Amphibian Species Distribution Modelling in Poland

Jinjin Lai
March, 2009

Course Title:	Geo-Information Science and Earth Observation for Environmental Modelling and Management
Level:	Master of Science (Msc)
Course Duration:	September 2007 - March 2009
Consortium partners:	University of Southampton (UK) Lund University (Sweden) University of Warsaw (Poland) International Institute for Geo-Information Science and Earth Observation (ITC) (The Netherlands)
GEM thesis number:	2007-10

Amphibian Species Distribution Modelling in Poland

by

Jinjin Lai

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

Thesis Assessment Board

Chairman: Dr. Yousif Hussin, ITC
External Examiner: Dr. Kasia Dabrowska-Zielinska, University of Warsaw
Internal Examiner: Dr. Joan Looijen, ITC
First Supervisor: Prof. Dr. A.K. (Andrew) Skidmore, ITC
Second Supervisor: Dr. A.G. (Bert) Toxopeus, ITC

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

Disclaimer

Abstract

The population decline of amphibians has become a worldwide attention during the past decades. To protect the endangered species, niche modelling has been used as a powerful tool of evaluating a species' potential habitat and identifying most important environmental variables for amphibian species. Based on this knowledge, proper conservation activities can be proposed to maintain amphibian species population.

Maximum Entropy Species Distribution Model (Maxent) was applied to predict amphibian species distribution by using species presence-only observation data in combination with environmental variables. The potential distribution was modelled for 18 amphibian species in Poland using maximum 22 different environmental predictors.

The most important factors turned out to be precipitation and soil temperature variables. Especially, soil temperature variables had great impact on both variables' relative importance to the model and predicted species spatial distribution. Aside from these variables, altitude, classified NDVI, insolation and proximity to pond also appeared to be considerable to explain amphibian species distribution.

Furthermore, the predicted species richness distribution was compared with the know species range from survey data. In general, potential habitat corresponded to the expected species range but it covered much broader suitable habitat area than the survey range.

Acknowledgements

It is a pleasure to thank many people who made this thesis possible.

First of all, many thanks to Erasmus Mundus – GEM MSc course, and its organizers and directors – Professors Andrew Skidmore from ITC, Peter Atkinson from University of Southampton, Petter Pilesjõ from Lund University and Katarzyna Dabrowska from University of Warsaw, as well as Prof. Liu Xuehua who introduced this course to me and opened the door to a new world. I have not only learned knowledge but also made good friends here. This experience is the most precious treasure for me and will be remembered forever.

I would like to express my sincere thanks to my thesis advisors Prof. Dr. Andrew Skidmore and Dr. Bert Toxopeus. They gave me an opportunity to explore a new field. Throughout my thesis period, they have shown a lot of concern and patience, providing knowledge, advice, and more important, support and encouragement. They are humorous also, which makes me relax during discussion. I am extremely thankful to them for guidance and support throughout the research phase.

Many thanks to Dr. Kasia Dabrowska-Zielinska from University of Warsaw, Prof. Włodzimierz Chętnicki from University of Bialystok, Maciej Pabijan and Beltran-Abaunza, who have provided reliable data and make this research possible. Especially Dr, Kasia Dabrowska-Zielinska, she is not only providing data, and always giving nice suggestions and good guiding as well.

I would like to thank to Mandar and Wang who gave me a lot of help during data processing and thesis research. Especially Mandar, who makes the tough thesis time interesting. Thanks to lulu who gives me chance to talk in Chinese daily. Thanks to all Gem friends.

A final special thanks to my parents, who brought me to this world and love and support me forever!

Table of contents

List of figures

List of tables

1. Introduction

1.1. Background and Significance

Amphibians are a class of active vertebrates which spread all over the world. It includes frogs, toads, salamanders, newts and caecilians. All amphibians are cold-blooded and most of them lay eggs. Amphibians play an important role in natural ecosystems as predators of invertebrates and as prey for small mammals, birds and snakes (Pearman, 1997).

Many amphibians live in both aquatic and terrestrial habitats and are particularly sensitive to natural ecosystem (Duellman & Trueb, 1994). Thus they are excellent indicators of the quality of the global environment.

Due to urbanization and climatic changes, habitat fragmentation and habitat availability decrease have become the main threat for amphibian species richness (Olsen, 2006; Linderman *et al*., 2005; Woodford, 2003). By the end of the 1990s, the population of many amphibian species had fallen down rapidly all over the world (Alford & Richards, 1999).

In response to ongoing population declines, broad scale inventory have been established to better understand current amphibian distribution and relative abundance across large landscapes (Corn, 2000).

The conventional approaches for observing species distributions, e.g. ground surveys, aerial photography, telemetry and satellite tracking (de Leeuw *et al*., 2002), require funds, time and manpower due to the fieldwork involved. It is also difficult to keep pace with landscape change because of the time-consuming process (Osborne *et al*., 2002).

GIS-based niche modelling

The problem of missing data could be resolved by using existing information on known species habitats to infer the locations of species elsewhere (Tole, 2006). GIS-based niche modelling approaches have been used to analyze species distribution (Arntzen, 2006; Weiers, 2004). They make good use of the limited observation data and provide a powerful tool to study the relationship between species distribution and environmental conditions (Guisan *et al*., 2002).

Use presence-only model

Due to time and space limitation of field survey, accurate absence data may be lacking. Even it is available it may be not as reliable as presence data (Anderson, 2003). A species may be recorded as absent at a given location for many reasons, e.g., the species is present but could not be detected; the species is absent but the habitat is suitable; or the habitat is truly unsuitable for the species. The former two situations can lead to identify false absences (Hirzel *et al*., 2002), which can decrease the reliability of predictive models (Chefaoui *et al*., 2005).

The absence data for amphibians are even more difficult to obtain compared to other species. All amphibians depend on moisture to prevent desiccation, and as a result can often be seen in winter, but are much rarer in other seasons. Particular species may be active only during certain times of the year or are active at night. Lack of consistent and up-to-date species occurrence data constrains the species distribution analysis (Weiers, 2004). The relationship between amphibian species richness and environmental parameters is still not clear (Qian *et al*., 2007).

Predicting species distributions from presence-only data and pseudo-absences is a powerful alternative when presence/absence data are unavailable or difficult to obtain (Zaniewski *et al*., 2002; Graham *et al*., 2004; Guisan and Thuiller, 2005). Different presence-only models have been developed (Manel *et al*., 1999).

The presence-only modelling techniques provide new solutions to study the relationships between amphibian species occurrence and environmental parameters to predict the potential species distributions.

1.2. Research Objectives

1.2.1. General Objective

The general objective of this study is to accurately predict the amphibian species distribution in Poland for effective species conservation activities. In order to achieve this, the following specific objectives are proposed.

1.2.2. Specific Objectives

- Determine environmental predictors for each species.
- Predict "probability of occurrence" and create presence-absence maps.

- Identify how the environmental predictors affect the model performance.
- Accurately estimate amphibian species richness distributions in Poland.

1.3. Research Questions

- Which environmental variables are ecologically important for affecting amphibian species distributions based on current studies?
- Which environmental variables are statistically important for determining amphibian species distribution modelling process?
- How to evaluate the model accuracy and reliability?
- How do the environmental predictors affect model performance?

1.4. Research Hypotheses

Hypothesis 1

H_0: The sensitivity (the proportion of correctly predicted presences) predicted by a distribution model is not higher than by a random prediction.

H_1: The sensitivity (the proportion of correctly predicted presences) predicted by a distribution model is higher than by a random prediction.

The formally stated hypothesis is as follows:

$$H_0:\ S_{sp_{dis}} = S_{sp_{ran}}$$

$$H_1:\ S_{sp_{dis}} > S_{sp_{ran}}$$

Where

S is sensitivity (the proportion of correctly predicted species presence)

sp is species presence

dis is distribution model

ran is ran prediction

Hypothesis 2

H_0: The accuracy (Kappa value) of the presence-absence map predicted by the distribution model which includes all environmental predictors is not higher than by the distribution model which does not include soil temperature predictors.

H_1: The accuracy (Kappa value) of the presence-absence map predicted by the distribution model which includes all environmental predictors is higher than by the distribution model which does not include soil temperature predictors.

The formally stated hypothesis is as follows:

$$\text{H}_0\text{:}\ K_{dis_{soiltem}} = K_{dis_{nosoiltem}}$$

$$\text{H}_1\text{:}\ K_{dis_{soiltem}} > K_{dis_{nosoiltem}}$$

Where
K is the Kappa value of predicted presence-absence map
dis is distribution model
soiltem is soil temperature predictors
nosoiltem is not including soil temperature predictors

Hypothesis 3
H_0: The mean observed species richness is the same as the mean predicted species richness.
H_1: The mean observed species richness is not the same as the mean predicted species richness.

The formally stated hypothesis is as follows:

$$\text{H}_0\text{:}\ M_{sr_{obs}} = M_{sr_{pre}}$$

$$\text{H}_1\text{:}\ M_{sr_{obs}} \neq M_{sr_{pre}}$$

Where
M is the mean value
sr is species richness
obs is observation
pre is model prediction

1.5. Research Approaches

The overall conceptual framework is summarized in Figure 1-1.

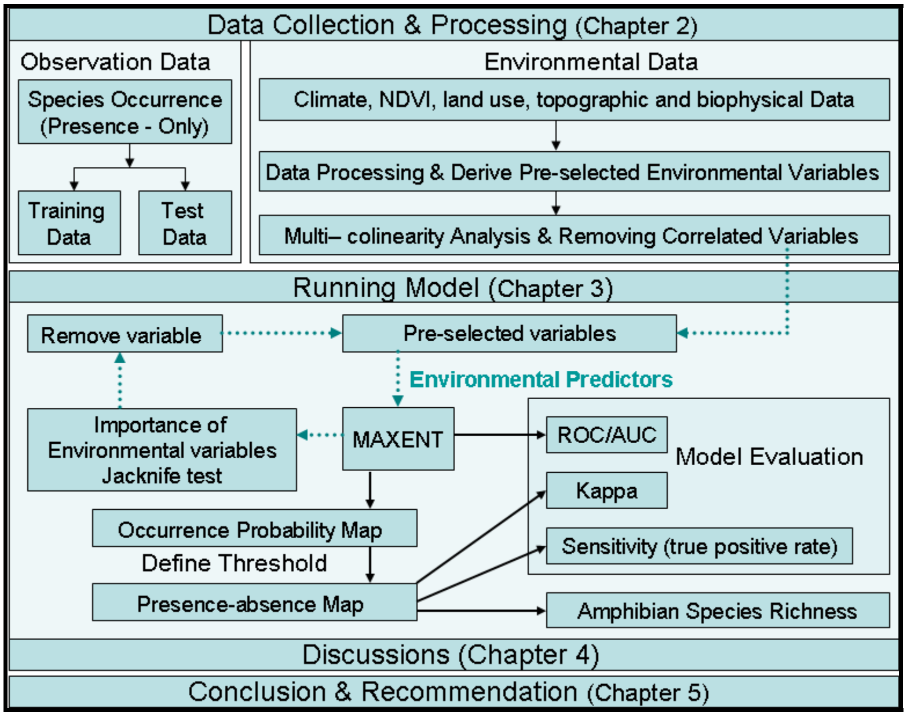


Figure 1-1 Conceptual framework[1] of this study

[1] The "" lines in the graph show the process of selecting environmental predictors. The details are described in Section 3.2.3.

2. Materials and Methods

2.1. Study Area

Poland is located in Central Europe. The geographic coordinates are between N 49-54°, E 14-24°, as shown in Figure 2-1.

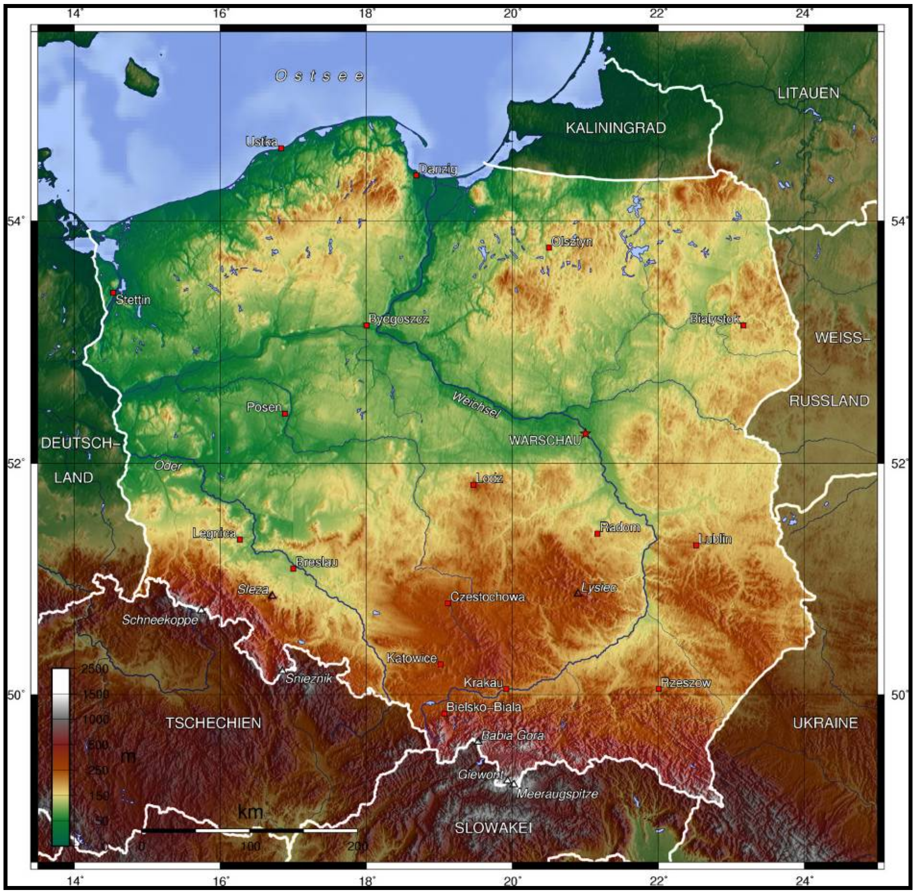


Figure 2-1 Topography map of Poland
Data source: http://gmt.soest.hawaii.edu/

The total area of Poland is 312,679 km^2. The total population is over 38 million. Forest covers 28% of Poland's land area. More than half of the land is devoted to agriculture. More than 1% of Poland's territory (3,145 km^2) is protected within 23 national parks. Wetland along with lakes and rivers in central Poland are legally protected, as are coastal areas in the north. There are over 120 landscape parks, nature reserves and other protected areas (http://en.wikipedia.org/wiki/ Poland).

The climate is mostly temperate throughout the country. The climate is oceanic in the north and west. It becomes gradually warmer and continental in south and east. Summers are generally warm with average temperatures between 20 °C and 27 °C. Winters are cold with average temperatures around 3 °C in the northwest and - 8 °C in the northeast. Precipitation falls down throughout the year, especially in the east; winter is drier than summer (http://en.wikipedia.org/wiki/ Poland).

2.2. Amphibian Species in Poland

There are 18 amphibian species in Poland, listed in Table 2-1. The pictures of the 18 species were showed in Figure 2-2.

Table 2-1 Amphibian Species in Poland (IUCN, 2006)

Code	Scientific Name	Common Name	Family	Red List
1	*Bombina bombina*	Fire-Bellied Toad	Bombinatoridae	LC[1]
2	*Bombina variegata*	Yellow-Bellied Toad	Bombinatoridae	LC[1]
3	*Bufo bufo*	Common Toad	Bufonidae	LC[1]
4	*Bufo calamita*	Natterjack Toad	Bufonidae	LC[1]
5	*Bufo viridis*	Green Toad	Bufonidae	LC[1]
6	*Hyla arborea*	European Tree Frog	Hylidae	LC[1]
7	*Pelobates fuscus*	Common Spadefoot	Pelobatidae	LC[1]
8	*Rana arvalis*	Moor Frog	Ranidae	LC[1]
9	*Rana dalmatina*	Agile Frog	Ranidae	LC[1]
10	*Rana esculenta*	Edible Frog	Ranidae	LC[1]
11	*Rana lessonae*	Pool Frog	Ranidae	LC[1]
12	*Rana ridibunda*	Eurasian Marsh Frog	Ranidae	LC[1]
13	*Rana temporaria*	European Common Frog	Ranidae	LC[1]
14	*Salamandra salamandra*	Fire Salamander	Salamandridae	LC[1]
15	*Triturus alpestris*	Alpine Newt	Salamandridae	LC[1]
16	*Triturus cristatus*	Great Crested Newt	Salamandridae	LC[1]
17	*Triturus montandoni*	Carpathian Newt	Salamandridae	LC[1]
18	*Triturus vulgaris*	Smooth Newt	Salamandridae	LC[1]

[1] Listed as Least Concern in view of its wide distribution, presumed large population, and because it is unlikely to be declining fast enough to qualify for listing in a more threatened category.

Bombina bombina
(Photo by Bert Toxopeus)

Bombina variegata
(Photo by Boris Timofeev)

Bufo bufo
(Photo by Bert Toxopeus)

Bufo calamita
(Photo by Bert Toxopeus)

Bufo viridis
(Photo by Bert Toxopeus)

Hyla arborea
(Photo by Bert Toxopeus)

Pelobates fuscus
(Photo by Bert Toxopeus)

Rana arvalis
(Photo by Horia Bogdan)

Rana dalmatina
(Photo by Wouter Beukema)

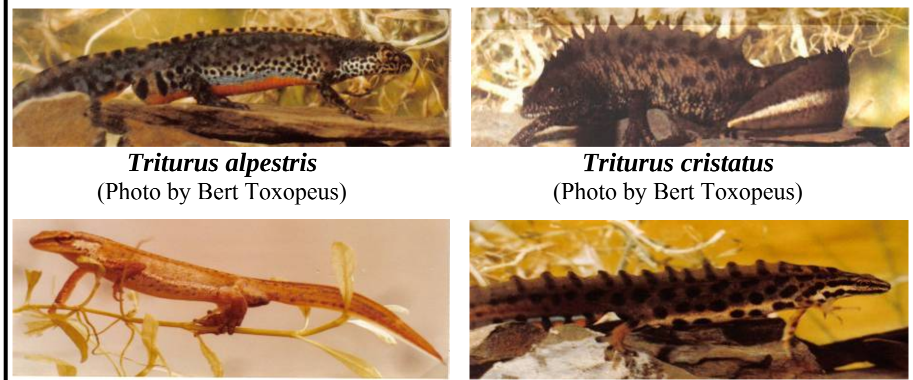
Rana esculenta
(Photo by Fabrizio Vigni)

Rana lessonae
(Photo by Jiří Mařík)

Rana ridibunda
(Photo by Christian Fischer)

Rana temporaria
(Photo by Bert Toxopeus)

Salamandra salamandra
(Photo by Bert Toxopeus)



Triturus alpestris
(Photo by Bert Toxopeus)

Triturus cristatus
(Photo by Bert Toxopeus)

Triturus montandoni
(Photo by Bert Toxopeus)

Triturus vulgaris
(Photo by Bert Toxopeus)

Figure 2-2 Photographs of the amphibian species in Poland

2.3. Species Occurrence Data

The amphibian species occurrence data (presence-only) included two independent parts.

2.3.1. National Atlas Data

All available presence data for the 18 species covered the whole country during 1970-2002 were summarized and provided in shape file polygon format. The whole country was divided by grids with the spatial resolution of 0.167×0.083 decimal degrees (approximately 10×10 km). The occurrence distributions for each species were shown in Figure 2-3. Except *Rana dalmatina*, all of the species spread out at least 100 grids or even more.

2.3.2. Local Field Survey Data

The presence data for central south area was collected during 2006-2008 (Bonk & Pabijan, 2006; Bonk & Pabijan, 2008), as shown in Figure 2-4. 73 localities were investigated and the occurrence data for 14 species were obtained (*Bombina variegata, Salmandra salmandra, Triturus montandoni* and *Rana dalmatina* occur outside of this area).

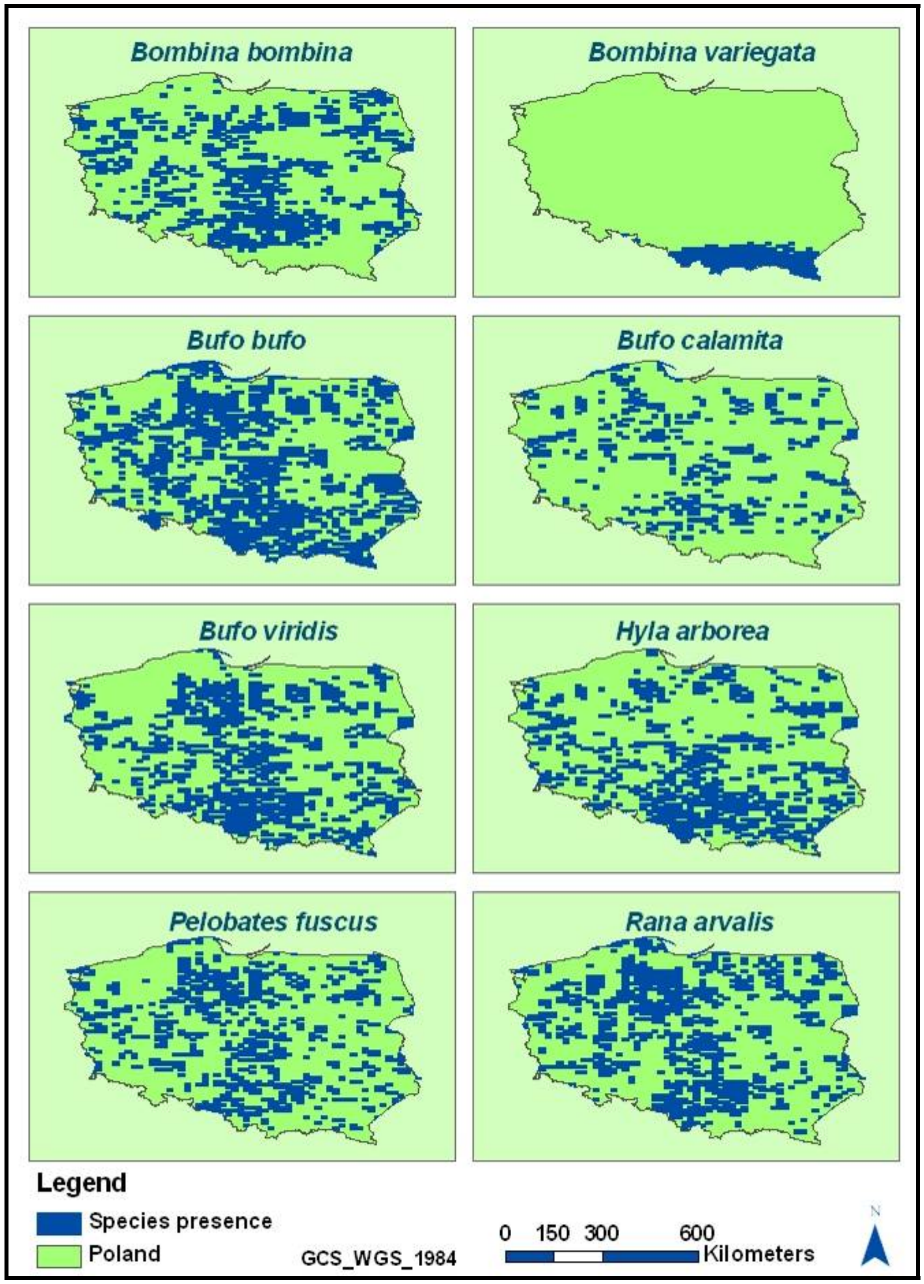


Figure 2-3 Species presence data from atlas

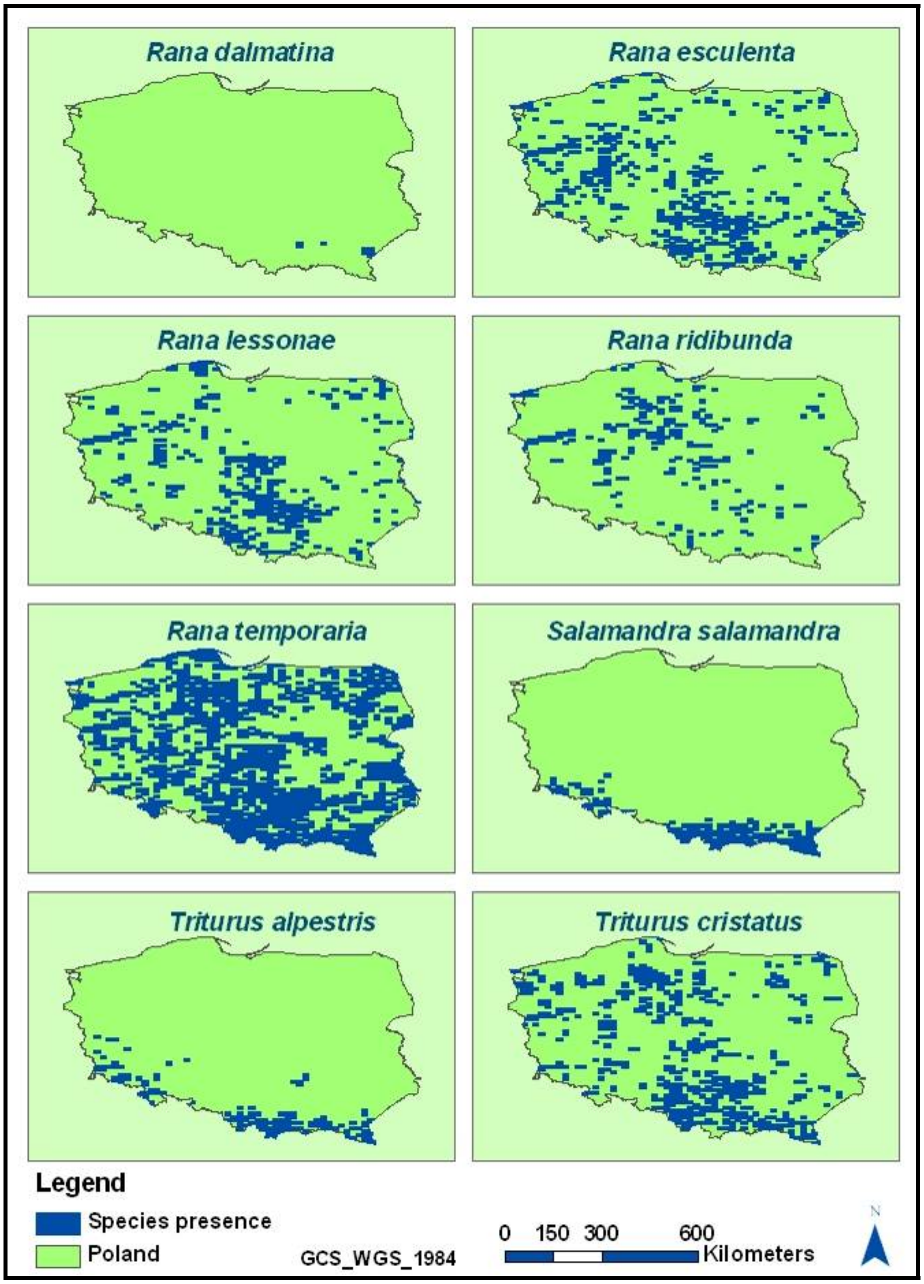


Figure 2-3: Species presence data from atlas (continued)

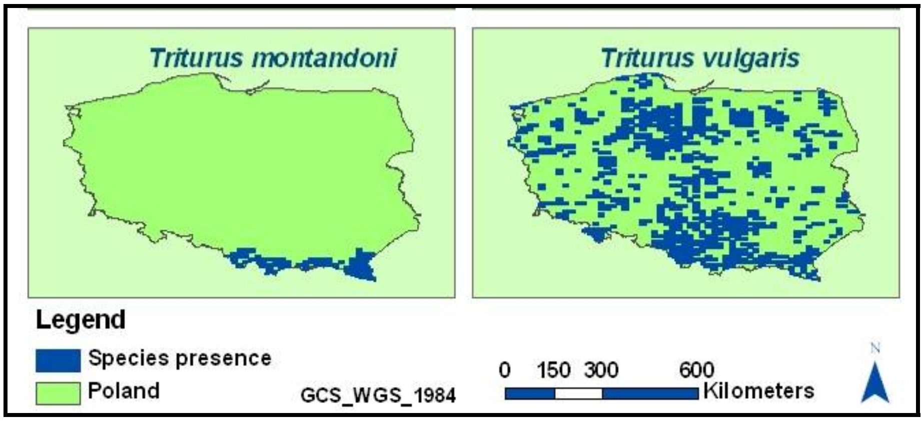


Figure 2-3: Species presence data from atlas (continued)

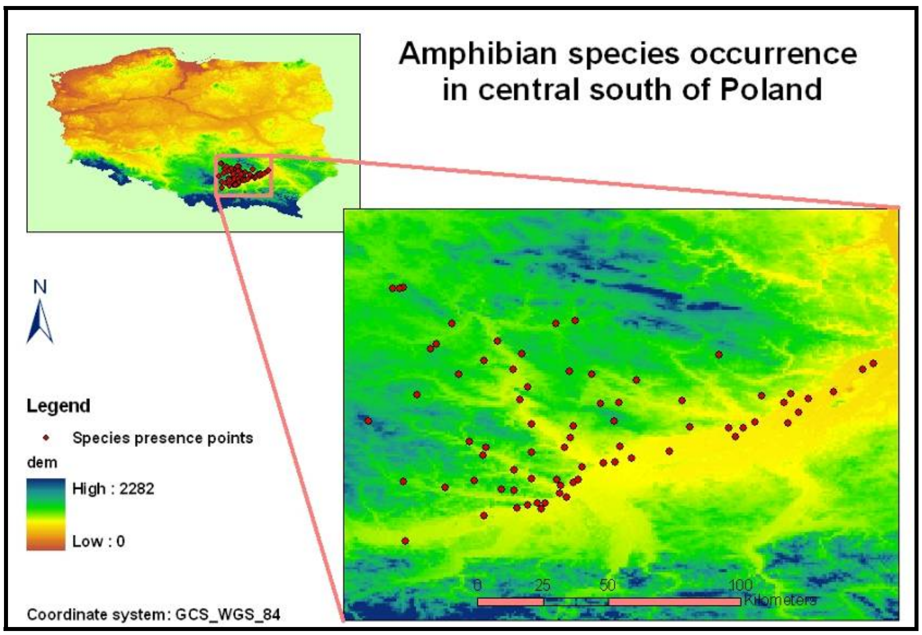


Figure 2-4 Species presence data from field survey

2.4. Pre-selected Environmental Variables

Species distributions are limited to a certain time and space due to environmental conditions (Gusian & Thuiller, 2005). Expert knowledge was used for selecting environmental variables.

Current studies indicate that climatic variables impact species distribution in both direct and indirect ways (Lennon *et al*., 2000; Badgley & Fox, 2000). The primary productivity may change because of less energy available for species to grow and reproduce, which is followed by more competition and influences population dynamics of animals. Long term NDVI images can show the overall productivity of ecosystem and could be used as an index for net primary production (NPP) (Oindo & Skidmore, 2002). Land cover changes, leading to fragmentation of the natural species habitats, have also been altering the species population (Joly, 2004). The proximity to water also has impact on amphibian distribution (Negga, 2007).

Based on previous studies (Negga, 2007; Wu, 2006; Arntzen, 2006; Sun, 2007), 36 environmental variables from six categories which have potential influences on amphibian species distribution were pre-selected and derived from multi-datasets, as summarize in Table 2-2. Classified NDVI, Corine map and Soil type are categorical variables; the rest are continuous variables. In order to assess the effect of environment on amphibian species distribution at different phase of their life cycle, e.g., aquatic and terrestrial phase, seasonal[2] climate, NDVI and soil data were used for analysis. The details of data processing were described in Table 2-2.

Table 2-2: Pre-selected Environmental Variables

Category	Variable used	Data format	Data Source
Climate	Seasonal mean precipitation Seasonal mean max/min air temperature	Raster	WORLDCLIM
	Seasonal mean insolation	Hard copy	Local database
NDVI	Seasonal mean NDVI	Raster	SPOT Vegetation
	Classified NDVI	Raster	Msc thesis
Land cover	Corine map	Raster	CLC 2000
Terrain	Altitude Slope Aspect	Raster	USGS/NASA SRTM data
Soil	Soil type	Vector	FAO
	Seasonal mean max/min soil temperature at 5cm depth	Hard copy	Local database
Proximity to water	Distance to pond Distance to river	Vector	USGS

[2] In this study, season 1 is from Jan to Mar, season 2 is from Apr to Jun, season 3 is from Jul to Sep and season 4 is from Oct to Dec.

2.4.1. Climate Data

2.4.1.1. Precipitation and Temperature

Precipitation and temperature data were downloaded from WORLDCLIM website. WORLDCLIM is a set of global climate layers (climate grids) with multi-spatial resolution (30 arc-second, 2.5 arc-minutes, 5 arc-minutes and 10 arc-minutes) and multi-temporal period (current representing 1950-2000 and future including 2020, 2050 and 2080), developed by Robert J. Hijmans, *et al*. (2005) from the Museum of Vertebrate Zoology, University of California. The data layers were generated through interpolation of average monthly climate data from weather stations where at least 10 years of data were available. Variables included are monthly total precipitation, and monthly mean, minimum and maximum temperature, and 19 derived bioclimatic variables. The data are widely used for mapping and spatial modelling in a GIS or other computer program. The details of the data are described in Hijmans' paper.

The current monthly total precipitation and monthly mean, minimum and maximum temperature data with 30 arc-second spatial resolutions were used in this study to generate the seasonal total precipitation and seasonal mean, minimum and maximum temperature.

2.4.1.2. Insolation data

The monthly insolation data were obtained from the local weather stations during 2005-2008. The station points were located in Figure 2-5. The raster layers were generated through interpolation of average monthly insolation data by *Moving Window Kriging* function in Arc GIS 9.3 (Hijmans, *et al.*, 205). The seasonal mean insolation was calculated for further analysis.

Figure 2-5 Location of weather stations in Poland

2.4.2. Normalize d Difference Vegetation Index (NDVI) Data

The NDVI images were obtained from the VEGETATION database. It consists of 360 layers, generated every 10 days from Apr 1998-Mar 2008, by SPOT-4 sensor with a spatial resolution of 1×1 km. The seasonal NDVI were calculated based on the 10-day-period data. The classified NDVI image, including 87 classed, were obtained from an Msc thesis work (Beltran Abaunza, 2009) and used in this study.

2.4.3. Land Cover

Corine Land Cover (CLC) 2000 was downloaded from European Environment Agency (EEA) website (http://dataservice.eea.europa.eu/dataservice/). It provides consistent information on land cover changes during the past decade across Europe. in TIFF format with the spatial resolution of 100×100 m (Table 2-3).

Table 2-3 Description of land cover classes

No.	Land Cover	Description
1	Urban fabric	Continuous and discontinuous urban fabric
2	Industrial area	Industrial or commercial units, road and rail networks and associated land, port areas and airports
3	Mine	Mineral extraction sites, dump sites and construction sites
4	Non-agricultural vegetated areas	Green urban areas and sport and leisure facilities
5	Arable land	Non-irrigated arable land, permanently irrigated land and rice fields
6	Permanent crops	Vineyards, fruit trees and berry plantations and olive groves
7	Pastures	Pastures
8	Heterogeneous agricultural areas	Annual crops associated with permanent crops, complex cultivation patterns, principally occupied by agriculture, with significant areas of natural vegetation and agro-forestry areas
9	Forests	Broad-leaved forest, coniferous forest and mixed forest
10	Scrub and/or herbaceous vegetation associations	Natural grasslands, moors and heathland, sclerophyllous vegetation and transitional woodland-shrub
11	Open spaces with little or no vegetation	Beaches, dunes, sands, bare rocks, sparsely vegetated areas, burnt areas and glaciers and perpetual snow
12	Inland wetlands	Inland marshes and peat bogs
13	Inland waters	Water courses and water bodies

Source: European Environment Agency (EEA) website

2.4.4. Topographical Data

Digital elevation model (DEM) was obtained from Shuttle Radar Topography Mission (SRTM) and downloaded from CGIAR-Consortium for Spatial Information (CSI) website (http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp) with a spatial resolution of 90×90 m. Slope and aspect were generated from the DEM data using *Spatial Analyst* function in Arc GIS 9.3.

2.4.5. Soil Data

2.4.5.1. Soil Type

Soil data (vector polygon) were obtained from the FAO database. The data were classified based on FAO85_LEV1 standard, as shown in Table 2-4. The data were converted into raster format in Arc GIS 9.3

Table 2-4 Description of soil types

Code	Soil Type	Code	Soil Type	Code	Soil Type
1	Town	6	Rendzina	11	Luvisol
2	Soil disturbed by man	7	Gleysol	12	Histosol
3	Water body	8	Phaeozem	13	Podzol
4	Cambisol	9	Lithosol	14	Arenosol
5	Podzoluvisol	10	Fluvisol	15	Ranker

2.4.5.2. Soil Temperature

The monthly max/min soil temperatures at 5cm depth were obtained from the local weather stations during 2005-2008. The station points were located in Figure 2-5. The raster layers were generated through interpolation of average monthly data by *Moving Window Kriging tool* in Arc GIS. The seasonal mean max/min soil temperatures were calculated for further analysis.

2.4.6. Proximity to Water

The water bodies and water courses were downloaded from USGS website (http://water.usgs.gov/) in vector format. "Proximity to water" represents distance from every pixel to water. Proximity to pond and proximity to river were calculated separately by using *Euclidean distance tool* in Arc GIS.

All variables were converted into raster layers in Arc GIS 9.3 and projected into GCS_WGS84 system with the spatial resolution of 0.0167×0.0083 decimal degrees (1×1km). All the raster layers were converted into ASCII format for further analysis.

2.5. Multi-colinearity Analysis

Multi-colinearity refers to linear inter-correlation among variables. High correlations among environmental variables may result in highly unstable performance of the Least Squares Estimator, which will lead to problems for running species distribution model.

Multi-colinearity can be detected by calculating Variance Inflation Factor (VIF):

$$VIF = \frac{1}{1 - R^2} \qquad \textbf{Equation 2-1}$$

Where
R^2 is coefficient of determination.

VIF indicates the inflation in the variance of each regression coefficient compared with a situation of orthogonality. The rule of thumb is that VIF>10 means multi-colinearity may influence the least squares estimator.

Multi-colinearity was analyzed by using the *linear regression* tool in SPSS 16.0 software. All the categorical variables (classified NDVI, Corine and Soil type) were excluded from the analysis as they can not be tested and VIF value for the rest 33 variables were calculated by Colinearity Diagnostics.

Table 2-5 Remaining environmental variables after multi-colinearity analysis

No.	Environmental variables	VIF
1	Proximity to pond	1.428
2	Proximity to river	1.686
3	The mean insolation in season 1	5.417
4	The mean insolation in season 2	4.211
5	The mean insolation in season 3	5.646
6	The mean NDVI in season 2	1.602
7	The mean NDVI in season 3	1.961
8	The mean NDVI in season 4	2.108
9	The mean precipitation in season 1	4.214
10	Slope	3.225
11	The mean min soil temperature at 5cm depth in season 1	3.849
12	The mean min soil temperature at 5cm depth in season 2	2.477
13	The mean max air temperature season 3	6.668
14	The mean min air temperature season 3	7.604

Removal of the variable with the highest values from the variable-list was followed by re-running colinearity diagnostics, till all the remaining values are below 10. The variables were omitted one by one because the values of all the remaining variables might change drastically after removing one variable. Expert knowledge was required to decide which variable should be removed in each step. The aim was to try to keep as many parameters with different meaning in the list to represent as much environmental information as possible. 14 environmental variables were finally left (Table 2-5).

Based on the expert knowledge, altitude, precipitation in spring, air temperature in summer, soil temperatures are important for certain amphibian species. These variables should be considered in the modelling process even though the VIF value is higher than 10. Based on this information and including the three categorical variables (classified NDVI, Corine and Soil type), 22 variables were pre-selected as the input of the model, as listed in Table 2-6.

Table 2-6 Pre-selected environmental variables used in model process

No.	Variable Name	Description
1	altitude	altitude
2	classified_ndvi	classified_ndvi (87 classes)
3	corine	corine map
4	dis_to_pond	Proximity to pond
5	dis_to_river	Proximity to river
6	inso_sea_1	The mean insolation in season 1
7	inso_sea_2	The mean insolation in season 2
8	inso_sea_3	The mean insolation in season 3
9	ndvi_sea_2	The mean NDVI in season 2
10	ndvi_sea_3	The mean NDVI in season 3
11	ndvi_sea_4	The mean NDVI in season 4
12	pre_sea_1	The mean precipitation in season 1
13	pre_sea_2	The mean precipitation in season 2
14	slope	Slope
15	soil_type	Soil type
16	sotmin_sea_1	The mean min soil temperature at 5cm depth in season 1
17	sotmin_sea_2	The mean min soil temperature at 5cm depth in season 2
18	sotmin_sea_3	The mean min soil temperature at 5cm depth in season 3
19	sotmax_sea_2	The mean max soil temperature at 5cm depth in season 2
20	sotmax_sea_3	The mean max soil temperature at 5cm depth in season 3
21	tmax_sea_3	The mean max air temperature season 3
22	tmin_sea_3	The mean min air temperature season 3

2.6. Spatial Resolution Differences and Solutions

To study the relationship between the species distribution and environmental conditions, the two datasets should have the same spatial resolution. As described in Section 2.3 and 2.4, for each 10×10 km species occurrence grid, there were one hundred 1×1 km cells for the environmental variables.

Instead of re-sampling the environmental variables into 10×10 km resolution, efforts had been made to solve this problem in different ways (Olivero *et al*., 2009). For each 10×10 km observation grid:

1. Randomly select one cell out of the hundred; use the cell value to represent the environmental condition in the grid.
2. Select the central cell of the grid; use the cell value to represent the environmental condition in the grid.
3. Take the mean value of all the hundred cells.

The results showed that it did not make much difference on modelling predictions no matter which method was used to determine the environmental value for each observation grid.

In this study, the value of central cell was selected to represent the environmental conditions for each grid. To achieve this, the central points (x, y coordinates) of grids were generated by the *feature to point* tool in Arc GIS 9.3. For each central point, it contains the occurrence information of the grid and represents the central cell of environmental variables in the grid as well.

3. Amphibian Species Distribution Modelling

3.1. Introduction

In this chapter, the Maxent model was used to predict the potential distribution of amphibian species in Poland.

3.2. Methodolo gy

3.2.1. Maximum Entropy (Maxent) model

The Maximum entropy (Maxent) is a machine-learning method used to obtain predictions or make inferences from incomplete information (Phillips *et al*., 2006). The main purpose is "to estimate (approximate) unknown probability distribution of a species" based on the Maximum-Entropy Principle (Phillips *et al*., 2006).

Entropy is defined by Shannon (1948), as “a measure of how much 'choice' is involved in the selection of an event". A distribution with higher entropy, involves more choices. Given a set of samples (species occurrences) and set of features (environmental variables), the Maxent model estimates niches by finding the distribution of probabilities closest to uniform (maximum entropy), constrained by the fact that feature values match their empirical average. Phillips *et al*.(2006) documented the main features of this software:

1. Species occurrences results are a range of probabilities between 0 and 1(if you take the logistic values). Planners can consider all sites in the landscape based on their degree of conservation suitability (e.g. Sarkar *et al*., 2004).
2. Both presence-only data and presence-absence data can be used.
3. Ability to use numerous and diverse environmental information. Environmental data may be both continuous and categorical.
4. Generate accurate models and provide an output which identifies the role of each environmental variable in the prediction model.

Maxent 3.2.1 was used, as shown in Figure 3-1. The model input requires both the species occurrence data and environmental layers:

1. The species occurrence data should be point data in MS Excel *.csv format. Three fields are defined: species’ name, longitude and latitude (in decimal degrees). If more than one species are being considered at a time, these are

put systematically one after the other. The occurrence points can be divided into training points and testing points. The proportions of training and testing points are user-defined.

In this study, the observation grids data (Section 2.3) were converted into points as the model input. The central points of grids were generated by the *feature to point* tool in Arc GIS 9.3 to represent each grid (Section 2.6), then converted into *.csv format. 75% of the sample points were used for training the model while 25% were used for testing.

2. All the environmental layers should be input in ASCII raster format. They must have the same projection system and the same geographic boundary and cell size.

   22 environment variables data were pre-processed (Section 2.5) and input in the model. All the data used during modelling process were raster layers with spatial resolution of 0.0083 decimal degrees (1×1 km).

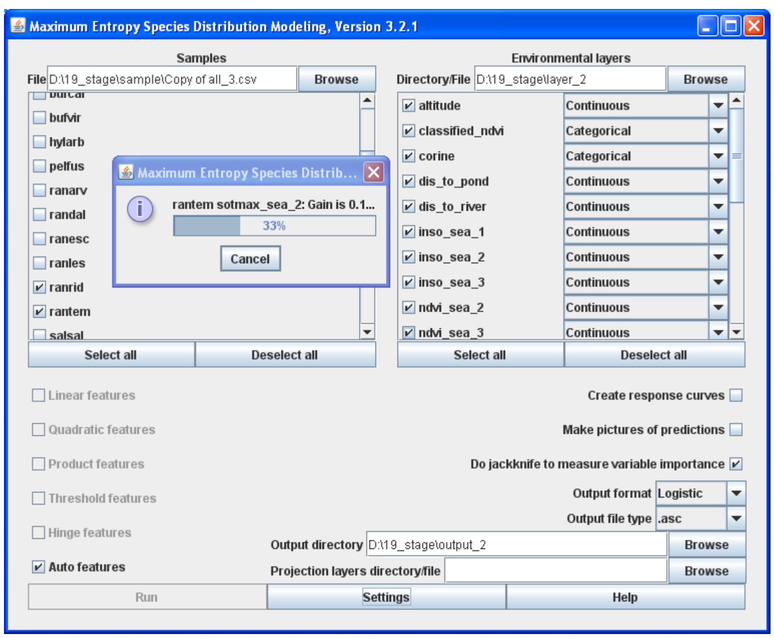


Figure 3-1 Maximum entropy species distribution modelling (Version 3.2.1)

The user specified parameters were set as follows: regularization multiplier = 0.2, maximum iteration = 1000, convergence threshold = 0.001, maximum number of background points = 3000.

The training model Progress starts after loading the environmental layers. The gain is closely related to deviation, a measure of goodness of fit used in generalized additive and generalized linear models. It starts at 0 and increases towards an asymptote during the run. During this process, the model is generating a probability distribution over pixels in the grid, starting from the uniform distribution and repeatedly improving the fit to the data. The gain is defined as the average log probability of the presence samples, minus a constant that makes the uniform distribution have zero gain. At the end of the run, the gain indicates how closely the model is concentrated around the presence samples; for example, if the gain is 2, it means that the average likelihood of the presence samples is $\exp(2) \approx 7.4$ times higher than that of a random background pixel.

3.2.2. Importance of Environmental Variables

Species distribution model should be able to identify which variables are making the greatest contribution to the model. There are two possible ways in which Maxent can be used to address it.

3.2.2.1. Relative Contribution of Environmental Variables

While the model is being trained, we can keep track of which environmental variables make the greatest contribution to the model. Each step of the Maxent algorithm increases the gain of the model by modifying the coefficient for a single feature; the program assigns the increases in the gain to the environmental variables that the feature depends on. The relative contributions of the variables are converted to percentages at the end of the training process.

However, the percent contribution values are only heuristically defined: they depend on the particular path that the code uses to obtain the optimal solution, and a different algorithm could lead to the same solution via a different path, resulting in different percent contribution values. Especially when there are highly correlated environmental variables, the percent contributions should be interpreted with more caution. Having a higher percentage contribution does not necessarily imply that the variable is more important to the species than other variables.

3.2.2.2. Jacknife Test

Maxent can run Jackknife operation to assess the importance of each environmental predictor variable. It excludes one environmental variable out of the model and runs a model using the remaining variables sequentially. It also runs a model using only the excluded variable in isolation. As a result, the gain contribution of each variable to the total gain of the model (inclusive of all variables) can be calculated. A variable which decreases the total gain of the model higher than all the other variables when excluded and as well as a variable which contributes the highest gain when used alone will be identified as the most important variable.

Compared with the percent contribution values, the results of Jacknife test are more reliable and can be easily interpreted.

3.2.3. Environmental Predictors Selection

Due to different environment requirements among species, even if the same variable is used to build the model, its behavior will vary along with the predicted species. Based on the relative importance of the environmental variables, environmental predictors can be selected for each species. The basic processes are listed below.

1. Use all the 22 pre-selected variables to run the model.
2. Check the Jacknife test results. Omit the variable which has most negative effect on the total gain (in other words, using remaining variables will get more gain than including this variable).
3. Use the remaining variables to run the model and check the Jacknife test results to omit another variable.
4. Repeat step 3 until all the remaining variables have positive effect to the total gain.

Take *Bombina bombina* for example. Firstly run the model and Jacknife test by using 22 variables (Figure 3-2). It can be clearly seen that Corine contributed very little on the total gain and can be removed. After omitting Corine, the minimum air temperature in sea 3 (Jul-Sep) showed little effect on the total gain (Figure 3-3), which can be excluded as the second step. The process were continued until the all the remaining variables have positive impact on the total gain.

The multi-colinearity could also be detected and reduced during this process. The total gain will not vary much if one of the correlated variables is excluded since the remaining correlated variables have similar effect on the model. In this case, we can omit this variable and the multi-correlation decrease as well.

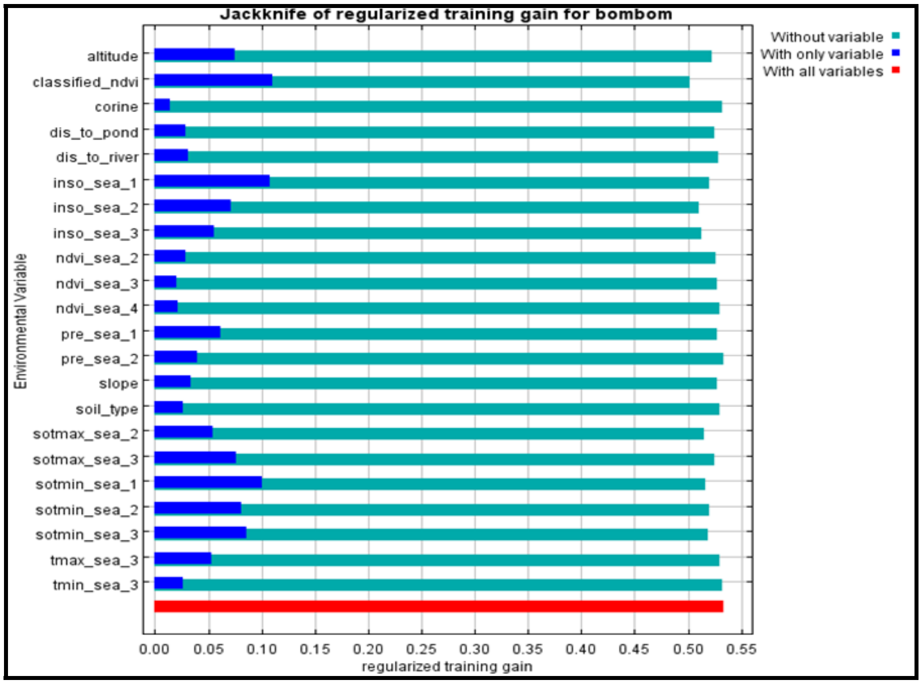


Figure 3-2 Jacknife test results for *Bombina bombina* by using 22 variables

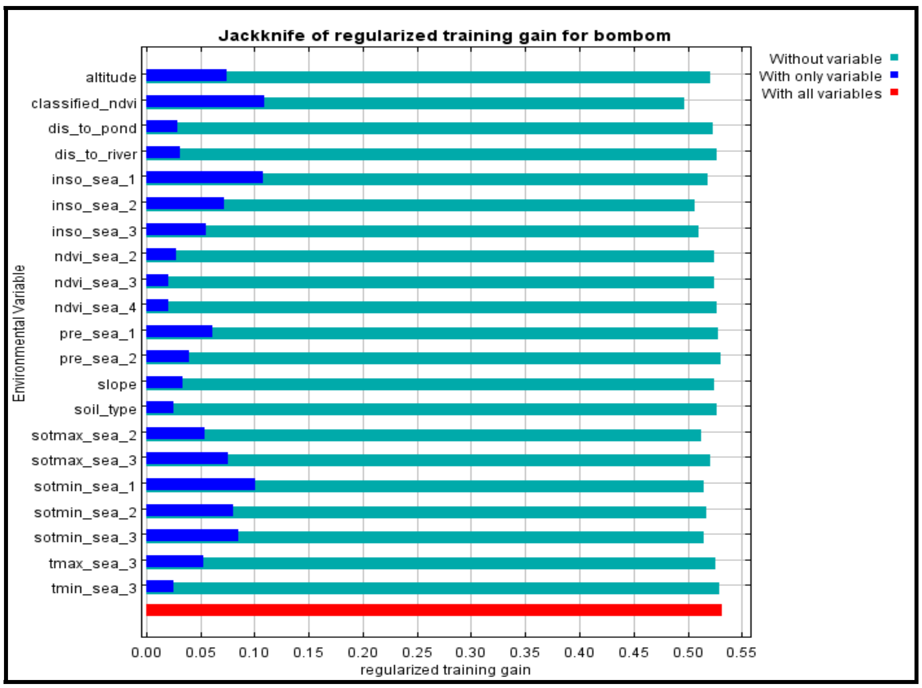


Figure 3-3 Jacknife test results for *Bombina bombina* after removing Corine

3.2.4. Presence-Absence Map and Amphibian Species Richness

The environmental predictors (Section 3.2.3) were used to run the Maxent model. The output is a continuous occurrence probability map. A threshold must be defined to determine the presence or absence of a species.

There are many studies working on selecting the optimal threshold for binary predictions of presence-only models, but no dominant rules has been developed yet (Phillips *et al*., 2006; Liu *et al*., 2005). Phillips *et al*. (2006) used the minimum value of training sampling presence points to decide the presence or absence of species. Objective approaches like maximum Kappa (Guisan *et al*., 1998) are frequently used. The 10 percentile training presence value was used as the threshold (Table3-1), which means that the ten percent of each species records with the lowest predicted model values will fall into the absence regions, and the presence regions will encompass the other 90% of the distribution records (Cameron *et al*., 2008). Using the cut off value, the occurrence probability map can be classified into binary (1 or 0) or presence-absence and the distribution for each species will be determined. Amphibian species richness for the whole country was calculated by overlaying the distribution map of all species.

Table 3-1 The 10 percentile training presence threshold for each species

No.	Species	Threshold	No.	Species	Threshold
1	*Bombina bombina*	0.452	10	*Rana esculenta*	0.463
2	*Bombina variegata*	0.546	11	*Rana lessonae*	0.529
3	*Bufo bufo*	0.441	12	*Rana ridibunda*	0.623
4	*Bufo calamita*	0.503	13	*Rana temporaria*	0.438
5	*Bufo viridis*	0.439	14	*Salamandra salamandra*	0.381
6	*Hyla arborea*	0.440	15	*Triturus alpestris*	0.590
7	*Pelobates fuscus*	0.450	16	*Triturus cristatus*	0.455
8	*Rana arvalis*	0.468	17	*Triturus montandoni*	0.394
9	*Rana dalmatina*	0.585	18	*Triturus vulgaris*	0.427

3.2.5. Model Evaluation

3.2.5.1. Receiver Operating Characteristics (ROC) Curve

Receiver Operating Characteristics (ROC) curve is a threshold independent method developed in signal processing and widely used for distribution model evaluation (Fielding & Bell, 1997; Graham & Hijmans, 2006; Phillips *et al*., 2006).

An ROC curve is obtained by plotting sensitivity (true positive rate) on the y-axis and 1-specificity (false positive rate) on the x-axis for all possible thresholds.

Sensitivity and specificity are determined by cross tabulating observed and predicted values of a model in a confusion matrix. For each particular threshold, sensitivity is the fraction of all positive samples that are classified as truth, while specificity is the fraction of all negative samples which are classified as false. The area under curve (AUC) value indicates the model accuracy (Phillips *et al*., 2006). For random prediction, AUC is 0.5.

If only presence data is available, ROC curves seems to be inapplicable since absence data is required to calculate specificity. To avoid this problem, Phillips *et al*. (2006) introduced a new concept of "distinguishing presence from random rather than presence from absence". For each pixel in the study area, a negative instance (random) is defined while a positive instance is defined for the pixels containing presence data. Then the model makes predictions without looking at the value (For further explanation, see Phillips *et al*., 2006). This process can also be interpreted as using pseudo-absence in ROC analysis (Wiley et al., 2003).

3.2.5.2. Kappa Statistics

Cohen's Kappa coefficient is a statistical measure of inter-rater agreement for categorical items (Cohen, 1960). It is a more robust measure than simple overall accuracy since it excludes the agreement occurring by chance (Liu *et al*., 2005). Kappa value can be calculated based on an error matrix (Table 3-2).

Table 3-2 Error matrix (observation and prediction)

		Observation		
		Presence	Absence	
Prediction	Presence	a	b	a + b
	Absence	c	d	c + d
		a + c	b + d	n

Where
a is number of correctly predicted occurrences
b is number of incorrectly predicted occurrences (commission error)
c is number of incorrectly predicted absences (omission error)
d is number of correctly predicted absences

The relative observed agreement (P_o) is:

$$P_0 = \frac{a+d}{n} \qquad \textbf{Equation 3-1}$$

The hypothetical probability of chance agreement (P_e) is:

$$P_e = p_{\cdot 1} \times p_{1\cdot} + p_{\cdot 2} \times p_{2\cdot} \qquad \textbf{Equation 3-2}$$

Where

$$p_{\cdot 1} = \frac{a+c}{n}$$ **Equation 3-3**

$$p_{1\cdot} = \frac{a+b}{n}$$ **Equation 3-4**

$$p_{\cdot 2} = \frac{b+d}{n}$$ **Equation 3-5**

$$p_{2\cdot} = \frac{c+d}{n}$$ **Equation 3-6**

The equation for Kappa is:

$$k = \frac{P_0 - P_e}{1 - P_e}$$ **Equation 3-7**

If the groups are in complete agreement, k=1; if there is no agreement among the groups, k≤0. Kappa value can be interpreted as in Table 3-3 (Landis & Koch, 1977).

Table 3-3 Interpretation of Kappa value

Kappa Value	Interpretation
< 0	No agreement
0.0 — 0.20	Slight agreement
0.21 — 0.40	Fair agreement
0.41 — 0.60	Moderate agreement
0.61 — 0.80	Substantial agreement
0.81 — 1.00	Almost perfect agreement

SPSS 16.0 software was use to compute Kappa value by selecting Analysis → Descriptive statistics → *Crosstabs tool*.

3.2.5.3. Sensitivity of Local Occurrence Data

A low omission rate (1-sensitivity) is a necessary condition for a good model (Anderson *et al.*, 2003). The distribution model was also validated by another independent dataset (local survey data, Section 2.3) by calculate the sensitivity of presence points for each species.

3.3. Results

3.3.1. Environmental Predictors Selection

Based on Section 3.2.3, environmental predictors for each species were obtained and listed in Table 3-4. The Jacknife test results were in Appendix A.

Table 3-4 Environmental Predictors[1] selection

No.	Species	Environmental Predictors
1	*Bombina bombina*	altitude, classified_ndvi, dis_to_pond, inso_sea_1, inso_sea_2, inso_sea_3, ndvi_sea_2, ndvi_sea_3, pre_sea_1, pre_sea_2, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2, sotmax_sea_3, tmax_sea_3
2	*Bombina variegata*	altitude, pre_sea_2, sotmin_sea_3
3	*Bufo bufo*	altitude, classified_ndvi, dis_to_pond, dis_to_river, inso_sea_1, inso_sea_2, inso_sea_3, ndvi_sea_2, ndvi_sea_3, pre_sea_1, pre_sea_2, soil_type, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2, sotmax_sea_3, tmax_sea_3, tmin_sea_3
4	*Bufo calamita*	altitude, classified_ndvi, dis_to_pond, dis_to_river, inso_sea_1, inso_sea_3, ndvi_sea_2, ndvi_sea_3, ndvi_sea_4, pre_sea_2, slope, soil_type, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2, tmax_sea_3, tmin_sea_3
5	*Bufo viridis*	altitude, classified_ndvi, dis_to_pond, inso_sea_1, inso_sea_2, inso_sea_3, ndvi_sea_2, ndvi_sea_3, ndvi_sea_4, pre_sea_1, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2
6	*Hyla arborea*	classified_ndvi, dis_to_pond, inso_sea_1, inso_sea_3, ndvi_sea_2, ndvi_sea_3, ndvi_sea_4, pre_sea_2, soil_type, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2, sotmax_sea_3, tmax_sea_3
7	*Pelobates fuscus*	altitude, classified_ndvi, dis_to_pond, inso_sea_1, inso_sea_2, inso_sea_3, ndvi_sea_2, pre_sea_1, pre_sea_2, slope, soil_type, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2, sotmax_sea_3, tmin_sea_3
8	*Rana arvalis*	altitude, classified_ndvi, dis_to_pond, dis_to_river, inso_sea_1, inso_sea_2, inso_sea_3, ndvi_sea_3, pre_sea_2, slope, soil_type, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2, sotmax_sea_3, tmax_sea_3
9	*Rana dalmatina*	classified_ndvi, corine, inso_sea_1, ndvi_sea_3, tmin_sea_3

[1] The descriptions of the environmental predictors are the same as Table 2-6.

10	*Rana esculenta*	classified_ndvi, dis_to_pond, dis_to_river, inso_sea_1, inso_sea_2, ndvi_sea_2, ndvi_sea_3, pre_sea_1, pre_sea_2, soil_type, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2, sotmax_sea_3, tmin_sea_3
11	*Rana lessonae*	altitude, classified_ndvi, dis_to_pond, inso_sea_1, inso_sea_2, inso_sea_3, ndvi_sea_3, pre_sea_2, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2, sotmax_sea_3, tmax_sea_3
12	*Rana ridibunda*	altitude, classified_ndvi, dis_to_pond, dis_to_river, inso_sea_1, inso_sea_2, ndvi_sea_3, pre_sea_1, pre_sea_2, slope, sotmin_sea_1, sotmin_sea_3, sotmax_sea_2, tmax_sea_3
13	*Rana temporaria*	altitude, classified_ndvi, corine, dis_to_pond, dis_to_river, inso_sea_1, inso_sea_2, inso_sea_3, ndvi_sea_2, ndvi_sea_3, ndvi_sea_4, pre_sea_1, pre_sea_2, slope, soil_type, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2, sotmax_sea_3, tmax_sea_3, tmin_sea_3
14	*Salamandra salamandra*	altitude, pre_sea_2, soil_type
15	*Triturus alpestris*	altitude, classified_ndvi, inso_sea_3, pre_sea_2, sotmin_sea_3
16	*Triturus cristatus*	altitude, classified_ndvi, inso_sea_1, inso_sea_2, pre_sea_1, pre_sea_2, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2, sotmax_sea_3, tmax_sea_3, tmin_sea_3
17	*Triturus montandoni*	altitude, pre_sea_2, sotmin_sea_3
18	*Triturus vulgaris*	altitude, classified_ndvi, inso_sea_1, inso_sea_2, inso_sea_3, ndvi_sea_2, ndvi_sea_4, pre_sea_2, sotmin_sea_1, sotmin_sea_2, sotmin_sea_3, sotmax_sea_2

3.3.2. Presence-absence Map and Amphibian Species Richness

The species occurrence probability maps were shown in Figure 3-4.

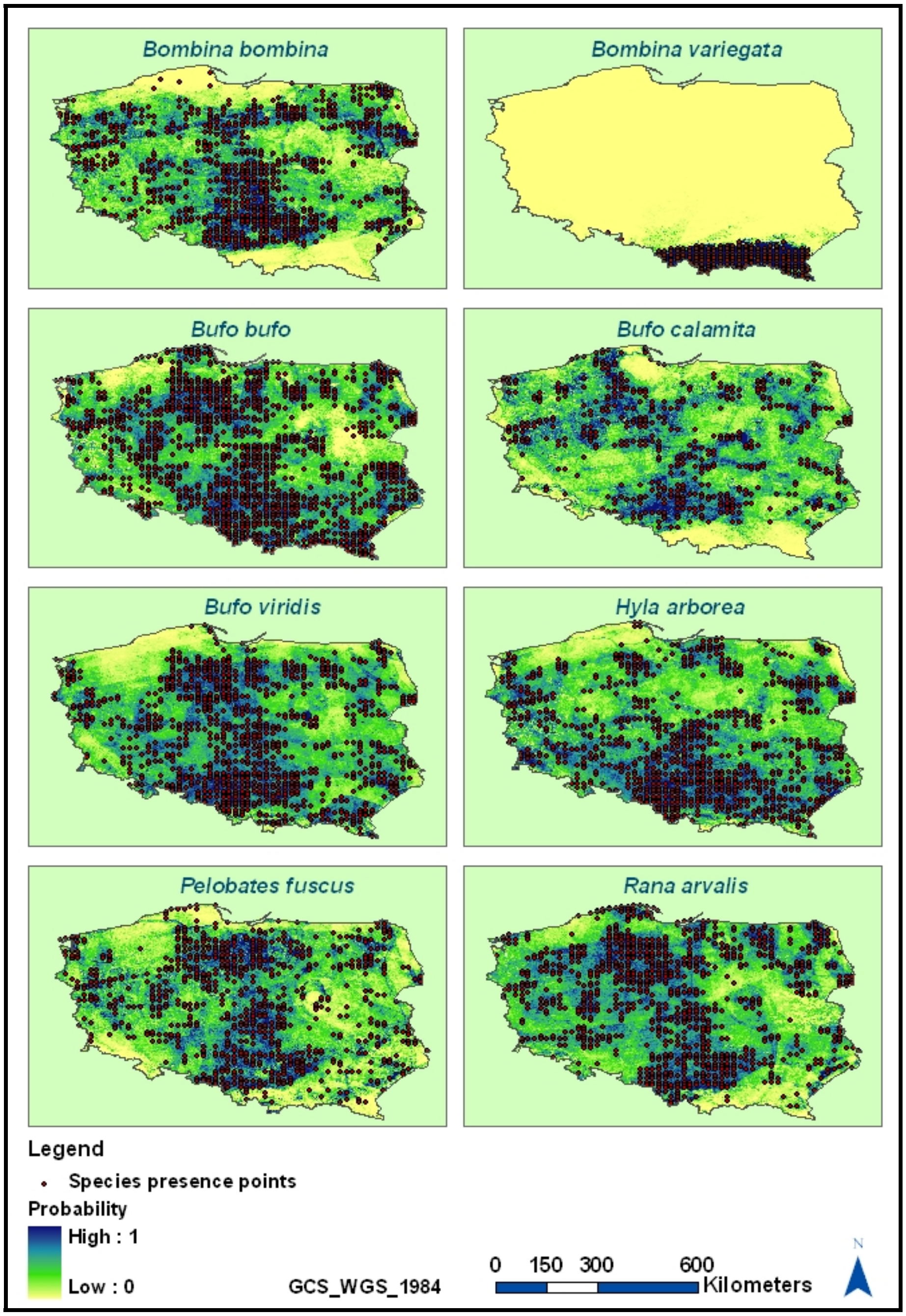


Figure 3-4 Amphibian species occurrence probability map in Poland

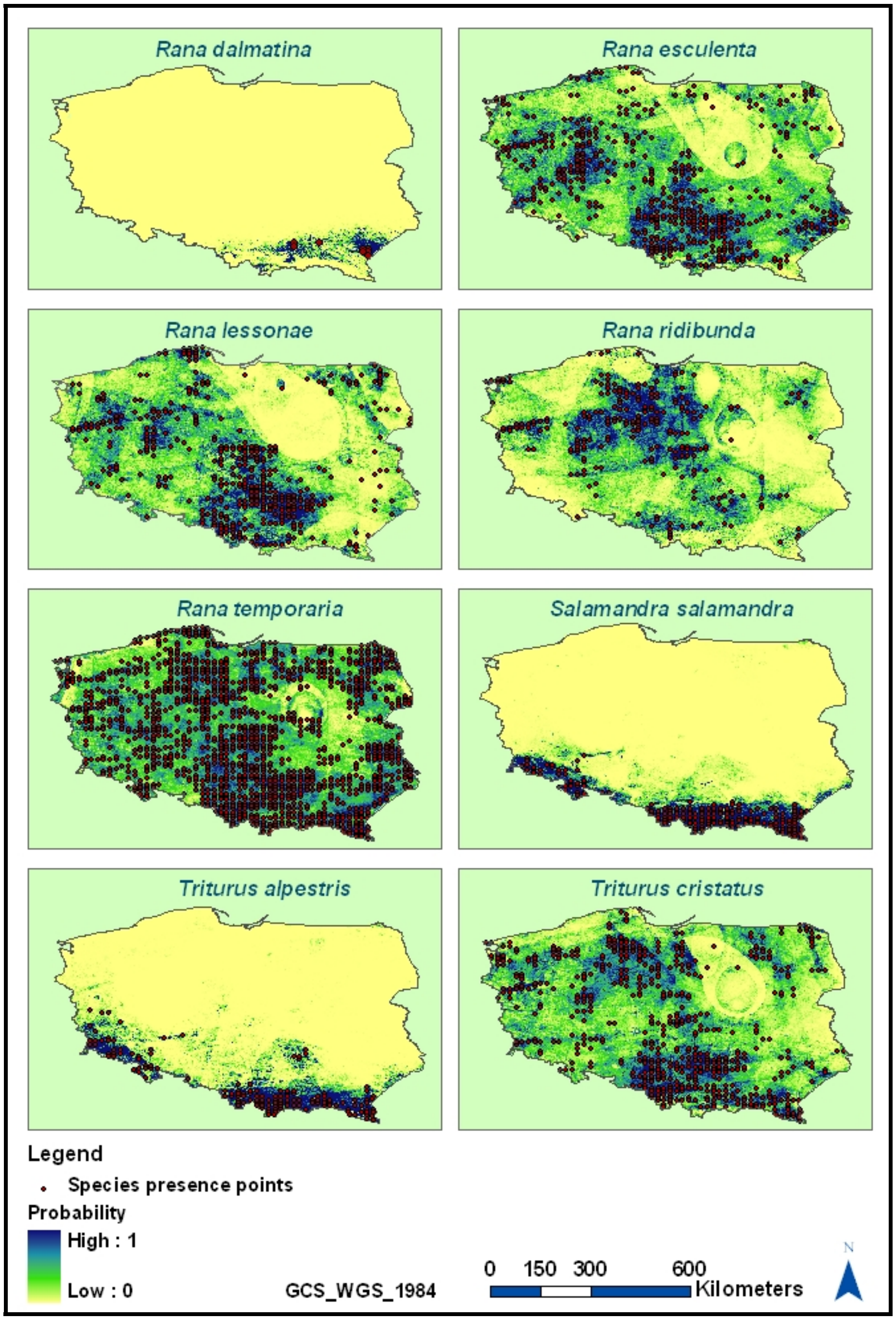


Figure 3-4 Amphibian species occurrence probability map in Poland (continued)

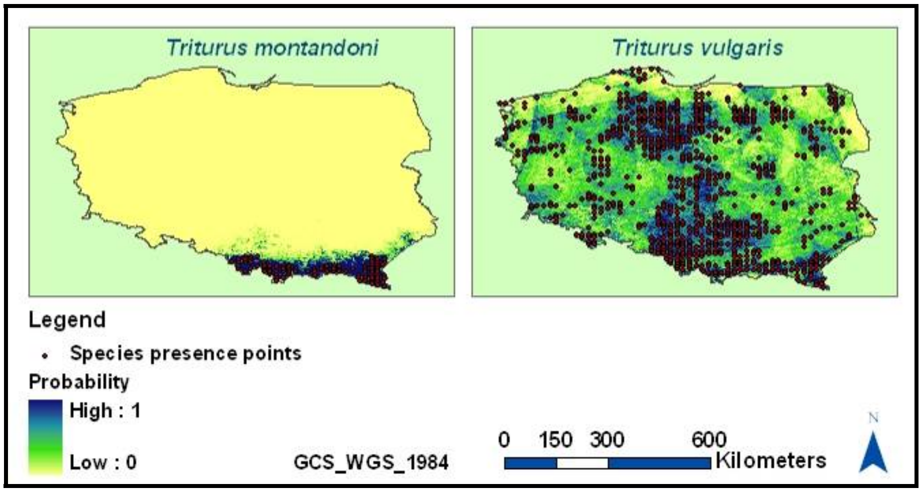


Figure 3-4 Amphibian species occurrence probability map in Poland (continued)

Using the threshold determined in Section 3.2.4 to reclassify the probability map, the species presence-absence maps were generated (Figure 3-5).

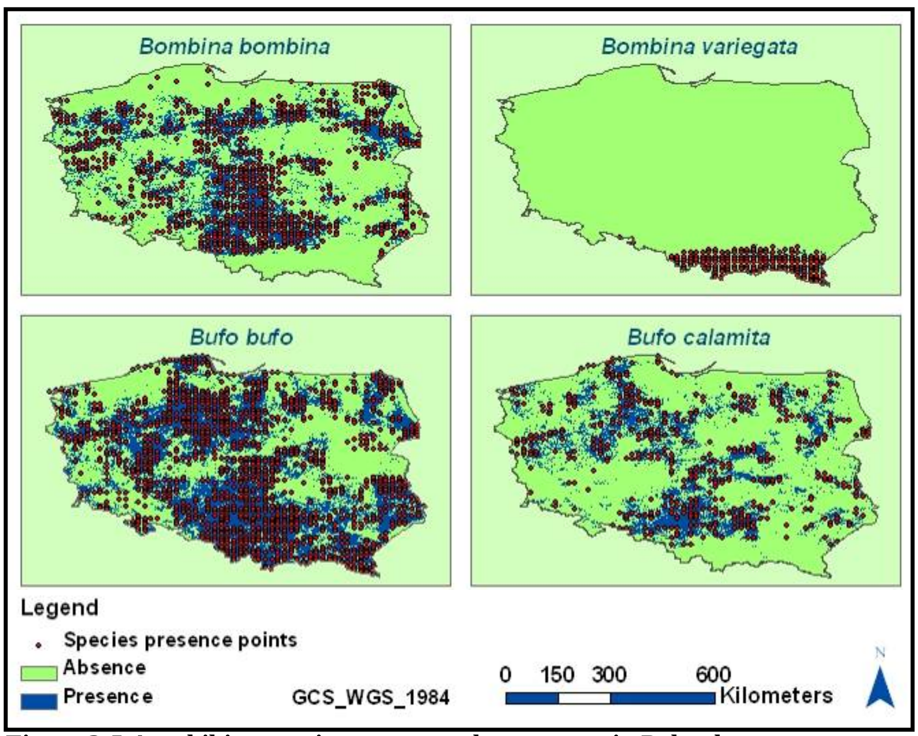


Figure 3-5 Amphibian species presence-absence map in Poland

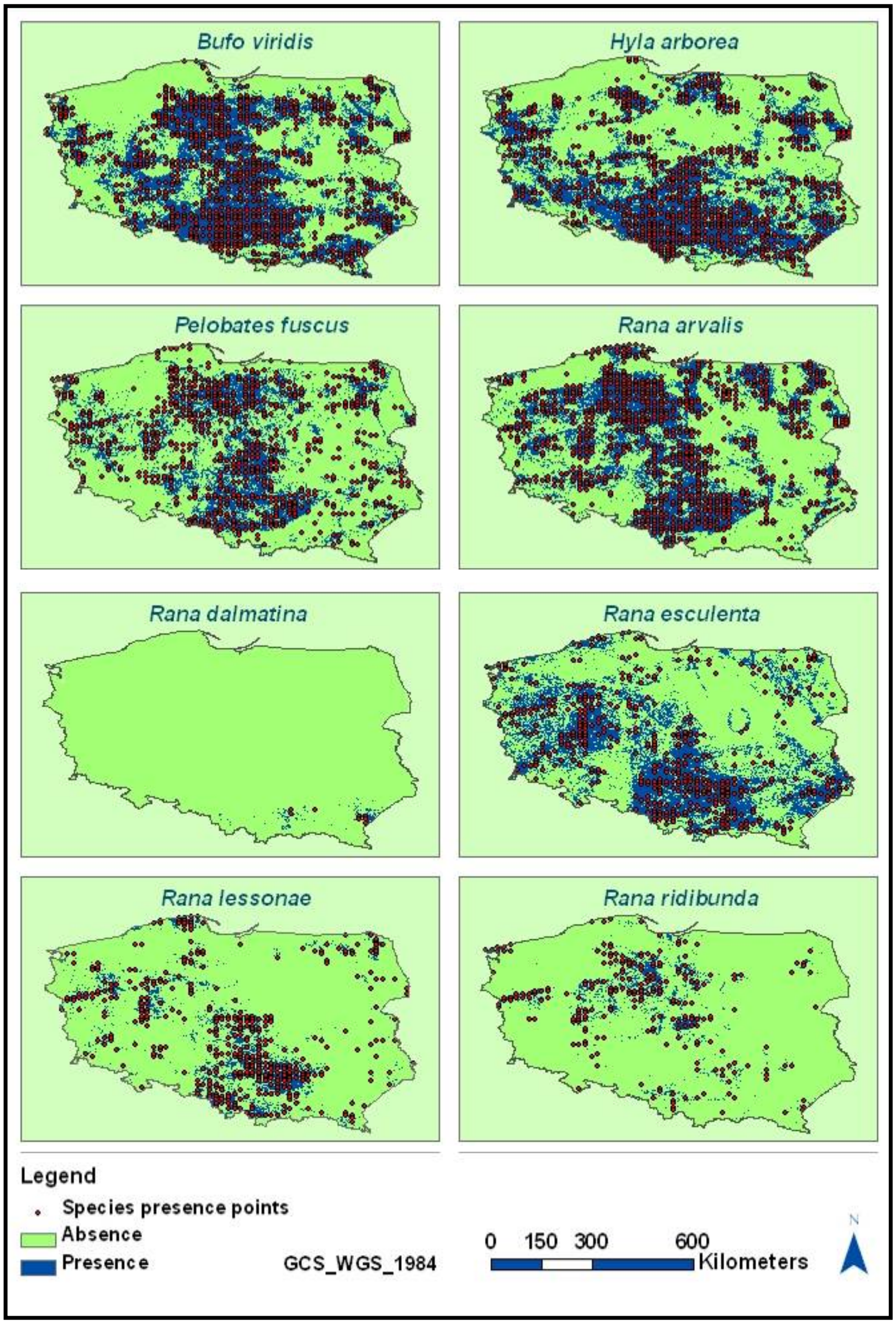


Figure 3-5 Amphibian species presence-absence map in Poland (continued)

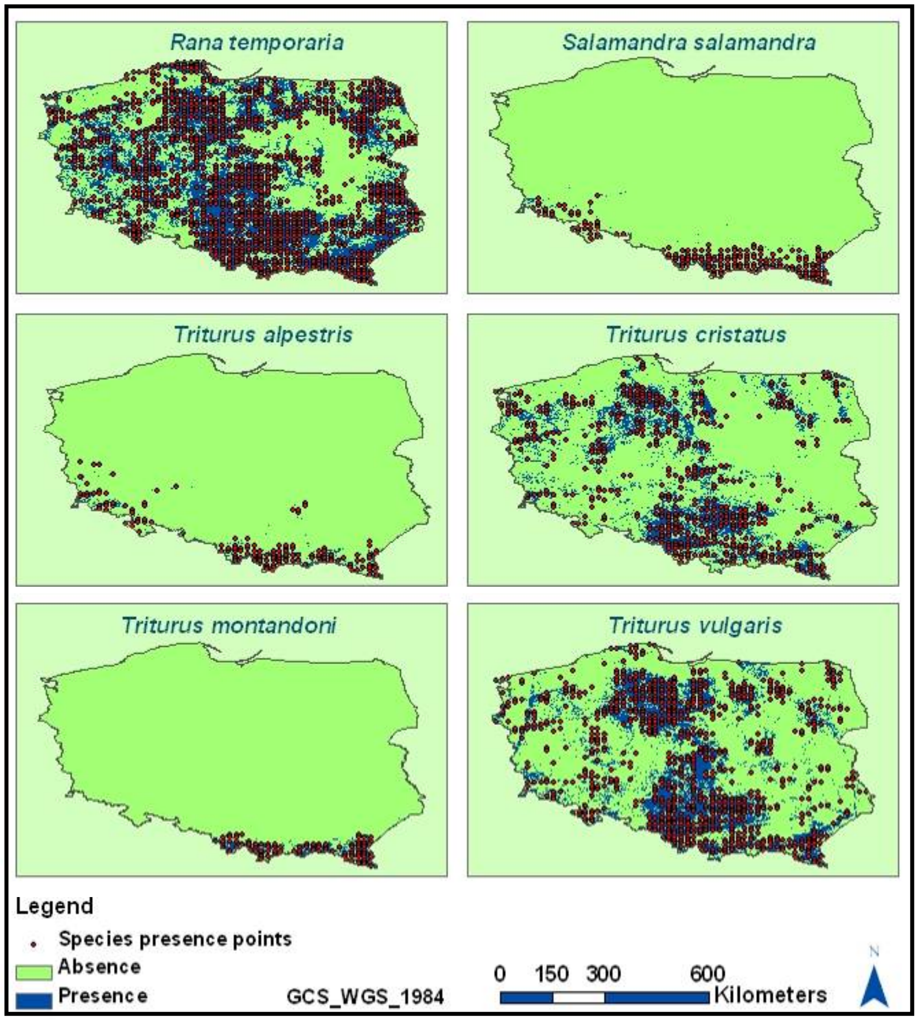


Figure 3-5 Amphibian species presence-absence map in Poland (continued)

From Figure 3-4, it can be clearly seen that most sample points located in the high-probability area. Only a few points scattered in the low-probability area. Figure 3-5 showed the similar pattern. All the presence-absence maps were overlaid to generate the amphibian species richness distribution (Figure 3-6). The areas of high species richness turned out to be the south-central part of the country.

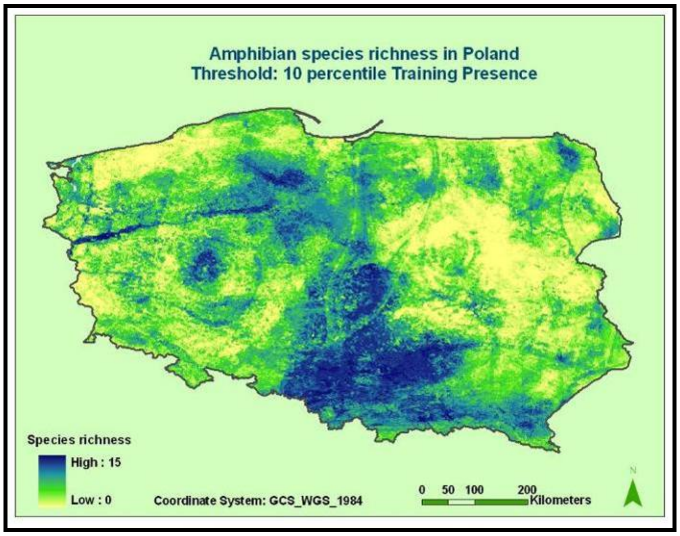


Figure 3-6 Amphibian Species Richness in Poland

3.3.3. Model Evaluation

3.3.3.1. ROC/AUC

AUC value for all the species were listed in Table 3-5.

Table 3-5 AUC and Kappa value for Amphibian Species

No.	Species	AUC (training)	AUC (testing)	Kappa
1	*Bombina bombina*	0.892	0.681	0.557
2	*Bombina variegata*	0.991	0.977	0.842
3	*Bufo bufo*	0.821	0.650	0.479
4	*Bufo calamita*	0.935	0.667	0.507
5	*Bufo viridis*	0.860	0.678	0.512
6	*Hyla arborea*	0.863	0.681	0.508
7	*Pelobates fuscus*	0.878	0.649	0.468
8	*Rana arvalis*	0.874	0.684	0.547
9	*Rana dalmatina*	0.998	0.994	0.886
10	*Rana esculenta*	0.919	0.716	0.469
11	*Rana lessonae*	0.954	0.731	0.556
12	*Rana ridibunda*	0.981	0.763	0.578
13	*Rana temporaria*	0.811	0.639	0.457
14	*Salamandra salamandra*	0.971	0.950	0.645
15	*Triturus alpestris*	0.987	0.960	0.622
16	*Triturus cristatus*	0.907	0.759	0.496
17	*Triturus montandoni*	0.983	0.976	0.677
18	*Triturus vulgaris*	0.866	0.683	0.506

For training data, the AUC values of all the 18 species were higher than 0.8; 10 of them were higher than 0.9. For testing data, 5 of them were higher than 0.9; the rest ones fluctuated between 0.6-0.8. The ROC curves were shown in Appendix B.

3.3.3.2. Kappa Results

Kappa values were also listed in Table 3-5. According to Table 3-3, the results for all 18 species were classified in three different agreement levels (Table 3-6).

Table 3-6 Kappa value for Amphibian Species

Class	Kappa Value	Interpretation	No. of Species
1	0.41 — 0.60	Moderate agreement	13
2	0.61 — 0.80	Substantial agreement	3
3	0.81 — 1.00	Almost perfect agreement	2

3.3.3.3. Sensitivity of Local Occurrence Data

The sensitivities of local presence points for 14 species were listed in Table 3-7. The results showed large variation: for *Bombina bombina, Hyla arborea* and *Rana temporaria,* more than 90% presence points were correctly predicted; for *Bufo calamita* and *Triturus alpestris,* they are rare in this area (2 and 1 occurrences) (Bonk & Pabijan, 2006; Bonk & Pabijan, 2008), the true positive rate were zero; for *Rana ridibunda*, the sensitivity was also zero. The reason will be discussed in Chapter 4.

Table 3-7 Sensitivity of species presence data

No.	Species	Observed presence points	Correctly predicted points	Sensitivity (%)
1	*Bombina bombina*	31	30	96.8%
3	*Bufo bufo*	55	46	83.6%
4	*Bufo calamita*	2	0	0.0%
5	*Bufo viridis*	24	14	58.3%
6	*Hyla arborea*	35	33	94.3%
7	*Pelobates fuscus*	21	17	81.0%
8	*Rana arvalis*	30	21	70.0%
10	*Rana esculenta*	37	28	75.7%
11	*Rana lessonae*	19	13	68.4%
12	*Rana ridibunda*	14	0	0.0%
13	*Rana temporaria*	50	50	100.0%
15	*Triturus alpestris*	1	0	0.0%
16	*Triturus cristatus*	16	12	75.0%
18	*Triturus vulgaris*	30	26	86.7%

4. Discussio ns

4.1. Maxent Model Performance

4.1.1. Binomial Test (*Hypothesis 1 testing*)

One-tailed binomial test can be used to identify whether the model predicted more accurately than random (Anderson *et al*., 2002). Binomial test procedure compares the observed frequencies of the two categories to the frequencies that are expected under a binomial distribution with a specified probability value.

If: there are t presence points, the omission rate is r and the proportional predicted area is a. "One-tailed binomial test can be used to determine the probability of having at least t× (1-r) successes out of the t trails, each with probability a." (Phillips *et al*., 2006)

Binomial tool in SPSS 16.0 software was used to achieve it (Table 4-1). With 95% confidence level ($P<0.05$), it can be concluded that the proportion of correctly predicted species presence is significantly higher by a distribution model than by a random prediction all the 18 species. The null hypothesis is rejected.

Table 4-1 Binomial test for the difference of sensitivity predicted by distribution model and random

No.	Species	P-value	No.	Species	P-value
1	*Bombina bombina*	<0.0001	10	*Rana esculenta*	<0.0001
2	*Bombina variegata*	<0.0001	11	*Rana lessonae*	<0.0001
3	*Bufo bufo*	<0.0001	12	*Rana ridibunda*	<0.0001
4	*Bufo calamita*	<0.0001	13	*Rana temporaria*	<0.0001
5	*Bufo viridis*	<0.0001	14	*Salamandra salamandra*	<0.0001
6	*Hyla arborea*	<0.0001	15	*Triturus alpestris*	<0.0001
7	*Pelobates fuscus*	<0.0001	16	*Triturus cristatus*	<0.0001
8	*Rana arvalis*	<0.0001	17	*Triturus montandoni*	<0.0001
9	*Rana dalmatina*	<0.0001	18	*Triturus vulgaris*	<0.0001

4.1.2. Model Performance and Kappa Value

The binomial test (Section 4.1.1) and ROC/AUC values (Section 3.3.3.1) show that the Maxent model performs better than random prediction for all the 18 amphibian species. The sensitivity calculation of local dataset also shows low omission errors for most species. Another advantage of the Maxent model is that it works well with small sample size compared with other species models. Take *Rana dalmatina* for example, only 8 sample points were available (6 for training and 2 for testing), however, the training AUC is 0.998 and the testing AUC was 0.994.

As a niche modelling approach, the Maxent model determines the presence of species by identifying the areas closely associated with observed presence localities (Guisan & Thuiller, 2005). When we calculate Kappa, random background points are selected as the pseudo-absences data. These locations might have similar environmental conditions as sample points and are predicted as presence, which leads to large false positive value (commission error) and low Kappa value.

The Kappa value also dependents on the proportion of all presences in the full validation dataset (Fielding & Bell, 1997; Allouche *et al*., 2006). Take *Rana dalmatina* for example. Kappa value was calculated by generating different number of random background points (Table 4-2). It can be clearly seen that the value reduces dramatically when the proportion of absence data become very large.

Table 4-2 Error Matrix of *Rana dalmatina* based on different number of pseudo-absence points

		Observation				
		Presence	Pseudo-absence			
		8 points	10 points	100 points	1000 points	3131 points
Prediction	Presence	7	0	0	7	14
	Absence	1	10	100	993	3117
	Kappa	-	0.8861	0.9284	0.6327	0.4808

The threshold for determining the presence or absence of species also affects the sensitivity and Kappa value. In table 3-7, the sensitivity of *Rana ridibunda* was 0 when the 10 percentile training presence value was used as the threshold. If we changed the threshold into the minimum training presence, 9 points of 14 were located in presence area. Thus sensitivity and Kappa value should be interpreted along with other information, e.g. quality of error matrix, threshold, when we do model evaluation.

Kappa value and AUC are also correlated with distribution range of species - the geographical area within which that species can be found. It could be described by the observed species distribution (occurrence data). Here the percentages of amphibian species occurrence area in the whole Poland are used to represent the distribution range (Figure 4-1). The higher the value is, the wider the distribution range could be. On the contrary, the lower the value is the narrower area the species occurs.

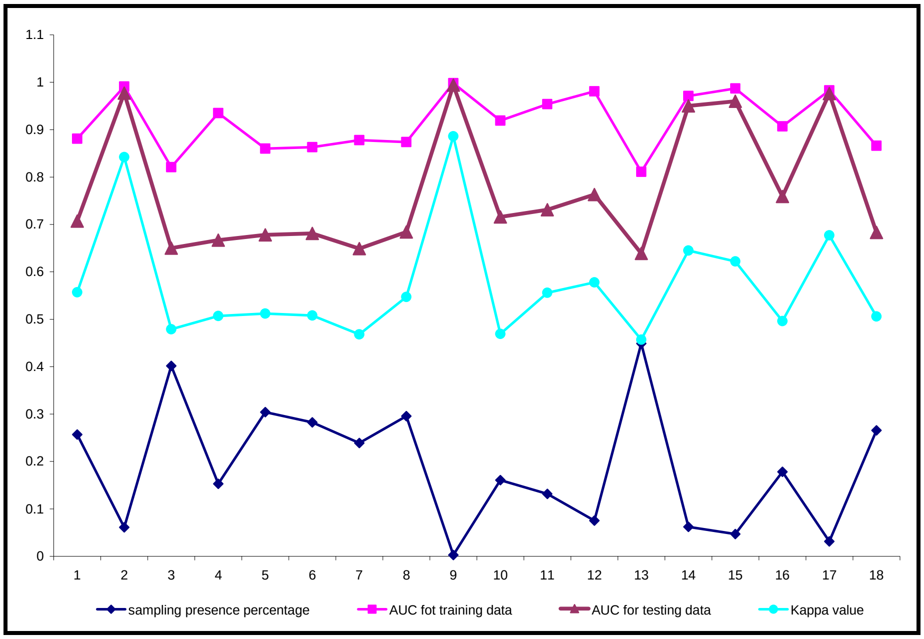


Figure 4-1 Species presence distribution, AUC and Kappa value

From Figure 4-1, the species which are wide spread-out have lower AUC value and Kappa result. The species which are limited in small area have higher AUC and Kappa value. As mentioned above (Section 3.2.1), the Maxent model predict a certain probability distribution which has the maximum entropy and subject to certain constraints derived from sampling points (Phillips *et al.*, 2006). For common species, it is difficult to differentiate the environmental requirements of the occurrences from the background areas. The accuracy of prediction could be low. If the species lives at specified area, it is easy to establish the relationship between presence points and environmental variables. The prediction result could be much more accurate.

4.2. Effect of Environmental Predictors on Model Performance

Based on Table 3-4 (Section 3.3.1), the frequency that each variable was selected as environmental predictor were summarized in Figure 4-2. The higher the frequency is, the more species that the variable may affect.

The mean min soil temperature at 5cm depth in summer and the mean precipitation in spring are the most important variables, used for 16 species out of the total 18. Classified NDVI, altitude and other soil temperature variables also affect the modelling process for most species. These variables may influence the Maxent model performance and prediction results.

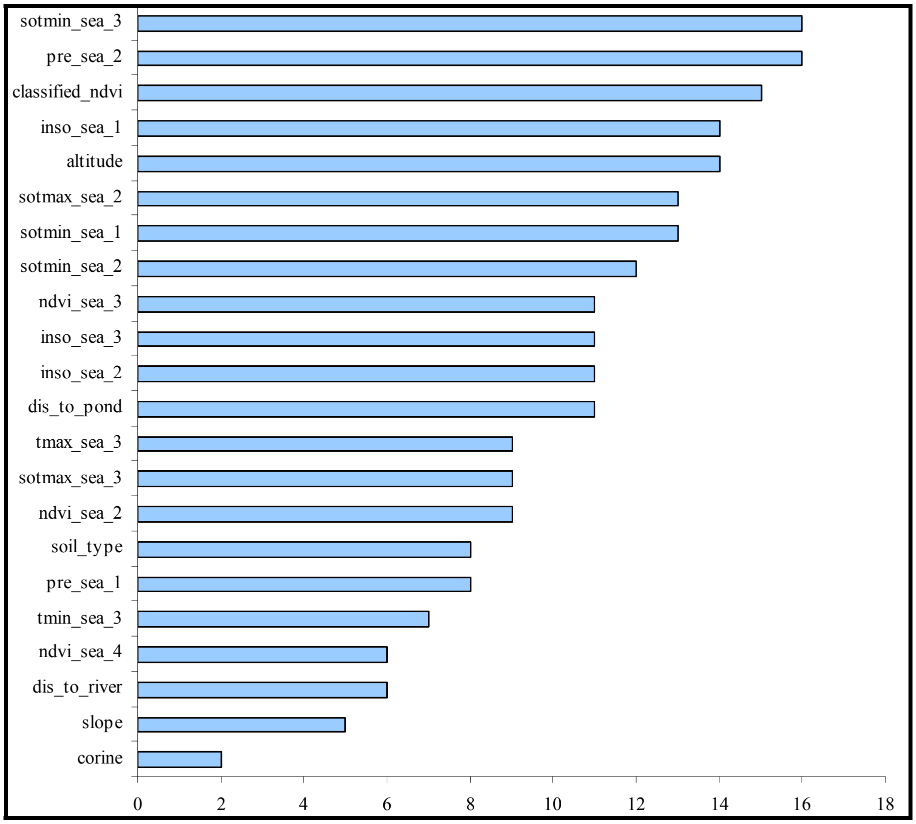


Figure 4-2 Frequency of each variable[5] selected as environmental predictors

[5] The descriptions of the variables are the same as in Table 2-6.

4.2.1. Effect of Soil Temperature Variables (*Hypothesis 2 testing*)

The effects of all the soil temperature variables (including the mean min soil temperature at 5cm depth in winter, spring and summer; the mean max soil temperature at 5cm depth in spring and summer) are studied here. Run the Maxent model by using environmental predictors with and without soil temperature layers and compare the results (Figure4-3, 4-4). The AUC value and Kappa reduce notably for all the species except *Triturus alpestris* and *Triturus montandoni*.

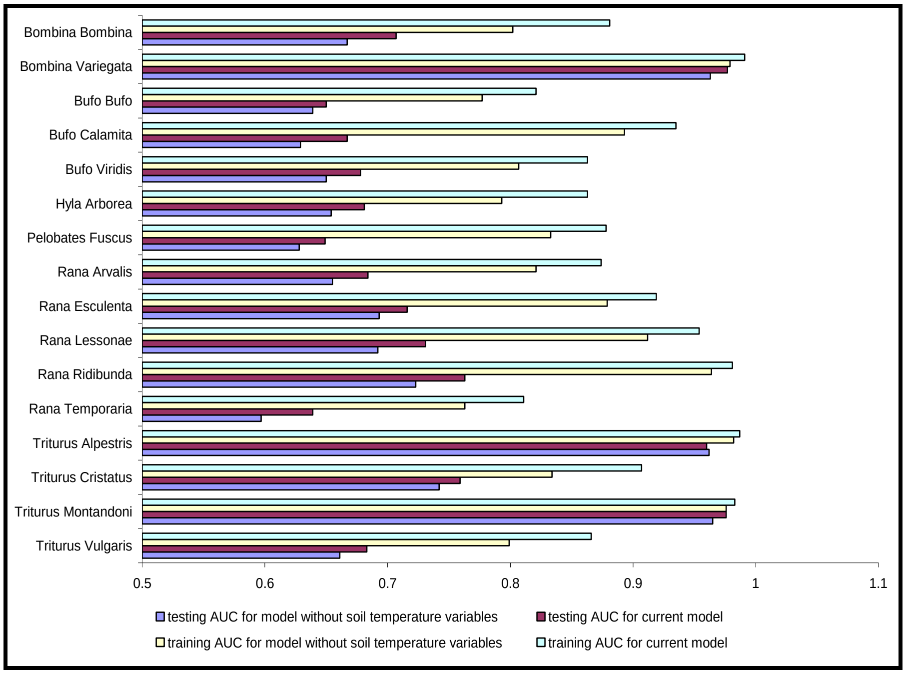


Figure 4-3 AUC of the models with and without soil temperature variables

The soil temperature variables affect prediction result by increasing the gain of the model. The interesting part is that the gain of other variables does not change after removing soil temperature variables, which means the contributions of independent variables to the model prediction are independent from each other. Consequently, the relative contribution of the remaining variables changes after removing soil temperature variables since the total gain has reduced. Take *Bombina bombina* for example. The comparison of percentage contribution values and Jacknife test was attached in Appendix C.

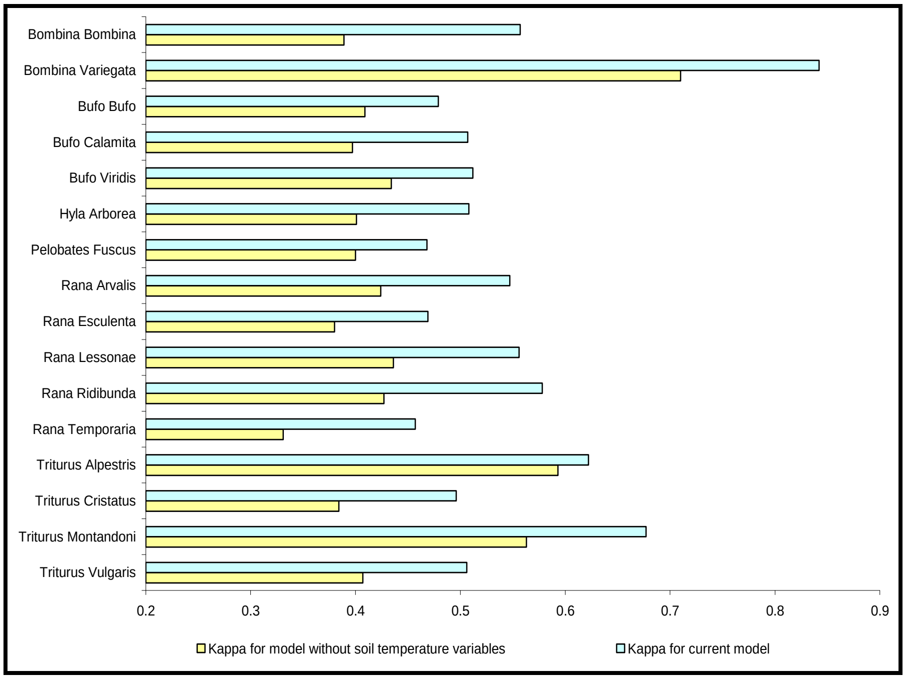


Figure 4-4 Kappa of the models with and without soil temperature variables

To test if the accuracy (Kappa value) of the presence-absence map predicted by the distribution model which includes all environmental predictors is higher than by the distribution model which does not include soil temperature predictors, the Z-value test is used to check the Kappa difference. The process includes three steps:

1. Generate error matrix

   r is number of rows and columns in error matrix

   X_{ii} is number of observations in row i and column i

   X_{i+} is marginal total of row i

   X_{+i} is marginal total of column i

   X_{ij} is number of observations in row i and column j

   X_{j+} is marginal total of row j

   N is total number of observations

2. Kappa statistics

$$\boxed{k = \frac{N\sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r} X_{i+}X_{+i}}{N - \sum_{i=1}^{r} X_{i+}X_{+i}} = \frac{\theta 1 - \theta 2}{1 - \theta 2}}$$

Equation 4-1

Where

$$\boxed{\theta 1 = \sum_{i=1}^{r} \frac{X_{ii}}{N}}$$

Equation 4-2

$$\boxed{\theta 2 = \sum_{i=1}^{r} \frac{X_{i+}X_{+i}}{N^2}}$$

Equation 4-3

$$\boxed{\sigma^2[k] = \frac{1}{N}\left[\frac{\theta 1(1-\theta 1)}{(1-\theta 2)^2} + \frac{2(1-\theta 1)(2\theta 1\theta 2 - \theta 3)}{(1-\theta 2)^3} + \frac{(1-\theta 1)^2(\theta 4 - 4\theta 2^2)}{(1-\theta 2)^4}\right]}$$

Equation 4-4

Where

$$\boxed{\theta 3 = \sum_{i=1}^{r} \frac{X_{ii}(X_{i+} + X_{+i})}{N^2}}$$

quation 4-5

$$\boxed{\theta 4 = \sum_{\substack{i=1 \\ j=1}}^{r} \frac{X_{ij}(X_{j+} + X_{+i})^2}{N^3}}$$

Equation 4-6

3. Calculate variance of kappa value

$$\boxed{Z = \frac{K1 - K2}{\sqrt{\sigma^2[K1] + \sigma^2[K2]}}}$$

Equation 4-7

The results are shown in Table 4-3. With 95% confidence level (Z<1.96), the Kappa values using model with and without soil temperature variables are significantly different. The null hypothesis is rejected. It can be concluded that the Kappa value of the presence-absence map predicted by the distribution model which includes all environmental predictors is higher than by the distribution model which does not include soil temperature predictors.

Table 4-3 Z test for the difference between Kappa values generated by model with and without soil temperature variables

No.	Species	Z value	No.	Species	Z value
1	*Bombina bombina*	17.19	10	*Rana esculenta*	9.36
2	*Bombina variegata*	31.39	11	*Rana lessonae*	48.25
3	*Bufo bufo*	6.49	12	*Rana ridibunda*	21.23
4	*Bufo calamita*	12.12	13	*Rana temporaria*	11.34
5	*Bufo viridis*	8.44	15	*Triturus alpestris*	5.37
6	*Hyla arborea*	10.48	16	*Triturus cristatus*	11.76
7	*Pelobates fuscus*	6.87	17	*Triturus montandoni*	24.84
8	*Rana arvalis*	12.29	18	*Triturus vulgaris*	15.59

4.3. Species Richness Distribution (*Hypothesis 3 testing*)

Species richness distribution can be generated by overlaying the presence-absence maps of all species (Section 3.3.2). Here species richness distributions are generated by both observation dataset and model results. The output maps are reclassified into 4 classes to represent different richness levels (Table 4-4, Figure4-5).

Table 4-4 Reclassification of species richness

Class	No. of Species	Species richness level
1	0-2	poor
2	3-6	fair
3	7-10	good
4	>10	excellent

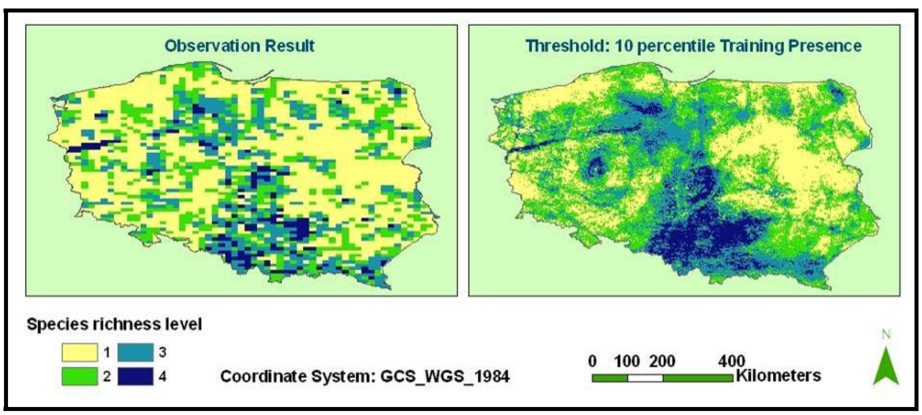


Figure 4-5 Observed and predicted amphibian species richness

From Figure 4-5, it can be seen that the two maps are in different spatial resolutions (the reason was described in Section 2.6). To get better comparisons of the maps, the layer generated from model prediction was re-sampled into the same size as the one derived from observation grids.

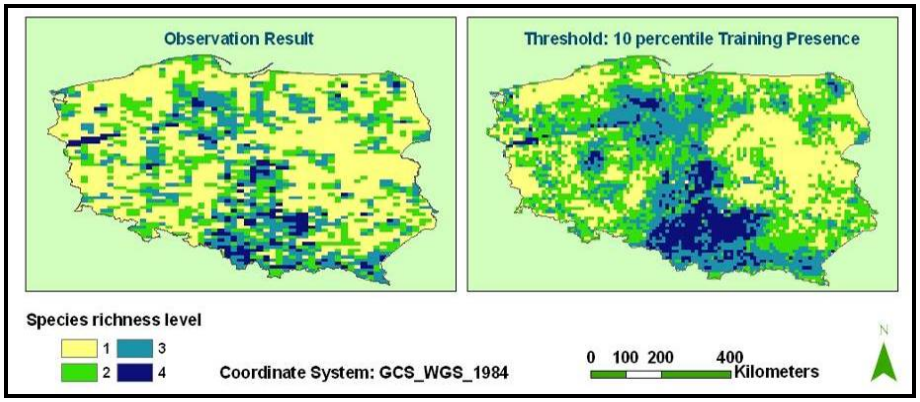


Figure 4-6 Observed and re-sampled predicted amphibian species richness

From both maps, the predicted richness distribution has similar spatial pattern as the observed one. However, the maps generated by model prediction results reveal more high-richness areas. To test the difference between observed species richness and predicted species richness, Chi-square test is applied by JMP 7 software. Both the predicted maps, before and after re-sampling, are compared with the observed map (Table 4-5, Figure 4-7).

Table 4-5 Chi-square test showing the difference of species richness maps generated by observation data and model prediction

Predicted map	DF	Chi-square	Prob>Chisq	RSquare(U)
without resample	9	152.688	<0.0001	0.2852
after resample	9	123.684	<0.0001	0.2293

With 95% confidence level, the predicted richness maps, both with and without re-sampling, are significantly different from the observed richness map. The prediction results show more higher-richness area. It can be concluded that the predicted richness is not the same as the observed richness. The null hypothesis is rejected.

This phenomena could be explained by two reasons: since observation data is presence-only data (Section 2.3) and background areas do not necessarily mean absence, the model result may predict the area which have similar environmental conditions with sampling points as presence area (under-sampling/over prediction); the season could be the effects of historical factors (e.g. limited dispersal, speciation, extinction). These factors act to limit species distribution to an area smaller than that in which its ecological needs are met (Soberon and Peterson, 2005). Even if some areas have the same environmental with sampling area and are predicted as presence, they might actually be absent due to historical effect. The output of the Maxent model shows the potential species habitat at landscape scale. The area which shows high richness could be used as reference information for conservation activities.

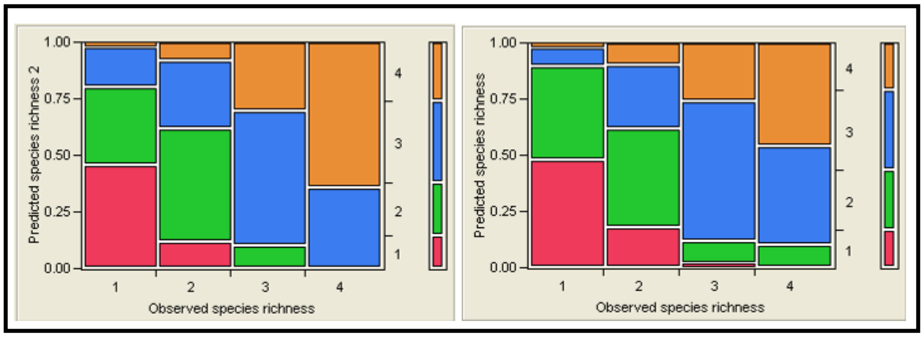


Figure 4-7 The Mosaic plot showing the percentage frequency occurrence

Another interesting part is the predicted species richness without re-sampling shows more association with the observed species richness than the one after re-sampling (Table 4-5). The reason could be that the re-sampling process reduces the data accuracy. The testing result could also be linked to the issue that the observation data and environmental parameters have different spatial resolution (Section 2.6). Instead of re-sampling the environmental layer into lower resolution to match the observation data, it is better to use it directly to make better use of the environmental information for modelling prediction.

4.4. Environmental Predictors Selection

The AUC results of current model (using environmental predictors) and model using 22 variables were listed in Figure 4-8. There were not much difference between AUC values but the current model used less variables.

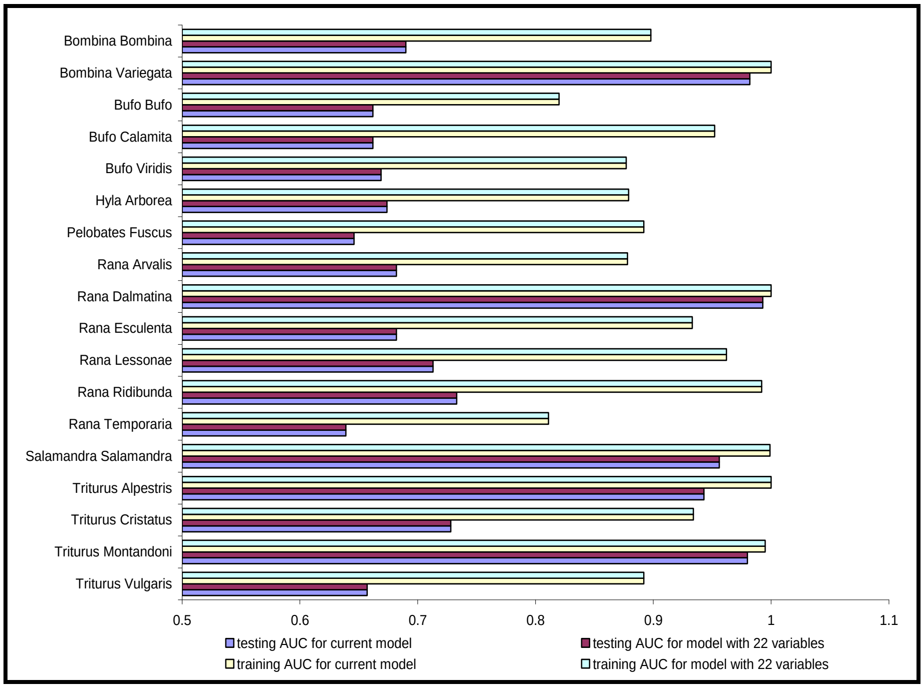


Figure 4-8 AUC of current model and model using 22 variables

Selecting environmental predictors to run the model has remarkable advantages:

1. Reduce input data with slightly loss of model accuracy.
2. Reduce the multi-correlations among remaining variables.

5. Overall Conclusion and Recommendation

5.1. Conclusion

The main objective of this research is to model the amphibian species distribution in Poland, identify the most important environmental predictors and predict the species richness distribution. The following general conclusions can be drawn.

- The accuracy (sensitivity) of the species distribution predicted by the Maxent model is higher than by a random prediction. For all the 18 species, AUC are higher than 0.8; among 10 of them, AUC are higher than 0.9.
- Soil temperature variables are the most important environmental predictors for most of the species (16). The Kappa value of the presence-absence map predicted by the distribution model which includes all environmental predictors is higher than by the distribution model which does not include soil temperature predictors.
- The mean predicted species richness is higher than the mean observed species richness because of under-sampling and historic reason, etc.
- Instead of using all available environmental data, select certain environmental predictors to run the model can reduce input data with slightly loss of model accuracy and reduce the multi-correlations among remaining variables as well.

5.2. Recommendation

- Soberon & Peterson (2005) described the factors which determine species distribution, including abiotic, biotic, dispersal and evolutionary capacity. Here only abiotic factors (e.g. climate, proximity to water) are concerned. It would be good if all these factors are incorporated together to predict the species distribution.
- Soil temperature and insolation variables were derived from climate stations using spatial interpolation (Hijmans, 2005) which may introduce uncertainty. If the soil temperature data with better quality is available, the model results might be improved.

6. References

Alford, R.A. & Richards, S.J. 1999. Global amphibian declines: a problem in applied ecology. *Annual Reviews in Ecology and Systematics*, 30, 133–165.

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.

Anderson, R. P., Lew, D. & Peterson, A. T. 2003. Evaluating predictive models of species distributions: criteria for selecting optimal models. *Ecological Modelling*, 162, 211–232.

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

Anderson, R.P., G´omez-Laverde, M. & Peterson, A.T., 2002. Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecol. Biogeogr*. 11, 131–141.

Arntzen, J.W. 2006. From descriptive to predictive distribution models: a working example with Iberian amphibians and reptiles. *Frontiers in Zoology*, 3(8).

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

Beltran-Abaunza, J.M., 2009. Method Development to Process Hyper Temporal Remote Sensing (RS) Images for Change Mapping. Master in Science Thesis, ITC, Enschede.

Bonk M., Pabijan M. 2006. Amphibians of Nida Basin - introduction to comparative studies in 1979-82 and 2006. VII Polish Herpetological Conference, Krakow.

Bonk M., Pabijan M. 2008. Amphibians of Southern Poland - comparative studies in 1979-82 and 2006-7. IX Polish Herpetological Conference, Krakow.

Cameron, A., Phillips, S., Golden1, C., Hijmans, R. J. & Louis, E. 2008. Identifying corridors for Lemur dispersal, for two future climate change scenarios.

Chefaoui, R. M., Hortal, J. & Lobo, J. M. 2005. Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian Copris species. *Biological Conservation*. 122, 327-338.

Cohen, J. 1960. A coefficient of agreement for nominal scales, Educational and Psychological Measurement conservation planning. *Journal of Biosciences*, 27, 37-46.

Corn, S. 2000. Amphibian research and monitoring initiative. *Froglog*, 41, 1–2.

Corsi, F., Skidmore, A. & de Leeuw, J. 2000. Modelling species distribution with GIS. In L. Boitani & T.K.Fuller (Eds). Research techniques in animal ecology (pp. 389-434). New York: Columbia University Press.

de Leeuw, A.J., Ottichilo, W.K., Toxopeus, A.G. & Prins, H.H.T. 2002. Application of remote sensing and geographic information systems in wildlife mapping and modelling. In A. Skidmore (Ed.), *Environmental modelling with GIS and remote sensing* (pp. 121-145). London: Taylor & Francis.

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

European Environment Agency. Corine Land cover 2000. [On line] Available from http://dataservice.eea.europa.eu/dataservice/[Accessed Oct 20, 2008]

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.

Gibbs, J.P. 1998. Distribution of woodland amphibians along a forest fragmentation gradient. *Landscape Ecology*, 13, 263-268.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T., 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol. Evol*. 19 (9), 497–503.

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

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

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

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

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones & A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International, *Journal of Climatology*. 25: 1965-1978 (pdf) (http://www.worldclim.org/)

Hirzel, A. H., Harsser J., Chessel D. & Perrin N. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data. *Ecology*. 83: 2027-2026.

Hutchinson, G.E. 1959. Homage to Rosalia or Why are there so many kinds of animals? *American Naturalist*, 870, 145-159.

Joly, P., Morand, C. & Cohas, A. 2004. Habitat fragmentation and amphibian conservation: building a tool for assessing landscape matrix connectivity. *Comptes-Rendus Biologie*, 326, 132-139.

Landis, J. R. & Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

Lennon, J.J., Greenwood, J.J.D. & Turner, J.R.G. 2000. Bird diversity and environmental gradients in Britain: a test of the species-energy hypothesis. *Journal of Animal Ecology*, 69, 581-598.

Linderman, M.A., An, L., Bearer, S., He, G., Ouyang, Z. & Liu, J. 2005. Modelling the spatio-temporal dynamics and interactions of households, landscapes, and giant panda habitat. *Ecological Modelling*, 183, 47-65.

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

Löfvenhaft, K., Runborg, S. & Sjogren-Gulve, P. 2004. Biotope patterns and amphibian distribution as assessment tools in urban landscape planning. *Landscape and Urban Planning*, 68, 403-427.

Manel, S., Dias, J.M. & Ormerod, S.J. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, 120, 337-347.

Mansfield, E.R. & Helms, B.H. 1982. Detecting Multicollinearity. *The American Statistician*, 36, 158-160.

Negga, H.E. 2007. Predictive modelling of amphibian distribution using ecological survey data: a case study of certral portugal. Master in Science Thesis, ITC, Enschede. 7-38.

Oindo, B.O., de By, R.A. & Skidmore, A.K. 2002. Interannual variability of NDVI and species richness in Kenya. I*nternational Journal of Remote Sensing*, 2, 172-180.

Olsen, L.M, Dale, V.H. & Foster, T. 2006. Landscape patterns as indicators of ecological change at Fort Benning, Georgia, USA. *Landscape and Urban Planning*, 79, 137-149.

Osborne, P.E. & Suárez-Seoane, S. 2002. Should data be partitioned spatially before building large-scale distribution models? *Ecological Modelling,* 157, 249-259.

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

Phillips, S.J., Anderson, R.P. & Schapire, R.E. 2006. Maximum entropy modeling of species geographical distribution. *Ecological Modeling,* 190, 231-259.

Pontius, R.G. & Batchu, K. 2003. Using the Relative Operating Characteristic to quantify certainty in Prediction of Location of Land Cover Change in India. *Transaction in GIS*, 7(4): 467-484.

Qian, H., Wang, X.H., Wang, S.L. & Li, Y.L. 2007. Environmental determinants of amphibian and reptile species richness in China. *Ecography*, 30, 471-482.

Sarkar, S. & Margules, C. 2002. Operationalizing biodiversity for Shannon, C.E. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.

Soberon, J. & Peterson, A.T. 2005. Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiversity Informatics*, 2, 1-10.

Stockwell, D. & Peter, D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Systems*, 13, 143-158.

Sun, Y. 2007. Influence of Human Activities on the Distribution of Amphibians and Reptiles in the Salamanca Province, Spain. Master in Science Thesis, ITC, Enschede, 3-40.

Tole, L. 2006. Choosing reserve sites probabilistically: A Colombian Amazon case study. *Ecological Modeling,* 194, 344–356.

Olivero, J., Real, R., Toxopeus, A.G. & Skidmore, A.K. 2009. Transferability of distribution models to a finer resolution scale: performance for the Favorability Function and Maxent. In print.

USGS. Online Database Access: http://water.usgs.gov/.[accessed on: Jan 10, 2009]

Weiers, S., Bock, M., Wissen, M. & Rossner, G. 2004. Mapping and indicator approaches for the assessment of habitats at different scales using remote sensing and GIS methods. *Landscape and Urban Planning*, 67, 43, 65.

Wiley, E.O., McNyset, K.M., Peterson, A.T., Robins, C.R. & Stewart, A.M., 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography*, 16 (3), 120–127.

Woodford, J.E. & Meyer, M.W. 2003. Impact of lakeshore development on green frog abundance. *Biological Conservation*, 110, 277-284.

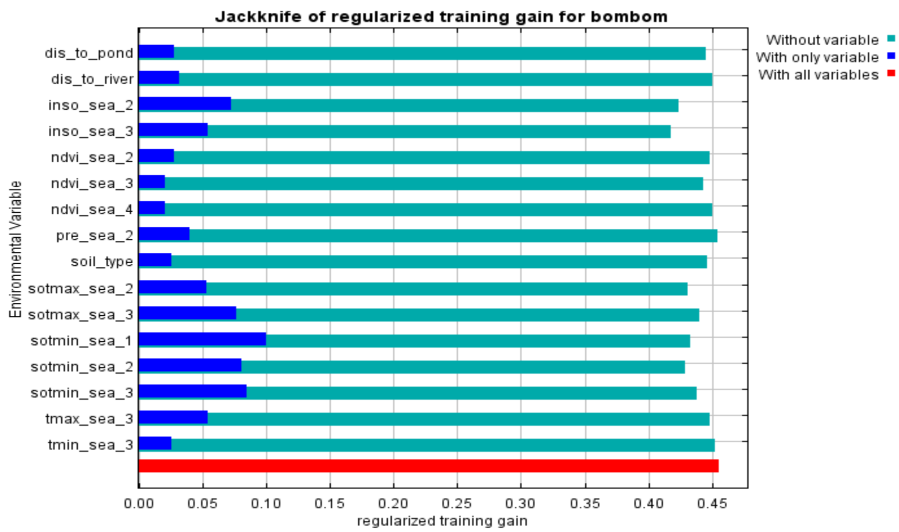
Wu, Y. 2006. Mapping amphibian distribution at national scale, using species environmental models. Msc Thesis. ITC. 4-26.

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

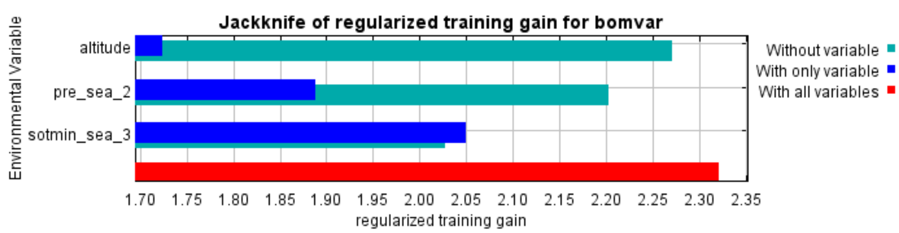
7. Appendixes

Appendixes A - Jacknife Test for Amphibian Species

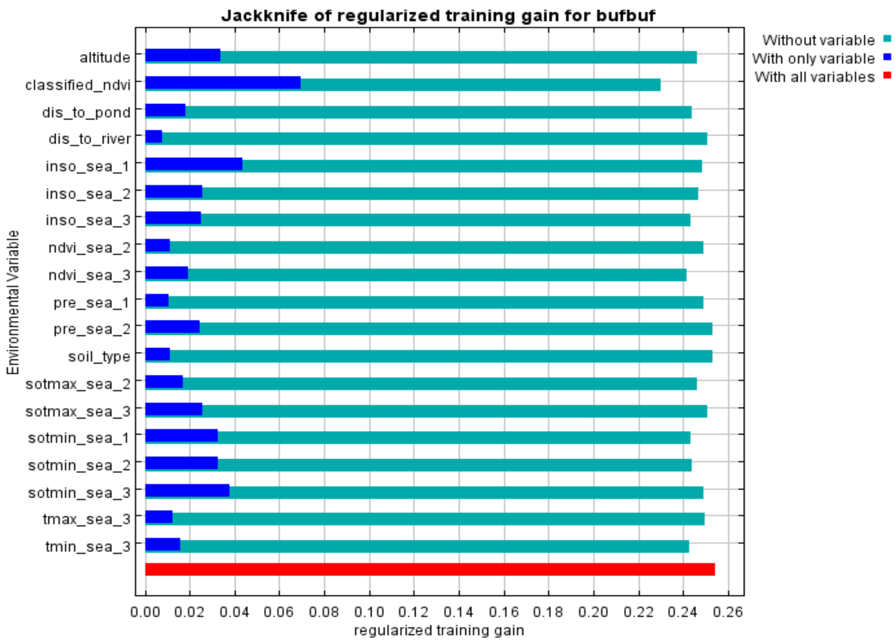
1. Bombina bombina



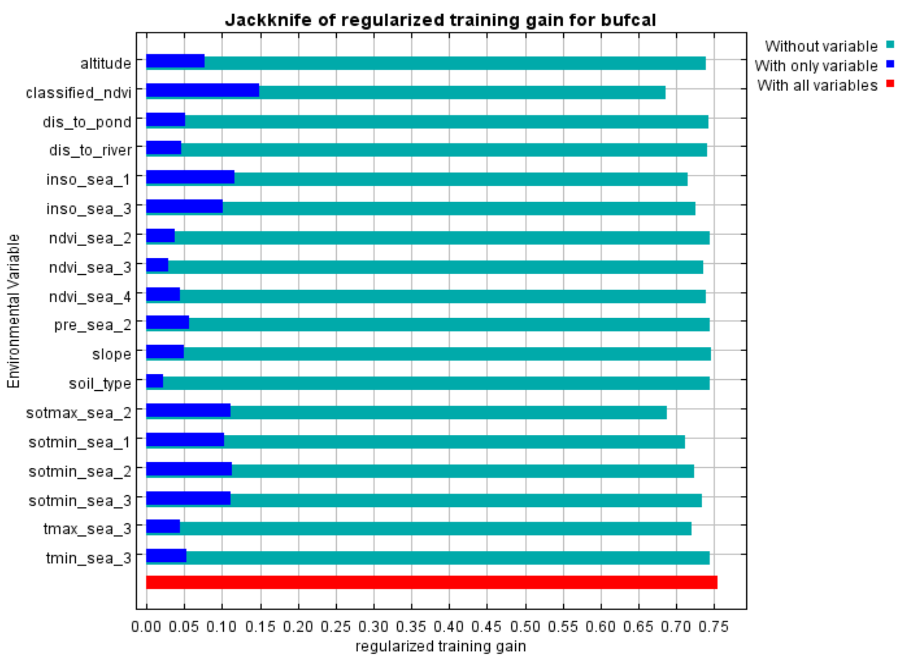
2. Bombina variegata



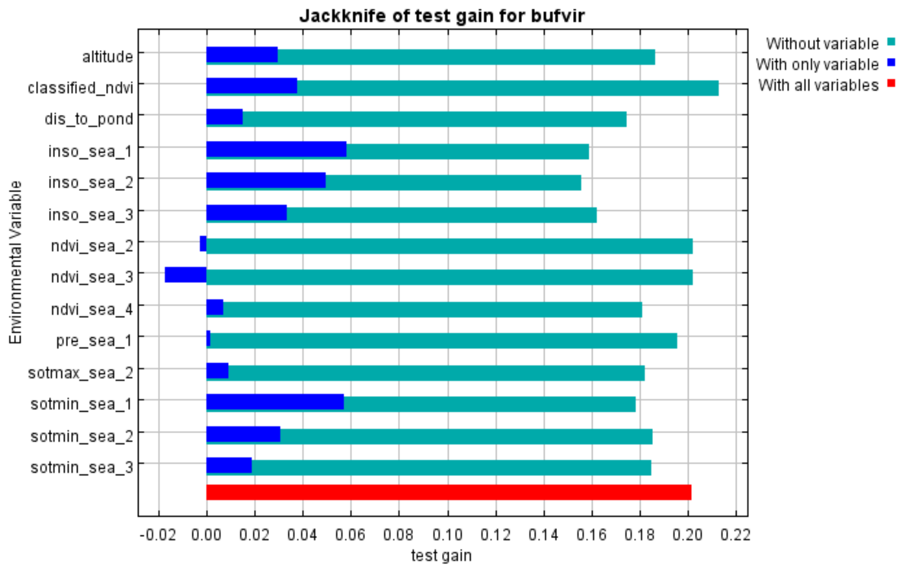
3. *Bufo bufo*



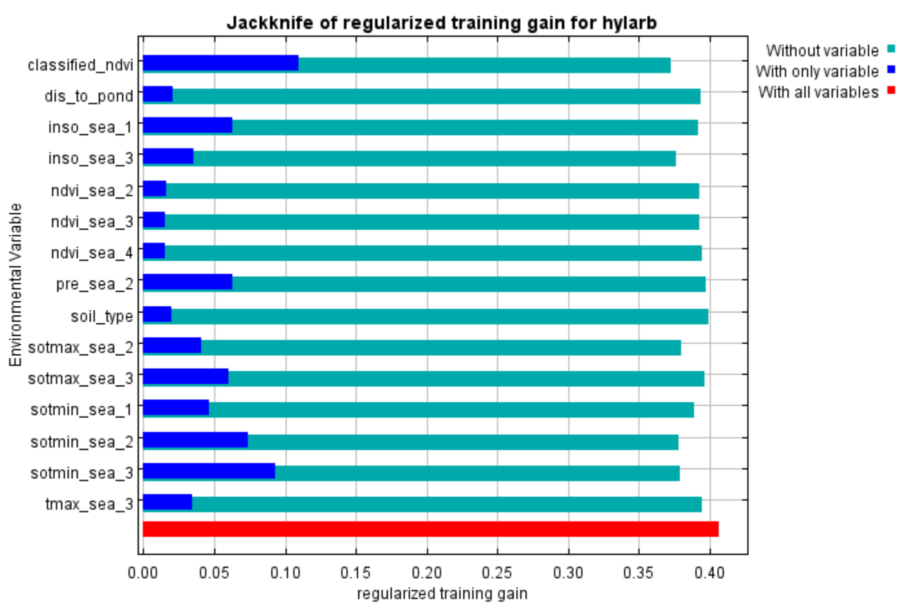
4. *Bufo calamita*



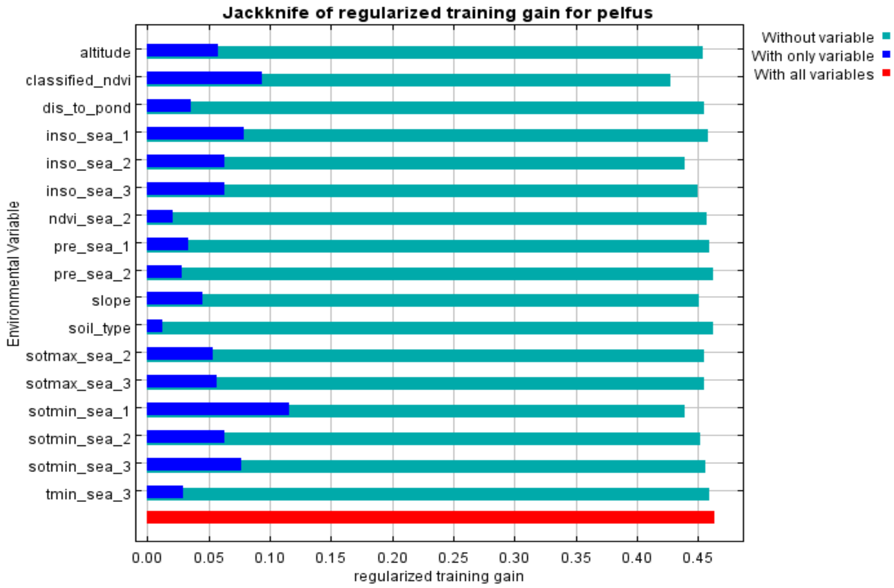
5. *Bufo viridis*



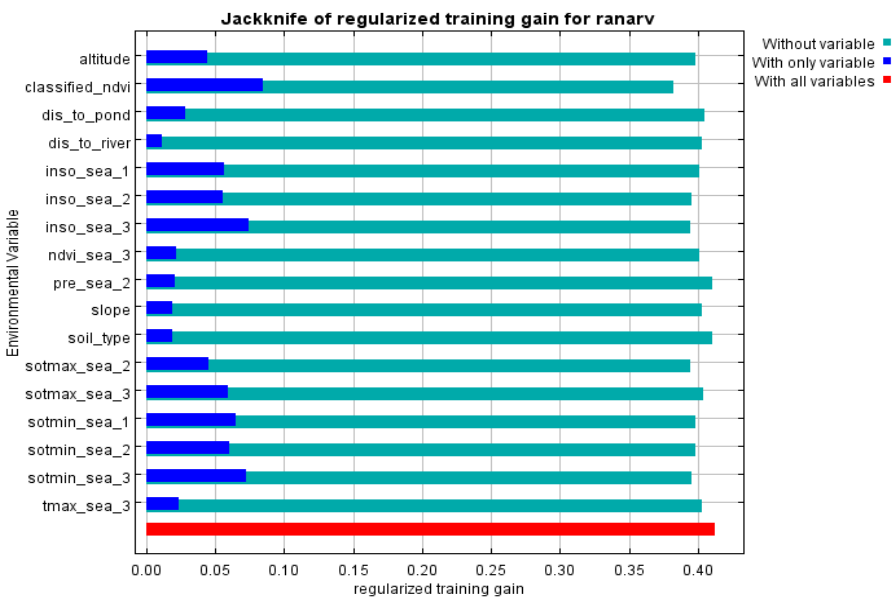
6. *Hyla arborea*



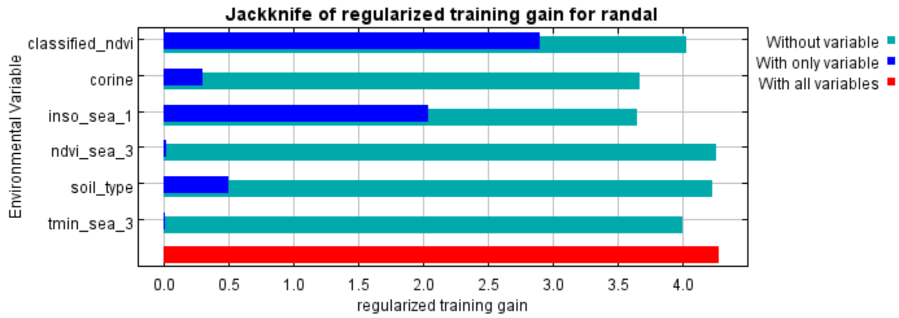
7. Pelobates Fuscus



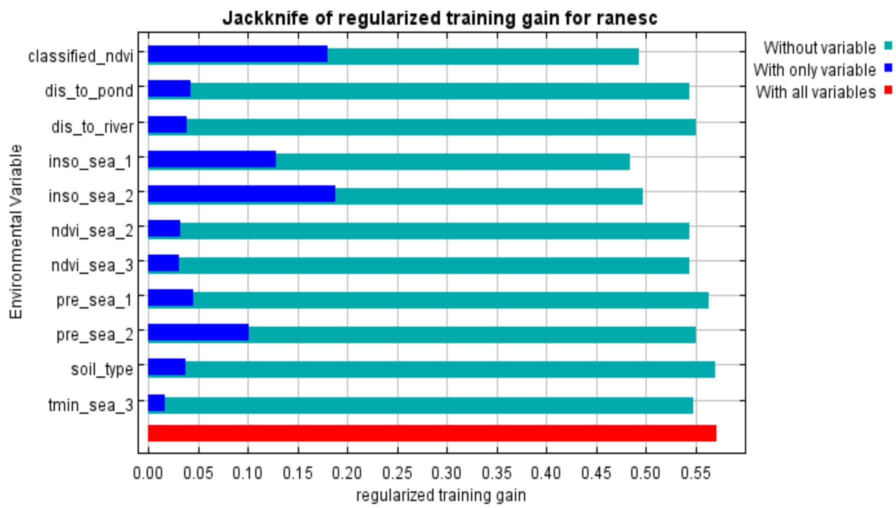
8. Rana Arvalis



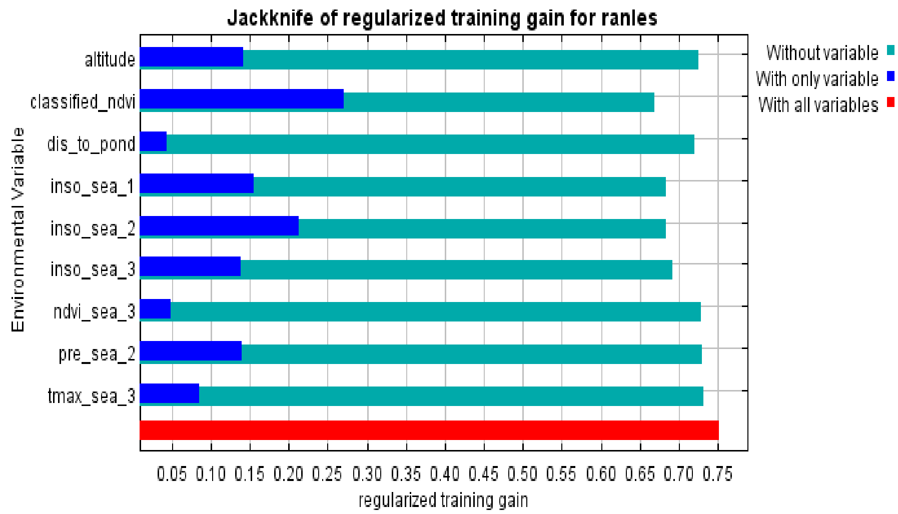
9. Rana dalmatina



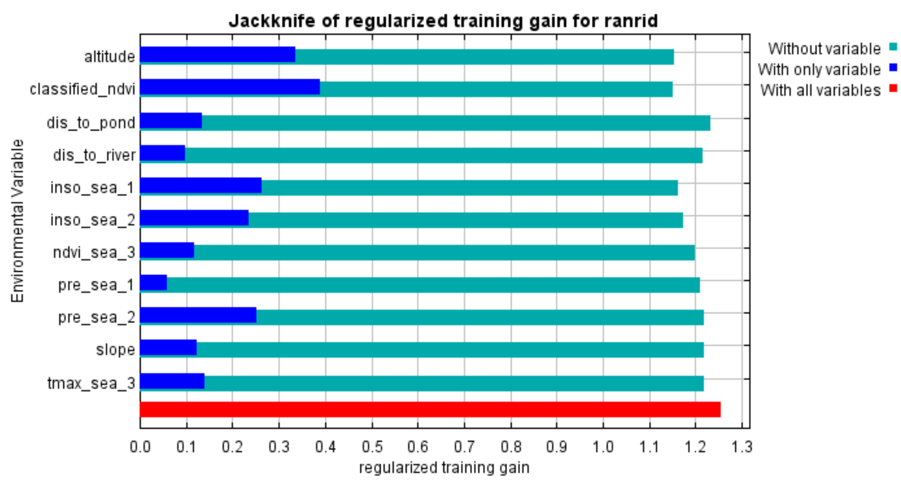
10. Rana Esculenta



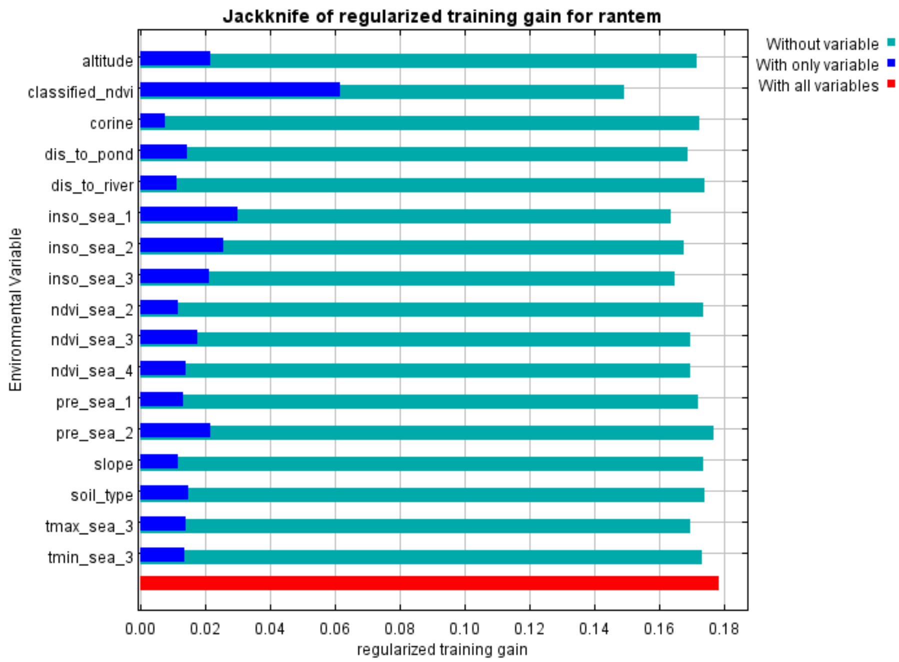
11. Rana lessonae



12. Rana ridibunda

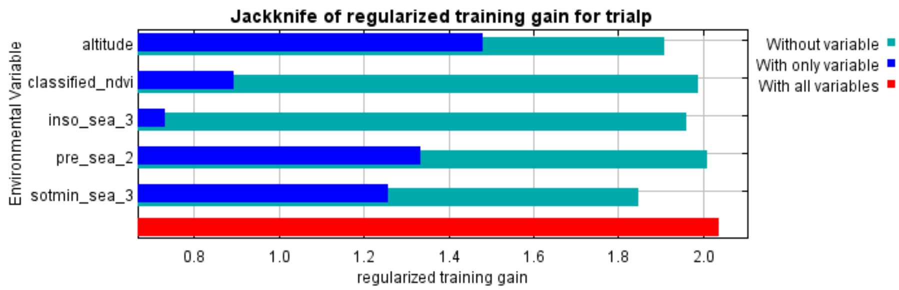


13. Rana temporaria

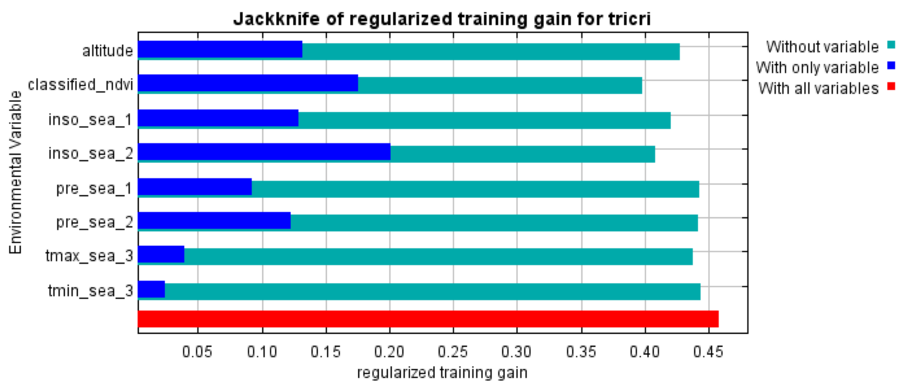


14. Salamandra salamandra

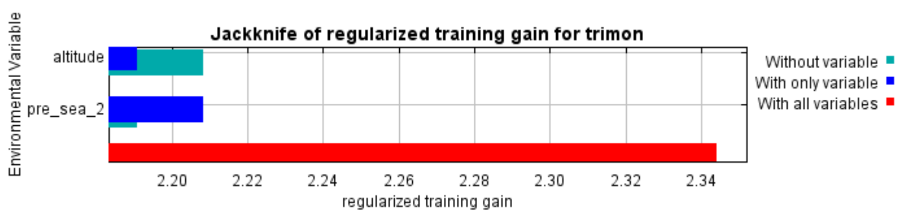
15. *Triturus alpestris*



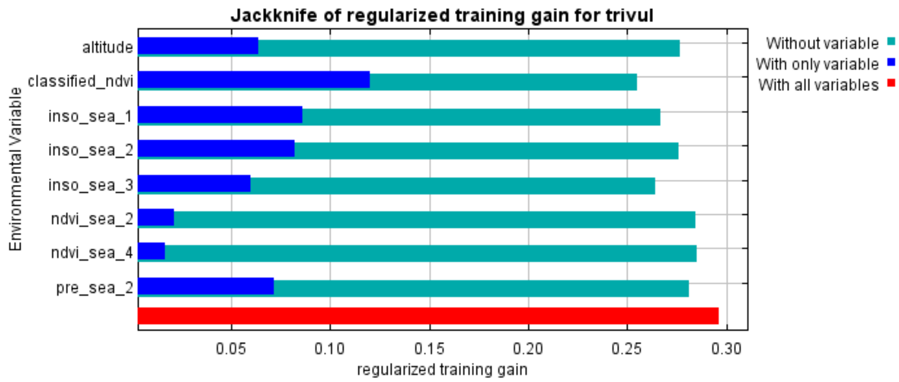
16. *Triturus cristatus*



17. *Triturus montandon*

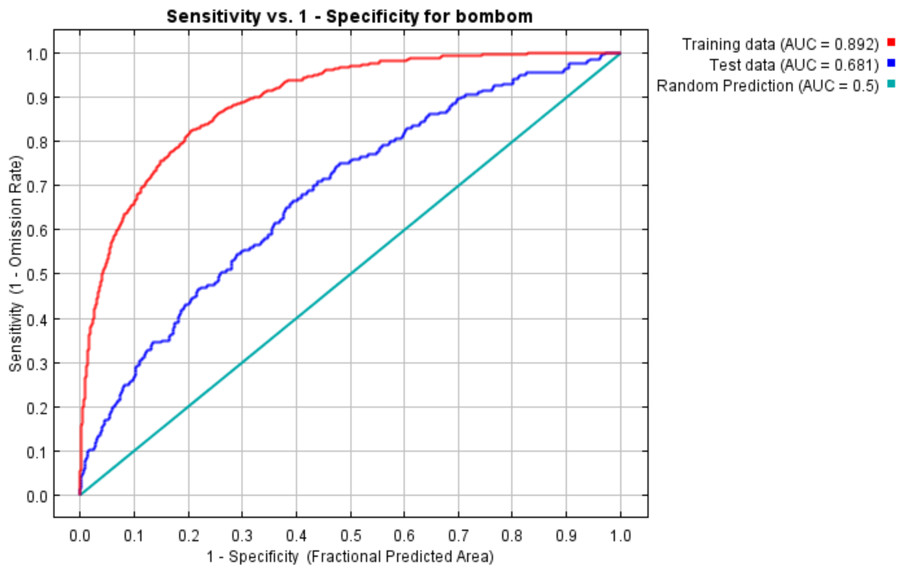


18. Triturus vulgaris

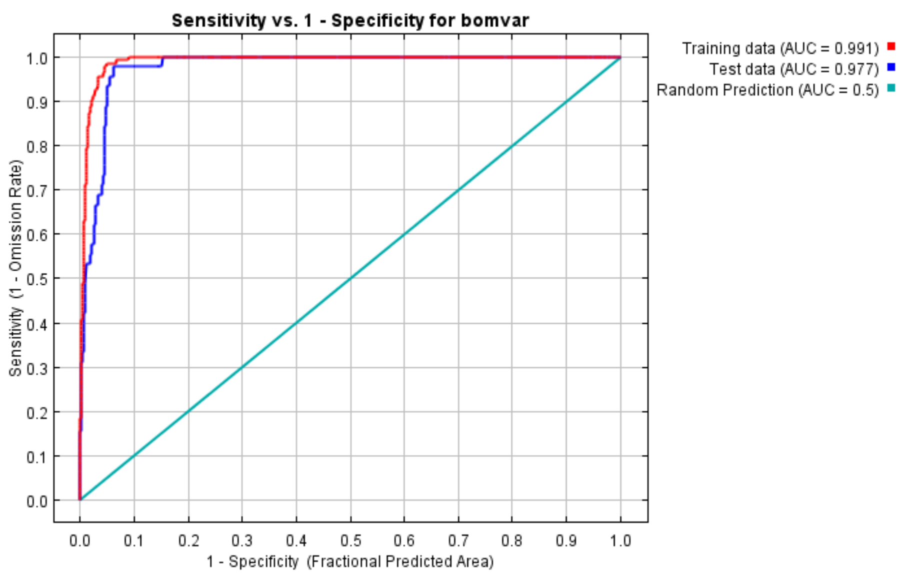


Appendix B - ROC Curves

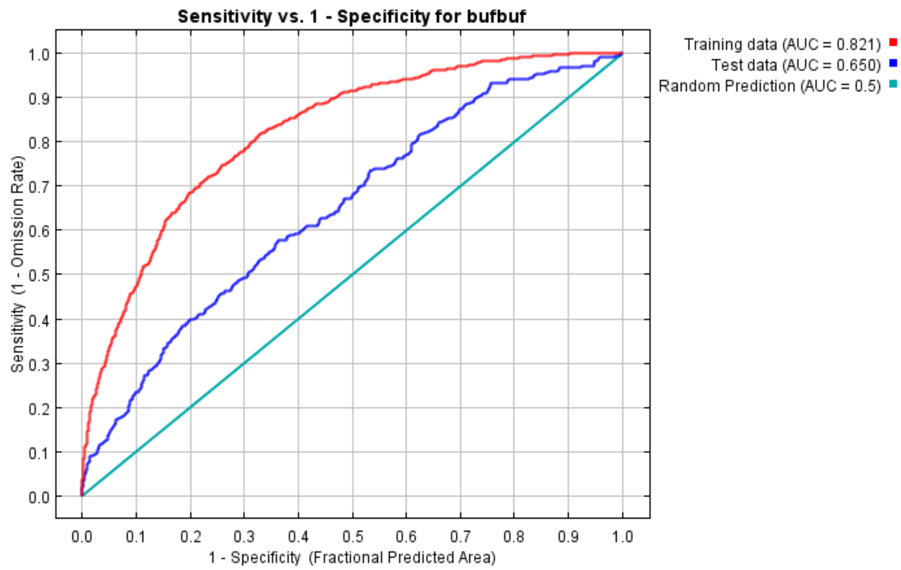
1. Bombina bombina



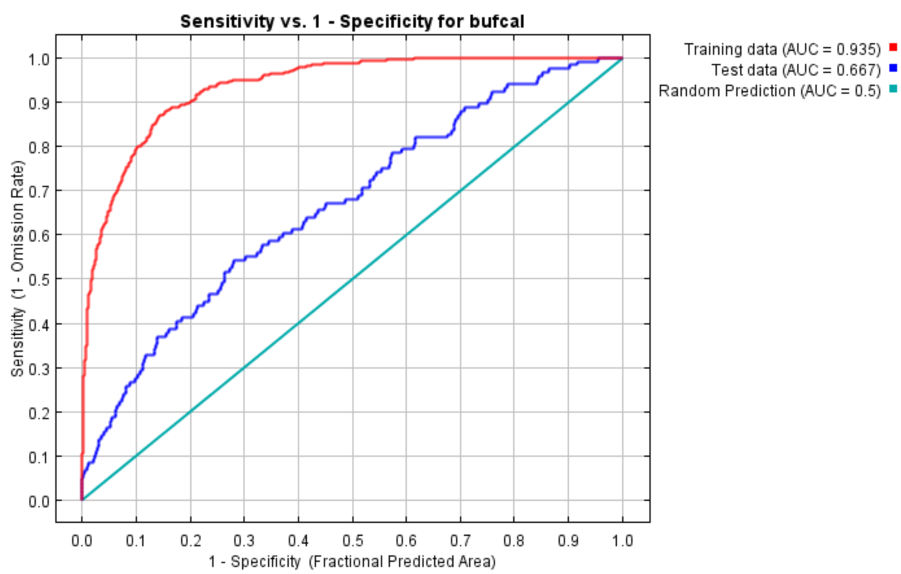
2. Bombina variegata



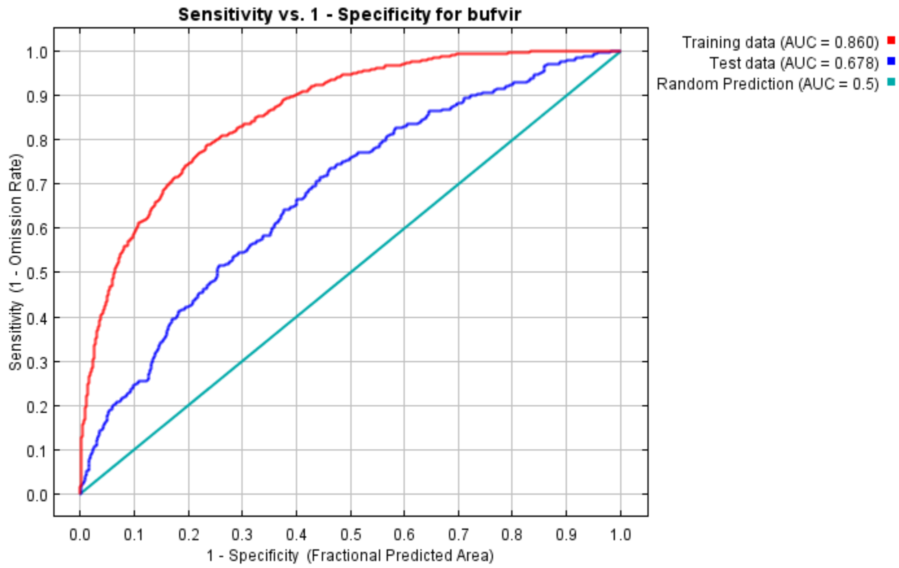
3. *Bufo bufo*



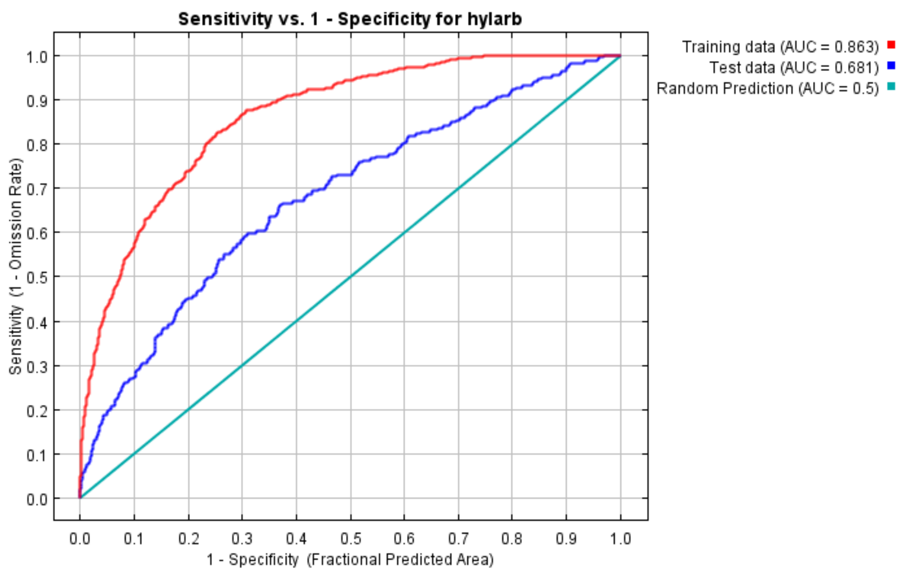
4. *Bufo calamita*



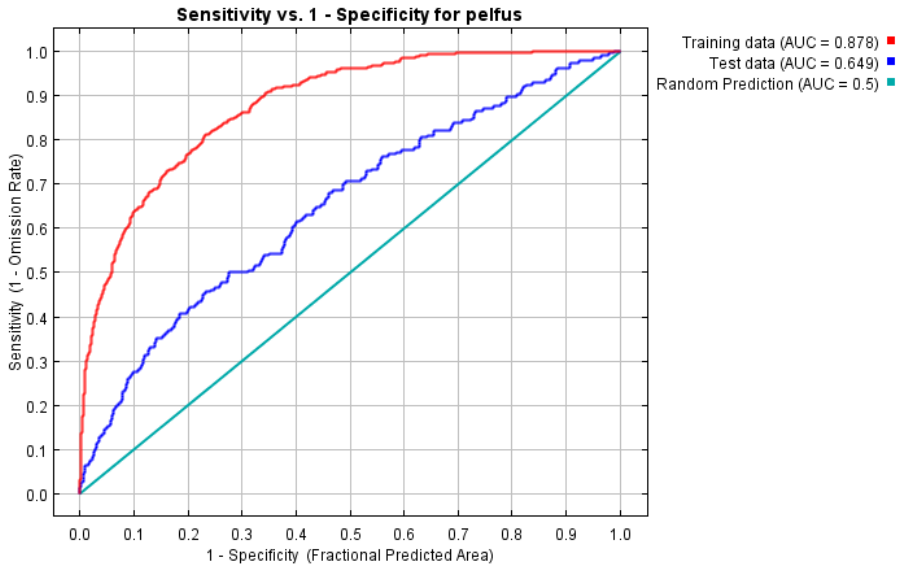
5. Bufo viridis



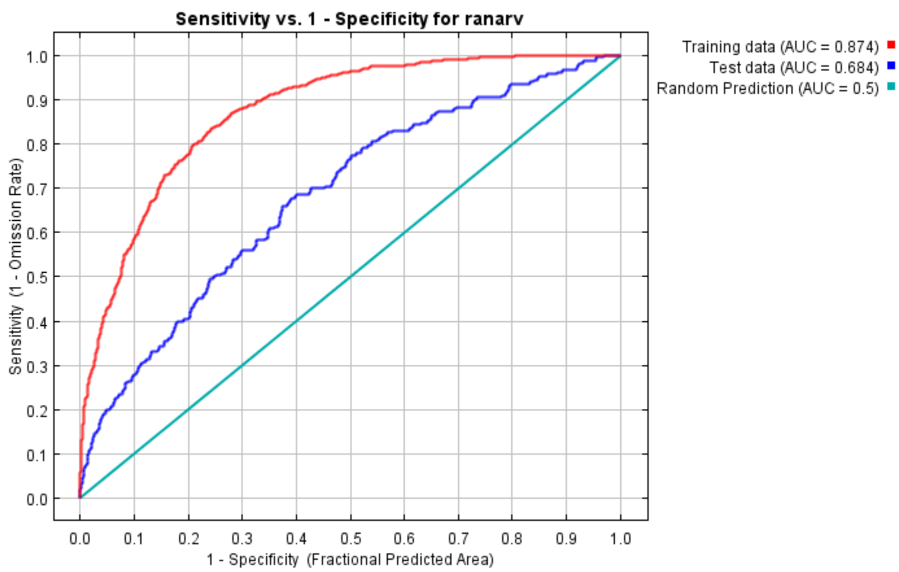
6. Hyla arborea



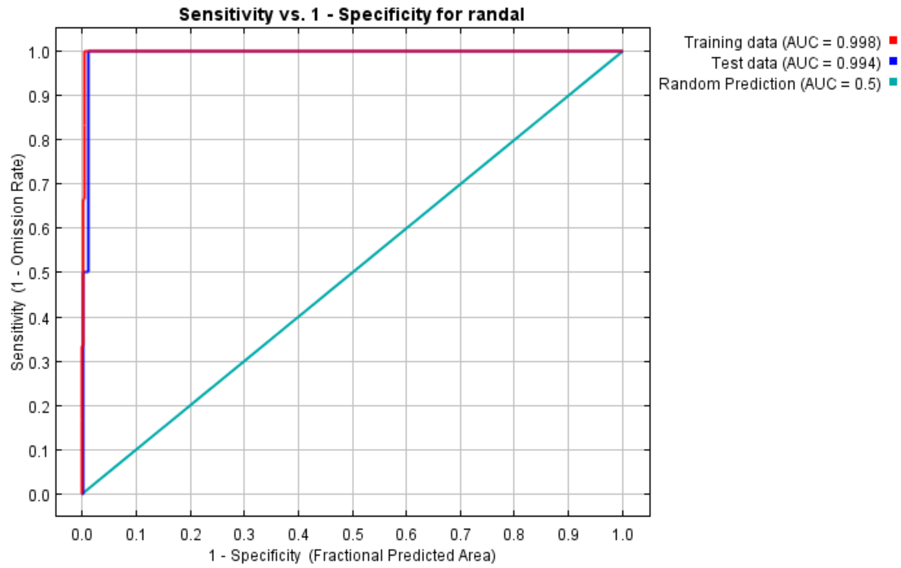
7. Pelobates Fuscus



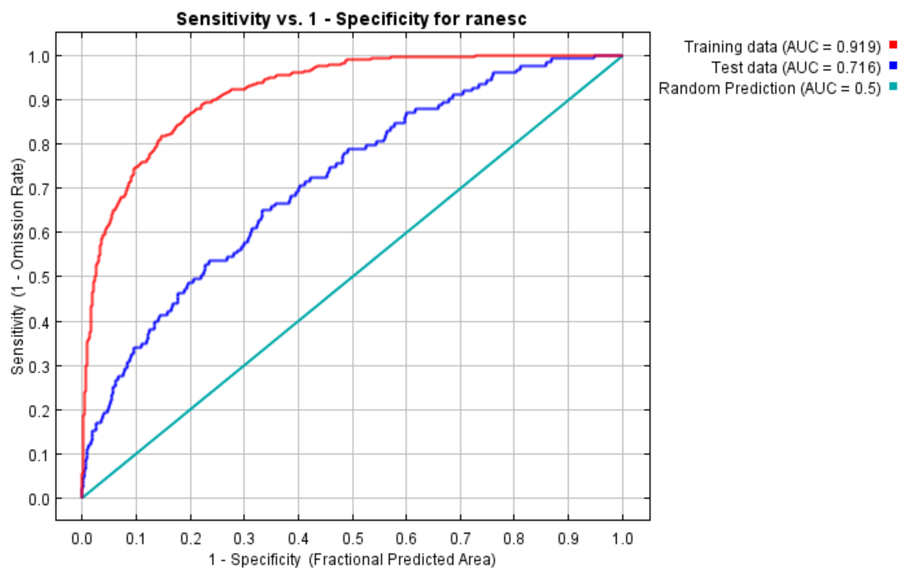
8. Rana Arvalis



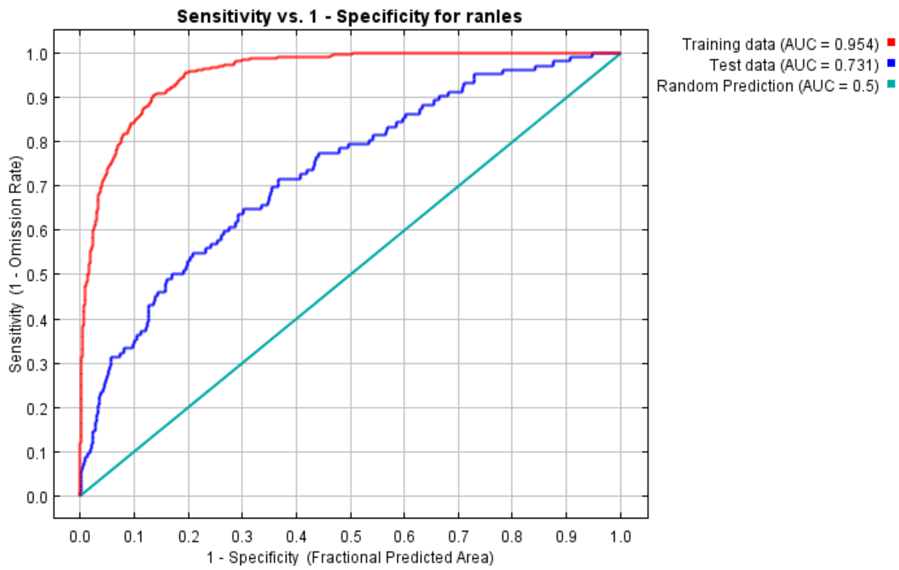
9. Rana dalmatina



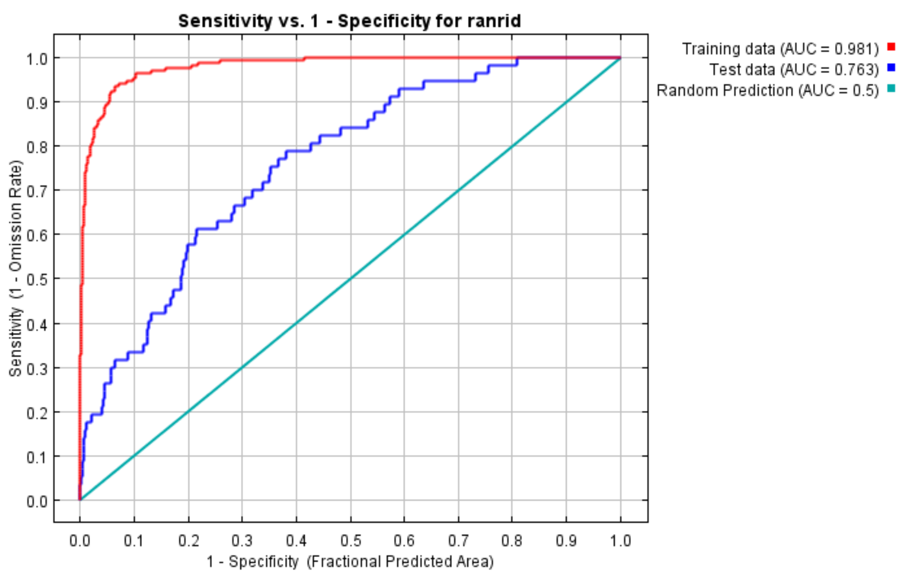
10. Rana Esculenta



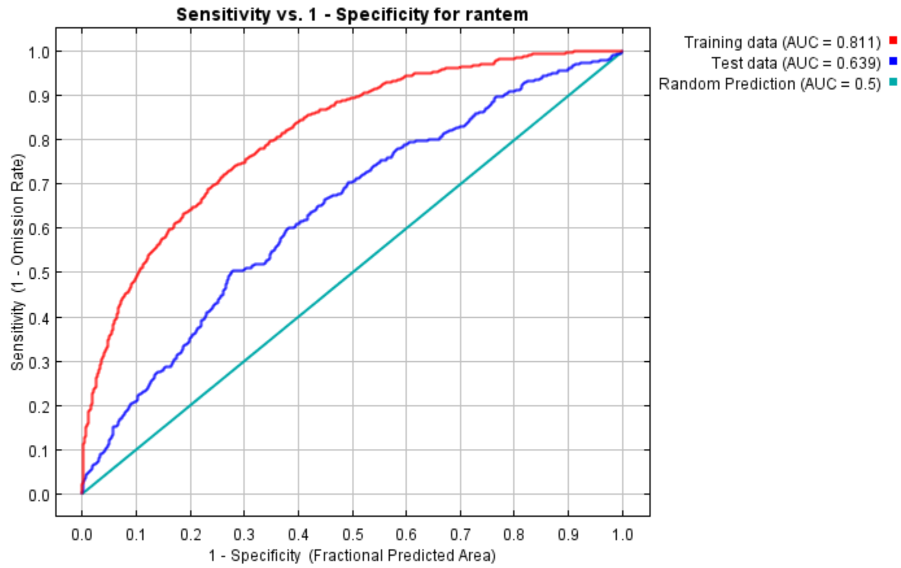
11. Rana lessonae



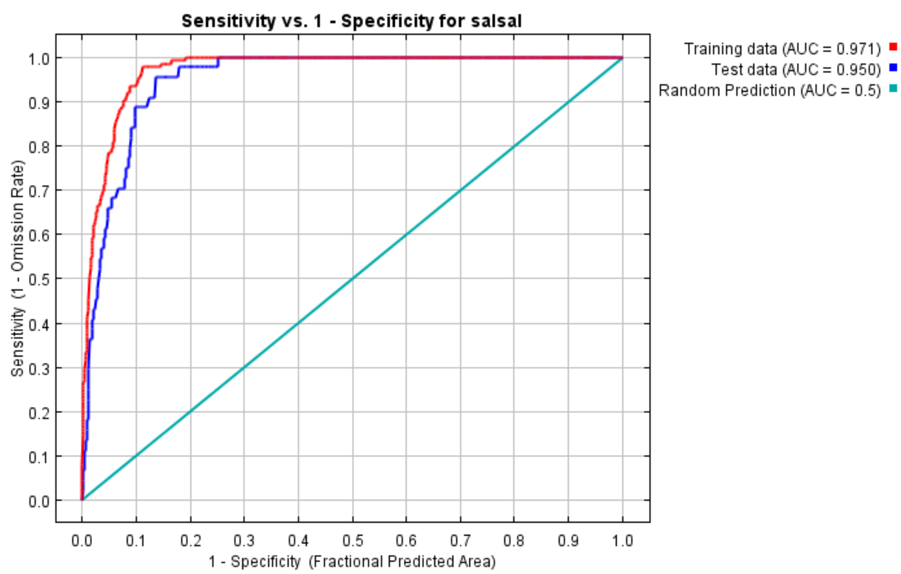
12. Rana ridibunda



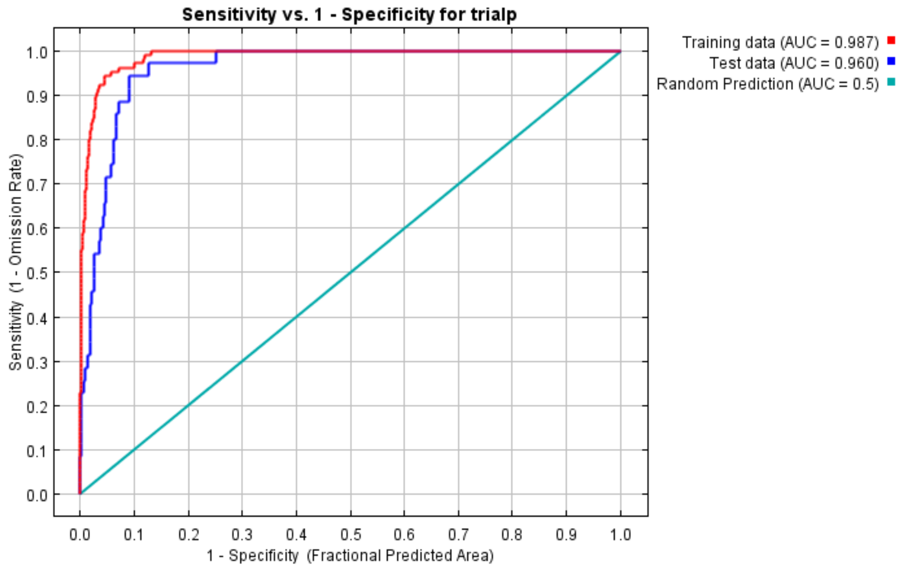
13. Rana temporaria



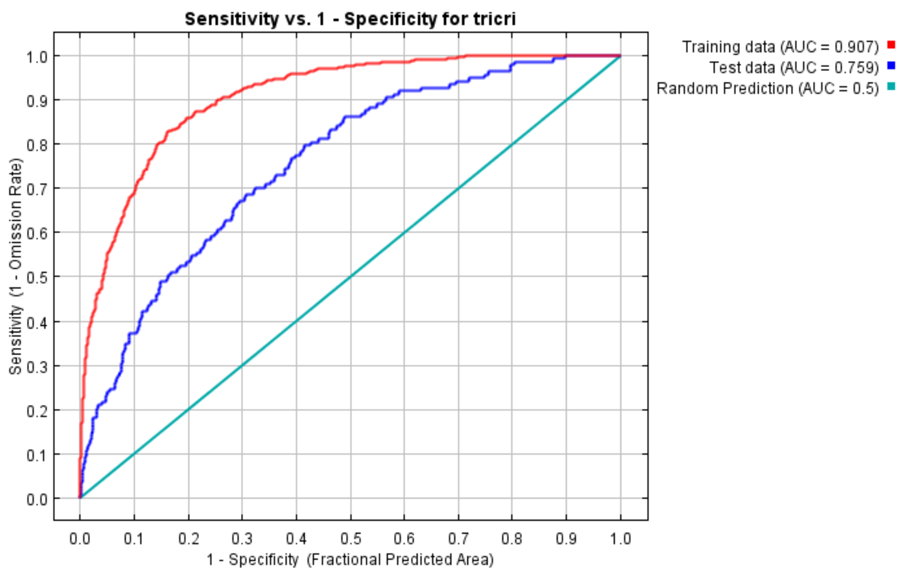
14. Salamandra salamandra



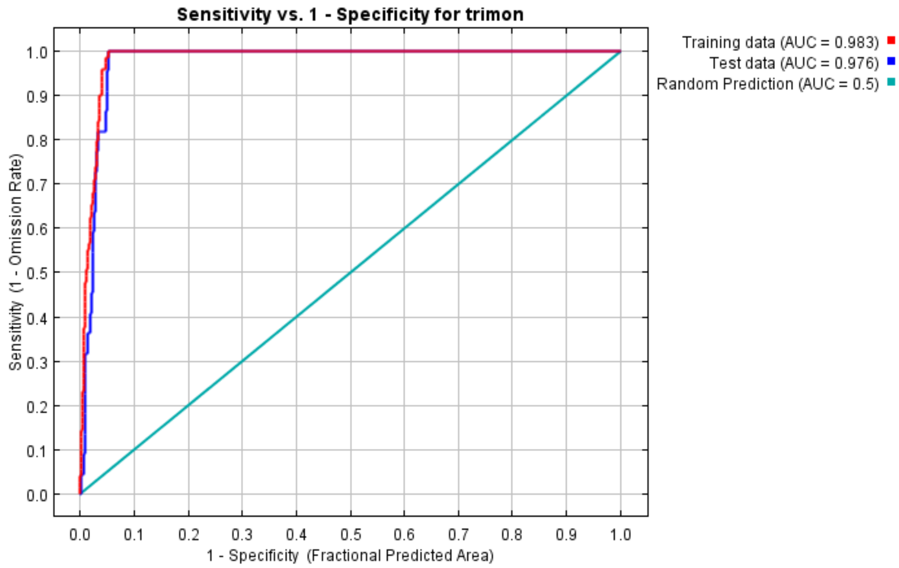
15. Triturus alpestris



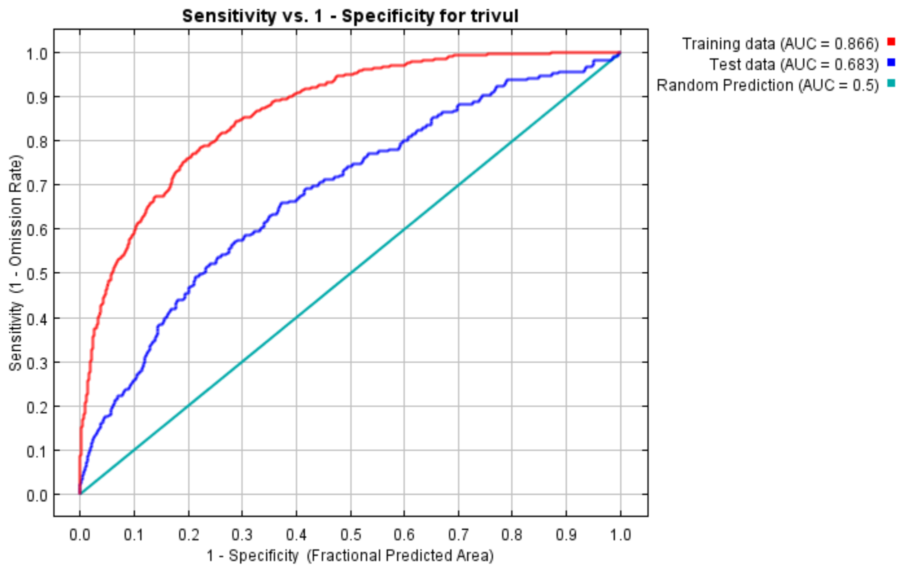
16. Triturus cristatus



17. Triturus montandon



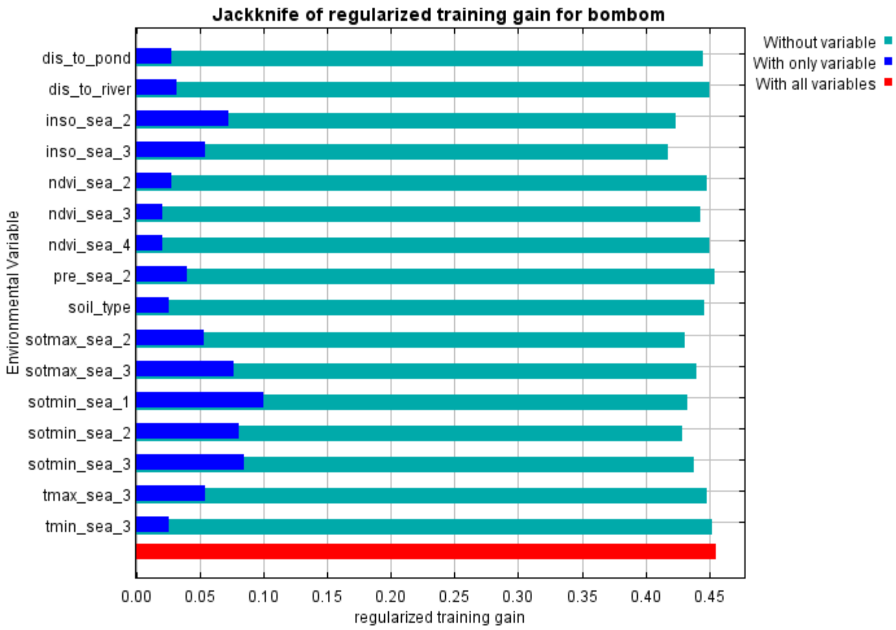
18. Triturus vulgaris



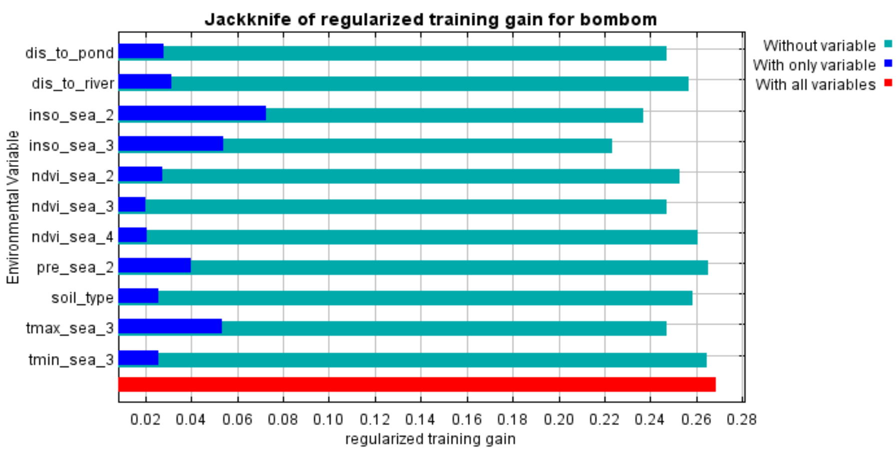
Appendix C - Comparisons of the Models with and without Soil Temperature Parameters

Take *Bombina bombina* for example.

1. Training Gain Comparison



(a) Current Model



(b) Model without Soil Temperature Variables

2. Percentage Contribution Value Comparison

(a) Current Model

Variable	Percent contribution
sotmin_sea_1	14.9
sotmin_sea_3	13.3
inso_sea_3	10.5
sotmin_sea_2	10.4
sotmax_sea_2	9.5
inso_sea_2	8.8
sotmax_sea_3	8.7
ndvi_sea_3	3.9
tmax_sea_3	3.5
dis_to_pond	3.3
ndvi_sea_2	3
pre_sea_2	2.5
ndvi_sea_4	2.4
dis_to_river	2.2
soil_type	2
tmin_sea_3	1

(b) Model without Soil Temperature Variables

Variable	Percent contribution
inso_sea_3	16.6
tmax_sea_3	16.4
inso_sea_2	14.5
ndvi_sea_2	9.5
pre_sea_2	8.9
ndvi_sea_3	8
dis_to_pond	7.9
dis_to_river	7.2
soil_type	4.4
ndvi_sea_4	4.2
tmin_sea_3	2.4