Modelling the habitat suitability of the brown bear (*Ursus arctos*) in the Prespa region.

Methods

We modelled the habitat suitability of brown bears in the Prespa region using 311 locations of detections of signs, considered as presence data and 1000 random pseudo-absences as background points. The covariates were extracted from the Corine Land Cover database (https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=download) and we calculated for each of the CLC classification, the percentage of a 1km x 1km grid cell covered by it (0 % meaning the cell had 0% of it covered by the defined land cover, and 100 % meaning the cell consisted uniquely of the defined land cover). The variables consisted therefore in 26 rasters of percentage of CLC land cover. We ran a first set of 10 "total models", each time running for 10 000 iterations. These total models included all 26 rasters of CLC land covers (Table 1). We selected the 6 variables that best explained the model by looking at their percentage of contribution to the model, and additionally took out the variables that did not have enough variation in the landscape to explain bears' habitat suitability. The percentage of contribution is a representation of how much better the model fits the data when a variable is included in the model. Finally, we ran a final model set consisting of only the best 6 variables that best explained the data. The model set consisted in 10 repetitions of this model, each running for 10 000 iterations. We plotted the estimated map of habitat suitability for bears in the Prespa region (Figure 1) and plotted the response curve of habitat suitability depending on each covariate (Figure 2). We estimated the percentage of contribution as well as the permutation importance for each of these variables (Tables 2 and 3). The permutation importance is another measure of the importance of variables by varying the values and measuring the decrease of accuracy in the model to fit the data. An increase in the permutation importance means that the model depends heavily of this variable. Finally, to validate the

model, we calculated the sensitivity which is the proportion of correctly predicted presences and the specificity, which is the proportion of absences correctly predicted (Table 4).

Table 1: All CLC covers that were present in the Prespa region and their corresponding code

CLC classification	Code
Continuous urban fabric	111
Discontinuous urban fabric	112
Industrial or commercial units	121
Road and rail networks and associated land	122
Airports	124
Mineral extraction sites	131
Green urban areas	141
Sport and leisure facilities	142
Non-irrigated arable land	211
Permanently irrigated land	212
Vineyards	221
Fruit trees and berry plantations	222
Pastures	231
Complex cultivation patterns	242
Land principally occupied by agriculture, with significant areas of natural vegetation	243
Broad-leaved forest	311
Coniferous forest	312
Mixed forest	313
Natural grasslands	321
Moors and heathland	322
Sclerophyllous vegetation	323
Transitional woodland-shrub	324
Beaches, dunes, sands	331
Sparsely vegetated areas	333
Inland marshes	411
Water bodies	512

Results

The variables that appear to best fit the data were water bodies, transitional woodlands-shrubs, natural grasslands, fruit trees and berry plantations, complex cultivation patterns and pastures. Water bodies appeared to negatively influence the suitability of bears' habitat, transitional woodlands didn't appear to have an influence on bears' habitat, however remains important in the model, given its score in variable permutation test. Natural grassland appeared to positively influence bears' habitat suitability, fruit trees and berry plantations appeared to negatively influence the habitat suitability such as complex cultivation patterns and pastures. The sensitivity appeared relatively correct for seven of the ten models, the specificity appeared to be lower, however, in the case of a model without information about true absences, such measure can be considered of a lesser importance than the sensitivity. As a conclusion, it appeared that the model performs relatively well and fits well to the data. This model is a first attempt at modelling brown bears' habitat suitability in the Prespa region, however some refinements need to be done at the variables level in order to better and more precisely reflect the distribution of the species in this region.

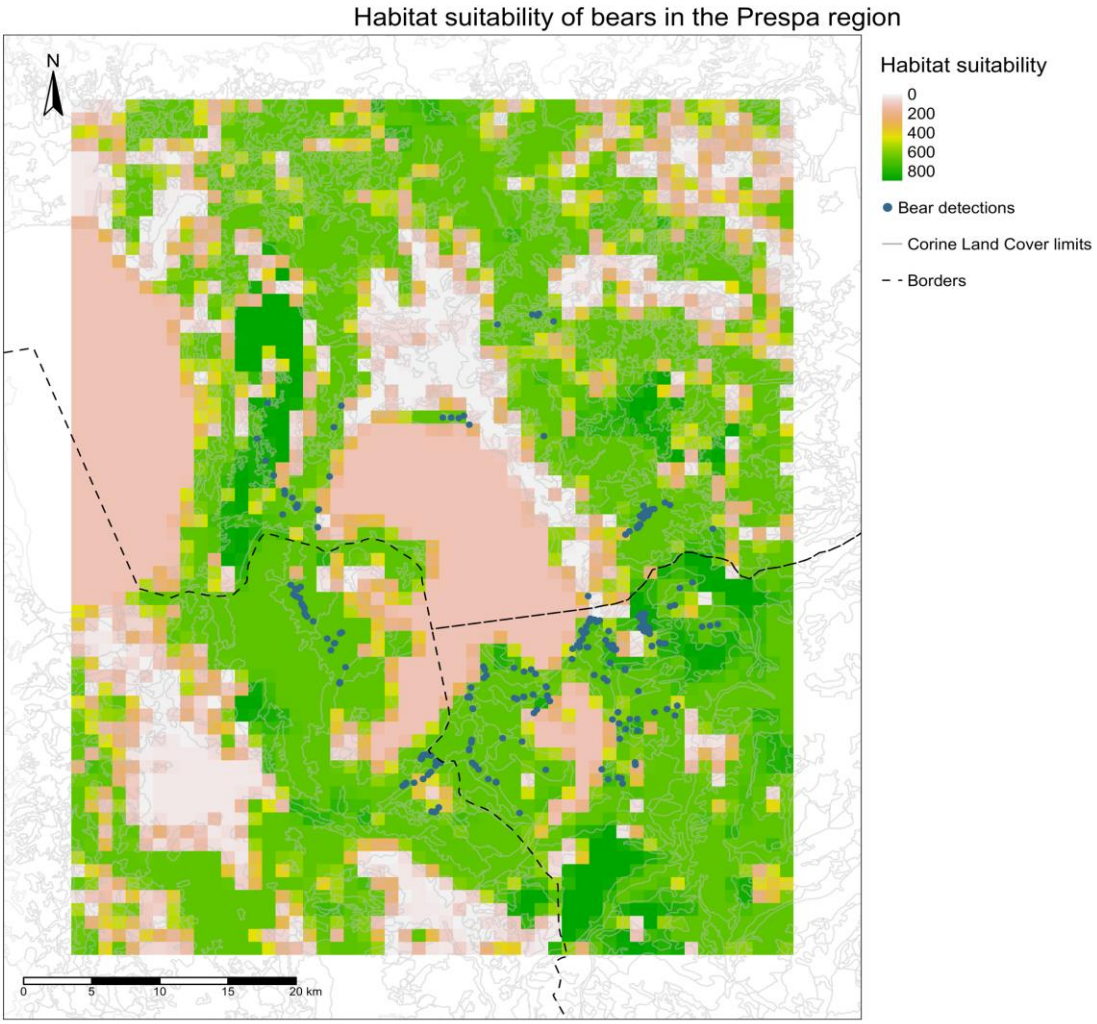


Figure 1: Habitat suitability map of the brown bear (Ursus arctos) in the Prespa region. Blue points represent the locations of detection of presence signs.

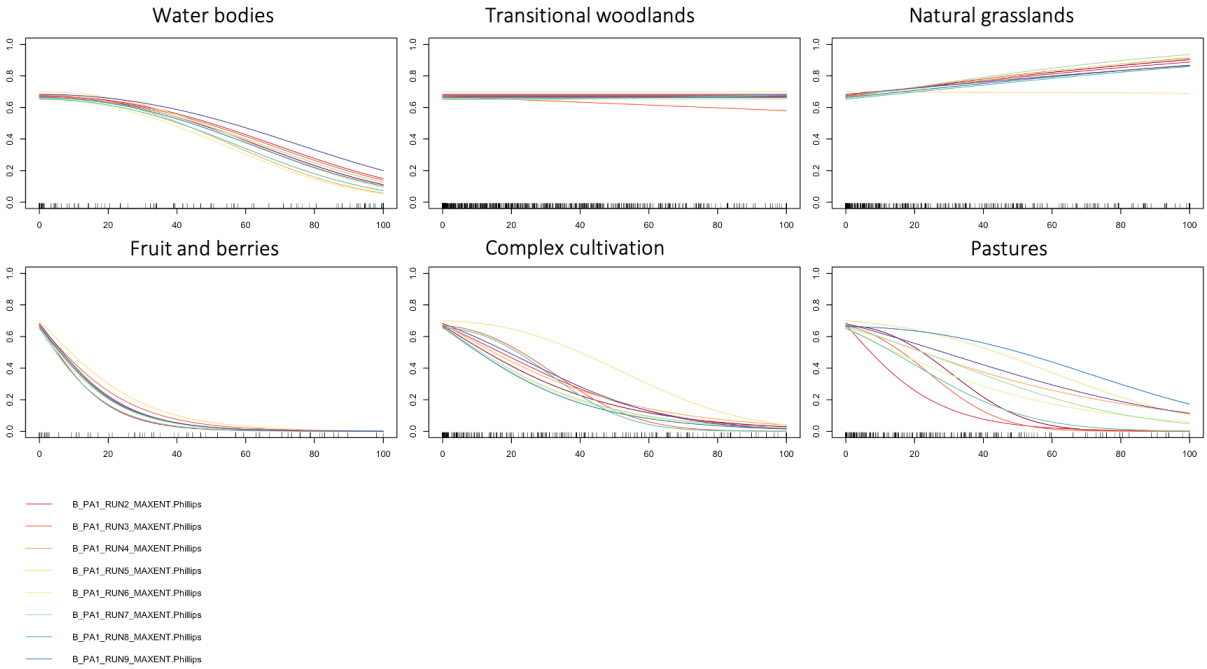


Figure 2: Response curves of the predicted habitat suitability of brown bear depending on the six selected covariates.

Table 2 : Variable contributions of the six selected variables for each ten runs of the selected model

CLC variable	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Fruits and berries	11.66	13.24	11.11	12.38	11.97	11.37	10.87	14.16	12.34	16.58
Pastures	18.47	22.28	18.35	11.59	12.57	7.97	10.97	15.29	5.70	11.03
Complex cultivation patterns	27.04	23.32	23.04	29.51	25.98	29.68	29.90	25.13	37.96	34.60
Natural grassland	11.93	12.72	17.76	7.30	5.97	14.67	16.60	9.21	9.57	10.90
Transitional woodland / shrub	6.75	9.41	10.33	7.72	9.94	6.20	10.09	7.55	5.37	7.65
Water bodies	24.16	19.03	19.41	31.51	33.58	30.11	21.57	28.66	29.07	19.24

Table 3: Permutation test for the six selected variables for each of the ten runs of the selected model

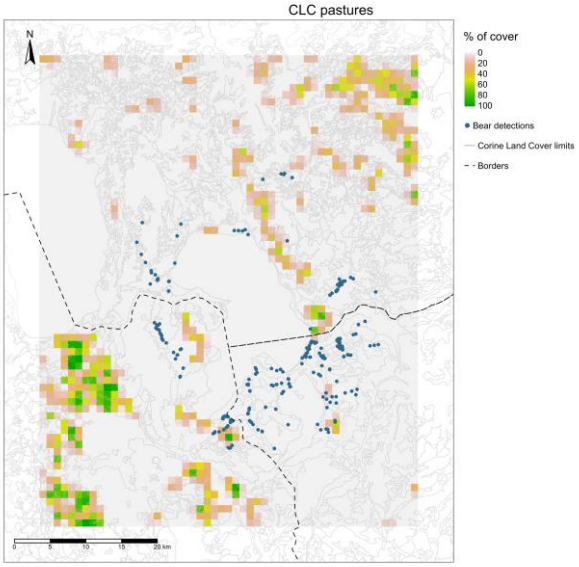
CLC variable	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Fruits and berries	4.36	10.37	8.38	9.43	7.89	9.01	8.13	5.59	8.97	9.65
Pastures	11.11	23.34	10.80	10.27	13.16	12.07	9.54	9.67	12.47	16.48
Complex cultivated patterns	9.24	15.56	13.51	23.04	24.38	20.17	24.70	17.67	26.76	18.21
Natural grassland	21.65	12.89	15.95	14.69	9.97	15.41	19.57	21.79	12.387	14.24
Transitional woodland / shrub	27.49	20.04	37.13	19.91	10.09	21.64	19.84	19.69	21.48	19.42
Water bodies	26.15	17.80	14.22	22.66	34.50	21.70	18.21	25.59	17.92	22.00

Table 4: Sensitivity and Specificity of each of the ten runs of the selected model

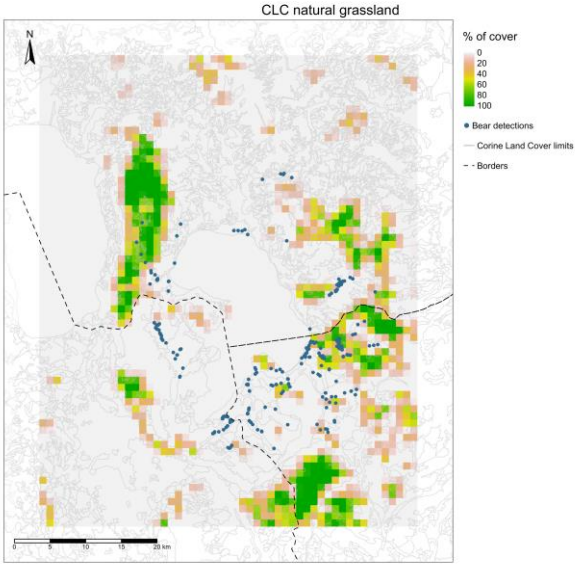
	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Testing.data	0.25	0.27	0.25	0.24	0.29	0.26	0.28	0.25	0.29	0.28
Cutoff	516	552	513	550	524	394.5	507	520	516	523
Sensitivity	85.81	30.97	63.87	29.68	84.52	90.97	83.23	40	83.23	87.74
Specificity	52	91.8	67.6	91	57.2	48.6	55	83.8	57.8	53.2

Appendix: Maps of the selected variables

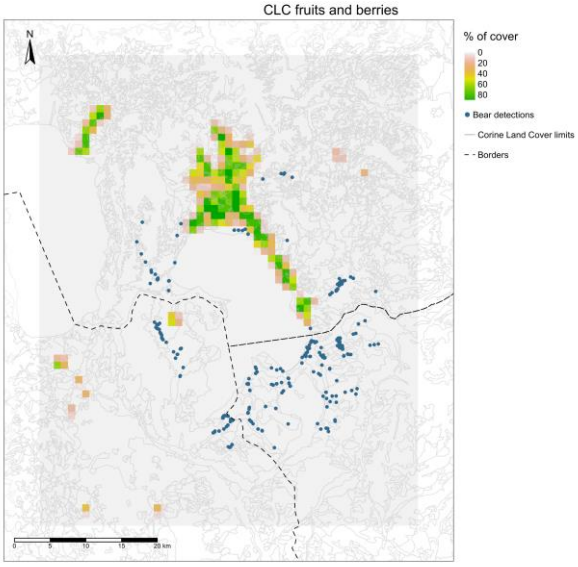
CLC pastures



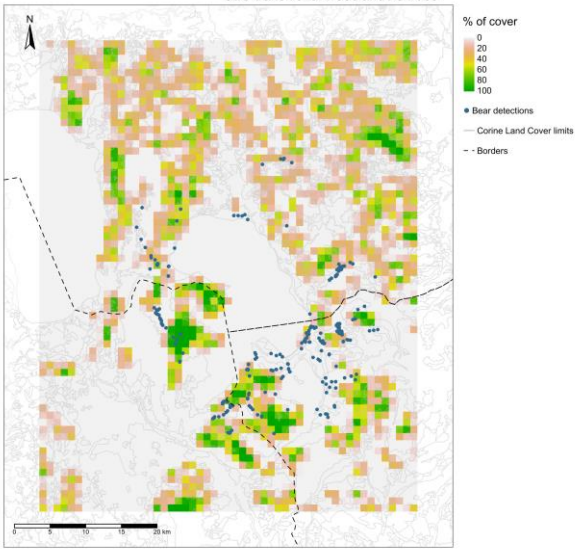
CLC natural grassland



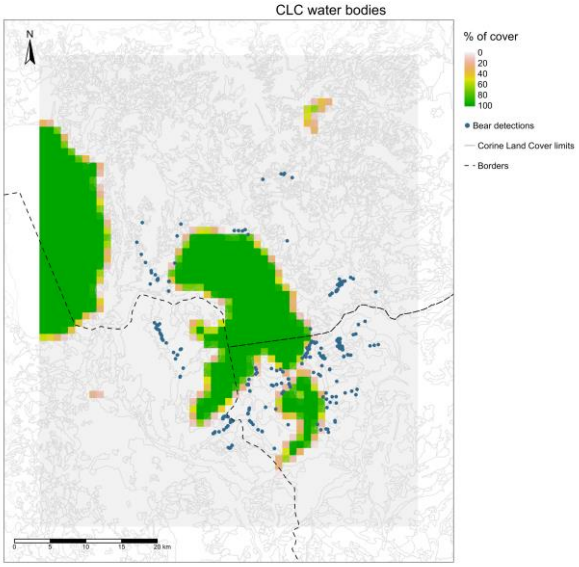
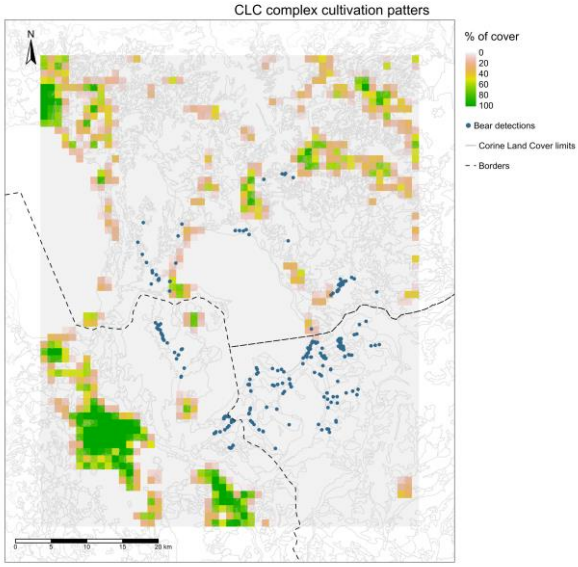
CLC fruits and berries



CLC transitional woodland / shrubs



CLC water bodies



CLC complex cultivation patters