

Investigating Habitat Association of Breeding Birds Using Public Domain Satellite Imagery and Land Cover Data

Abdulhakim Mohamed Abdi

MÜNSTER, 2010



Investigating Habitat Association of Breeding Birds Using Public Domain Satellite Imagery and Land Cover Data

A Case of the Corn Bunting *Miliaria calandra* in Spain

by

Abdulkhkim Mohamed Abdi
hakim.abdi@uni-muenster.de

Thesis presented to Universität Münster in partial fulfillment of the
requirements for the degree of Master of Science in Geospatial Technologies

Münster, North Rhine-Westphalia, Germany

© Abdulkhkim Mohamed Abdi, 2010

Programme Title

Geospatial Technologies

Degree

Master of Science

Course Duration

September 2008 – March 2010

Erasmus Mundus Consortium Partners

Institut für Geoinformatik
Universität Münster, DE

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa, PT

Departamento de Lenguajes y Sistemas Informáticos
Universitat Jaume I, ES



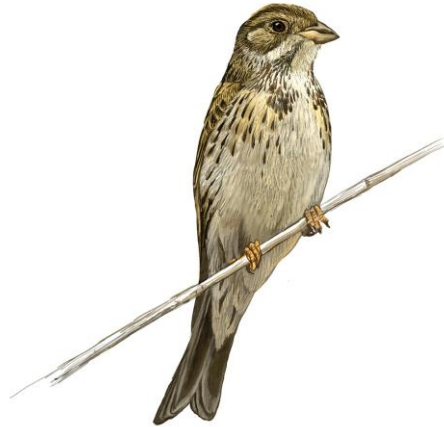
Education and Culture

Erasmus Mundus



ifgi
Institut für Geoinformatik
Universität Münster





Supervisor

Prof. Dr. Edzer Pebesma
Institute for Geoinformatics
University of Muenster

Co-supervisors

Prof. Dr. Pedro Cabral
Higher Institute of Statistics and Information Management
New University of Lisbon

Prof. Dr. Mario Caetano
Higher Institute of Statistics and Information Management
New University of Lisbon
Portuguese Geographic Institute

Prof. Dr. Filiberto Pla
Department of Computer Languages and Systems
University of Jaume I

This document describes work undertaken as part of a programme of study at Universität Münster, Universidade Nova de Lisboa and Universitat Jaume I. All views and opinions expressed therein remain the sole responsibility of the author, and not necessarily represent those of the universities.

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public. The drawing of the corn bunting in the preceding page is copyright of the Catalan breeding bird atlas.

Abstract

Twenty-five years after the implementation of the Birds Directive in 1979, Europe's farmland bird species and long-distance migrants continue to decrease at an alarming rate. Farmland supports more bird species of conservation concern than any other habitat in Europe. Therefore, it is imperative to understand farmland species' relationship with their habitats.

Bird conservation requires spatial information; this understanding not only serves as a check on the individual species' populations, but also as a measure of the overall health of the ecosystem as birds are good indicators of the state of the environment. The target species in this study is the corn bunting *Miliaria calandra*, a bird whose numbers in northern and central Europe have declined sharply since the mid-1970s.

This study utilizes public domain data, namely Landsat imagery and CORINE land cover, along with the corn bunting's presence-absence data, to create a predictive distribution map of the species based on habitat preference. Each public domain dataset was preprocessed to extract predictor variables. Predictive models were built in R using logistic regression.

Three models resulted from the regression analysis; one containing the satellite-only variables, one containing the land cover variables and a combined model containing both satellite and land cover variables. The final model was the combined model because it exhibited the highest predictive accuracy (AUC=0.846) and the least unexplained variation (RD=276.11). The results have shown that the corn bunting is strongly influenced by land surface temperature and the modified soil adjusted vegetation index. Results have also shown that the species strongly prefers non-irrigated arable land and areas containing vegetation that has high moisture content while avoiding areas with steep slopes and areas near human activity.

This study has shown that the combination of public data from different sources is a viable method in producing models that reflect species' habitat preference. The development of maps that are comprised of information from both satellites and land cover datasets are of importance for species whose habitat requirements are poorly known.

Acknowledgements

I am much obliged to all my advisors, in particular Prof. Dr. Edzer Pebesma for helping develop my research skills starting with the “Advanced Research Methods and Skills” course and his valuable comments during the thesis project.

My sincere gratitude goes to my co-advisors Prof. Dr. Pedro Cabral and Prof. Dr. Mario Caetano for their technical assistance with the final draft, for sharing their GIS and remote sensing expertise, and for their encouragement and support throughout the master’s programme.

I acknowledge the supportive, understanding and challenging environments that ISEGI and IFGI offered. In particular, I would like to thank Prof. Dr. Marco Painho, Prof. Dr. Werner Kuhn, Prof^a. Ana Cristina Costa, Dr. Christoph Brox.

I am also grateful to:

The Catalan breeding bird atlas, in particular Dr. Lluís Brotons, for providing the data, without which, this study would not have been possible.

Dr. Veronique St-Louis of Brown University for her assistance with satellite image texture analysis.

All the SCGIS, R-SIG-GEO and R-SIG-ECO list members who answered my questions and sometimes sent long detailed emails to help me with particular problems.

My partner Martina for her help with the initial draft and her love and support throughout the master’s programme.

My fellow Mundus colleagues for their professional advice, assistance and friendship.

Last but not least, no amount of words can express my gratitude for all the sacrifices my parents have made for the sake of my education. Aabo iyo Hooyo, I can’t thank you enough for all that you have done for me.

Dedication

To my parents, Mohamed Abdi Hassan and Fadumo Hussein Mohamud;

To my grandmother, Habiba Amir Omar.

List of Figures

- Figure 1: The Corn Bunting *Miliaria calandra* (Photograph by Raúl Baena Casado)
- Figure 2: Overview of the study area in Google Earth.
- Figure 3: Thesis design flowchart
- Figure 4: MSAVI vs. NDVI
- Figure 5: NDVI values compared to Land Surface Temperature and Land Surface Emissivity
- Figure 6: Topographic variables employed in the study
- Figure 7: The 27 CORINE Land Cover 2000 classes in the study area.
- Figure 8: The eight landscape metrics that were extracted from CLC2000
- Figure 9: Corn bunting presence-absence points
- Figure 10: Habitat suitability map derived from satellite data
- Figure 11: Habitat suitability map derived from land cover data
- Figure 12: Habitat suitability map derived from a combination of satellite and land cover data
- Figure 13: ROC plot of the satellite-only, CLC2000-only and combined model
- Figure 14: Comparison between map produced by the Catalan breeding bird atlas and the final map produced in this study.
- Figure 15: Distance to human activity extracted from CLC2000
- Figure 16: Distance to roads extracted from CLC2000
- Figure 17: Boxplots of the relationship between selected Landsat derivatives and CLC2000
- Figure 18: Graphical plots of the association of satellite predictors with the response
- Figure 19: Graphical plots of the association of land cover predictors with the response
- Figure 20: Graphical plots of the association of anthropogenic predictors with the response
- Figure 21: Graphical plots of the association of topographic predictors with the response
- Figure 22: Importance of each variable in the satellite and the land cover model

List of Tables

Table 1: Landsat 7 ETM+ band characteristics

Table 2: Tasseled cap transformation coefficients for Landsat ETM+ (Liang 2004)

Table 3: Mean, minimum and maximum values of predictor variables in occupied squares

Table 4: Results of the bivariate descriptive statistics

Table 5: Summary results of the logistic regression analysis for the satellite model

Table 6: Summary results of the logistic regression analysis for the land cover model

Table 7: Summary results of the logistic regression analysis for the combined model

Table 8: Variance inflation factor values for all the predictor variables

Table 9: Logistic regression output for the maximal model.

List of Acronyms

AIC	Akaike Information Criterion
AUC	Area Under the Curve of the Receiver Operating Characteristic
AVHRR	Advanced Very High Resolution Radiometer
BLF	Broad-leaved Forest
CAP	Common Agricultural Policy
CBBA	Catalan Breeding Bird Atlas
CCP	Complex Cultivation Patterns
CLC2000	CORINE Land Cover 2000
CORINE	Coordination of Information on the Environment
CV	Coefficient of Variation
DEM	Digital Elevation Model
DN	Digital Number
EEA	European Environmental Agency
EEC	European Economic Community
EPSG	European Petroleum Survey Group
ETM	Enhanced Thematic Mapper
EU	European Union
EVI	Enhanced Vegetation Index
FTBP	Fruit Trees and Berry Plantations
GWR	Geographically Weighted Regression
L1T	Level 1 Terrain Corrected
LSE	Land Surface Emissivity
LST	Land Surface Temperature
MSAVI	Modified Soil Adjusted Vegetation Index
NDVI	Normalized Difference Vegetation Index
NIAL	Non-irrigated Arable Land
NOAA	National Oceanic and Atmospheric Administration
PANV	Principally Agricultural with Natural Vegetation
PAR	Photosynthetically Active Radiation
PIL	Permanently Irrigated Land
R	The R Environment for Statistical Computing
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SRTM	Shuttle Radar Topography Mission
SVEG	Sclerophyllous Vegetation
TCT	Tasseled Cap Transformation
TM	Thematic Mapper
TWS	Transitional Woodland-shrub
UK	United Kingdom
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
VIF	Variance Inflation Factor

Table of Contents

AUTHOR'S DECLARATION.....	i
Abstract.....	ii
Acknowledgements.....	iii
Dedication.....	iv
List of Figures.....	v
List of Tables.....	vi
List of Acronyms.....	vii
Chapter 1 Introduction.....	1
1.1 Background and significance.....	1
1.1.1 The decline of farmland breeding birds in Europe.....	1
1.1.2 The Corn Bunting <i>Miliaria (Emberiza) calandra</i>	3
1.2 Species distribution modeling.....	4
1.3 Statement of problem.....	6
1.4 Study area.....	7
1.5 Research objectives.....	7
1.6 Research questions.....	8
1.7 Thesis organization.....	8
Chapter 2 Data.....	10
2.1 Satellite imagery.....	10
2.2 Satellite imagery preprocessing.....	11
2.2.1 Texture analysis.....	11
2.2.2 Calculation of vegetation indices.....	11
2.2.3 Tasseled cap transformation.....	13
2.2.4 Land surface temperature.....	14
2.2.5 Topographic variables.....	17
2.3 Land cover data.....	17
2.4 Land cover data preprocessing.....	18
2.4.1 Anthropogenic variables.....	18
2.4.2 Landscape metrics.....	19

2.5 Catalan breeding bird atlas	21
2.6 Analysis tools.....	22
Chapter 3 Methodology	23
3.1 Bivariate descriptive statistics	23
3.2 Multiple logistic regression	24
3.3 Multicollinearity diagnosis	25
3.4 Variable selection	26
3.5 Assessing goodness of fit and model validation	26
3.6 Model evaluation and selection	28
Chapter 4 Results	30
4.1 Overlay analysis	30
4.2 Bivariate descriptive statistics	31
4.3 Satellite model	32
4.4 Land cover model	34
4.5 Combined model.....	36
4.6 Model selection.....	38
Chapter 5 Discussion	39
5.1 Satellite imagery	39
5.2 Land cover dataset	40
5.2.1 Non-irrigated arable land	40
5.2.2 Permanently irrigated land.....	40
5.3 Final model	41
5.4 Variable selection	42
5.5 Comparison to the CBBA map	42
Chapter 6 Conclusions and Recommendations.....	44
References.....	47
Appendix A : Anthropogenic variables	54
Appendix B : Descriptive statistics.....	56
Appendix C : Statistical analysis.....	61
Appendix D : R Code.....	65

Chapter 1 Introduction

1.1 Background and significance

In 1979, the European Economic Community passed the *Council Directive 79/409/EEC on the conservation of wild birds*, otherwise known as the Birds Directive, which aims “at providing long-term protection and conservation of all bird species naturally living in the wild within the European territory of the Member States” (EEC 1979). Among other things, the directive seeks the protection and management of wild birds through the creation of protected areas and habitat maintenance. However, 25 years after the implementation of the Birds Directive, farmland bird species and long-distance migrants continue to decrease at an alarming rate (Birdlife-International 2004). This has been credited to detrimental land use policies such as the Common Agricultural Policy (Birdlife-International 2004) that have promoted the intensification of farmlands through crop specialization (monocultures), pesticide use and the eradication of uncultivated areas in order to maximize productivity (Donald et al. 2001).

Farming evolved over the last 10,000 years and spread across the forested European landscape up to the point that over 50% of the European continent is used for farming. Several organisms have adapted to this new landscape and are now open-country specialists that use farmland as their primary habitat (Donald et al. 2002). Gradually, agricultural landscape began to support a large amount of biological diversity and eventually became its own ecosystem, sustained by humans through traditional farming systems that employ low-input techniques such as lack of irrigation, large fallow areas and relatively low potential yields. This habitat supports more bird species of conservation concern than any other habitat in Europe (Stoate et al. 2003).

1.1.1 The decline of farmland breeding birds in Europe

The decline in farmland bird species first became obvious in the 1980s, which was the decade that displayed a steady rise in EU agricultural output and the adoption of intensive agriculture to maximize yield (Siriwardena et al. 1998; Donald et al. 2001). The use

of intensive agricultural practices is characterized by a high amount of heavy soils and extensive irrigation of the landscape which doubles the yield of certain crops (Stoate et al. 2000). In contrast, areas that are extensively farmed are high in biodiversity (Tucker and Heath 1994; Benton et al. 2003) and are characterized by thin soils, no irrigation, high fallow areas and low yields (Stoate et al. 2000).

Some species have been extinct as breeders in certain countries; for example the Red-backed Shrike *Lanius collurio* in the UK and the Roller *Coracias garrulus* in the Czech Republic (Tucker and Heath 1994). While others, such as the Corncrake *Crex crex* in France (Deceuninck 1998), have been identified as endangered. The declines in breeding birds not only have implications for Europe but also contribute to declines in biodiversity for Africa and Asia as those continents host many migratory European species during winter months.

Birds are good indicators of the state of the environment because they are highly mobile, well-studied, easily monitored and occupy a range of habitats (Tankersley 2004; Gregory et al. 2005). Due to their high mobility, birds can also respond quickly to changes in landscape and local vegetation (Coreau and Martin 2007; Vallecillo et al. 2008). Accordingly, it is critical to understand species' relationship with their habitats to determine which areas are more favorable than others. Bird conservation requires spatial information; this understanding not only serves as a check on the individual species' populations but also as a measure of the overall health of the ecosystem.

The term "bird atlas" has appeared in ornithological vocabulary to mean aggregated distribution maps based on rectangular presence/absence grids produced from field surveys; currently many countries have their own breeding bird atlases. However, such projects may take several years to complete and there often is a large temporal lag between two inventories because it is a time and effort consuming process that is often limited to small spatial extents (Norris and Pain 2002; St-Louis et al. 2006). Therefore, there is a need for a rapid and more effective process to map species distributions that is relatively affordable, accurate, and that can be applied frequently.

1.1.2 The Corn Bunting *Miliaria (Emberiza) calandra*.

The target species in this study is the corn bunting *Miliaria calandra* (Figure 1), which is a bird of low-intensity arable landscapes (Taylor and O'Halloran 2002). The northern and central European corn bunting population has declined sharply since the mid-1970s (Tucker and Heath 1994) particularly in Britain (Brickle et al. 2000), Poland (Orlowski 2005) and Ireland (Taylor and O'Halloran 2002) while southern European breeding densities, particularly in Spain, Portugal and Turkey are stable (Diaz and Telleria 1997). Declines in northern corn bunting populations have been attributed to the process of farmland intensification (Brickle et al. 2000; Donald et al. 2001) mentioned in the preceding section. However, relatively few studies have been conducted on the habitat requirements and breeding density of corn buntings in southern Europe (Brambilla et al. 2009).



Figure 1: The Corn Bunting *Miliaria calandra*. (Photograph by Raúl Baena Casado)

1.2 Species distribution modeling

Birds, like all mobile organisms, have favorite habitats in which to breed, spend winter months and refuel while on migration. In order to effectively conserve a species it is vital to know these habitats and their spatial dimensions. Several studies have utilized geospatial technologies in bird distribution research. However in this thesis only indirect methods of mapping species will be discussed. Indirect methods involve the use of land cover mapping and other remote sensing techniques based on habitat requirements to predict the distribution of species (Nagendra 2001). The advantages of using satellite imagery include large areal coverage and fine spatial and temporal resolutions (Griffiths et al. 2000) while national land cover datasets have proven to link birds to habitat classes and vice versa (Fuller et al. 2005).

St-Louis et al (2006) used linear regression models to evaluate the correlation between high-resolution satellite image texture and bird point count data, the results have shown that different methods described 57% to 76% of variability in species richness. A similar study by Bellis et al (2008) assessed the relationship between greater rhea *Rhea americana* group size against normalized difference vegetation index (NDVI) and texture measures from Landsat Thematic Mapper (TM) imagery. Their results had shown that “rhea group size was most strongly positively correlated with texture variables derived from near infrared reflectance measurement”. The use of outputs resulting from the characterization and identification of upland vegetation using satellite imagery in bird abundance–habitat models was performed by Buchanan et al (2005). Their results showed that bird abundances forecasted using Landsat Enhanced Thematic Mapper (ETM) derived vegetation data was similar to that acquired when field-collected data were used for one bird species.

A study on the use of unclassified satellite imagery in the study of habitat selection of three bird species was undertaken by Erickson et al (2004) in a method that uses Landsat TM spectral values. Foody (2005) applied geographically weighted regression (GWR) on NDVI and temperature variables derived from Advanced Very High Resolution Radiometer (AVHRR) of the National Oceanic and Atmospheric Administration (NOAA) and his research

indicated the ability to characterize aspects of biodiversity from coarse spatial resolution remote sensing data and highlight the need to accommodate for the effects of spatial non-stationarity in the relationship. Wallin et al (1992) monitored potential breeding habitat for the red-billed quelea *Quelea quelea* using NDVI calculations derived from AVHRR.

A combination of land cover maps derived from Landsat ETM imagery, digital elevation models (DEM) were utilized by Hale (2006) to model the distribution and abundance of Bicknell's thrush *Catharus bicknelli* that resulted in spatially explicit predictions of probability of species' presence. Habitat selection criteria for the loggerhead shrike *Lanius ludovicianus* were derived from one province and applied to Landsat TM imagery covering another province by Jobin et al (2005) in order to evaluate the availability of suitable breeding habitats. Laurent et al (2005) investigated the potential of using unclassified spectral data in the predicting the distribution of three bird species using Landsat ETM imagery and point count data.

The effectiveness of combining Landsat TM satellite imagery, topographic data and a Geographic Information System (GIS) in bird species richness modeling was investigated by Luoto et al (2004) where they concluded that a spatial grid system containing different environmental variables derived from remote sensing data creates consistent datasets that can be used when predicting species richness. Nohr and Jorgensen (1997) concluded that "there is a positive correlation between avian parameters and satellite image features with the highest value obtained when correlating avian data with combined data from Landsat TM images on landscape diversity and integrated NDVI (INDVI) derived from AVHRR imagery".

Knowledge of the range and distribution of species at risk of extinction is crucial. In Senapathi et al (2007) the loss of habitat that the critically endangered Jerdon's courser *Rhinoptilus bitorquatus* suffered from 1991 to 2000 was quantified using classified Landsat TM and ETM imagery. Their results have shown that the species' breeding habitat has been decreasing at an annual rate of 1.2-1.7%.

Apart from unclassified satellite imagery, habitat variables can be extracted from land cover maps and be used to predict the distribution of species. Seoane et al (2004b)

compared the capacity of two general land cover maps and “two more accurate structural vegetation maps” in forecasting the distribution of bird species.

A review of studies in bird-habitat relationships using satellite imagery in the last thirty years was presented in Gottschalk et al (2005), where 120 publications were examined. A noteworthy conclusion of the review was that the potential of using the geospatial tools of remote sensing and GIS “might exist in their application in limited access ecosystems and where coarse and quick but quantitative estimates with statistical confidence limits on biodiversity are needed to achieve wildlife conservation and management objectives”.

1.3 Statement of problem

Conservation work is sometimes done by non-profit organizations that cannot afford expensive methodologies with their limited resources. On the other hand, national agencies and environmental lobby groups might find themselves in situations that require the rapid production of results to decision makers. Since Coordination of Information on the Environment (CORINE) land cover and Landsat datasets are free and bird distribution and habitat suitability models can be derived from them relatively quickly, they present themselves as important conservation tools.

The problem is that the potential of public domain data is not fully exploited. Public data is under-used because of its coarse output compared to more detailed, and more expensive, data.

Although as Gottschalk (2005) demonstrated, there is no lack in research that deals with the use of geospatial tools in the prediction of species’ distribution, there is, however, a lack in comparative research that assesses land cover data and satellite imagery in habitat modeling. There is additionally a need to evaluate the viability and accuracy of distribution maps from public sources because of their potential as primary sources of environmental and physical data for habitat modeling and conservation studies.

1.4 Study area

The study area is located in the province of Lerida in the western part of the Autonomous Community of Cataluña, Spain. The area is covered by low-intensive cereal crops and small remains of the original dry-shrub vegetation. The study area covers approximately 1,514 square kilometers and is a stepic landscape comprised of non-irrigated cropland and dry forests with land use devoted to extensive agriculture and dry pastureland (Sundseth and Sylwester 2009). The study area also encompasses part of the Lerida plain, which is an area of steppes and pseudo-steppes on the eastern edge of the river Ebro basin (Ponjoan et al. 2008).

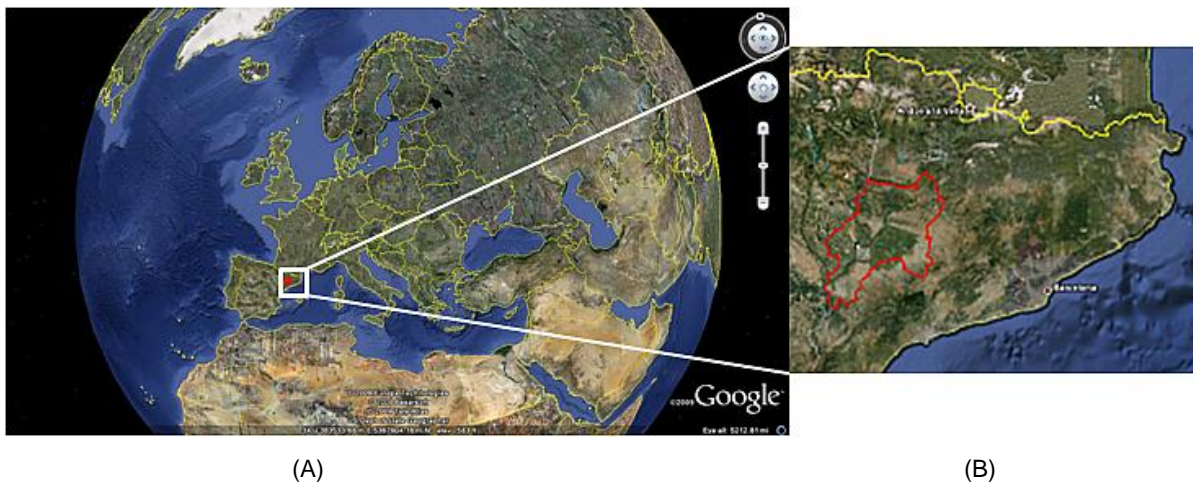


Figure 2: Overview of the study area in Google Earth. (A): The white box shows location of the study area within Europe. (B): The red outline shows location of the study area within Cataluña.

1.5 Research objectives

Notwithstanding previous research, there is an inadequate amount of information on the relationship between the occurrence of farmland bird species and predictor variables extracted from a combination of public domain sources. This study aims to develop a model of the probability of occurrence of the corn bunting based on habitat preference.

The specific objectives of the study are:

1. Assess the predictive power of variables derived from public domain satellite imagery and general-purpose land cover data in modeling the distribution of the corn bunting based on habitat preference.
2. Compare predictions produced by satellite imagery against predictions obtained from the land cover data.
3. Examine the potential of combining both public data sources in the modeling process.

1.6 Research questions

1. Can the distribution of the corn bunting be predicted by solely using data derived from public domain satellite imagery?
2. Can the distribution of the corn bunting be predicted by solely using data derived from a general-purpose land cover dataset?
3. What is the relative performance of the model resulting from data based on land cover data against the model resulting from data based on satellite imagery?
4. How does a combined model perform against the individual land cover and satellite models?
5. Which approach would be more effective in predicting the distribution of the corn bunting?

1.7 Thesis organization

The design of the thesis encompasses twelve steps that were employed in order to answer the research questions and fulfill the research objectives. The steps were divided into three general categories: acquisition, GIS & remote sensing analysis and statistical analysis. The first category involves the acquisition of the response and the explanatory predictor variables. The second category involves the computation and extraction of the predictor raster images using GIS and remote sensing methods. The third category involves the use of the R in a series of statistical analyses that culminates in the creation of a habitat suitability map in a GIS environment.

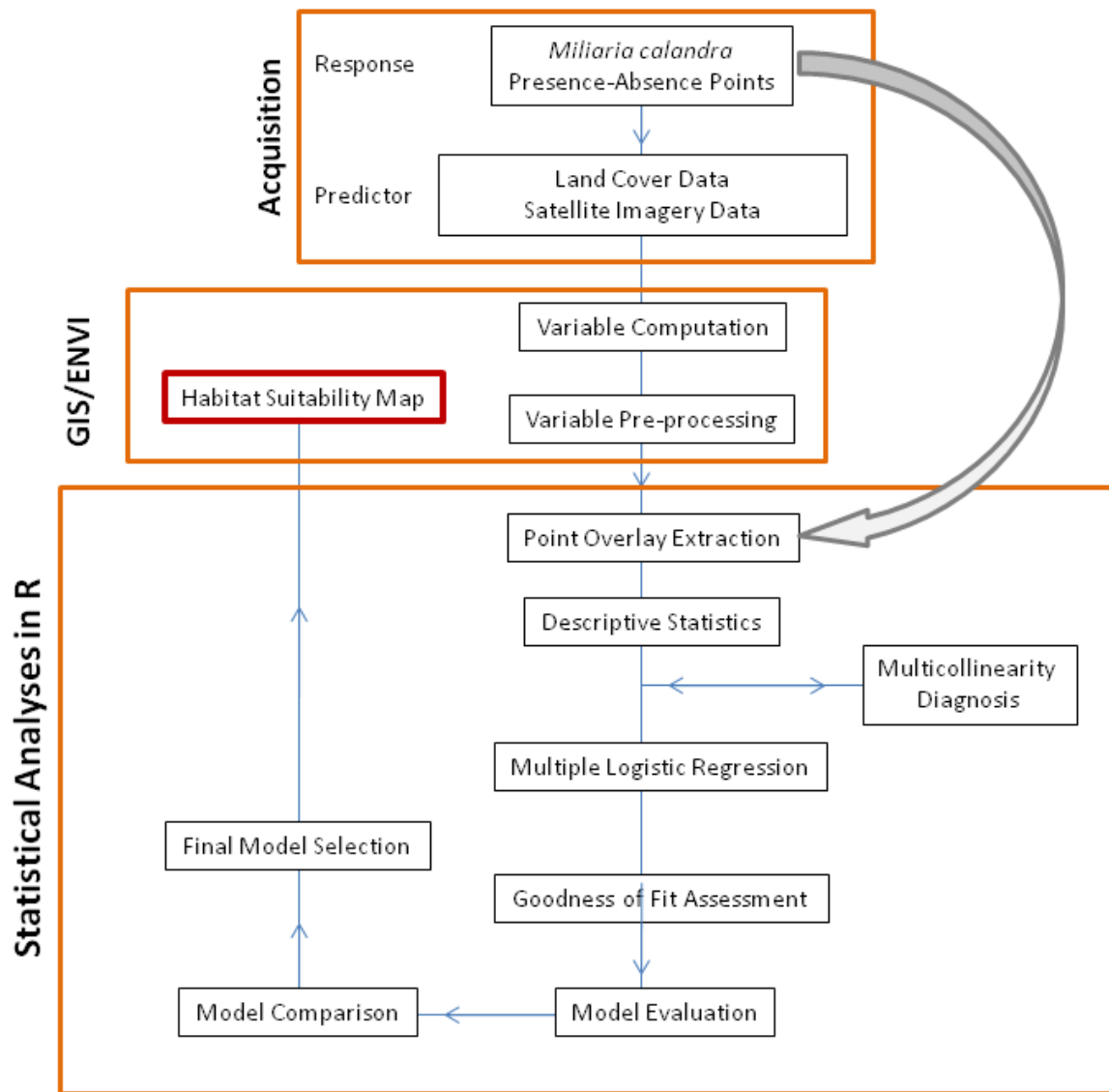


Figure 3: Thesis design flowchart

Chapter 2 Data

This chapter describes how the predictor variables from satellite imagery and land cover data were extracted and preprocessed for use in the statistical procedure that follows.

2.1 Satellite imagery

Imagery from the Enhanced Thematic Mapper Plus (ETM+) sensor onboard Landsat 7 satellite was downloaded from the United States Geological Survey (USGS) Global Visualization Viewer version 7.26. Two scenes from path-198, row-31 for the month of June (01/06 & 17/06) of the year 2001 were used for this study to temporally coincide with the survey period. Imagery used is a standard level-one terrain-corrected (L1T) product that has also undergone radiometric and geometric correction. This product level was chosen because the L1T correction employs ground control points and digital elevation models to achieve complete geodetic accuracy (USGS 2009).

Table 1: Landsat 7 ETM+ band characteristics

	Band	Spatial resolution (m)	Lower limit (μm)	Upper limit (μm)	Bandwidth (nm)	Bits per pixel	Gain	Offset
1	BLUE	30	0.45	0.52	70	8	0.786	26.19
2	GREEN	30	0.53	0.61	80	8	0.817	26.00
3	RED	30	0.63	0.69	60	8	0.639	24.50
4	NIR	30	0.75	0.90	150	8	0.939	24.50
5	MIR	30	1.55	1.75	200	8	0.128	21.00
6	THERMAL	60	10.40	12.50	2100	8	0.066	0.00
7	MIR	30	2.10	2.35	250	8	0.044	20.34
8	PAN	15	0.52	0.90	380	8	0.786	26.19

Atmospheric correction was performed using the Quick Atmospheric Correction (QUAC) method available in the ENVI 4.7 image processing software. QUAC is a method for atmospherically correcting multispectral imagery in the visible, near infrared and through mid-infrared region (0.4 – 2.5 μm). The method was chosen because of its ability to determine atmospheric compensation parameters directly from information contained within the scene without the need for ancillary information and also allows for any view or solar elevation angle resulting in accurate reflectance spectra (ITTVIS 2009). Clouds were masked and the imagery underwent pixel-by-pixel averaging to produce a single representative image for the month.

2.2 Satellite imagery preprocessing

2.2.1 Texture analysis

Image texture represents the visual effect produced by the spatial distribution of tonal variability (pixel values) in a given area (Baraldi and Parmiggiani 1995). Satellite image texture can thus serve as a substitute for habitat structure because variability in the reflectance among adjacent pixels can be caused by horizontal variability in plant growth (St-Louis et al. 2009). Due to the size of the study area, a 3x3 pixel local statistic was selected to calculate first order texture measures of mean, standard deviation and coefficient of variation. The mean computes the average texture value, the standard deviation assesses the variability of texture and the coefficient of variation is standard deviation of pixel values divided by the mean and gives a measure of the variability in image texture as a percent of the mean. St-Louis et al (2006) has indicated that first order standard deviation to be the best predictor amongst the first order texture measures.

2.2.2 Calculation of vegetation indices

Photosynthesis in green vegetation requires the absorption of solar radiation in the region 400–700 nm (called photosynthetically active radiation or PAR) for use as an energy source (Alados-Arboledas et al. 2000). Beyond the PAR, in the near-infrared region, the absorption decreases significantly and the vegetation instead reflects radiation. Due to this

strong difference in absorption and reflectance, a relatively simple algorithm, the Normalized Difference Vegetation Index (NDVI) was developed (Tucker 1979):

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

(Equation 1)

Where NIR refers to the near-infrared band (ETM4) and RED refers to the visible red band (ETM3). The resultant reflectance values are in the form of ratios of the reflected over the incoming radiation. NDVI ranges between -1 and +1; negative values indicate lack of vegetation while positive values indicate the presence of vegetation.

NDVI has been proven to be correlated with ecological and physical conditions such as land cover, vegetation composition, species richness and productivity of many species (Wallin et al. 1992; Sanz et al. 2003; Seto et al. 2004; Foody 2005; Pettorelli et al. 2005).

Modified Soil Adjusted Vegetation Index (MSAVI) was also added as a predictor variable because the algorithm possesses a correction factor that can be adjusted according to vegetation density (Liang 2004; Qi et al. 1994). MSAVI has been shown to enhance the dynamic range of the vegetation signal, producing greater vegetation sensitivity (Qi et al. 1994). It is defined as:

$$MSAVI = NIR + 0.5 - \sqrt{(NIR + 0.5)^2 - 2(NIR - RED)}$$

(Equation 2)

The correction factor (0.5) is generally used for most applications and represents areas with intermediate vegetation densities. The amount of detail produced by MSAVI compared to NDVI is highlighted in Figure 4.

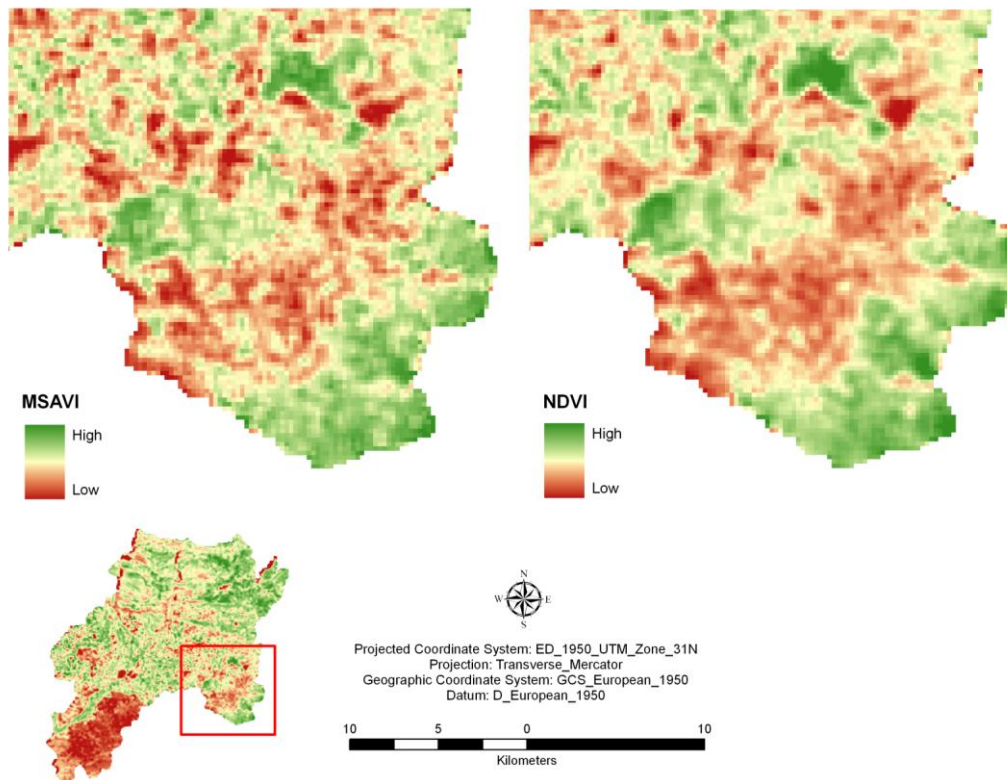


Figure 4: MSAVI vs. NDVI

2.2.3 Tasseled cap transformation

The tasseled cap transformation (TCT; Crist and Kauth 1986) translates multispectral bands into a feature space that denotes the physical characteristics of the ground cover (Liang 2004). TCT returns six bands, the first three of which: brightness, greenness and wetness are of relevance. The brightness band corresponds to overall reflectance, greenness is a measure of vegetation health and structure and the wetness band measures soil moisture and vegetation density (Crist 1983). The first three TCT bands have been shown to explain up to 97% of the spectral variance in individual Landsat scenes (Huang et al. 2002) and strongly correlate with avian composition and tree cover (Ranganathan et al. 2007).

Table 2: Tasseled cap transformation coefficients for Landsat ETM+ (Liang 2004)

Feature	Band 1	Band 2	Band 3	Band 4	Band 5	Band 7
Brightness	0.3561	0.3972	0.3904	0.6966	0.2286	0.1596
Greenness	-0.3344	-0.3544	-0.4556	0.6966	-0.0242	-0.2630
Wetness	0.2626	0.2141	0.0926	0.0656	-0.7629	-0.5388
Fourth	0.0805	-0.0498	-0.1050	-0.1327	-0.5752	-0.7775
Fifth	-0.7252	-0.0202	0.6683	0.0631	-0.1494	-0.0274
Sixth	0.4000	-0.8172	0.3832	0.0602	-0.1095	0.0985

2.2.4 Land surface temperature

The first step of obtaining LST involves accounting for the land surface emissivity (LSE) of the study area. Surface emissivity is a quantification of the intrinsic ability of a surface in converting heat energy into above-surface radiation and depends on the physical properties of the surface and on observation conditions (Sobrino et al. 2001). LSE was calculated following the procedure by Sobrino et al (2004).

LSE can be extracted by using NDVI considering three different cases (1) bare ground (2) fully vegetated and (3) mixture of bare soil and vegetation (Sobrino et al. 2004). Since the study area falls within the third case, the following equation is used to extract LSE:

$$\varepsilon = 0.004P_v + 0.986$$

(Equation 3)

Where ε is the LSE and P_v is the proportion of vegetation obtained and is calculated by:

$$P_v = \left[\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right]^2$$

(Equation 4)

Where :

$NDVI_{max} = 0.5$ and $NDVI_{min} = 0.2$

The next step involves calculating the at-sensor radiance (L_λ), which is the amount of energy that reaches the satellite sensor:

$$L_\lambda = \frac{(L_{Max} - L_{Min})}{(Q_{CalMax} - Q_{CalMin})} * (DN - Q_{CalMin}) + L_{Min}$$

(Equation 5)

Where:

DN = the quantized calibrated pixel value in DN

LMin = the spectral radiance that is scaled to QCalMin in watt/m² * ster * μm

LMax = the spectral radiance that is scaled to QCalMax in watt/m² * ster * μm

QCalMin = the minimum quantized calibrated pixel value (corresponding to LMin) in DN

QCalMax = the maximum quantized calibrated pixel value (corresponding to LMax) in DN

The at-sensor radiance is in turn converted to the effective at-satellite temperatures of the viewed Earth-atmosphere system under an assumption of unity emissivity (USGS 2009). This is also referred to as blackbody temperature and denotes a surface that absorbs all the electromagnetic radiation that reaches it. The blackbody temperature is calculated by:

$$T_B = \frac{K_2}{\left(\ln \left(\frac{K_1}{L_\lambda + 1} \right) \right)}$$

(Equation 6)

Where:

K_1 = Calibration constant 1 (666.09 watt/m² * ster * μm)

K_2 = Calibration constant 2 (1282.71 K)

L_λ = At-sensor radiance calculated from Equation 5.

A final step involving correction for spectral emissivity is necessary according to the nature of the surface:

$$LST = \frac{T_B}{1 + \left(\frac{\lambda T_B}{\rho}\right) \ln \varepsilon TM_6}$$

(Equation 7)

Where:

TB = Blackbody temperature from Equation 6.

λ = Wavelength of emitted radiance (11.5 μ m)

$\rho = h \times c / \sigma = 1.438 \times 10^{-2}$ mK (σ =Boltzmann constant= 1.38×10^{-23} J/K, h =Planck's constant= 6.626×10^{-34} Js, c =velocity of light= 2.998×10^8 m/s)

$\ln \varepsilon$ = Land surface emissivity calculated from Equation 3.

TM₆ = Landsat thermal band 6 in DN

All LST retrieval algorithms and descriptions apart from LSE estimation are according to the Landsat Science Data User's Handbook (USGS 2009)

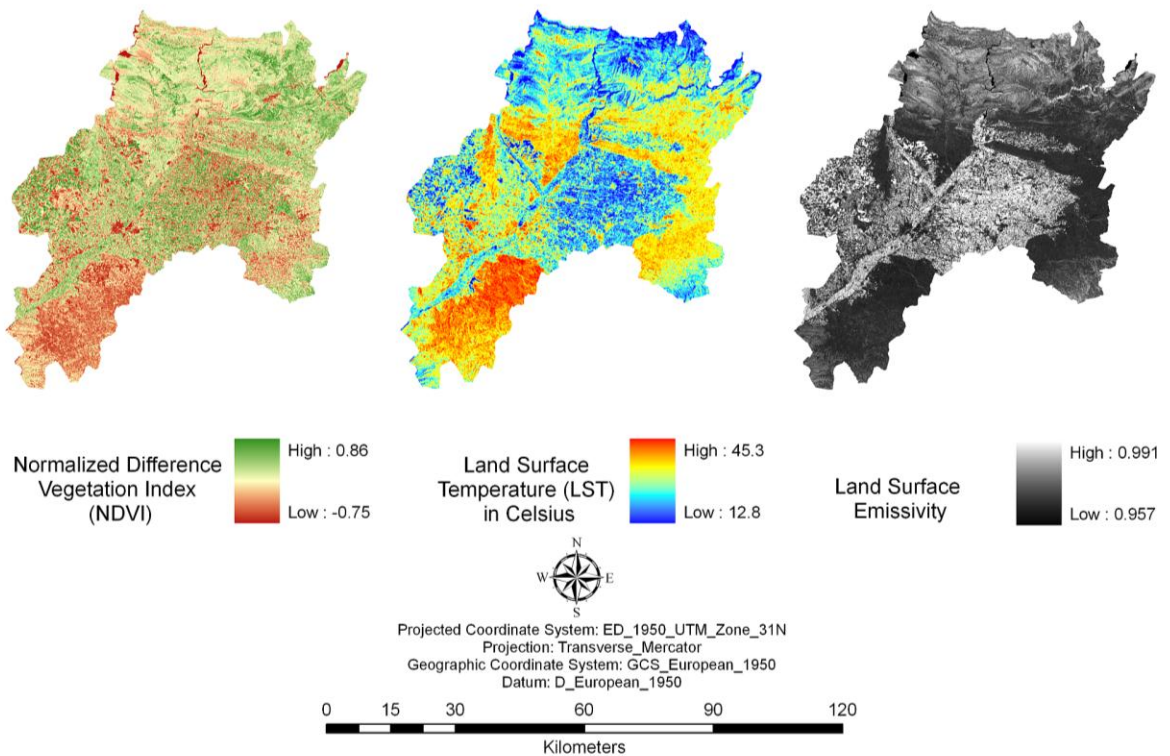


Figure 5: NDVI values compared to Land Surface Temperature and Land Surface Emissivity

2.2.5 Topographic variables

Topography indirectly affects the distribution of species by modifying the relationships of birds with vegetation or by modifying the vegetation types (Seoane et al. 2004a). Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) resampled to 250m was downloaded from the CGIAR-CSI database.

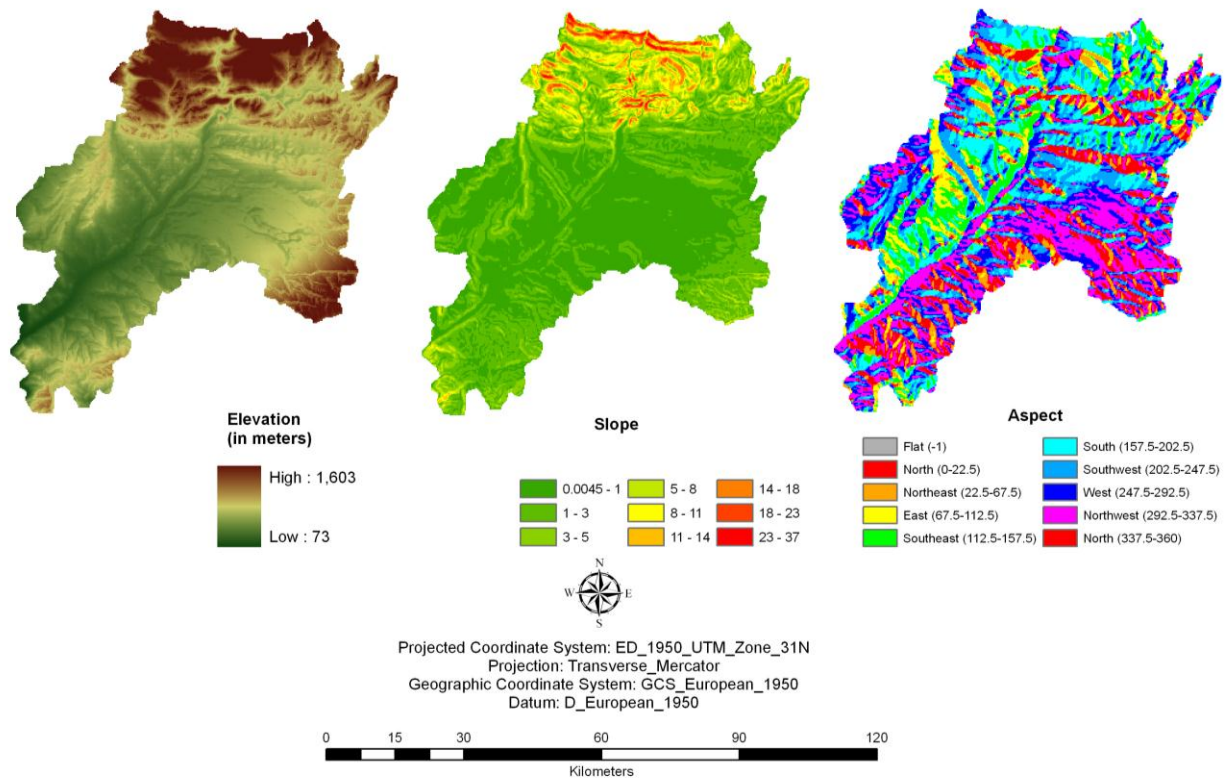


Figure 6: Topographic variables employed in the study

2.3 Land cover data

CORINE Land Cover (CLC) data for the year 2000 (CLC2000; dated 01/01/2002) was downloaded from the EEA's online portal. CLC is a pan-European project that aims to produce distinctive and comparable land cover data set for Europe. CLC has a total of 44 land cover classes out of which 27 classes occur in the study area (Figure 7).

2.4 Land cover data preprocessing

2.4.1 Anthropogenic variables

Anthropogenic factors such as road density are important measures for predicting bird assemblages in agricultural eco-regions (Whited et al. 2000). A vector shapefile of the major roads in the study area was obtained from ESRI Data and Maps 2002 and the Euclidian distance to roads was calculated. One anthropogenic factor was extracted from the land cover map: distance to human activity. This was done by rasterizing the CLC2000 map and extracting only the CLC codes which correspond to human activity:

- Continuous urban fabric
- Discontinuous urban fabric
- Industrial or commercial units
- Construction sites
- Mineral extraction sites

This was followed by calculating the Euclidian distance of each pixel to the above land cover classes. Due to space limitations, the figures of the anthropogenic variables are exhibited in Appendix A.

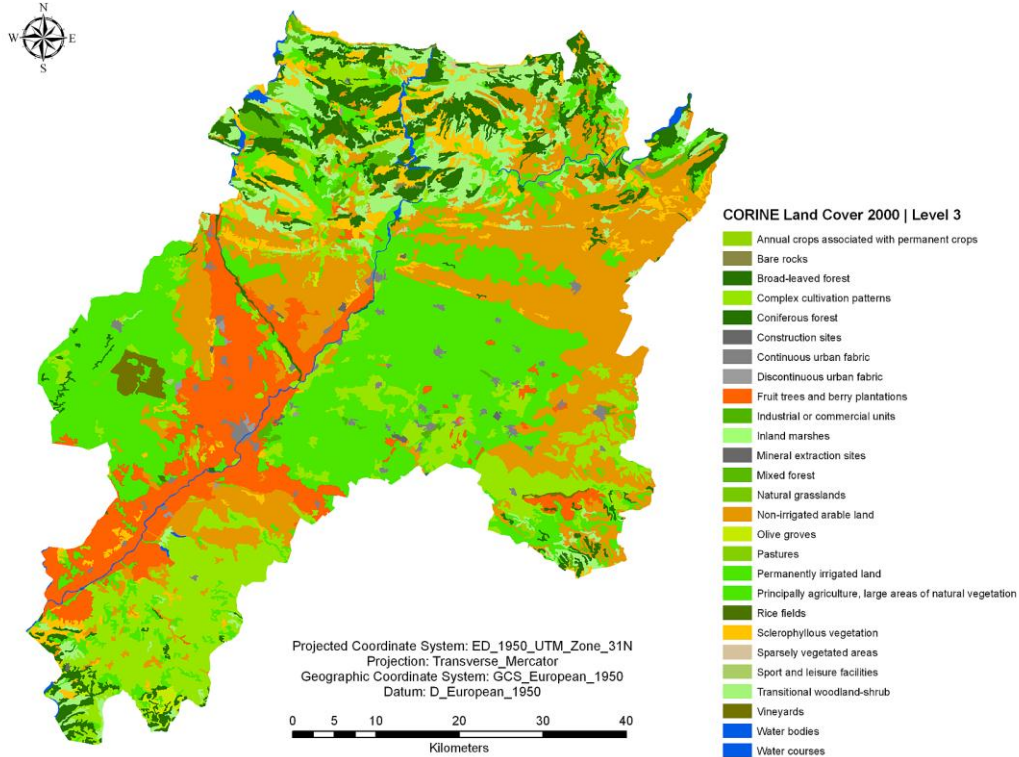


Figure 7: The 27 CORINE Land Cover 2000 classes in the study area.

2.4.2 Landscape metrics

Landscape metrics are indices developed for categorical map patterns that quantify specific spatial characteristics of patches, classes of patches, or entire landscape mosaics (Smith et al. 2003). They help explain how spatial patterns of landscapes influence the most important ecological processes (Carrao and Caetano 2002) and have also been applied in an urban context (Cabral et al. 2005).

Compositional metrics were calculated from the CLC2000 data and included the proportions of habitat types and landscape richness. Local statistics were calculated using a 3x3 pixel moving window to quantify the landscape metrics with 0 signifying the absence of the metric in the window and 1 signifies that the window is fully covered by the metric (Figure 8). Following are the eight dominant habitat types calculated using this method:

- Broad-leaved Forest
- Complex Cultivation Patterns
- Fruit Trees and Berry Plantations
- Non-irrigated Arable Land
- Permanently Irrigated Land
- Sclerophyllous Vegetation
- Principally Agricultural with Natural Vegetation
- Transitional Woodland-shrub

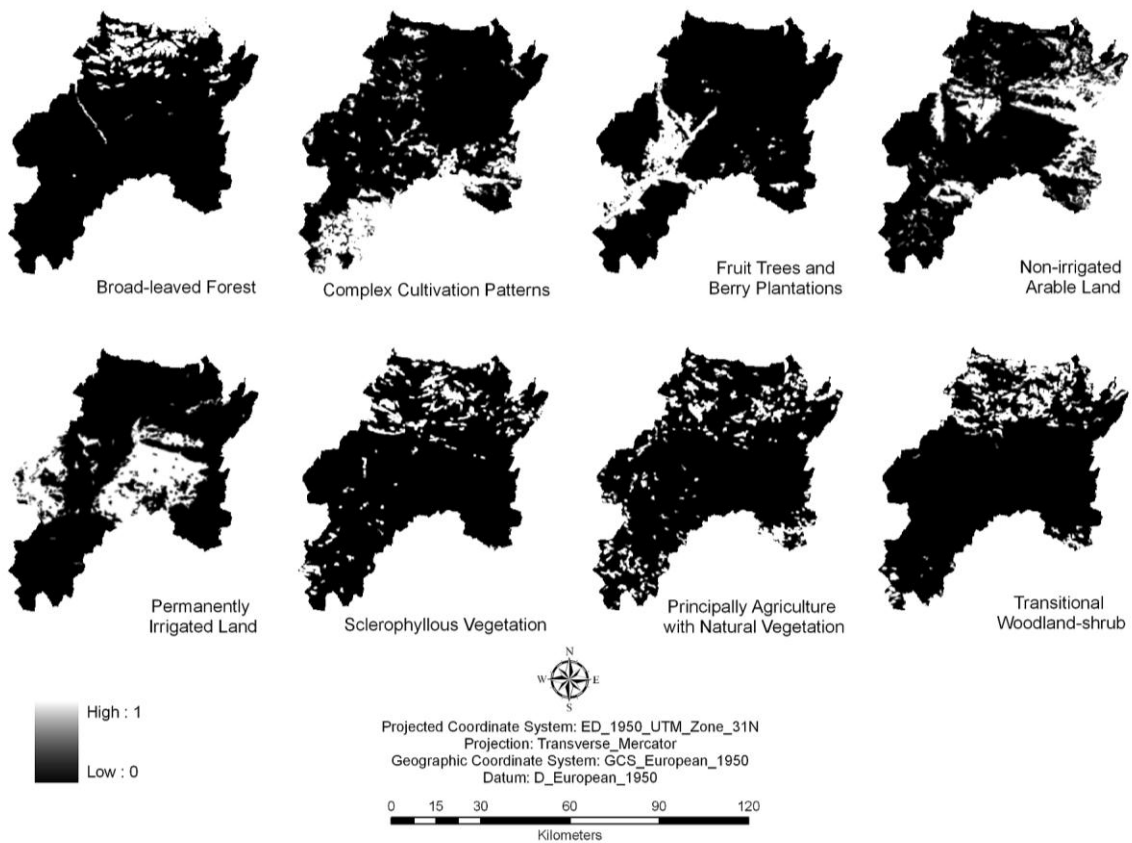


Figure 8: The eight landscape metrics that were extracted from CLC2000

2.5 Catalan breeding bird atlas

Data was provided by the Catalan breeding bird atlas (CBBA) in the form of presence/absence records of eight bird species. Surveys were conducted in the summer breeding season (March 1st to July 30th) in the years 1999-2002. Surveys were conducted between sunrise and 11 am, and between 6 pm and sunset. The survey plots were 1 km x1 km UTM squares in which two 1-hour surveys were conducted and the presence or absence of each species recorded. The CBBA does not allow the use of tapes or lures to increase the attract species (Brotons et al. 2008). The assignment of the category “Confirmed Breeding” was performed following guidelines set by the European Ornithological Atlas Committee and includes (Brotons et al. 2008):

- Anti-predatory displays
- Nest used during current breeding season
- Recently fledged young
- Adult carrying fecal sacs or food
- Nest with eggs or bird incubating
- Nest with young; or young of nidifugous species

Since the records of all eight bird species were spread out over the five months and indeed over all four years, a subset of one farmland species, the corn bunting *Miliaria calandra* was selected as a response variable (Figure 9).

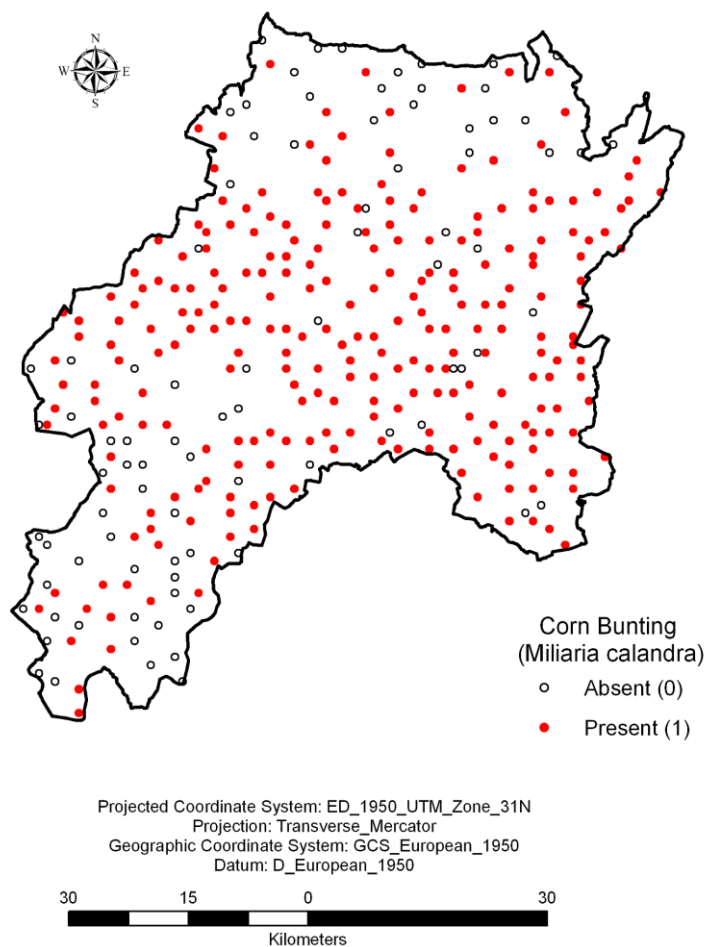


Figure 9: Corn bunting presence-absence points

2.6 Analysis tools

The primary tool for statistical analysis is the R Environment for Statistical Computing version 2.9.2 (R Development Core Team 2009) using the R Commander graphical user interface version 1.5-3 (Fox 2005). Open Office 3.1 was used to manipulate tabular data. GIS analysis, creation and visualization of predictive surfaces were conducted using ArcGIS 9.3. Satellite image analysis was done in ENVI 4.7. The EPSG:23031 projection was retrieved from the EPSG list provided in the `rgdal` package (Bivand et al. 2008; Keitt et al. 2009) and used to project both the response and the explanatory predictor variables.

Chapter 3 Methodology

This chapter describes the statistical analyses that were employed to produce probability maps of the occurrence of the corn bunting based on habitat preference. In this and subsequent chapters, the statistical terminology 'response variable' refers to the corn bunting. It is the target species whose response was modeled based on a set of predictor variables. The entire R code used in this study is presented in Appendix D.

3.1 Bivariate descriptive statistics

Bivariate descriptive statistics were calculated to gauge the relationship between each predictor and the response variable. Furthermore, the relationship of satellite derivatives with the land cover codes is presented in Appendix B.

Regression coefficients explain the amount of contribution of each predictor variable in terms of the log odds of the response variable. A positive coefficient expresses a directly proportional relationship while a negative coefficient expresses an inversely proportional relationship. The magnitude of the coefficient describes the strength of influence of that predictor variable. The standard error assesses the precision of the regression coefficient measurements and is an approximation of the standard deviation of the coefficients. The Z-value is basically the value of each coefficient divided by its standard error. The square of the Z-value is approximately a chi-square statistic with one degree of freedom called the Wald statistic (Kleinbaum and Klein 2002). The presence of high multicollinearity between the predictor variables causes an inflation of the standard errors causing lower values of the Wald statistic and creating Type II errors (Menard 2002). A p-value of 0.05 means that there is 5% likelihood that the model results would be produced in a random distribution, so there is 95% likelihood that the variable in question has a significant effect on the model.

3.2 Multiple logistic regression

The statistical method employed in this study is multiple logistic regression. There are several statistical methods that use binary data for mapping the distribution of species based on habitat preference. However, they exhibit certain drawbacks.

Artificial neural networks (ANN) demonstrate a good predictive capability but an assessment of the relative contribution of the predictors is quite difficult. Methods such as ecological niche factor analysis (ENFA; Hirzel et al. 2002) offered in the Biomapper software, while offering good predictions, obscures the internal workings of the algorithm so the process which has resulted in the predictions is unclear. Others such as genetic algorithm for rule-set prediction (GARP; Stockwell and Peters 1999) use presence-only data and create random pseudo-absences for presence-absence modeling. The flaw in this method is that pseudo-absences points might be allocated to areas that possess favorable habitats. Brotons et al (2004b) have shown that the use of recorded absence data yields better predictions than pseudo-absences and they recommend their inclusion into habitat modeling algorithms.

Therefore logistic regression, implemented through R, stands out as a viable method that offers the combination of methodological transparency, assessment of predictor contribution, and allows the use of recorded absence data.

Logistic regression is a binomial generalized linear model that predicts the probability of occurrence of an event using a binary response variable and multiple covariates (Hosmer and Lemeshow 2000). The probability distribution is fitted to the sigmoid logistic curve and the outcome is between 0 and 1. Imagine that π is the probability of an event occurring, hence the logit of Y from a set of predictor variables ($X_1 \dots X_n$) is:

$$\mathit{logit} = \ln \left[\frac{\pi}{1 - \pi} \right] = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots b_n X_n$$

(Equation 8)

Where b_0 is a constant (the y-intercept) and $b_1, b_2, b_3 \dots b_n$ are the regression coefficients estimated by the maximum likelihood method. The formula above states that the response variable represents the input of all the variables in the model. The response is transformed to a logit variable and a maximum likelihood approximation is implemented. The logit variable is the natural log of the odds of the response being 1 or 0, hence estimating the odds whether an event (represented by the response) will occur. Hence, the probability of Y occurring is given by:

$$\pi = \frac{e^{b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots b_n X_n}}{1 + e^{b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots b_n X_n}}$$

(Equation 9)

The logistic models were fitted using the `glm` function of the `stats` (R Development Core Team 2009) package.

3.3 Multicollinearity diagnosis

Multicollinearity refers to extreme correlation between the predictor variables. This leads to a situation where the regression model fits the data well, but none of the predictors has any significant impact in predicting the dependent variable because they basically share the same information (Ho 2006). Sometimes predictors in high correlation that individually explain a significant portion of deviance can appear non-significant due to the collinearity (Guisan et al. 2002). Pearson correlation coefficients can be computed using the `cor` function in R, however that pair-wise approach is limited to only two variables at a time and does not account for correlation between multiple variables. Therefore, the variance inflation factor, VIF (Brauner and Shacham 1998) has been computed for each variable to detect multicollinearity. VIF is calculated as:

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

(Equation 10)

The expression $1-R^2$ is the tolerance and R^2 is the proportion of variance the predictor variables explain in the response variable. The function `vif` in the `Design` package (Harrell 2009) was used to compute VIF values.

3.4 Variable selection

Models that have too few predictor variables can introduce bias in the inference process, while models that possess too many variables could yield poor precision or identification of effects that are actually non-existent (Burnham and Anderson 2004). Since there are several derivatives of Landsat bands in use, the problem of multicollinearity can lead to a high degree of unreliability in the estimated regression coefficients (Kleinbaum and Klein 2002), therefore the satellite model underwent a stepwise selection process to pick variables that significantly contribute to the model's ability to describe the data. All the satellite predictor variables were placed in the model and then an iterative forward-backward elimination (Pearce and Ferrier 2000a) of the non-significant variables was performed. Then, variables with high VIF values were removed one at a time until all the variables have VIF values below 10 which is the threshold below which multicollinearity is not of concern (Brauner and Shacham 1998).

3.5 Assessing goodness of fit and model validation

A goodness of fit assessment describes how well a given model fits the data by measuring the deviation between observed values and the values produced by the model. Two measures of goodness of fit are used here: Pearson Chi-square and the Likelihood ratio test.

Pearson Chi-square (χ^2) test statistic evaluates H_0 that the independent variables are not in a linear relationship to the log-odds of the response. This test evaluates improvement contributed by the independent variables compared to H_0 :

$$\chi^2 = \sum_{b=1}^n \frac{(O_b - E_b)^2}{E_b}$$

(Equation 11)

Where O the observation, E is the expectation, n is the amount of possible results.

Logistic models provide a better fit to the data if improvement over the null model is exhibited (Hosmer and Lemeshow 2000). The likelihood ratio test is based on the disparity between deviance of the intercept-only model minus the deviance of the full model. The test was performed using the `lrtest` function of the `lmtest` package (Zeileis and Hothorn 2002). Likelihood is the probability of the response's observed values to be predicted from the predictor variables. The likelihood ratio test statistic is given by:

$$\textit{likelihood ratio} = -2\ln L_1 - (-2\ln L_2)$$

(Equation 12)

Where L_1 and L_2 denote the maximized likelihood values for models 1 and 2 respectively; this is a χ^2 distributed statistic with degrees of freedom equal to the number of predictors and is a measure of how poorly the model predicts the decisions. It is a probability that ranges from 0 to 1. The log likelihood of this probability produces a value between 0 for no significance, and $-\infty$ for high significance, however by multiplying that value by -2 , the high significance value would be $+\infty$.

In ordinary least squares, the coefficient of determination, R^2 , serves as a statistic that ranges from zero to one and summarizes the overall strength of a given model. There is no such statistic for logistic regression but a number of pseudo- R^2 statistics have been proposed in the last three decades (Hu et al. 2006). One of them, the Nagelkerke R^2 , implemented through the `lrm` function in the `Design` package (Harrell 2009), is used here. Pseudo- R^2 is defined as the proportion of the variance of the response variable that is explained by the independent variables (Hu et al. 2006).

The model performance is estimated by measuring the true error rate. The predicted probabilities of the chosen models are corroborated with the actual values to determine if high probabilities are associated with incidents (1) and low probabilities with non-incidents (0). Since the dataset has quite a limited set of observations, a cross validation resampling technique was chosen to evaluate performance. The K-Fold Cross-Validation performs K random splits of the dataset, with each split retained for testing and the remaining K-1 for training. By training and testing the model on separate subsets of the data, an idea of the model's prediction strength is obtained (Tibshirani and Tibshirani 2009). Each K-1 split produces an error rate; hence, the true error (E) is estimated as an average of the separate error rates:

$$E = \frac{1}{K} \sum_{x=1}^K E_x$$

(Equation 13)

The benefit of this method is that all records are used for both training and testing. The cross validation was performed using the `cv.glm` function in the `boot` package (Canty and Ripley 2009).

3.6 Model evaluation and selection

Model predictive power was evaluated using area under the curve (AUC) of the receiver operating characteristic (ROC) which relates sensitivity (true positive) on the y-axis against the corresponding 1 minus specificity values (false positive) on the x-axis for a wide range of threshold levels (Pearce and Ferrier 2000b). The closer the AUC value is to 1.0, the better the model performance. The AUC index is significant due to the single measure of general accuracy it provides that is not reliant on a particular threshold (Deleo 1993). AUC analysis was performed using functions in the `Presence-Absence` package (Freeman and Moisen 2008).

Models were compared using the Akaike Information Criterion, AIC (Akaike 1973) which offers a clear-cut comparison between models that is not reliant on a hypothesis testing context (Burnham and Anderson 1998). This method is preferred because it extracts more information from the data regarding the relative strength of evidence for each variable and model (Young and Hutto 2002). AIC is described by the following formula:

$$AIC = -2A + 2b$$

(Equation 14)

The first part, A, is the probability of the data given a model and the second part, b, is the number of parameters in the model. The first part approximates how well the model fits the data. The second part is a penalty which relies on the number of parameters used. Smaller values of the AIC indicate a better fit of the model to the observed data.

Chapter 4 Results

4.1 Overlay analysis

The corn bunting occupies 251 1x1 UTM squares which represents 73.8% of the total number of squares in the study area. The predictor variables in the form of ASCII files were imported into R using the `readGDAL` function of the `rgdal` package (Keitt et al. 2009) and corn bunting presence points were then overlaid on the ASCII files using the `overlay` function of the `sp` package (Pebesma and Bivand 2005). Table 3 shows the mean, minimum and maximum of the predictor variables in the occupied squares.

Table 3: Mean, Minimum and Maximum values of predictor variables in occupied squares

Variable	Mean	Min	Max	Variable	Mean	Min	Max
lst	28.57	23.28	33.37	band7cv	0.325	0.075	0.831
band1m	26.63	14.88	45.88	bright	130.51	32.69	182.54
band2m	34.96	15.22	62.11	green	3.84	-42.22	36.72
band3m	46.83	15.77	90.22	wet	-63.35	-98.06	-28.40
band4m	90.23	26.66	125.55	dem	329.64	134.22	832.44
band5m	77.65	24.22	116	slope	1.92	0.082	15.11
band7m	52.69	15.77	93.55	aspect	209.73	0.721	358.91
ndvi_m	0.335	-0.188	0.638	panv	0.054	0	0.888
msavi_m	0.449	-0.225	0.760	blf	0.020	0	0.666
band1sd	7.38	1.33	17.17	ccp	0.143	0	1
band2sd	9.96	1.69	22.52	ftbp	0.081	0	1
band3sd	16.90	3.37	37.51	nial	0.303	0	1
band4sd	13.33	3.29	38.62	pil	0.170	0	1
band5sd	16.28	3.13	33.47	sveg	0.046	0	0.888
band7sd	16.92	2.85	35.75	tws	0.030	0	0.666
band1cv	0.273	0.057	0.635	lcrich	1.75	1	4
band2cv	0.282	0.067	0.615	wetdist	11200.68	24.90	41443.13
band3cv	0.363	0.096	0.839	humdist	3931.08	0	13221.3
band4cv	0.153	0.034	1.17	roadsdist	1828.72	0	8961.72
band5cv	0.214	0.054	0.948				

4.2 Bivariate descriptive statistics

Bivariate descriptive statistics involves concurrently examining two variables to conclude if there is a relationship between them (Appendix B). The results of the descriptive statistics are summarized in Table 4. The regression coefficients produced by NDVI and MSAVI have strong positive correlation with the response variable. The Non-irrigated Arable Land (NIAL) coefficient also produced strong positive correlation while Broad-leaved forest (BLF) and Transitional Woodland Shrub (TWS) coefficients produced strong negative correlation with the response variable which is reasonable considering the fact that corn buntings strongly favor open arable landscape and avoid wooded areas. All the predictors that produced strong correlation with the response also exhibited low (<0.05) p-values.

Table 4: Results of the bivariate descriptive statistics

Variable	Coefficient	S.E.	Z	p-Value	Variable	Coefficient	S.E.	Z	p-Value
LST	0.182	0.052	3.45	0.0006	band7cv	0.995	0.964	1.03	0.3024
band1m	-0.023	0.016	-1.42	0.1543	bright	0.0077	0.0052	1.47	0.1418
band2m	-0.0098	0.012	-0.79	0.4287	green	0.029	0.0079	3.71	0.0002
band3m	-0.0057	0.0077	-0.75	0.4529	wet	0.00065	0.0093	0.07	0.9445
band4m	0.045	0.0091	5.02	0.0000	dem	-0.0015	0.00065	-2.29	0.0220
band5m	0.0016	0.0078	0.21	0.8299	slope	-0.252	0.047	-5.27	0.0000
band7m	-0.0076	0.0081	-0.93	0.3506	aspect	-0.0012	0.0012	-1.01	0.3113
NDVI	2.97	0.920	3.23	0.0012	panv	-0.522	0.775	-0.67	0.5003
MSAVI	2.31	0.677	3.41	0.0006	blf	-2.45	0.895	-2.74	0.0061
band1sd	0.0060	0.036	0.16	0.8689	ccp	-0.393	0.390	-1.01	0.3142
band2sd	0.017	0.027	0.64	0.5230	ftbp	-0.479	0.460	-1.04	0.2981
band3sd	0.027	0.016	1.63	0.1034	nial	2.69	0.594	4.54	0.0000
band4sd	0.012	0.019	0.64	0.5234	pil	1.13	0.397	2.86	0.0042
band5sd	-0.0055	0.018	-0.30	0.7639	sveg	-0.449	0.798	-0.56	0.5737
band7sd	0.022	0.018	1.26	0.2092	tws	-2.46	0.722	-3.42	0.0006
band1cv	1.17	1.24	0.94	0.3449	lcrich	-0.519	0.157	-3.29	0.0010
band2cv	1.39	1.22	1.14	0.2539	wetdist	0.000076	1.9E-05	3.94	0.0001
band3cv	1.92	0.934	2.06	0.0396	humdist	-0.000126	3.7E-05	-3.39	0.0007
band4cv	-1.00	1.005	-1.00	0.3184	roadsdist	-0.000063	6.9E-05	-0.92	0.3591
band5cv	-0.99	1.049	-0.95	0.3430					

4.3 Satellite model

The minimal adequate model for the satellite predictor variables is summarized in Table 5 and the resultant map in Figure 10. The selected model contains seven predictor variables.

Table 5: Summary results of the logistic regression analysis for the satellite model

	Coefficient	S.E.	Z	p-Value
(Intercept)	-12.74107	3.00470	-4.24000	0.0002
band4m	0.04256	0.01538	2.76700	0.00565
msavi_m	3.43200	0.97912	3.50500	0.00046
band1sd	-0.09312	0.05626	-1.65500	0.09787
band5cv	5.25594	1.73581	3.02800	0.00246
dem	0.00319	0.00126	2.52900	0.01144
slope	-0.30126	0.07579	-3.97500	0.00007
lst	0.28615	0.07197	3.97600	0.00007
ND	388.24	df	338	
RD	302.81	df	331	
AIC	318.81			
Pearson ChiSq	92.4039	PCC	0.7905	
L.R.	85.43	AUC	0.8095	
R2	0.331	CV Error	0.1520	

The AUC value for this model was 0.81 with 79% of the points accurately classified. The Nagelkerke pseudo- R^2 statistic was 0.33 (95% Confidence Interval: $0.251 \leq R^2 \leq 0.410$), which means that approximately 33% of the variation in the response is explained by the model. K-Fold Cross Validation yielded an error rate of 0.15. The model performed 30% better than a random model. The residual deviance (318.81) is well below the degrees of freedom (331) indicating that there is no over-dispersion in the model.

The importance of each variable is presented in visual form in Appendix C using `plot.anova.Design` function of the `Design` package (Harrell 2009).

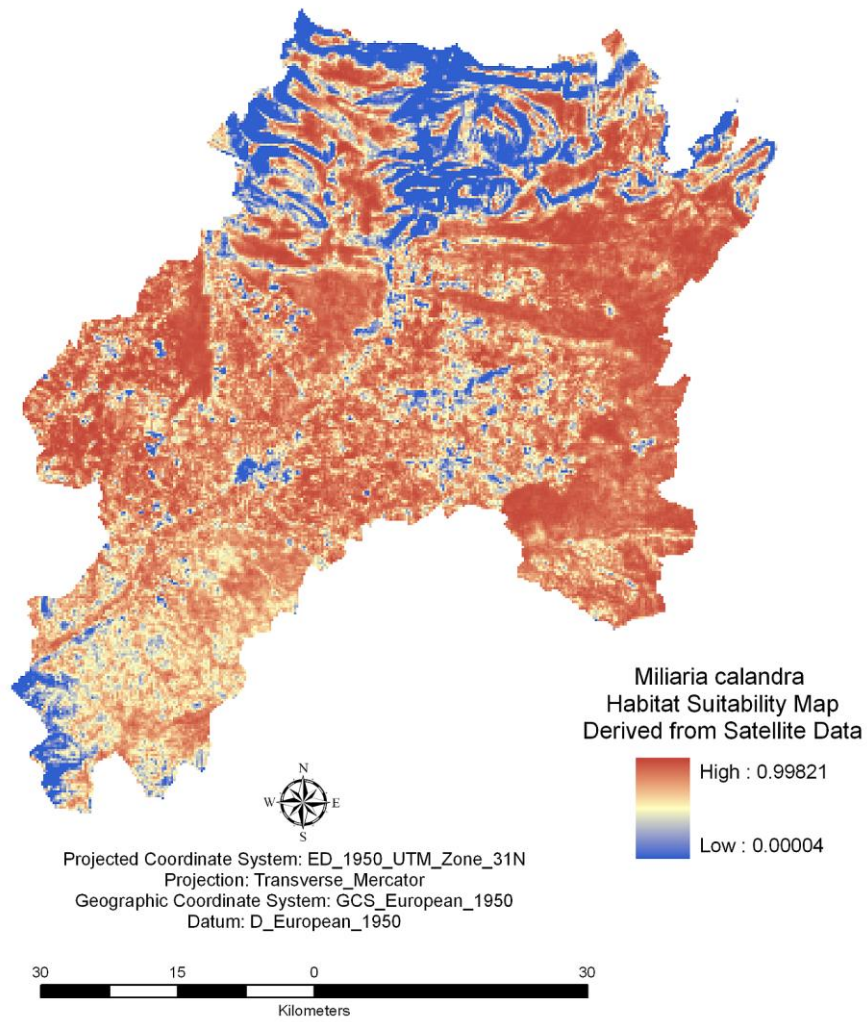


Figure 10: Habitat suitability map derived from satellite data

4.4 Land cover model

The logistic regression model for the land cover predictor variables is summarized in Table 6 and the resultant map in Figure 11. The selected model contains eight predictor variables.

Table 6: Summary results of the logistic regression analysis for the land cover model

	Coefficient	S.E.	Z	p-Value
(Intercept)	-0.70370	0.45650	-1.542	0.12320
panv	1.20700	0.86130	1.401	0.16120
ccp	1.13600	0.53980	2.105	0.03530
ftbp	1.29100	0.61390	2.103	0.03550
nial	4.03200	0.72160	5.587	2.31E-008
pil	2.39500	0.55740	4.297	0.0002
sveg	1.92400	0.97690	1.970	0.04890
humdist	-0.00008	0.00005	-1.687	0.09170
wetdist	0.00005	0.00002	2.394	0.01660
ND	388.24	df	338	
RD	304.17	df	330	
AIC	322.17			
Pearson ChiSq	88.0229	PCC	0.7964	
L.R.	84.07	AUC	0.8103	
R2	0.322	CV Error	0.1543	

The AUC value for this model was 0.81 with 79.6% of the points accurately classified. The Nagelkerke pseudo- R^2 statistic was 0.32 (95% Confidence Interval: $0.242 \leq R^2 \leq 0.401$), which means that approximately 32% of the variation in the response is explained by the model. K-Fold Cross Validation yielded an error rate of 0.15. The model performed 31% better than a random model. The residual deviance indicates the absence of over-dispersion.

Because of the coarse resolution of the CLC2000, the probability map comes out coarse as well. Although the land cover map does provide valuable information it is not as visually appealing as map derived from the satellite imagery. It is interesting to note that

even without the topographic data the land cover model assumes an unfavorable habitat in the higher altitudes with steep slopes.

The importance of each variable is presented in a visual form in Appendix C using `plot.anova.Design` function.

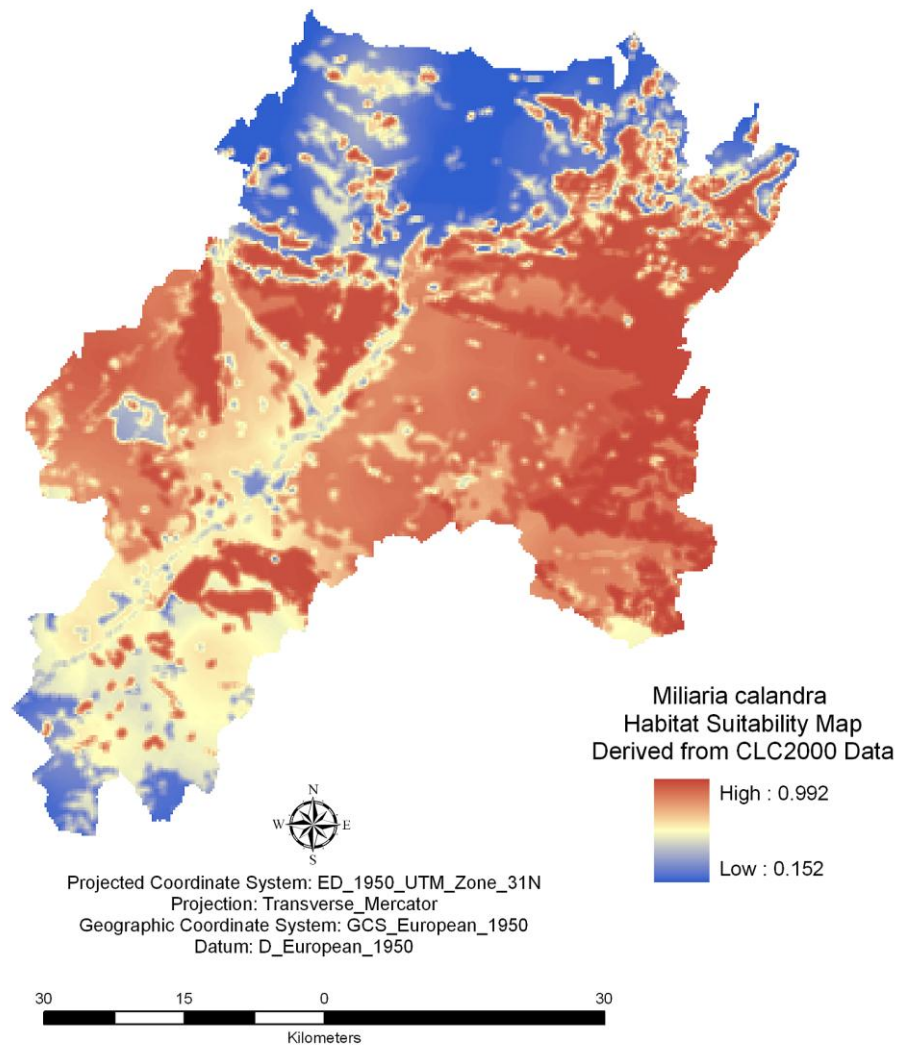


Figure11: Habitat suitability map derived from land cover data

4.5 Combined model

The logistic regression model for the combined predictor variables is summarized in Table 7 and the resultant map in Figure 12. The selected model contains twelve predictor variables.

Table 7: Summary results of the logistic regression analysis for the combined model

	Coefficient	S.E.	Z	p-Value
(Intercept)	-12.16	3.49500	-3.48	0.0050
band4m	0.02018	0.01692	1.193	0.23285
msavi_m	3.28700	1.08300	3.034	0.00241
band1sd	-0.14150	0.06071	-2.331	0.01974
band5cv	4.60000	1.85600	2.478	0.01320
dem	0.00262	0.00145	1.802	0.07152
slope	-0.21100	0.08201	-2.573	0.01008
lst	0.32840	0.09584	3.426	0.00061
nial	1.97500	0.61340	3.220	0.00128
pil	1.65800	0.68590	2.417	0.01566
sveg	1.63300	1.06100	1.539	0.12387
humdist	-0.00009	0.00005	-1.762	0.07812
wetdist	0.00004	0.00002	1.677	0.09360
ND	388.24	df	338	
RD	276.11	df	326	
AIC	302.11			
Pearson ChiSq	90.6919	PCC	0.8171	
L.R.	112.13	AUC	0.8462	
R2	0.413	CV Error	0.1433	

The AUC value for this model was 0.84 with 81.7% of the points accurately classified. The Nagelkerke pseudo- R^2 statistic was 0.41 (95% Confidence Interval: $0.335 \leq R^2 \leq 0.490$), which means that approximately 41% of the variation in the response is explained by the model. K-Fold Cross Validation yielded an error rate of 0.14. The model performed 35% better than a random model. The saturated model with all 39 variables included that does not account for high VIFs (Appendix C) displays a smaller deviance

(RD=240.19) due to the number of parameters in the model because deviance corresponds to -2 times the log likelihood of the data under the model and measures how the model predicts the decisions. Since smaller residual deviance is better, it is tempting to select this model, however, the p-values and the inflated standard errors due to the presence of multicollinearity has led to its rejection.

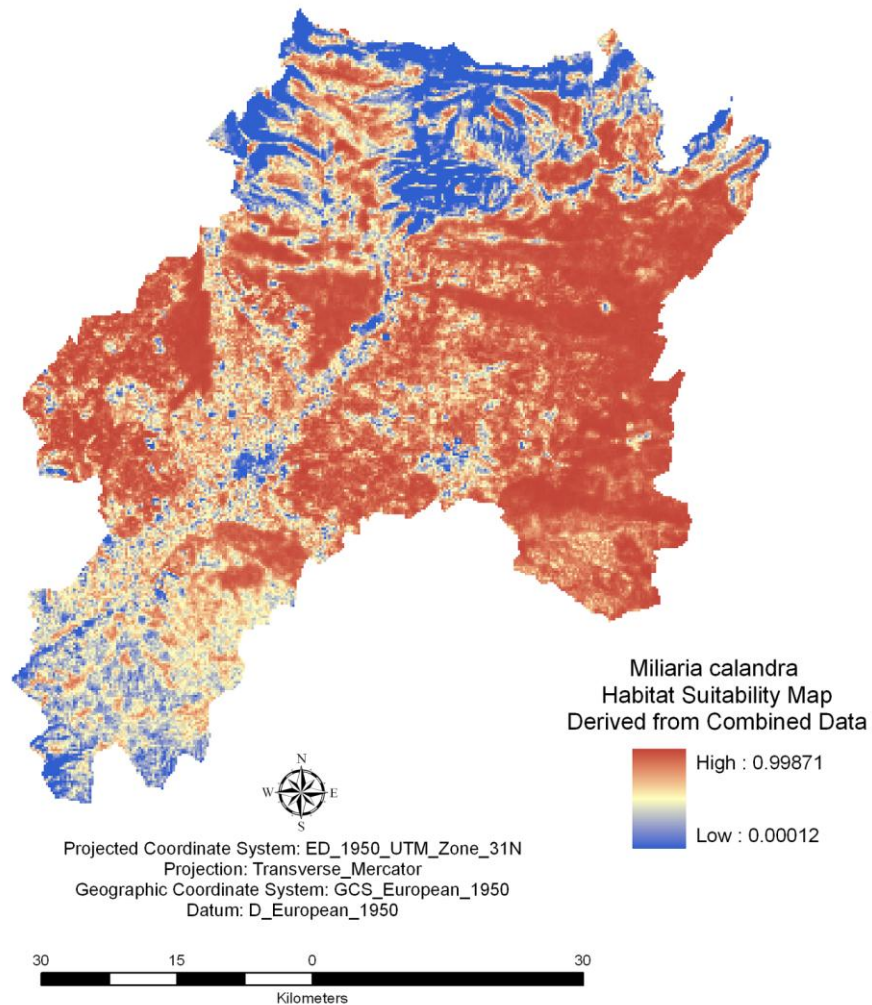


Figure 12: Habitat suitability map derived from a combination of satellite and land cover data

4.6 Model selection

A comparative receiver operating characteristic (ROC) plot provided in Figure 13 displays the performance of the combined model in relation to the satellite and land cover models.

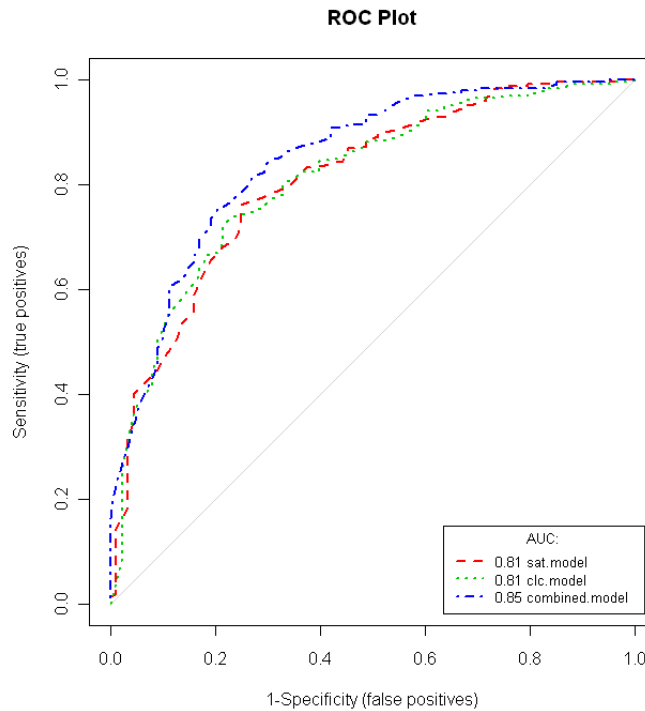


Figure 13: ROC plot of the satellite-only, CLC2000-only and combined model

The value of the area under the ROC (AUC) measures the ability of the model's predictions to distinguish between positive and negative cases and hence evaluates the predictive accuracy of the model. The ROC curve that is closest to the upper-left corner of Figure 13 is the one with the best predictive performance. The combined model (AUC=0.85) has a better predictive performance than the other two. Additionally, this model explains more variation ($R^2=0.41$) in the response than the other two models and has a lower cross validation error rate ($E=0.1433$). The AIC of the combined model is (AIC=302.11) which is much lower than the satellite (AIC=318.81) and the land cover (AIC=322.17) models. Based on these facts the combined model was chosen as the final model.

Chapter 5 Discussion

This chapter will discuss in detail the results obtained. The satellite model will be discussed in the first section, followed by the land cover model and the final combined model. The fourth section talks about the importance of selecting viable predictor variables. The last section compares the final model from this study and the final corn bunting probability map from the Catalan breeding bird atlas.

5.1 Satellite imagery

Amongst the satellite variables, land surface temperature (LST: $p=0.00007$) had the strongest influence on the corn bunting because of the variable's ability in discriminating the thermal signature of dry, non-irrigated arable land. Additionally, intensified agricultural fields exhibit low temperature in summer breeding months due to heavy irrigation; therefore, LST has potential, in dry environments such as Lerida, to discriminate favorable habitats from non-favorable ones for species such as the corn bunting.

The mean value of the near infrared band 4 (band4m: $p=0.0056$) and the coefficient of variation (CV) of the mid-infrared band 5 (band5cv: $p=0.0024$) exhibited a strong positive correlation and high significance in describing the corn bunting occurrence. Band 4 is responsive to photosynthetically active vegetation and the quantity of biomass while band 5 is responsive to vegetation moisture content (St-Louis et al. 2009). This suggests that texture features in the infrared region are likely to detect variation in vegetation structure.

An interesting result was the relationship of the corn bunting with the standard deviation of band 1 (band1sd: $p=0.097$), removal of this variable increased both the residual deviance and AIC score. The bunting had a negative relationship with band1sd because the spectral range of band 1 ($0.45-0.52\mu\text{m}$) is ideal for detecting urban and man-made features. However a surprising outcome was the exclusion of NDVI from the final model due to its insignificance ($p=0.799$), one possible reason can be attributed to the overlap in information between NDVI, band 4 and MSAVI.

Several satellite variables were excluded from the final models because of the high level of multicollinearity between them. One approach that might address this inconvenience would be to use groups of satellite derivatives in separate models as it would reduce the correlation between different textures of the same band and ensure distinct contribution of each variable.

5.2 Land cover dataset

The CLC2000 dataset, despite (or because of) being public domain, has a couple of disadvantages. For starters it takes several years to produce one country-wide (and indeed Europe-wide) CLC2000 dataset as only three (1990, 2000, and 2006) have been produced in the last 20 years. And secondly, because of the low resolution, CORINE does not discriminate between the differences in vegetation structure. These are the areas where satellite imagery outperforms land cover data due to the high temporal and spatial resolution of available satellites.

Although the creation of reliable land cover datasets is both time and effort (e.g. ground-truthing) consuming, the lure of new, more efficient classification algorithms, expert knowledge, in-field verification make them a promising products in identifying species' habitat requirements. In order to be effective, they need to be produced on a yearly basis so that temporal variations in species' habitat preference could be recorded.

5.2.1 Non-irrigated arable land

There was a distinctive preference for non-irrigated arable land (NIAL: $p=2.31E-008$) landscape metric which is corroborated by earlier research (Diaz and Telleria 1997; Stoate et al. 2000; Brambilla et al. 2009). The exclusion of NIAL had the greatest effect on the model, increasing the AIC by an average of 43.56 and the deviance by an average of 41.56 in the stepwise variable selection process.

5.2.2 Permanently irrigated land

Permanently irrigated land metric (PIL: $p=0.0002$) is an interesting category because it is a relatively new phenomenon in Cataluña (Brotos et al. 2004a; Moreno-Mateos et al.

2009). Compared to NIAL, there was reduced preference for areas that were predominantly comprised of this habitat. The land cover map (and the combined map) shows that there is increased preference for grassy fringes where permanently irrigated land meets other more favorable habitats such as non-irrigated arable land.

Water used on permanently irrigated land often comes from wetlands that are eventually drained. This water eventually drains down to lower parts of valleys to form other wetlands (Moreno-Mateos et al. 2009) but due to the agricultural intensification process there is a possibility that this runoff carries components of pesticides (Matson et al. 1997; Firbank et al. 2008).

5.3 Final model

Selection of the “best” model embodies an understanding of the phenomena that influence the behavior of species. The final model must be one that is the most parsimonious and biologically reasonable to describe the relationship between the response and the predictor variables (Hosmer and Lemeshow 2000). The combined model is comprised of information from both the satellite variables and the land cover data and is the statistical model of choice in indentifying the breeding habitat selection of the corn bunting. The model produced the least unexplained variation (RD=276.11) in the response and the highest predictive accuracy (AUC=0.84) amongst all the models in this study. It displays the effect of multiple factors that are not limited by data source on the habitat selection behavior of the corn bunting. For example, the selection of MSAVI shows the effect of soil background is important factor in habitat selection of the corn bunting. Increased reflectance from the underlying soil can be caused by anthropogenic effects such as livestock grazing, mowing and periodic burning.

The final model was also able to explain more variation in the response ($R^2=41\%$) than the other two model. The proportion of explained variance is small because measures such as R^2 rely on the extent and distribution of the predictors. They tend exhibit low values in logistic regression even if the regression displays a perfect relationship (Cox and Wermuth 1992).

Note that the saturated model containing all 39 predictor variables (Table 9) neither adheres to the principle of parsimony nor is biologically interpretable. It also contains many parameters that are statistically insignificant.

The final map shows that the corn bunting avoids habitats that are comprised of steep slopes, near human activity and urban infrastructure and areas comprised wholly of intensely irrigated landscapes. The map also shows that the corn bunting favors habitats that are comprised of non-irrigated arable land, close to areas where vegetation moisture content is high, and dry, open areas near grassland and away from dense cover.

5.4 Variable selection

Depending on the objectives of the research, it is important to select variables that are of biological or ecological importance with regards to the response. On the one hand, avoiding errors caused by subjective land cover classification allows for the use of the full range of information contained in the satellite imagery (Laurent et al. 2005) and the creation of indices that are biologically relative to species. On the other hand, the advantage of the inclusion of land cover data lies in its ability to produce spatial metrics that can help explain how landscapes influence the most important ecological processes (Carrao and Caetano 2002). An optimal model would be one that produced the highest predictive performance. Therefore, the best modeling approach would be to combine the predictive variables extracted from both satellite and land cover sources.

5.5 Comparison to the CBBA map

Comparison of this study's final map with the map produced by the CBBA reveals a general similarity (Figure 14). There is an overall agreement of the corn bunting's preference for NIAL and the evasion of areas that are built up and where human activity is high. However, the map produced in this study exhibits more detail because of the CBBA's broad scale (UTM 1x1 km) and owing to the spatial resolution of the Landsat sensor (30m) and the CLC2000 used here. Although the CBBA included topographic details into their analysis, their final map does not display in enough detail the bunting's avoidance of areas with high

slope. This is probably because the CBBA used the mean value of the slope in each UTM square which considerably dilutes the amount of information in the digital elevation model.

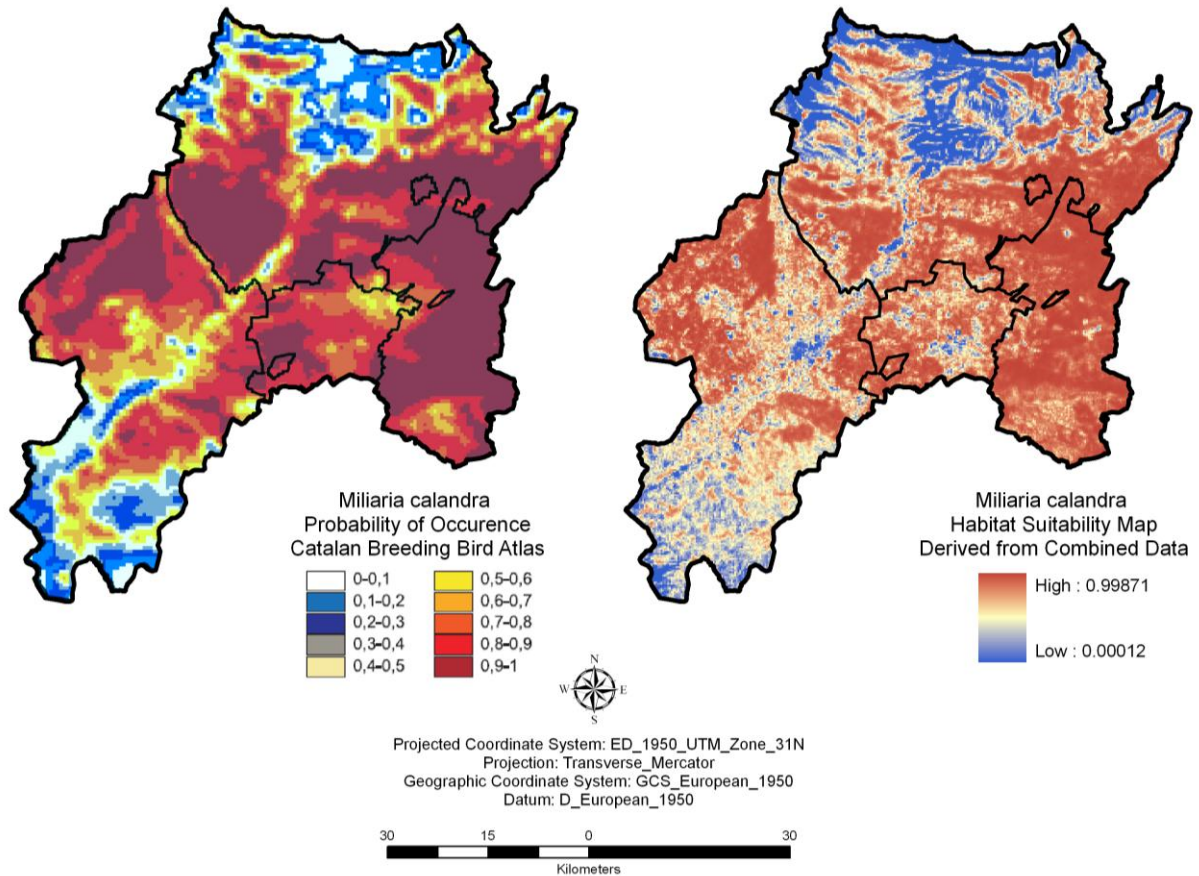


Figure 14: Comparison between map produced by the Catalan breeding bird atlas and the final map produced in this study.

Chapter 6 Conclusions and Recommendations

The decline in the breeding populations of European farmland birds is a witness phenomenon to the impact of humans on the biodiversity of agricultural systems. Traditional, low intensity farming has been abandoned in favor of intensified, high yield farming supported by the Common Agricultural Policy of the European Union (Donald et al. 2001; Donald et al. 2002). These changes in the agricultural landscape have resulted in a continuous decreasing trend in the breeding numbers of farmland bird species. Predictive distribution modeling of species of concern is important in order to assess the significance of habitats from a conservation perspective. Therefore there is a need to monitor this decline using tools that are accurate, expedient and practical. The geospatial tools of remote sensing and GIS address this need by monitoring processes which influence species both directly and indirectly.

The objectives of this study, which were based on a set of research questions, were all fulfilled. Based on the results of this study, the first and second research question can be comfortably answered in the affirmative. The third question of this study aimed at assessing the predictive performance of variables derived from the satellite imagery and land cover data. It has been shown that predictor variables extracted from satellite imagery such as satellite image texture and vegetation indices were able to produce a habitat preference map for the corn bunting that had an 81% predictive accuracy based on the AUC value. Similarly, landscape metrics and distance variables extracted from the CLC2000 dataset were also able to produce a map that had 81% predictive accuracy. However, the satellite model had a slightly lower residual deviance (318.81 vs. 322.17).

Regarding the fourth research question, the combined model performed better (AUC=0.85) than both the land cover and the satellite model. As for the fifth question, this study reinforces the conclusion of Seoane et al (2004a) that the selection of predictor variables should be based on the grounds of data availability and that the best predictive accuracy is obtained when combining spectral and thematic data.

Variables selected for this study were derived directly from public domain satellite imagery and land cover data and could serve as substitutes that assess habitat suitability and/or the availability of food. Research into the use of proxies for food conditions in predicting the occurrence and density of bird species has been studied before (Pebesma et al. 2005), however, comparative research on the potential of publicly available data to act as surrogates is lacking. Saveraid et al (2001) proposed that the use of satellite data alone is not sufficient in modeling bird distribution and that habitat structure variables are also necessary. Landscape metrics are compositional quantifications extracted from CLC2000 that can describe the structure of a landscape and thus provide a number of potential predictor variables.

The final logistic regression model had a predictive accuracy of 85% based on the AUC. The corn bunting had a strong positive correlation with the modified soil adjusted vegetation index, the coefficient of variation image texture of band 5 and the non-irrigated arable land landscape metric. Each of these parameters serves as an ecological surrogate in the modeling process. It must be noted that the resultant models are only applicable for the scale, spatial and temporal resolution in which they have been developed. This approach does not allow dependence of the response to vary spatially. The spatial coverage of the predictions must not include areas that are beyond the environmental space of the data used to build the model.

This study has shown that the combination of public data from different sources is a viable method in producing models that reflect species' habitat preference. The development of maps that are comprised of information from both satellites and land cover datasets are of importance for species that have indeterminate ranges (Seoane et al. 2004b) and for monitoring the spatial dynamics of protected areas. This not only aids in the identification and maintenance of important habitats as the Bird Directive has stipulated but would also identify trends in bird numbers.

However, some important ecological and methodological aspects may have been missed; even though this study had its core focus on free land cover and satellite data, there are certain areas where research could be furthered:

1. The addition of climatic variables may enhance the model further by quantifying the effect of precipitation and daily temperature on site-selection behavior of the species. Climatic variables will also aid in the study of the effects of climate change on species.
2. The use of the enhanced vegetation index (EVI) could also significantly aid species' habitat modeling because of improved sensitivity over other vegetation indices and its ability to correct both for atmospheric influences and ground reflectance (Jiang et al. 2008).
3. Inclusion of the intensification quantifications (agricultural yields, pesticide use, amount of water used for irrigation, hectares of monocultures, etc.) in farmland bird distribution models would be the next step in the research on the decline of farmland bird species. This enables direct correlation between the levels of farm intensification and breeding bird diversity.
4. Recreation of the models by including spatial effects that allows the dependence of the response on the predictors to fluctuate spatially as proposed by Foody (2005).
5. The use of presence-only and background environmental data to model the distribution. This technique centers on the ecological relationship between locations where species are recorded and the rest of the study area.
6. Use the same dataset to create models from different statistical methods such as generalized additive models, classification and regression trees, generalized boosting models, niche-based models, etc. The BIOMOD package (Thuiller et al. 2009) provides such approaches.
7. Using a multi-scale approach would allow a more in-depth analysis because birds might choose habitats at different scales depending on the size of the breeding territory (Graf et al. 2005).

It is hoped that this study encourages the development of habitat models using data from the public domain as a cost-efficient and practical alternative to expensive data.

References

1. Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267-281). Akademiai Kiado, Budapest, Hungary
2. Alados-Arboledas, L., Olmo, F.J., Alados, I., & Perez, M. (2000). Parametric models to estimate photosynthetically active radiation in Spain. *Agricultural and Forest Meteorology*, 101, 187-201
3. Baraldi, A., & Parmiggiani, F. (1995). An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 293-304
4. Bellis, L.M., Pidgeon, A.M., Radeloff, V.C., St-Louis, V., Navarro, J.L., & Martella, M.B. (2008). Modeling Habitat Suitability for Greater Rheas based on Satellite Image Texture. *Ecological Applications*, 18, 1956-1966
5. Benton, T.G., Vickery, J.A., & Wilson, J.D. (2003). Farmland biodiversity: is habitat heterogeneity the key? *TRENDS in Ecology and Evolution*, 18, 182-188
6. Birdlife-International (2004). Birds in the European Union: a status assessment. In Wageningen, The Netherlands: Birdlife International
7. Bivand, R.S., Pebesma, E.J., & Gomez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. New York, NY: Springer
8. Brambilla, M., Guidali, F., & Negri, I. (2009). Breeding-season habitat associations of the declining Corn Bunting *Emberiza calandra* - a potential indicator of the overall bunting richness. *Ornis Fennica*, 41-50
9. Brauner, N., & Shacham, M. (1998). Role of range and precision of the independent variable in regression of data. *American Institute Of Chemical Engineers Journal*, 603-611
10. Brickle, N., Harper, D., Aebischer, N., & Cockayne, S. (2000). Effects of agricultural intensification on the breeding success of corn buntings *Miliaria calandra*. *Journal of Applied Ecology*, 742-755
11. Brotons, L., Herrando, S., Estrada, J., Pedrocchi, V., & Martin, J.L. (2008). The Catalan Breeding Bird Atlas (CBBA): methodological aspects and ecological implications. *Revista Catalana d'Ornitologia*, 118-137
12. Brotons, L., Manosa, S., & Estrada, J. (2004a). Modelling the effects of irrigation schemes on the distribution of steppe birds in Mediterranean farmland. *Biodiversity and Conservation*, 1039-1058
13. Brotons, L., Thuiller, W., Araujo, M.B., & Hirzel, A.H. (2004b). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 437-448
14. Buchanan, G., Pearce-Higgins, J., Grant, M., Robertson, D., & Waterhouse, T. (2005). Characterization of moorland vegetation and the prediction of bird abundance using remote sensing. *Journal of Biogeography*, 697-707

15. Burnham, K.P., & Anderson, D.R. (1998). *Model Selection and Inference: A Practical Information-Theoretical Approach*. New York: Springer-Verlag
16. Burnham, K.P., & Anderson, D.R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods Research*, 33, 261-304
17. Cabral, P., Gilg, J.-P., & Painho, M. (2005). Monitoring urban growth using remote sensing, GIS, and spatial metrics. In W. Gao (Ed.), *SPIE Optics & Photonics: Remote sensing and modeling of ecosystems for sustainability*. San Diego, CA: SPIE
18. Canty, A., & Ripley, B. (2009). boot: Bootstrap R (S-Plus) Functions. In
19. Carrao, H., & Caetano, M. (2002). The Effect of Scale on Landscape Metrics. In, *International Symposium of Remote Sensing of the Environment*. Buenos Aires
20. Coreau, A., & Martin, J.-L. (2007). Multi-scale study of bird species distribution and of their response to vegetation change: a Mediterranean example. *Landscape Ecology*, 747-764
21. Cox D.R., & Wermuth N. (1992). A comment on the coefficient of determination for binary responses. *American Statistician*, 46, 1-4.
22. Crist, E.P. (1983). The Thematic Mapper Tasseled Cap - A preliminary formulation. In, *Ninth International Symposium on Machine Processing of Remotely Sensed Data*. Purdue University, West Lafayette, IN, USA: IEEE
23. Crist, E.P., & Kauth, R.J. (1986). The tasseled cap demystified. *Photogrammetric Engineering and Remote Sensing*, 81-86
24. Deceuninck, B. (1998). The Corncrake (*Crex crex*) in France. In N. Schaeffer & U. Mammen (Eds.), *International Corncrake Workshop*. Hilpoltstein, Germany
25. Deleo, J.M. (1993). Receiver operating characteristic laboratory (ROCLAB): software for developing decision strategies that account for uncertainty. In, *Second International Symposium on Uncertainty Modelling and Analysis* (pp. 318-325). College Park, MD: IEEE Computer Society Press
26. Diaz, M., & Telleria, J.L. (1997). Habitat selection and distribution trends of corn buntings in the Iberian Peninsula. In P. Donald & N.J. Aebischer (Eds.), *The Ecology and Conservation of Corn Buntings *Miliaria calandra**. (pp. 151-161). Peterborough: JNCC
27. Donald, P.F., Green, R.E., & Heath, M.F. (2001). Agricultural intensification and the collapse of Europe's farmland bird populations. *Proceedings of the Royal Society B*, 25-29
28. Donald, P.F., Pisano, G., Rayment, M.D., & Pain, D.J. (2002). The Common Agricultural Policy, EU enlargement and the conservation of Europe's farmland birds. *Agriculture, Ecosystems and Environment*, 167-182
29. EEC, T.C.o.E.C. (1979). Council Directive 79/409/EEC of 2 April 1979 on the conservation of wild birds. In E.E. Community (Ed.), 409. Brussels
30. Erickson, W.P., Nielson, R., Skinner, R., Skinner, B., & Johnson, J. (2004). Applications of Resource Selection Modeling Using Unclassified Landsat Thematic Mapper Imagery. In S. Huzubazar (Ed.), *1st International Conference on Resource Selection* (pp. 130-140). Laramie, Wyoming: Weston EcoSystems Technology, Inc., Cheyenne, Wyoming, USA

31. Firbank, L.G., Petit, S., Smart, S., Blain, A., & Fuller, R.J. (2008). Assessing the impacts of agricultural intensification on biodiversity: a British perspective. *Philosophical Transactions of the Royal Society B*, 777-787
32. Foody, G.M. (2005). Mapping the richness and composition of British breeding birds from coarse spatial resolution satellite sensor imagery. *International Journal of Remote Sensing*, 26, 3943-3956
33. Fox, J. (2005). The R Commander: A Basic-Statistics Graphical User Interface to R. *Journal of Statistical Software*, 14, 1-42
34. Freeman, E.A., & Moisen, G. (2008). PresenceAbsence: An R Package for Presence Absence Analysis. *Journal of Statistical Software*, 23, 1-31
35. Fuller, R.M., Devereux, B.J., Gillings, S., Amable, G.S., & Hill, R.A. (2005). Indices of bird-habitat preference from field surveys of birds and remote sensing of land cover: a study of south-eastern England with wider implications for conservation and biodiversity assessment. *Global Ecology and Biogeography*, 14, 223-239
36. Gottschalk, T.K., Huettmann, F., & Ehlers, M. (2005). Thirty years of analysing and modelling avian habitat relationships using satellite imagery data: a review. *International Journal of Remote Sensing*, 26, 2631-1656
37. Graf, R.F., Bollmann, K., Suter, W., & Bugmann, H. (2005). The importance of spatial scale in habitat models: capercaillie in the Swiss Alps. *Landscape Ecology*, 703-717
38. Gregory, R.D., van Strien, A., Vorisek, P., Gmelig Meyling, A.W., Noble, D.G., Foppen, R.P.B., & Gibbons, D.W. (2005). Developing indicators for European birds. *Philosophical Transactions of the Royal Society B*, 269-288
39. Griffiths, G.H., Lee, J., & Eversham, B.C. (2000). Landscape pattern and species richness; regional scale analysis from remote sensing. *International Journal of Remote Sensing*, 21, 2685-2704
40. Guisan, A., Thomas C. Edwards, J., & Hastie, T. (2002). Generalized linear models and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 89-100
41. Hale, S.R. (2006). Using Satellite Imagery to Model Distribution and Abundance of Bicknell's Thrush (*Catharus bicknelli*) in New Hampshire's White Mountains. *The Auk*, 123, 1038-1051
42. Harrell, F.E. (2009). Design: Design Package. . In
43. Hirzel, A.H., Hausser, J., Chessel, D., & Perrin, N. (2002). Ecological-niche factor analysis: How to compute habitat-suitability map without absence data. *Ecology*, 83, 2027-2036
44. Ho, R. (2006). *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*. Boca Raton: Taylor and Francis Group
45. Hosmer, D.W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley and Sons, Inc.
46. Hu, B., Shao, J., & Palta, M. (2006). Pseudo-R² in logistic regression model. *Statistica Sinica*, 16, 847-860
47. Huang, C., Wylie, B., Yang, L., Homer, C., & Zylstra, G. (2002). Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. *International Journal of Remote Sensing*, 23, 1741-1748

48. ITTVIS (2009). Atmospheric Correction Module: QUAC and FLAASH User's Guide. In
49. Jiang, Z., Huete, A.R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112, 3833-3845
50. Jobin, B., Grenier, M., & Laporte, P. (2005). Using satellite imagery to assess breeding habitat availability of the endangered loggerhead shrike in Quebec. *Biodiversity and Conservation*, 81-95
51. Keitt, T.H., Bivand, R., Pebesma, E., & Rowlingson, B. (2009). rgdal: Bindings for the Geospatial Data Abstraction Library. In
52. Kleinbaum, D.G., & Klein, M. (2002). *Logistic Regression: A Self-Learning Text*. New York: Springer
53. Laurent, E., Shi, H., Gatzliolis, D., LeBouton, J., Walters, M., & Liu, J. (2005). Using the spatial and spectral precision of satellite imagery to predict wildlife occurrence patterns. *Remote Sensing of Environment*, 249-262
54. Liang, S. (2004). *Quantitative Remote Sensing of Land Surfaces*. Hoboken, New Jersey: John Wiley & Sons, Inc.
55. Luoto, M., Virkkala, R., Heikkinen, R.K., & Rainio, K. (2004). Predicting Bird Species Richness Using Remote Sensing in Boreal Agricultural-Forest Mosaics. *Ecological Applications*, 14, 1946-1962
56. Matson, P.A., Parton, W.J., Power, A.G., & Swift, M.J. (1997). Agricultural Intensification and Ecosystem Properties. *Science*, 277, 504-509
57. Menard, S. (2002). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage Publications
58. Moreno-Mateos, D., Pedrocchi, C., & Comin, F.A. (2009). Avian communities' presence in recently created agricultural wetlands in irrigated landscapes of semi-arid areas. *Biodiversity and Conservation*, 811-828
59. Nagendra, H. (2001). Using remote sensing to assess biodiversity. *International Journal of Remote Sensing*, 22, 2377-2400
60. Nohr, H., & Jorgensen, A.F. (1997). Mapping of biological diversity in Sahel by means of satellite image analyses and ornithological surveys. *Biodiversity and Conservation*, 545-566
61. Norris, K., & Pain, D.J. (Eds.) (2002). *Conserving Bird Biodiversity: General principles and their application*. Cambridge, UK: Cambridge University Press
62. Orłowski, G. (2005). Endangered and declining bird species of abandoned farmland in south-western Poland. *Agriculture, Ecosystems and Environment*, 231-236
63. Pearce, J., & Ferrier, S. (2000a). An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, 127-147
64. Pearce, J., & Ferrier, S. (2000b). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, 225-245
65. Pebesma, E.J., & Bivand, R.S. (2005). Classes and methods for spatial data in R. *R News*, 5, 9-13

66. Pebesma, E.J., Duin, R.N.M., & Burrough, P.A. (2005). Mapping sea bird densities over the North Sea: spatially aggregated estimates and temporal changes. *Environmetrics*, 573-587
67. Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.-M., Tucker, C.J., & Stenseth, N.C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *TRENDS in Ecology and Evolution*, 20, 503-510
68. Phillips, S.J., Anderson, R.P., & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 231-259
69. Ponjoan, A., Bota, G., Morena, E.L.G.D.L., Morales, M.B., Wolff, A., Marco, I., & Manosa, S. (2008). Adverse Effects of Capture and Handling Little Bustard. *Journal of Wildlife Management*, 72, 315-319
70. Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48, 119-126
71. R Development Core Team. (2009). R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing
72. Ranganathan, J., Chan, K.M.A., & Daily, G.C. (2007). Satellite Detection of Bird Communities in Tropical Countryside. *Ecological Applications*, 17, 1499-1510
73. Sanz, J.J., Potti, J., Moreno, J., Merino, S., & Frias, O. (2003). Climate change and fitness components of a migratory bird breeding in the Mediterranean region. *Global Change Biology*, 461-472
74. Saveraid, E.H., Debinski, D.M., Kindscher, K., & Jakubauskas, M.E. (2001). A comparison of satellite data and landscape variables in predicting bird species occurrences in the Greater Yellowstone Ecosystem, USA. *Landscape Ecology*, 71-83
75. Senapathi, D., Vogiatzakis, I.N., Jeganathan, P., Gill, J.A., Green, R.E., Bowden, C.G.R., Rahmani, A.R., Pain, D., & Norris, K. (2007). Use of remote sensing to measure change in the extent of habitat for the critically endangered Jerdon's Courser *Rhinoptilus bitorquatus* in India. *Ibis*, 328-337
76. Seoane, J., Bustamante, J., & Diaz-Delgado, R. (2004a). Competing roles for landscape, vegetation, topography and climate in predictive models of bird distribution. *Ecological Modelling*, 209-222
77. Seoane, J., Bustamante, J., & Diaz-Delgado, R. (2004b). Are existing vegetation maps adequate to predict bird distributions? *Ecological Modelling*, 137-149
78. Seto, K.C., Fleishman, E., Fay, J.P., & Betrus, C.J. (2004). Linking spatial patterns of bird and butterfly species richness with Landsat TM derived NDVI. *International Journal of Remote Sensing*, 25, 4309-4324
79. Siriwardena, G.M., Baillie, S.R., Buckland, S.T., Fewster, R.M., Marchant, J.H., & Wilson, J.D. (1998). Trends in the abundance of farmland birds: a quantitative comparison of smoothed Common Birds Census indices. *Journal of Applied Ecology*, 35, 24-43
80. Smith, M.J.d., Goodchild, M.F., & Longley, P.A. (2003). *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Leicester: Matador on behalf of The Winchelsea Press

81. Sobrino, J.A., Jimenez-Munoz, J.C., & Paolini, L. (2004). Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of Environment*, 434-440
82. Sobrino, J.A., Raissouni, N., & Li, Z.-L. (2001). A Comparative Study of Land Surface Emissivity Retrieval from NOAA Data. *Remote Sensing of Environment*, 256-266
83. St-Louis, V., Pidgeon, A., Radeloff, V., Hawbaker, T., & Clayton, M. (2006). High-resolution image texture as a predictor of bird species richness. *Remote Sensing of Environment*, 299-312
84. St-Louis, V., Pidgeon, A.M., Clayton, M.K., Locke, B.A., Bash, D., & Radeloff, V.C. (2009). Satellite image texture and a vegetation index predict avian biodiversity in the Chihuahuan Desert of New Mexico. *Ecography*, 468-480
85. Stoate, C., Borralho, R., & Araujo, M. (2000). Factors affecting corn bunting *Miliaria calandra* abundance in a Portuguese agricultural landscape. *Agriculture, Ecosystems and Environment*, 219-226
86. Stoate, C., Borralho, R., & Araujo, M. (2003). Abundance of Four Lark Species in Relation to Portuguese Farming Systems. *Ornis Hungarica*, 297-301
87. Stockwell, D.R.B., & Peters, D.P. (1999). The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Systems*, 13, 143-158
88. Sundseth, K., & Sylwester, A. (2009). Assessment of similarity between protected and unprotected territory at the NUTS-2 level: Spain Case Study. In, *Towards a green infrastructure for Europe. Integrating Natura 2000 sites into the wider countryside*. Brussels
89. Tankersley, R. (2004). Migration of birds as an indicator of broad-scale environmental condition. *Environmental Monitoring and Assessment*, 55-67
90. Taylor, A.J., & O'Halloran, J. (2002). The Decline of the Corn Bunting *Miliaria calandra*, in the Republic of Ireland. *Biology and Environment: Proceedings of the Royal Irish Academy*, 102B, 165-175
91. Thuiller, W., Lafourcade, B., Engler, R., & Araujo, M.B. (2009). BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, 369-373
92. Tibshirani, R.J., & Tibshirani, R. (2009). A Bias Correction for the Minimum Error Rate in Cross-validation. *The Annals of Applied Statistics*, 3, 822-829
93. Tucker, C.J. (1979). Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment*, 8, 127-150
94. Tucker, G.M., & Heath, M.F. (1994). *Birds in Europe: Their Conservation Status*. Cambridge, UK: BirdLife International
95. USGS (2009). Chapter 11: Landsat 7 Science Data Users Handbook. Retrieved November 18, 2009, from <http://landsathandbook.gsfc.nasa.gov/handbook/>
96. Vallecillo, S., Brotons, L., & Herrando, S. (2008). Assessing the response of open-habitat bird species to landscape changes in Mediterranean mosaics. *Biodiversity and Conservation*, 103-119
97. Wallin, D.O., Elliott, C.C.H., Shugart, H.H., Tucker, C.J., & Wilhelmi, F. (1992). Satellite remote sensing of breeding habitat for an African weaver-bird. *Landscape Ecology*, 7, 87-99

98. Whited, D., Galatowitsch, S., Tester, J.R., Schik, K., Lehtinen, R., & Husveth, J. (2000). The importance of local and regional factors in predicting effective conservation: Planning strategies for wetland bird communities in agricultural and urban landscapes. *Landscape and Urban Planning*, 49, 49-65
99. Young, J.S., & Hutto, R.L. (2002). Use of Regional-scale Exploratory Studies to Determine Bird-habitat Relationships. In J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Hafler, M.G. Raphael, W.A. Wall & F.B. Samson (Eds.), *Predicting species occurrences: issues of accuracy and scale* (pp. 107-119). Washington, DC: Island Press
100. Zeileis, A., & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. *R News*, 2, 7-10

Appendices

Appendix A: Anthropogenic variables

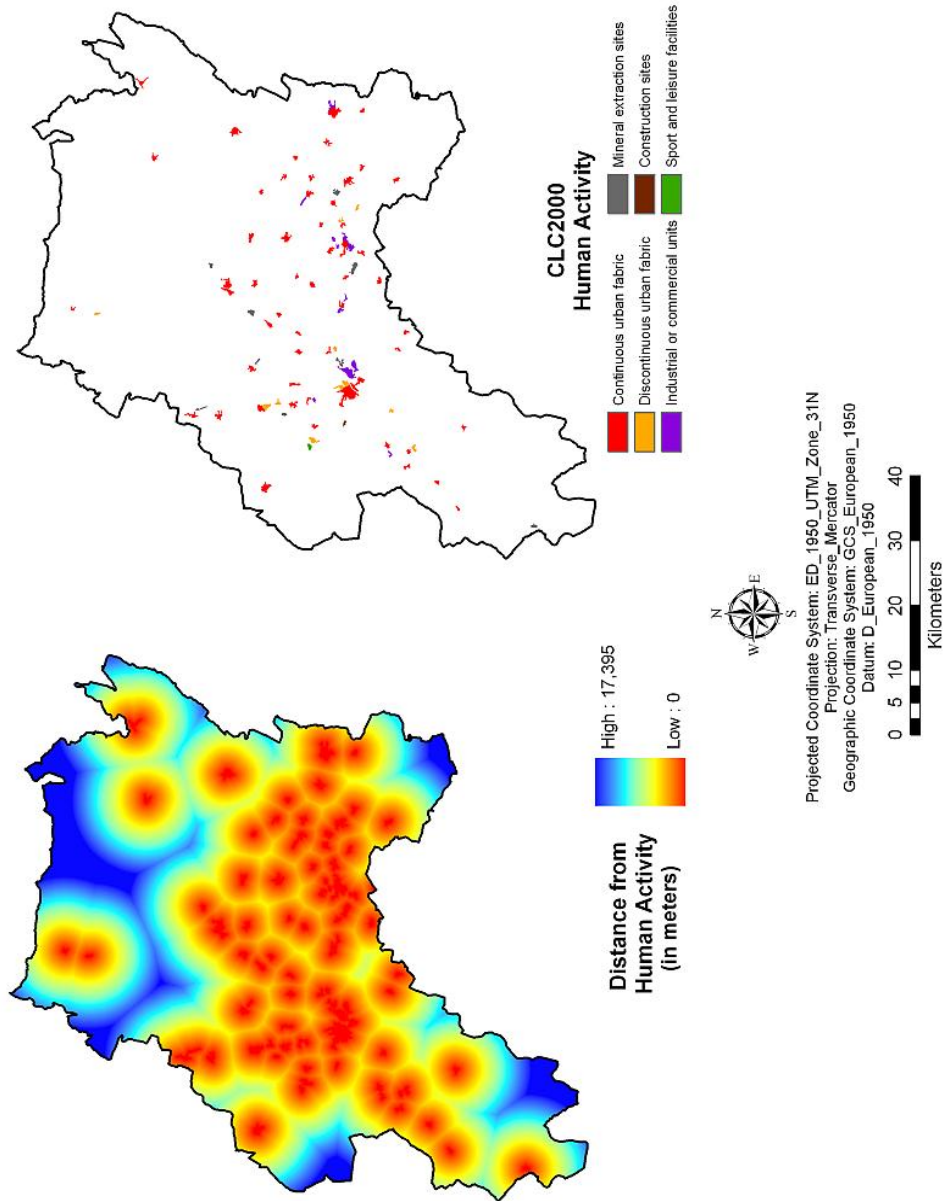


Figure 15: Distance to human activity extracted from CLC2000

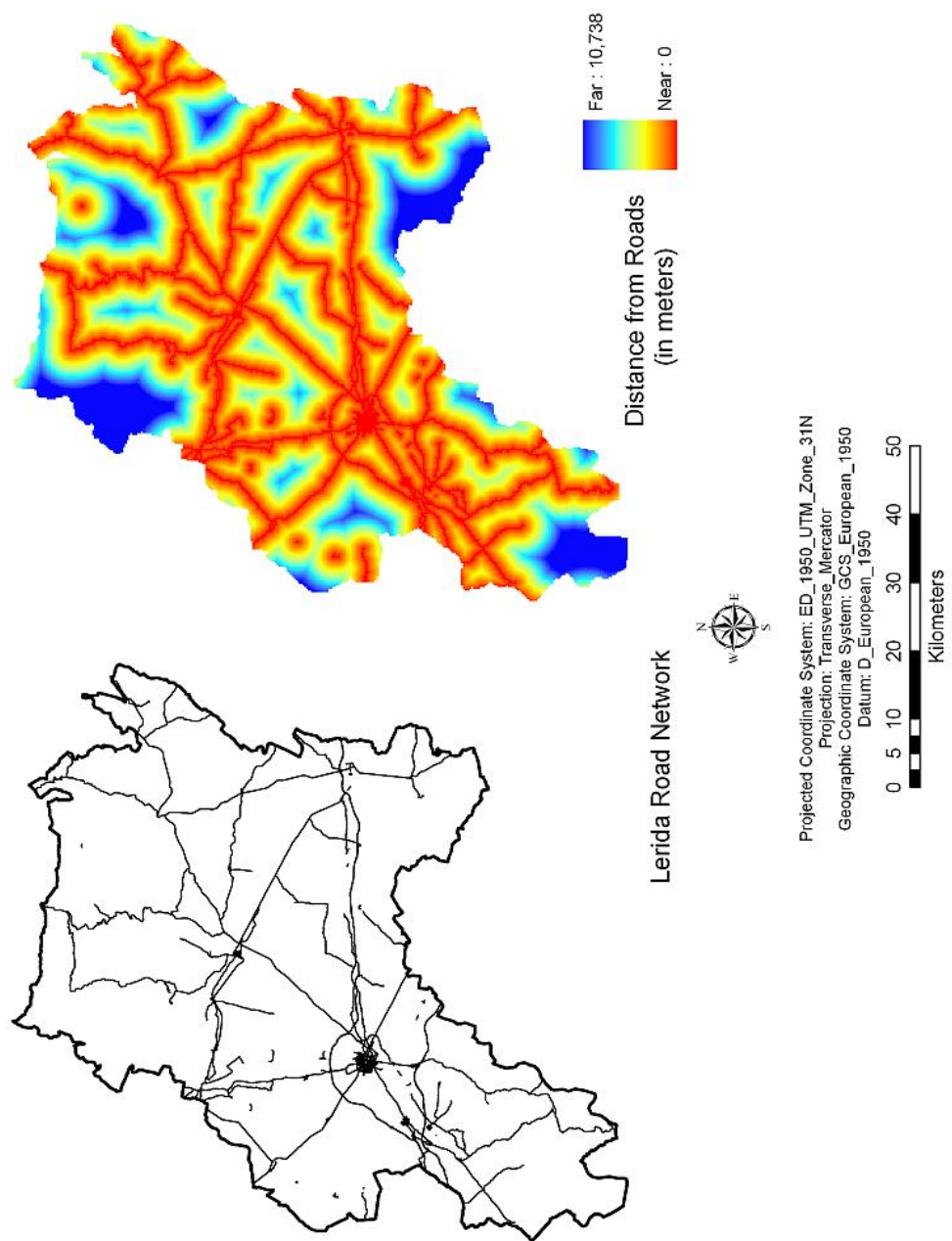


Figure 16: Distance to roads extracted from CLC2000

Appendix B: Descriptive statistics

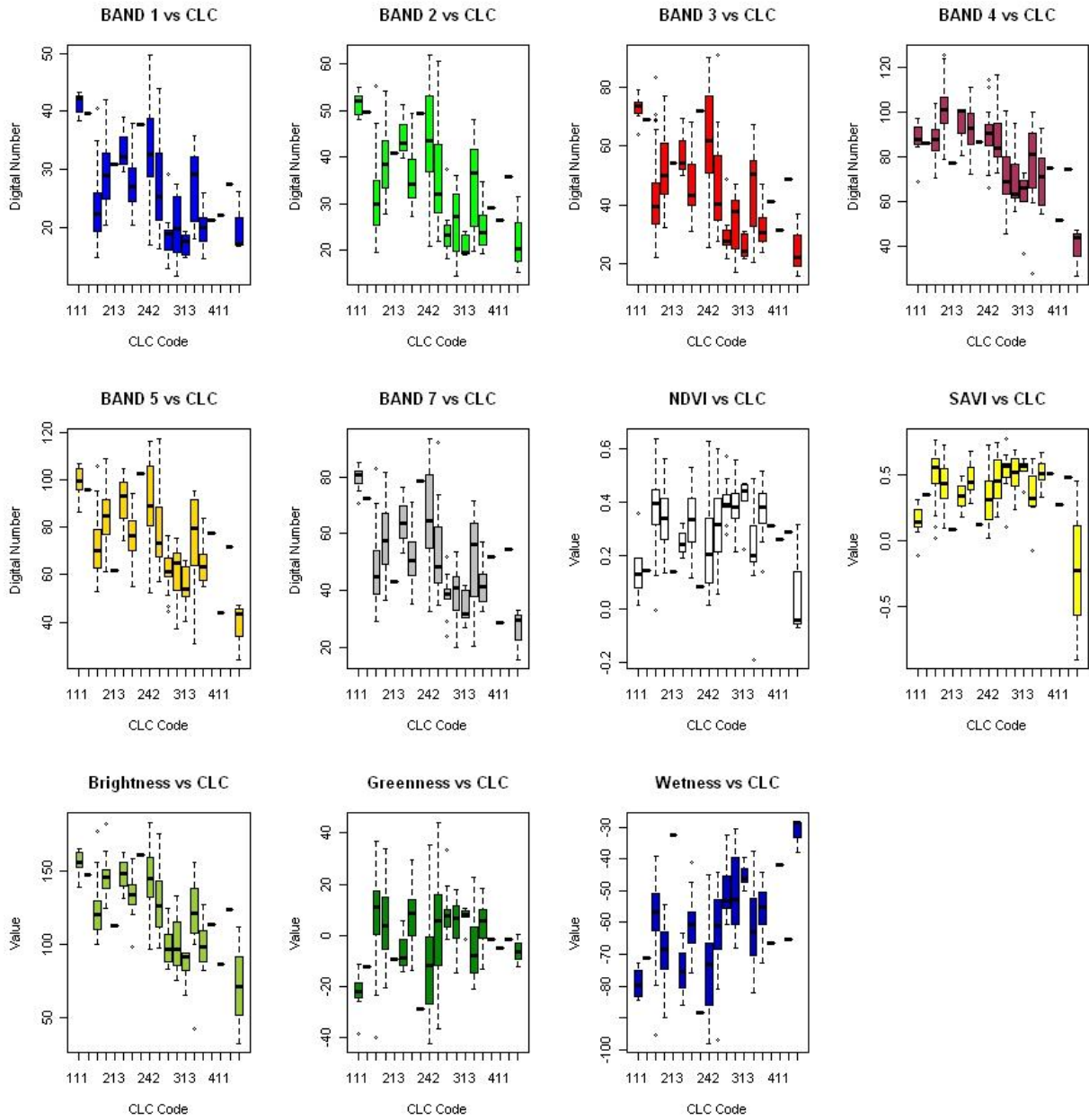


Figure 17: Boxplots of the relationship between selected Landsat derivatives and CLC2000

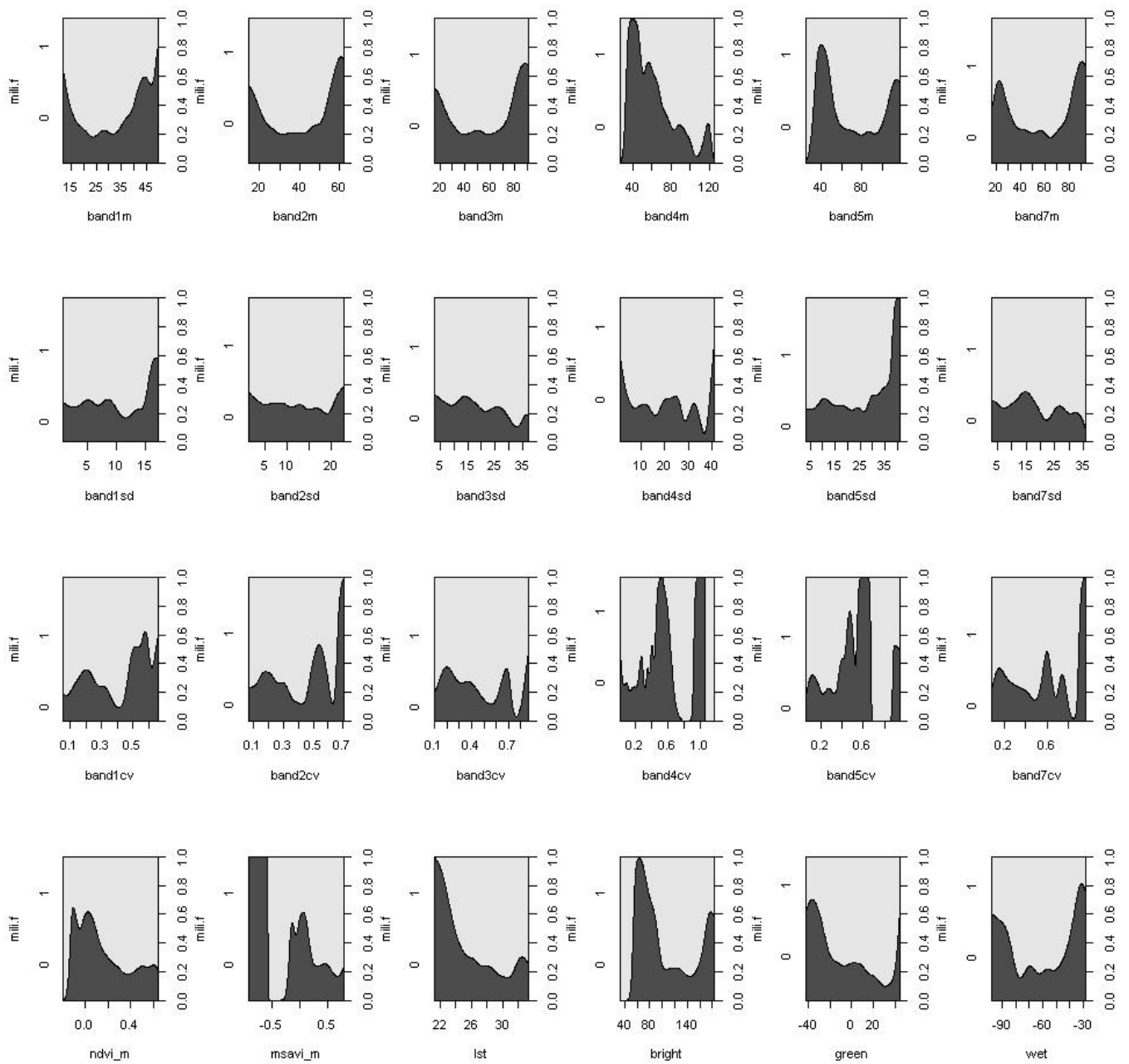


Figure 18: Graphical plots of the association of satellite predictors with the response

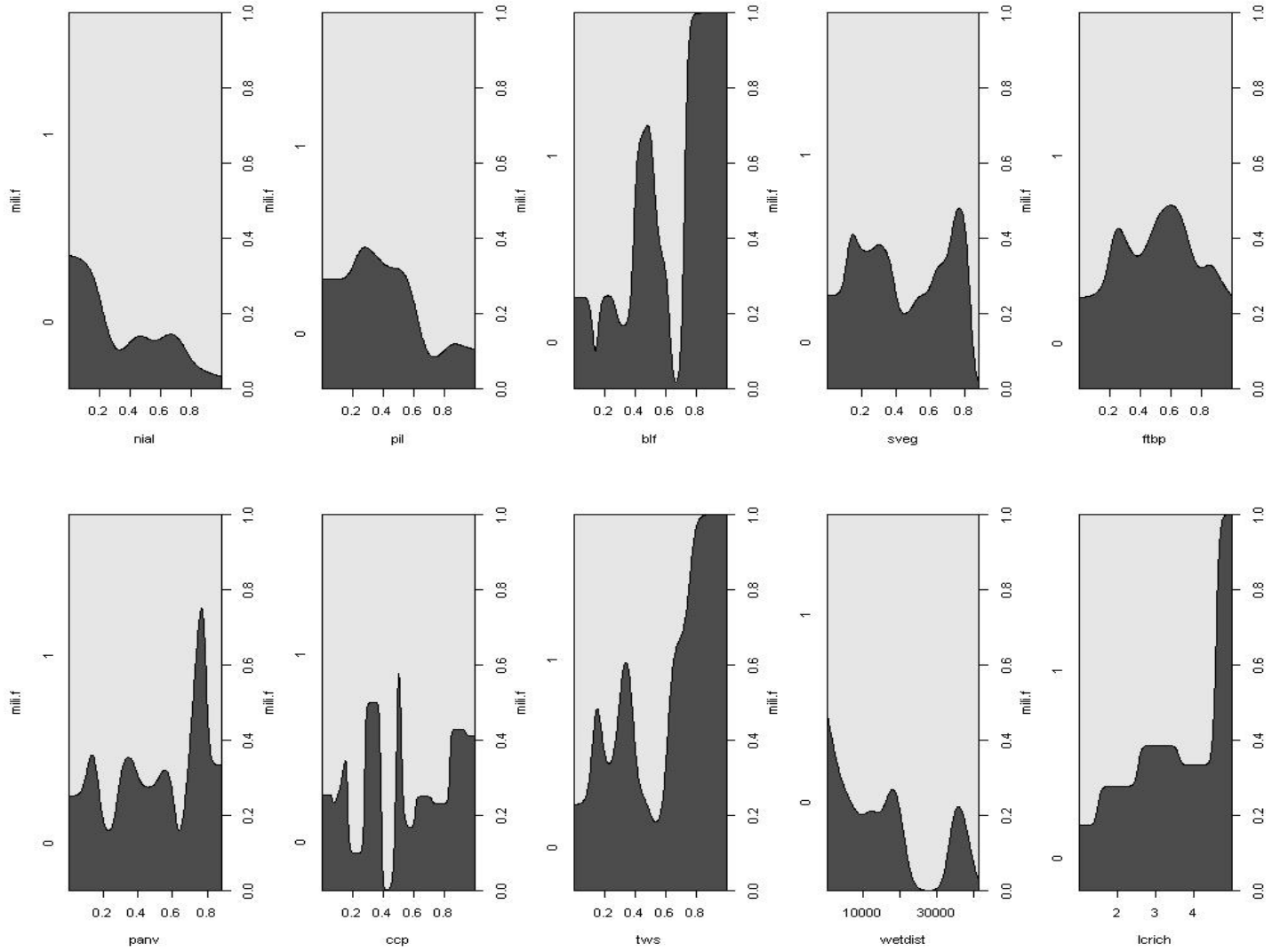


Figure 19: Graphical plots of the association of land cover predictors with the response

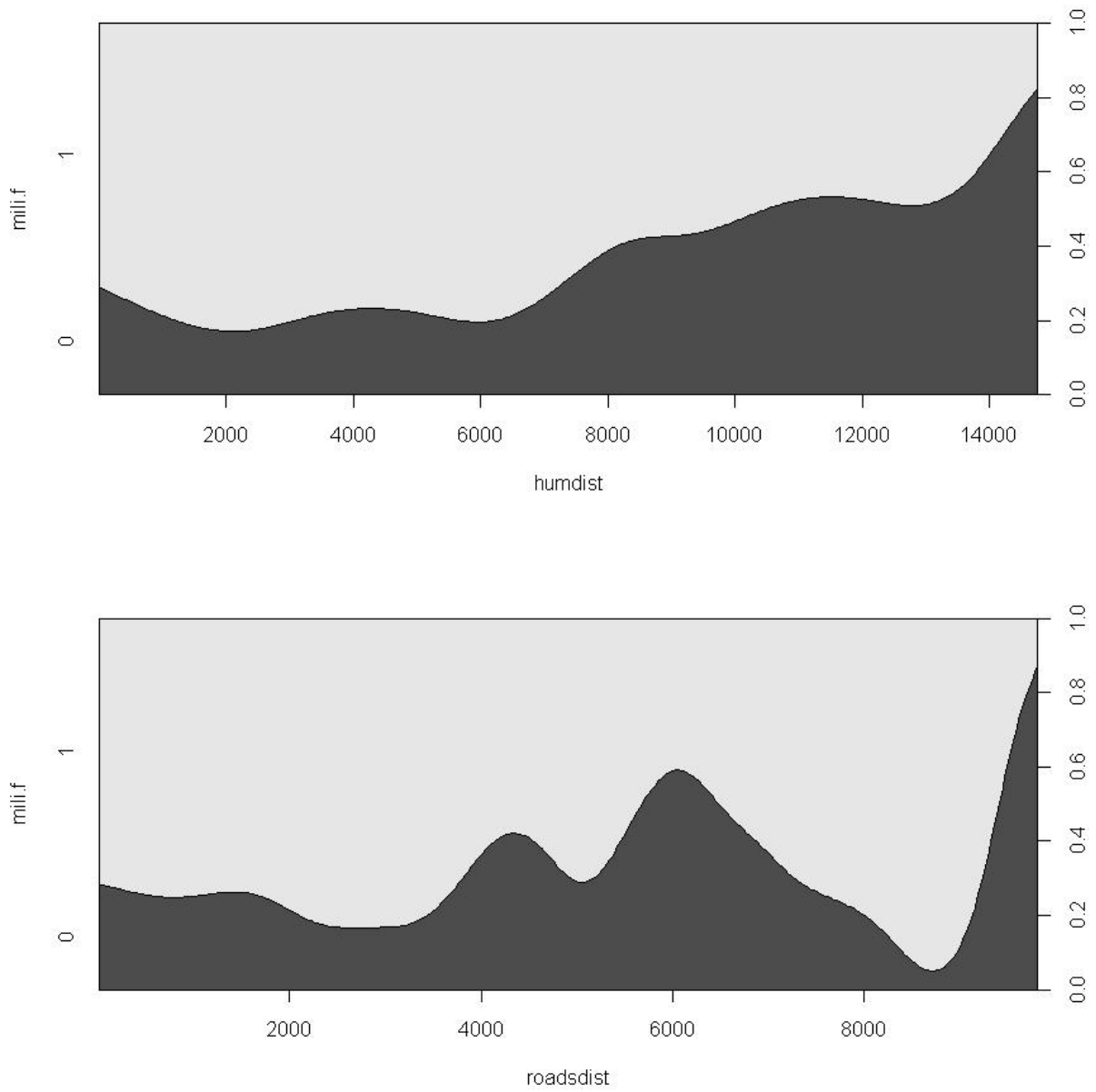


Figure 20: Graphical plots of the association of anthropogenic predictors with the response

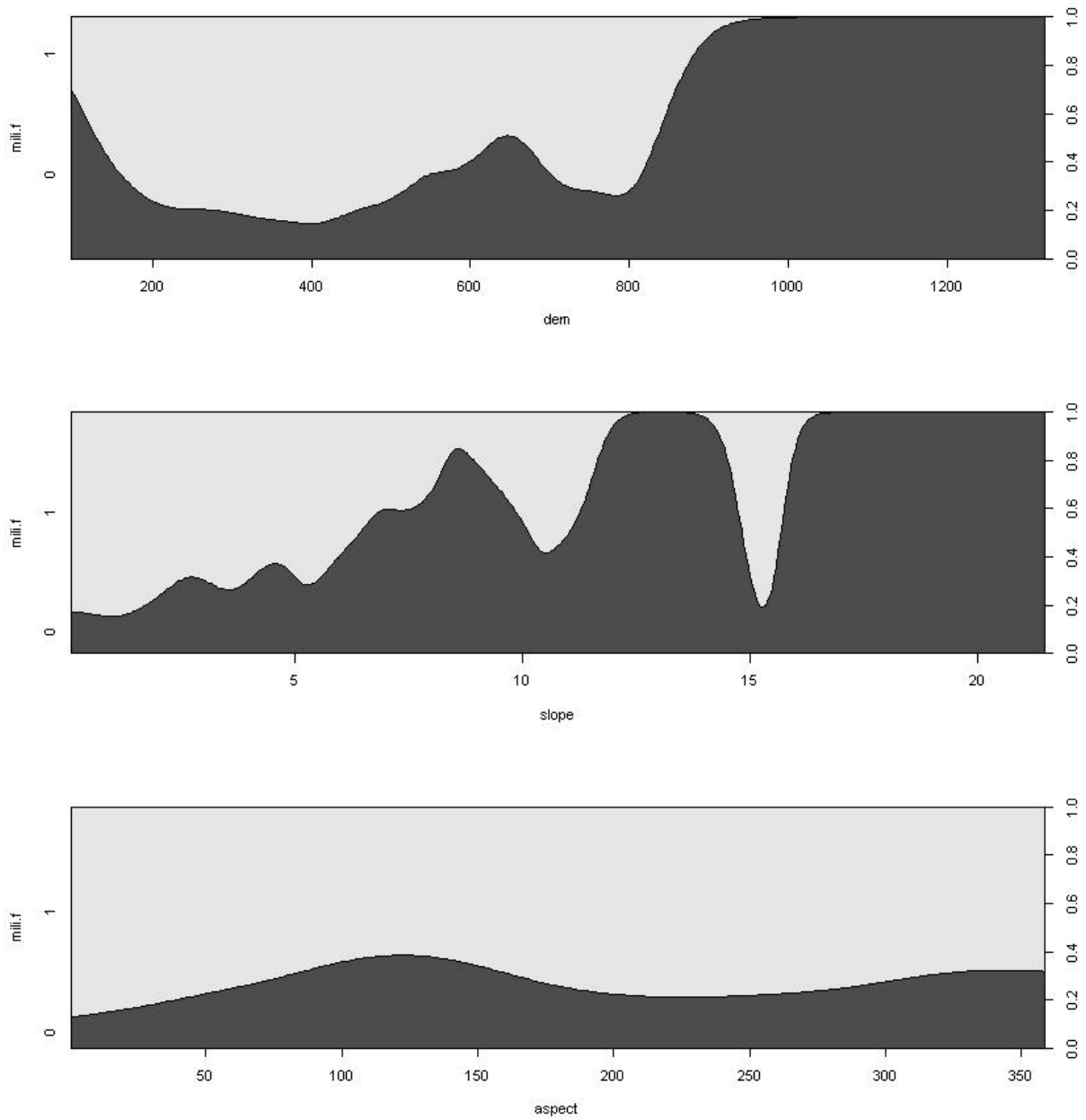


Figure 21: Graphical plots of the association of topographic predictors with the response

Appendix C: Statistical analysis

Table 8: Variance inflation factor values for all the predictor variables

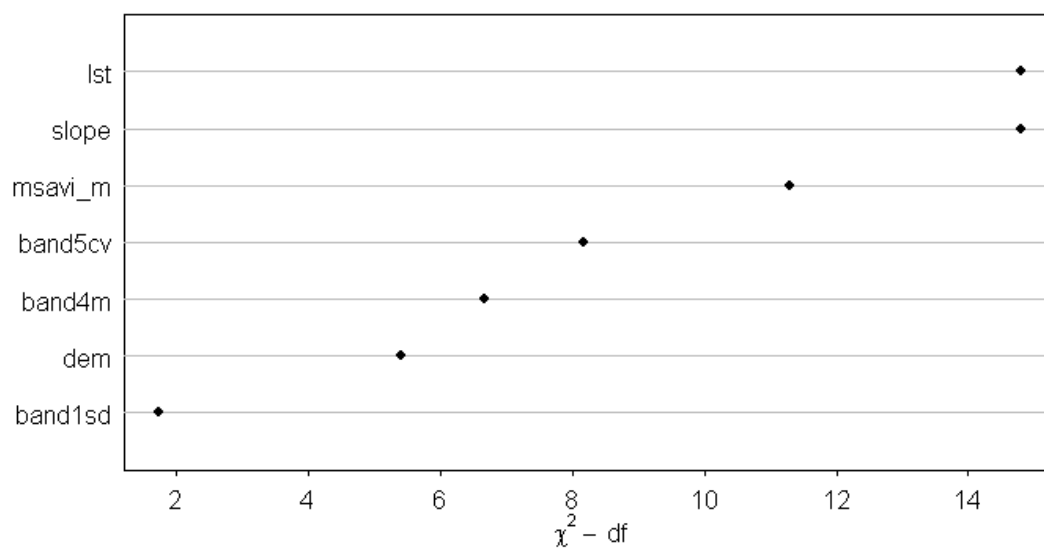
Variable	VIF	Variable	VIF
band2sd	447.14	pil	8.27
band5cv	385.90	msavi_m	7.96
band3sd	346.27	nial	6.75
band7cv	313.02	ccp	4.75
band3m	302.32	dem	4.61
band7sd	301.91	ftbp	3.71
band2m	258.07	slope	3.63
band2cv	228.15	lst	3.50
band5sd	227.54	tws	2.02
band7m	223.35	sveg	1.90
band3cv	186.67	lcrich	1.88
band5m	151.87	panv	1.84
band1sd	138.42	wetdist	1.73
band1cv	96.11	humdist	1.72
band4cv	91.64	blf	1.65
band4sd	39.20	roadsdist	1.36
green	28.23	aspect	1.15
wet	23.83		
band4m	20.76		
bright	20.67		
ndvi_m	11.82		

Table 9: Logistic regression output for the maximal model.

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-18.340	6.163	-2.975	0
band1m	-0.08405	0.35250	-0.238	0.81154
band2m	-0.01809	0.39560	-0.046	0.96353
band3m	0.11890	0.20020	0.594	0.55260
band4m	0.04127	0.05741	0.719	0.47223
band5m	0.26340	0.14390	1.830	0.06725
band7m	-0.33570	0.19200	-1.748	0.08042
ndvi_m	4.20500	4.20900	0.999	0.31772
msavi_m	3.67800	2.70000	1.362	0.17308
band1sd	1.01200	0.87520	1.157	0.24735
band2sd	-1.69700	0.87720	-1.934	0.05306
band3sd	0.58170	0.42060	1.383	0.16667
band4sd	-0.25510	0.16310	-1.563	0.11794
band5sd	0.14930	0.38040	0.392	0.69477
band7sd	-0.08025	0.42750	-0.188	0.85110
band1cv	-38.970	24.370	-1.599	0.10977
band2cv	61.640	29.810	2.068	0.03868
band3cv	-25.300	18.510	-1.367	0.17171
band4cv	27.400	14.640	1.871	0.06129
band5cv	-27.320	30.730	-0.889	0.37394
band7cv	20.690	23.650	0.875	0.38167
dem	0.00356	0.00190	1.873	0.06106
slope	-0.25520	0.10140	-2.517	0.01184
aspect	-0.00363	0.00184	-1.975	0.04832
bright	-0.03477	0.03425	-1.015	0.31004
green	-0.02572	0.05558	-0.463	0.64349
wet	-0.02620	0.06288	-0.417	0.67690
panv	-0.24090	1.28600	-0.187	0.85143
blf	-0.35170	1.49400	-0.235	0.81384
ccp	-0.19960	1.04400	-0.191	0.84831
ftbp	-0.39730	1.08900	-0.365	0.71526
nial	0.96270	1.10200	0.874	0.38226
pil	1.32300	1.15100	1.149	0.25066

sveg	0.64090	1.44500	0.444	0.65729
tws	-0.24310	1.37000	-0.177	0.85916
lcrich	0.01965	0.28730	0.068	0.94549
wetdist	0.00005	0.00003	1.846	0.06492
roadsdist	-0.00002	0.00011	-0.171	0.86423
humdist	-0.00011	0.00006	-1.802	0.07147
lst	0.35570	0.13610	2.614	0.00896
ND	388.24	df	338	
RD	240.19	df	299	
AIC	320.19	Kappa	0.5824	
Pearson ChiSq	96.6630	PCC	0.8466	
L.R.	140.94	AUC	0.8926	
R2	0.499	CV Error	0.1757	

Satellite Model Variable Importance



Land Cover Model Variable Importance

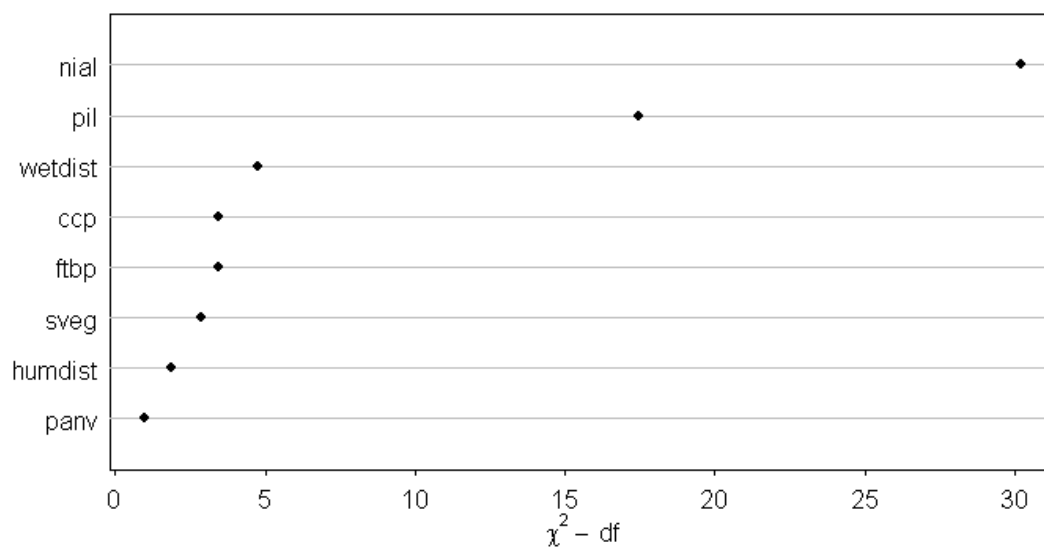


Figure 22: Importance of each variable in the satellite and the land cover model

Appendix D: R Code

```
-----  
# title           : abdi_thesis.R  
# purpose        : Habitat suitability and species distribution mapping  
# author         : Abdulhakim M. Abdi  
# last update    : 20 January 2010  
# response       : Miliaria calandra presence/absence data  
# explanatory    : Landsat bands, satellite image texture, vegetation  
                  indices, land surface temperature, CLC2000 landscape  
                  metrics  
# outputs        : Predictive map of habitat suitability for M. calandra  
-----  
  
# initialize required libraries:  
library(maptools)  
library(gstat)  
library(geoR)  
library(rgdal)  
library(lattice)  
library(spatstat)  
library(rpart)  
library(MASS)  
library(gbm)  
library(nnet)  
library(mda)  
library(Design)  
library(Hmisc)  
library(reshape)  
library(plyr)  
library(splancs)  
library(adehabitat)  
library(car)  
library(PresenceAbsence)  
library(boot)  
  
# set working directory  
setwd("C:/GeoData/Exercise")  
  
#set data source  
lerida <- read.csv("miliaria.lerida.csv", h=T, sep="," , dec=".")  
mili <- read.csv("mili.csv", h=T, sep="," , dec=".")  
str(lerida)  
  
#see how many presences and absences are there  
summary(factor(lerida$mili))  
  
predictors = readGDAL("asc/ndvi.asc")  
predictors$lst = readGDAL("asc/lst.asc")$band1
```

```

predictors$wetdist = readGDAL("asc/wetdist.asc") $band1
predictors$humdist = readGDAL("asc/humdist.asc") $band1
predictors$band1m = readGDAL("asc/band1_m.asc") $band1
predictors$band2m = readGDAL("asc/band2_m.asc") $band1
predictors$band3m = readGDAL("asc/band3_m.asc") $band1
predictors$band4m = readGDAL("asc/band4_m.asc") $band1
predictors$band5m = readGDAL("asc/band5_m.asc") $band1
predictors$band7m = readGDAL("asc/band7_m.asc") $band1
predictors$band1cv = readGDAL("asc/band1_cv.asc") $band1
predictors$band1sd = readGDAL("asc/band1_sd.asc") $band1
predictors$band2cv = readGDAL("asc/band2_cv.asc") $band1
predictors$band2sd = readGDAL("asc/band2_sd.asc") $band1
predictors$band3cv = readGDAL("asc/band3_cv.asc") $band1
predictors$band3sd = readGDAL("asc/band3_sd.asc") $band1
predictors$band4cv = readGDAL("asc/band4_cv.asc") $band1
predictors$band4sd = readGDAL("asc/band4_sd.asc") $band1
predictors$band5cv = readGDAL("asc/band5_cv.asc") $band1
predictors$band5sd = readGDAL("asc/band5_sd.asc") $band1
predictors$band7cv = readGDAL("asc/band7_cv.asc") $band1
predictors$band7sd = readGDAL("asc/band7_sd.asc") $band1
predictors$panv = readGDAL("asc/panv.asc") $band1
predictors$blf = readGDAL("asc/blf.asc") $band1
predictors$ccp = readGDAL("asc/ccp.asc") $band1
predictors$ftbp = readGDAL("asc/ftbp.asc") $band1
predictors$nia1 = readGDAL("asc/nial.asc") $band1
predictors$pil = readGDAL("asc/pil.asc") $band1
predictors$sveg = readGDAL("asc/sveg.asc") $band1
predictors$twc = readGDAL("asc/twc.asc") $band1
predictors$lcrich = readGDAL("asc/lcrich.asc") $band1
predictors$dem = readGDAL("asc/dem.asc") $band1
predictors$slope = readGDAL("asc/slope.asc") $band1
predictors$aspect = readGDAL("asc/aspect.asc") $band1
predictors$roadsdist = readGDAL("asc/roadsdist.asc") $band1
predictors$bright = readGDAL("asc/bright.asc") $band1
predictors$green = readGDAL("asc/green.asc") $band1
predictors$wet = readGDAL("asc/wet.asc") $band1
predictors$msavi = readGDAL("asc/msavi.asc") $band1
predictors$clc00 = readGDAL("asc/clc00.asc") $band1
predictors$ndvi = predictors$band1
predictors$band1=NULL
proj4string(predictors) <- CRS("+init=epsg:23031")
str(predictors)

# attach XY coordinates
coordinates(lerida)=~X+Y

# ED 1950 UTM Zone 31N:
proj4string(lerida) <- CRS("+init=epsg:23031")

```



```

# overlay presence absence points on the predictors
predictors.ov = overlay(predictors, lerida)
lerida$lst = predictors.ov$lst
lerida$band1m = predictors.ov$band1m
lerida$band2m = predictors.ov$band2m
lerida$band3m = predictors.ov$band3m
lerida$band4m = predictors.ov$band4m
lerida$band5m = predictors.ov$band5m
lerida$band7m = predictors.ov$band7m
lerida$ndvi_m = predictors.ov$ndvi
lerida$msavi_m = predictors.ov$msavi
lerida$band1sd = predictors.ov$band1sd
lerida$band2sd = predictors.ov$band2sd
lerida$band3sd = predictors.ov$band3sd
lerida$band4sd = predictors.ov$band4sd
lerida$band5sd = predictors.ov$band5sd
lerida$band7sd = predictors.ov$band7sd
lerida$band1cv = predictors.ov$band1cv
lerida$band2cv = predictors.ov$band2cv
lerida$band3cv = predictors.ov$band3cv
lerida$band4cv = predictors.ov$band4cv
lerida$band5cv = predictors.ov$band5cv
lerida$band7cv = predictors.ov$band7cv
lerida$bright = predictors.ov$bright
lerida$green = predictors.ov$green
lerida$wet = predictors.ov$wet
lerida$dem = predictors.ov$dem
lerida$slope = predictors.ov$slope
lerida$aspect = predictors.ov$aspect
lerida$panv = predictors.ov$panv
lerida$blf = predictors.ov$blf
lerida$ccp = predictors.ov$ccp
lerida$ftbp = predictors.ov$ftbp
lerida$snial = predictors.ov$snial
lerida$pil = predictors.ov$pil
lerida$sveg = predictors.ov$sveg
lerida$twts = predictors.ov$twts
lerida$lcrich = predictors.ov$lcrich
lerida$wetdist = predictors.ov$wetdist
lerida$humdist = predictors.ov$humdist
lerida$roadsdist = predictors.ov$roadsdist
lerida$clc00 = predictors.ov$clc00
str(lerida)

# take a look at the mean digital number distribution per land cover code
par(mfrow=c(3, 4))
boxplot(band1m~clc00, data=lerida, col=c("blue"), main="BAND 1 vs CLC",
xlab="CLC Code", ylab="Digital Number")

```

```

boxplot(band2m~clc00, data=lerida, col=(c("green")), main="BAND 2 vs CLC",
xlab="CLC Code", ylab="Digital Number")
boxplot(band3m~clc00, data=lerida, col=(c("red")), main="BAND 3 vs CLC",
xlab="CLC Code", ylab="Digital Number")
boxplot(band4m~clc00, data=lerida, col=(c("maroon")), main="BAND 4 vs
CLC", xlab="CLC Code", ylab="Digital Number")
boxplot(band5m~clc00, data=lerida, col=(c("gold")), main="BAND 5 vs CLC",
xlab="CLC Code", ylab="Digital Number")
boxplot(band7m~clc00, data=lerida, col=(c("grey")), main="BAND 7 vs CLC",
xlab="CLC Code", ylab="Digital Number")
boxplot(ndvi_m~clc00, data=lerida, col=(c("white")), main="NDVI vs CLC",
xlab="CLC Code", ylab="Value")
boxplot(msavi_m~clc00, data=lerida, col=(c("yellow1")), main="SAVI vs
CLC", xlab="CLC Code", ylab="Value")
boxplot(bright~clc00, data=lerida, col=(c("yellowgreen")),
main="Brightness vs CLC", xlab="CLC Code", ylab="Value")
boxplot(green~clc00, data=lerida, col=(c("green4")), main="Greenness vs
CLC", xlab="CLC Code", ylab="Value")
boxplot(wet~clc00, data=lerida, col=(c("blue3")), main="Wetness vs CLC",
xlab="CLC Code", ylab="Value")

# export into CSV
write.table(lerida,file="leridaex.csv",sep="," ,row.names=F, col.names=T)
lerida.im <- read.csv("leridaex.csv", h=T, sep="," , dec=".")
summary(lerida.im)
fix(lerida.im)

# There seems to be a row (340) that has NA values, so it has to be
removed
lerida.nona = na.omit(lerida.im)

# Plot conditional density plot of the binary outcome on the continuous x
variable.
# Miliaria calandra as factor response
mili.f = factor(lerida.nona$mili)

# plot satellite variables for Miliaria
par(mfrow=c(4,6))
cdplot(mili.f~band1m, data=lerida.nona)
cdplot(mili.f~band2m, data=lerida.nona)
cdplot(mili.f~band3m, data=lerida.nona)
cdplot(mili.f~band4m, data=lerida.nona)
cdplot(mili.f~band5m, data=lerida.nona)
cdplot(mili.f~band7m, data=lerida.nona)
cdplot(mili.f~band1sd, data=lerida.nona)
cdplot(mili.f~band2sd, data=lerida.nona)
cdplot(mili.f~band3sd, data=lerida.nona)
cdplot(mili.f~band4sd, data=lerida.nona)
cdplot(mili.f~band5sd, data=lerida.nona)

```

```

cdplot(mili.f~band7sd, data=lerida.nona)
cdplot(mili.f~band1cv, data=lerida.nona)
cdplot(mili.f~band2cv, data=lerida.nona)
cdplot(mili.f~band3cv, data=lerida.nona)
cdplot(mili.f~band4cv, data=lerida.nona)
cdplot(mili.f~band5cv, data=lerida.nona)
cdplot(mili.f~band7cv, data=lerida.nona)
cdplot(mili.f~ndvi_m, data=lerida.nona)
cdplot(mili.f~msavi_m, data=lerida.nona)
cdplot(mili.f~lst, data=lerida.nona)
cdplot(mili.f~bright, data=lerida.nona)
cdplot(mili.f~green, data=lerida.nona)
cdplot(mili.f~wet, data=lerida.nona)

# plot topographic variables for Miliaria
par(mfrow=c(3,1))
cdplot(mili.f~dem, data=lerida.nona)
cdplot(mili.f~slope, data=lerida.nona)
cdplot(mili.f~aspect, data=lerida.nona)

# plot anthropogenic variables for Miliaria
par(mfrow=c(2,1))
cdplot(mili.f~humdist, data=lerida.nona)
cdplot(mili.f~roadsdist, data=lerida.nona)

# plot land cover variables for Miliaria
par(mfrow=c(2,5))
cdplot(mili.f~nial, data=lerida.nona)
cdplot(mili.f~pil, data=lerida.nona)
cdplot(mili.f~blf, data=lerida.nona)
cdplot(mili.f~sveg, data=lerida.nona)
cdplot(mili.f~ftbp, data=lerida.nona)
cdplot(mili.f~panv, data=lerida.nona)
cdplot(mili.f~ccp, data=lerida.nona)
cdplot(mili.f~tws, data=lerida.nona)
cdplot(mili.f~wetdist, data=lerida.nona)
cdplot(mili.f~lcrich, data=lerida.nona)

## Individual variable relation to response

lst.lrm = lrm(mili~lst, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band1m.lrm = lrm(mili~band1m, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band2m.lrm = lrm(mili~band2m, data=lerida.nona,

```

```

method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band3m.lrm = lrm(mili~band3m, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band4m.lrm = lrm(mili~band4m, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band5m.lrm = lrm(mili~band5m, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band7m.lrm = lrm(mili~band7m, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

ndvi.lrm = lrm(mili~ndvi_m, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

msavi.lrm = lrm(mili~msavi_m, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band1sd.lrm = lrm(mili~band1sd, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band2sd.lrm = lrm(mili~band2sd, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band3sd.lrm = lrm(mili~band3sd, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band4sd.lrm = lrm(mili~band4sd, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band5sd.lrm = lrm(mili~band5sd, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band7sd.lrm = lrm(mili~band7sd, data=lerida.nona,
method="lm.fit", model=TRUE, x=TRUE, y=TRUE,

```

```

linear.predictors=TRUE, se.fit=TRUE)

band1cv.lrm = lrm(mili~band1cv, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band2cv.lrm = lrm(mili~band2cv, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band3cv.lrm = lrm(mili~band3cv, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band4cv.lrm = lrm(mili~band4cv, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band5cv.lrm = lrm(mili~band5cv, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

band7cv.lrm = lrm(mili~band7cv, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

bright.lrm = lrm(mili~bright, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

green.lrm = lrm(mili~green, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

wet.lrm = lrm(mili~wet, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

dem.lrm = lrm(mili~dem, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

slope.lrm = lrm(mili~slope, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

aspect.lrm = lrm(mili~aspect, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

```

```

panv.lrm = lrm(mili~panv, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

blf.lrm = lrm(mili~blf, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

ccp.lrm = lrm(mili~ccp, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

ftbp.lrm = lrm(mili~ftbp, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

nial.lrm = lrm(mili~nial, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

pil.lrm = lrm(mili~pil, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

sveg.lrm = lrm(mili~sveg, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

twsl.lrm = lrm(mili~twsl, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

lcrich.lrm = lrm(mili~lcrich, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

wetdist.lrm = lrm(mili~wetdist, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

humdist.lrm = lrm(mili~humdist, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

roadsdist.lrm = lrm(mili~roadsdist, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)

```

```

clc00.lrm = lrm(mili~clc00, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)
## END of individual variable relation to response

#data distribution
attach(lerida.nona)
ddist = datadist(band1m, band2m, band3m, band4m, band5m, band7m, ndvi_m,
msavi_m, band1sd, band2sd, band3sd, band4sd, band5sd, band7sd, band1cv,
band2cv, band3cv, band4cv, band5cv, band7cv, dem, slope, aspect,
bright, green, wet, panv, blf, ccp, ftbp, nial, pil, sveg, tws,
lcrich, wetdist, roadsdist, humdist, lst)
options(datadist='ddist')

#####
# Miliaria satellite imagery regression

sat.var = lerida.nona[c("band1m", "band2m", "band3m", "band4m", "band5m",
"band7m", "ndvi_m", "msavi_m", "band1sd", "band2sd", "band3sd", "band4sd",
"band5sd", "band7sd", "band1cv", "band2cv", "band3cv", "band4cv",
"band5cv",
"band7cv", "bright", "green", "wet", "lst", "dem", "slope", "aspect")]
sat.var.out <- glm(sat.var,data=lerida.nona)
vif(sat.var.out)

mili.sat.full =
formula(mili~band1m+band2m+band3m+band4m+band5m+band7m+ndvi_m+msavi_m+
band1sd+band2sd+band3sd+band4sd+band5sd+band7sd+band1cv+
band2cv+band3cv+band4cv+band5cv+band7cv+dem+slope+aspect+
bright+green+wet+lst)

temp.sat.modell = glm(mili.sat.full, binomial(link = "logit"),
data=lerida.nona)
drop1(temp.sat.modell, test="Chisq")
anova(temp.sat.modell, test="Chisq")

sat.modell = stepAIC(temp.sat.modell, scope= list(mili.sat.full),
direction="both")
summary(sat.modell)

# satellite model 1 Pearson Chi-Square
sum((sat.modell$y - sat.modell$fitted.values)^2/sat.modell$fitted.values)

#LRM of modell
mili.sat1= formula(mili ~ band4m + band5m + band7m + msavi_m + band1sd +
band2sd + band3sd + band4sd + band1cv + band2cv + band3cv +
band4cv + band5cv + band7cv + dem + slope + aspect + lst)

```

```

sat.modell.lrm = lrm(mili.sat1, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)
sat.modell.lrm
# Hosmer-Lemeshow Goodness of Fit
resid(sat.modell.lrm, 'gof')

## Presence Absence Package
mili$modell = sat.modell$fitted.values
modell.cmx = cmx(mili, threshold=0.5, which.model=1, na.rm=FALSE)
Kappa(modell.cmx)
pcc(modell.cmx)

auc.roc.plot(mili, threshold=101, which.model=1, model.names="model 1",
na.rm=TRUE, xlab="1-Specificity (false positives)",
ylab="Sensitivity (true positives)", main="ROC Plot",
color=TRUE, line.type=TRUE, lwd=1, mark.numbers=TRUE,
obs.prev=NULL, add.legend=TRUE, legend.text=NULL,
add.opt.legend=TRUE, pch=NULL)

presence.absence.accuracy(mili, threshold=0.5, find.auc=TRUE,
which.model=1)

###

# analysis of deviance
anova(sat.modell, test="Chisq")
drop1(sat.modell, test="Chisq")

# get the log odds
sat.modell$linear.predictors

# residuals
sat.modell.res = residuals(sat.modell)
hist(sat.modell.res)
plot(sat.modell.res)

# 95% confidence interval for coefficients
confint(sat.modell)

# exponentiate the coefficients = odds ratio
exp(coef(sat.modell))

# 95% CI for exponentiated coefficients (odds ratio)
exp(confint(sat.modell))

# predicted values can also use: fitted(model3)
sat.modell.predict = predict(sat.modell, type="response")

```



```

plot(sat.modell.predict)
plot(fitted(sat.modell), residuals(sat.modell))

# outlier test
outlier.test(sat.modell)

# k-folds cross validation (model validation)
modell.cv = cv.glm(lerida.nona, sat.modell, K=10)
modell.cv$delta

sat.modell.val = validate(sat.modell.lrm, method="crossvalidation", B=10,
bw=FALSE, rule="aic",
type="residual", sls=0.05, aics=0, pr=FALSE, Dxy.method='somers2')

#####
###
### Satellite Model 2

summary(sat.modell)

vif.sat1 = glm(mili~band4m+band5m+band7m+msavi_m+band1sd+band2sd+band3sd+
band4sd+band1cv+band2cv+band3cv+band4cv+band5cv+band7cv+dem+slope+
aspect+lst, binomial(link = "logit"), data=lerida.nona)
vif(vif.sat1)

# remove band2sd
vif.sat1 = glm(mili~band4m+band5m+band7m+msavi_m+band1sd+band3sd+
band4sd+band1cv+band2cv+band3cv+band4cv+band5cv+band7cv+dem+slope+
aspect+lst, binomial(link = "logit"), data=lerida.nona)
vif(vif.sat4)

# remove band7m
vif.sat1 = glm(mili~band4m+band5m+msavi_m+band1sd+band3sd+
band4sd+band1cv+band2cv+band3cv+band4cv+band5cv+band7cv+dem+slope+
aspect+lst, binomial(link = "logit"), data=lerida.nona)
vif(vif.sat1)

# remove band1cv
vif.sat1 = glm(mili~band4m+band5m+msavi_m+band1sd+band3sd+
band4sd+band2cv+band3cv+band4cv+band5cv+band7cv+dem+slope+
aspect+lst, binomial(link = "logit"), data=lerida.nona)
vif(vif.sat1)

# remove band3cv
vif.sat1 = glm(mili~band4m+band5m+msavi_m+band1sd+band3sd+
band4sd+band2cv+band4cv+band5cv+band7cv+dem+slope+
aspect+lst, binomial(link = "logit"), data=lerida.nona)
vif(vif.sat1)

```

```

# remove band7cv
vif.sat1 = glm(mili~band4m+band5m+msavi_m+band1sd+band3sd+
band4sd+band2cv+band4cv+band5cv+dem+slope+
aspect+lst, binomial(link = "logit"), data=lerida.nona)
vif(vif.sat1)

# remove band4cv
vif.sat1 = glm(mili~band4m+band5m+msavi_m+band1sd+band3sd+
band4sd+band2cv+band5cv+dem+slope+
aspect+lst, binomial(link = "logit"), data=lerida.nona)
vif(vif.sat1)

# remove band3sd
vif.sat1 = glm(mili~band4m+band5m+msavi_m+band1sd+
band4sd+band2cv+band5cv+dem+slope+
aspect+lst, binomial(link = "logit"), data=lerida.nona)
vif(vif.sat1)

# End of Multicollinearity Analysis
#####

summary(vif.sat1)

#Remove insignificant variables band5m+band4sd+band2cv+aspect

mili.sat2 = formula(mili~band4m+msavi_m+band1sd+
band5cv+dem+slope+lst)
sat.model2 = glm(mili.sat2, binomial(link = "logit"), data=lerida.nona)
summary(sat.model2)
#

# satellite model 2 Pearson Chi-Square
sum((sat.model2$y - sat.model2$fitted.values)^2/sat.model2$fitted.values)

#LRM of model2
sat.model2.lrm = lrm(mili.sat2, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)
sat.model2.lrm
resid(sat.model2.lrm, 'gof')

## Presence Absence Package
mili$model2 = sat.model2$fitted.values
model2.cmx = cmx(mili, threshold=0.5, which.model=4, na.rm=FALSE)
Kappa(model2.cmx)
pcc(model2.cmx)

auc.roc.plot(mili, threshold=101, which.model=2, model.names="model 4",
na.rm=TRUE, xlab="1-Specificity (false positives)",

```

```

ylab="Sensitivity (true positives)", main="ROC Plot",
color=TRUE, line.type=TRUE, lwd=1, mark.numbers=TRUE,
obs.prev=NULL, add.legend=TRUE, legend.text=NULL,
add.opt.legend=TRUE, pch=NULL)

presence.absence.accuracy(mili, threshold=0.5, find.auc=TRUE,
which.model=2)
###

# analysis of deviance
anova(sat.model2, test="Chisq")
drop1(sat.model2, test="Chisq")

# get the log odds
sat.model2$linear.predictors

# residuals
sat.model2.res = residuals(sat.model2)
hist(sat.model2.res)
plot(sat.model2.res)

# 95% confidence interval for coefficients
confint(sat.model2)

# exponentiate the coefficients = odds ratio
exp(coef(sat.model2))

# 95% CI for exponentiated coefficients (odds ratio)
exp(confint(sat.model2))

# predicted values can also use: fitted(model2)
sat.model2.predict = predict(sat.model2, type="response")
plot(sat.model2.predict)
plot(fitted(sat.model2), residuals(sat.model2))

# outlier test
outlier.test(sat.model2)

# Goodness of fit: likelihood ratio test
lrtest(sat.modell1, sat.model2)

# k-folds cross validation (model validation)
model2.cv = cv.glm(lerida.nona, sat.model2, K=10)
model2.cv$delta

sat.model2.val = validate(sat.model2.lrm, method="crossvalidation", B=10,
bw=FALSE, rule="aic",
type="residual", sls=0.05, aics=0, pr=FALSE, Dxy.method='somers2')

```

```

#####
### Miliaria and land cover variables

mili.clc = formula(mili~panv+blf+ccp+ftbp+nial+pil+sveg+tws+
lcrich+humdist+wetdist+roadsdist)
clc.model = glm(mili.clc, binomial(link = "logit"),
data=lerida.nona)
summary(clc.model)
exp(coef(clc.model))

clc.step = stepAIC(clc.model, scope= list(mili.clc), direction="both")
summary(clc.step)
exp(coef(clc.step))

clc.step.formula = formula(mili~panv+ccp+ftbp+nial+pil+sveg+
humdist+wetdist)
summary(glm(clc.step.formula, binomial(link = "logit"),
data=lerida.nona))

vif(clc.step)

# CLC model Pearson Chi-Square
sum((clc.step$y - clc.step$fitted.values)^2/clc.step$fitted.values)

#LRM of model3 (with Wald values)
clc.step.lrm = lrm(clc.step.formula, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)
clc.step.lrm
resid(clc.step.lrm, 'gof')

## Presence Absence Package
mili$model3 = clc.step$fitted.values
model3.cmx = cmx(mili, threshold=0.5, which.model=3, na.rm=FALSE)
Kappa(model3.cmx)
pcc(model3.cmx)

auc.roc.plot(mili, threshold=101, which.model=3,
na.rm=TRUE, xlab="1-Specificity (false positives)",
ylab="Sensitivity (true positives)", main="ROC Plot",
color=TRUE, line.type=TRUE, lwd=1, mark.numbers=TRUE,
obs.prev=NULL, add.legend=TRUE, legend.text=NULL,
add.opt.legend=TRUE, pch=NULL)

presence.absence.accuracy(mili, threshold=0.5, find.auc=TRUE,
which.model=3)
###

# analysis of deviance

```

```

anova(clc.step, test="Chisq")
drop1(clc.step, test="Chisq")

# get the log odds
plot(clc.step$linear.predictors)

# residuals
hist(residuals(clc.step))
plot(residuals(clc.step))

# 95% confidence interval for coefficients
confint(clc.step)

# exponentiate the coefficients = odds ratio
exp(coef(clc.step))

# 95% CI for exponentiated coefficients (odds ratio)
exp(confint(clc.step))

# predicted values can also use: fitted(model3)
plot(predict(clc.step, type="response"))
plot(fitted(clc.step), residuals(sat.model3))

# outlier test
outlier.test(clc.step)

# k-folds cross validation (model validation)
clc.cv = cv.glm(lerida.nona, clc.step, K=10)
clc.cv$delta

clc.step.val = validate(clc.step.lrm, method="crossvalidation", B=10,
bw=FALSE, rule="aic",
type="residual", sls=0.05, aics=0, pr=FALSE, Dxy.method='somers2')

#####
#####
# Miliaria and the Full Model

mili.full =
formula(mili~band1m+band2m+band3m+band4m+band5m+band7m+ndvi_m+msavi_m+
band1sd+band2sd+band3sd+band4sd+band5sd+band7sd+band1cv+
band2cv+band3cv+band4cv+band5cv+band7cv+dem+slope+aspect+
bright+green+wet+panv+blf+ccp+ftbp+nial+pil+sveg+twst+
lcrich+wetdist+roadsdist+humdist+lst)
full.model = glm(mili.full, binomial(link = "logit"), data=lerida.nona)
summary(full.model)
mili$model7 = full.model$fitted.values
presence.absence.accuracy(mili, threshold=0.5, find.auc=TRUE,
which.model=7)

```

```

combo.cv = cv.glm(lerida.nona, full.model, K=10)
combo.cv$delta

combined = formula(mili~band4m+msavi_m+band1sd+band5cv+dem+slope+lst+
nial+pil+sveg+humdist+wetdist)
combo.model = glm(combined, binomial(link = "logit"), data=lerida.nona)
summary(combo.model)

vif(combo.model)

# Full model Pearson Chi-Square
sum((combo.model$y -
combo.model$fitted.values)^2/combo.model$fitted.values)

#LRM of combined model
combo.model.lrm = lrm(combined, data=lerida.nona,
method="lrm.fit", model=TRUE, x=TRUE, y=TRUE,
linear.predictors=TRUE, se.fit=TRUE)
combo.model.lrm
resid(combo.model.lrm, 'gof')

par(mfrow=c(2,6))
plot.Design(combo.model.lrm)

#univarLR takes a multivariable model fit object from Design and
#re-fits a sequence of models containing one predictor at a time.
#It prints a table of likelihood ratio chi^2 statistics from these fits.

univarLR(combo.model.lrm)

## Presence Absence Package
mili$model6 = combo.model$fitted.values
model6.cmx = cmx(mili, threshold=0.5, which.model=6, na.rm=FALSE)
Kappa(model6.cmx)
pcc(model6.cmx)

auc.roc.plot(mili, threshold=101, which.model=c(1,2,3),
na.rm=TRUE, xlab="1-Specificity (false positives)",
ylab="Sensitivity (true positives)", main="ROC Plot",
color=TRUE, line.type=TRUE, lwd=1, mark.numbers=TRUE,
obs.prev=NULL, add.legend=TRUE, legend.text=NULL,
add.opt.legend=TRUE, pch=NULL)

presence.absence.accuracy(mili, threshold=0.5, find.auc=TRUE,
which.model=6)
###

# analysis of deviance
anova(combo.model, test="Chisq")

```

```

drop1(combo.model test="Chisq")

# get the log odds
sat.model3$linear.predictors

# residuals
hist(residuals(combo.model))
plot(residuals(combo.model))

# 95% confidence interval for coefficients
confint(combo.model)

# exponentiate the coefficients = odds ratio
exp(coef(combo.model))

# 95% CI for exponentiated coefficients (odds ratio)
exp(confint(combo.model))

# predicted values can also use: fitted(model3)
pred.comb = predict.glm(combo.model, type="response", se.fit=TRUE)
plot(lerida.nona$mili, pred.comb$fit, xlab="M. calandra PA",
ylab="Predicted")
lines(lerida.nona$mili, pred.comb$fit - 1.96 * pred.comb$se.fit, lty=2)
lines(lerida.nona$mili, pred.comb$fit + 1.96 * pred.comb$se.fit, lty=2)
plot(lerida.nona$mili, fitted(combo.model))

plot(predict(sat.model3, type="response"))
plot(fitted(sat.model3), residuals(sat.model3))

# outlier test
outlier.test(combo.model)

# k-folds cross validation (model validation)
combo.cv = cv.glm(lerida.nona, combo.model, K=10)
combo.cv$delta

sat.model3.val = validate(clc.step.lrm, method="crossvalidation", B=10,
bw=FALSE, rule="aic",
type="residual", sls=0.05, aics=0, pr=FALSE, Dxy.method='somers2')

```