d. Greater probabilities of presence mean lesser cost in dispersal for probability map from SDM.

I classified the environmental variable to provide a cost of travelling across a class of environmental variable based on the inferred relationship between the variable and ability to disperse in a class (Table 2-3).

To test the hypothesis of the combined effect of variables, I used the ensemble model of the three SDM analysis (Section 2.3.3). The model outputs were stretched between 0 to 1, and the unweighted average of the output was used to ensemble the models. The values indicated the suitability index for the species; greater the value of the pixel, easier for an individual to traverse the pixel. The ensemble model was then classified according to the probability of presence thus giving empirical evidence of impedance cost.

2.3.7. Hypotheses on Gene flow and effective distance

Simultaneously with multiple hypotheses of cost raster, I also hypothesized the possible path of gene flow across the landscape. I used the two competing hypotheses to illustrate possible dispersal path of species and gene-flow, i.e. accumulated cost along least cost path and random multiple paths (Figure 2-2).

a. Accumulated Cost along Least Cost Path: Intuitively, least cost path refers to the most effective path of dispersal with least resistance for the dispersing individuals (Figure 2-2 a). The accumulated cost is uncorrelated to Euclidean distance between the points and thus a better predictor for gene flow than the length of the least cost path (Etherington & Holland, 2013). So, the accumulated cost was used as a measure of effective distance between two sites. The effective distance was calculated on ArcMap 10.6 using the “Cost Distance” function in Spatial Analysis Tool.
b. Random Multiple Path: Rather than following a single path, a gene can flow from one location to other location via various random path (Figure 2-2 b.). Analogous to electrical current (Chandra et al., 1997), the flow of gene can spread over the landscape with varying resistance between two points following multiple paths (McRae, 2006; McRae, Dickson, Keitt, & Shah, 2008). Based on the graph theory, the resistance distance as a measure of effective distance (McRae et al., 2008) can then be calculated as a sum of a sequence of resistance along a path in series and across multiple paths in parallel circuits (Equation 1).

$$1/R=1/R_1+1/R_2+1/R_n \text{..........................Equation 2}$$

The effective distance between the site impeded by the resistance of the landscape and multiple paths were calculated in program CIRCUITSCAPE (McRae, 2006; McRae et al., 2008).

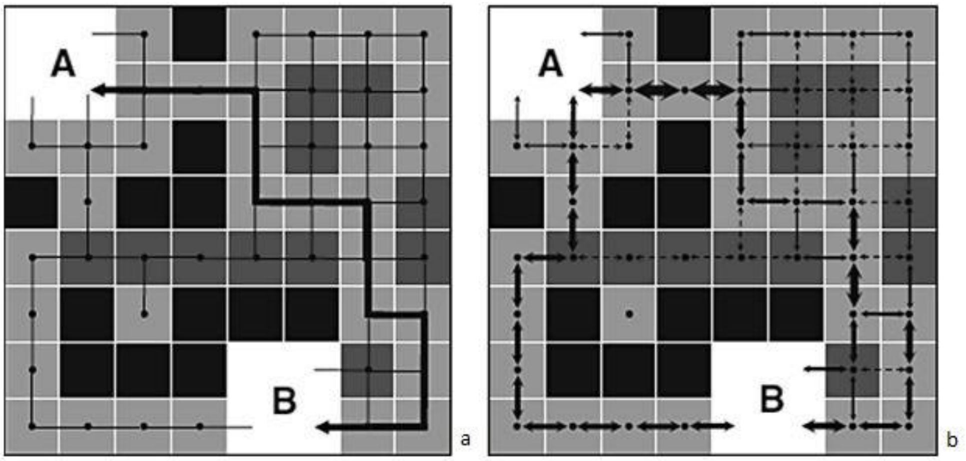


Figure 2-1: A comparison between Least Cost Path (a) and Resistance Distance (b) across a resistance layer (Darker colour represent greater resistance). Adapted from (Spear et al., 2010).

Although the two paths appear to be different, both paths are derived using the same resistance raster and complement each other (McRae, 2006). The least cost path can be considered as a special case of resistance distance when only one possible path exists between two sites.

2.3.8. Gene flow estimation

Because of the panmictic nature of the population in *P. cretensis* in Crete (Section 1.1.2), I used "Proportion of shared alleles (D_{ps})"-an individual based genetic distance that measures the proportion of similar alleles between two individuals (Wagner & Fortin, 2016), an indirect estimate of gene flow to estimate the gene flow. The genetic distance is calculated as $1\text{-}D_{ps}$. I used D_{ps} because it does not assume the equilibrium of population and thus relaxes assumptions of population equilibrium. I used the mean genetic distance of the individual pairs calculated by ten bootstrapping runs. I performed all the genetic distance analysis between individuals in Microsatellite Analyzer (MSA 4.05; Dieringer & Schlotterer, 2003). I used R Studio to average the genetic distance obtained from MSA.

2.3.9. Linking Landscape Distance and Genetic Distance

I used Mantels test and Partial Mantel Test to answer the multiple hypotheses on cost allocation, the path of gene flow on the effect of genetic distance. Mantel test (Mantel, 1967) is a linear correlation between two distance matrices which can be used to test the important variable for the gene flow in the landscape (Legrende & Fortin, 2010). Partial Mantel Test is a partial correlation accounting for a random variable which is also expected to be correlated to the genetic distance. I constrained the effect of geographic distance to be constant to test the effect of the variable only using Partial Mantel Test. I used "ecodist" package in R to calculate the Mantel and Partial Mantel Correlation Coefficient (r_M) between two matrices to test the variable important for restricting gene flow.

3. RESULTS

3.1. Species Distribution Model

3.1.1. Multicollinearity Test

VIF test of multicollinearity presented nine continuous variables that were the least correlated to each other (Table 3-1). Most of the bio-climatic variables were removed due to collinearity-only precipitation of driest quarter and iso-thermality were found to be below the threshold of significant correlation for climatic variables. Topographic variables remained the set of environmental predictors found to be the least affected by multicollinearity to each other (Table 3-1).

Table 3-1: Candidate Set of Environmental Variables after Multicollinearity Test

Environmental Variable	Variation Inflation Factor
Precipitation of driest quarter (Bio17)	7.03
Altitude	6.34
Iso-thermality (Bio3)	1.72
Annual NDVI	1.67
Distance to Road	1.52
Distance to River	1.42
Actual Evapotranspiration	1.28
Slope	1.26
North	1.01

3.1.2. Predicted Range and Current Range

3.1.2.1. GLM Model

The distribution range of *P. cretensis* was created with “fair” accuracy using GLM model (Table 3-2). The model predicted only a small area of the high suitability in the peninsula of Chania and south of it. The model predicted a few areas to be suitable for the species (Figure 3-1).

Table 3-2: Model Evaluation Statistics for all models analysed and ensemble model (the threshold for TSS is Sensitivity equals Specificity)

S. No	Model	TSS	AUC
1	GLM	0.47 (0.07)	0.78 (0.04)
2	BRT	0.66 (0.14)	0.81 (0.05)
3	Maxent	0.40 (0.21)	0.93 (0.04)
	Ensemble (Unweighted)	0.51 (0.14)	0.90 (0.04)

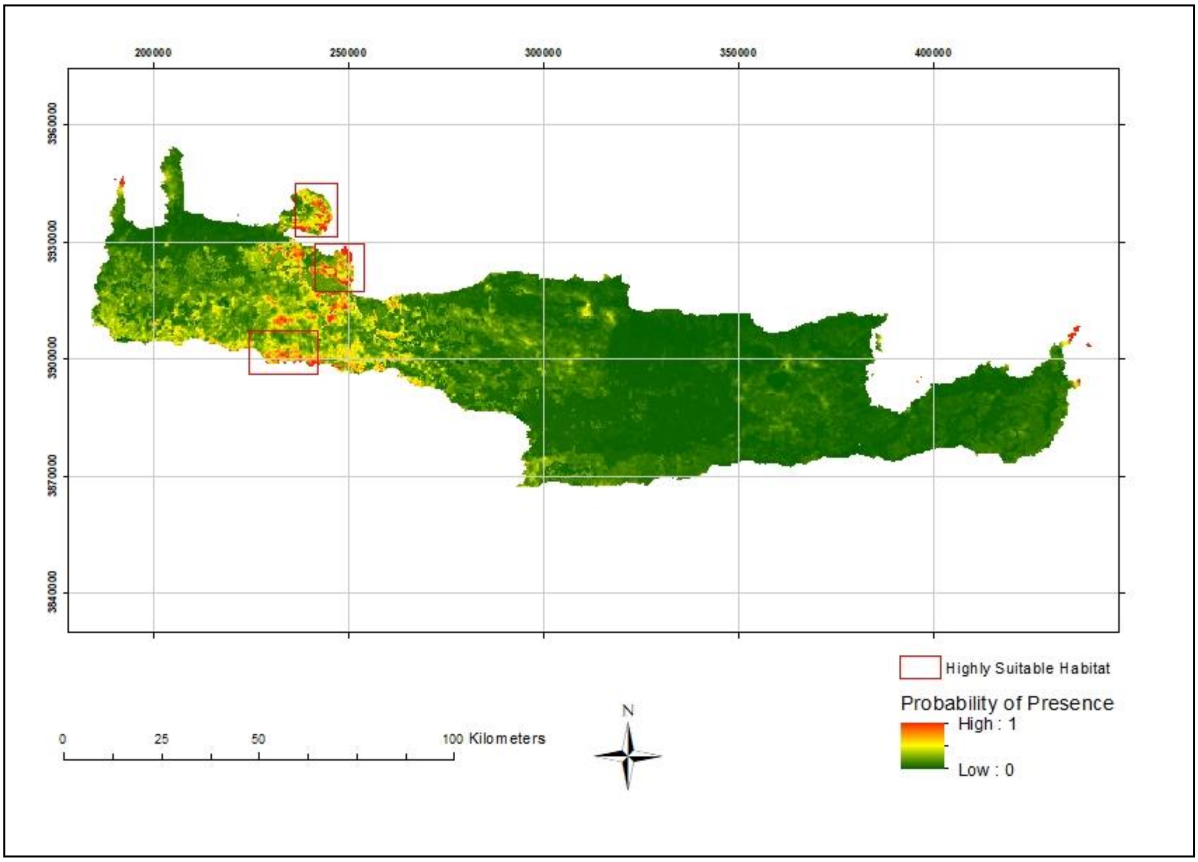


Figure 3-1: Model prediction by GLM

Test of variable importance showed isothermality to be the major variable to determine the distribution of the species. Distance to river and landcover followed isothermality were top three predictors of *P. cretensis* in Crete by GLM model (Figure 3-2)

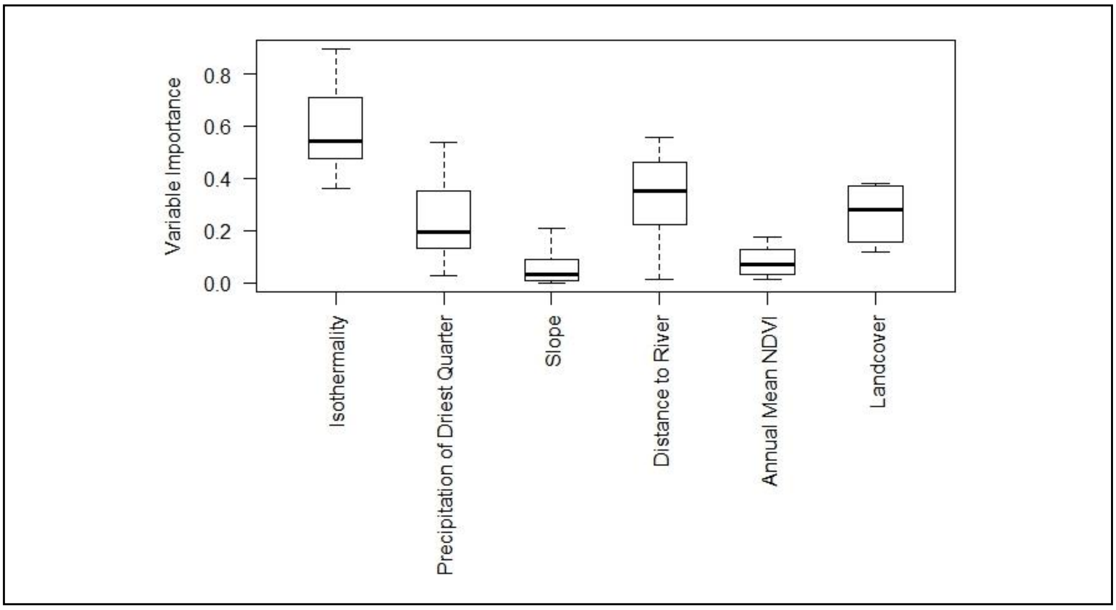


Figure 3-2: Measure of Variable Importance for GLM model.

3.1.2.2. BRT Model

The distribution range of *P. cretensis* was created with "good" accuracy using BRT model (Table 3-2). The model predicted the White Mountain to be the very suitable habitat of the species. Apart from the mountain, coastal regions on north east Crete was also predicted to be suitable habitat for *P. cretensis*. The model also predicted only a few areas to be suitable for the species (Figure 3-3). It also predicted a few moderately suitable habitat patches outside the current extent of *P. cretensis*.

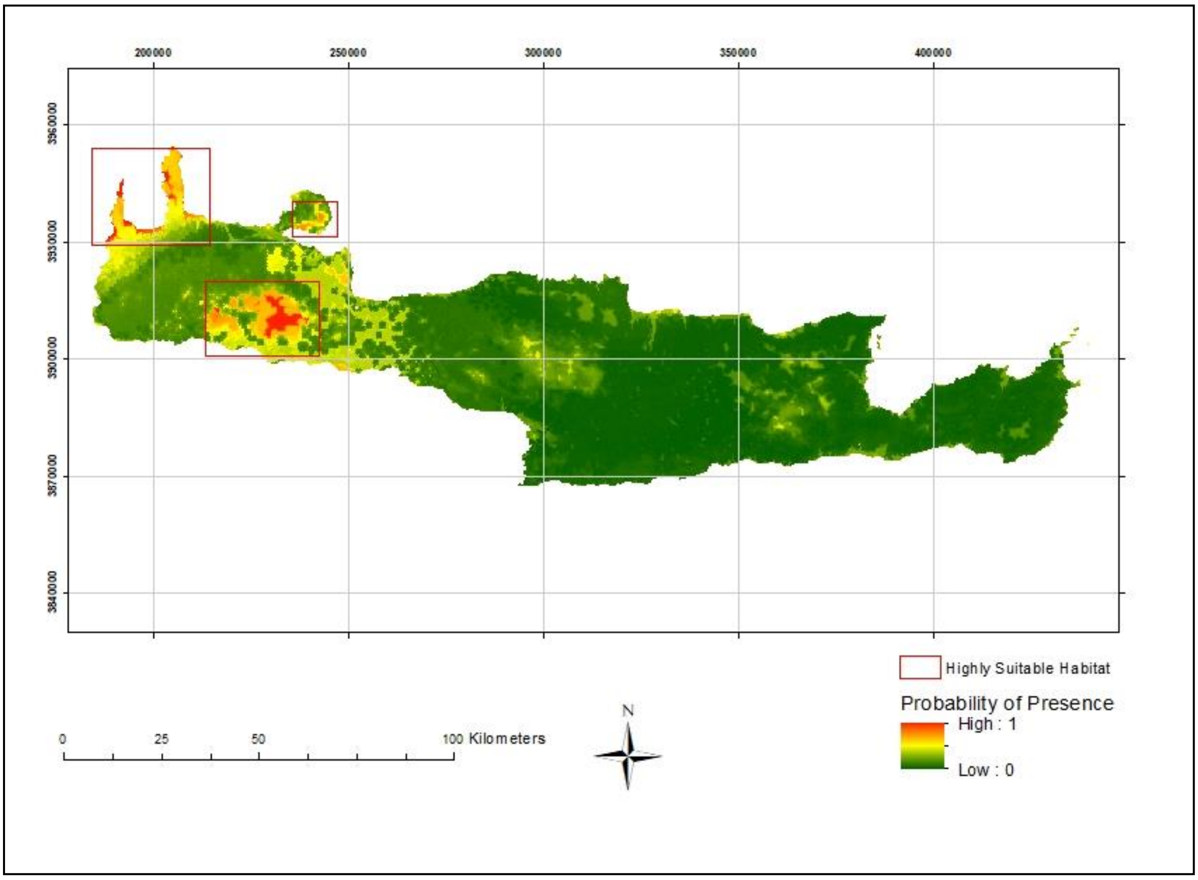


Figure 3-3: Model prediction by BRT

Test of variable importance showed actual evapotranspiration to be the major variable to determine the distribution of the species. Isothermality, altitude, and distance to river followed AET as top four predictors of *P. cretensis* in Crete by BRT model.

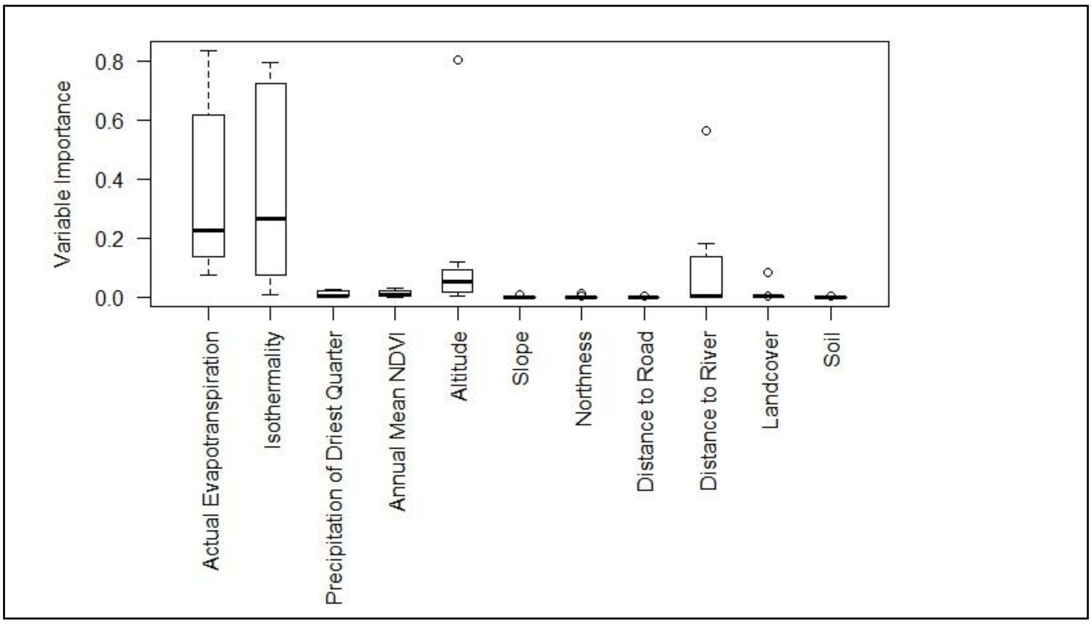


Figure 3-4: Measure of Variable Importance for BRT model.

3.1.2.3. Maxent Model

The distribution range of *P. cretensis* was created with "fair" accuracy using Maxent model (Table 3-2). The model predicted the White Mountain to be the very suitable habitat of the species. Apart from the mountain, the coastal regions on north east Crete and Chania were also predicted to be suitable habitat for *P. cretensis*. The model also predicted only a few areas to be suitable for the species (Figure 3-5). It also predicted a few moderately suitable habitat patches outside the current extent of *P. cretensis*.

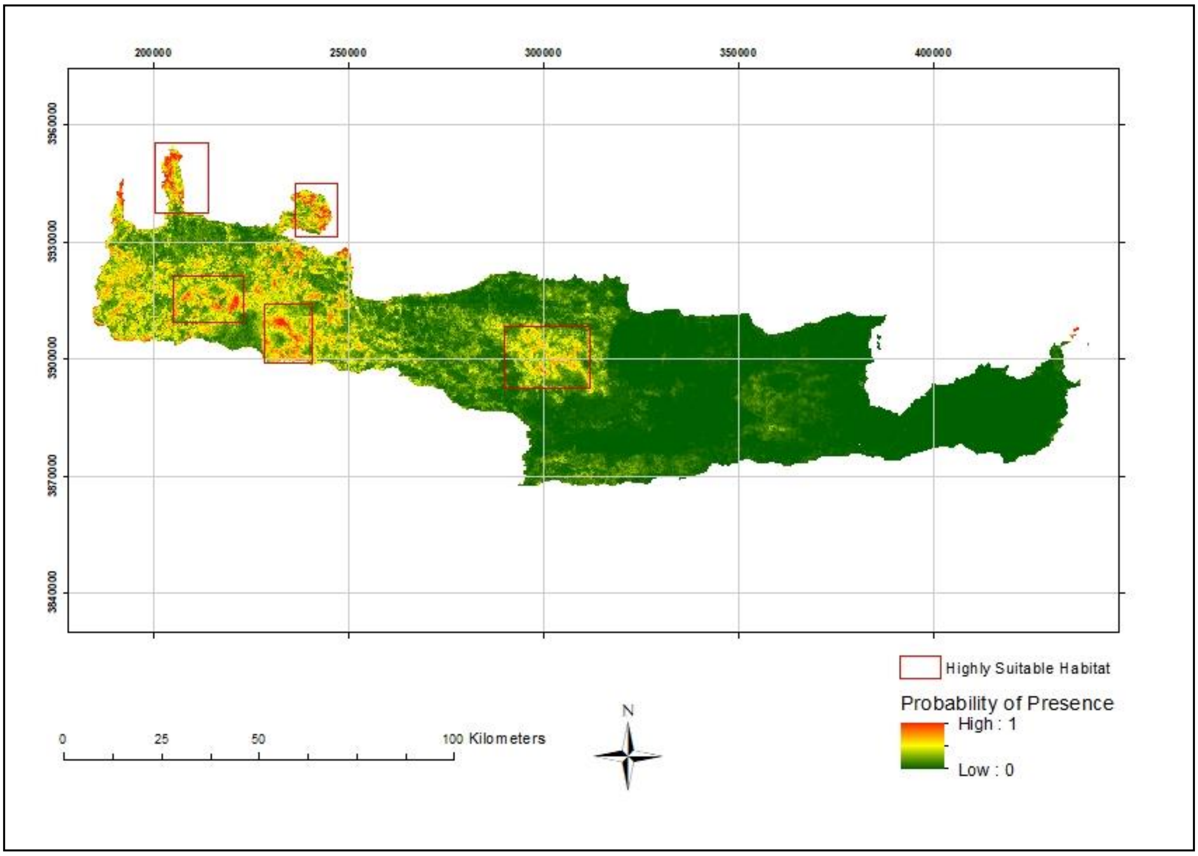


Figure 3-5: Model prediction by Maxent

Test of variable importance showed actual evapotranspiration to be the major variable to determine the distribution of the species. Isothermality, landcover, and altitude followed AET as top four predictors of *P. cretensis* in Crete by Maxent model.

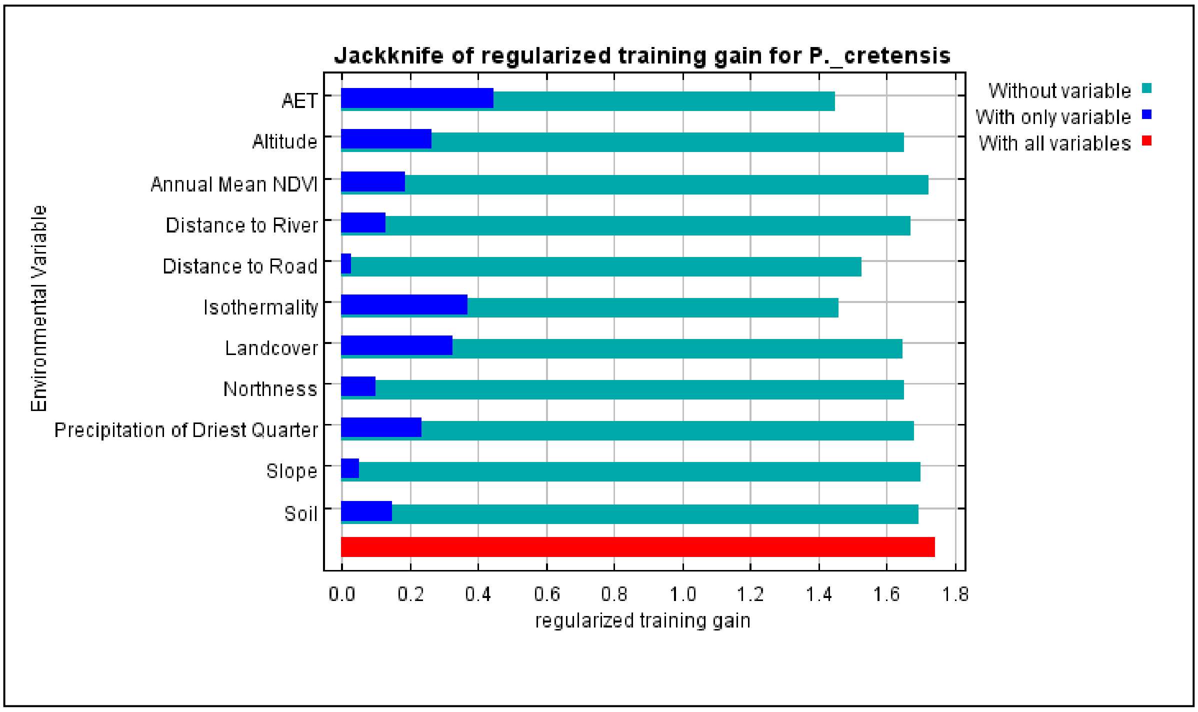


Figure 3-6: Measure of Variable Importance for Maxent model

3.1.2.4. Ensemble Model

The distribution range of *P. cretensis* was created with "fair" accuracy with an ensemble model (Table 3-2). The model predicted the White Mountain to be the very suitable habitat of the species. Apart from the mountain, the coastal region on north east Crete and Chania were also predicted to be suitable habitat for *P. cretensis*. The model also predicted only a few areas to be suitable for the species (Figure 3-7). It also predicted a few moderately suitable habitat patches outside the current extent of *P. cretensis*.

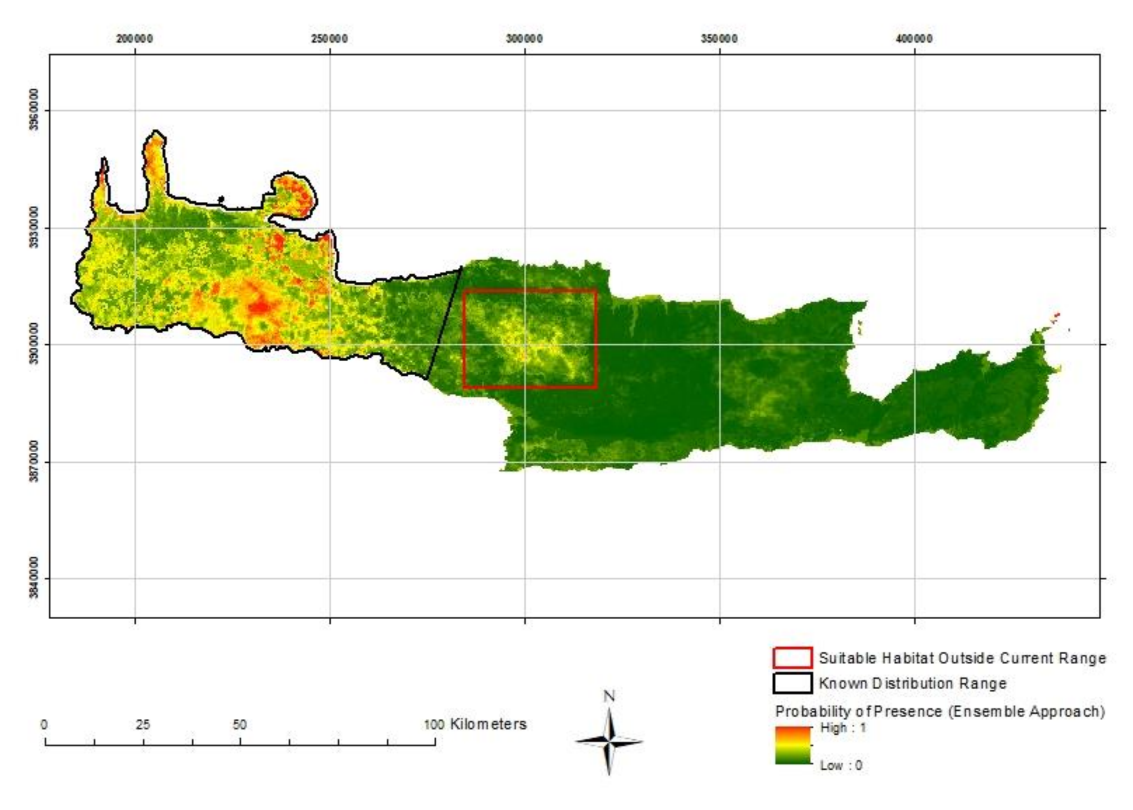


Figure 3-7: Ensemble Model Prediction overlaid with the current distribution range and suitable habitat outside the current range (the distribution on eastern islets are not considered)

The ensemble model predicted Isothermality to be the most important variable for determining the distribution of *P. cretensis*. Landcover and distance to river were the next contributing variables. AET with the other variables were top four contributing variables.

Table 3-3: Variable Importance Score in Ensemble Model

Variable	BRT Score	Maxent Score	GLM Score	Average Score	% Score
Isothermality	10	10	11	10.3	16.9
Distance to River	8	6	10	8	13.1
Landcover	6	9	9	8	13.1
AET	11	11	0	7.3	12
Annual Mean NDVI	7	4	7	6	9.8
Altitude	9	8	0	5.7	9.2
Precipitation of Driest Quarter	5	3	8	5.3	8.7
Slope	3	1	6	3.3	5.4
Northness	4	5	0	3	4.9
Distance to Road	2	7	0	3	4.9
Soil	1	2	0	1	1.6

3.2. Landscape Genetics Analysis

3.2.1. Descriptive Statistics for Population Genetics

Ninety-two individuals (Appendix 6) have at least 50% of genotypic information available across 13 analysed loci. All loci were found to be moderately to highly polymorphic with 10-27 alleles per locus (Table 3-3). Mean observed heterozygosity across all alleles was found to be 0.74 (S. E= 0.03) and mean expected heterozygosity was 0.9 (SE=0.007). Fisher's Test of Hardy-Weinberg Equilibrium (Chi sq.= 106.58, df=26, p<0.00) and linkage disequilibrium (57 out of 75 tests between loci were significant at p<0.05) suggested that there is evidence of deviation from the equilibrium.

Table 3-4: Descriptive statistics for samples loci (N is the number of samples, Na represent the number of alleles, Ne is the expected number of alleles, Ho is observed heterozygosity, and He represents expected Heterozygosity)

	B6	**C9**	**Lv3-19**	**Lv4-72**	**Pb10**	**Pli4**	**Pm16**	**Pm27**	**Pmeli19**	**Pmeli2**	**Pod1B**	**Pod2**	**Pod8**
N	83	89	92	86	91	90	60	88	90	90	89	92	90
Na	16	16	15	13	25	27	10	17	16	24	17	21	19
Ho	0.76	0.78	0.80	0.79	0.85	0.62	0.72	0.56	0.50	0.73	0.80	0.86	0.88
He	0.89	0.89	0.90	0.89	0.94	0.91	0.84	0.89	0.88	0.94	0.88	0.90	0.90

3.2.2. Influence of landscape variables and SDM on genetic distance

Mantels test on geographic distance and the genetic distance failed to show any isolation due to distance alone (Table 3-5). Various models were then run to determine the importance of selected variables to determine the gene flow *P. cretensis.* I modelled the accumulated cost within the least cost path (Figure 3-8a.) between sites to test the hypothesis if the least cost path for isothermality, landcover, altitude or a species distribution model better contributed for a pattern of genetic distance. However, accumulated cost along the least cost path did not have a significant association to the gene flow between sites when checked with or without the dependence of geographical distance on the variable (Table 3-5)

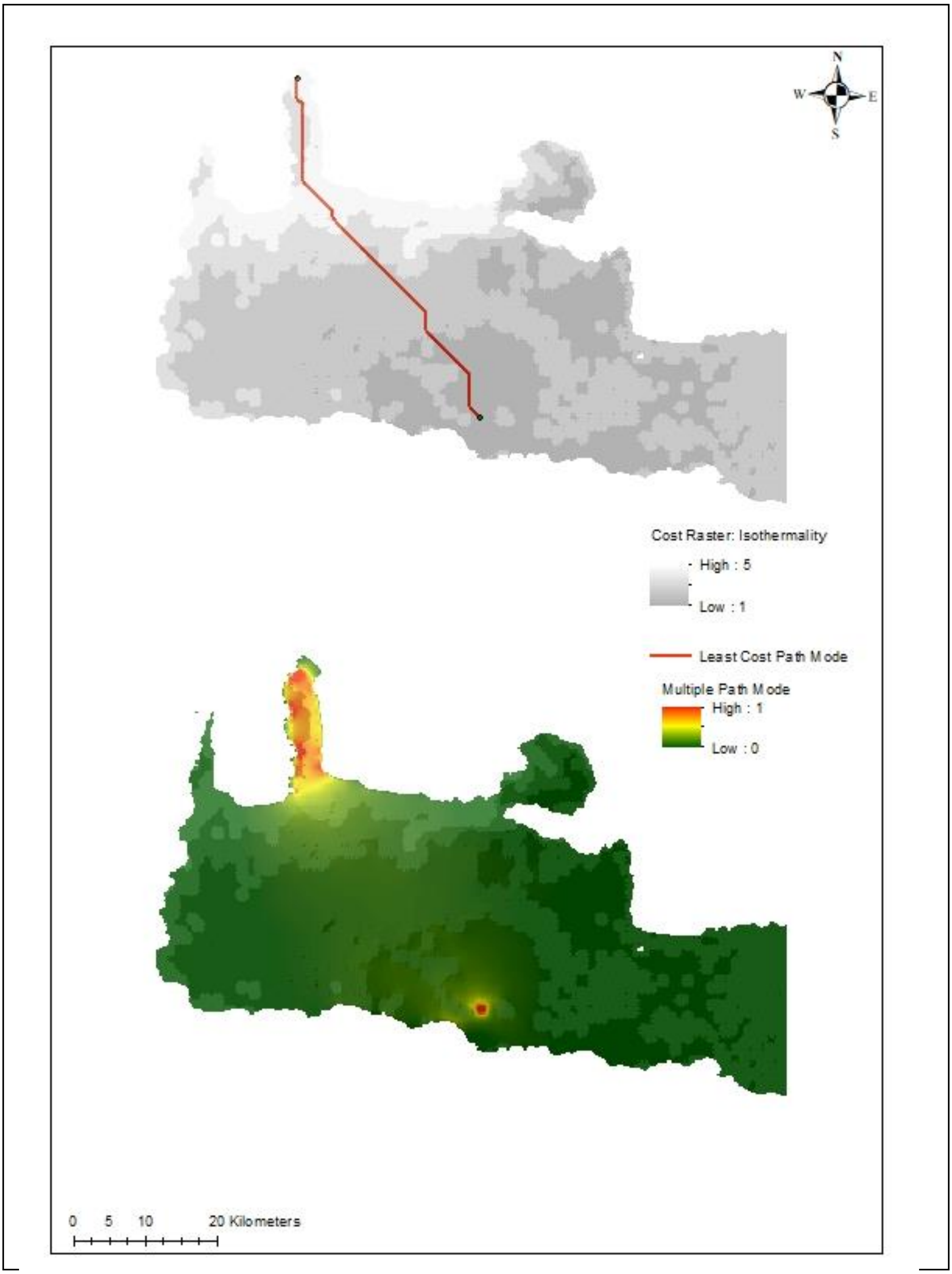


Figure 3-8: Modes of gene flow for *P. cretensis*: Least Cost Path (a) and the probability of movement along multiple paths (b). The cost raster (isothermality) in grey shades and one source and one destination is given.

The resistance model that explains the importance of possible multiple paths (Figure 3-8b.) to contribute to the gene flow also did not show any association for the selected variables. Only a marginal correlation between altitude and the gene flow was seen when the geographic distance did not add further cost to the altitude (Table 3-5). Also, the cost parameter derived from present Species Distribution Model failed to explain the gene flow of *P. cretensis* within the landscape in either accumulated cost along the least cost path or multiple path hypotheses (Table 3-5).

Table 3-5 Variable importance for gene flow

Variable	Test	Model	r_M	*p-value*
	Mantel Test	Genetic Distance~ Geographic Distance	-0.002	0.93
Isothermality	Mantel Test	Genetic Distance~ Accumulated Cost	-0.02	0.50
	Partial Mantel Test	Genetic Distance~ Accumulated Cost	-0.02	0.49
	Mantel Test	Genetic Distance~ Resistance Distance	-0.03	0.36
	Partial Mantel Test	Genetic Distance~ Resistance Distance	-0.04	0.28
Landcover	Mantel Test	Genetic Distance~ Accumulated Cost	-0.01	0.64
	Mantel Test	Genetic Distance~ Resistance Distance	-0.03	0.45
Altitude	Mantel Test	Genetic Distance~ Accumulated Cost	-0.02	0.40
	Partial Mantel Test	Genetic Distance~ Accumulated Cost	-0.05	0.13
	Mantel Test	Genetic Distance~ Resistance Distance	-0.05	0.12
	Partial Mantel Test	Genetic Distance~ Resistance Distance	-0.07	***0.04***
SDM	Mantel Test	Genetic Distance~ Accumulated Cost	-0.006	0.82
	Partial Mantel Test	Genetic Distance~ Accumulated Cost	-0.032	0.28
	Mantel Test	Genetic Distance~ Resistance Distance	-0.027	0.40
	Partial Mantel Test	Genetic Distance~ Resistance Distance	-0.039	0.24

4. DISCUSSION

4.1. Models for Range Limits Detection

Correlative approaches of environmental predictors and abundance, or presence and/or absence of species has been used to model the distribution of a species (Arundel, 2005; Cumming, 2002; Gaston, 2009; Herkt, 2007; Rotenberry, Preston, & Knick, 2006; Sexton et al., 2009; Zabalaga, 2008). Presence-absence type response data has received a wide application for these correlative approaches after the advent of ecological niche modelling (Sexton et al., 2009). In these modelling methods, the variables that have a significant effect on explaining the distribution range was considered as variables important for the range limit. Cumming, (2002) used a similar approach to explain the climatic variables that are contributing the range limits of 50 Ixodidae ticks to conclude climatic variables as a significant predictor of range limits over vegetation related variables. Similarly, Rotenberry et al. (2006) used the contribution of the environmental variable on PCA eigenvector of the predictor variables to determine a robust species range limits which could not have been obtained from environmental variables alone. Arundel (2005) explained the limiting environmental factor as the variable that significantly explained the difference between potential and release distribution- expanding the limits of a variable while meeting criteria for all other variables. For *P. cretensis,* Herkt (2007) used Maxent distribution model to explain its curious distribution attributed to vegetation and landcover.

Although the inference seems logical the impact of climatic variables only cannot be explained as a lone contributor for species range limits, rather an uncommon one. This notion has particularly given rise to a state sometimes called as “Range Boundary Disequilibrium” (Sexton et al., 2009) when a species range is more commonly limited by factors other than abiotic variables only. Demographic stochasticity or failure to adapt to the marginal environmental condition because of gene flow from other population might also contribute to limiting a species range limit (Alleaume-Benharira, Pen, & Ronce, 2006; Kirkpatrick & Barton, 1997). These models also neglect the importance of interspecific competition (Bullock, Edwards, Carey, & Rose, 2000; MacArthur, 1972) or predation (Bruelheide & Scheidel, 1999) which has been shown to limit the geographical range. Although the impact of competition has been modelled using SDM successfully (Cadena & Loiselle, 2007), it has not been routinely used to test the hypothesis of range limits (Sexton et al., 2009).

The correlative approach provides a likely mean of range limits, but it does not yield a mechanistic explanation for the process involved in the factors to limit the geographical range (Korner et al., 2016; Morin, Augspurger, & Chuine, 2007). Mechanistic models or process-based model addresses this deficiency of the former modelling approach (Chuine & Beaubien, 2001; Sexton et al., 2009) but are computationally intensive for many process parameter estimations and it is prone to have limited data availability (Dormann et al., 2012). Also, these models are unable to incorporate the effect of adaptation, migration and gene flow on the population that are an equally important factor of range limits (Gaston, 2009; Sexton et al., 2009) which is lately receiving interest in ecologist molecular biologists. Indeed, the gene flow has been shown to restrict the distribution in diverse taxon including plant (Angert, 2009; Herrera & Bazaga, 2008; Sexton et al., 2016), fish (Leclerc, Mailhot, Mingelbier, & Bernatchez, 2008), amphibians (Micheletti & Storfer, 2017), reptiles (Hagerty et al., 2011), and mammals (Wasserman, Cushman, Schwartz, & Wallin, 2010). As these models incorporate abiotic factors (e.g., climatic variables), process (dispersal) and their interaction (adaptation and micro-evolution), these models make them a better choice to model the range limits of a species.

4.2. SDM and Fundamental Niche of *P. cretensis*

Statistically, species distribution modelling predicts the environmental condition suitable to a species from a multi-dimensional environmental space based on the geographical location of the species. This prediction is then projected onto a geographical space to define the potential distribution of the species (Pearson, 2010). This multi-dimensional environmental space has a close affinity with the ecological niche-Fundamental Niche and Realized Niche (Hutchinson, 1957; Pearson, 2010). Deriving the environmental space unconstrained by the biotic interaction produces a "Fundamental or Grinellian Niche" of the species (Pulliam, 2000). "Realized or Hutchinsonian Niche" is produced when the sampled environmental space explicitly contains biotic factors like predation or competition (Hutchinson, 1957; Morin et al., 2007; Pulliam, 2000) which can affect the distribution of the species. The contrast is obvious; a model with biotic interaction produces the realized niche while excluding it produces the fundamental niche (Guisan & Thuiller, 2005). The contrast allows for the direct comparison if the species are in a state of equilibrium with the environment. In that sense, "Fundamental Niche" obtained from an ensemble of three different models (Figure 3-1) can be compared with 'known" realized niche (Figure 1-1). The visual comparison (Figure 3-7) showed that there is a discrepancy between what is the present range and potential range stating a disequilibrium between the environmental variables and the present distribution. The difference is enough to explain the insufficiency of the direct influence of only abiotic variables for the present range of the species. This statement contrast to the conclusion stated by Herkt (2007) which stated that the climatic variables are enough to explain the curious distribution of *P. cretensis*. Given the similar nature of the environmental variable and method used, the greater the sample size in the present study might have caused the difference. It has been shown that increasing sample size generally increases model performance and reduces variability in predictive accuracy (Hernandez et al., 2006; Stockwell & Peterson, 2002; Wisz et al., 2008). While Maxent has been shown to perform well with small sample size (Hernandez et al., 2006; Pearson et al., 2007), the greater sample size generally adds the predictive performance of Maxent (Hernandez et al., 2006) explaining the better performance predicted on the present study. Herkt (2007) included hyper-temporal NDVI as a vegetation index which was shown to have significant importance for predicting the distribution of the species. This variable, however, was not included in this study because the discrete variable had too many classes with only a few classes having enough sample size of presence point to provide accurate prediction (Jarnevich, Stohlgren, Kumar, Morisette, & Holcombe, 2015). The present study included Annual Mean NDVI as a vegetation index which did not appear to be a top contributor (Table 3-3) although it had some significance for the model.

4.2.1. Variable Importance for Distribution of *P. cretensis*

Selection of a candidate model for species distribution is dependent on the researcher's interpretation of the ecological response. Thus, it is possible that the variable importance might be different among different researchers. The present study estimated isothermality, land cover and distance to river to be the top-ranked three environmental predictors to explain the distribution of the species in an ensemble framework. Owing to the small size of islands, insular biodiversity cannot fully exploit the physiological limit of the climatic factors. The daily oscillation in relation to seasonal oscillation i.e. isothermality thus become the important climatic variable to define the range of insular species (Nix, 1986). The increase in isothermality represents lower seasonal oscillation relative to the daily oscillation or more stable temperature throughout the year preferred by cold-blooded reptile for a stable thermo-regulation. The way the landcover types were presented in the study was the direct impact of human on the landscape. It showed that the species least preferred the low disturbed agricultural area which might be related to the large extent of the vegetated area not suitable for the species. Open and less vegetated landcover type were the most preferred landcover type which again relates to its thermoregulation; mosaic vegetation and open area means enough microhabitat for basking and cooling off and protection from predators. Increased preference for the increased distance to river indicates its preference to dry-mesic to dry soil conditions. Of quick conservation concern is its low preference towards low disturbed agricultural practices like an olive plantation and heterogenous agricultural area. Although heterogenous agro-landscape are known to increase the regional biodiversity (Fahrig et al., 2011), the species responded negatively to this landcover type (Appendix 3,4 and 5) demonstrating the importance of species-specific conservation actions.

It is, however, important to be noted that the variable importance of the three different models was different, and so was the model predictions. The variable importance also changed across similar studies. Herkt (2007) found hyper temporal NDVI, land cover and soil to be an important predictor for *P. cretensis*. Zabalaga (2008) modelled the young clade of *P. cretensis* to find actual evapotranspiration to be the most important variable to predict the distribution which, however, the study was not done across the entire range of species. This discrepancy in the variable importance is not uncommon in species distribution modelling. Syphard & Franklin (2009) critically examined the difference in spatial predictions among species distribution to conclude that variable selection varied according to the model and most of the variable contribution can be weighted in favour of climatic variables. The explanatory power of variable increases as the predictor change slowly in space (i.e., high spatial autocorrelation) (Segurado, Araújo, & Kunin, 2006), a property which is much redundant in global climatic variables although all environmental predictors are spatially patterned (Syphard & Franklin, 2009).

4.2.2. SDM, errors, and uncertainty

The geographic sampling of the species presence is a subset of space and time of possible species occurrence (Jarnevich et al., 2015). The observation location might have been a moderately suitable or even unsuitable area for the species, yet the model represented it as a part of the equally probable sample in environmental space (Jarnevich et al., 2015). This might have introduced a risk of over-prediction of the species range. Conversely, the selected pseudo-absent are might have a viable population of the species yet be undetected (MacKenzie et al., 2002) thus giving false absence. The fundamental niche of species as explained by Hutchinson (1957) states the species responses to an n-dimensional hyperspace. To represent all these variables is impossible because there may be an infinite number of possible variables that may affect the distribution because of the limited availability of these variables or full understanding of species' ecology (Pearson, 2010). The error due to missing a variable can either have so significance or differ largely based on the importance of the unmodelled predictor (Harris et al., 2013). Of importance of these variables is incorporation of a sand fly distribution which might possibly cause leishmaniasis in *P. cretensis* and thus are important biological control for the species towards east (Lymberakis 2018, *pers. comm.*). Other variables that were not realized to be important during the study like heat load index, specific soil properties like percentage of clay, loam or silt, availability of basking rocks and cover from predators, etc might carry important local scale information that may contribute towards more accurate prediction of distribution.

Although these errors are prevalent in the model, the predictive accuracy of the model is enough to answer the question that the abiotic variables are not the probable reason to limit the range of *P. cretensis*. Alternatively, it is used as empirical evidence of cost parameterization in this study to hypothesize the impact of landscape on gene flow of *P. cretensis*. Jarnevich et al. (2015) also supported the use of these distribution models as a hypothesis to be tested and validated rather than considering it as a definitive explanation of the species distribution.

4.3. Environmental variables and gene flow

SDM and the environmental variables that were considered important for the distribution as predicted by SDM were tested if they were also important for the gene flow too. As evident from the result, it is unlikely that these variables had a significant contribution to the gene flow. Altitude only showed a marginal significance towards increased genetic similarity with increased distance between altitude. Although it seems unusual, the increased similarity on increased distance also called as "counter-gradient gene flow" (Sexton, Hangartner, & Hoffmann, 2014) is not uncommon. Strong directional gene flow along the variable (Kirkpatrick & Barton, 1997) can cause increased similarity towards the increased distance. To conclude the causation, however, is difficult as the significance might have occurred due to a random chance.

With the available evidence, it is probable that there is an unrestricted gene flow due to random mating (Sexton et al., 2014) provided that the environmental variables used are the only variables that are likely to affect the gene flow. This assumption looks unlikely because the environmental variables used here are more appropriate to model the species distribution at the regional scale and are unlikely to change across a large distance. The genetic distance is also very high among the individuals suggesting a clustered population structure with limited gene flow among the population. Therefore, it is more likely that the available environmental variables could not explain the resistance in gene flow. Although the insufficiency might be contributed to the spatial scale of the climatic variables, there can be some other uncertainty in the model that can be the cause of the inconclusive pattern of the gene flow in this study for local-scale variables like land cover.

4.3.1. Cost Parameterization on Landscape

Except for the SDM cost parameterization which is empirical, all the parameterization for the cost was done using the realized impact of the environmental variable on the cost of gene flow along the environmental gradient (Section 2.3.6). This parameterization technique is subjective and depends on the experience of the investigator and the literature on the biology of the species (Spear et al., 2010). Given the lack of literature on the cost impeded by environmental variables for the gene flow, this study can be considered as a starting point for solving these questions. Also, the only linear cost was assumed on this study which removed the possibility of unimodal cost response to be included in the model. Recent evidence has shown the unimodal cost might have significant evidence to explain the gene flow (Trumbo, Spear, Baumsteiiger, & Storfer, 2013). A failure to include non-linear relationship might also have caused the insignificance of the selected variables. The use of the discrete cost parameterization might also have caused the decreased support for the association as it has been shown that the significance level dropped down when the continuous variable was converted to a discrete variable (Cushman & Landguth, 2010)

4.3.2. Population equilibrium and genetic distance

The analysis and interpretation of measures of gene flow are dependent on the basic assumptions of population equilibrium like Hardy-Weinberg equilibrium of genotype frequency and linkage equilibrium that indicates independent assortment of alleles during mating (Waits & Storfer, 2016). Although this assumption can be relaxed by using a proportion of shared alleles as a measure of genetic distance (Micheletti & Storfer, 2017; Waits & Storfer, 2016), this genetic distance measure might not be as robust as when the population was in equilibrium.

Clustering of individuals into population might have fulfilled the basic assumption of population equilibrium within the population and some population-based genetic distance measure like F_{st} would have been a better way to explain the relationship between landscape and population. I attempted to use a genetic base of structuring population (Pritchard, Stephens, & Donnelly, 2000) without considering spatial configuration which could only determine two population of the species within the study area (Appendix 7), description of which is beyond the scope of the present study. Population structuring that uses both spatial and genetic data (Corander, Sirén, & Arjas, 2008) might have better clustering powers to hold the assumption of population genetics true for robust estimation.

5. CONCLUSION AND RECOMMENDATION

The main objective of the study was to model the distribution of *P. cretensis* to test if the range of the species was determined by the climatic variables only or was gene flow restricting its distribution. While I could not explain the variable important to restrict the gene flow, this study presents some of the crucial insights for the impact of environmental variables on gene flow.

The model with fair accuracy predicted that the species had not attained its state of equilibrium with the climatic condition explaining non-significance of climatic variables only to explain the distribution limits of the species. However, the distribution model presented some highly useful information on the ecology of *P. cretensis*. Isothermality as a temperature related variable was the most important variable as it relates to thermoregulation of the species. Land cover was found to be the second important variable with selective deterrence towards urban and heterogeneous agricultural landscape. The species was found away from the water source as consistent with its preferred dry habitat.

The genetic distance between individuals was high indicating a low level of gene flow in the landscape. The low level of gene flow also suggested a clustered population of the species. But the apparent cause of the limited gene flow could not be defined by the environmental variables tested for causality. Likely, the gene flow is affected by some finer scale environmental variable.

SDM has a broad application on the study of species range limit either directly using correlative or mechanistic approach and as empirical data for determining the cost of dispersal or gene flow in similar studies. It can also be the method to select the variable that might be considered as limiting factors for these studies when coupled with genetic tools. It is important to consider the ecological relevance of the environmental variable and parameterize this relevance into appropriate cost which adds the complexity of the study. If enough information on the ecology of the species can be obtained and the various hypothesis of cost and impact of environmental variables be made, it is possible to define the cryptic barrier of *P. cretensis*. Based on the present study, the following recommendations are put forward:

i. Environmental variables that influence micro-habitat of *P. cretensis* like heat load index, specific soil properties like percentage of clay, loam or silt, availability of basking rocks and cover from predators, etc might be variables that can be defined using geo-information or remote sensing across a large landscape can be used to better predict the distribution of *P. cretensis*.
ii. Cost parameterization of a variable is an important factor in determining the restriction impeded by the variable in gene flow. Increased expert opinion can help to define an improved parameterization. Use of response curves from SDM might help to provide important information on the cost of dispersal.
iii. Models of gene flow incorporating barriers with infinite cost like gorges, roads or urban areas might present the importance of physical barrier to restrict gene flow if they are the likely cause.
iv. Use of spatial clustering techniques can help define the population to remove the disequilibrium among population allowing the use of robust distance measures.
v. A robust test of the gene under selection pressure may be appropriate to remove such loci to provide a robust estimate of gene flow.

LIST OF REFERENCES

Adriaensen, F., Chardon, J. P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., & Matthysen, E. (2003). The application of "least-cost" modelling as a functional landscape model. *Landscape and Urban Planning*, *64*(4), 233–247. https://doi.org/10.1016/S0169-2046(02)00242-6

Alleaume-Benharira, M., Pen, I. R., & Ronce, O. (2006). Geographical patterns of adaptation within a species ' range : interactions between drift and gene flow. *Journal of Evolutionary Biology*, *19*, 203–215. https://doi.org/10.1111/j.1420-9101.2005.00976.x

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Amos, W., Hoffman, J. I., Frodsham, A., Zhang, L., Best, S., & Hill, A. V. S. (2007). Automated binning of microsatellite alleles : problems and solutions. *Molecular Ecology Notes*, *7*, 10–14. https://doi.org/10.1111/j.1471-8286.2006.01560.x

Angert, A. L. (2009). The niche, limits to species' distributions, and spatiotemporal variation in demography across the elevation ranges of two monkeyflowers. *Proceedings of the National Academy of Sciences*, *106*, 19693–19698. https://doi.org/10.1073/pnas.0901652106

Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, *33*, 1677–1688. https://doi.org/10.1111/j.1365-2699.2006.01584.x

Araújo, M. B., & New, M. (2006). Ensemble forecasting of species distributions. *Ensemble Forecasting of Species Distribution*, *22*, 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Araújo, M. B., & Pearson, R. G. (2005). Equilibrium of species' distributions with climate. *Ecography*, *28*, 693–695.

Arundel, S. T. (2005). Using spatial models to establish climatic limiters of plant species' distributions. *Ecological Modelling*, *182*, 159–181. https://doi.org/10.1016/j.ecolmodel.2004.07.016

Austin, M. . (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, *157*, 101–118. https://doi.org/10.1016/S0304-3800(02)00205-3

Austin, M. P., Nicholls, A. O., & Margules, C. R. (1990). Measurement of the Realized Qualitative Niche : Environmental Niches of Five Eucalyptus Species. *Ecological Monographs*, *60*, 161–177.

Bates, J. M., & Granger, C. W. J. (1969). The Combination of Forecasts. *Operational Research Quarterly*, *20*, 451–468.

Brehme, C. S., Hathaway, S. A., & Fisher, R. N. (2018). An objective road risk assessment method for multiple species : ranking 166 reptiles and amphibians in California. *Landscape Ecology*, *33*, 911–935. https://doi.org/10.1007/s10980-018-0640-1

Breiman, L. (2001). Statistical Modeling : The Two Cultures. *Statistical Science*, *16*, 199–215.

Bridle, J. R., & Vines, T. H. (2007). Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology & Evolution*, *22*(3), 140–147. https://doi.org/10.1016/j.tree.2006.11.002

Briers, R. A. (2003). Range limits and parasite prevalence in a freshwater snail. *Proceedings of the Royal Society of London B*, *270*, 178–180. https://doi.org/10.1098/rsbl.2003.0046

Bruelheide, H., & Scheidel, U. (1999). Slug herbivory as a limiting factor for the geographical range of Arnica montana. *Journal of Ecology*, *87*, 839–848.

Bruford, M. W., Cheesman, D. J., Coote, T., Green, H. A. A., Haines, S. A., O'Ryan, C., & Williams, T. R. (1996). Microsatellites and their application to conservation genetics. In T. B. Smith & R. K. Wayne (Eds.), *Molecular Genetic Approaches in Conservation* (pp. 278–297). New York: Oxford University Press.

Bruford, M. W., Hanotte, O., & Burke, T. (1998). Multi and single-locus DNA fingerprinting. In A.. Hoelzel (Ed.), *Molecular genetic analysis of populations: a practical approach* (pp. 225–269). IRL Press.

Bullock, J. M., Edwards, R. J., Carey, P. D., & Rose, R. J. (2000). Geographical separation of two Ulex species at three spatial scales : does competition limit species ' ranges ? *Ecography*, *23*, 257–271.

Cadena, C. D., & Loiselle, B. A. (2007). Limits to elevational distributions in two species of emberizine finches : disentangling the role of interspecific competition, autoecology , and geographic variation in the environment. *Ecography*, 491–504. https://doi.org/10.1111/j.2007.0906-7590.05045.x

Case, T. J., Holt, R. D., McPeek, M. A., & Keitt, T. H. (2005). The Community Context of Species ' Borders : Ecological and Evolutionary Perspectives of species ' borders : The community context ecological evolutionary perspectives. *Oikos*, *108*, 28–46.

Chandra, A. K., Raghavan, P., Ruzzo, W. L., Smolensky, R., & Tiwari, P. (1997). The electrical resistance of a graph captures its commute and cover time. *Computational Complexity*, *6*, 312–340. https://doi.org/10.1007/BF01270385

Chatfield, C. (1995). Model Uncertainty, Data Mining, and Statistical Inference. *Journal of the Royal Statistical Society. Series A*, *158*, 419–466.

Chuine, I., & Beaubien, E. G. (2001). Phenology is a major determinant of tree species range. *Ecology Letters*, *4*, 500–510.

Clark, D. P. (2005). *Molecular Biology*. Elsevier Academic Press.

Corander, J., Sirén, J., & Arjas, E. (2008). Bayesian spatial modeling of genetic population structure. *Computational Statistics*, *23*(1), 111–129. https://doi.org/10.1007/s00180-007-0072-x

Cumming, G. S. (2002). Comparing climate and vegetation as limiting factors for species ranges of African ticks. *Ecology*, *83*, 255–268.

Cushman, S. A., & Landguth, E. L. (2010). Scale-dependent inference in landscape genetics. *Landscape Ecology*, *25*, 967–979. https://doi.org/10.1007/s10980-010-9467-0

Darwin, C. (1859). *On The Origin of Species by means of Natural Selection*. D. Appleton and Company, 443 and 445 Broadway.

De'Ath, G., & Fabricius, K. (2000). Classification and Regression Trees : A Powerful Yet Simple Technique for Ecological Data Analysis. *Ecology*, *81*, 3178–3192. https://doi.org/10.2307/177409

Deleo, J. M. (1993). Receiver operating characteristic laboratory (ROCLAB): software for developing decision strategies that account for uncertainty. In *Proceedings of the Second International Symposium on Uncertainty Modelling and Analysis* (pp. 318–325). College Park, MD: IEEE Computer Society Press.

DeRivera, C. E., Ruiz, G. M., Hines, A. H., & Jivoff, P. (2005). Biotic resistance to invasion: native predator limits abundance and distribution of an introduced crab. *Ecology*, *86*, 3364–3376.

Dieckmann, U., & Doebeli, M. (1999). On the origin of species by sympatric speciation. *Nature*, *400*, 354–357.

Dieringer, D., & Schlotterer, C. (2003). microsatellite analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, *3*, 167–169. https://doi.org/10.1046/j.1471-8286

Doebeli, M., & Dieckmann, U. (2003). Speciation along environmental gradients. *Nature*, *421*, 259–264. https://doi.org/10.1038/nature01312.Published

Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., … Singer, A. (2012). Correlation and process in species distribution models : bridging a dichotomy. *Journal of Biogeography*, *39*, 2119–2131. https://doi.org/10.1111/j.1365-2699.2011.02659.x

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, *17*, 1170–1188. https://doi.org/10.1111/j.1365-294X.2007.03659.x

Elith, J., Graham, C. H., Anderson, R. P., Dudık, M., Ferrier, S., Guisan, A., … Zimmermann, N. E. (2006). Novel methods improve prediction of species ' distributions from occurrence data. *Ecography*, *29*, 129–151.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Etherington, T. R., & Holland, E. P. (2013). Least-cost path length versus accumulated-cost as connectivity measures. *Landscape Ecology*, *28*, 1223–1229. https://doi.org/10.1007/s10980-013-9880-2

Fahrig, L., Baudry, J., Brotons, L., Burel, F. G., Crist, T. O., Fuller, R. J., … Martin, J. L. (2011). Functional landscape heterogeneity and animal biodiversity in agricultural landscapes. *Ecology Letters*, *14*, 101–112. https://doi.org/10.1111/j.1461-0248.2010.01559.x

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

Franklin, J., & Miller, A. J. (2009). *Mapping species distributions: Spatial inference and prediction*. Cambridge University Press.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, *28*, 337–407.

Gaston, K. J. (2009). Geographic range limits : achieving synthesis. *Proceedings of the Royal Society B*, *276*, 1395–1406. https://doi.org/10.1098/rspb.2008.1480

Gaylord, B., & Gaines, S. D. (2000). Temperature or Transport ? Range Limits in Marine Species Mediated Solely by Flow. *The American Naturalist*, *155*, 769–789.

Goldberg, E. E., & Lande, R. (2006). Ecological and reproductive character displacement on an environmetal gradient. *Evolution*, *60*, 1344–1357.

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Statistical Reports*, *84*, 2809–2815.

Grove, A. T., & Rackham, O. (1993). Threatened landscapes in the Mediterranean: examples from Crete. *Landscape and Urban Planning*, *24*, 279–292. https://doi.org/10.1016/0169-2046(93)90107-O

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions : Setting the scene. *Ecological Modelling*, *157*, 89–100.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C. C., Aspinall, R., & Hastie, T. (2006). Making better biogeographical predictions of species ' distributions. *Journal of Applied Ecology*, *43*, 386–392. https://doi.org/10.1111/j.1365-2664.2006.01164.x

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*, 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

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

Gutiérrez, N. L., & Defeo, O. (2005). Spatial patterns in population dynamics of the scallop *Psychrochlamys patagonica* at the northern edge of its range. *Journal of Shellfish Research*, *24*, 877–882.

Hagerty, B. E., Nussear, K. E., Esque, T. C., & Tracy, C. R. (2011). Making molehills out of mountains: Landscape genetics of the Mojave desert tortoise. *Landscape Ecology*, *26*, 267–280. https://doi.org/10.1007/s10980-010-9550-6

Harris, R. M. B., Porfirio, L. L., Hugh, S., Lee, G., Bindoff, N. L., Mackey, B., & Beeton, N. J. (2013). To be or not to be? Variable selection can change the projected fate of a threatened species under future climate. *Ecological Management and Restoration*, *14*, 230–234. https://doi.org/10.1111/emr.12055

Herkt, M. (2007). *Modelling Habitat Suitability To Predict The Potential Distribution Of Erhard's Wall Lizard Podarcis erhardii On Crete*. University of Twente. MSc Thesis.

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

Herrera, C. M., & Bazaga, P. (2008). Adding a third dimension to the edge of a species' range: Altitude and genetic structuring in mountainous landscapes. *Heredity*, *100*, 275–285. https://doi.org/10.1038/sj.hdy.6801072

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/joc.1276

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: Defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics*, *10*(9), 639–650. https://doi.org/10.1038/nrg2611

Holt, R. D. (2003). On the evolutionary ecology of species' ranges. *Evolutionary Ecology Research*, *5*, 159–178.

Holt, R. D., Keitt, T. H., Lewis, M. A., Maurer, B. A., & Taper, M. L. (2005). Theoretical models of species' borders : single species approaches. *Oikos*, *108*, 18–27.

Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, *22*, 415–427. https://doi.org/10.1101/SQB.1957.022.01.039

IUCN. (2012). *IUCN Red List categories and criteria* (Second). Gland, Switzerland and Cambridge, UK.

Jarnevich, C. S., Stohlgren, T. J., Kumar, S., Morisette, J. T., & Holcombe, T. R. (2015). Ecological Informatics Caveats for correlative species distribution modeling. *Ecological Informatics*, *29*, 6–15. https://doi.org/10.1016/j.ecoinf.2015.06.007

Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species. *Global Ecology and Biogeography*, *21*, 498–507. https://doi.org/10.1111/j.1466-8238.2011.00683.x

Kirkpatrick, M., & Barton, N. H. (1997). Evolution of a Species' Range. *The American Naturalist*, *150*, 1–23.

Korner, C., Basler, D., Hoch, G., Kollas, C., Lenz, A., Randin, C. F., … Zimmermann, N. E. (2016). Where, why and how ? Explaining the low-temperature range limits of temperate tree species. *Journal of Ecology*, *104*, 1076–1088. https://doi.org/10.1111/1365-2745.12574

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill Irwin.

Leclerc, É., Mailhot, Y., Mingelbier, M., & Bernatchez, L. (2008). The landscape genetics of yellow perch (*Perca flavescens*) in a large fluvial ecosystem. *Molecular Ecology*, *17*, 1702–1717.

https://doi.org/10.1111/j.1365-294X.2008.03710.x

Legrende, P., & Fortin, M.-J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, *10*, 831–844. https://doi.org/10.1111/j.1755-0998.2010.02866.x

Louthan, A. M., Doak, D. F., & Angert, A. L. (2015). Where and When do Species Interactions Set Range Limits? *Trends in Ecology and Evolution*, *30*, 780–792. https://doi.org/10.1016/j.tree.2015.09.011

Lymberakis, P. (2009). Podarcis cretensis. *The IUCN Red List of Threatened Species 2009*, *8235*. Retrieved from http://www.iucnredlist.org/details/157252/0

Lymberakis, P., & Poulakakis, N. (2010). Three Continents Claiming an Archipelago: The Evolution of Aegean's Herpetofaunal Diversity. *Diversity*, *2*, 233–255. https://doi.org/10.3390/d2020233

Lymberakis, P., Poulakakis, N., Kaliontzopoulou, A., Valakos, E., & Mylonas, M. (2008). Two new species of *Podarcis* (Squamata; Lacertidae) from Greece. *Systematics and Biodiversity*, *6*, 307–318. https://doi.org/10.1017/S1477200008002727

MacArthur, R. H. (1972). *Geographic Ecology: Patterns in the Distribution of Species*. Harper and Row, Publishers, Inc.,

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

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, *18*, 189–197. https://doi.org/10.1016/S0169-5347(03)00008-9

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*, 209–220.

Marquardt, D. W. (1970). Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics*, *12*, 591–612.

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models, Second Edition*. Taylor & Francis.

McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*, 1551–1561.

McRae, B. H., Dickson, B. G., Keitt, T. H., & Shah, V. B. (2008). Using Circuit Theory To Model Connectivity in Ecology, Evolution, and Conservation. *Ecology*, *89*, 2712–2724. https://doi.org/10.1890/07-1861.1

Mehlman, D. W. (1997). Change in avian abundance across the geographical range in response to environmental change. *Ecological Applications*, *7*, 614–624.

Micheletti, S. J., & Storfer, A. (2015). A test of the central-marginal hypothesis using population genetics and ecological niche modelling in an endemic salamander (*Ambystoma barbouri*). *Molecular Ecology*, *24*, 967–979. https://doi.org/10.1111/mec.13083

Micheletti, S. J., & Storfer, A. (2017). An approach for identifying cryptic barriers to gene flow that limit species' geographic ranges. *Molecular Ecology*, *26*, 490–504. https://doi.org/10.1111/mec.13939

Morin, X., Augspurger, C., & Chuine, I. (2007). Process-based modelling of species distribution: What limits temperate tree species' range boundaries? *Ecology*, *88*, 2280–2291.

Naimi, B., & Araújo, M. B. (2016). SDM: A reproducible and extensible R platform for species distribution modelling. *Ecography*, *39*, 368–375. https://doi.org/10.1111/ecog.01881

Nix, H. (1986). A biogeographic analysis of Australian elapid snakes. In R. Longmore (Ed.), *Atlas of Elapid Snakes of Australia. Australian Flora and Fauna 7th series* (pp. 4–15). Canberra: Australian Government Publishing Service.

Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, *4*, 347–354.

Panagos, P., Christos, K., Cristiano, B., & Ioannis, G. (2014). Seasonal monitoring of soil erosion at the regional scale: An application of the G2 model in Crete focusing on agricultural land uses. *International Journal of Applied Earth Observation and Geoinformation*, *27*, 147–155. https://doi.org/10.1016/j.jag.2013.09.012

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*, 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Pearson, H. (2006). What is a gene? *Nature*, *441*, 399–401.

Pearson, R. G. (2010). Species ' Distribution Modeling for Conservation Educators and Practitioners. *Lessons in Conservation*, *3*, 54–89.

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Peterson, A. T. (2007). Predicting species distributions from small numbers of occurrence records : a test case using cryptic geckos in Madagascar predicting species distributions from small numbers of occurrence records : a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, *34*, 102–117. https://doi.org/10.1111/j.1365-2699.2006.01594.x

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., & Dudík, M. (2008). Modeling of species distribution with Maxent: new extensions and a comprehensive evaluation. *Ecograpy*, *31*, 161–175. https://doi.org/10.1111/j.2007.0906-7590.05203.x

Polechová, J., & Barton, N. H. (2015). Limits to adaptation along environmental gradients. *PNAS*, *112*, 6401–6406. https://doi.org/10.1073/pnas.1421515112

Pometti, C., Bessega, C., Cialdella, A., Ewens, M., Saidman, B., & Vilardi, J. (2018). Spatial genetic structure within populations and management implications of the South American species *Acacia aroma* (Fabaceae), 1–14.

Poulakakis, N., Lymberakis, P., Antoniou, A., Chalkia, D., Zouros, E., Mylonas, M., & Valakos, E. (2003). Molecular phylogeny and biogeography of the wall-lizard *Podarcis erhardii* (Squamata: Lacertidae). *Molecular Phylogenetics and Evolution*, *28*, 38–46. https://doi.org/10.1016/S1055-7903(03)00037-X

Poulakakis, N., Lymberakis, P., Valakos, E., Zouros, E., & Mylonas, M. (2005). Phylogenetic relationships and biogeography of Podarcis species from the Balkan Peninsula , by Bayesian and maximum likelihood analyses of mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, *37*, 845–857. https://doi.org/10.1016/j.ympev.2005.06.005

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population Structure using multilocus genotype data. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Pulliam, H. R. (2000). On the relationship between niche and distribution. *Ecology Letters*, *3*, 349–361. https://doi.org/10.1046/j.1461-0248.2000.00143.x

Rakham, O., & Moody, J. (1997). *The Making of the Cretan Landscape*. Manchester University Press.

Rotenberry, J. T., Preston, K. L., & Knick, S. T. (2006). GIS-Based niche modeling for mapping species' habitat. *Ecology*, *87*, 1458–1464.

Rousset, F. (2008). GENEPOP ' 007 : a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, *8*, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Schurr, F. M., Pagel, J., Cabral, J. S., Groeneveld, J., Bykova, O., O'Hara, R. B., … Zimmermann, N. E. (2012). How to understand species ' niches and range dynamics : a demographic research agenda for biogeography. *Journal of Biogeography*, *39*, 2146–2162. https://doi.org/10.1111/j.1365-2699.2012.02737.x

Segurado, P., & Araújo, M. B. (2004). An evaluation of methods for modelling species distributions. *Journal of Biogeography*, *31*, 1555–1568.

Segurado, P., Araújo, M. B., & Kunin, W. E. (2006). Consequences of spatial autocorrelation for niche-based. *Journal of Applied Ecology*, *43*, 433–444. https://doi.org/10.1111/j.1365-2664.2006.01162.x

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, *68*, 1–15. https://doi.org/10.1111/evo.12258

Sexton, J. P., Hufford, M. B., Bateman, A. C., Lowry, D. B., Meimberg, H., Strauss, S. Y., & Rice, K. J. (2016). Climate structures genetic variation across a species' elevation range: A test of range limits hypotheses. *Molecular Ecology*, *25*, 911–928. https://doi.org/10.1111/mec.13528

Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and Ecology of Species Range Limits. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 415–436. https://doi.org/10.1146/annurev.ecolsys.110308.120317

Spear, S. F., Balkenhol, N., Fortin, M. J., McRae, B. H., & Scribner, K. (2010). Use of resistance surfaces for landscape genetic studies: Considerations for parameterization and analysis. *Molecular Ecology*, *19*, 3576–3591. https://doi.org/10.1111/j.1365-294X.2010.04657.x

Spear, S. F., Cushman, S. A., & McRae, B. H. (2016). Resistance Surface Modeling in Landscape Genetics. In N. Balkenhol, S. A. Cushman, A. Storfer, & L. P. Waits (Eds.), *Landscape Genetics: Concepts, Methods, Application* (pp. 129–148). Wiley Blackwell.

Stockwell, D. R. ., & Peterson, A. T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, *148*, 1–13. https://doi.org/10.1016/S0304-3800(01)00388-X

Stockwell, D. R. B. (2006). Improving ecological niche models by data mining large environmental datasets for surrogate models. *Ecological Modelling*, *192*, 188–196. https://doi.org/10.1016/j.ecolmodel.2005.05.029

Storfer, A., Murphy, M. A., Spear, S. F., Holderegger, R., & Waits, L. P. (2010). Landscape genetics: Where are we now? *Molecular Ecology*, *19*, 3496–3514. https://doi.org/10.1111/j.1365-294X.2010.04691.x

Syphard, A. D., & Franklin, J. (2009). Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors. *Ecography*, *32*, 907–918. https://doi.org/10.1111/j.1600-0587.2009.05883.x

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD-a platform fro ensemble forecasting of species distribution. *Ecography*, *32*, 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Toonen, R. J., & Hughes, S. (2001). Increased Throughput for Fragment Analysis on an ABI Prism 377 Automated Sequencer using a membrane comb and STRand software. *Biotechniques*, *31*, 1320–1324.

Torres, J., Brito, J. C., Vasconcelos, M. J., Catarino, L., Gonçalves, J., & Honrado, J. (2010). Ensemble models of habitat suitability relate chimpanzee (Pan troglodytes) conservation to forest and landscape dynamics in Western Africa. *Biological Conservation*, *143*, 416–425. https://doi.org/10.1016/j.biocon.2009.11.007

Trumbo, D. R., Spear, S. F., Baumsteiiger, J., & Storfer, A. (2013). Rangewide landscape genetics of an endemic Pacific northwestern salamander. *Molecular Ecology*, *22*, 1250–1266. https://doi.org/10.1111/mec.12168

Vogiatzakis, I. N., & Rackham, O. (2008). Crete. In *Mediterranean Island Landscape* (pp. 245–270).

Wagner, H. H., & Fortin, M.-J. (2016). Basics of Spatial Data Analysis: Linking Landscape and Genetic Data for Landscape Genetics Studies. In N. Balkenhol, S. A. Cushman, A. Storfer, & L. P. Waits (Eds.), *Landscape Genetics: Concepts, Methods, Application* (pp. 77–97). Wiley Blackwell.

Waits, L. P., & Storfer, A. (2016). Basics of population genetics: quantifying neutral and adaptive genetic variation for landscape genetics studies. In N. Balkenhol, S. A. Cushman, A. Storfer, & L. P. Waits (Eds.), *Landscape Genetics: Concepts, Methods, Application* (pp. 35–57). Wiley Blackwell.

Wang, I. J., Glor, R. E., & Losos, J. B. (2013). Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, *16*, 175–182. https://doi.org/10.1111/ele.12025

Wang, Y., Yang, K., Bridgman, C., & Lin, L.-K. (2008). Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecology*, *23*, 989–1000. https://doi.org/10.1007/s10980-008-9262-3

Wasserman, T. N., Cushman, S. A., Schwartz, M. K., & Wallin, D. O. (2010). Spatial scaling and multi-model inference in landscape genetics: Martes americana in northern Idaho. *Landscape Ecology*, *25*, 1601–1612. https://doi.org/10.1007/s10980-010-9525-7

Whitlock, M. C., & McCauley, D. E. (1999). Indirect measures of gene flow and migration : $F_{ST} \neq 1$ /(4 Nm+1). *Heredity*, *82*, 117–125.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*, 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

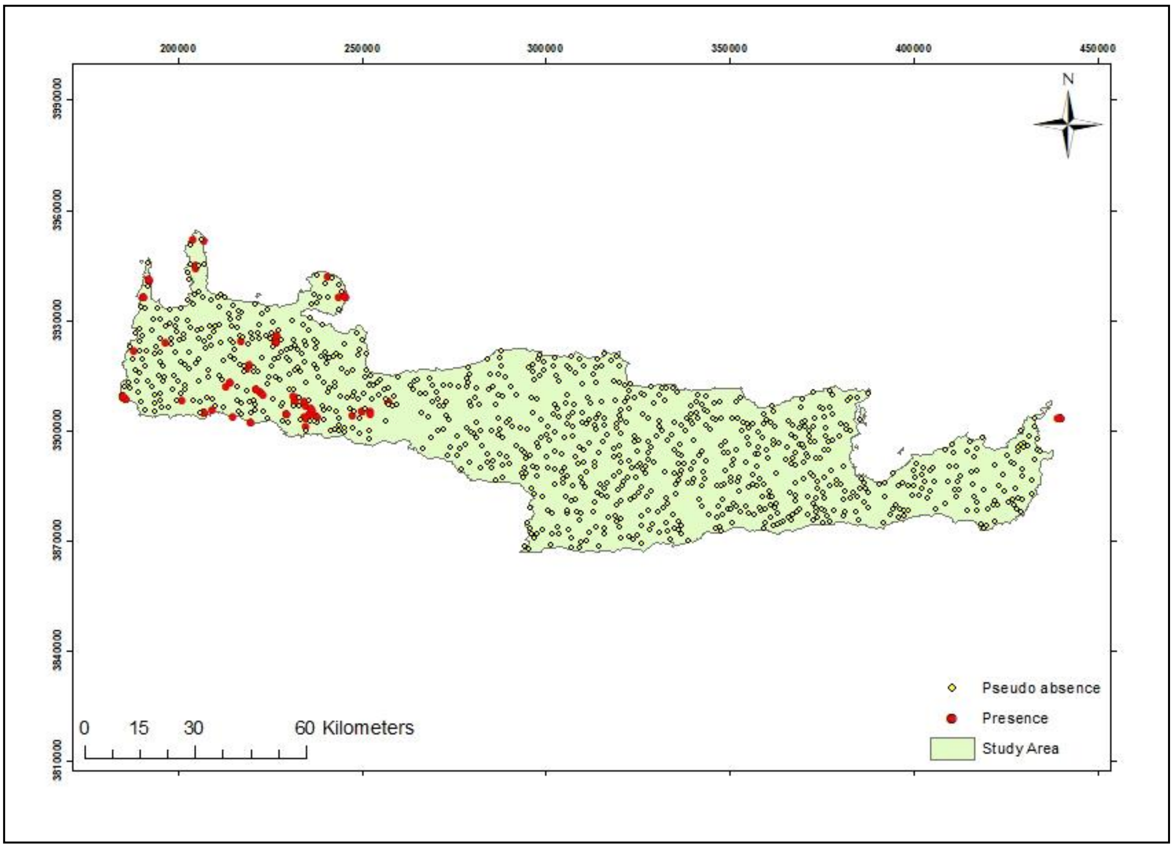
Wright, S. (1943). Isolation by distance. *Genetics*, *28*, 114–138.

Yackulic, C. B., Chandler, R., Zipkin, E. F., Royle, J. A., Nichols, J. D., Grant, E. H. C., & Veran, S. (2013). Presence-only modelling using MAXENT : when can we trust the inferences ? *Methods in Ecology and Evolution*, *4*, 236–243. https://doi.org/10.1111/2041-210x.12004

Zabalaga, N. A. (2008). *Influence of vegetation types and environmental variables in structuring Podarcis erhardii spatial heterogeneity in Crete, Greece Influence of vegetation types and environmental variables in structuring Podarcis erhardii spatial heterogeneity in Crete*. University of Twente, MSc Thesis.

Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: A review. *Landscape Ecology*, *27*(6), 777–797.

APPENDICES



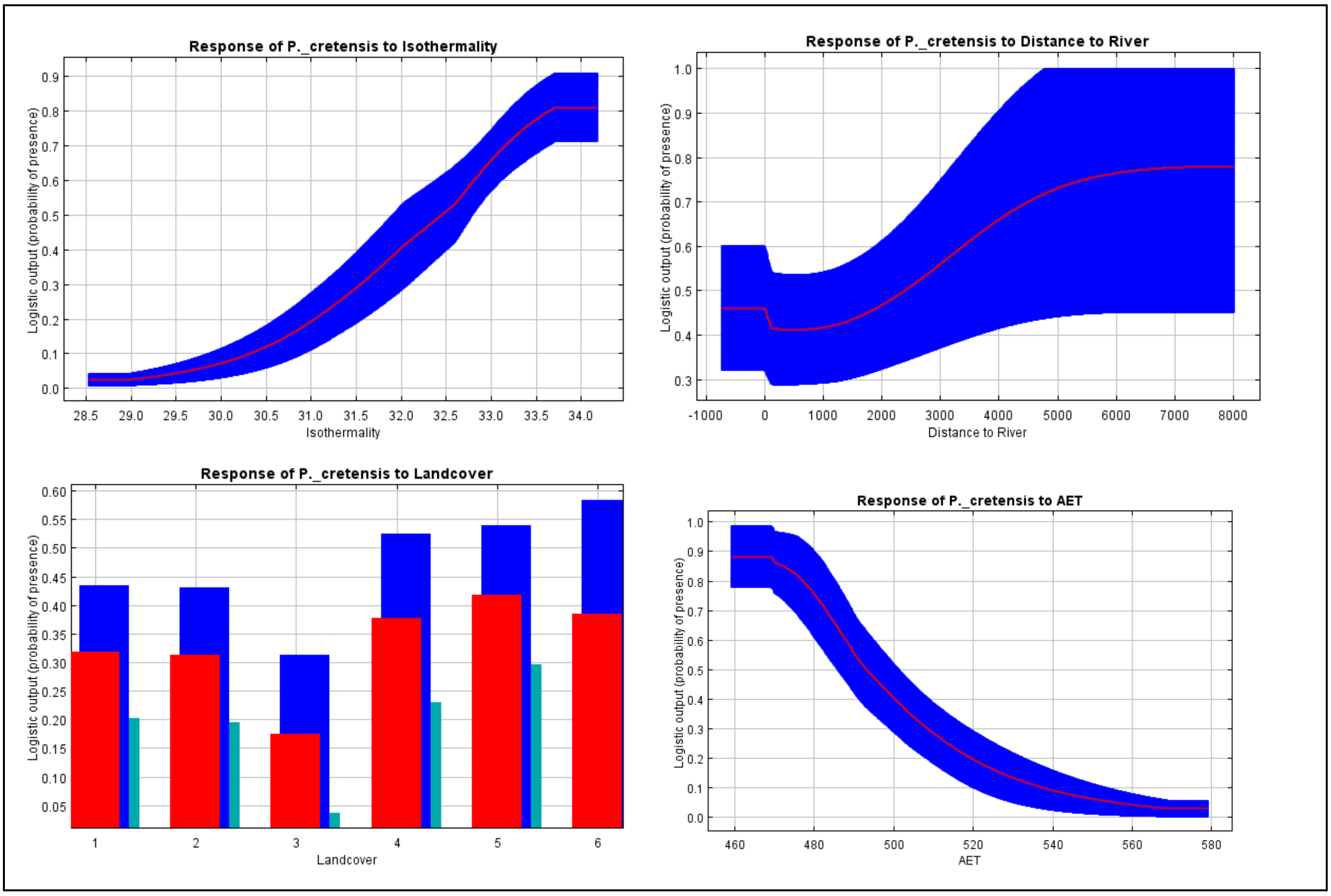
Appendix 1: Presence and Pseudo-absence data used in Species Distribution Modelling (Only Presence data was used for Maxent)

Locus	Primer (fluorescent dye)	Sequence (5’-3’)	Conditions	Reference
B6	B6-F (ROX)	CTG CTG CTT CAA TCA CAC TC	1.5 mM MgCl 94 °C, 1 min 50 °C, 1 min 72 °C, 1 min 35 cycles	(Nembrini and Oppliger, 2003)
	B6-R	GCC TTG CCT CTC CAG AAC		
C9	C9-F (TAMRA)	CAT TGC TGG TTC TGG AGA AAG	1.5 mM MgCl 94 °C, 1 min 50 °C, 1 min 72 °C, 1 min 35 cycles	(Nembrini and Oppliger, 2003)
	C9-R	CCT GAT GAA GGG AAG TGG TG		
Lv3-19	Lv3-19-F (TAMRA)	CTG TTG CTA TTT TGT ATG CTT AC	1.5 mM MgCl 94 °C, 1 min 50 °C, 1 min 72 °C, 1 min 35 cycles	(Boudjemadi et al., 1999)
	Lv3-19-R	CCT GTG ACT GTC CTC AGA GG		
Lv4-72	Lv4-72-F (ROX)	CCC TAC TTG AGT TGC CGT C	1.5 mM MgCl 94 °C, 1 min 47 °C, 1 min 72 °C, 1 min 35 cycles	(Boudjemadi et al., 1999)
	Lv4-72-R	CTT TGC AGG TAA CAG AGT AG		
Pb10	Pb10-F (ROX)	AGT GGA ATC GGC TGC AAT AC	1.5 mM MgCl 94 °C, 1 min 47 °C, 1 min 72 °C, 1 min 35 cycles	(Pinho et al., 2004)
	Pb10-R	ACC AGT CCC AGG AAT TTA GG		
Pb47	Pb47-F (FAM)	CTT GGT GGT TAA CAA TGT TGG C	3 mM MgCl 94 °C, 1 min 45 °C, 1 min 72 °C, 1 min 35 cycles	(Pinho et al., 2004)
	Pb47-R	GTG AGC TAA TAC AAC TCT CCA C		
Pb50	Pb50-F (TET)	GGA TGT TTC AGC ATG CTT GG	1.5 mM MgCl 94 °C, 1 min 47 °C, 1 min 72 °C, 1 min 35 cycles	(Pinho et al., 2004)
	Pb50-R	AGA CCT CAC TGG GCC ATT AC		
Pli4	Pli4-F (FAM)	TCA GTT CAT GCA TAA GGT CCA	1.5 mM MgCl 94 °C, 1 min	(Bloor et al., 2010)

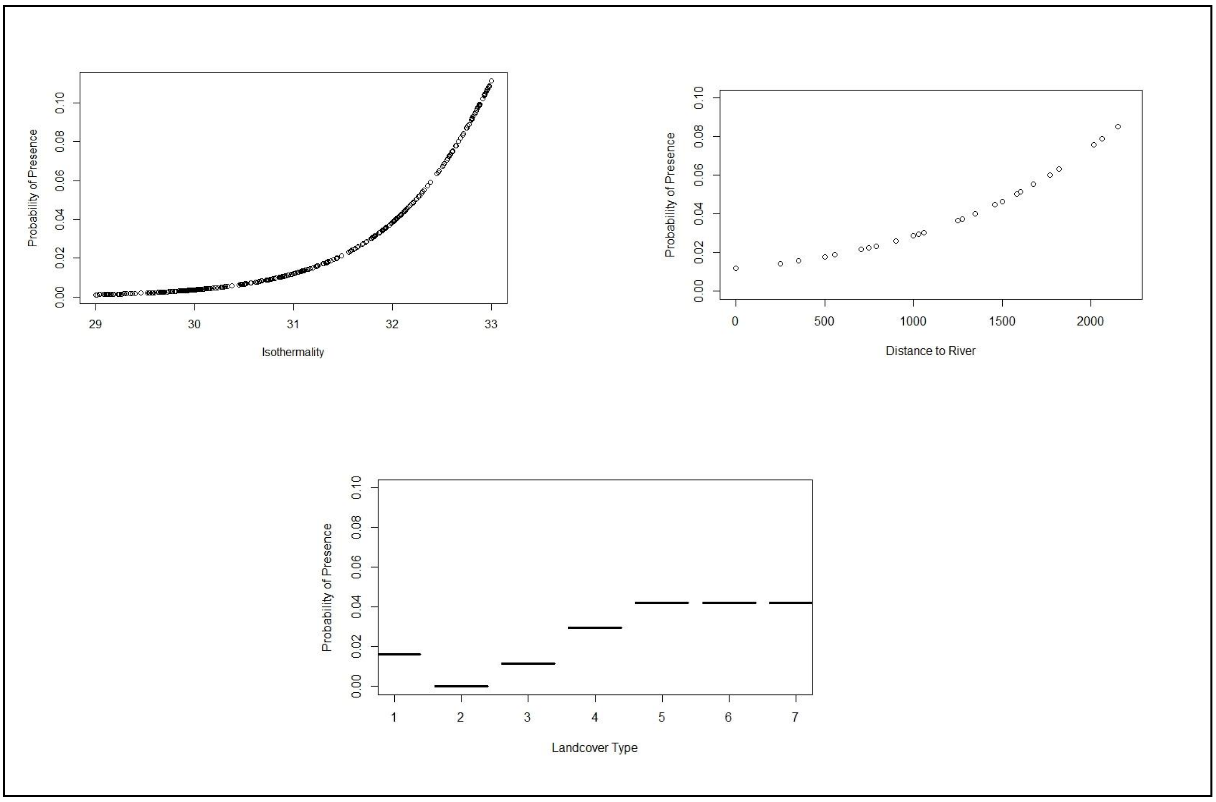
	Pli4-R	TTC GGC ATT TTT CTT CAG GT	50 °C, 1 min 72 °C, 1 min 35 cycles	
Pm16	Pm16- (TAMRA)	GGG ATG GAG AAA GAT GGC G	1.5 mM MgCl 94 °C, 1 min 50 °C, 1 min 72 °C, 1 min 35 cycles	(Richard et al., 2012)
	Pm16-R	GCA CTT GCC TAC TGG TCA TAC		
Pm27	Pm27-F (HEX)	TCC ATG AGC TCC ACA CAC G	1.5 mM MgCl 94 °C, 1 min 50 °C, 1 min 72 °C, 1 min 35 cycles	(Richard et al., 2012)
	Pm27-R	TCC ACA GCC ACT TAC GGA C		
Pmeli02	Pmeli02-F (TAMRA)	AGT GGA ATC GGC TGC AAT AC	1.5 mM MgCl 94 °C, 1 min 55 °C, 1 min 72 °C, 1 min 35 cycles	(Huyghe et al., 2010)
	Pmeli02-R	ACC AGT CCC AGG AAT TTA GG		
Pmeli19	Pmeli19-F (FAM)	TTC CAA GTC TGA TTC ACT CCA A	1.5 mM MgCl 94 °C, 1 min 57.4 °C, 1 min 72 °C, 1 min 35 cycles	(Huyghe et al., 2010)
	Pmeli19-R	AGC TGC AAG CAC CTA GCA AT		
Pod1A	Pod1A-F (FAM)	TGA GAA GCA CAT CTG CAC AC	1.5 mM MgCl 94 °C, 1 min 57.4 °C, 1 min 72 °C, 1 min 35 cycles	(Poulakakis et al., 2005)
	Pod1A-R	TGA ACG CAT AAT GGC TGA AGG		
Pod1B	Pod1B-F (FAM)	CCT TCA GCC ATT ATG CGT TCA TC	1.5 mM MgCl 94 °C, 1 min 57.4 °C, 1 min 72 °C, 1 min 35 cycles	(Poulakakis et al., 2005)
	Pod1B-R	AGG ATG GGG ATA ACC CCA GT		
Pod2	Pod2-F (FAM)	GGC AAT GTT CCT GCA TGA CG	1.5 mM MgCl 94 °C, 1 min 57.4 °C, 1 min 72 °C, 1 min 35 cycles	(Poulakakis et al., 2005)
	Pod2-R	TGG GAC AAA AAG GCA GAA CG		

Pod3	Pod3-F (TET)	TTA TCA GAC GTT GGG GAA AG	3 mM MgCl 94 °C, 1 min 57.4 °C, 1 min 72 °C, 1 min 35 cycles	(Poulakakis et al., 2005)
	Pod3-R	GCA CTT CAA CCC GAG GTC TG		
Pod8	Pod8-F (TET)	CCT CTA ACT ATC TGT TGC TGC TG	1.5 mM MgCl 94 °C, 1 min 58 °C, 1 min 72 °C, 1 min 35 cycles	(Poulakakis et al., 2005)
	Pod8-R	CAC AAA GGG TAT CGA AGG AGG		

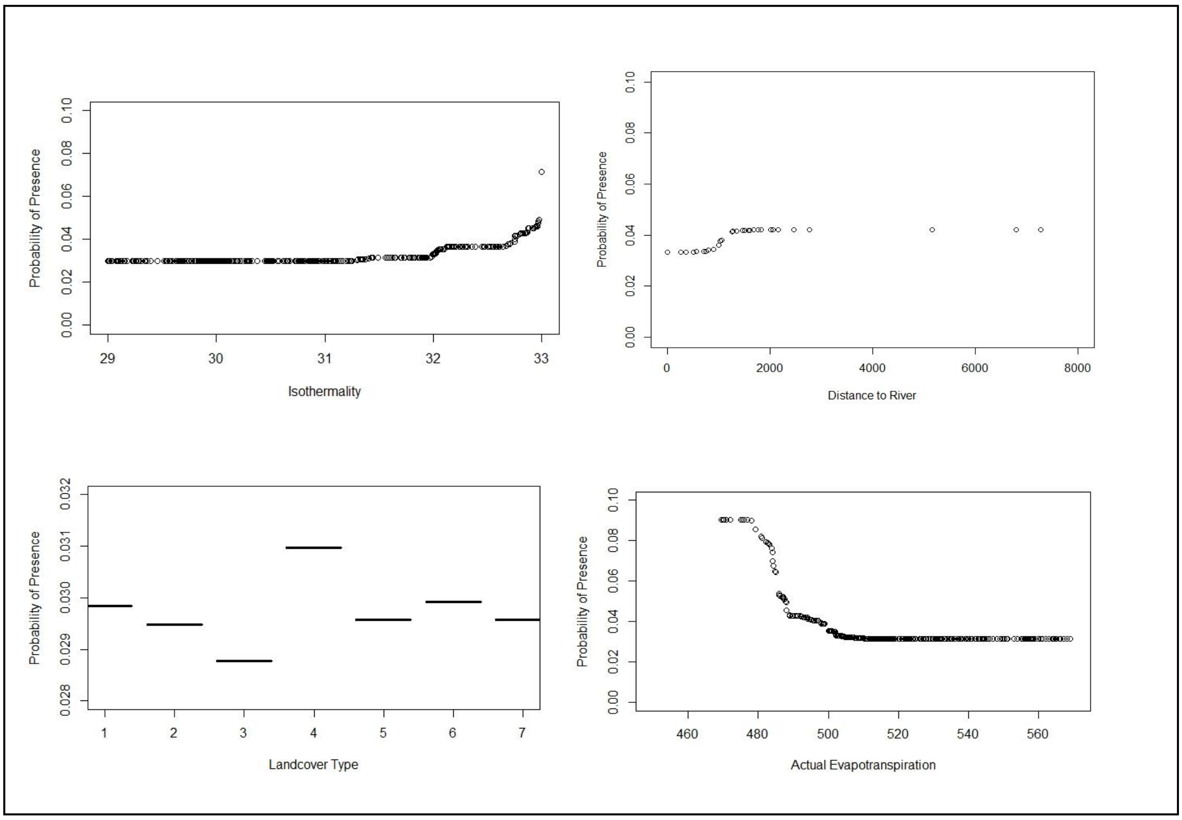
Appendix 2: Primers and condition used in PCR amplification of microsatellite loci



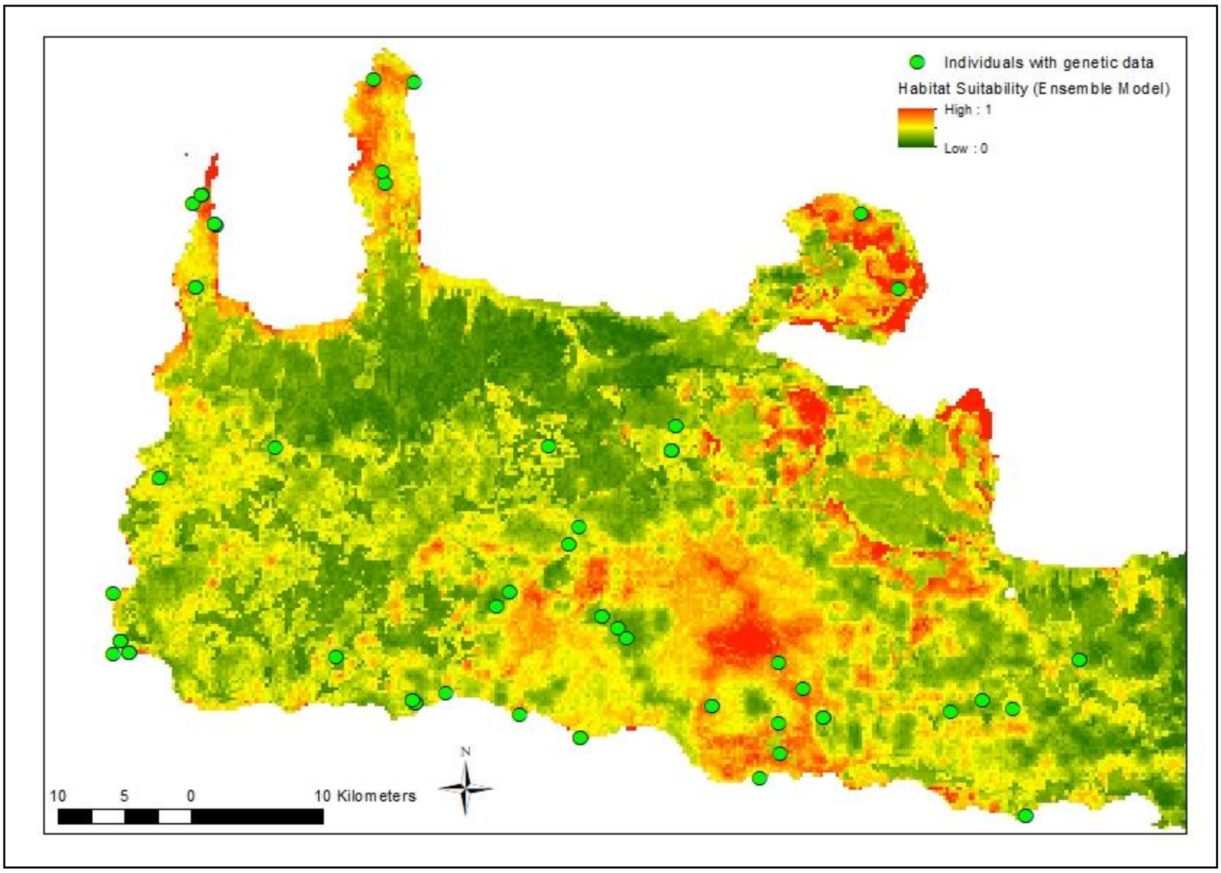
Appendix 3: Response Curve for Maxent Model



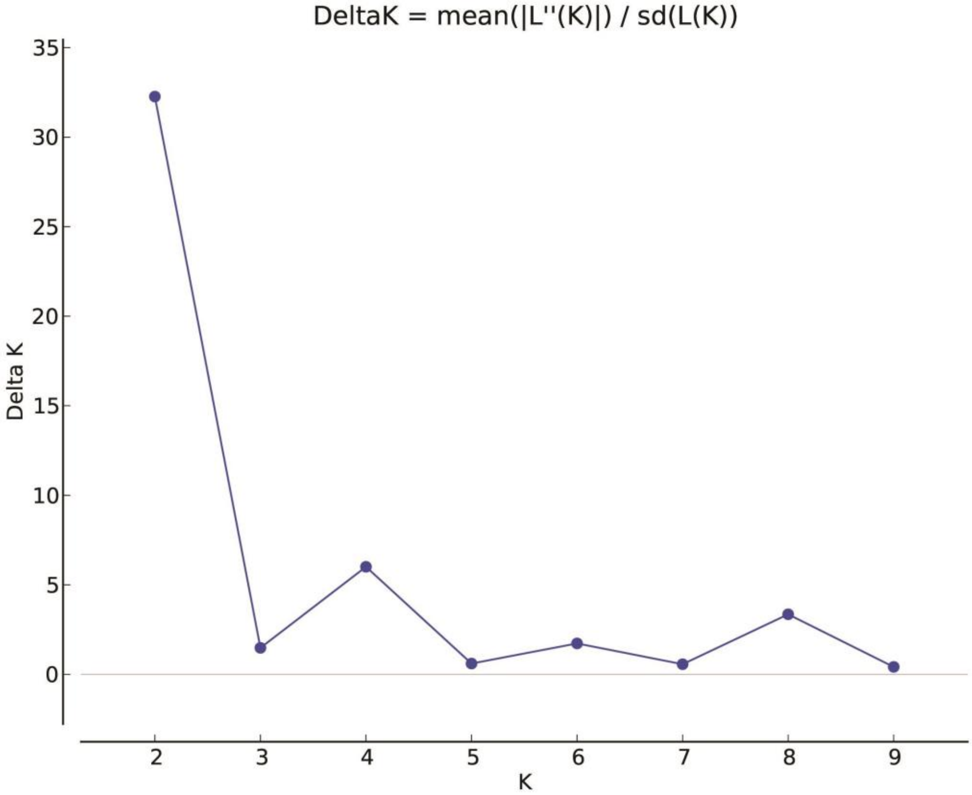
Appendix 4: Response Curve for GLM Model

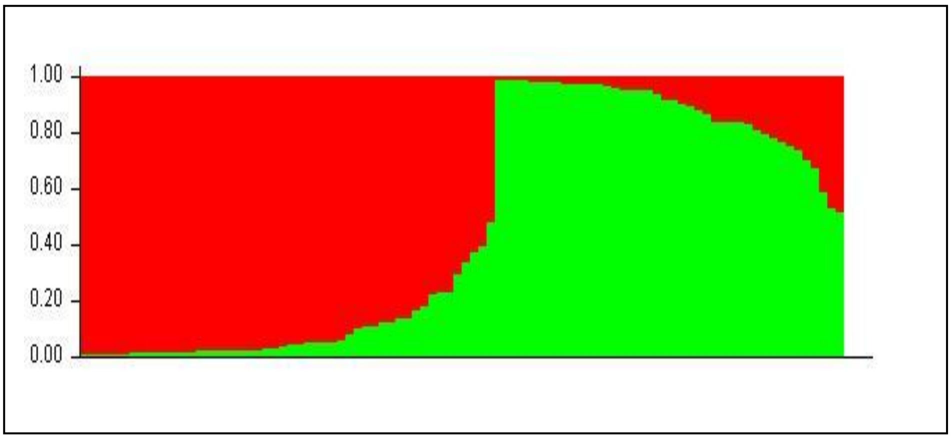


Appendix 5: Response Curve for BRT Model



Appendix 6: Individuals with genetic information overlaid with ensemble distribution model





Appendix 7: Evanno's delta K method to determine the genetic cluster. A population of 2 had the best support (above). The population cluster of 2 showing admixture.