

Bird Diversity Modeling Using Geostatistics and GIS

Abdulkhakim Abdi* and Anand Nandipati

*Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa, Campus de
Campolide, Lisbon 1070-312, Portugal

Email: m2008149@isegi.unl.pt

ABSTRACT

Portugal has a diverse landscape that provides a variety of habitats for birds. Many species of birds that breed in Portugal migrate south in the autumn and return in the spring. Portugal also serves as a wintering ground for several northern species. Since migrating bird species sometimes cover long distances and several habitats, they can serve as indicators of the overall health of the environment. This study investigate the possibility of using geostatistics to predict the diversity of bird species over a given area and using these predictions as a basis for the conservation of ecosystems. The study involves the use of 2007 data from the Portuguese common bird census conducted by Sociedade Portuguesa para o Estudo das Aves (SPEA) and using European Environmental Agency's CORINE Land Cover classification. Geostatistical interpolation techniques were used to build prediction models. GIS was used to extract both habitat information and the predicted diversity values. The results showed the habitat preferences based on the predictions. The combination of diversity modeling and habitat characterization can be used to aid conservation efforts in identifying key habitats of importance for each category. The study will also conduct comparative analysis of the interpolators and an assessment of their accuracy.

Keywords: bird migration, biodiversity, conservation, kriging, inverse distance weighting, habitat characterization, breeding birds, CORINE Land Cover.

1 INTRODUCTION

The 2002 Rio de Janeiro World Summit of Sustainable Development set a goal “to achieve by 2010 a significant reduction in the current rate of loss of biological diversity”. Therefore, in order to achieve this objective there is a need to identify species that are at risk or in deteriorating numbers. Breeding bird species occupy a variety of habitats, many of them created and/or altered by humans, and thus have the potential to serve as indicators of the sustainability of land use activities. Similarly, migrant birds react to changes in environmental conditions by making regional alteration to their migration routes and stopovers (Tankersley, 2004). Furthermore, these environmental conditions may correspond to rainfall patterns, food availability (Sanderson et al, 2006) and habitat degradation.

BirdLife International¹ held its World Conference in Buenos Aires in 2008 and launched a “State of the World’s Birds” publication which declared that “common birds are in decline across the world, providing evidence of a rapid deterioration in the global environment that is affecting all life on earth” (BirdLife International, 2008).

¹ BirdLife International is a global Partnership of conservation organisations that strives to conserve birds, their habitats and global biodiversity, working with people towards sustainability in the use of natural resources.

1.1 Study Rationale

Europe serves as a breeding ground for many Afro-Palaearctic species and studies have shown that many of these species are in decline due to human pressure (Sanderson et al, 2006). This decline is solely due to the propagation of human-induced activities; chief among them is habitat loss and fragmentation. The latter term refers to the condition of a landscape after it is divided into solitary leftover plots.

A partnership between BirdLife International and the European Bird Census Council (EBCC) established the Pan-European Common Bird Monitoring Scheme (PECBMS) in January 2002 to investigate the application of breeding bird population dynamics as biodiversity indicators in Europe. The focus of the PECBMS has been breeding bird populations in agricultural land and woodland (as illustrated by Figure 1) because both are greatly influenced by human activity and because approximately 44% of the rural European landscape is agricultural (Telleria et al, 2003) and 30% consists of temperate and boreal forest (Tucker & Evans, 1997).

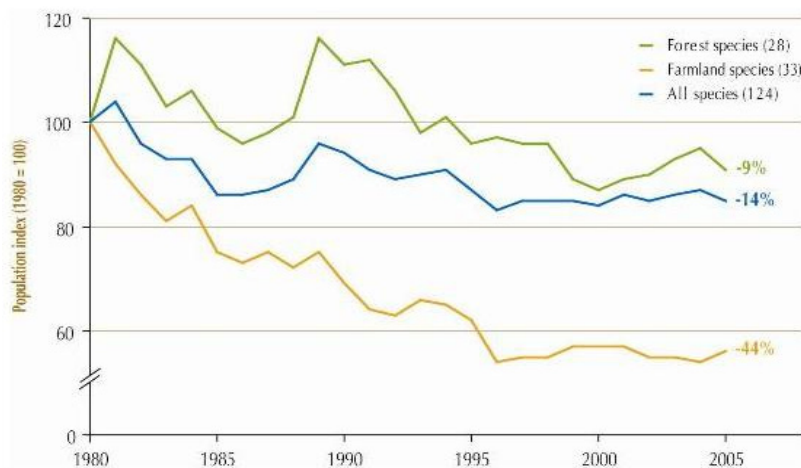


Figure 1: The decline of European common birds (BirdLife International, 2008).

Portugal is of importance because it hosts many of the common species that reside in the aforementioned habitats. Traditionally farmed landscapes often consist of complex and extensive mosaics with high levels of biodiversity (Wolff et al, 2001). Traditional Portuguese farming which focuses on low-input agriculture is steadily abandoned for more intensified methods due to high economic demand.

1.1.1 Objective

The purpose of this study is to compare the interpolation methods available in ArcGIS in the modelling of bird species richness and to overlay CORINE land cover classes with the interpolation results in order to extract habitat preference from each interpolated point. Using root mean square error, the accuracy of the interpolators will be determined. It is hoped that this will enable the prediction of diversity of species across habitats and aid in formulating conservation measures. An important note to consider is that this study uses data based on a set collection method, established by the providing agency (see Section 2), in which the authors have no control.

1.2 Study Area

The study area is located south of the city of Lisbon, Portugal, in the Distrito of Setúbal (as illustrated by Figure 2). Under the 1973 Hayford Gauss IPCC projection, it lies between latitudes $38^{\circ}40'47.744''\text{N}$ and $38^{\circ}24'26.917''\text{N}$ and between longitudes $9^{\circ}13'26.79''\text{W}$ and $8^{\circ}58'15.33''\text{W}$. The area encompasses 333.82 square kilometres and lies in four Portuguese 10km by 10km National UTM Grid squares: MC85, MC86, MC87 and MC95. This area was chosen because of the contiguity of the UTM squares and the consistent temporal distribution of the data.

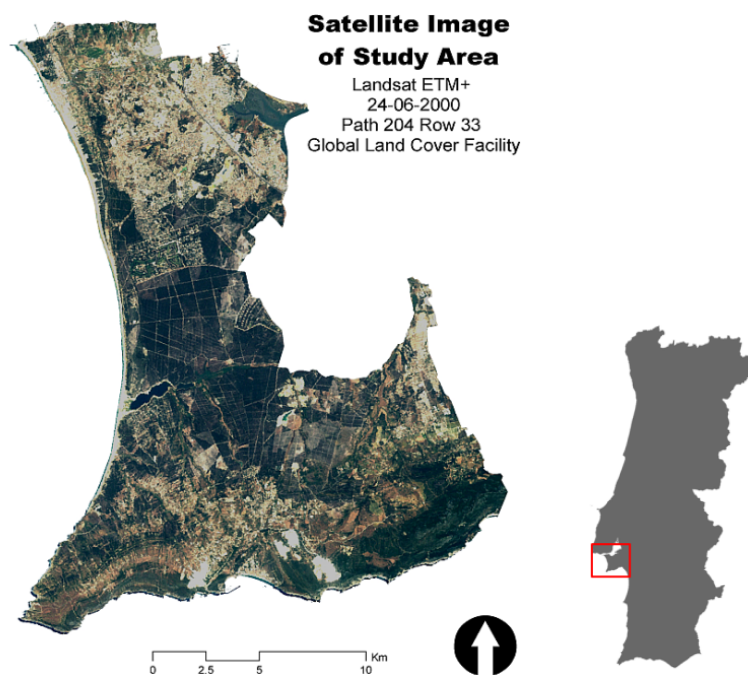


Figure 2: Satellite image of the study area.

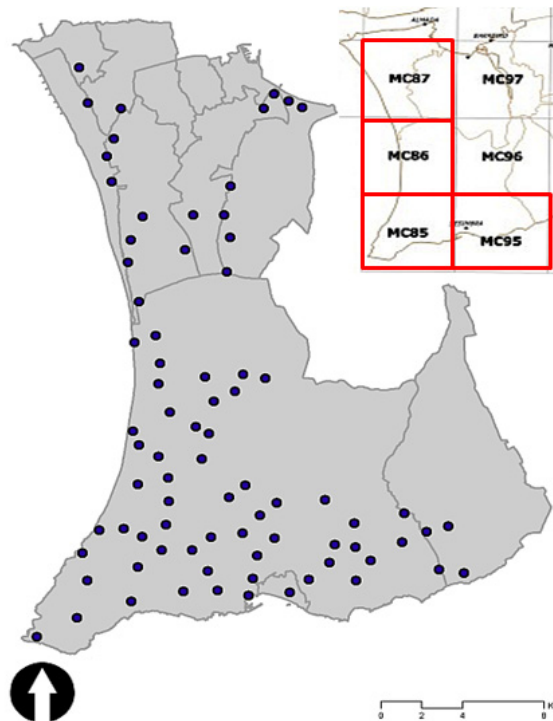


Figure 3: Overview of the study area with bird count locations.

2 METHODOLOGY

Material for this study was provided by Sociedade Portuguesa para o Estudo das Aves (SPEA) as part of the organization's Censo de Aves Comuns (Common Bird Census) which is the Portuguese contribution to the PECBMS. Data included twenty habitat characterizations and point counts in each 10km UTM square that spans the study area (Figure 3). Quality control measures for data collection established by SPEA involved:

- A compulsory greater-than one kilometre spacing between observation points,
- The designation of two recording distances (< 25m and > 25m) to avoid double counting,
- A representative proportionality of major habitat types between the UTM square and each point observation, and
- The non-placement of observation points at habitat transition boundaries.

Two visits (in this study: April 15th and May 18th) on clear days were conducted per year with each visit spanning between dawn and 11:00 hours. For consistency, the same predetermined points were used for each visit.

Initial data was provided in the form of two Microsoft Excel files containing data on the point counts (xlsA) and habitat characterization (xlsB) for each location (as illustrated by Figure 4).

Software used in the study includes Microsoft Excel, GeoMS and ArcGIS.

Year	UTM	Visit	Point	Species	Sp<25m	Sp>25m	Year	Point	UTM	D_HABP1	D_HABP2	D_HABP3	D_HABP4	D_HABP5	
2004	MC85	1	1	16530	0	4	2004	1	MC85	C		4	1	1	1
2004	MC85	1	1	6700	0	1	2004	1	MC87	A		1	1	1	2
2004	MC85	1	1	10010	0	1	2004	2	MC85	A		9	1	2	2
2004	MC85	1	1	15910	3	1	2004	2	MC87	A		1	1	1	2
2004	MC85	1	2	15670	0	2	2004	3	MC85	D		3	2	1	1
2004	MC85	1	2	15910	5	2	2004	3	MC87	A		1	1	1	2
2004	MC85	1	2	11870	0	1	2004	4	MC85	C		4	2	1	1
2004	MC85	1	2	32767	0	1	2004	4	MC87	E		2	1	2	2
2004	MC85	1	3	15910	1	4	2004	5	MC85	A		11	2	1	1
2004	MC85	1	3	6870	1	0	2004	5	MC87	A		1	1	1	2
2004	MC85	1	3	11870	0	1	2004	6	MC85	E		2	1	2	2
2004	MC85	1	3	32767	0	1	2004	6	MC87	A		1	1	1	2
2004	MC85	1	4	16530	0	2	2004	7	MC85	A		4	1	1	2
2004	MC85	1	4	6700	0	2	2004	7	MC87	A		1	1	1	2
2004	MC85	1	4	10010	0	3	2004	8	MC85	A		9	2	1	2
2004	MC85	1	4	15910	4	7	2004	8	MC87	A		1	1	1	2
							2004	9	MC85	C		4	2	1	1
							2004	9	MC87	A		1	1	1	2
							2004	10	MC85	A		4	2	1	2

Figure 4: Species point count data (xlsA) and habitat codes (xlsB).

2.1 Preprocessing

The data had to undergo preprocessing in order to make it ready for analysis. The files that were provided lacked the format necessary for input into the GeoMS and ArcGIS software. Therefore, the files were organized in a manner that would preserve necessary information while discarding irrelevant ones. The first step was to organize the data into a consistent temporal format. The year 2007 was selected henceforth because both the observations in that year covered all four squares. The next step was to aggregate the observed species for each point; the underlying principle is to get the total number of species per observation point which would be a measure of the diversity. The data was then split into two files based on the months of the visits (April and May) and then converted into a *.dbf format. Furthermore, in order to use the data in GeoMS, duplicate files were created for each visit that each contained three columns: X-coordinates, Y-coordinates and Species Diversity.

The location data was provided in two formats; high resolution, scanned topographic maps of each square and moderate resolution, scanned field topographic maps. The latter were used as a reference base since they contained the actual points, the former were used as digitizing surface since they were identical to the field maps. The images were georeferenced to an outline of the study area downloaded from the Portuguese Geographic Institute.

All points were digitized as accurately as possible and in the sequence of the original visit (as illustrated by Figure 5). The XY coordinates of each point were obtained post-digitization. The points overlaid on the georeferenced topographic images (as illustrated by Figure 6).

Each point was given a unique ID which includes the point number and the square to which it belongs in order to avoid confusion since the number sequence for each square was identical.



Figure 5: Digitizing the observation points.

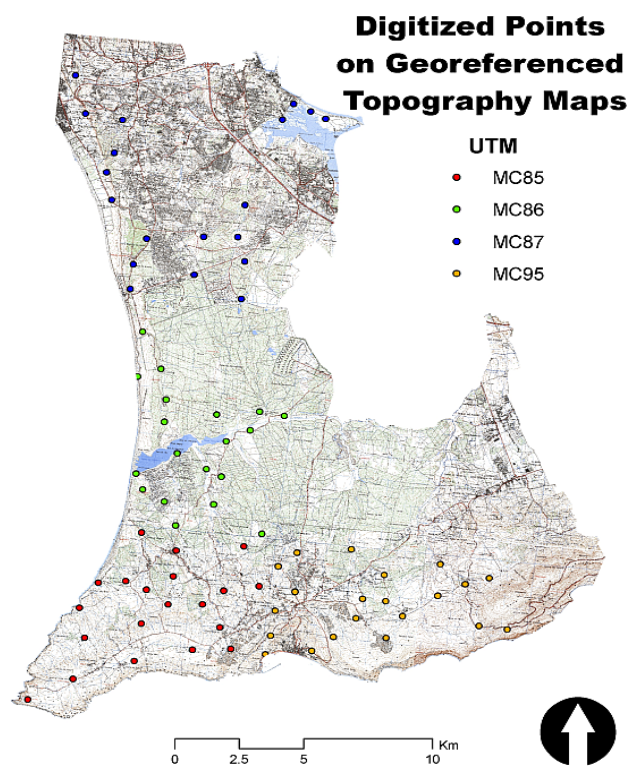


Figure 6: All the digitized points presented on the georeferenced topographic maps in a color-coded format that represents the four UTM squares.

2.2 Exploratory Data Analysis

Prior to employing geostatistical analysis, the data had to undergo Exploratory Data Analysis (EDA). The purpose of the EDA is the inspection of the data properties for reasons of detection of

patterns in the dataset. GeoMS was employed to explore the data (as illustrated by Figure 7 and 8). In figures below both histograms display positive skewness.

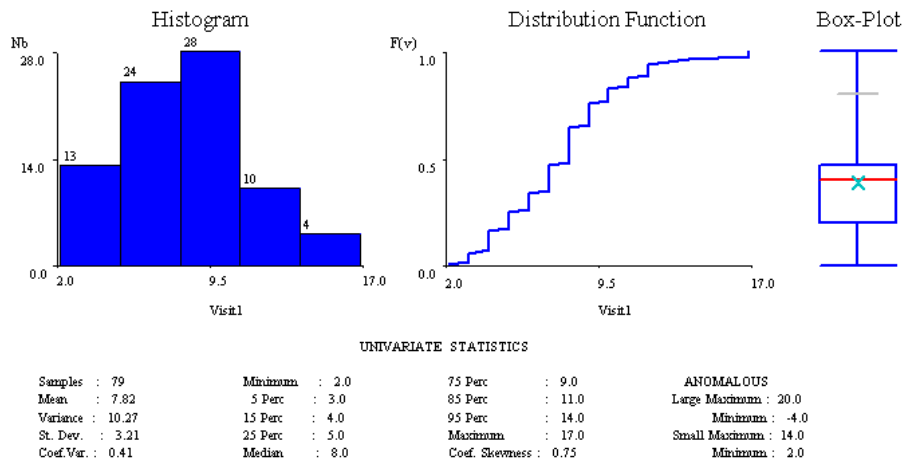


Figure 7: EDA of the first visit.

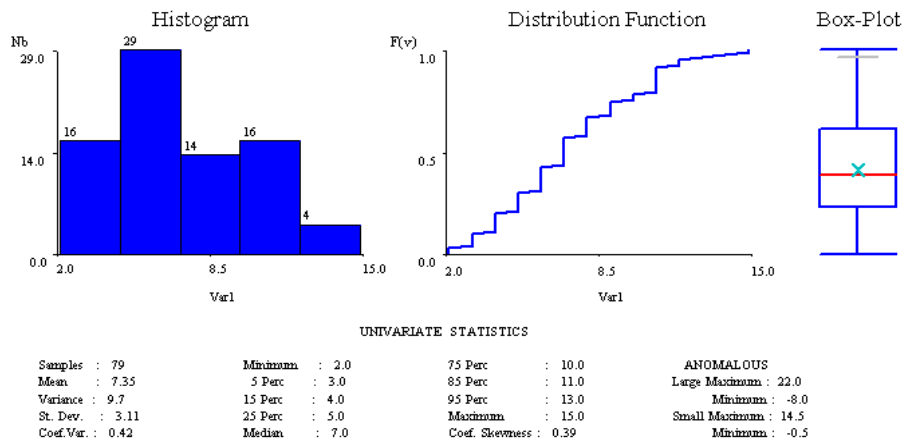


Figure 8: EDA of the second visit.

2.3 Geostatistical Interpolation

Two interpolators were used in this study; Inverse Distance Weighting and Ordinary Kriging. Inverse Distance Weighting is a deterministic method which works on the principle of identification of a neighborhood around the point to be interpolated and a weighted average is calculated of the values within the neighborhood, the weights decrease with distance. Ordinary Kriging is similar to Inverse Distance Weighting in that it calculates the surrounding measured values to derive a

prediction. However, unlike Inverse Distance Weighting, the weighted calculations are not based solely on the distance but also on the overall spatial composition between the measured points. In order for the spatial composition in the weights to be used, the spatial autocorrelation must be quantified. This means that a statistical connection between values at sampled points is ascertained and this connection is then employed to make predictions concerning unsampled points.

2.3.1 Inverse Distance Weighting

The Inverse Distance Weighting (IDW hereafter) function of the Geostatistical Analyst extension in ArcMap was used to produce the interpolated surfaces. The ten nearest neighbours and a power value of 2 were used in the creation of the interpolated surface (as illustrated by Figure 9). Subsequent to the creation of the prediction surfaces, Spatial Analyst's Raster Calculator (as illustrated by Figure 10) was used to determine which observation points had an increase or decrease in bird species between the two visits based on the interpolation. This means that areas where there was high diversity during the first visit experienced a reduction in species due to migrants moving further north and breeders moving to preferred breeding habitats. Similarly, breeding habitats which displayed low bird diversity during the migration season receive a boost in the number of species during the second visit.

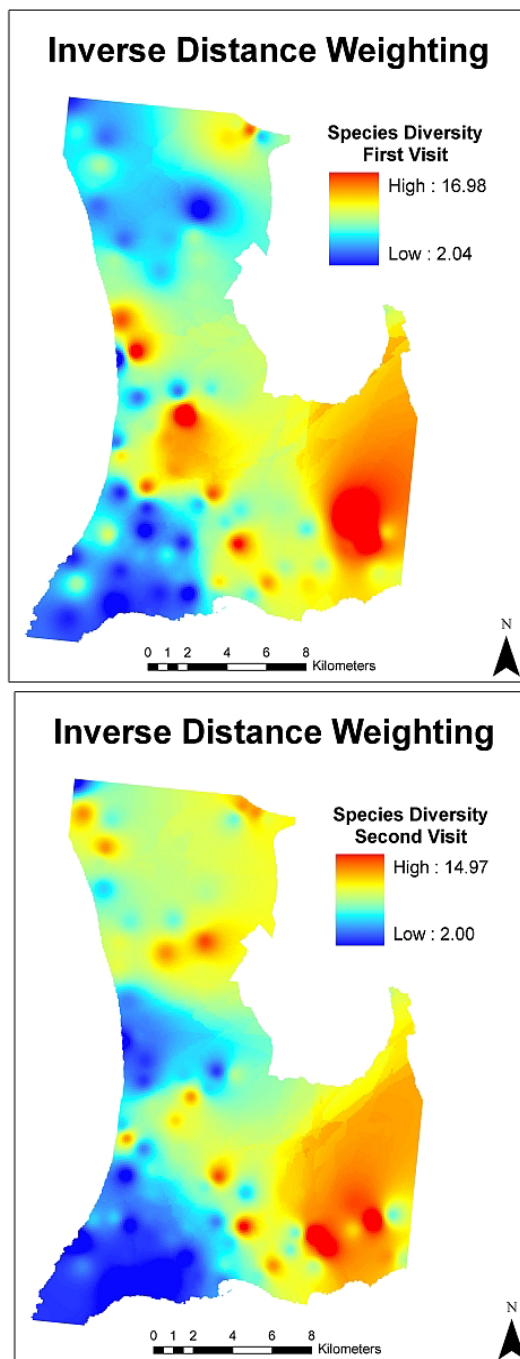


Figure 9: IDW interpolation for the first and second visit displaying species diversity.

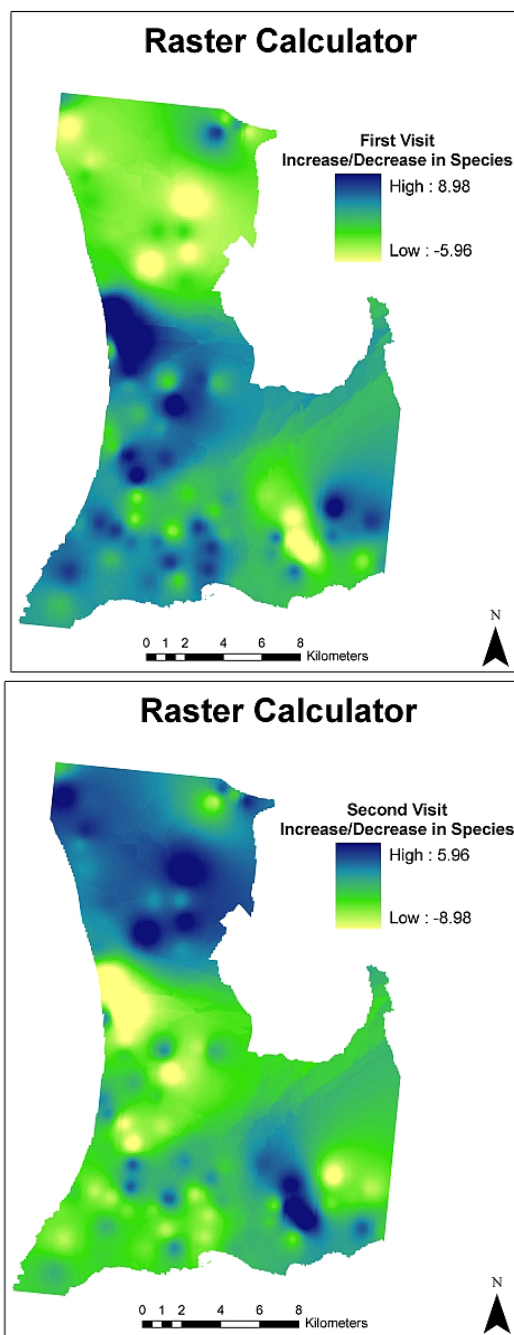


Figure 10: Raster calculator used to determine increase or decrease in species between the two counts, see section 2.3.1 for further clarification.

2.3.2 Ordinary Kriging: Structural Analysis

Prior to the Ordinary Kriging (OK) interpolation, a semivariogram was built which displayed the spatial continuity in the data. The mathematical model produced by the semivariogram was then used in calculating estimates of the surface at different points. The semivariogram is a function that provides half the variance of the continuity between all points separated by h and is inversely proportional to the distance of points. Semivariogram will be near the variance of the entire surface when it reaches the Sill at a distance known as the Range which is the distance where there is no evidence of spatial dependence. The computation of all likely combination of different pairs at various distances will be performed. The initial distance used is called Lag and will be modified through the dataset. Each point is evaluated against all other points to establish the variance between the points that have identical Lag distance and their geographical position. This process is repeated until every possible distance is examined. An isotropic semivariogram model was assumed due to the fact that bird species are abundant and dispersed across the region in various habitats; additionally the EDA did not reveal trends in any particular direction.

Theoretically, the semivariogram should be zero at the origin, this was not the case with the semivariograms produced in this study and a nugget had to be applied. The nugget is the height of the jump from the origin and it represents either variation at distances smaller than the sample spacing or measurement error. Figures 11 and 12 illustrate the semivariograms produced using GeoMS.

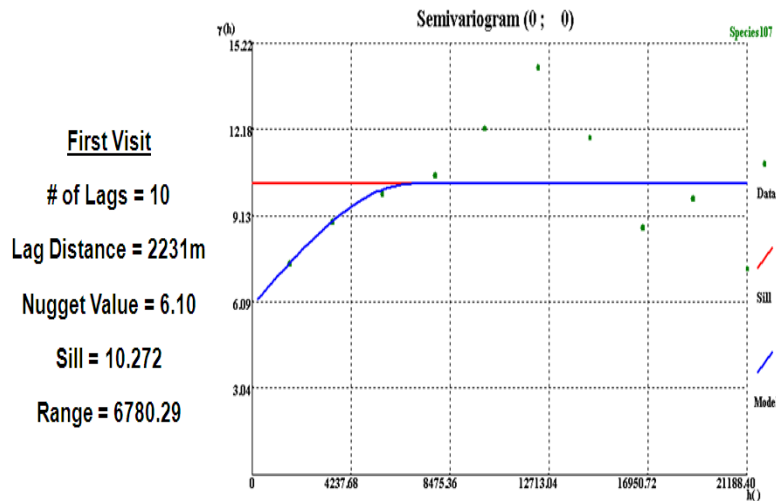


Figure 11: Semivariogram of the first visit.

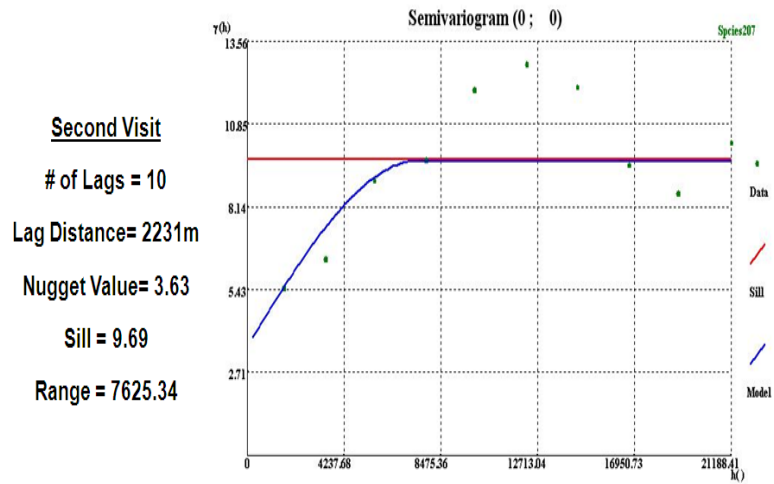


Figure 12: Semivariogram of the second visit.

2.3.3 Ordinary Kriging: Interpolation

Ordinary Kriging provides both a prediction and a standard error of the prediction at unsampled locations. The standard error provides a level of the uncertainty encountered at each predicted location. Figure 13 illustrates the OK prediction for the first and second visit.

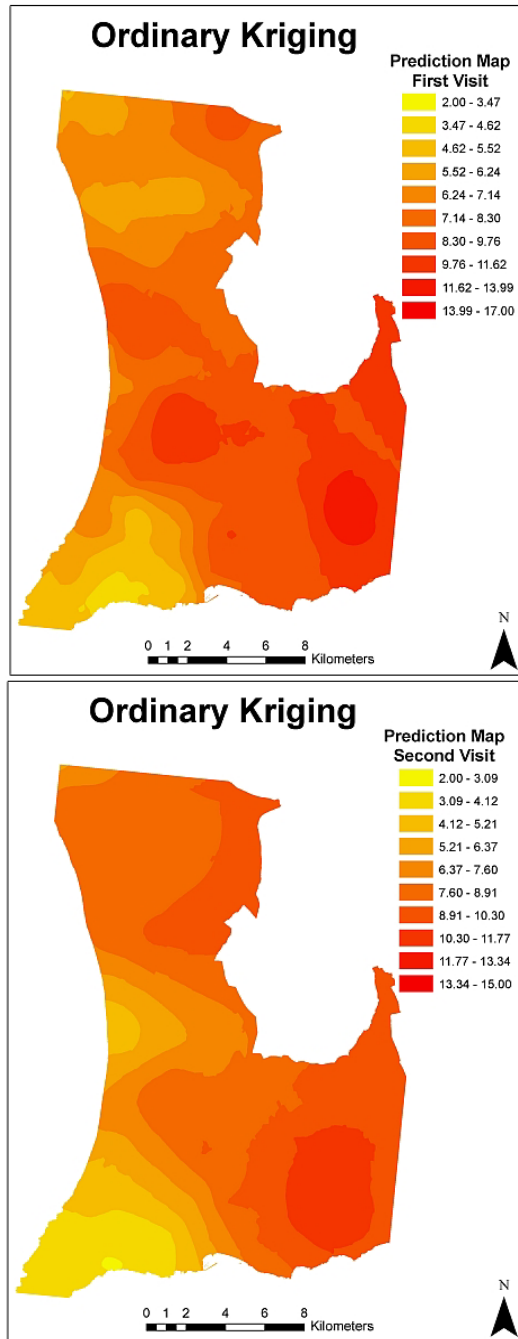


Figure 13: Ordinary Kriging interpolation displaying species diversity for the first and second visit.

2.4 Incorporating Land Cover Information

Although habitat information for each point was included with the provided dataset, it was found that this data differed from the CORINE Land Cover (CLC) project of the European Environmental Agency (EEA). Therefore, a decision was made to utilize the CLC as it provides the basis for integrated decision-making and environmental assessment for the EEA. Comparison was made by overlaying the CLC map (as illustrated by Figure 14) with the observation points and extracting the polygon attributes into the points that contained XY coordinates. This resulted in point features that contained CLC information which in turn could be overlaid on IDW (as illustrated by Tables 1 and 2) and OK (as illustrated by Tables 3 and 4) prediction maps to determine species diversity per habitat type.

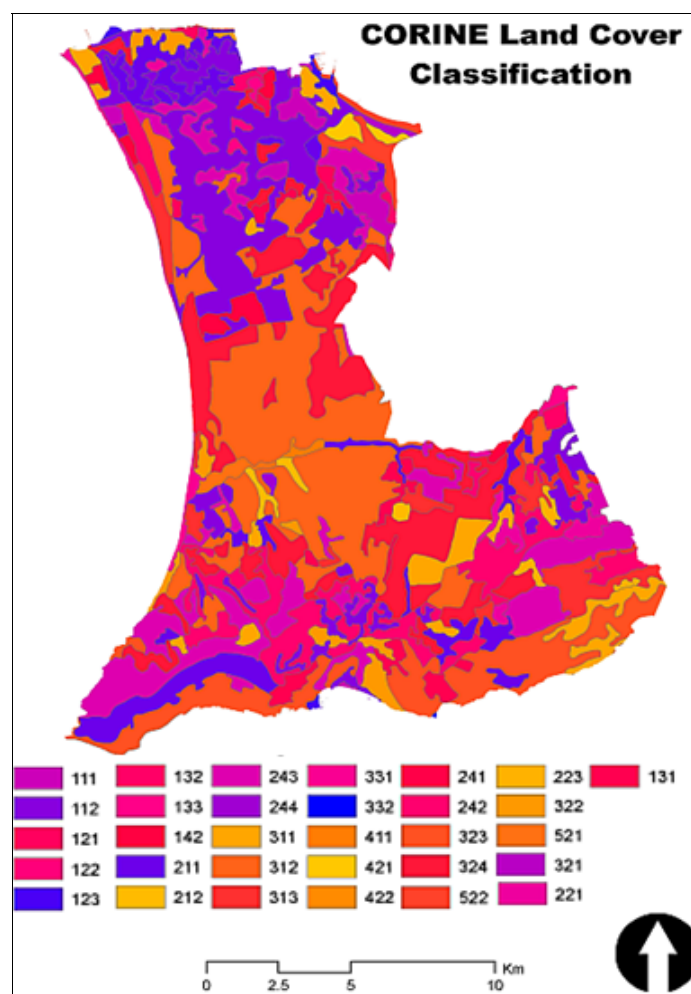


Figure 14: CLC classes were used to derive the habitat information.

PointID	CLC_CODE	LABEL
4MC86	112	Discontinued urban fabric
10MC86	212	Permanently irrigated land
1MC85	242	Complex cultivation patterns
4MC85	242	Complex cultivation patterns
15MC85	243	Agricultural land with significant natural vegetation
12MC85	243	Agricultural land with significant natural vegetation
17MC95	243	Agricultural land with significant natural vegetation
17MC86	312	Coniferous forest
16MC86	312	Coniferous forest
2MC86	312	Coniferous forest
18MC95	322	Moors and heathland
5MC95	324	Transitional wood-shrub
18MC86	324	Transitional wood-shrub
7MC86	324	Transitional wood-shrub
16MC87	421	Salt marshes

Table 1: The 15 highest bird diversity points for the first visit extracted from the IDW interpolation. Highlighted habitats are those that do not occur in the 15 highest for the second visit.

PointID	CLC_CODE	LABEL
4MC87	112	Discontinued urban fabric
7MC95	242	Complex cultivation patterns
6MC85	242	Complex cultivation patterns
14MC87	243	Agricultural land with significant natural vegetation
3MC85	243	Agricultural land with significant natural vegetation
8MC85	243	Agricultural land with significant natural vegetation
8MC95	311	Broad-leaved forest
9MC95	312	Coniferous forest
20MC87	312	Coniferous forest
2MC87	312	Coniferous forest
12MC87	312	Coniferous forest
14MC86	312	Coniferous forest
2MC85	312	Coniferous forest
3MC86	313	Mixed forest
3MC95	323	Sclerophyllous vegetation

Table 2: The 15 highest bird diversity points for the second visit extracted from the IDW interpolation. Highlighted habitats are those that do not occur in the 15 highest for the first visit.

PointID	CLC_CODE	LABEL
4MC86	112	Discontinued urban fabric
6MC95	211	Non-irrigated arable land
10MC86	212	Permanently irrigated land
8MC86	212	Permanently irrigated land
4MC95	243	Agricultural land with significant natural vegetation
1MC95	243	Agricultural land with significant natural vegetation
9MC86	312	Coniferous forest
16MC86	312	Coniferous forest
17MC86	312	Coniferous forest
12MC86	312	Coniferous forest
18MC95	322	Moors and heathland
7MC86	324	Transitional wood-shrub
19MC86	324	Transitional wood-shrub
5MC95	324	Transitional wood-shrub
18MC86	324	Transitional wood-shrub

Table 3: The 15 highest bird diversity points for the first visit extracted from the OK interpolation. Highlighted habitats are those that do not occur in the 15 highest for the second visit.

PointID	CLC_CODE	LABEL
4MC87	112	Discontinued urban fabric
13MC87	112	Discontinued urban fabric
6MC95	211	Non-irrigated arable land
7MC95	242	Complex cultivation
4MC95	243	Agricultural land with significant natural vegetation
14MC87	243	Agricultural land with significant natural vegetation
8MC95	311	Broad-leaved forest
9MC95	312	Coniferous forest
2MC87	312	Coniferous forest
1MC87	312	Coniferous forest
20MC87	312	Coniferous forest
10MC87	313	Mixed forest
19MC87	321	Natural grasslands
5MC95	324	Transitional wood-shrub
5MC87	324	<i>Transitional wood-shrub</i>

Table 4: The 15 highest bird diversity points for the second visit extracted from the OK interpolation. Highlighted habitats are those that do not occur in the 15 highest for the first visit.

3 RESULTS

The two interpolators gave similar predictions about the diversity of species in different habitats. The highlighted rows of tables 1 through 4 display habitats that did not occur in either the first visit (with respect to the second) or second visit (with respect to the first) for each interpolator. This signifies a particular preference of these habitats. For the first visit IDW predicted an increase in diversity in permanently irrigated land, moors and heathland and transitional woodland-shrub while OK predicted a diversity increase in permanently irrigated land and moors and heathland. For the second visit IDW predicted an increase in diversity in broad-leaved forest, mixed forest and sclerophyllous vegetation while OK predicted diversity increase in complex cultivations, broad-leaved forest, mixed forest and natural grasslands.

It is taken into account that the first visit corresponds to the migration season while the second visit corresponds to the breeding season. The predictions presented by both interpolators corroborate earlier studies that displayed increased richness and diversity in breeding birds across the gradient: low shrublands → tall shrublands → conifer → mixed → deciduous → agricultural areas (Moreira et al, 2001).

3.1 Error Calculation

In order to calculate error in the interpolated surfaces, the root mean square error (RMSE) was employed. RMSE is a measure of the mean error across an interpolated surface and is used to estimate the difference between the predicted and observed values.

The RMSE of the IDW interpolation for the first visit was 2.997 and for the second visit 2.518 (as illustrated by figure 15). RMSE values of the OK for the first visit and second visit were 2.888 and 2.448 respectively (as illustrated by Figure 16). It is deduced from these numbers that OK is the more accurate method for modeling species diversity. Furthermore, OK has an additional advantage in that it provides a Prediction Standard Error Map (as illustrated by Figure 17) that visually displays prediction accuracy across the surface.

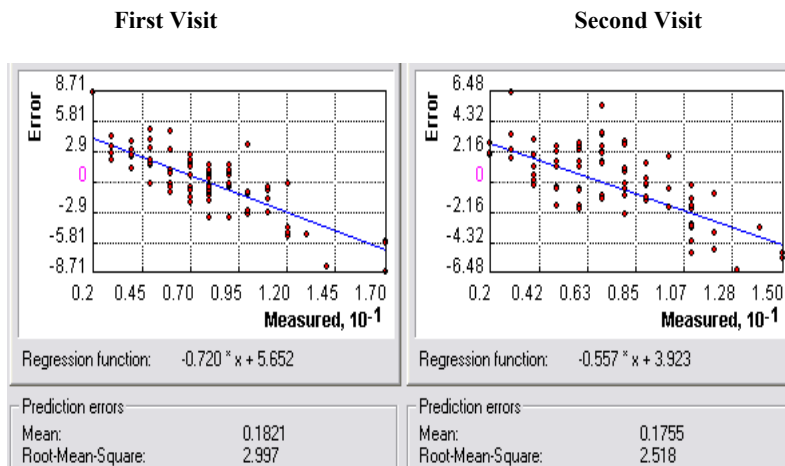


Figure 15: Prediction errors for the IDW interpolation.

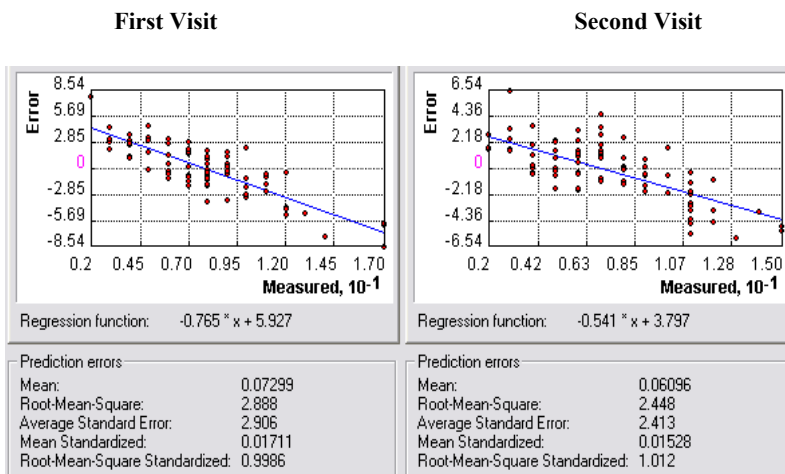


Figure 16: Prediction errors for the OK interpolation.

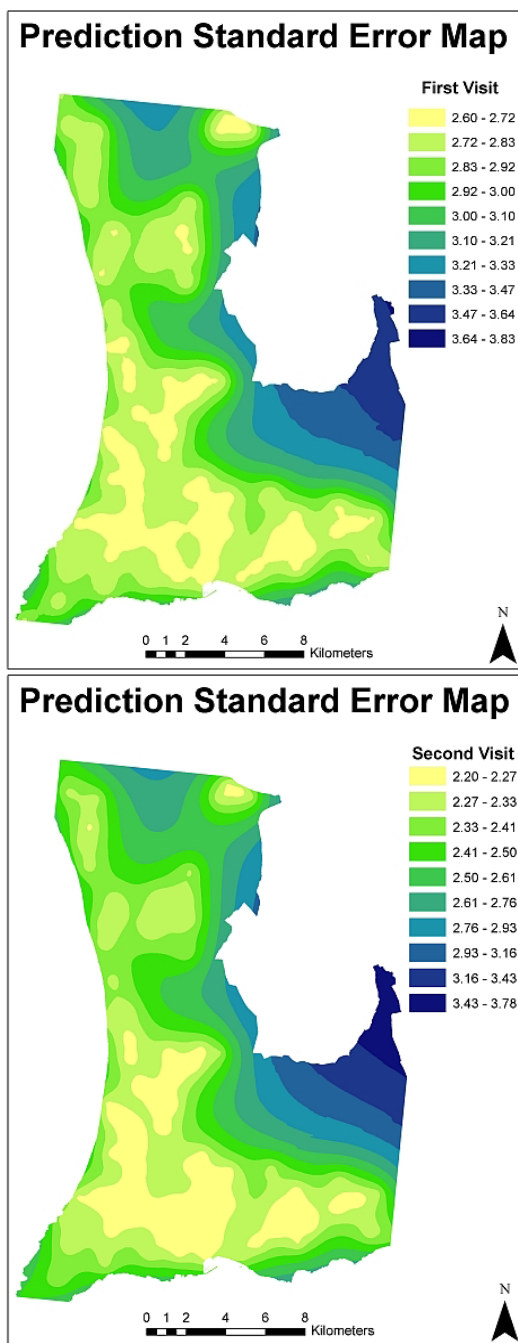


Figure 17: Prediction Standard Error map for the Ordinary Kriging interpolation.

4 CONCLUSIONS

Although the selected study area is not representative of the overall Portuguese landscape, it does provide an insight into the diverse types of habitats available to breeding and migratory birds in the country. Migratory birds are able to alter their migration routes in response to environmental changes and this might cause the failure of a migration and subsequently an unproductive breeding season. Since Portugal itself provides both migration stopovers and breeding habitats for birds, it is crucial to protect habitats, such as the ones identified in this study, in order to stop declines in biodiversity and achieve the 2002 Rio de Janeiro World Summit of Sustainable Development's goal to achieve a significant reduction in the current rate of loss of biological diversity by 2010.

Even though the margin of accuracy between IDW and OK is rather small, and both interpolators produced similar results, this study has demonstrated the potential of using geostatistical interpolation in the prediction of species diversity. Additionally, the study has shown the importance of incorporating habitat data and coordinates into bird counts; many countries have annual, semi-annual or quarterly bird counts that are recorded with diligence but often without such records. The inclusion of habitats and coordinates will not only aid ornithological research by providing spatially explicit quantitative and qualitative data but will also support the overall conservation of both the habitat and the species.

ACKNOWLEDGEMENTS

This work has been supported by the European Commission, Erasmus Mundus Programme, M.Sc. in Geospatial Technologies, project no. 2007-0064. We would like to thank Ana Meirinho of Sociedade Portuguesa para o Estudo das Aves for providing the data, Prof. Doutor Pedro Cabral and Prof^a. Ana Cristina Costa of Instituto Superior de Estatística e Gestão de Informação (ISEGI), Universidade Nova de Lisboa, for their invaluable assistance while working on this project.

BIBLIOGRAPHY

- BirdLife International, 2008 Europe-wide monitoring schemes highlight declines in widespread farmland birds. Retrieved December 26, 2008, from BirdLife State of the world's birds : <http://www.biodiversityinfo.org/sowb/casestudy.php?r=state&id=22>
- Moreira, F., Ferreira, P. G., Rego, F. C., & Bunting, S., 2001 Landscape changes and breeding bird assemblages in northwestern Portugal: the role of fire. *Landscape Ecology* , 175-187.
- Sanderson, F. J., Donald, P. F., Pain, D. J., Burfield, I. J., & van Bommel, F. P., 2006 Long-term population declines in Afro-Palaearctic migratory birds. *Biological Conservation* , 93-105.
- Tankersley, J. R., 2004 Migration of Birds As An Indicator of Broad-Scale Environmental Condition. *Environmental Monitoring and Assessment* , 55-67.
- Telleria, J. L., Baquero, R., & Santos, T., 2003 Effects of forest fragmentation on European birds: implications of regional differences in species richness. *Journal of Biogeography* , 621-628.
- Tucker, G. M., & Evans, M. I., 1997 Habitats for birds in Europe: a conservation strategy for the wider environment. Birdlife International. Cambridge, United Kingdom. ISBN 0946888329.
- Wolff, A., Paul, J.-P., Martin, J.-L., & Bretagnolle, V., 2001 The benefits of extensive agriculture to birds: the case of the little bustard. *Journal of Applied Ecology* , 963-975.