LUND UNIVERSITY

Integrating Open Access Geospatial Data to Map the Habitat Suitability of the Declining Corn Bunting (Miliaria calandra)

Abdi, Hakim

Published in:
ISPRS International Journal of Geo-Information

DOI:
10.3390/ijgi2040935

2013

Link to publication

Citation for published version (APA):
Abdi, H. (2013). Integrating Open Access Geospatial Data to Map the Habitat Suitability of the Declining Corn Bunting (Miliaria calandra). *ISPRS International Journal of Geo-Information*, *2*(4), 935-954. https://doi.org/10.3390/ijgi2040935

Total number of authors:
1

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

ISPRS Int. J. Geo-Inf. **2013**, 2, 935-954; doi:10.3390/ijgi2040935

OPEN ACCESS
ISPRS International
Journal of
Geo-Information
ISSN 2220-9964
www.mdpi.com/journal/ijgi/

Article

Integrating Open Access Geospatial Data to Map the Habitat Suitability of the Declining Corn Bunting (*Miliaria calandra*)

Abdulhakim M. Abdi [1,2,*]

[1] Institute for Geoinformatics, Universität Münster, Weseler Straße 253, Münster 48151, Germany; E-Mail: hakim.abdi@uni-muenster.de

[2] Institute of Statistics and Information Management, Universidade Nova de Lisboa, Lisbon 1070-312, Portugal

* Current Address: Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, Lund 223 62, Sweden; E-Mail: hakim.abdi@nateko.lu.se

Received: 31 July 2013; in revised form: 13 August 2013 / Accepted: 30 August 2013 / Published: 27 September 2013

Abstract: The efficacy of integrating open access geospatial data to produce habitat suitability maps for the corn bunting (*Miliaria calandra*) was investigated. Landsat Enhanced Thematic Mapper Plus (ETM+), Shuttle Radar Topography Mission (SRTM) and Corine (Coordination of Information on the Environment) land cover data for the year 2000 (CLC2000) were processed to extract explanatory variables and divided into three sets; Satellite (ETM+, SRTM), CLC2000 and Combined (CLC2000 + Satellite). Presence-absence data for *M. calandra*, collected during structured surveys for the Catalan Breeding Bird Atlas, were provided by the Catalan Ornithological Institute. The dataset was partitioned into an equal number of presence and absence points by dividing it into five groups, each composed of 88 randomly selected presence points to match the number of absences. A logistic regression model was then built for each group. Models were evaluated using area under the curve (AUC) of the receiver operating characteristic (ROC). Results of the five groups were averaged to produce mean Satellite, CLC2000 and Combined models. The mean AUC values were 0.69, 0.81 and 0.90 for the CLC2000, Satellite and the Combined model, respectively. The probability of *M. calandra* presence had the strongest positive correlation with land surface temperature, modified soil adjusted vegetation index, coefficient of variation for ETM+ band 5 and the fraction of non-irrigated arable land.

Keywords: agricultural intensification; Corine land cover; corn bunting; Landsat; *Miliaria calandra*; open access data; open source geospatial software; species distribution modeling

1. Introduction

European farmland bird species and long-distance migrants continue to decrease at an alarming rate [1] despite the passing, in 1979, of the Council Directive 79/409/EEC on the conservation of wild birds, also known as the "Birds Directive". Examples of declining farmland bird species include the corncrake (*Crex crex*) in France [2] and the grey partridge (*Perdix perdix*) in Italy [3]. Some species have become extinct as regular breeders in certain countries, such as the red-backed shrike (*Lanius collurio*) in the United Kingdom and the roller (*Coracias garrulus*) in the Czech Republic [4].

Four farmland-specialized bunting species in Western Europe have also suffered particularly severe declines. For example, the Ortolan bunting (*Emberiza hortulana*) population in Finland declined 72% between 1984 and 2002 [5], and in Poland, its population declined 20% between 2000 and 2010 [6]. The population collapse of the cirl bunting (*Emberiza cirlus*) in Britain between 1970 and 1990 coincided with unprecedented large-scale changes in agricultural practice, leading to losses in the species' breeding habitat [7,8]. The yellowhammer (*Emberiza citrinella*) suffered significant declines in the United Kingdom [9], Poland [10] and Sweden [11] since the mid-1960s, due to losses of both suitable breeding habitat and nearby wintering sites [12]. Similarly, northern and central European populations of the corn bunting (*Miliaria calandra*) have declined sharply since the mid-1970s [4], notably in Britain [13], Poland [14], Germany [15] and Ireland [16]. The declines of these species have been credited to detrimental land use policies, such as the Common Agricultural Policy (CAP) [1], that have been in effect in the European Union since the early 1960s. CAP promotes the maximization of agricultural productivity through the intensification of farmland; a process that results in monocultures, pesticide use and the eradication of uncultivated areas [17].

Geospatial technologies provide valuable tools to map bird distribution by linking independent observations with habitat and other environmental information extracted from satellite imagery or land cover data. To assess the relationship between farmland bird species and their habitats, Fuller *et al.* [18] applied a clustering method to link species' occurrence data from the atlas of breeding birds of Britain to a national land cover dataset and found that maps of species clusters exhibited patterns associated with habitat assemblages. In another study, Kuczynski *et al.* [19] combined the proportional cover of Corine land cover classes and digital elevation data to study the habitat use of the great grey shrike (*Lanius excubitor).* Recently, Kosicki and Chylarecki [6] combined data from Corine land cover, a digital elevation model, normalized difference vegetation index (NDVI) and climatic data from Worldclim to predict the habitat of the *E. hortulana* in Poland.

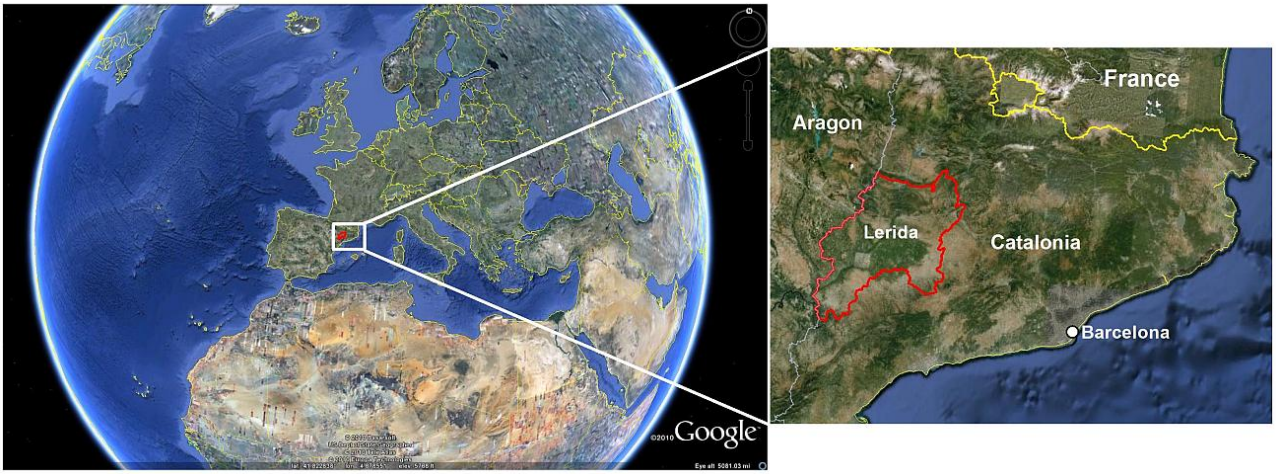
This study aims to better understand the habitat requirement of *M. calandra* in north-eastern Spain through the fusion of open access data derived from Corine land cover, Landsat and the Shuttle Radar Topography Mission as surrogates for land use, habitat structure and topography, respectively. Declines in Northern European *M. calandra* populations have been attributed to farmland intensification [13,17], while Southern European breeding densities, particularly in Spain, Portugal and

Turkey, are fairly stable [20,21], due to the prevalence of traditional low-intensity farming. Furthermore, most of the studies on the distribution ecology of *M. calandra* have been conducted in Northern and Central Europe [22]. Hence, there is an opportunity for the study of the habitat requirement of this species in a part of the continent that is relatively unaffected by the declining trend.

2. Study Area

The study area is located in the province of Lerida in the western part of the autonomous community of Catalonia in Spain (Figure 1). The area consists of the comarques of Noguera, Segria, Urgell and Pla d'Urgell, covers approximately 3,920 square kilometers and is composed mainly of non-irrigated cropland and dry forests with extensive agriculture and dry pastures. Intensive agricultural practices are slowly being implemented in the region to replace the economically unviable low-intensity extensive farming methods that dominate the landscape of Lerida [23,24].

Figure 1. Location of Lerida (red outline) in relation to Europe and Catalonia.



3. Materials and Methods

3.1. Open Access Geospatial Data

Since the 1970s, Landsat data has been widely used to map the habitat preference and distribution of various species (e.g., [25–28]). Imagery from the Enhanced Thematic Mapper Plus (ETM+) sensor onboard the Landsat 7 satellite were downloaded from the USGS Global Visualization Viewer, version 7.26, available from glovis.usgs.gov [29]. Two scenes from path-198, row-31 for the month of June (1 June 2006 and 17 June) of the year 2001 were used for this study to temporally coincide with the survey period of the bird data. The dataset used was a standard level one terrain-corrected (L1T) product that underwent radiometric and geometric correction. Atmospheric correction was performed on the Landsat data using the "Landsat" package [30] in the software program R [31], then projected into the European Datum 1950/Universal Transverse Mercator zone 31 North (ED50/UTM 31N) projection. A brief description of the Landsat bands is given in Table 1.

Table 1. Description of the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) bands used in this study. Optical data ranges from the beginning of the visible (400 nm) until the end of the thermal infrared (1 mm) portion of the electromagnetic spectrum.

Band	Spatial Resolution (m)	Lower Limit (μm)	Upper Limit (μm)	Designation	Optical Domain
1	30	0.45	0.52	Blue	Visible
2	30	0.52	0.60	Green	Visible
3	30	0.63	0.69	Red	Visible
4	30	0.77	0.90	Near Infrared	Infrared
5	30	1.55	1.75	Shortwave Infrared	Infrared
6	60	10.40	12.50	Thermal Infrared	Thermal
7	30	2.09	2.35	Shortwave Infrared	Infrared

The Shuttle Radar Topography Mission (SRTM) [32] digital elevation model (DEM) was included as a variable because *M. calandra* abundance is linked directly to habitat diversity, which is, in turn, correlated with topographic variability [33]. The 90 m DEM dataset was downloaded from the CGIAR-CSI database [34] and projected into the ED50/UTM 31N projection. The dataset was then resampled to 30 m to match the resolution of the ETM+ imagery using the nearest neighbor method. Elevation of the study area ranges from 1,603 m in the northern parts to 73 m around the river Segre.

Corine (Coordination of Information on the Environment) land cover 100 m data for the year 2000 (CLC2000) [35] was downloaded from the European Environmental Agency (EEA) website and projected into the ED50/UTM 31N projection. The dataset was rasterized at the original 100 m resolution, then resampled to 30 m using the nearest neighbor method to match the spatial resolution of the ETM+ imagery. Corine is a pan-European project that aims to produce a distinctive and comparable land cover dataset for Europe. The project has a total of 44 land cover classes, out of which, 26 occur in the study area (Table 2). The entire CLC2000 dataset is available from www.eea.europa.eu/data-and-maps [35].

3.2. Explanatory Variables

Explanatory variables in this study describe the biogeophysical aspects of the landscape and were used in statistical analysis to predict the occurrence of *M. calandra* at unsurveyed locations. A total of 27 explanatory variables (Table 3) were derived from open access geospatial datasets described in the preceding section. The following section details the different subgroups of explanatory variables.

Table 2. Coordination of Information on the Environment (Corine) land cover 2000 (CLC2000) classes that occur in the study area. Classes included in the analysis represent over 95% of the total surface area.

CLC2000 Classes	Area in km^2	Percent of Study Area
Permanently irrigated land	1,440	26.48
Non-irrigated arable land	1,182	21.74
Complex cultivation patterns	923	16.97
Fruit trees and berry plantations	452	8.31
Agricultural with natural vegetation	396	7.27
Transitional woodland-shrub	331	6.08
Sclerophyllous vegetation	240	4.41
Broad-leaved forest	240	4.41
Mixed forest	56	1.04
Water bodies	50	0.92
Continuous urban fabric	38	0.69
Vineyards	26	0.48
Water courses	16	0.29
Olive groves	12	0.23
Natural grasslands	8	0.15
Industrial or commercial units	8	0.14
Discontinuous urban fabric	5	0.10
Mineral extraction sites	5	0.09
Bare rocks	3	0.05
Inland marshes	2	0.04
Annual crops associated with perm crops	2	0.03
Pastures	1	0.02
Rice fields	1	0.02
Sparsely vegetated areas	1	0.02
Sport and leisure facilities	0.5	0.01
Construction sites	0.2	0.004

Table 3. Univariate descriptive statistics of all extracted explanatory variables grouped by data source and their associated acronyms. SD, standard deviation; CV, coefficient of variation; LST, land surface temperature; MSAVI, modified soil adjusted vegetation index; SRTM, Shuttle Radar Topography Mission.

	Variable	Coefficient	S.E.	Wald	*p*
Landsat ETM+	Standard deviation ETM+1: BAND1SD	−1.106	0.636	−1.74	0.009
	Standard deviation ETM+2: BAND2SD	0.017	0.027	0.63	0.523
	Standard deviation ETM+3: BAND3SD	0.027	0.016	1.69	0.103
	Standard deviation ETM+4: BAND4SD	0.012	0.019	0.63	0.523
	Standard deviation ETM+5: BAND5SD	−0.005	0.018	−0.28	0.763
	Standard deviation ETM+7: BAND7SD	0.022	0.018	1.22	0.209
	Coefficient of variation ETM+1: BAND1CV	1.172	1.240	0.95	0.344
	Coefficient of variation ETM+2: BAND2CV	1.397	1.225	1.14	0.253
	Coefficient of variation ETM+3: BAND3CV	1.922	0.934	2.06	0.039
	Coefficient of variation ETM+4: BAND4CV	−1.003	1.005	−1.00	0.318
	Coefficient of variation ETM+5: BAND5CV	−2.994	1.049	−2.85	0.003
	Coefficient of variation ETM+7: BAND7CV	0.995	0.964	1.03	0.302
	LST (℃)	2.182	1.052	2.07	<0.001
	MSAVI	2.311	0.677	3.41	<0.001
SRTM	Digital elevation model: DEM (meters)	−0.001	0.007	−0.15	0.022
	Slope: SLOPE (degrees)	−0.252	0.047	−5.36	0.001
	Aspect: ASPECT	−0.001	0.001	−1.00	0.311
CLC2000	Agricultural with natural vegetation: PANV *	−0.522	0.775	−0.67	0.500
	Broad-leaved forest: BLF *	−2.453	0.895	−2.74	0.206
	Complex cultivation patterns: CCP *	−0.393	0.390	−1.01	0.314
	Fruit trees and berry plantations: FTBP *	−0.479	0.460	−1.04	0.298
	Non-irrigated arable land: NIAL *	2.694	0.594	4.54	<0.001
	Permanently irrigated land: PIL *	1.139	0.397	2.87	0.004
	Sclerophyllous vegetation: SVEG *	−0.449	0.798	−0.56	0.573
	Transitional woodland-shrub: TWS *	−2.469	0.722	−3.42	0.600
	Distance to wet areas: WETDIST (meters)	7.69e−05	1.95e−05	3.94	<0.001
	Distance to human activity: HUMDIST (meters)	−1.26e−04	3.74e−05	−3.37	<0.001

* = Fractional cover

3.2.1. Satellite Image Texture (ETM+)

Saveraid *et al.* [36] proposed that the use of satellite data alone is not sufficient in modeling bird distribution and that habitat structure variables are also necessary. In order to extract habitat structure from the satellite data, first order texture variables were produced by running a 3 × 3 pixel (~8,100 m^2) local statistic across ETM+ imagery of the study area to calculate standard deviation (SD) and coefficient of variation (CV). This resulted in two texture measurements for each of the six ETM+ visible and infrared bands: BAND*x*SD and BAND*x*CV, where *x* is the Landsat band used (see Tables 1 and 2). Standard deviation assesses the variability of image texture, and the coefficient of variation (standard deviation of pixel values divided by their mean) gives a measure of the variability in image texture as a percent of the mean. Satellite image texture variables have been used to quantify landscape heterogeneity and to model the spatial patterns of birds [37] by providing data on the relationship between bird distribution and habitat structure.

3.2.2. Vegetation Index (ETM+)

Spectral vegetation indices have been widely used in habitat mapping [38], and they provide important information about the condition of the vegetation. The modified soil adjusted vegetation index (MSAVI) was chosen to minimize the soil background effect and enhance the dynamic range of the vegetation signal, producing greater vegetation sensitivity in areas that have significant portions of bare soil [39]. MSAVI was calculated from the red (ETM+3) and near-infrared (ETM+4) bands using the methods described in Qi *et al.* [39].

3.2.3. Land Surface Temperature (ETM+)

Land surface temperature (LST) measures the energy efficiency of terrestrial ecosystems by quantifying radiated thermal energy [40]. It is seldom included as a variable in ecological studies [41], but it can offer valuable information about anthropogenic or natural modifications to ecosystem energy budgets [42]. LST was derived from the thermal infrared band (ETM+6) using methods described by Sobrino *et al.* [43] and NASA [44].

3.2.4. Landscape Metrics (CLC2000)

Landscape metrics describe the spatial pattern of habitats and have been used to classify the habitat suitability of different bird species [45,46]. Metrics were calculated to quantify fractions of the eight dominant habitat types in the study area; dominance in this context was established if a CLC2000 class covered an area greater than 100 m^2. For each metric, a 3 × 3 local statistic was used to calculate the proportion (between 0 and 1) of a given CLC2000 class within the window. Additionally, distance metrics were obtained by calculating the Euclidean distance of each pixel from superimposed layers consisting of regions occupied by humans and wet areas (rivers, lakes and marshes). See De Smith *et al.* [47] for details on the derivation of landscape metrics.

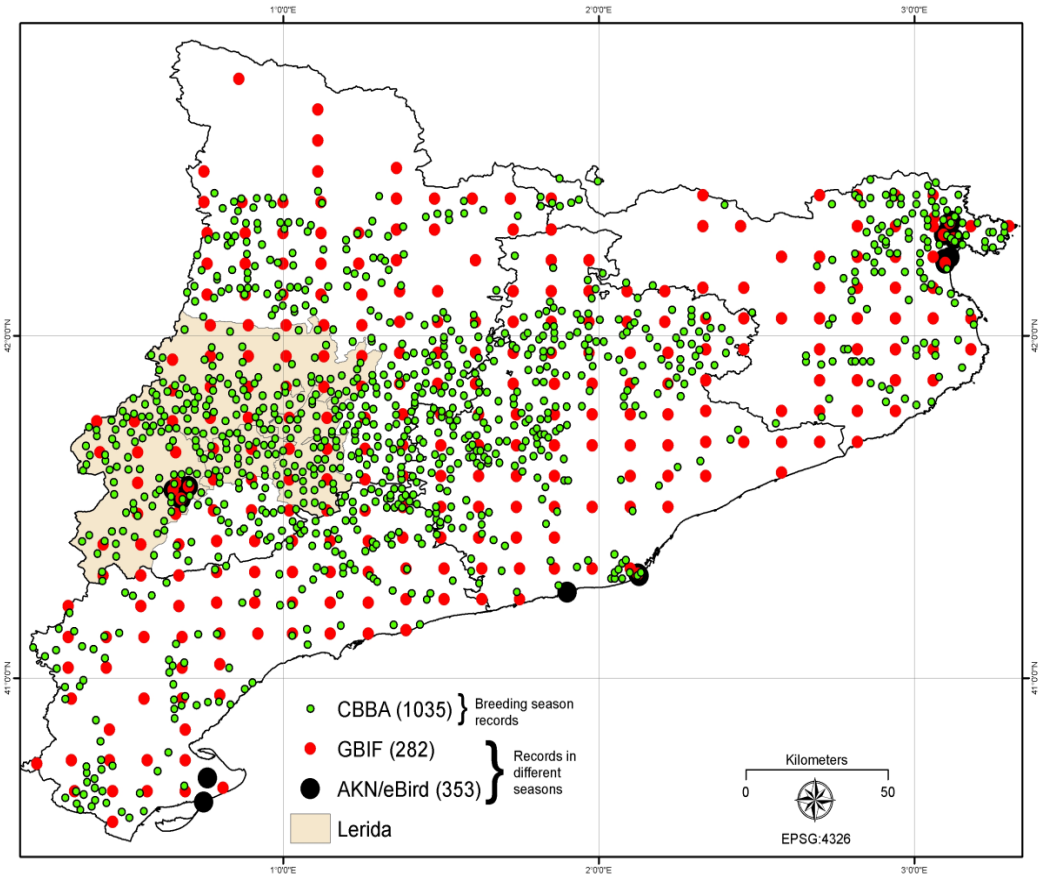
3.2.5. Topography (SRTM)

Topographic heterogeneity directly affects habitat selection by influencing temperature and humidity gradients and indirectly by modifying the vegetation composition [48]. The inclination of the surface slope in degrees and aspect (downslope direction) of the study area were extracted from the DEM.

3.3. Open Access Bird Data

This study was commenced with the intention to only use open access data in mapping the habitat of *M. calandra*. Therefore, data on *M. calandra* presence was downloaded from two open access sources, the Global Biodiversity Information Facility (GBIF; www.gbif.org) [49] and the Avian Knowledge Network/eBird (AKN/eBird; www.avianknowledge.net) [50]. However, it was found that the quality of these data was too poor to perform robust analysis on the habitat suitability of this particular species in Catalonia (Figure 2). Hence, these open access data were not used in this study.

Figure 2. Comparison of corn bunting observations in Catalonia from one closed access (Catalan Breeding Bird Atlas (CBBA)) and two open access sources (Global Biodiversity Information Facility (GBIF) and Avian Knowledge Network/eBird (AKN/eBird)).



Data from the Catalan Breeding Bird Atlas

Data was requested from the Catalan Ornithological Institute (ICO), which provided presence and absence records of *M. calandra* from the Catalan Breeding Bird Atlas (CBBA) [51]. CBBA surveys were conducted during the breeding season (1 March to 30 July) in the years 1999–2002. Surveys were conducted by experienced professionals between sunrise and 11 am and between 6 pm and sunset. The survey plots were 1 km × 1 km UTM squares in which two 1-hour surveys were conducted and the presence or absence of the species was recorded [51]. The total number of *M. calandra* locations in the study area was 339, out of which, 251 (74%) were presence points and 88 (26%) were absence points. Further details on the field sampling methodology can be found in Brotons *et al.* [51].

3.4. Free and Open Source Geospatial Software

Due to the limited availability of funds, only free and open source software were used in this study. Statistical modeling was conducted using the R Environment for Statistical Computing version 2.9.2 [31]; the latest version of R is available from www.r-project.org [31]. R was also used to project the data to the ED50/UTM 31N projection using the "rgdal" package [52]. Spatial analysis, including calculation of landscape metrics and visualization of the final maps, was conducted using Whitebox Geospatial Analysis Tools (WGAT) version 0.12, available at www.uoguelph.ca/~hydrogeo/Whitebox [53]. Topographic analysis was done in the System for Automated Geoscientific Analyses (SAGA) version 2.03 [54]. SAGA is available from www.saga-gis.org [54]. All work was done in Ubuntu version 9.10. Ubuntu is a Linux-based open source operating system that can be downloaded from www.ubuntu.com [55].

3.5. Modeling

There are three times (251) as many presence points as there are absence points (88) in the bird data, and this disparity will introduce bias in the modeling process. In order to minimize this, the dataset was partitioned into an equal number of presence and absence points. Five groups were created (Figure 3), each composed of 88 randomly selected presence points to match the number of absences in the dataset. All the explanatory variables were extracted at each one of these points using the "overlay" function in R. The modeling framework is presented in Figure 4.

Each of the five groups comprises three subgroups: (1) a Satellite group that comprises the explanatory variables derived from the ETM+ imagery and the SRTM data, (2) a CLC2000 group that comprises the landscape metrics derived from the Corine land cover data and (3) a Combined group that consists of both the Satellite and CLC2000 variables.

A logistic regression [56] model was built for each of the five groups, and a stepwise elimination process based on Akaike Information Criterion (AIC) [57] optimization was applied to remove insignificant variables. Then, a stepwise variance inflation factor (VIF) [58] elimination was applied using the "vif" function in R. All variables with VIF values greater than 5 [59] were discarded to avoid multicollinearity (Table A1 in the Appendix). Models were evaluated using Cohen's kappa [60], area under the curve (AUC) of the receiver operating characteristic (ROC) [61] with a threshold of 0.50 to

denote suitable habitat. Resultant probability maps were averaged to produce mean CLC2000, Satellite and Combined maps (Figure 5).

Figure 3. The 251 presence points were divided into five groups, each composed of 88 randomly selected presence points (in black) to match the 88 absence points (in red).

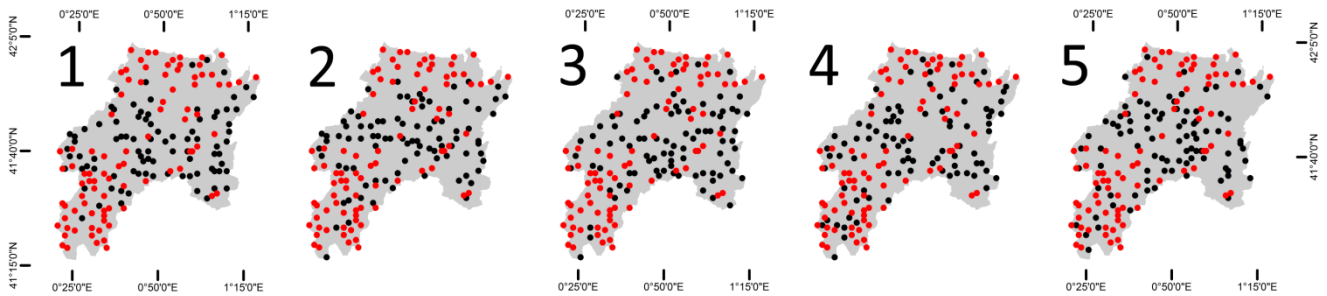


Figure 4. Framework of this study.

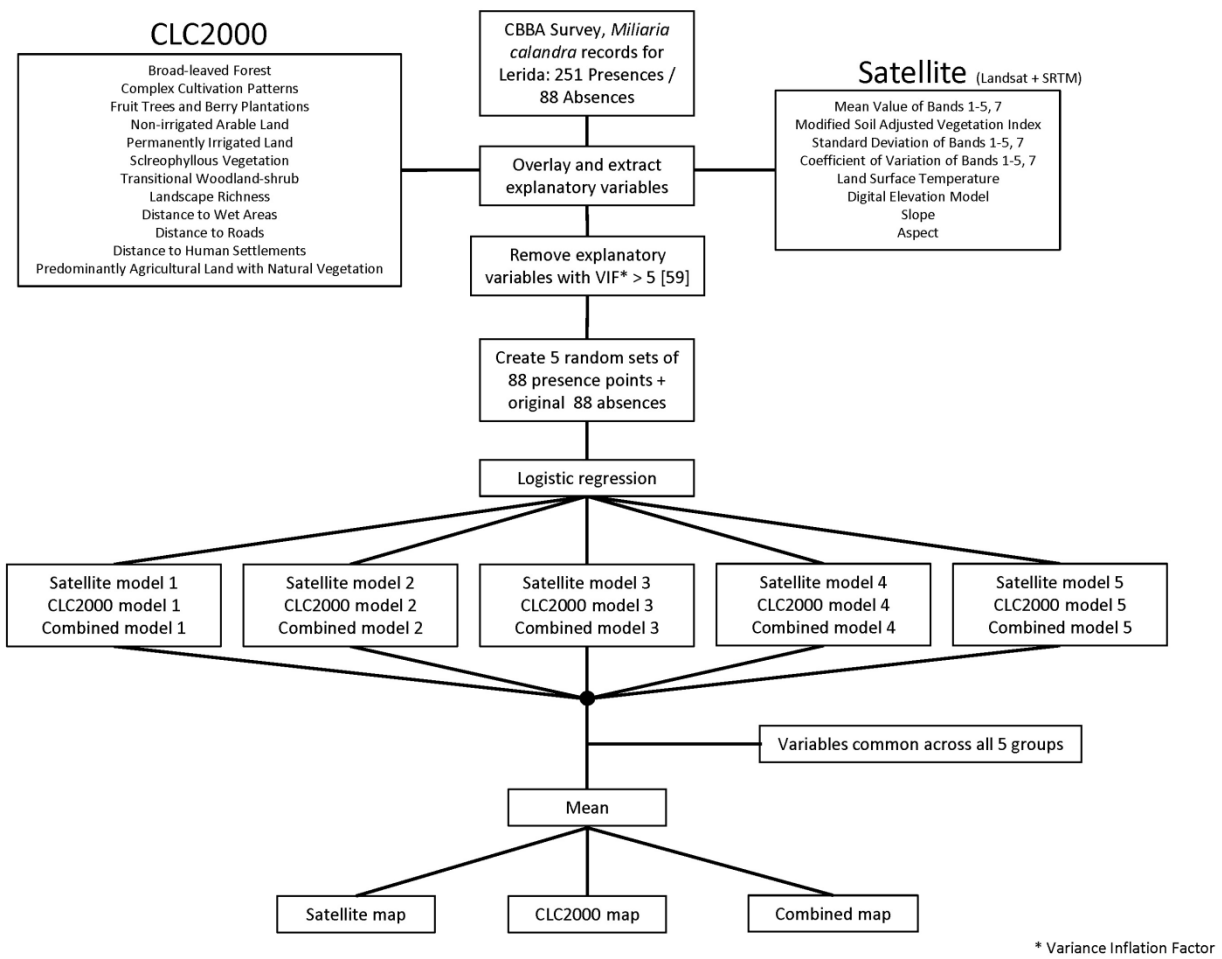
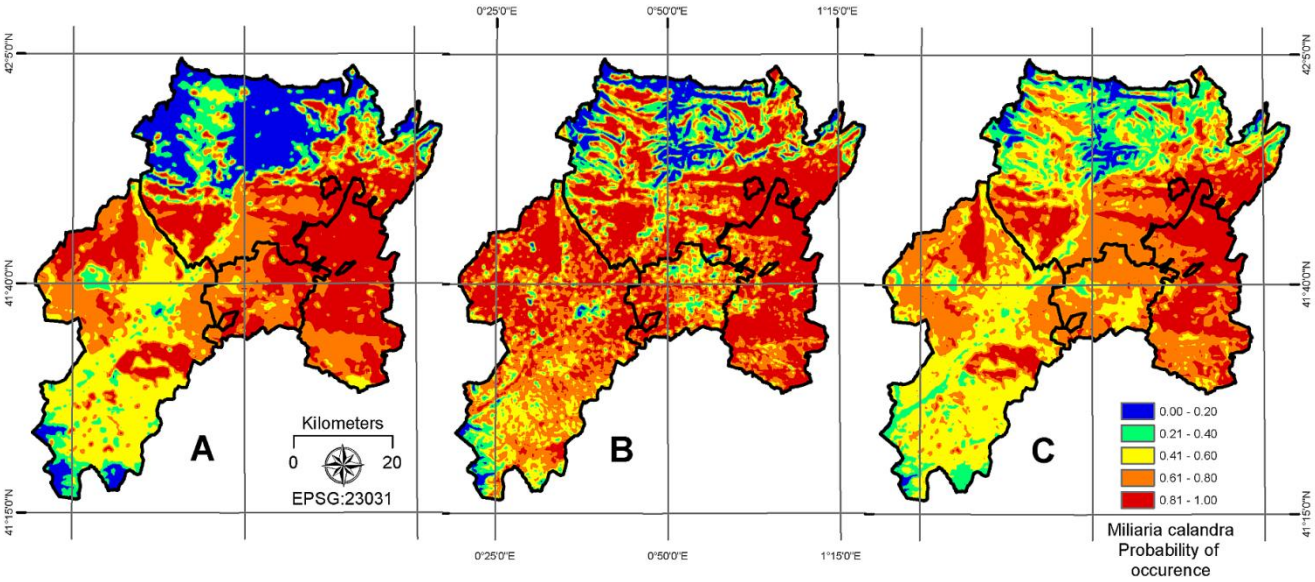


Figure 5. Mean probability maps according to the (**A**) CLC2000 (**B**) Satellite and (**C**) Combined models. Note the difference in texture between A and B owing to the coarser resolution of the CLC2000 data and the finer resolution of the combined Landsat and SRTM data. C produces a harmonized output that captures key variations of the previous two without compromising resolution.



4. Results

The CLC2000, Satellite and Combined model outputs for all five groups were averaged to produce three final maps (Figure 5). The mean AUC values of 0.69, 0.81 and 0.90 for the CLC2000, Satellite and the Combined Model, respectively, were above the random model threshold (0.50), indicating that the explanatory variables influence *M. calandra* habitat selection.

The mean logistic regression model for the CLC2000 group is summarized in Table 4. AUC value for this model was 0.69. The model assumes unfavorable habitat in the steeps slopes of higher altitudes, and there was a distinctive preference for the non-irrigated arable land ($p < 0.001$) landscape metric. The exclusion of non-irrigated arable land had the greatest effect on the model by increasing the AIC by an average of 43.56. In contrast, *M. calandra* exhibited reduced preference for areas that were predominantly composed of permanently irrigated land ($p = 0.002$), but the map displays favorability for grassy fringes, where permanently irrigated land meets habitats, such as non-irrigated arable land.

The mean logistic regression model for the Satellite group is summarized in Table 5. AUC value for this model was 0.81. Land surface temperature ($p < 0.001$), surface slope ($p < 0.001$) and modified soil adjusted vegetation index ($p < 0.001$) had the strongest influence on the independent variable. *M. calandra* is a ground nesting species and prefers uncultivated land; therefore, it is not surprising that land surface temperature would have a strong influence on habitat preference during the breeding season, as it is able to discriminate the thermal signature of dry, non-irrigated land. The correction factor in the modified soil adjusted vegetation index algorithm enhances the vegetation signal in areas with low vegetation density. The coefficient of variation of the ETM+ band 5 ($p = 0.002$) also exhibited a strong positive influence on *M. calandra* occurrence, because vegetation moisture content is

discernible in the shortwave infrared region between 1.55 and 1.75 microns, suggesting that texture features in the infrared region are likely to detect variation in vegetation structure. An interesting result was the negative relationship of *M. calandra* with the standard deviation of the ETM+ band 1 (p = 0.097); removal of this variable significantly increased the model AIC score. The negative relationship could be attributed to the fact that the spectral range of ETM+ band 1 captures the reflection of urban landscapes.

Table 4. Output of the mean CLC2000 model.

Variable	Coefficient	S.E.	Wald	*p*
(Intercept)	−0.703	0.456	−1.542	0.123
PANV	1.207	0.861	1.402	0.161
CCP	1.136	0.539	2.108	0.035
FTBP	1.291	0.613	2.106	0.035
NIAL	4.032	0.721	5.592	<0.001
PIL	2.395	0.557	4.300	0.002
SVEG	1.924	0.976	1.971	0.048
HUMDIST	−8e−05 *	5e05 **	−1.6e−10	0.091
WETDIST	−5e−05	2e05	−2.501	0.016
AIC	355	AUC	0.69	

* −1e−05 = −0.00001

** 1e05 = 100000

Table 5. Output of the mean Satellite model.

Variable	Coefficient	S.E.	Wald	*p*
(Intercept)	−12.741	3.004	−4.241	0.002
MSAVI	3.432	0.979	3.506	<0.001
BAND1SD	−0.093	0.056	−1.661	0.097
BAND5CV	5.255	1.735	3.029	0.002
DEM	0.003	0.001	3.001	0.011
SLOPE	−0.301	0.075	−4.013	<0.001
LST	0.286	0.071	4.028	<0.001
AIC	310	AUC	0.81	

The mean logistic regression model for the Combined group consisted of nine explanatory variables (Table 6, Figure 6) and had the highest AUC of 0.90. The comparative ROC plot in Figure 7 displays the effect of multiple factors not limited by data source on *M. calandra* habitat selection behavior. For example, the selection of the modified soil adjusted vegetation index indicates the effect of soil background as an important factor in habitat selection. Increased reflectance from the underlying soil could be caused by the operation of heavy machinery or other anthropogenic disturbances. The Combined model indicates that *M. calandra* avoids areas with steep slopes, areas near human activity and urban infrastructure and areas entirely composed of intensely irrigated land. The model also shows

that, during the breeding season, *M. calandra* habitat suitability is positively influenced by non-irrigated arable land and dry, open areas far from forest cover.

Table 6. Output of the mean Combined model. Bold letters in parentheses link each variable with the corresponding plot in Figure 6.

Variable	Coefficient	S.E.	Wald	*p*
(Intercept)	−12.16	3.495	−3.479	0.005
MSAVI (A)	3.287	0.883	3.723	0.002
LST (B)	0.328	0.095	3.453	<0.001
BAND5CV (C)	4.600	1.856	2.478	0.013
BAND1SD (D)	−0.141	0.060	−2.350	0.019
SLOPE (E)	−0.211	0.082	−2.573	0.010
BLF (F)	−0.002	0.001	−2.001	0.051
NIAL (G)	1.975	0.613	3.222	0.001
PIL (H)	1.658	0.685	2.420	0.015
HUMDIST (I)	−9e−05	5e−05	−1.801	0.048
AIC	285	AUC	0.90	

Figure 6. The nine explanatory variables common to the Combined model in all five groups: (**A**) Modified Soil Adjusted Vegetation Index; (**B**) Land Surface Temperature in ℃; (**C**) Coefficient of Variation of ETM+ Band 5; (**D**) Standard Deviation of ETM+ Band 1; (**E**) SRTM Slope in degrees; (**F**) Fraction of Broad-leaved Forest; (**G**) Fraction of Non-irrigated Arable Land; (**H**) Fraction of Permanently Irrigated Land; (**I**) Distance to Human Activity in meters.

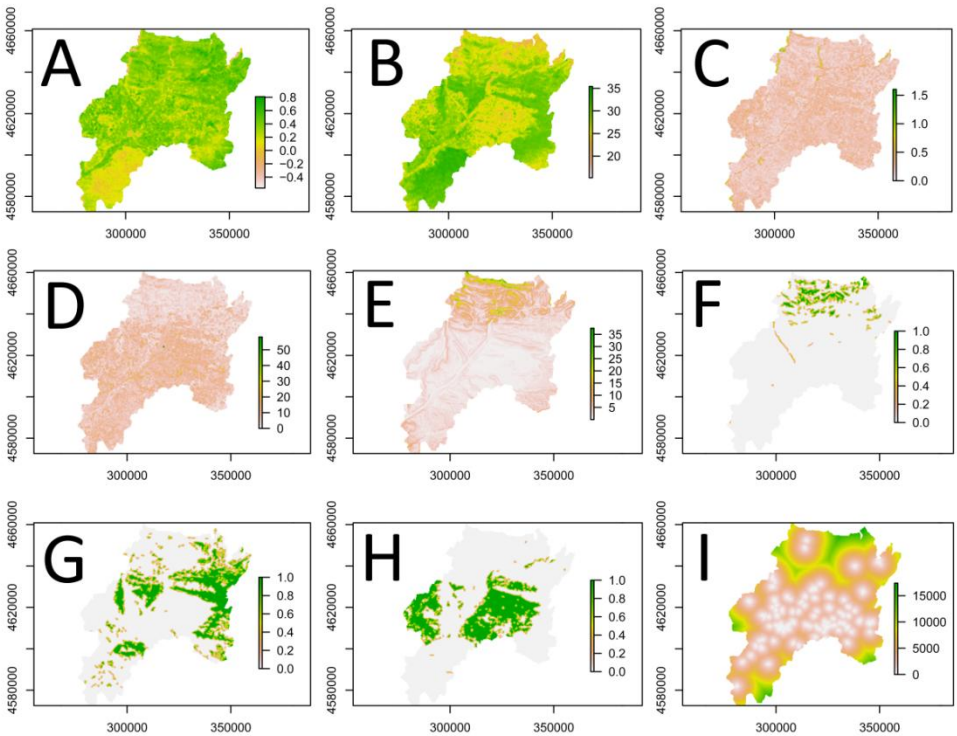
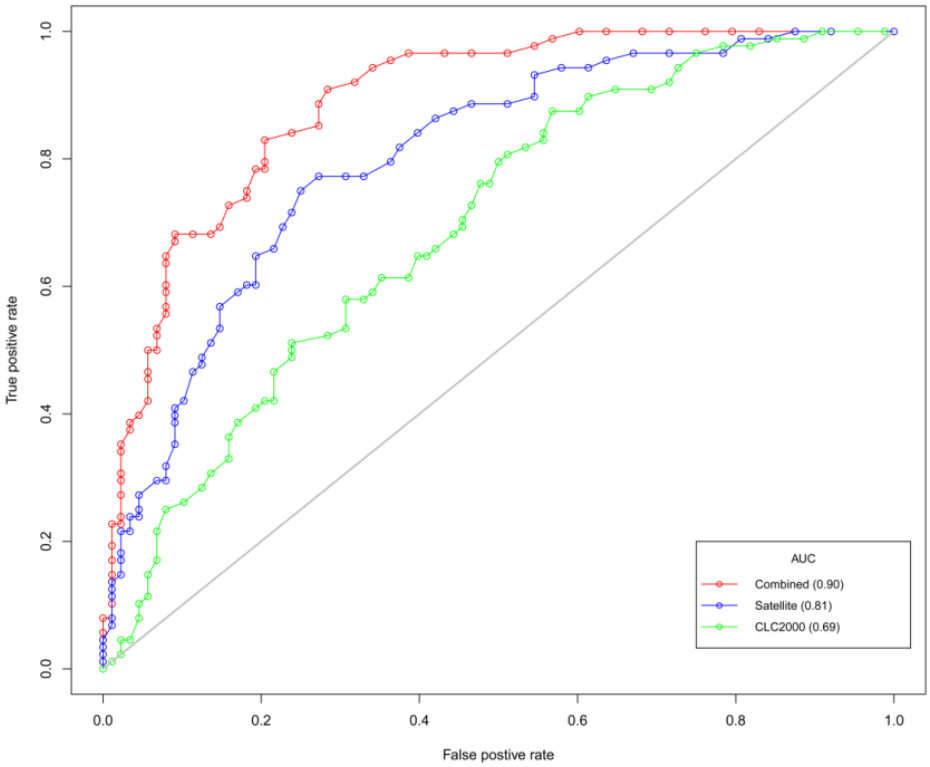


Figure 7. Comparative receiver operating characteristic (ROC) plot of the mean area under the curve (AUC) values for the five groups.



5. Discussion

Over the last forty years, changes in the European agricultural landscape have resulted in a continuously 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, and there is a need to monitor this decline using tools that are easily available, affordable and practical. The fusion of open access data, both biological and geophysical, from different sources is a viable method in producing models that reflect species' habitat preference.

Explanatory variables selected for this study were derived from open access geospatial data sources to assess *M. calandra* habitat suitability during the breeding season. Image texture represents the visual effect produced by the spatial distribution of tonal variability (pixel values) in a given area [62] and can serve as a substitute for habitat structure, because variability in the reflectance among adjacent pixels can be caused by horizontal variability in plant growth [37]. MSAVI possesses a correction factor that adjusts according to vegetation density, which has been shown to enhance the dynamic range of the vegetation signal and produce greater vegetation sensitivity [39]. Topography indirectly affects the distribution of species by modifying the relationships of birds with vegetation or by modifying the vegetation types [63]. Landscape metrics quantify specific spatial characteristics of patches, classes of patches or entire landscape mosaics from categorical map patterns [47] and help explain how spatial patterns of landscapes influence ecological processes [64]. Anthropogenic distance metrics are important measures for predicting bird assemblages in agricultural eco-regions [65], since some species prefer to breed in areas far from intensive human disturbance. Non-irrigated arable land is an extensive land cover class [66] that includes infrequently irrigated fallow land or areas under crop

rotation that are cultivated with cereals, legumes, fodder crops and root crops. On the other hand, permanently irrigated land is an intensive land cover class [66] that includes areas under constant irrigation using permanent infrastructure, such as irrigation channels and drainage networks. The varieties of crops grown on permanently irrigated land cannot be cultivated without an artificial water supply.

The mean Combined logistic regression model had an accuracy of 0.90 based on AUC. *M. calandra* had a strong positive correlation with LST, MSAVI, BAND5CV and non-irrigated arable land (NIAL). These parameters served as ecological surrogates in the modeling process. Land cover variables provided direct causal relationship with *M. calandra* distribution; for example, during the breeding season, the species prefers to nest in uncultivated land and avoids pastures, which explains the favorability of NIAL over permanently irrigated land (PIL). Pastures and intensified agricultural fields exhibit low temperature in summer breeding months, due to heavy irrigation. Here, LST proves useful in discriminating between intensely irrigated and non-irrigated land. However, urban environments also exhibit high LST, and the inclusion of BAND1SD, which is sensitive to the highly variable spectral signature of urban landscapes, provided additional information about habitat suitability.

Research on the declines of the four farmland bunting species in Europe generally involved a variety of statistical modeling of field-collected independent variables (e.g., the presence and absence of species [8] or the abundance and density [5]) and explanatory variables (e.g., crop type, land use [7], tree density and vegetation structure [21] or proportions of certain land use features [12]). The absence of geolocated components in these studies creates difficulty in understanding the spatial patterns of species distributions. In contrast, the logic behind the modeling approach utilized in this study is that the presence or absence of the species at a particular location is a function of the explanatory variables that represent the species' environment at that location. The use of variables derived from remote sensing and land cover data as proxies predicting avian habitat suitability is not a new field of research, but previous studies have not taken full advantage of the information contained within open access remote sensing data. For example, in the study by Kosicki and Chylarecki [6], NDVI was the only biophysical variable extracted from remote sensing data. Previous studies on *M. calandra* habitat association that did not use open access geospatial data have yielded results similar to the outcomes of this study; Brambilla *et al.* [22] found that the proportion of arable land positively influenced the habitat preference of breeding *M. calandra*, while Stoate *et al.* [21] found that 'treeless cereal cropland' supported the highest breeding densities of the species.

The data used in this study were provided by the ICO, and the procurement of biodiversity data from such sources often has to undergo a request-and-decision process that results in approval or rejection, depending on the quality of the submitted proposals, the affiliation and academic level of the requester, the cooperativeness of the data-holding institution and the sensitivity of the data (e.g., IUCN Red List species). Closed-access data, such as the CBBA, are often of a higher scientific quality and are more extensive than open access, readily available data, such as GBIF and AKN/eBird (Figure 2). During the process of developing a methodology for this project, and prior to gaining permission to use ICO data, the author had experienced either rejections or no responses to requests for data access from several European organizations. Hence, the scope of the project and the methodology had to be constrained. It is thus imperative for conservation organizations with large high quality

scientific data to invest in web-based infrastructure that facilitates access to biodiversity data, so that the research community, particularly those with limited finances, can use them.

6. Conclusion

The development of distribution maps that consist of information from both Earth observation and land cover datasets are of importance for species that have indeterminate ranges [67] and for monitoring the spatial dynamics of threatened species. This data-fusion approach helps in the identification and maintenance of important habitats as the Bird Directive stipulates and, in combination with readily accessible open access biodiversity data, also facilitates the rapid delivery of environmental information to decision makers.

Acknowledgments

This work is a part of a graduate thesis conducted by the author and supported by the European Commission, Erasmus Mundus Program, Master of Science in Geospatial Technologies, Framework Partnership Agreement 2007-0064/001 FRAME MUNB123. The author would like to thank the Catalan Ornithological Institute for allowing access to the bird data and to Lluís Brotons of the Forest Technology Center of Catalonia for assisting in its acquisition. The author is also grateful for the support of the Institute of Statistics and Information Management at Universidade Nova de Lisboa and the Institute for Geoinformatics at Universität Münster. Finally, the two anonymous reviewers provided very constructive and helpful comments that improved the manuscript.

Conflict of Interest

The author declares no conflict of interest.

References

1. Birdlife-International. *Birds in the European Union: A Status Assessment*, 1st ed.; Birdlife International: Wageningen, The Netherlands, 2004.
2. Deceuninck, B. *The Corncrake (Crex crex) in France*; Schaeffer, N., Mammen, U., Eds.; International Corncrake Workshop: Hilpoltstein, Germany, 1998.
3. De Leo, G.A.; Focardi, S.; Gatto, M.; Cattadori, I.M. The decline of the grey partridge in Europe: Comparing demographies in traditional and modern agricultural landscapes. *Ecol. Model.* **2004**, *177*, 313–335.
4. Tucker, G.M.; Heath, M.F. *Birds in Europe: Their Conservation Status*; BirdLife International: Cambridge, UK, 1994.
5. Vepsäläinen, V.; Pakkala, T.; Piha, M.; Tiainen, J. Population crash of the ortolan bunting *Emberiza hortulana* in agricultural landscapes of southern Finland. *Ann. Zoologici. Fennici* **2005**, *42*, 91–107.
6. Kosicki, J.Z.; Chylarecki, P. Habitat selection of the Ortolan bunting *Emberiza hortulana* in Poland: Predictions from large-scale habitat elements. *Ecol. Res.* **2012**, *27*, 347–355.

7. Evans, A.; Smith, K. Habitat selection of Cirl Buntings *Emberiza cirlus* wintering in Britain. *Bird Study* **1994**, *41*, 81–87.
8. Wotton, S.R.; Langston, R.H.W.; Gibbons, D.W.; Pierce, A.J. The status of the Cirl Bunting *Emberiza cirlus* in the UK and the Channel Islands in 1998. *Bird Study* **2000**, *47*, 138–146.
9. Stoate, C.; Moreby, S.J.; Szczur, J. Breeding ecology of farmland Yellowhammers *Emberiza citrinella*. *Bird Study* **1998**, *45*, 109–121.
10. Golawski, A.; Dombrowski, A. Habitat use of Yellowhammers *Emberiza citrinella*, Ortolan Buntings *E. hortulana*, and Corn Buntings *Miliaria calandra* in farmland of east-central Poland. *Ornis Fennica* **2002**, *79*, 164–172.
11. Wretenberg, J.; Lindström, Å.; Svensson, S.; Pärt, T. Linking agricultural policies to population trends of Swedish farmland birds in different agricultural regions. *J. Appl. Ecol.* **2007**, *44*, 933–941.
12. Whittingham, M.J.; Swetnam, R.D.; Wilson, J.D.; Chamberlain, D.E.; Freckleton, R.P. Habitat selection by yellowhammers *Emberiza citrinella* on lowland farmland at two spatial scales: Implications for conservation management. *J. Appl. Ecol.* **2005**, *42*, 270–280.
13. Brickle, N.; Harper, D.; Aebischer, N.; Cockayne, S. Effects of agricultural intensification on the breeding success of corn buntings *Miliaria calandra*. *J. Appl. Ecol.* **2000**, *37*, 742–755.
14. Orlowski, G. Endangered and declining bird species of abandoned farmland in south-western Poland. *Agric. Ecosyst. Environ.* **2005**, *111*, 231–236.
15. Eisloeffel, F. The Corn Bunting (Miliaria calandra) in South-West Germany: Population Decline and Habitat Requirements. In *The Ecology and Conservation of Corn Buntings (Miliaria calandra)*; Donald, P.F., Aebischer, N.J., Eds.; Joint Nature Conservation Committee: Peterborough, UK, 1997; pp. 170–173.
16. Taylor, A.J.; O'Halloran, J. The decline of the Corn Bunting *Miliaria calandra*, in the Republic of Ireland. *Biology Environ. Proc. Royal Irish Acad.* **2002**, *102*, 165–175.
17. Donald, P.F.; Green, R.E.; Heath, M.F. Agricultural intensification and the collapse of Europe's farmland bird populations. *Proc. Royal Soc. B* **2001**, *268*, 25–29.
18. Fuller, R.M.; Devereux, B.J.; Gillings, S.; Amable, G.S. 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 Ecol. Biogeogr.* **2005**, *14*, 223–239.
19. Kuczynski, L.; Antczak, M.; Czechowski, P.; Grzybek, J.; Jerzak, L.; Zablocki, P.; Tryjanowski, P. A large scale survey of the great grey shrike *Lanius excubitor* in Poland: Breeding densities, habitat use and population trends. *Ann. Zoologici Fennici* **2010**, *47*, 67–78.
20. Diaz, M.; Telleria, J.L. Habitat Selection and Distribution Trends of Corn Buntings in the Iberian Peninsula. In *The Ecology and Conservation of Corn Buntings Miliaria Calandra*; Donald, P., Aebischer, N.J., Ed.; JNCC: Peterborough, UK, 1997; pp. 151–161.
21. Stoate, C.; Borralho, R.; Araujo, M. Factors affecting corn bunting *Miliaria calandra* abundance in a Portuguese agricultural landscape. *Agric. Ecosyst. Environ.* **2000**, *77*, 219–226.
22. Brambilla, M.; Guidali, F.; Negri, I. Breeding-season habitat associations of the declining Corn Bunting *Emberiza calandra*—A potential indicator of the overall bunting richness. *Ornis Fennica* **2009**, *86*, 41–50.

23. Brotons, L.; Manosa, S.; Estrada, J. Modelling the effects of irrigation schemes on the distribution of steppe birds in Mediterranean farmland. *Biodivers. Conserv.* **2004**, *13*, 1039–1058.
24. Moreno-Mateos, D.; Pedrocchi, C.; Comin, F.A. Avian communities' presence in recently created agricultural wetlands in irrigated landscapes of semi-arid areas. *Biodivers. Conserv.* **2009**, *18*, 811–828.
25. Reeves, H.M.; Cooch, F.G.; Munro, R.E. Monitoring arctic habitat and goose production by satellite imagery. *J.Wildl. Manag.* **1976**, *40*, 532–541.
26. Cannon, R.W.; Knopf, F.L.; Pettinger, L.R. Use of Landsat data to evaluate Lesser Prairie Chicken habitats in western Oklahoma. *J. Wildl. Manag.* **1982**, *46*, 915–922.
27. Lauver, C.L.; Whistler, J. A hierarchical classification of Landsat TM imagery to identify natural grassland areas and rare species habitat. *Photogramm. Eng. Remote Sens.* **1993**, *59*, 627–634.
28. Shirley, S.; Yang, Z.; Hutchinson, R.; Alexander, J.; McGarigal, K.; Betts, M. Species distribution modelling for the people: Unclassified landsat TM imagery predicts bird occurrence at fine resolutions. *Divers. Distrib.* **2013**, *19*, 855–866.
29. Houska, T.R.; Johnson, A.P. *GloVis Ver. 8.17.1.*; US Geological Survey General Information Product Series 137; Earth Resources Observation and Science (EROS) Center: Reston, VA, USA, 2012.
30. Goslee, S.C. Analyzing remote sensing data in R: The Landsat package. *J. Stat. Softw.* **2011**, *43*, 1–25.
31. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2009.
32. Rabus, B.; Eineder, M.; Roth, A.; Bamler, R. The shuttle radar topography mission—A new class of digital elevation models acquired by spaceborne radar. *ISPRS J. Photogramm. Remote Sens.* **2003**, *57*, 241–262.
33. Oliver, T.; Roy, D.B.; Hill, J.K.; Brereton, T.; Thomas, C.D. Heterogeneous landscapes promote population stability. *Ecol. Lett.* **2010**, *13*, 473–484.
34. Jarvis, A.; Reuter, H.; Nelson, A.; Guevara, E. Hole-Filled SRTM for the Globe Version 4. In CGIAR-CSI SRTM 90 m Database. Available online: http://srtm.csi.cgiar.org (accessed on 12 November 2009).
35. Büttner, G.; Feranec, J.; Jaffrain, G.; Mari, L.; Maucha, G.; Soukup, T. The CORINE land cover 2000 project. *EARSeL eProc.* **2004**, *3*, 331–346.
36. Saveraid, E.H.; Debinski, D.M.; Kindscher, K.; Jakubauskas, M.E. A comparison of satellite data and landscape variables in predicting bird species occurrences in the Greater Yellowstone Ecosystem, USA. *Landsc. Ecol.* **2001**, *16*, 71–83.
37. St-Louis, V.; Pidgeon, A.M.; Clayton, M.K.; Locke, B.A.; Bash, D.; Radeloff, V.C. Satellite image texture and a vegetation index predict avian biodiversity in the Chihuahuan Desert of New Mexico. *Ecography* **2009**, *32*, 468–480.
38. Horning, N.; Robinson, J.; Sterling, E.; Turner, W.; Spector, S. *Remote Sensing for Ecology and Conservation*; Oxford University Press: Oxford, UK, 2010; pp. 108–111.
39. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sens. Environ.* **1994**, *48*, 119–126.
40. Lambin, E.F.; Ehrlich, D. The surface temperature-vegetation index space for land cover and land-cover change analysis. *Int. J. Remote Sens.* **1996**, *17*, 463–487.

41. Quattrochi, D.A.; Luvall, J.C. Thermal infrared remote sensing for analysis of landscape ecological processes: Methods and applications. *Landsc. Ecol.* **1999**, *14*, 577–598.
42. Kerr, J.T.; Ostrovsky, M. From space to species: Ecological applications for remote sensing. *Trends Ecol. Evol.* **2003**, *18*, 299–305.
43. Sobrino, J.A.; Jiménez-Muñoz, J.C.; Paolini, L. Land surface temperature retrieval from Landsat TM 5. *Remote Sens. Environ.* **2004**, *90*, 434–440.
44. NASA. Band 6 Conversion to Temperature. In *Landsat 7 Science Data Users Handbook*; National Aeronautics and Space Administration: Greenbelt, MD, USA, 2009; pp. 117–120.
45. Gustafson, E.J.; Parker, G.R.; Backs, S.E. Evaluating spatial pattern of wildlife habitat: A case study of the wild Turkey (Meleagris gallopavo). *Am. Midl. Nat.***1994**, *131*, 24–33.
46. Fauth, P.T.; Gustafson, E.J.; Rabenold, K.N. Using landscape metrics to model source habitat for Neotropical migrants in the Midwestern US. *Landsc. Ecol.* **2000**, *15*, 621–631.
47. De Smith, M.J.; Goodchild, M.F.; Longley, P.A. Grid-based Statistics. In *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*; Troubador Publishing Ltd.: Leicester, UK, 2007; pp. 194–199.
48. Vander Haegen, W.M.; Dobler, F.C.; Pierce, D.J. Shrubsteppe bird response to habitat and landscape variables in Eastern Washington, USA. *Conserv. Biol.* **2000**, *14*, 1145–1160.
49. Edwards, J.L.; Lane, M.A.; Nielsen, E.S. Interoperability of biodiversity databases: Biodiversity information on every desktop. *Science* **2000**, *289*, 2312–2314.
50. Sullivan, B.L.; Wood, C.L.; Iliff, M.J.; Bonney, R.E.; Fink, D.; Kelling, S. eBird: A citizen-based bird observation network in the biological sciences. *Biol. Conser.* **2009**, *142*, 2282–2292.
51. Brotons, L.; Herrando, S.; Estrada, J.; Pedrocchi, V.; Martin, J.L. The Catalan Breeding Bird Atlas (CBBA): Methodological aspects and ecological implications. *Revista Catalana d'Ornitologia* **2008**, *24*, 118–137.
52. Keitt, T.H.; Bivand, R.; Pebesma, E.; Rowlingson, B. *Rgdal: Bindings for the Geospatial Data Abstraction Library. R Package Version 0.6–25. 2010*. Available online: http://cran.r-project.org/web/packages/rgdal/index.html (accessed on 25 February 2010),
53. Lindsay, J. *Whitebox Geospatial Analysis Tools*; Department of Geography, University of Guelph: Guelph, Ontario, Canada, 2009. Available online: http://www.uoguelph.ca/~hydrogeo/Whitebox/ (accessed on 29 December 2009).
54. Cimmery, V. *User Guide for SAGA. Version 2.0, 2007*. Available online: http://saga-gis.org/ (accessed on 15 January 2010).
55. Gagné, M. *Moving to Ubuntu Linux*; Pearson Education: Upper Saddle River, NJ, USA, 2006.
56. Hosmer, D.W.; Lemeshow, S. *Applied Logistic Regression*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 2000.
57. Akaike, H. *Information Theory as an Extension of the Maximum Likelihood Principle*; Petrov, B.N.; Csaki, F., Eds.; Akademiai Kiado: Budapest, Hungary, 1973; pp. 267–281.
58. Brauner, N.; Shacham, M. Role of range and precision of the independent variable in regression of data. *Am. Inst. Chem. Eng. J.* **1998**, *44*, 603–611.
59. O'Brien, R.M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quantity* **2007**, *41*, 673–690.
60. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychological Meas.* **1960**, *20*, 37–46.

61. Deleo, J.M. Receiver Operating Characteristic Laboratory (ROCLAB): Software for Developing Decision Strategies that Account for Uncertainity. In Proceedings of Second International Symposium on Uncertainty Modelling and Analysis, College Park, MD, 25–28 April 1993; IEEE Computer Society Press: College Park, MD, USA, 1993; pp. 318–325.
62. Baraldi, A.; Parmiggiani, F. An investigation of the textural characteristics associated withgray level cooccurrence matrix statistical parameters. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 293–304.
63. Seoane, J.; Bustamante, J.; Diaz-Delgado, R. Competing roles for landscape, vegetation, topography and climate in predictive models of bird distribution. *Ecol. Model.* **2004**, *171*, 209–222.
64. Carrao, H.; Caetano, M. The Effect of Scale on Landscape Metrics. In Proceedings of International Symposium of Remote Sensing of the Environment, Buenos Aires, Argentina, 8–12 April 2002.
65. Whited, D.; Galatowitsch, S.; Tester, J.R.; Schik, K.; Lehtinen, R.; Husveth, J. The importance of local and regional factors in predicting effective conservation: Planning strategies for wetland bird communities in agricultural and urban landscapes. *Landsc. Urban. Plann.* **2000**, *49*, 49–65.
66. Martí-Raguè, X.; Lescrauwaet, A.-K.; Borg, M.; Valls, M. *Indicators Guidelines: To Adopt an Indicators-Based Approach to Evaluate Coastal Sustainable Development*; Government of Catalonia: Barcelona, Spain, 2007.
67. Seoane, J.; Bustamante, J.; Diaz-Delgado, R. Are existing vegetation maps adequate to predict bird distributions? *Ecol. Model.* **2004**, *175*, 137–149.

Appendix

Table A1. Variance inflation factor (VIF) values for the nine explanatory variables. VIF measures the increase of variance in a regression coefficient that is caused by multicollinearity (high correlation of two of more explanatory variables).

Explanatory Variable	VIF
MSAVI	3.820
BAND1SD	0.095
BAND5CV	2.850
SLOPE	0.165
LST	0.283
BLF	0.438
NIAL	2.061
PIL	2.051
HUMDIST	4.3e−5

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).