



# IBIS

BRITISH ORNITHOLOGISTS' UNION'S  
International Journal of Avian Science

## Evaluating species distribution models to predict nest occurrence of Eleonora's falcon (*Falco eleonora*) in the Aegean Sea

Journal:	<i>Ibis</i>
Manuscript ID:	IBIS-2010-OP-281
Manuscript Type:	Original Paper
Date Submitted by the Author:	22-Dec-2010
Complete List of Authors:	Kassara, Christina; University of Patras, Department of Biology Dimalexis, Anastasios; Hellenic Ornithological Society Fric, Jakob; Hellenic Ornithological Society Karris, Georgios; Technological Educational Institution (TEI) of the Ionian Islands Barboutis, Christos; University of Crete, Department of Biology; University of Crete, Natural History Museum of Crete Sfenthourakis, Spyros; University of Patras, Department of Biology
Category:	ecology: habitat, ecology: niche, conservation biology

SCHOLARONE™  
Manuscripts

1  
2  
3 993 **Title:** Evaluating species distribution models to predict nest occurrence of Eleonora's  
4 falcon (*Falco eleonora*) in the Aegean Sea.  
5  
6  
7 995

8  
9 996 **Authors:** Christina Kassara<sup>1\*</sup>, Anastasios Dimalexis<sup>2</sup>, Jakob Fric<sup>2</sup>, Georgios Karris<sup>3</sup>,  
10 Christos Barboutis<sup>4</sup> and Spyros Sfenthourakis<sup>1</sup>  
11

12  
13 998 <sup>1</sup> Section of Animal Biology, Department of Biology, University of Patras, Patras GR-  
14 26500, Greece.  
15 999

16  
17 1000 <sup>2</sup> Hellenic Ornithological Society, 24 Vas. Herakliou, GR-10682 Athens, Greece.

18  
19 1001 <sup>3</sup> Department of Environmental Technology and Ecology, Technological Educational  
20 Institution (TEI) of the Ionian Islands, 2 Kalvos Square, GR-29100 Zakyntos,  
21 Greece.  
22 1002  
23 1003  
24

25  
26 1004 <sup>4</sup> Department of Biology & Natural History Museum of Crete, University of Crete, P.O.  
27 Box 2208 GR-71409 Herakleion, Crete, Greece.  
28 1005

29  
30 1006 \* Corresponding author [ckassara@upatras.gr](mailto:ckassara@upatras.gr)  
31

32 1007  
33

34 1008 **Running title:** Modelling Eleonora's falcon nest occurrence  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 1009 **Summary**  
4

5 1010 1. A key element in the study and management of Earth's biodiversity is the  
6  
7 1011 knowledge of species distributions and the environmental conditions that shape  
8  
9 1012 habitat suitability. Eleonora's falcon exhibits a highly fragmented distribution pattern,  
10  
11 1013 which poses a constraint to systematic monitoring of its breeding colonies. Thus,  
12  
13 1014 predictive modelling of nest presence in a broad spatial extent could be proven a  
14  
15 1015 valuable tool for nation-wide and even, international plans.  
16  
17

18 1016 2. We explored the performance of different, yet well-established, methodologies for  
19  
20 1017 modeling nest-site selection by Eleonora's falcon at the local scale. In particular, we  
21  
22 1018 developed presence-absence and presence-only models, namely GLMs and Maxent,  
23  
24 1019 to predict nest occurrence as a function of the topography of the nesting territory.  
25  
26 1020 Nest occurrence data covered a large geographical range, the Aegean Sea, within  
27  
28 1021 which the majority of the global population is encountered.  
29  
30

31 1022 3. Both modeling approaches predicted successfully unoccupied areas in islands of  
32  
33 1023 substantial size, especially after having incorporated spatial terms in the model  
34  
35 1024 building process.  
36  
37

38 1025 4. We conclude that predictive models characterized by flexibility and/or making use  
39  
40 1026 of absence data, as well as incorporating nest clustering, can result in robust  
41  
42 1027 predictions of the nest occurrence of Eleonora's falcons in Greek breeding colonies  
43  
44 1028 and eventually, facilitate management policies.  
45  
46

47 1029  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 1030 **Introduction**  
4

5 1031 A key element in the study and management of Earth's biodiversity is the knowledge  
6  
7 1032 of species distributions and the environmental conditions that shape habitat  
8  
9  
10 1033 suitability. Occurrence data are usually sporadic, while many species exhibit a  
11  
12 1034 widespread, yet habitat-specific, pattern. Modern statistical approaches, coupled with  
13  
14 1035 GIS tools, have enabled scientists to develop a variety of methodologies and  
15  
16 1036 eventually predict species occurrence over broad geographic ranges. GIS data-  
17  
18 1037 sources, generally available free or at a low cost, minimize sampling effort of  
19  
20 1038 environmental data. GIS tools provide rapid, automated and sophisticated modules  
21  
22 1039 for data preparation, analysis and map creation. However, in many cases, the grain  
23  
24 1040 of the analysis is actually restricted to that of the available GIS data, which does not  
25  
26 1041 necessarily portray the species-environment relationships in sufficient detail (Guisan  
27  
28 1042 *et al.* 2007). Yet, a larger grain size might provide more meaningful results in terms of  
29  
30 1043 animal conservation and management. Therefore, a good knowledge of the species  
31  
32 1044 biology and ecology is essential, so that the model building process is based on a  
33  
34 1045 robust and sound conceptual frame given the objectives of the study (Guisan &  
35  
36 1046 Zimmermann 2000, Brotons *et al.* 2004).  
37

38  
39 1047 Another important issue in species distribution modeling is the presence of spatial  
40  
41 1048 structure in the model residual errors, but mainly in the response variable being  
42  
43 1049 modeled (i.e. species occurrence). Presence records are likely to be positively  
44  
45 1050 correlated in space either due to exogenous reasons, like sampling bias and spatial  
46  
47 1051 autocorrelation of environmental variables, or due to endogenous reasons, like  
48  
49 1052 aggregation patterns inherent to the species biology (Legendre 1993). In the former  
50  
51 1053 case, spatial autocorrelation is usually addressed by appropriate subsampling of the  
52  
53 1054 data (e.g. Graf *et al.* 2006, Kaliontzopoulou *et al.* 2008), while in the latter case the  
54  
55 1055 observed spatial pattern of species presence is incorporated in the model with the  
56  
57 1056 inclusion of spatial terms (e.g. De Frutos *et al.* 2007).  
58  
59  
60

1  
2  
3 1057 Statistical methods mainly used for nest-site selection are classified into two broad  
4  
5 1058 classes; namely the presence-absence models and the presence-only models  
6  
7 1059 according to the availability of absence records (for review see Guisan & Zimmerman  
8  
9 1060 2000). Presence-only methods are currently receiving special attention (Elith *et al.*  
10  
11 1061 2006), since data on absence are most often lacking from ecological studies. If  
12  
13 1062 presence-only methods can produce comparable results to presence-absence  
14  
15 1063 methods, then they could constitute a powerful tool in species conservation and  
16  
17 1064 management by minimizing sampling effort in terms of time, logistics and cost.  
18  
19 1065 Recently there has been a significant increase of papers comparing the performance  
20  
21 1066 of various methodologies used to predict species occurrence (e.g. Manel *et al.* 1999,  
22  
23 1067 Elith & Graham 2009). Their findings cannot be easily generalized since model  
24  
25 1068 performance may vary according to the species being studied (Guisan *et al.* 2007),  
26  
27 1069 the region of interest (Manel *et al.* 1999), the range size (McPherson *et al.* 2004), the  
28  
29 1070 sample size (Pearce & Ferrier 2000), and the ratio of presence to total number of  
30  
31 1071 observations, i.e. prevalence (Manel *et al.* 2001). Thus, results of predictive  
32  
33 1072 performance should be evaluated given the specifications of the analysis in question.  
34  
35 1073 With this study, after having considered the available knowledge on the species  
36  
37 1074 habitat requirements, we aim to (1) identify factors that influence nest-site selection  
38  
39 1075 by Eleonora's falcon at the local scale and (2) explore the performance of alternative  
40  
41 1076 methodologies in predicting the species presence in breeding colonies. We assess  
42  
43 1077 model performance with threshold-independent metrics of accuracy and discuss our  
44  
45 1078 findings considering previous studies with implications to the species study.  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 1079 **Materials and Methods**

4  
5 1080 *The study species*

6  
7 1081 Eleonora's falcon (*Falco eleonorae* Gén , 1839) is a medium sized migratory falcon  
8  
9 1082 of the hobby group, which breeds exclusively in the Mediterranean basin, the Canary  
10  
11 1083 Islands and the NW coast of Africa. Given its wide range, it is listed as "Least  
12  
13 1084 Concern" in the IUCN 2010 list (Birdlife International 2010).

14  
15 1085 According to the most recent assessment, its global population accounts for ca.  
16  
17 1086 15,000 pairs (Papakonstantinou 2007) of which Greece, and more specifically the  
18  
19 1087 Aegean Sea, constitutes the centre of the species' range hosting more than 80% of  
20  
21 1088 the total breeding pairs (12,300 pairs; Dimalexis *et al.* 2008). The typical habitat of  
22  
23 1089 Eleonora's falcon colonies in Greece consists of sea-cliffs and rocky uninhabited  
24  
25 1090 islets. Mean colony size is 54 pairs (Dimalexis *et al.* 2008), while the nests are mainly  
26  
27 1091 located in crevices, under bushes and boulders, offering protection from intense solar  
28  
29 1092 radiation and strong winds, factors which have been proven crucial for breeding  
30  
31 1093 success across its breeding colonies (Wink *et al.* 1982, Dolz & Dies 1987, Ristow *et*  
32  
33 1094 *al.* 1983, Badami 1995, Mocci Demartis & Peddis 1999, Bonn n 2004).

34  
35  
36  
37 1095 To our knowledge, until now only one study has attempted to relate the nest  
38  
39 1096 distribution pattern of Eleonora's falcon with environmental parameters, but in a  
40  
41 1097 spatially restricted area (Urios & Mart nez-Abra n 2006). In the current study we use  
42  
43 1098 nest occurrence records gathered during a monitoring project implemented in the  
44  
45 1099 summer periods between 2004 and 2007, involving twenty three uninhabited islets in  
46  
47 1100 the Aegean Sea (Dimalexis *et al.* 2008). We believe that, given the species  
48  
49 1101 fragmented distribution and the difficulties in accessing its breeding colonies in terms  
50  
51 1102 of weather conditions and geomorphology of the areas in question, predictive  
52  
53 1103 modelling of nest presence in a broader spatial extent could be proven a valuable  
54  
55 1104 tool for long-term management policies.

56  
57  
58  
59 1105

60 1106 *Nest occurrence and topographic data*

1  
2  
3 1107 After a preliminary data screening, but also due to the availability of topographic data,  
4  
5 1108 we restricted the data pool to nine islets for the analyses described in continuation,  
6  
7 1109 the surface of which was searched thoroughly for the presence of nests.

8  
9  
10 1110 The islets in question lie between  $35^{\circ} 19'N$  -  $39^{\circ} 12'N$  and  $23^{\circ} 27'E$  -  $26^{\circ} 48'E$ . Their  
11  
12 1111 mean surface is  $0.01 \text{ km}^2$ , their substrate is mainly formed by calcareous leptosols  
13  
14 1112 (Panagos & Van Liedekerke 2004) and their elevation does not exceed 110m a.s.l.  
15  
16 1113 During the breeding seasons of 2004-2007, mean air temperature was  $23.8^{\circ}\text{C}$  and  
17  
18 1114 the prevailing winds were of NW direction, with mean speed of 23.1 km/hr at 2m  
19  
20 1115 a.s.l., based on meteorological data provided by the National Observatory of Athens.

21  
22 1116 In the current study nest presence was associated with continuous topographic  
23  
24 1117 variables extracted by a Digital Elevation Model (Hellenic Military Geographic  
25  
26 1118 Service, [www.gys.gr](http://www.gys.gr)) in the Greek Grid projection EGSA87 (Table 1). Prior to the  
27  
28 1119 model building process a grid matrix of rectangular cells, matching the resolution size  
29  
30 1120 of the topographic variables (i.e.  $30 \times 30\text{m}$ ), was superimposed on maps illustrating  
31  
32 1121 the nest location on each islet. Grid cells containing at least one nest during the  
33  
34 1122 course of the monitoring project were classified as occupied, while the rest were  
35  
36 1123 classified as unoccupied, resulting in 153 presence and 991 absence cells for model  
37  
38 1124 building. Since the geomorphology of the islets varied significantly, we further divided  
39  
40 1125 them into two groups (islets smaller or larger than  $0.100 \text{ km}^2$ ) and proceeded with the  
41  
42 1126 analyses separately for each islet group.

43  
44  
45  
46 1127

#### 47 48 49 1128 *Statistical analyses*

50  
51 1129 Through the years, a variety of models have emerged aiming to describe and predict  
52  
53 1130 the current and future distribution of the species as a function of environmental  
54  
55 1131 predictors, which in turn are directly or indirectly linked to the species presence in a  
56  
57 1132 given geographical region (Guisan & Zimmerman 2000). We chose two different  
58  
59 1133 modelling approaches from the families of presence-absence and presence-only  
60

1  
2  
3 1134 methods, to test their performance in predicting Eleonora's falcon nest occurrence at  
4  
5 1135 the local scale.

6  
7 1136 Statistical analyses and model assessment were performed in SPSS 17.0, model  
8  
9 1137 development in R 2.10.1 (R Development Core Team 2009) and Maxent 3.3.1,  
10  
11 1138 whereas data preparation, spatial analyses and map construction in ArcGIS 9.2  
12  
13 1139 (ESRI 2006). Moran's index statistics were calculated in SAM v3.0 (Rangel *et al.*  
14  
15 1140 2006), while the autocovariate terms were built in R 2.10.

16  
17  
18 1141

19  
20  
21 1142 *2.3.1. Presence-absence models*

22  
23 1143 We developed presence-absence models using General Linear Models (GLM;  
24  
25 1144 McCullagh & Nelder 1989) with a binomial error distribution and a logit link function.  
26  
27 1145 In this context, the probability of occurrence of an event ( $P$ ) is estimated according to  
28  
29 1146 the formula (eqn 1):

30  
31  
32 1147 
$$P(z) = \frac{1}{1 + \exp(z)}$$
 Equation 1

33  
34  
35 1148 where  $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i$  i.e. the sum of contributions of all independent variables  
36  
37 1149 ( $x_i$ ) to the model.

38  
39  
40 1150 By transforming the above equation into the natural log of the odds ratio for success,  
41  
42 1151 the binomial probability of the occurrence of an event can also be expressed as a  
43  
44 1152 linear function of the independent variables as follows (eqn 2)

45  
46  
47 1153

48  
49 1154 
$$\ln\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i$$
 Equation 2

50  
51  
52 1155

53  
54 1156 GLMs offer greater modelling capabilities in comparison to classic linear regression  
55  
56 1157 models, since they allow for the inclusion of non-linear effects among variables and  
57  
58 1158 for non-parametric distributions of the independent variables. However, GLMs are  
59  
60 1159 sensitive to the presence of multicollinearity among the explanatory variables



1  
2  
3 1160 (Graham 2003), presence of outliers and lack of independence in the errors, issues  
4  
5 1161 that should be addressed accordingly (Fielding & Bell 1997).  
6

7 1162 Prior to GLM construction, we used Mann-Whitney U-tests to check for statistically  
8  
9 1163 significant differences between occupied and unoccupied cells with regards to the  
10  
11 1164 candidate topographic variables. We also investigated the spatial pattern of  
12  
13 1165 Eleonora's falcon nest distribution using Moran's index (Moran's I). For clustered  
14  
15 1166 distributions, Moran's I is close to +1, for stratified distributions close to -1, while for  
16  
17 1167 random distributions it is close to 0. We created correlograms, in which Moran's I  
18  
19 1168 values were plotted against distances between localities (lags), to identify the  
20  
21 1169 distance at which spatial autocorrelation was minimized (i.e. observations were  
22  
23 1170 spatially independent). We tested the significance of Moran's I values at different lags  
24  
25 1171 by performing 199 Monte Carlo permutations on the raw data-set. In addition, we  
26  
27 1172 calculated Cook's distance to identify any outliers (i.e. Cook's distance > 1), in which  
28  
29 1173 case they were removed from the analysis. Given the lack of multicollinearity (i.e.  
30  
31 1174 VIF > 10; Poirazidis *et al.* 2004), all available explanatory variables were considered in  
32  
33 1175 the model building process.  
34  
35  
36

37 1176 In particular, we fitted GLMs using forward stepwise logistic regression (hereafter  
38  
39 1177 LR), after a modification for variable inclusion developed by Engler *et al.* (2004). This  
40  
41 1178 technique resembles the typical forward stepwise procedure, in which the model  
42  
43 1179 initially contains no variables and variables are added sequentially until a final model  
44  
45 1180 is obtained, but giving priority to those variables that are most influential on the model  
46  
47 1181 performance based on the reduction in residual deviance. In this way, the model  
48  
49 1182 building procedure is no longer sensitive to the order in which variables enter the  
50  
51 1183 model; hence, under- or over-fitting is avoided (Pearce & Ferrier 2000).  
52  
53

54 1184 After having verified the existence of spatial autocorrelation, both in the response  
55  
56 1185 variable and in the residuals of the nonspatial logistic regression models (LR), we  
57  
58 1186 investigated two alternative solutions commonly cited in literature. In one case, we  
59  
60 1187 incorporated a third degree polynomial of the central latitude and longitude of each

1  
2  
3 1188 grid cell to the final logistic regression model (hereafter LRsp; Lichstein *et al.* 2002)  
4  
5 1189 and, in another case, we included an autocovariate term (hereafter LRar; Augustín *et*  
6  
7 1190 *al.* 1996), estimated for each grid cell as the average nest presence of its neighboring  
8  
9  
10 1191 cells. The neighbors' effect was restricted up to that distance where spatial  
11  
12 1192 autocorrelation of the nest occurrence was no longer significant. Therefore, for the  
13  
14 1193 big islet group, the autocovariate term was calculated taking into account 49  
15  
16 1194 neighboring grid cells (window size 7 x 7), while for the small islet group, the  
17  
18 1195 autocovariate term was calculated considering 25 neighboring grid cells (window size  
19  
20 1196 5 x 5).

22 1197

### 25 1198 *2.3.2. Presence-only models*

27 1199 Among the available presence-only techniques we chose Maxent (Phillips *et al.*  
28  
29 1200 2004), a niche-based technique that has recently gained popularity, to predict the  
30  
31 1201 habitat suitability for nesting for Eleonora's falcon on the islets in question. Maxent  
32  
33 1202 resembles other statistical modelling techniques, like GLMs, since it calculates the  
34  
35 1203 unknown distribution of a species over a geographical region of interest from a  
36  
37 1204 sample of localities of known occurrence and spatially explicit environmental  
38  
39 1205 conditions. This technique has been found to perform equally well or even better than  
40  
41 1206 other presence-only and/or presence-absence methods, especially when sample  
42  
43 1207 sizes are small (Elith *et al.* 2006).

46 1208 In the model building process, presence data were added as samples, while absence  
47  
48 1209 data were added as background points. We retained the default settings for model  
49  
50 1210 parameterization. We run Maxent under the 'auto features' mode, which allows for  
51  
52 1211 maximum model flexibility (Phillips & Dudík 2008), and we chose logistic values as  
53  
54 1212 output (range 0-1) to generate habitat suitability values reflecting the estimated  
55  
56 1213 probability of occurrence. The contribution of each predictor to the final model was  
57  
58 1214 assessed by a jack-knife analysis of the training gain, which is a measure of the  
59  
60 1215 likelihood of the training samples (Phillips *et al.* 2006).

1  
2  
3 1216

4  
5 1217 *2.4. Model comparison and evaluation*

6  
7 1218 We compared model performance using independent data by testing the agreement  
8  
9 1219 between the observed and predicted nest occurrence by means of two different  
10  
11 1220 threshold-independent measures; namely the AUC score and Pearson's correlation  
12  
13 1221 coefficient. We did not consider threshold-dependent metrics that are typically  
14  
15 1222 derived by classification matrices, since their performance is sensitive to model  
16  
17 1223 quality and species prevalence (Freeman & Moisen 2008). The two chosen accuracy  
18  
19 1224 measures were computed through a cross-validation procedure. The training set,  
20  
21 1225 used to calibrate the model, consisted of 75% of the available data, while the test set,  
22  
23 1226 used to evaluate the final models, consisted of the remaining 25% (Huberty 1994 in  
24  
25 1227 Fielding & Bell 1997). Absence data were disproportionately higher than presence  
26  
27 1228 data in each training and test set (prevalence 0.376 and 0.105 for the small and big  
28  
29 1229 islet group, respectively). Data partitioning was conducted ten times at random,  
30  
31 1230 ensuring that prevalence remained the same across the resulting data-sets. Model  
32  
33 1231 calibration and evaluation for Maxent and the three different was conducted with the  
34  
35 1232 same training and test data-sets.

36  
37 1233 As a measure of model predictive power, we used the Area under the ROC curve  
38  
39 1234 (AUC), which is rather sensitive to both species prevalence and decision threshold  
40  
41 1235 effects (Hanley & McNeil 1982). A ROC curve is created by plotting sensitivity  
42  
43 1236 against the omission rate (1-sensitivity) at all possible thresholds of presence-  
44  
45 1237 absence classification, and the area beneath the curve corresponds to AUC. AUC  
46  
47 1238 ranges from 0.5, for models that predict no better than random, to 1.0, for models  
48  
49 1239 with perfect predictive power.

50  
51 1240 However, being a rank-based measure, AUC does not account for the degree to  
52  
53 1241 which the predicted values have been calibrated. In contrast, Pearson's correlation  
54  
55 1242 coefficient (Elith *et al.* 2006) considers whether big differences in predicted values  
56  
57 1243 correspond to big differences in probability of occurrence (Phillips & Dudík 2008).  
58  
59  
60

1  
2  
3 1244 Altogether, we built forty models for six small islets and forty for three big islets. Intra-  
4  
5 1245 group differences in modelling performance among the three GLMs (LR, LRsp and  
6  
7 1246 LRar) and Maxent were tested with Wilcoxon signed rank tests.  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

IBIS Review Copy

1  
2  
3 1247 **Results**

4  
5 1248 *Nest location in relation to topography*

6  
7 1249 Nest-site characteristics differed according to the size of the islet being studied. On  
8  
9 1250 small islets, nests were found mainly in the interior of the islets and at locations of  
10  
11 1251 high inclination, as compared to the available sites. On the contrary, on big islets,  
12  
13 1252 nests were located closer to the coastline, in crevices of SE-facing cliffs receiving  
14  
15 1253 greater amounts of solar radiation as compared to the available sites (Table 2). On  
16  
17 1254 both islet groups nest distribution was non-random (Moran's  $I > 0$  at  $P \leq 0.05$ , Fig. 1)  
18  
19 1255 with a mean nest spacing of 30m.  
20  
21

22 1256

23  
24  
25 1257 *Predicted nest presence*

26  
27 1258 *Presence-absence models*

28  
29 1259 According to the final LR models, on small islets of the Aegean Sea the probability of  
30  
31 1260 nest occurrence increases with elevation (Table 3). On big islets, nest-site selection  
32  
33 1261 depends on the curvature of the terrain, the distance to the coastline and the incident  
34  
35 1262 solar radiation. In particular, according to the final model, it is more likely to find nests  
36  
37 1263 close to the coastline, in concave surfaces and with increased incident solar radiation  
38  
39 1264 (Table 3).  
40  
41

42 1265

43  
44 1266 *Presence-only models*

45  
46 1267 For the small islet group, Maxent was in agreement with LR, since it identified  
47  
48 1268 elevation as the most important variable, followed by slope. On the other hand, for  
49  
50 1269 the big islets, Maxent highlighted slope orientation along the North-South axis, solar  
51  
52 1270 radiation and distance to the coastline as the most influential factors (Table 4), in  
53  
54 1271 contrast to GLM results. However, a visual inspection of the response curves of the  
55  
56 1272 aforementioned explanatory variables revealed the same tendency with regards to  
57  
58 1273 the probability of nest occurrence as a function of these variables.  
59  
60

1  
2  
3 1274 Predictive maps of nest presence for all 4 modelling techniques, illustrating average  
4  
5 1275 predicted values, are presented for one islet per islet group (Figs 2 and 3).  
6

7 1276  
8

9  
10 1277 *Model comparison and evaluation*

11 1278 Examination of the spatial structure of errors of the non-spatial GLMs (LR) showed  
12  
13 1279 the lack of independence in the model residuals for both islet groups. However, the  
14  
15 1280 addition of spatial terms in the final models addressed this issue successfully (Fig. 4).  
16  
17

18 1281 The LRsp and LRar for the small and the big islet group, respectively, performed  
19  
20 1282 better, since these two models explained a larger amount of unexplained variance  
21  
22 1283 according to Nagelkerke's  $R_N^2$  (Table 5).  
23  
24

25 1284 Significance levels of the two accuracy measures, AUC and Pearson's correlation  
26  
27 1285 coefficient, revealed that, within the big islet group, values were significant at  $P \leq 0.05$   
28  
29 1286 for all test data-sets, while within the small islet group a considerable percentage of  
30  
31 1287 non-significant values was observed, indicating LR and Maxent as the worst  
32  
33 1288 performing models. In particular, Pearson's correlation coefficient was not statistically  
34  
35 1289 significant for 70% of LR models, 10% of LRsp models, 20% of LRar models and  
36  
37 1290 90% of Maxent models, while the corresponding percentages for AUC scores were  
38  
39 1291 70%, 0%, 0% and 80%.  
40  
41

42 1292 Within the small islet group predictive performance was found significantly higher for  
43  
44 1293 the LRsp model as assessed with the AUC score, while based on Pearson's  
45  
46 1294 correlation coefficient the two spatial models (LRsp, LRar) performed equally well  
47  
48 1295 and better than the rest (Wilcoxon signed rank tests, Table 6).  
49  
50

51 1296 In the case of the big islet group, both measures of accuracy indicated that model  
52  
53 1297 performance was similar among all models but LR, which performed worse (Wilcoxon  
54  
55 1298 signed rank tests, Table 6).  
56

57 1299 As a rule of thumb, AUC scores  $>0.75$  suggest good predictive ability (Phillips &  
58  
59 1300 Dudík 2008). Consequently, and in agreement with Pearson's correlation coefficients,  
60  
1301 for the small islet group we consider LRsp and LRar models being of average

1  
2  
3 1302 predictive power, and LR together with Maxent of poor predictive power. For the big  
4  
5 1303 islet group, AUC scores suggest relatively good model accuracy in all cases.  
6  
7 1304 Overall, probabilities of nest occurrence were not clearly discriminated among  
8  
9 1305 occupied and unoccupied cells, as implied by the low Pearson's correlation  
10  
11 1306 coefficients. A closer look at the distribution of probability values revealed that  
12  
13 1307 discrimination of true presences was relatively poor, since almost half of these cases  
14  
15 1308 were assigned to probability values lower than 0.5 (Figs 5 and 6). Therefore, AUC  
16  
17 1309 scores were mainly influenced by the omission rate, i.e. false positive fraction.  
18  
19 1310 Nonetheless, mean predicted probability values for the occupied cells were  
20  
21 1311 significantly higher compared to the unoccupied cells in all cases (Mann-Whitney U-  
22  
23 1312 tests,  $P \leq 0.05$ ).  
24  
25  
26 1313 Finally, average model predictions were more consistent among the four models for  
27  
28 1314 the big than the small islet group. In both cases, mean predictions between the  
29  
30 1315 spatial GLMs (LRsp, LRar) were the most highly correlated, while Maxent average  
31  
32 1316 predictions were more similar to LR (Table 7).  
33  
34 1317 Predicted values were more variable for the LRsp and Maxent replicate models  
35  
36 1318 within the small and big islet groups, respectively, as expressed by the Coefficient of  
37  
38 1319 Variation (Mann-Whitney U-tests,  $P \leq 0.05$ ). In all cases, model variability in predicted  
39  
40 1320 values was higher for unoccupied cells in comparison to the occupied ones (Mann-  
41  
42  
43  
44 1321 Whitney U-tests,  $P \leq 0.05$ ).  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 1322 **Discussion**

4  
5 1323 *Nest location in relation to topography*

6  
7 1324 Eleonora's falcons tend to occupy the same nest in consecutive years (Ristow, Wink  
8  
9 1325 & Wink 1979). According to our findings, nest-site selection is influenced, *inter alia*,  
10  
11 1326 by the topography of the area at the local scale. On small islets, most nests were  
12  
13 1327 located towards the interior of the islets, at high elevations and steep slopes. On big  
14  
15 1328 islets, breeding pairs preferred concave locations closer to the coastline, at  
16  
17 1329 intermediate elevations of medium inclination facing SE and receiving relatively  
18  
19 1330 higher solar radiation. These topographic variables, in combination with  
20  
21 1331 meteorological parameters, have been pinpointed as important factors of nest  
22  
23 1332 location on a western Mediterranean island (Urios & Martínez-Abrain 2006), as well  
24  
25 1333 as of breeding performance in various Mediterranean colonies (Ristow & Wink 1985,  
26  
27 1334 Badami 1995, Bonnín 2004), although at the micro-scale level. In particular, elevated  
28  
29 1335 sites have been related to protection from wave action, steep slopes to lower risk of  
30  
31 1336 intruder invasion, and concave sites to sufficient visual protection from predators and  
32  
33 1337 conspecifics (Walter 1979, Urios & Martínez-Abrain 2006).

34  
35 1338 Preference in terrain orientation is inconsistent among other study areas in the  
36  
37 1339 Mediterranean Sea and some have linked it to the direction of the incoming migrant  
38  
39 1340 flow (Mayol 1977), the main food source for Eleonora's falcon during the young-  
40  
41 1341 raising period, and others to protection from intense sun irradiation and increased  
42  
43 1342 exposure to the prevailing winds (Urios & Martínez-Abrain 2006). In our case, there  
44  
45 1343 was no apparent association to the direction of the migrant flow, but instead, taking  
46  
47 1344 also into consideration the mean wind direction during the summer period in the area,  
48  
49 1345 nests located in SE-facing terrain are more protected against both sun and wind. In  
50  
51 1346 addition, we also noticed a preference towards hotter surfaces, a characteristic that  
52  
53 1347 has been also observed in another Greek breeding colony in the past (Walter 1979).  
54  
55 1348 Due to the strong northern winds blowing in the study areas during the summer  
56  
57 1349 period, hotter surfaces could counterbalance the chilling effect of such wind  
58  
59  
60



1  
2  
3 1350 conditions during egg incubation. The availability of such data at a finer scale could  
4  
5 1351 shed more light into the nesting requirements of the species at the microhabitat level.  
6

7  
8 1352

9  
10 1353 *Predicting nest occurrence at the local level*

11 1354 Presence-absence (LR) and presence-only models (Maxent) distinguished the same  
12  
13 1355 topographic predictor of nest presence on the small islets, i.e. elevation. Wave  
14  
15 1356 action, as expressed through the preference of higher elevations and more remote  
16  
17 1357 sites from the coastline, seems to play a predominant role in nest-site selection on  
18  
19 1358 small islets. Given their restricted surface and their relatively smoother overall relief,  
20  
21 1359 wave action may have a larger impact on nests found on such islets than on larger  
22  
23 1360 ones, especially during the summer period when wind force is considerable as  
24  
25 1361 explained above.  
26  
27

28  
29 1362 In the case of the big islets, different priority was given to the influence of topographic  
30  
31 1363 variables given the applied methodology, although there was some overlap.  
32

33 1364 According to GLM, curvature of the terrain appears to be the most influential  
34  
35 1365 parameter in nest-site selection, while distance to the coastline and incident solar  
36  
37 1366 radiation is also important, but to a lesser degree. Taking into account the substrate  
38  
39 1367 of the islets in question, being mainly calcareous, sites close to the coastline are  
40  
41 1368 prone to wave erosion. Such sites, given that they are also located at an adequate  
42  
43 1369 elevation above sea level, constitute suitable nesting places since they are protected  
44  
45 1370 from both wave action and potential intruders.  
46  
47

48  
49 1371 On the other hand, Maxent identified terrain orientation along the North-South axis as  
50  
51 1372 the most significant factor, followed by solar radiation and distance to the coastline.  
52

53 1373 Such a discrepancy makes sense if we consider that, being a presence-only method,  
54  
55 1374 Maxent focuses only on the information conveyed by positive occurrence locations  
56  
57 1375 and it, therefore, disregards differences among occupied and unoccupied cells.  
58

59 1376 Returning to the primary question of this study, whether nest presence can be  
60  
1377 adequately predicted via topographic factors at the local scale, the results presented

1  
2  
3 1378 here allow us to conclude that on islets of substantial size (i.e. larger than 0.100 km<sup>2</sup>)  
4  
5 1379 we can be quite confident for the absence of nests from particular sites with specific  
6  
7 1380 topographic characteristics. On smaller islets, nest-selection seems to be influenced  
8  
9  
10 1381 by additional factors that have not been considered in the current analysis (Araújo &  
11  
12 1382 Williams 2000). It is equally possible that nest-selection is the product of more  
13  
14 1383 complex interactions, while topography may only be a distal predictor (Austin 2002).  
15  
16 1384 Poor model fit could have also resulted if the spatial resolution of the chosen  
17  
18 1385 topographic variables was not detailed enough to portray the interaction of nest-site  
19  
20 1386 selection with the environment (Guisan & Zimmermann 2000) or in the case where  
21  
22 1387 nest-site selection occurs in multiple scales (Johnson 1980, Foody 2004). Although  
23  
24 1388 multi-scale approaches have proved more realistic and efficient in large-scale studies  
25  
26 1389 (Gibson *et al.* 2004), they would be impractical in cases of restricted spatial extent as  
27  
28 1390 the small islets studied here.  
29  
30  
31 1391 As a final remark to the adequacy of topographic variables in predicting nest  
32  
33 1392 occurrence of Eleonora's falcon, we should emphasize that in cases where the  
34  
35 1393 studied populations are not in an equilibrium state with their environment, predictive  
36  
37 1394 modelling, based on the premise of stationarity, will fail to illustrate the species-  
38  
39 1395 environment relationship (Brotons *et al.* 2004). Non-equilibrium is likely to occur with  
40  
41 1396 sudden changes of habitat quality, forcing individuals to occupy sub-optimal sites.  
42  
43  
44 1397 We consider that the models built in the current study were not biased by equilibrium  
45  
46 1398 issues for two main reasons. First, nest-site tenacity of Eleonora's falcon in Greek  
47  
48 1399 colonies has been observed, not only during the course of the monitoring project  
49  
50 1400 from which the data used in this analysis were withdrawn, but also during long-term  
51  
52 1401 studies in the past (Ristow *et al.* 1983). . Second, the species fitness has been  
53  
54 1402 related to nest-site selection as mentioned above. Yet future studies, distinguishing  
55  
56 1403 among experienced and inexperienced pairs and focusing on the former, could  
57  
58 1404 reveal stronger relationships between nest presence and environmental variables.  
59  
60

1  
2  
3 1405  
4

5 1406 *Considering nest clustering in predictive models*  
6

7 1407 In both islet groups the distribution of nests exhibited a clustered pattern. Eleonora's  
8 falcon is a gregarious raptor, which breeds in colonies of even up to 400 pairs  
9 (Dimalexis *et al.* 2008) on Greek islands, although nest spacing has been found to  
10 vary considerably. Presence of high nest densities may be indicative of nest-site  
11 quality and availability as it has been observed at Morocco breeding colonies (Walter  
12 1979). In such cases, tolerance to neighbours is compensated by the advantages  
13 offered by communal defence of the colony from aerial trespassers, the efficiency of  
14 bird hunting in groups during the nestling-raising period and by increased food  
15 availability (Walter 1979, Ristow *et al.* 1982, Rosén *et al.* 1999). At this point, we  
16 should mention that Eleonora's falcon is in essence an insectivorous species and  
17 that its sociability has emerged due to insect-hunting needs in the wintering grounds,  
18 while colonial breeding has appeared later in evolutionary times (Ristow 2004).  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

33 1419 The importance of considering nest clustering during the model development was  
34 also reflected in the improvement of model performance with the addition of spatial  
35 terms. For avian species it has been suggested that conspecifics may take  
36 advantage of both the presence of neighbouring pairs and their reproductive success  
37 in order to evaluate nest-site quality and eventually choose where to settle (see  
38 references in Danchin & Wagner 1997). If this hypothesis holds true, then immature  
39 Eleonora's falcons that are commonly present in breeding colonies (Walter 1979)  
40 would indirectly learn how to assess habitat suitability and therefore optimize their  
41 breeding performance by settling close to experienced pairs. However, in a Greek  
42 colony young Eleonora's falcons usually occupied inferior nesting sites, while their  
43 reproductive success was lower compared to older falcons (Ristow *et al.* 1983).  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1425 Eleonora's falcons that are commonly present in breeding colonies (Walter 1979)  
1426 would indirectly learn how to assess habitat suitability and therefore optimize their  
1427 breeding performance by settling close to experienced pairs. However, in a Greek  
1428 colony young Eleonora's falcons usually occupied inferior nesting sites, while their  
1429 reproductive success was lower compared to older falcons (Ristow *et al.* 1983).  
1430 Therefore, further study is needed to test the presence of such a mechanism in the  
1431 case of Eleonora's falcon considering also the availability of suitable nesting areas in  
1432 conjunction with territoriality and intra-specific relations.

1  
2  
3 1433 Whatever the underlying reasons and processes, nest aggregation in Eleonora's  
4  
5 1434 falcons is an endogenous characteristic as in other raptor species (De Frutos *et al.*  
6  
7 1435 2007). Consequently, the incorporation of the spatial dimension in modelling the  
8  
9  
10 1436 probability of nest occurrence is well-justified (Augustín, Muggleston & Buckland  
11  
12 1437 1996).

1438

1439 *Model comparison and evaluation*

1440 We applied two modelling techniques, which have never been used before for  
1441 predictive modelling of the target species, but have been very popular in similar  
1442 studies, GLMs and Maxent. Overall, GLMs performed slightly better than Maxent, as  
1443 suggested by the two metrics of accuracy used, namely AUC and Pearson's  
1444 correlation coefficient. The choice between presence-absence models and presence-  
1445 only methods has been the subject of many studies and different conclusions can be  
1446 drawn according to the study-species. If the species being modeled is in equilibrium  
1447 with its environment (Brotons *et al.* 2004), absence corresponds to low habitat  
1448 suitability and, therefore, improves model accuracy in presence-absence studies  
1449 (Hirzel *et al.* 2001). In this study we have not found huge differences in model  
1450 performance between GLM and Maxent and this could be attributed to the fact that,  
1451 although it relies only on presence data, Maxent can reconstruct complex interactions  
1452 between the response variable and the environmental predictors, which reflect  
1453 species-environment relationships in a more realistic way. In other words, function  
1454 complexity in Maxent was balanced against a more spherical and accurate  
1455 knowledge of the species distribution, considered in GLM. Therefore, models  
1456 combining presence-absence data and allowing function flexibility, such as the  
1457 Generalized Additive Models (GAM), could prove more efficient in the future.  
1458 Previous studies have also emphasized that the success in predicting species  
1459 distributions depends on the distribution pattern itself, where more generalist species  
1460 are less accurately modelled than narrow-distributed species (Brotons *et al.* 2004).

1  
2  
3 1461 Widespread species could either present regional variations in habitat use (Segurado  
4  
5 1462 & Araújo 2004) or be influenced by environmental factors operating at different  
6  
7 1463 scales (Brotons *et al.* 2004). Again, in such cases presence-absence models are  
8  
9 1464 expected to provide more accurate results than presence-only models (Brotons *et al.*  
10  
11 1465 2004). We therefore believe that further studies, which consider a wider geographical  
12  
13 1466 extent and/or islands of different geomorphology within Eleonora's falcon breeding  
14  
15 1467 distribution, could shed more light on this aspect.

16  
17  
18 1468 The addition of spatial terms removed the spatial structure of model errors and  
19  
20 1469 improved the predictive power of GLMs substantially, as has been found in similar  
21  
22 1470 studies (Araújo & Williams 2000, De Frutos *et al.* 2007), due to intrinsic processes  
23  
24 1471 that determine nest spacing. Spatial autocorrelation was accounted for only by the  
25  
26 1472 GLMs, with the inclusion of either the geographic coordinates of occurrence localities  
27  
28 1473 or an autocovariate term to the final models. Although the problem of spatial  
29  
30 1474 autocorrelation has been emphasized in predictive modelling based on maximum  
31  
32 1475 entropy techniques, few studies have addressed this issue, either by incorporating  
33  
34 1476 spatial terms (De Marco *et al.* 2008) or by proper data subsampling (Kaliontzopoulou  
35  
36 1477 *et al.* 2008). Consequently, we strongly recommend the investigation of spatially  
37  
38 1478 explicit maximum entropy models in predictive modelling of Eleonora's falcon nest  
39  
40 1479 occurrence when absence data are lacking.

41  
42  
43 1480 An issue that is worth of special attention in predictive modelling is the choice of data  
44  
45 1481 sets for model assessment. The ideal scenario involves the use of independent test  
46  
47 1482 data in both space and time. However, such data sets are rarely available and,  
48  
49 1483 therefore, data partitioning has become common practice. In our case, the data pool  
50  
51 1484 was split into training and test data sets totally at random. Variability in predicted  
52  
53 1485 values suggested that prediction of nest occurrence is more sensitive to the choice of  
54  
55 1486 absence records rather than that of presence records. Nonetheless, the overall  
56  
57 1487 model's efficiency in discriminating unoccupied cells was higher when prevalence  
58  
59 1488 was low (i.e. big islet group). In general, model performance balances between  
60

1  
2  
3 1489 generality and accuracy. Therefore, given the objectives of a study, one should  
4  
5 1490 compromise between the cost of predicting false positives (specificity) and false  
6  
7 1491 negatives (sensitivity) (Fielding & Bell 1997). Low prevalence has been found to  
8  
9 1492 optimize specificity in the expense of sensitivity (Manel *et al.* 2001). However, GLMs  
10  
11 1493 tend to optimize predictions over the larger data sample (Fielding & Bell 1997). Our  
12  
13 1494 findings are in agreement with this, especially for the big islet group which had the  
14  
15 1495 lowest prevalence and the higher prediction success for absences. Still, over-  
16  
17 1496 predicting a species true distributional range might be unfavourable for conservation  
18  
19 1497 purposes, given the relative costs involved (Manel *et al.* 2001, VanDerWal *et al.*  
20  
21 1498 2009).

22  
23  
24 1499 As for the chosen accuracy measures, the AUC score, although widely used, has  
25  
26 1500 received criticism regarding its adequacy in evaluating model performance. Its main  
27  
28 1501 drawbacks lie on the fact that it weighs omission and commission errors equally and  
29  
30 1502 that it does not consider the goodness-of-fit of the models in question (Lobo *et al.*  
31  
32 1503 2008, Phillips & Dudík 2008). However, as shown in our study when it is used in  
33  
34 1504 conjunction with other measures like Pearson's correlation coefficient, model  
35  
36 1505 assessment becomes more robust (Elith *et al.* 2006, Elith & Graham 2009).  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 1506 **Conclusions**  
4

5 1507 To our knowledge, this is the first study to model nest occurrence of Eleonora's  
6 falcon over a broad geographic area in a systematic way and with the application of  
7  
8 1508 falcon over a broad geographic area in a systematic way and with the application of  
9  
10 1509 different methodological approaches, therefore providing the basis for future studies.  
11  
12 1510 According to our findings, taking into account certain topographic features along with  
13  
14 1511 population processes (i.e. nest aggregation) one can predict with satisfactory  
15  
16 1512 success areas void from nests by applying either classic logistic regression methods  
17  
18 1513 (GLM) or niche-based methods (Maxent). In the lack of absence data the latter  
19  
20 1514 comprise a powerful tool in the conservation and management of the species. In  
21  
22 1515 order to be able to extrapolate our results successfully to other geographic areas,  
23  
24 1516 more sophisticated methods which integrate non-linear processes occurring at  
25  
26 1517 multiple scales, are suggested for future studies. In addition, we consider worth  
27  
28 1518 investigating the effect of implementing distribution models by using nest records of  
29  
30 1519 experienced and inexperienced pairs in isolation. Finally, based on our findings, we  
31  
32 1520 strongly recommend the use of complementary accuracy measures in order to obtain  
33  
34 1521 a clear picture of the model predictive and discriminatory power.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 1522 **Acknowledgements**  
4

5 1523 Data collection was conducted in the frame of and funded by the LIFE-Nature Project  
6  
7 1524 “Conservation measures for *Falco eleonora* in Greece” (LIFE 03NAT/GR/000091)  
8  
9  
10 1525 coordinated by the Hellenic Ornithological Society (HOS-Birdlife-Greece). We would  
11  
12 1526 like to thank Portolou Danae, Latsoudis Panagiotis, Bourdakis Stratis, Xirouchakis  
13  
14 1527 Stavros, Georgiakakis Panagiotis and all field ornithologists, volunteers and boat  
15  
16 1528 captains for their assistance in fieldwork.  
17

18 1529 Special thanks to Sinos Giokas (Univ. of Patras) for his advice on several statistical  
19  
20 1530 issues. Finally, we would like to express our gratitude to C. Lagouvardos, National  
21  
22 1531 Observatory of Athens, for the provision of meteorological data for the study areas.  
23

24 1532 This work was funded by the European Union (LIFE/03NAT/GR000091) and the  
25  
26 1533 Leventis Foundation.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1534 **References**

- 1535 Araújo, M. B. & Williams, P. H. 2000. Selecting areas for species persistence using  
1536 occurrence data. *Biol. Conserv.* 96: 331-345.
- 1537 Augustín, N. H., Muggleston, M. A. & Buckland, S. 1996. An autologistic model for  
1538 the spatial distribution of wildlife. *J. Appl. Ecol.* 33: 339-347.
- 1539 Austin, M. P. 2002. Spatial prediction of species distribution: an interface between  
1540 ecological theory and statistical modelling. *Ecol. Model.* 157: 101-118.
- 1541 Badami, A. 1995. Relazioni tra Fattori Ecologici e Riproduzione nel Falco Della  
1542 Regina (*Falco eleonora*). *Supplemento alle ricerche di biologia della Selvaggina*  
1543 XXII: 115-119.
- 1544 BirdLife International 2010. Species factsheet: *Falco eleonora*. URL:  
1545 <http://www.birdlife.org> (accessed 15 June 2010).
- 1546 Bonnín, J. 2004. Recompte i paràmetres reproductors de la població balear del Falcó  
1547 marí *Falco eleonora* a l'any 2004. *Anuari Ornitològic de les Îles Balears* 19: 1-9.
- 1548 Brotons, L., Thuiller, W., Araújo, M. B. & Hirzel, A. H. 2004. Presence-absence versus  
1549 presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27:  
1550 437-448.
- 1551 Danchin, E. & Wagner, R. 1997. The evolution of coloniality: the emergence of new  
1552 perspectives. *TREE* 12(9): 342-347.
- 1553 De Frutos, A., Olea, P. P. & Vera, R. 2007. Analyzing and modelling spatial  
1554 distribution of summering lesser kestrel: The role of spatial autocorrelation. *Ecol.*  
1555 *Model.* 200: 33-44.
- 1556 De Marco, P. Jr, Diniz-Filho, J. A. F. & Bini, L. M. 2008. Spatial analysis improves  
1557 species distributions modelling during range expansion. *Biol. Lett.* 4: 577-580.
- 1558 Dimalexis, A., Xirouchakis, S., Portolou, D., Latsoudis, P., Karris, G., Fric, J.,  
1559 Georgiakakis, P., Barboutis, C., Bourdakis, S., Ivovic, M., Kominos, T. & Kakalis, E.  
1560 2008. The status of Eleonora's Falcon (*Falco eleonora*) in Greece. *J. Ornithol.* 149:  
1561 23-30.

- 1  
2  
3 1562 Dolz, J. C. & Dies, N. 1987. El halcón de Eleonor (*Falco eleonora*, G  n  ) en las  
4  
5 1563 islas Columbretes. *Islas Columbretes: Contribuci  n al estudio de su medio natural*  
6  
7 1564 (eds L. A. Matilla, J. L. Carretero & A. M. Garcia-Carrascosa), pp. 241-262.  
8  
9 1565 Generalitat Valencia, Valencia.  
10  
11 1566 Elith, J. & Graham, C. H. 2009. Do they? How do they? Why do they? - On finding  
12  
13 1567 reasons for differing performances of species distribution models. *Ecography* 32: 66-  
14  
15 1568 67.  
16  
17 1569 Elith, J., Graham, C. H., Anderson, R. P., Dud  k, M., Ferrier, S., Guisan, A., Hijmans,  
18  
19 1570 R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle,  
20  
21 1571 B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC.,  
22  
23 1572 Peterson, A. T., Philips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E.,  
24  
25 1573 Sobreon, J., Willimas, S., Wisz, M. S. & Zimmermann, N. E. 2006. Novel methods  
26  
27 1574 improve prediction of species' distributions from occurrence data. *Ecography* 29:  
28  
29 1575 129-151.  
30  
31 1576 Engler, R., Guisan, A. & Rechsteiner, L. 2004. An improved approach for predicting  
32  
33 1577 the distribution of rare and endangered species from occurrence and pseudo-  
34  
35 1578 absence data. *J. Appl. Ecol.* 41: 263-274.  
36  
37 1579 ESRI 2006. ArcGIS Desktop for Windows. Version 9.2. Redlands CA, USA.  
38  
39 1580 Fielding, A. H. & Bell, J. F. 1997. A review of methods for the assessment of  
40  
41 1581 prediction errors in conservation presence/absence models. *Envir. Conserv.* 24: 38-  
42  
43 1582 49.  
44  
45 1583 Foody, G. M. 2004. Spatial nonstationarity and scale-dependency in the relationship  
46  
47 1584 between species richness and environmental determinants for the sub-Saharan  
48  
49 1585 endemic avifauna. *Global Ecol. Biogeogr.* 13: 315-320.  
50  
51 1586 Freeman, E. A. & Moisen, G. G. 2008. A comparison of the performance of threshold  
52  
53 1587 criteria for binary classification in terms of predicted prevalence and kappa. *Ecol.*  
54  
55 1588 *Mod.* 217: 48-58.  
56  
57  
58  
59  
60

- 1  
2  
3 1589 Gibson, L. A., Wilson, B. A., Cahill, D. M. & Hill, J. 2004. Spatial prediction of rufous  
4  
5 1590 bristlebird habitat in a coastal heathland: a GIS-based approach. *J. Appl. Ecol.* 41:  
6  
7 1591 213-223.  
8  
9  
10 1592 Graf, R. H. Bollmann, K., Sachot, S., Suter, W. & Bugmann, H. 2006. On the  
11  
12 1593 generality of habitat distribution models: a case study of capercaillie in three Swiss  
13  
14 1594 regions. *Ecography* 29: 319-328.  
15  
16 1595 Graham, M. H. 2003. Confronting multicollinearity in ecological multiple regression.  
17  
18 1596 *Ecology* 84: 2809-2815.  
19  
20 1597 Guisan, A. & Zimmermann, N. E. 2000. Predictive habitat distribution models in  
21  
22 1598 ecology. *Ecol. Model.* 135: 147-186.  
23  
24 1599 Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S. & Peterson, A.  
25  
26 1600 T. 2007. What matters for predicting the occurrences of trees: techniques, data, or  
27  
28 1601 species' characteristics? *Ecol. Monogr.* 77: 615-630.  
29  
30 1602 Hanley, J. A. & McNeil, B. J. 1982. The meaning and use of the Area under a  
31  
32 1603 Receiver Operating Characteristic (ROC) Curve. *Radiology* 143: 29-36.  
33  
34 1604 Hirzel, A. H., Helfer, V. & Metral, F. 2001. Assessing habitat-suitability models with a  
35  
36 1605 virtual species. *Ecol. Model.* 145: 111-121.  
37  
38 1606 Johnson, D. H. 1980. The comparison of usage and availability measurements for  
39  
40 1607 evaluating resource preference. *Ecology* 61: 65-71.  
41  
42 1608 Kaliontzopoulou, A., Brito, J. C., Carretero, M. A., Larbes, S. & Harris, D. J. 2008.  
43  
44 1609 Modelling the partially unknown distribution of wall lizards (*Podarcis*) in North Africa:  
45  
46 1610 ecological affinities, potential areas of occurrence, and methodological constraints.  
47  
48 1611 *Can. J. Zool.* 86: 992-1001.  
49  
50 1612 Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:  
51  
52 1613 1659-1673.  
53  
54 1614 Lichstein, J. W., Simons, T. R., Shiner, S. A., & Franzreb, K. E. 2002. Spatial  
55  
56 1615 autocorrelation and autoregressive models in ecology. *Ecol. Monogr.* 72: 445-463.  
57  
58  
59  
60

- 1  
2  
3 1616 Lobo, J. M., Jimenez-Valverde, A. & Real, R. 2008. AUC: a misleading measure of  
4  
5 1617 the performance of predictive distribution models. *Global Ecol. Biogeogr.* 17: 145-  
6  
7 1618 151.  
8  
9  
10 1619 Manel, S., Diaz, J. M. & Ormerod, S. J. .1999. Comparing discriminant analysis,  
11  
12 1620 neural networks and logistic regression for predicting species distributions: a case  
13  
14 1621 study with a Himalayan river bird. *Ecol. Model.* 120: 337-347.  
15  
16 1622 Manel, S., Williams, H. C. & Osmerod, S. J. 2001. Evaluating presence-absence  
17  
18 1623 models in ecology: the need to account for prevalence. *J. Appl. Ecol.* 38: 921-931.  
19  
20 1624 Mayol. J. 1977. Estudios sobre el halcón de Eleonor, *Falco eleonora*, en las islas  
21  
22 1625 Baleares, *Ardeola* 23: 103-136.  
23  
24  
25 1626 McCullagh, P. & Nelder. J. A. 1989. *Generalized Linear Models*. Chapman and Hall,  
26  
27 1627 London.  
28  
29 1628 McPherson, J. M., Jertz, W. & Rogers D. J. 2004. The effects of species' range sizes  
30  
31 1629 on the accuracy of distribution models: ecological phenomenon or statistical artefact?  
32  
33 1630 *J. Appl. Ecol.* 41: 811-823.  
34  
35  
36 1631 Mocci Demartis A. & Peddis, S. 1999. Osservazioni relative all' influsso dei venti sul  
37  
38 1632 successo del ciclo riproduttivo del Falco della Regina (*Falco eleonora*), in  
39  
40 1633 Sardegna. *Rendiconti Seminario della Facolta dis Scienze dell' Universita di Cagliari*  
41  
42 1634 69: 109-112.  
43  
44  
45 1635 Panagos, P. & Van Liedekerke, M. 2004. The European Soil Database (Distribution  
46  
47 1636 Version 2). URL: <http://eusoils.jrc.ec.europa.eu/> (accessed 2 June 2009).  
48  
49 1637 Papakonstantinou, C. 2007. *Eleonora's falcon: Ruling the Aegean Skies*. Hellenic  
50  
51 1638 Ornithological Society, Athens.  
52  
53 1639 Pearce, J. & Ferrier, S. 2000. An evaluation of alternative algorithms for fitting  
54  
55 1640 species distribution models using logistic regression. *Ecol. Model.* 128: 127-147.  
56  
57 1641 Phillips, S. J. & Dudík, M. .2008. Modeling of species distributions with Maxent: new  
58  
59 1642 extensions and a comprehensive evaluation. *Ecography* 31: 161-175.  
60

- 1  
2  
3 1643 Phillips. S. J., Dudík, M. & Schapire, R. E. 2004. A maximum entropy approach to  
4  
5 1644 species distribution modeling. *Proceedings of the Twenty-First International*  
6  
7 1645 *Conference on Machine Learning*, pp. 655-662. ACM Press, New York.  
8  
9  
10 1646 Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of  
11  
12 1647 species geographic distributions. *Ecol. Model.* 190: 231-259.  
13  
14 1648 Poirazidis, K., Goutner, V., Skartsi, T. & Stamou, G. 2004. Modelling nesting habitat  
15  
16 1649 as a conservation tool for the Eurasian Black Vulture (*Aegypius monachus*) in Dadia  
17  
18 1650 National Reserve, northeastern Greece. *Biol. Conserv.* 118: 235-148.  
19  
20 1651 R Development Core Team 2009. R: A language and environment for statistical  
21  
22 1652 computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-  
23  
24 1653 07-0, URL: <http://www.R-project.org>.  
25  
26  
27 1654 Rangel, T. F. L. V. B., Diniz-Filho, J. A. F. & Bini, L. M. 2006. Towards an integrated  
28  
29 1655 computational tool for spatial analysis in macroecology and biogeography. *Global*  
30  
31 1656 *Ecol. Biogeogr.* 15: 321-327.  
32  
33  
34 1657 Ristow, D. 2004. On the insect diet of Eleonora's falcon *Falco eleonora* and its  
35  
36 1658 importance for coloniality. *Raptors worldwide. World Working Group on Birds of*  
37  
38 1659 *Prey/MME-Birdlife* (eds R. D. Chancellor & B.-U. Meyburg), pp. 705-712. London.  
39  
40 1660 Ristow, D. & Wink, M. 1985. Breeding success and conservation management of  
41  
42 1661 Eleonora's falcon. *Conservation Studies on Raptors. ICBP Technical Publication*  
43  
44 1662 *No.5* (eds I. Newton & R. D. Chancellor), pp. 147-152. Cambridge.  
45  
46  
47 1663 Ristow, D., Wink, C. & Wink, M. 1979. Site tenacity and pair bond of the Eleonora's  
48  
49 1664 falcon. *Il-Meril* 20: 16-18.  
50  
51 1665 Ristow, D., Wink, C. & Wink, M. 1982. Biology of Eleonora's falcon (*Falco*  
52  
53 1666 *eleonora*): 1. Individual and social defense behavior. *Raptor Res.* 16: 65-70.  
54  
55  
56 1667 Ristow, D., Wink, C. & Wink, M. 1983. Biologie des Eleonorenfalken (*Falco*  
57  
58 1668 *eleonora*): 11. Die Anpassung des Jagdverhaltens an die vom Wind abhängigen  
59  
60 1669 Zugvogelhäufigkeiten. *Die Vogelwarte* 32: 7-13.

- 1  
2  
3 1670 Rosén, M., Hedenstrom, A., Badami, A., Spina, F. & Akesson, S. 1999. Hunting flight  
4  
5 1671 behaviour of the Eleonora's Falcon *Falco eleonora*. *J. Av. Biol.* 30: 342-350 30, 342-  
6  
7 1672 350.  
8  
9 1673 Segurado, P. & Araújo, M. B. 2004. An evaluation of methods for modelling species  
10  
11 1674 distributions. *J. Biogeogr.* 31: 1555-1568.  
12  
13 1675 Urios, G. & Martinez-Abraín, A. 2006. The study of nest-site preferences in  
14  
15 1676 Eleonora's falcon through digital terrain models on a western Mediterranean island.  
16  
17 1677 *J. Ornithol.* 147: 13-23.  
18  
19 1678 VanDerWal, J., Shoo, L. P., Graham, C. & Williams, S. E. 2009. Selecting pseudo-  
20  
21 1679 absence data for presence-only data: How far you should stray from what you know?  
22  
23 1680 *Ecol. Model.* 220: 589-594.  
24  
25 1681 Walter, H. 1979. *Eleonora's Falcon: adaptations to prey and habitat in a social raptor.*  
26  
27 1682 University of Chicago Press, Chicago and London.  
28  
29 1683 Wink, M., Wink, C. & Ristow, D. 1982. Biologie des Eleonorenflaken (*Falco*  
30  
31 1684 *eleonora*): 10. Der Einfluß der Horstlage auf den Bruterfolg. *J. Ornithol.* 123: 401-  
32  
33 1685 408.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1686

1687 Table 1. Description of the explanatory topographical variables used for modeling  
 1688 nest-site selection of Eleonora's falcon.

1689

<b>Variable</b>	<b>Description</b>
Elev	Ground elevation (m)
Slope	Terrain slope (degrees)
Acos	Cosine of terrain aspect, representing northness <sup>1</sup> , where negative values correspond to southward-facing slopes
Asin	Sine of terrain aspect, representing eastness <sup>1</sup> , where negative values correspond to west-facing slopes
Curv	Terrain curvature, where negative values represent concave surfaces
Solar	Mean incident solar radiation for August and September 2004-2007 (WH/m <sup>2</sup> )
Dist	Distance to coastline (m)

1690 <sup>1</sup>Poirazidis *et al.* 2004

1691

1692 Table 2. Comparison of topographic variables between occupied and unoccupied  
 1693 grid cells, by means of Mann-Whitney U-tests for each islet group (\* significant at  
 1694  $P \leq 0.05$ , \*\* significant at  $P \leq 0.01$ , standard deviations in parenthesis).

1695

<b>Predictors</b>	<b>Small islets</b>			<b>Big islets</b>		
	Occupied	Unoccupied	U	Occupied	Unoccupied	U
Elev	13.796 (±10.988)	8.263 (±7.975)	2,694**	15.969 (±14.337)	28.334 (±27.992)	24,541**
Slope	13.412 (±7.672)	10.665 (±6.697)	3,155*	18.308 (±8.356)	15.080 (±9.652)	23,552**
Acos	0.078 (±0.743)	0.202 (±0.682)	3,523	-0.592 (±0.543)	0.031 (±0.658)	14,730**
Asin	-0.009 (±0.676)	0.045 (±0.707)	3,703	-0.218 (±0.560)	0.053 (±0.752)	26,307*
Curv	0.849 (±1.716)	0.488 (±1.739)	3,375	-0.341 (±0.941)	0.208 (±1.294)	17,704**
Solar	4,176.044 (±298.183)	4,149.162 (±232.822)	3,515	4,355.862 (±296.591)	4,041.545 (±413.846)	14,071**
Dist	27.620 (23.557)	19.121 (17.190)	3,077*	31.861 (25.462)	63.662 (53.227)	20,293**

1696



1697

1698 Table 3. Model summary of the nonspatial GLM (LR) for two islet groups.

1699

<b>Predictors</b>	<b>b</b>	<b>S.E.</b>	<b>p</b>
<i>Small islets</i>			
Intercept	-1.178	0.248	0.000
Elev	0.062	0.017	0.000
<i>Big islets</i>			
Intercept	35.074	9.844	0.000
Curv	-0.293	0.110	0.008
Dist	-0.023	0.005	0.000
Solar	-0.022	0.004	0.000
Solar <sup>2</sup>	3.360x10 <sup>-6</sup>	6.438x10 <sup>-7</sup>	0.000

1700

1701

1702 Table 4. Average percentage contribution of the topographic variables according to  
 1703 the Maxent models (contributions greater than 15% are highlighted in bold).

1704

<i>Predictors</i>	<i>Small islets</i>	<i>Big islets</i>
Acos	13.431	<b>31.271</b>
Asin	14.101	4.699
Curv	10.340	5.067
Dist	6.950	<b>19.069</b>
Elev	<b>38.768</b>	10.871
Slope	<b>15.950</b>	3.684
Solar	0.459	<b>25.339</b>

1705

1706

1707 Table 5. Model comparison between three GLM methods, one nonspatial (LR) and  
 1708 two spatial (LRsp, LRar).

1709

<b>Models</b>	<b>Total</b>			
	<b>observations (presence)</b>	<b>Adjusted-<math>D^2</math></b>	<b>Nagelkerke's <math>R_N^2</math></b>	<b>AIC</b>
<i>Small islets</i>				
LR	181 (68)	0.059	0.103	229.370
LRsp	181 (68)	0.209	0.330	211.450
LRar	181 (68)	0.144	0.236	211.220
<i>Big islets</i>				
LR	810 (85)	0.264	0.336	408.750
LRsp	810 (85)	0.343	0.427	380.160
LRar	810 (85)	0.350	0.432	363.870

1710

1711

1712 Table 6. Average model performance results as assessed by means of two accuracy  
 1713 measures, the AUC score and Pearson's correlation coefficient, based on 10  
 1714 independent test data-sets (\*significant values for all test data sets at  $P \leq 0.05$ ).

1715

<b><i>Models</i></b>	<b><i>Test AUC</i></b>	<b><i>Pearson's r</i></b>
<i>Small islets</i>		
LR	0.638 ( $\pm 0.070$ )	0.259 ( $\pm 0.099$ )
LRsp	0.733* ( $\pm 0.052$ )	0.423 ( $\pm 0.091$ )
LRar	0.703* ( $\pm 0.084$ )	0.403 ( $\pm 0.134$ )
Maxent	0.573 ( $\pm 0.091$ )	0.113* ( $\pm 0.146$ )
<i>Big islets</i>		
LR	0.836* ( $\pm 0.044$ )	0.472* ( $\pm 0.069$ )
LRsp	0.897* ( $\pm 0.028$ )	0.494* ( $\pm 0.076$ )
LRar	0.899* ( $\pm 0.023$ )	0.523* ( $\pm 0.083$ )
Maxent	0.895* ( $\pm 0.039$ )	0.512* ( $\pm 0.058$ )

1716

1717 Table 7. Comparison of the average predictions of nest occurrence for four modeling  
 1718 techniques as assessed by Pearson's correlation coefficient (\*significant at  $P \leq 0.05$ ).

1719

<b>Models</b>	<b>LR</b>	<b>LRsp</b>	<b>LRar</b>	<b>Maxent</b>
<i>Small islets</i>				
LR	-	0.556*	0.655*	0.632*
LRsp		-	0.916*	0.537*
LRar			-	0.572*
<i>Big islets</i>				
LR	-	0.888*	0.856*	0.866*
LRsp		-	0.949*	0.811*
LRar			-	0.812*

1720

1721

1722

1723 **List of figures with caption**

1724 Fig 1. Moran's I values at different distance intervals for the presence of nests on  
1725 small and big islets. Statistically significant aggregation of nests was observed up to  
1726 135m distance for the small islets and up to 180m for the big islets (i.e. significant  
1727 positive values at  $P \leq 0.05$ ).

1728 Fig 2. Representative maps of average predicted values of nest occurrence for the  
1729 small islet group based on four modelling techniques; logistic regression (LR), logistic  
1730 regression with geographic coordinates as spatial terms (LRsp), logistic regression  
1731 with an autocovariate term (LRar) and Maxent.

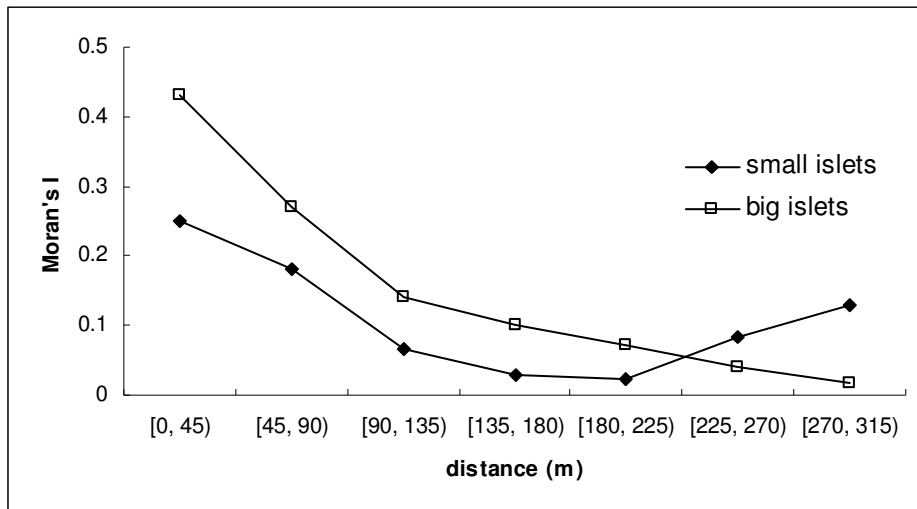
1732 Fig 3. Representative maps of average predicted values of nest occurrence for the  
1733 big islet group based on four modelling techniques; logistic regression (LR), logistic  
1734 regression with geographic coordinates as spatial terms (LRsp), logistic regression  
1735 with an autocovariate term (LRar) and Maxent.

1736 Fig 4. Spatial autocorrelation of the residuals in three GLMs (a) for the small islets  
1737 and (b) for the big islets. While the non-spatial GLM (LR) suffered from positively  
1738 autocorrelated errors, their spatial structure was successfully relaxed with the  
1739 addition of spatial terms (LRar and LRsp models).

1740 Fig 5. The cumulative frequency distribution of the probability of nest occurrence, for  
1741 three GLM techniques (LR, LRar, LRsp, see text for abbreviations) and a niche-  
1742 based method, Maxent, for the small islet group.

1743 Fig 6. The cumulative frequency distribution of the probability of nest occurrence as  
1744 modeled by three GLM techniques (LR, LRar, LRsp, see text for abbreviations) and a  
1745 niche-based method, Maxent, for the big islet group.

1746 Fig. 1



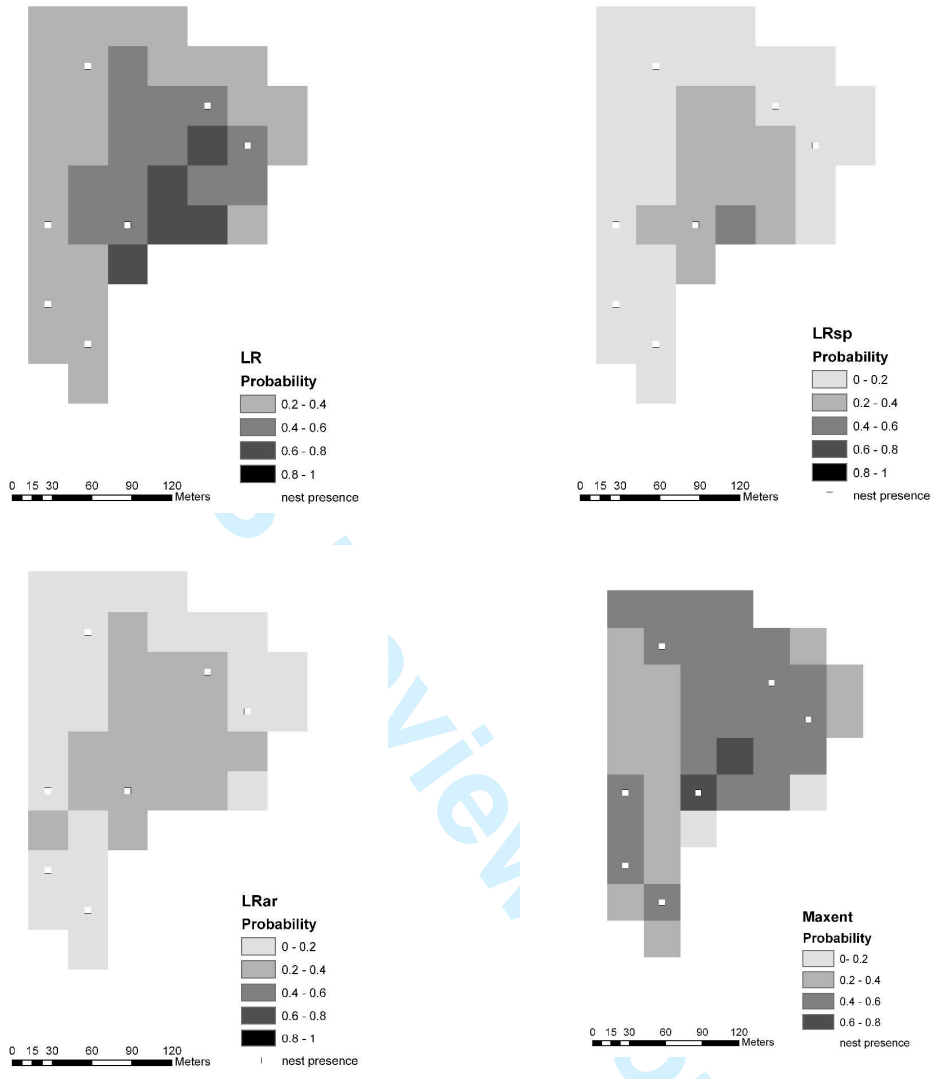
1747

1748

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1749

1750 Fig. 2

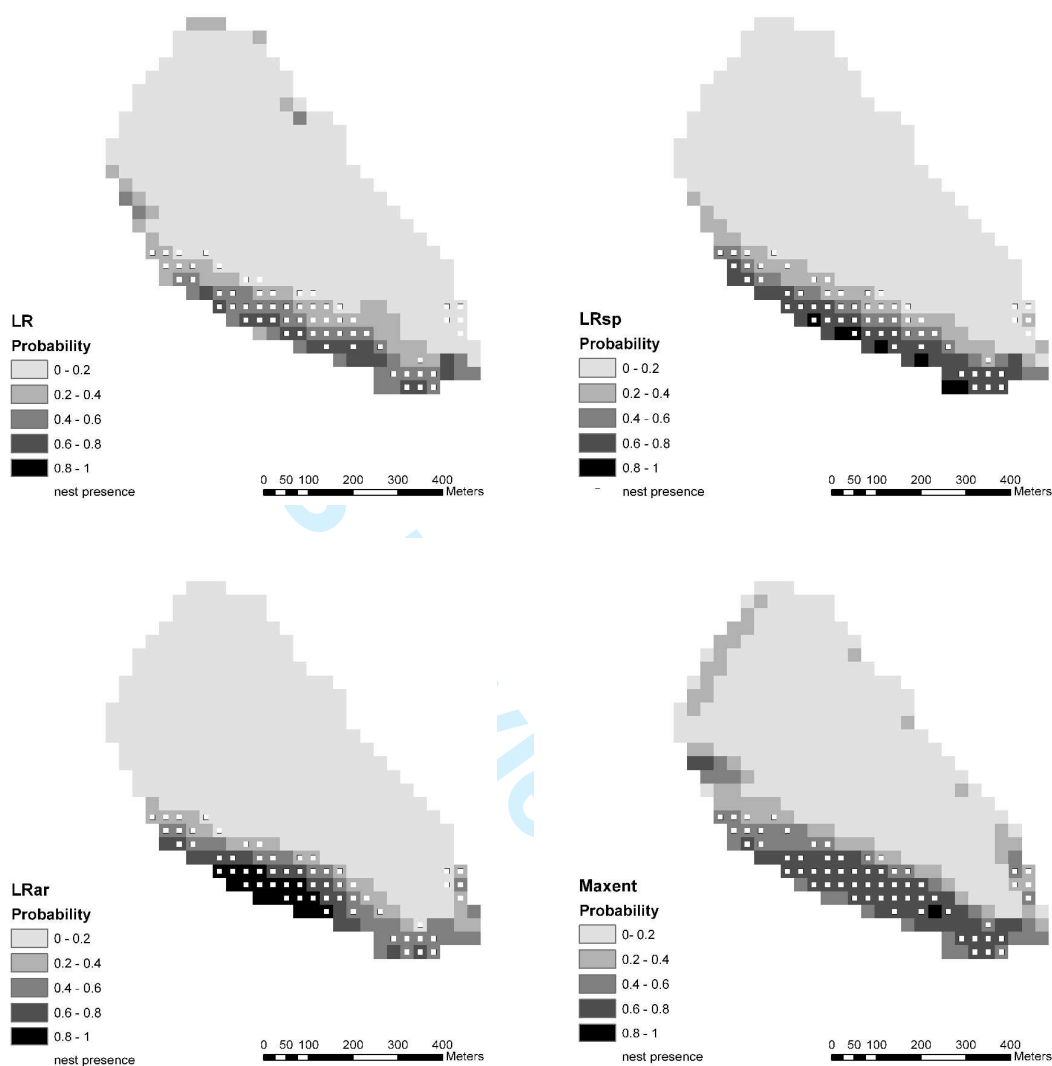


1751



1752

1753 Fig. 3

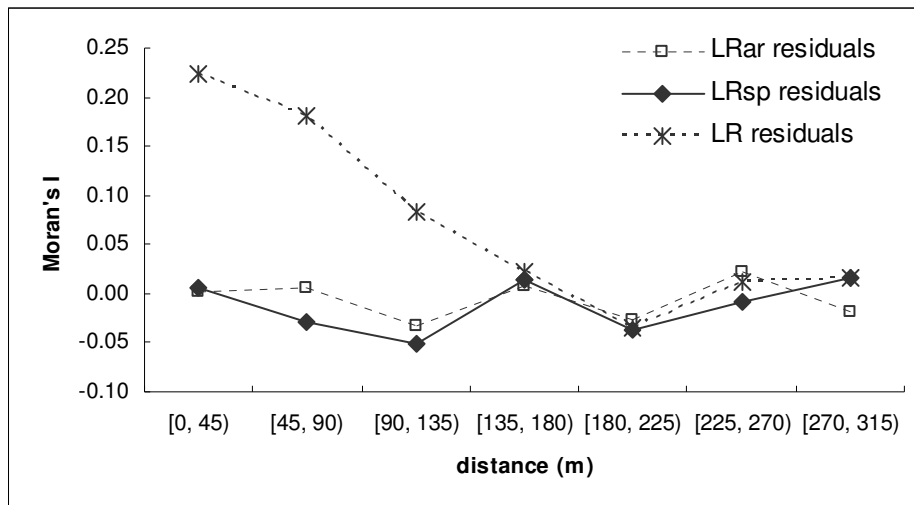


1754

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

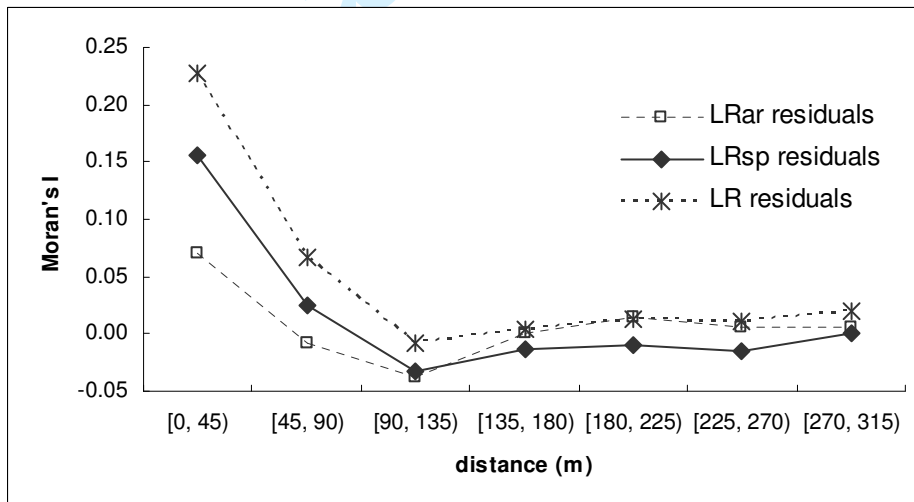
1755

1756 Fig. 4a



1757

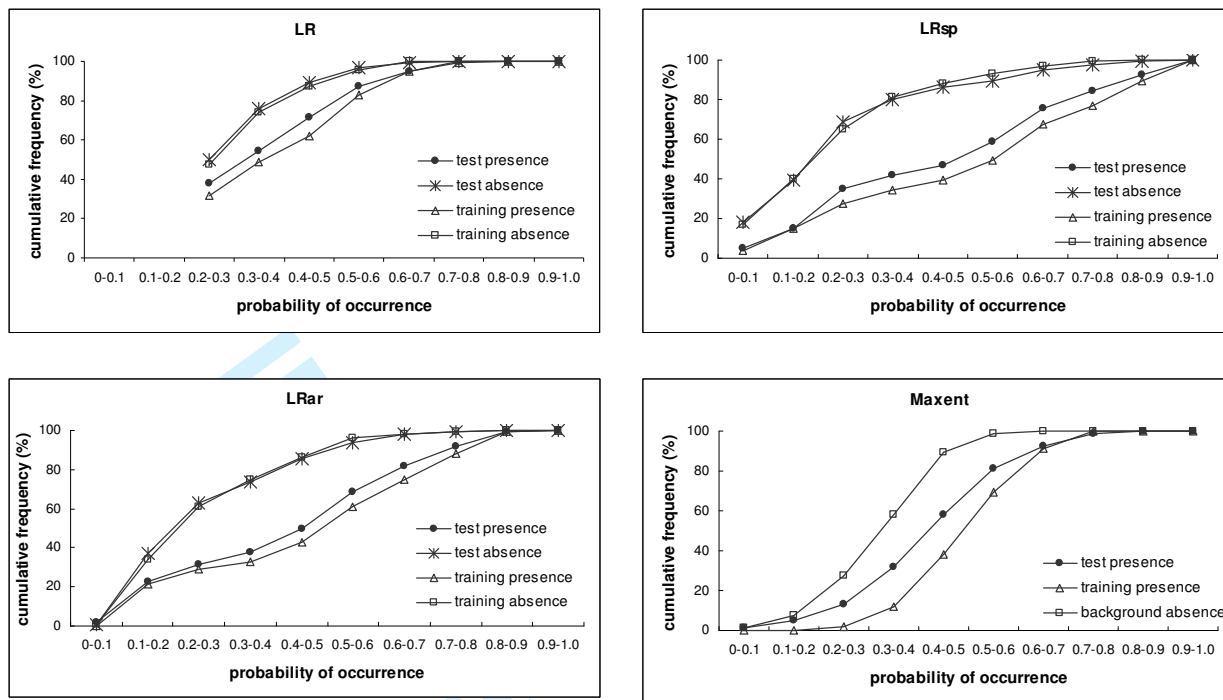
1758 Fig. 4b



1759

1760

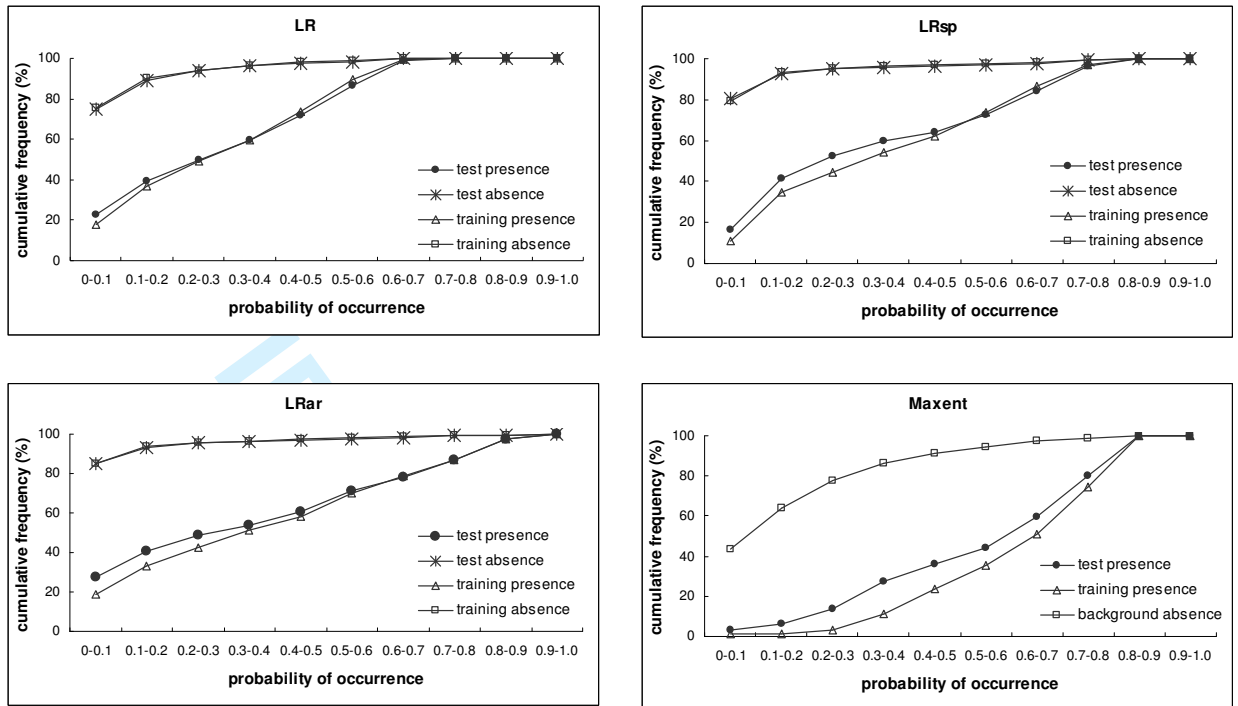
1761 Fig. 5



1762

1763

1764 Fig. 6



1765

1766