



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

School of Biology

**ASSESSING THE SCALE OF INFLUENCE OF LANDUSE
DESCRIPTORS ON BIRD DISTRIBUTION IN SWITZERLAND**

*1st step project,
Master in Behavior, Evolution and Conservation, University of Lausanne, 2020*

by

Aurélien ROCHAT

Director : Prof. Antoine Guisan

Supervisor : Dr. Olivier Broennimann

Department of Ecology and Evolution

December 2020

Abstract

Biodiversity is increasingly affected by human activities and thus it is becoming urgent to preserve natural habitats. For this purpose, predictions from species distribution models are very useful in conservation. However, to obtain correct maps we need to select the appropriate environmental predictors reflecting information at optimal scale. In this context, we propose a two steps framework to model the distribution of 170 Swiss breeding birds using landuse descriptors. Firstly, using univariate models, we selected the best scale of influence for each variable. Secondly, using these scales, we fitted a multivariate model to map species potential distributions. The distance of influence of the environment was highly heterogeneous between variables and between species, since their respective scale showed large differences. This suggests that these scales are variable- and species-specific. In general, wet habitat and human infrastructures were the most relevant variables in our multivariate model. Overall, our study highlighted the importance of choosing the appropriate scale for each predictor. In addition, we also showed that birds are especially affected by human factors and that natural habitats such as wet lands need protection. As perspectives, our findings could help management decisions in conservation planning.

Keywords : SDMs, habitat suitability, focal variable, highly mobile species

Introduction

Loss of biodiversity has increased recently and has become a major issue in ecology and conservation and more globally for our society (e.g., Maxwell et al., 2016; IPBES, 2019). This decrease in diversity has heavy consequences which are already visible on Earth (Haddad et al., 2015; Foley et al., 2015; Newbold et al., 2015). Since the main drivers of this so-called 6th extinction are human activities, it is our responsibility to try to stop this erosion of diversity (IPBES, 2019).

To keep intact all living organisms, we should enhance the management and the conservation of ecosystems by the creation of new protected areas. To reach this goal, we need both precise and up-to-date species distribution maps (Guisan et al., 2013; Pollock et al., 2020). By giving to the practitioners clear, informative and easy-to-read maps, we could contribute to prevent biodiversity loss. Ideally, scientists should provide predicted distributions that can be implemented in conservation planning to protect species and their habitats - as in the ValPar national project (see www.valpar.ch for more information) (Honeck et al., 2020, Pollock et al., 2020).

Among the animal kingdom, we know the importance of birds, being a large and highly diverse group that colonized almost all ecosystems, with mostly mobile species with specific needs and different home ranges and habitats (Schoener, 1968; Haskell et al., 2002; McPherson et al., 2019). Therefore, they are good indicators of the biodiversity situation (Maggini et al., 2014; Keller et al., 2020). In addition, lots of data are freely available thanks to numerous and widely avian monitoring programs (e.g. "Monitoring Häufige Brutvögel" by the Swiss Ornithological Institute, see www.vogelwarte.ch).

Some methods already exist to improve management and conservation planning. To protect the biodiversity, one option is to map the habitats, and in order to map them, we need to model them. To do so, we can use Species Distribution Models (SDMs). SDMs are powerful tools which benefit from recent advances in modelling techniques (Guisan et al., 2013; Guisan & Thuiller, 2005; Guisan & Zimmermann, 2000; Pollock et al., 2020). They are now widely used in different fields with concrete applications (Guisan et al., 2013).

However, generalization and simplification must be done when applied to numerous and heterogeneous species, which leads to a loss of information or bias (Hortal et al., 2015; Fritsch et al., 2020). A difficult step is to select the appropriate environmental variables, reflecting environmental information at optimal scale (Chase et al., 2018). This is especially the case with highly mobile species as birds since the distance of influence of the surroundings can vary a lot between variables and between species (Scherrer et al., 2019; Tehrani et al., 2020). As a consequence, when dealing with such species, it is not sufficient to account only for the environmental information from the observation site, since we also need to know the influence of the neighborhood. Indeed, within its home range, a bird can reach landscape elements in the surroundings around its observation site. Therefore, it is problematic to consider the same scale for all predictors whose influences might be different. This practice leads often to overestimations of species distribution (Fournier et al., 2017).

To avoid this problem, one method - sometimes called a “multi-scale approach” (e.g. Graf et al., 2005) – consists of taking into account the correct scale for each predictor in function of the ecology and specific behavior of a species. This solution gives more flexibility to the model which improves the predictions (Fournier et al., 2017). Indeed, when modeling with a large set of environmental variables, each of them can have different aspects and play a specific role in the habitat networks (Vicente et al., 2014). Thus, instead of using the same distance of influence for all predictors, each of them keeps

its optimal scale. To integrate these variables in the model, we determine their optimal scale of influence by using a focal function of different sizes. For each pixel, circular moving windows of concentric various radius quantify the proportion of each land use class (another value such as the median or the majority could have also been calculated) in surroundings cells. Thus, by summarizing the proportion of each variable in the neighborhood, the optimal radius could then be determined with statistical models. In doing so, we capture the influence of the environment and are able to find the optimal focal size for each variable and each species depending on their ecology and behavior (Bellamy et al., 2013; Scherrer et al., 2019; Tehrani et al., 2020; Bellamy et al., 2020).

Modelling species habitats to understand the importance of ecosystems networks is part of a wider goal of saving our biodiversity. In that perspective, this 1st step research takes part into a Swiss national project called *ValPar*, of which one of the goals is to map ecosystems - also called Green Infrastructures (GI) (Honeck et al., 2020). In the following study, we model the distribution of Swiss breeding birds by comparing the influence of environmental variables of different scales on the predictions. More precisely, in this 1st step project, using the framework from Scherrer et al. (2019), our aim is first to identify the optimal neighborhood distance (or best focal size) for each species and each landscape variable. We then provide prediction maps from the models incorporating these variables at their optimal scale.

Our results suggest that the optimal focal size is species-specific and that some variables such as wet lands and human infrastructures are more relevant. This approach was never done before at such a national scale and with so many species. Future works could benefit from our results by enhancing some SDMs with appropriate scale for a specific specie in relation to its range size or by re-using our modelling framework.

Methods

Species data

We used data from a breeding bird survey ("Monitoring Häufige Brutvögel" MHB) organized annually by the Swiss Ornithological Institute since 1999. For our study, we used more specifically the data from 2013 to 2016 which cover the 170 breeding birds of Switzerland. In total, we had 368'390 observations (average counts) made in 2318 different sampling sites of 1 km² covering all the country. Each single observation is an average of the annual counts made during this time period (2013-2016). These data are not publicly available.

Environmental variables

As part of the national *ValPar* project, the study area covered all Switzerland with a total surface of 41'285 km². Our environmental variables are represented by raster layers of 41'285 pixels of 1000 m x 1000 m covering all the country.

Our environmental data included two types of variables: focal and non-focal. Each variable is at a 1000 m resolution and a detailed description is shown in Table 1. The eleven focal variables were all land use and land cover classes. We used a circular moving window of different radius to capture the

proportion of each class in the neighborhood around each cell. The eight radius measured 0, 1000, 2000, 3000, 5000, 10'000, 15'000 and 25'000 m. More precisely, we used raster layers from different federal sources (or rasterizations at 25m of vector sources; see details in Table 1) that we aggregated at a 1000 m resolution as our base layer (i.e. not accounting for information in surrounding pixels; which corresponds to the radius size of zero in the analyses hereafter). Then, using a circular moving window of increasing size, we calculated the proportion of pixels of each land use class that are inside this window and put this value in the central pixel. We took into account only pixels that had at least half of their surface in the circle (for details about the method used, see the R code in Appendix). The 8 radius were chosen to be representative of the variation of home ranges in birds, covering approximately the whole variation in sizes from minimum to maximum values (Schoener, 1968; Haskell et al., 2002; McPherson et al., 2019). The maximum radius of 25 km corresponds to the maximum home range with a squared area of 2500 km² ($= 2 \times 25^2$).

We further used 6 non-focal variables representing climate, topography and human influence: average temperature [°C] and average precipitation [mm], slope [°] and topographic position [unitless], human traffic road noise [dB] and population density [habitants/ha].

Most of the variables were chosen to be similar to the ones used by Scherrer et al. (2019) in their study on bats, since they have an influence on highly mobile species like bats or birds. We verified their pairwise correlations (Appendix; Figure S1). Some variables were adapted to be more specific to birds (the correspondence with those from Scherrer et al. are given in Table 1). In the human influence category, we selected the average traffic road noise instead of the distance to roads, since the noise perturbations are known to disturb the breeding behavior of birds (Meillère et al., 2015). The other human influence variable is population density, which also account for the presence of pets in proximity, especially cats who are bird predators (Loss et al., 2013).

Table 1 : Name and description of each environmental variable used in the models.

* If not precised, all the variables were at a 100 m resolution (from the source) and were aggregated to a 1000 m resolution.

The corresponding variables of the ones used in Scherrer et al. (2019) are : Popdens is a proxy for buildingsED Euclidian distance between the closest building and the center of the cell. Roadnoise is a proxy for roadsED Euclidian distance between the closest road and the center of the cell. Water (standwater and runwater) is a proxy for waterED Euclidian distance between the closest water source (stream or lake) and the center of the cell. Foreedge is a proxy for forestED Euclidian distance between the closest forest and the center of the cell. Foreedge is a proxy for canopy Variance of canopy height at a 100 m resolution, calculating from a 1 m resolution raster masked by forest. Rock (or bare) is a proxy for ndvi Normalized difference vegetation index at 100 m resolution. Aggregate from 10 m resolution. Rock is a proxy for ndvi_focal Mean of normalized difference vegetation index at different focal scale.

Type	Name	Description – each layer is at a 1000 m resolution *
Land cover/ land use (focal variable)	alppast	Proportion of alpine pasture at different focal scale (Source : GeoStat OFS)
	confor	Proportion of coniferous forest (Source : NFI)
	culti	Proportion of agricultural land (GeoStat OFS)
	decfor	Proportion of deciduous forest (NFI)
	foredge	Proportion of cells including a forest edge or treeline (TLM3D)
	huminfrastr	Proportion of human habitations and infrastructures (GeoStat OFS)
	open	Proportion of open area (meadows, pasture) (GeoStat OFS)
	rock	Proportion of naked area (mainly rocks, bare ground) (TLM3D)
	runwater	Proportion of running water (river) (TLM3D)
	standwater	Proportion of standing water (lake) (TLM3D)
	wethabitat	Proportion of wet habitat (TLM3D)
Climatic	tmin	Mean of daily minimal temperature between May and August—averaged from 1981 to 2010 (CHCLIM25)
	prec	Sum of daily precipitation between May and August—averaged from 1981 to 2010 (CHCLIM25)
Human influence	roadnoise	Daytime road traffic noise average between 6h - 22h (BAFU)
	popdens	Population density in 2013 (OFS StatPop13)
Topographic	slope	Slope inferred from a digital elevation model at 25 m resolution. Aggregate to 1000 m resolution (WSL)
	topos	Topographic position at 25 m resolution aggregate to 1000 m resolution (WSL)

Univariate models

The optimal window size of each focal variable was found by running univariate regression models (GLM with a Poisson distribution) for each focal distance and for each species (Bellamy et al., 2013; Scherrer et al., 2019). We run (170 species x 11 focal variables x 8 window sizes) 14'960 models using the *stats* package in R (3.6.3 version; R Core Team, 2020). As we have no independent dataset, we performed for each of the 14'960 models 100 split-sample cross-validation runs with 80% calibration and 20% evaluation. To find the optimal scale, we selected the one that has the highest Spearman correlation between the observations and the predictions of the model (Guisan & Zimmermann, 2000). We also selected the optimal scale based on the adjusted explained deviance (D^2 or explD^2) of the GLM and compared these results with the ones found with the correlation approach (Appendix; Table S1). To test if the optimal focal sizes differ significantly between variables, we used the *multcompView* package (version 0.1-8) with the default parameters (i.e. Tukey HSD test with a p-value threshold of 0.05). Overall, we followed the framework from Scherrer et al. (2019). However, there is a divergence from their work, since we used average counts and not presence-absence observations. For this reason, we used Spearman correlation and adjusted explD^2 to select the best models instead of AUC (Guisan & Zimmermann, 2000). This is actually an improvement compared to their study, because the information we use to characterize species is not binary presence-absence, but includes the effect of abundance. As a consequence, analyzing counts data prevented us of using the *biomod2* package, so that we had to code our own implementations of the models (Appendix; R code).

Multivariate model

To predict bird distributions over Switzerland, we calibrated models using the 6 non-focal variables and the 11 focal variables with their optimal focal size as found in the univariate models. To do so, we implemented a multivariate GLM model with a Poisson distribution using a lasso regression from the *glmnet* package (version 3.0.2). We included in the model the linear and quadratic terms for all explanatory variables. The advantage of the lasso regression method is that all the variables can be incorporated directly into the model without preselecting them. The lasso regression will select some variables and discard the others by putting their coefficient to zero. The method includes an iterative procedure in which increasing value of a penalization term called “lambda” are calculated and a model is fitted. In each iteration, each variable gets a coefficient which is optimized to fit the model while the other coefficients are fixed. This process continues until convergence (i.e. the stability of all coefficients) (Friedman et al., 2019).

In addition to that, we did a K-fold cross-validation to select the right model among the 100 models that were fitted with the *glmnet*. To select the best model along the increasing lambda values, we could choose between *lambda min* (minimum mean cross-validated error), or *lambda 1se* (most regularized model which means that the error is within one standard error of the minimum). To obtain at the end the right model, the choice was made with *lambda 1se*, which excludes more variables than with *lambda min* (Friedman et al., 2010).

Finally, using the predictions of this model, birds' environmental suitability values were generated for all the points of Switzerland using the *raster* package, version 3.0.12 (Appendix; R code).

Results

As we cannot present the results for all the species (since it is a short article for a 1st step project), we had to make a selection. Therefore, in addition to general results, we present in details only some examples that are relevant for our analyses and discussion. These examples are shown in Box 1 and Box 2. Results for all species can be found in Appendix (Table S1 and S2).

Univariate models – scale selection

For each of the 170 breeding birds and the 11 focal variables, we found the appropriate focal scale (among the 8 radius sizes) by selecting the one that has the highest Spearman correlation between the real observations and the predictions from the GLM models (Appendix; Table S1). In overall, we observed no specific distance of influence that would be optimal for all the variables (Figure 1). Indeed, we observed high heterogeneity in optimal scales - both between and within variables - although some variables presented a smaller range of focal sizes compared to others (e.g. alppast and confor). These heterogeneous results confirm that the scales of influence of environmental variables are species-specific (with variation within variables) and variable-specific (with variation between variables). Indeed, we tested if these optimal focal sizes differ significantly between variables and visualized the multiple paired comparisons between them. (Figure S2). Our results (both with letter-based representation and graphical grouping T) suggest some significant differences (with p-values < 0.05). Therefore, this justifies the use of a unique combination of variables for each species in the multivariate model by selecting the optimal scale among the eight environmental layers. For instance, confor, alppast, rock and culti (which are grouped together in letter “a”) had significant smaller focal sizes than variables in groups f and g (Figure S2). On the opposite, open had a significant higher focal size than the seven smallest variables (from confor to foredge).

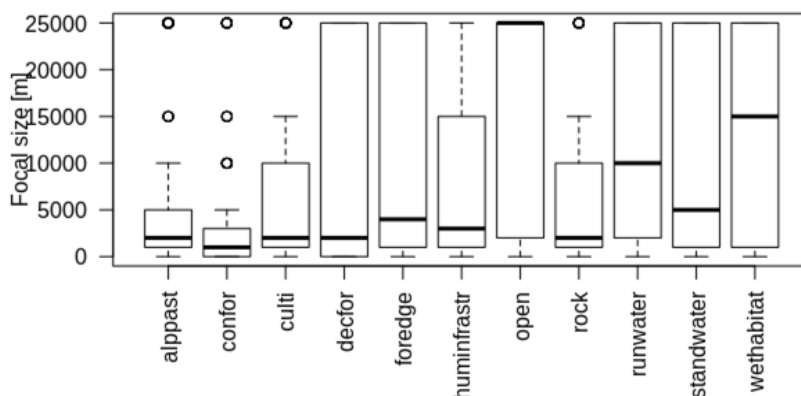


Figure 1 Boxplot of the optimal focal sizes of each focal variable for the 170 species, as found with the univariate models. For variable names and details see Table 1.

Differences in detected optimal focal size between variables are not explained by a same pattern of differences in model accuracy (i.e. correlations between predictions and observations have an interval of values almost identical for all variables; Figure S3). However, we noted an exception for run water, stand water and wet habitat (all in the water category) since their model accuracy are clearly lower.

In general, the method to determine the optimal radius of a focal variable (correlations between observation and model prediction vs. explained deviance) had not a strong influence (see Table S1).

Interestingly, with these detected focal sizes we may obtain some information on the home range of each species – related to their ecology – and on the distance of influence of each landuse descriptor – related to the type of habitat. For instance, the median of the coniferous forest (confor) showed the lowest optimal scale indicating a small distance of influence whereas open area (open) showed the largest (Figure 1). Among all the species, the Short-toed Treecreeper has the smallest scales and the Common Woodpigeon the largest (Table S1). Thus, it is a way to document the ecology of birds in relation to their habitat and their distance of influence.

The overall median of the detected optimal focal size across species and variables was 3000 m. Compared to non-focal variables usually used in other SDMs, this suggests that the neighborhood indeed has an influence. Focal variables should thus need to be taken into account to improve the predictive power of the models of mobile species. Nevertheless, the distance to calculate focal variables is not general, and each variable has its own scale of influence, which is even species-specific. The large optimal focal size we detected could be mainly explained by the high mobility of birds compared to other taxa often used in SDMs such as plants that have theoretically a home range size of zero.

To investigate in more details the optimal focal sizes and to see whether they are linked to some birds' traits, we did supplementary analyses for 3 groups that differ in their ecology – common, water, altitude and alpine area – each composed of 8 species. The “common” group included the most common species in Switzerland with more than 20'000 counts in total (from our data set) : Eurasian Blackcap, Eurasian Blackbird, Great Tit, Coal Tit, House Sparrow, Common Chaffinch, European Robin and Northern Wren. The “water group” included the most frequent water birds and ducks found on lakes (based also on the number of counts) : Mallard, Mute Swan, Common Coot, Common Moorhen, Great Cormorant, Little Grebe, Great Crested Grebe and Goosander. The “altitude and alpine group” included species that live in mountains area and at high altitude (Mullarney et al., 1999) : Alpine Accentor, Red Crossbill, Yellow-billed Cough, Red-billed Cough, Howfinch, Rock Ptarmigan, White-winged Snowfinch and Black Grouse.

On Figures 2 to 4, we observe that the focal sizes differed clearly between groups. These patterns showed also dissimilarities compared to the general results (Figure 1). Common birds have very small

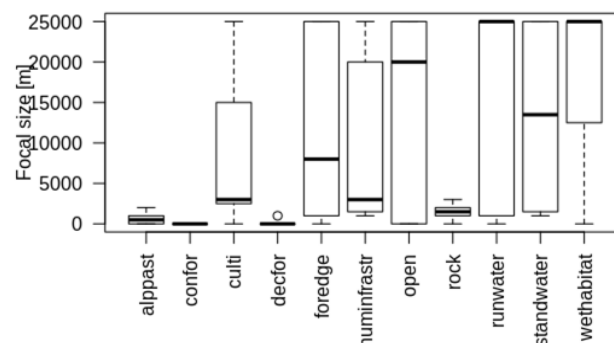


Figure 2 Boxplot of the optimal focal size for each variable, for the common group.

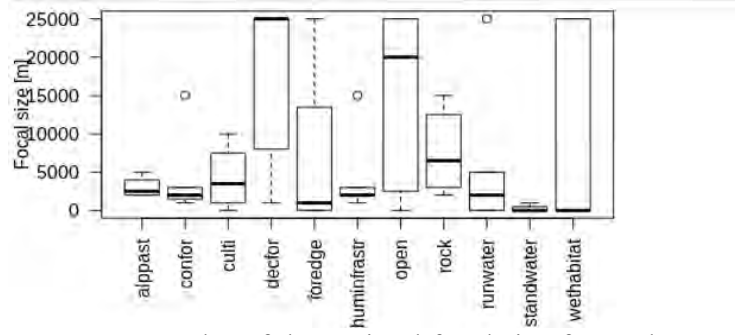


Figure 3 Boxplot of the optimal focal size for each variable, for the water group.

optimal scales for alppast, confor, decfor and rock, both in the median and the variance (Figure 2). On the contrary, the distances of influence (and the variance) of open, run water, stand water and wet habitat are much larger. The biggest differences with the water birds group are seen with decfor, run water, stand water and wet habitat which have almost opposite focal sizes. In this group, it is stand water that has the smallest distance of influence and decfor the largest (Figure 3). In the third group (alpine birds), their overall focal sizes are quite similar to the common birds, with even more extreme values (both for smallest and largest scales) but less variation within variables (Figure 4). The largest distances of influence are also seen with open, run water, stand water and wet habitat.

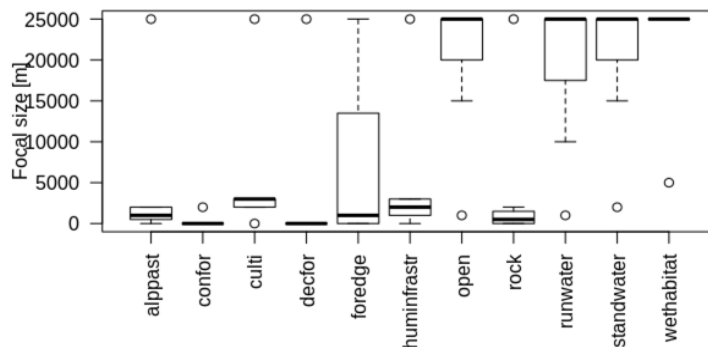


Figure 4 Boxplot of the optimal focal size for each variable, for the altitude group.

In the results for all univariate models (i.e. not only the ones with optimal radius), for a given species, model accuracy as a function of radius is similar when based on correlation between observation and predictions or with the explained deviance (see Figure 5 for an example species). We observed clear patterns, with similar response shape, minimum and maximum values, indicating that these response curves have ecological meanings. Thus, although these patterns show high variability between variables and between species, they give important information about how the influence of the neighborhood varies in function of the distance.

Box 1 The Common Blackbird

Below we illustrate the procedure of the selection of the optimal focal size of environmental predictions with the Common Blackbird, with the human infrastructures variable as example (Figure 5). To select the optimal focal size, we took the radius that had the highest correlation (i.e. the better model accuracy) which is here with 3000 m. The same procedure is done separately for all the other variables. With the explained deviance instead of the correlation, we found the optimal scale at 5000 m. Although we found a different focal size with this method, both curve shapes looked similar with a sharp increase at the beginning (for small distances), followed by a maximum peak, and with a decreased at the end. These unimodal curves - often found in biology - seems to highlight an ecological optimum reached at the corresponding focal size.

On Figure 6 are (a) the optimal scales for all variables and (b) their corresponding correlation representing the model accuracy for the Common Blackbird. Being part of the common birds group, we observed the same high distances of influence for open, run water, stand water and wet habitat, with foredge in addition. Regarding the correlations, human infrastructures has the highest one, followed by culti, decfor and rock whereas run water and wet habitat have the smallest ones.

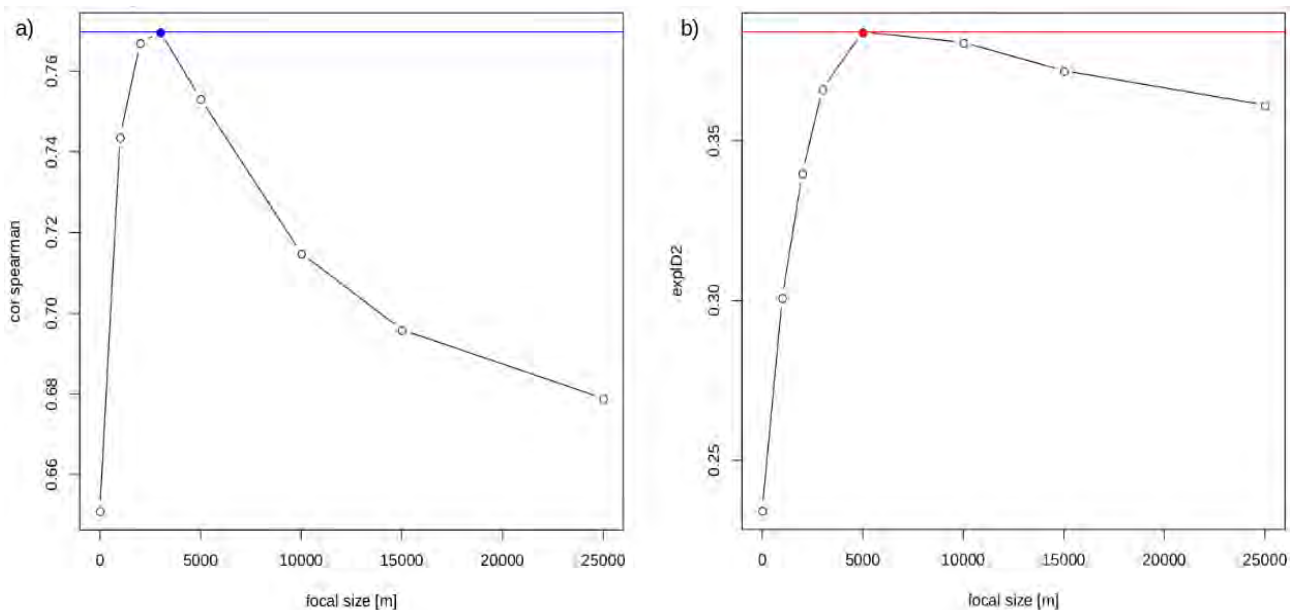


Figure 5 Illustration of focal variable selection for the Common Blackbird

a) Spearman correlation between the observations and the predictions (representing the model accuracy) and b) adjusted explained deviance (expl D2), for each focal size of the human infrastructures variable. Results are from the univariate models with the Common Blackbird (no 96 in the list). The colored lines indicate the maximum value of either the deviance or the correlation which corresponds to the optimal focal size used in further multivariate model.

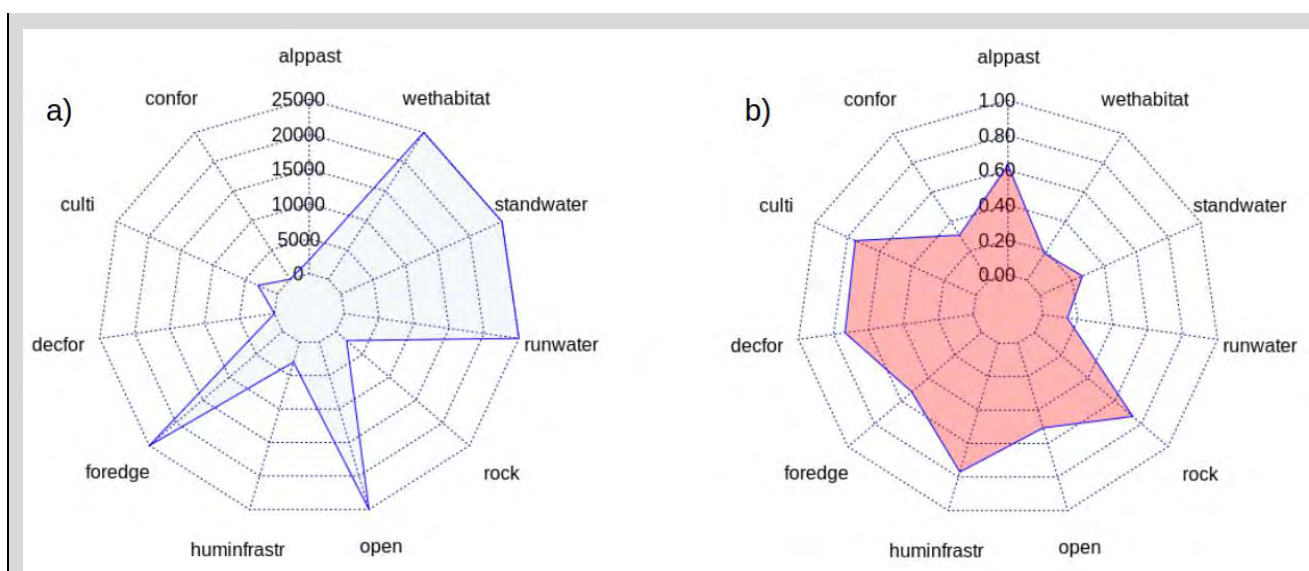


Figure 6 Optimal focal sizes and correlations for the Common Blackbird

a) The optimal radius of each focal variable found with the univariate models and b) the associated Spearman correlation of these models. The blue polygon represents the optimal radius [m, 0 to 25'000] and the red polygon represents the Spearman correlation between the observations and the predictions from these models [0 to 1]. For variable names see Table 1.

Multivariate model – species distribution

Predicted species distributions

To measure the model accuracy, we computed the correlations between the observations and the predictions from the model. The results were good in general, with a median of 0.56, a minimum of 0.11 for Icterine Warbler and a maximum of 0.87 for Common Blackbird (Figure S4 and Table S2).

We observed that the more common a species is, the higher the model accuracy (Figure S5). The correlation between the total number of observations per species (sum of average counts) and model accuracy (the correlation between observation and prediction of the multivariate model) was 0.56.

Using the predictions of the abundance *glmnet* models, we generated species distribution maps. They show the predicted counts for all pixels of Switzerland (e.g. Figure 7) and not the probability of occurrence as usually done in SDMs. Such predictions could not have been generated for some rare species : Cetti's Warbler 14, Hooded Crow x Carrion Crow 38, Tufted duck 55, European Bee-eater 75, Spotted Crake 91, Red-crested Pochard 113, White-backed Woodpecker 120, Tawny Pipit 136 and Ruddy Shelduck 157 (each number indicates the corresponding ID “i” of the species used in the R code and in Tables S1 and S2 found in Appendix). In these cases, since there were not enough observations, the lasso regression could not work. We discuss alternative modeling solutions (e.g. with the Zero-Inflated Poisson Model) in the Discussion.

Box 2 Examples of predicted distributions maps

As an example, we presented the distribution maps for three species, one from each of the specific group (Common Blackbird for “common”, Yellow-billed Chough for “altitude” and Mute Swan for “water”) on Figures 7 to 9.

The Common Blackbird is the third most common species in Switzerland according to our data set, with 30'743 counts in total. It has the highest correlation (0.87) between the observations and the predictions from the multivariate model (Table S2). This bird is widely present in all the country except at high altitude (Figure 7). The maximum value of the abundance color scale indicates in green regions where there are potentially 4 individuals.

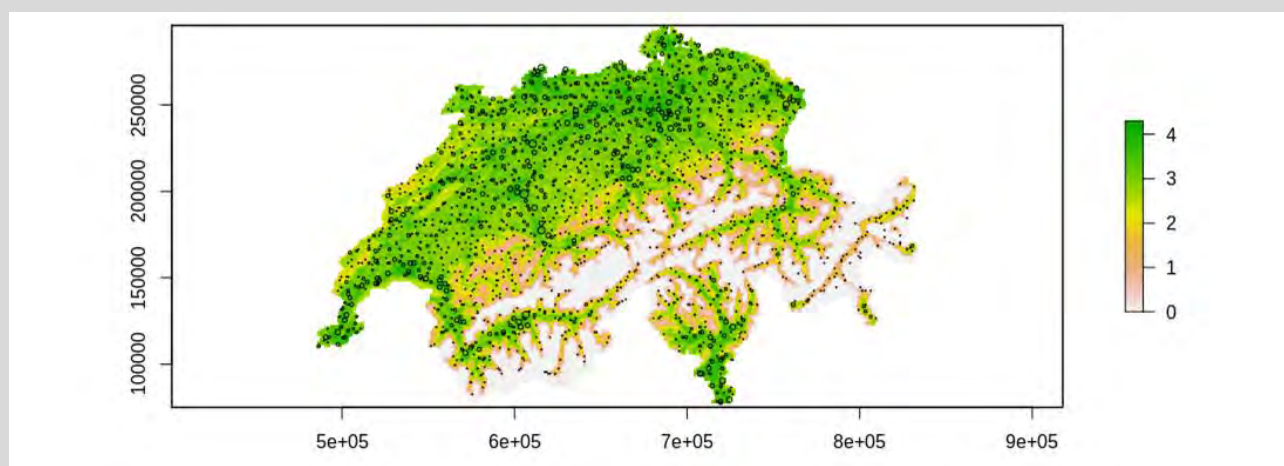


Figure 7 Predicted species distribution of the Common Blackbird from the multivariate model. Predictions are represented on an abundance color scale [number of predicted counts]. Black points represented the real observations and their respective sizes indicated the average counts at each observation site.

The Yellow-billed Chough is less frequent than the Common Blackbird as seen on the abundance scale of the map, with maximum values in green reaching only 1 averaged count (Figure 8). Indeed, in our data set this species has only 978 observations in total. Its predicted distribution is mainly in alpine regions at high altitude. The correlation between the real observations and the predictions from the model is 0.58 (Table S2).

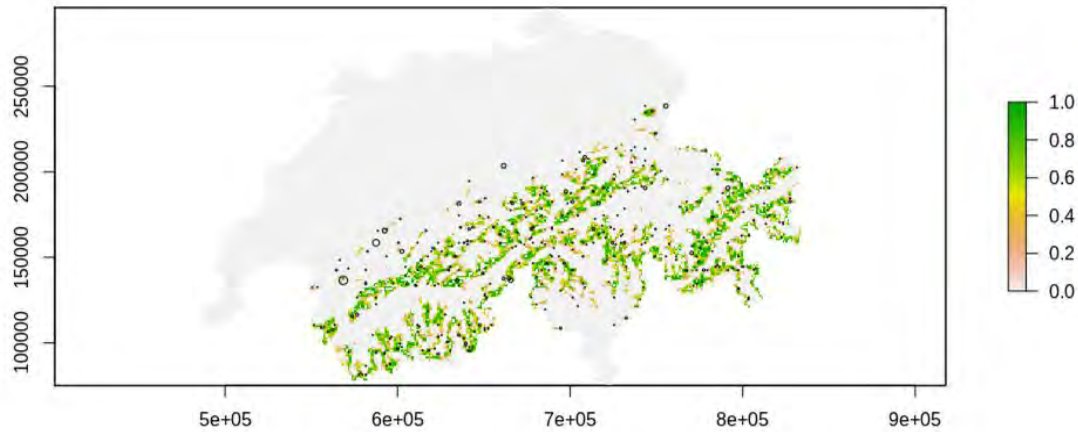


Figure 8 Predicted species distribution of the Yellow-billed Chough from the multivariate model. Predictions are represented on an abundance color scale [number of predicted counts]. Black points represented the real observations and their respective sizes indicated the average counts at each observation site.

The Mute Swan is a common water bird that is found mainly closed to stand water (Figure 9). Indeed, its predicted distribution corresponds well to the localization of the main lakes, with an optimal scale for the stand water variable of 0 m (Figure 3 and Table S1).

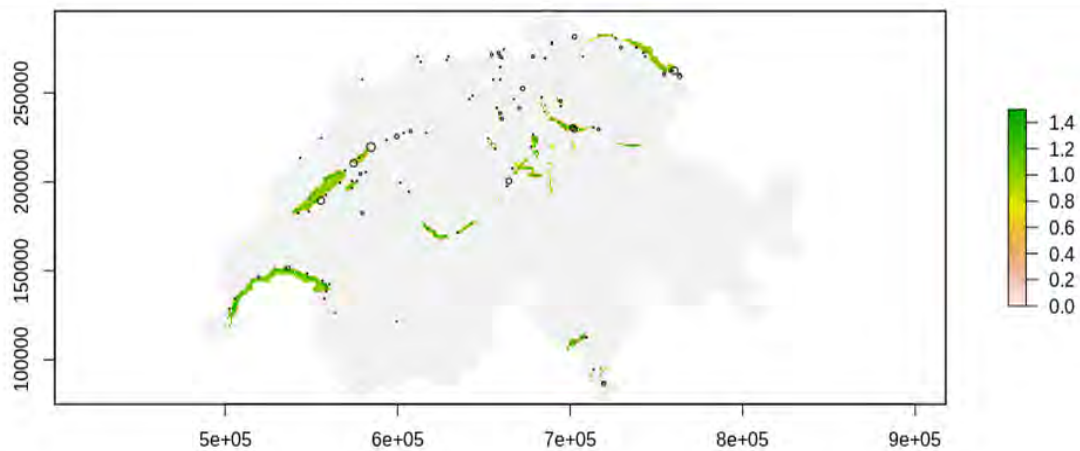


Figure 9 Predicted species distribution of the Mute Swan from the multivariate model. Predictions are represented on an abundance color scale [number of predicted counts]. Black points represented the real observations and their respective sizes indicated the average counts at each observation site.

Variable importance

Our multivariate models were calibrated with a matrix containing all the 17 variables plus their quadratic terms. As this matrix was not scaled, the absolute value of the coefficients extracted from the model cannot be compared between them. This did not affect the predictions and the species distribution maps shown before, but complicated the comparisons between variables.

To be able to compare the variable importance through the coefficients, we looked only at the 11 focal variables, since even without scaling their values are quite similar, being all proportion of land use

(i.e. between 0 and 1). We also conducted this analysis separately with their quadratic terms. However, even with these considerations, variables are not scaled and thus their coefficients are directly influenced by the magnitude of their own values. Indeed, for two variables having the same importance in the model, if one has small values, its coefficient will be higher than for the other variable in order to compensate its contribution to the prediction value. We took this in account in our discussion and are aware that we should be careful in our conclusion.

Moreover, the proportion of land use tends to decrease with the increase in focal size. Therefore, there are 8 different groups of proportion whose combination is specific for each species. Low values of a variable could indicate a large focal size, which will lead to a high coefficient. In the same way, coefficients of quadratic terms are necessary higher than their correspondence in linear terms as their values (between 0 and 1) become smaller with the square operation.

Nevertheless, beside this aspect of scaled data, conducting separate analyses could still give us some useful information about the importance of each variable in the multivariate model and their influence on a particular species. As a consequence, in the following figure we represented only the focal variables (from 1 to 11) and their quadratic terms (from 18 to 29) from the 34 columns of the matrix.

As described before, we looked at which variables influence the predictions and how much. Indeed, for each variable the lasso regression gives a coefficient which represents its importance or weight among all the variables. The predictions obtained by the combination of all the 17 variables are often driven by only one or two main variables that have big coefficients compared to the other. If one variable has a huge coefficient, it does not necessary mean that the correlation (between predictions and observations) is high. It could be just because this variable is more relevant than the other. Thus, this variable is not always a good predictor of the distribution of the species, but still it has the highest coefficient in the model.

From the multivariate models, we extracted for each species the 34 variable coefficients. We compared their absolute values and looked at which variable is the most important per species i.e. which has the highest coefficient among all variables (Figure 10).

Overall, the most relevant variables in the model were wet habitat and human infrastructures. Indeed, the absolute values of their coefficient and their frequency as highest coefficient (i.e. most important variable) showed – both with linear and quadratic terms - that wet habitat is the most important predictor, followed by human infrastructures. On the other side, cultivated areas were never found as the most important variable with linear terms whereas deciduous forest and open areas were in this case with quadratic terms.

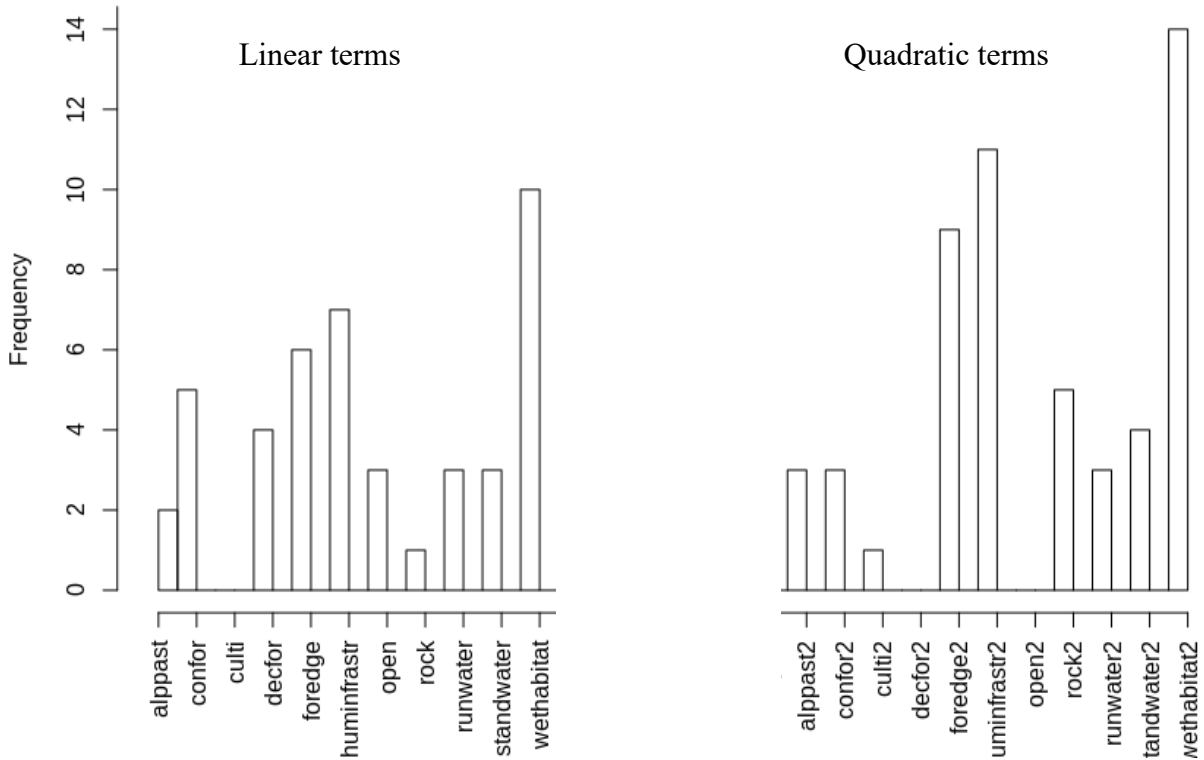


Figure 10 Variable importance

Histogram showing for each variable the frequency when it is the most important variable in the multivariate model. Each bar represents for each variable (left side) and its quadratic term (right side) the number of time when its coefficient has the highest absolute value compared to the other ones. Data are from all the bird species.

Discussion

The scales of influence of the landuse descriptors were very heterogeneous in general, both between variables as well as between species. We observed no general clear trend, which confirms that the scale of influence is species-specific. Within variables, no clear pattern was detected either, since some variables are more important for some species and less for others. Among all variables, coniferous forest has both the smallest focal size and variance, open area the highest focal size and deciduous forest the largest variance.

An advantage of using this “multi-scale approach” compared to classic ones is to improve the predicted distribution (Fournier et al., 2017, Bellamy et al., 2020), as it allows capturing more precisely the influence of the neighborhood on the observations. This local influence on the habitat is species-specific, highlighting the importance of such method for identifying the adequate area needed to protect a particular species (Bellamy et al., 2013). Thus, selecting the appropriate scale for each environmental variable could improve the results of SDMs and therefore should be included in the future in similar studies. This is especially the case when species have different scales of influence for each environmental predictor as for mobile species like birds or bats (Scherrer et al., 2019, Tehrani et al., 2020).

Our findings highlighted the importance of protecting wet habitats since it is a main predictor of bird distribution. Compared to other studies made in Switzerland, our results confirm the importance of wet lands for birds (Maggini et al., 2014). Human infrastructures have also a big impact on birds, having both a large scale of influence and a large coefficient. In both cases, we should not forget that the coefficients from the lasso regression (and so the variable importance) correspond only to the optimal scale found with univariate models. Thus, when comparing the importance of variables, two variables might have the same importance but they might be effective at different distances.

We did not expect such high differences in importance between variables (compared to other studies such as Scherrer et al., 2019). Indeed, the coefficients of wet habitat are far greater than all the other. This discrepancy may be explained by the fact that many birds linked to wet habitats have become rare. For such species with a big imbalance in the number of observation (many absences), the accuracy of the models might be low. Thus, despite a high coefficient, we could not really rely on these predictions. More probably, as discussed before, it is due to the matrix used in the model which was not scaled, so that each variable has different range of values. Nevertheless, we have to be careful not to over interpret the importance of a variable through its coefficient. In this context, we also have to account for the focal size of the variable as it could nuance our interpretation. It remains however that our results clearly show that a predictor can play an important role but only at a small scale, whereas others might have a larger scale of influence but are less important for the predictions.

Limitations and future perspectives

In this study, many focal variables were found to be optimal at the largest radius of 25 km. This was not expected since we had not anticipated that some birds' home ranges would be larger than 25 km. Indeed, for several species (as for the Common Blackbird shown in example), there seems to be a limit as some variables often reached the maximum focal size. Potentially, as we may have underestimated the maximum home range of some birds, we missed here some larger distances that could be optimal for some species. Thus, we should have selected larger values of focal sizes to test if this observed limit is still reached. Another aspect to improve the precision of the selected scale would be to consider more different focal sizes (i.e. more intervals between minimum and maximum expected home range) to obtain a more continuous view of the distance of influence.

More generally, the averaged counts which composed our species data are made at only some sampling sites of the studied area. Thus, based on this unique source of observation, our predicted distributions did not necessary represent the real distributions. Indeed, some species may not have been detected during the monitoring. Moreover, counts resulting from field observations depend on the probability to see the different species and could vary a lot. Therefore, to avoid this problem, further studies could use occupancy models which incorporate the detectability of species (Kéry et al., 2010).

Another issue in the modeling part is due to rare and under-observed species. Too many zeroes (absences) in the average counts is often problematic for the model which is not able to fit correctly the response. (Cunningham & Lindenmayer, 2005; Chiogna & Gaetan, 2007) In our results, 9 species (as listed before) were in this case, having only one or two observations in total. It is quite paradoxical since it is precisely these rare species that are the most in need of distribution maps to protect them.

Indeed, as seen before, common species have better predictions because many observations improve the predictive power of the models.

To study rare species and avoid this issue, one possibility - as done in Scherrer et al. (2019) - is to use Ensemble of Small Models (ESM) (Breiner et al., 2015). For species with very few occurrences, it is an appropriate and robust solution since many small models are put together to obtain final predictions. Nevertheless, this was not possible in our study as we did not have presence-absence observations, and that ESM are not yet implemented for count data.

As a second possibility, we could ideally have implemented a Zero-Inflated Poisson (ZIP) model. Taking into account the high number of zero in the observations, this method is well adapted to rare species (Schaub et al., 2011; Cunningham & Lindenmayer, 2005). However, this was far too complex to do in such a first-step project. A third possibility was to remove these rare species from the data and thus to keep only species with enough presences observations. However, this would have biased the dataset towards common species with potentially different ecological characteristics, ultimately lowering the relevance and generality of our findings.

As a perspective, further studies could test if our findings about the distance of influence are potentially linked with some birds' traits (Storchova & Horak, 2018). Indeed, bird's dietary preferences or morphology may have an impact on the scales of environmental variables. We expect differences between birds groups, which would ideally confirm some hints that we already found in our study between common, water and alpine birds for example.

To conclude, our study improved our knowledge about the scale of influence of environmental variables for birds in Switzerland. Indeed, we have seen that we should adapt the choice of the focal size for each landuse descriptor according to the species. More precisely, we found the optimal scale of each landuse descriptor for all Swiss breeding birds. This could help future work, for example in predicting the effect of land use or climate change on species distribution. Indeed, these perturbations are modifying increasingly the distribution of birds (as shown in the "European Breeding Bird Atlas 2", Keller et al., 2020). Finally, these results could be included as maps into conservation planning (e.g. Ramel, 2018). Hopefully, similar studies with other taxa (e.g. Asian hornet as in Barbet-Massin et al., 2018) or regions (e.g. high latitude as in Niittynen & Luoto, 2018) might also benefit from our modeling framework which is easily adaptable.

Acknowledgements

I would like to thank Olivier Broennimann for his great help and support during this project. I am also thankful to Antoine Guisan and Mathieu Chevalier which make this project possible and provide some data.

References

Barbet-Massin, M., Rome, Q., Villemant, C., & Courchamp, F. (2018). Can species distribution models really predict the expansion of invasive species? *PLOS ONE*, 13(3), e0193085. doi: 10.1371/journal.pone.0193085

- Bellamy, C., Boughey, K., Hawkins, C., Reveley, S., Spake, R., Williams, C., & Altringham, J. (2020). A sequential multi-level framework to improve habitat suitability modelling. *Landscape Ecology*, 35(4), 1001–1020. doi: 10.1007/s10980-020-00987-w
- Bellamy, C., Scott, C., & Altringham, J. (2013). Multiscale, presence-only habitat suitability models: Fine-resolution maps for eight bat species. *Journal of Applied Ecology*, 50(4), 892–901. doi: <https://doi.org/10.1111/1365-2664.12117>
- Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, 6(10), 1210–1218. doi: <https://doi.org/10.1111/2041-210X.12403>
- Chase, J. M., McGill, B. J., McGlinn, D. J., May, F., Blowes, S. A., Xiao, X., ... Gotelli, N. J. (2018). Embracing scale-dependence to achieve a deeper understanding of biodiversity and its change across communities. *Ecology Letters*, 21(11), 1737–1751. doi: <https://doi.org/10.1111/ele.13151>
- Chiogna, M., & Gaetan, C. (2007). Semiparametric zero-inflated Poisson models with application to animal abundance studies. *Environmetrics*, 18(3), 303–314. doi: 10.1002/env.830
- Cunningham, R. B., & Lindenmayer, D. B. (2005). Modeling Count Data of Rare Species: Some Statistical Issues. *Ecology*, 86(5), 1135–1142. doi: 10.1890/04-0589
- Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., ... Snyder, P. K. (2005). Global Consequences of Land Use. *Science*, 309(5734), 570–574. doi: 10.1126/science.1111772
- Fournier, A., Barbet-Massin, M., Rome, Q., & Courchamp, F. (2017). Predicting species distribution combining multi-scale drivers. *Global Ecology and Conservation*, 12, 215–226. doi: 10.1016/j.gecco.2017.11.002
- Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1–22. doi: 10.18637/jss.v033.i01
- Friedman J, Hastie T, Tibshirani R (2019). glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models. R package version 3.0-2, URL <http://CRAN.R-project.org/package=glmnet>
- Fritsch, M., Lischke, H., & Meyer, K. M. (2020). Scaling methods in ecological modelling. *Methods in Ecology and Evolution*, 11(11), 1368–1378. doi: 10.1111/2041-210X.13466
- Graf, R. F., Bollmann, K., Suter, W., & Bugmann, H. (2005). The Importance of Spatial Scale in Habitat Models: Capercaillie in the Swiss Alps. *Landscape Ecology*, 20(6), 703–717. doi: 10.1007/s10980-005-0063-7
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, 8(9), 993–1009. doi: <https://doi.org/10.1111/j.1461-0248.2005.00792.x>

- Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., ... Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16(12), 1424–1435. doi: <https://doi.org/10.1111/ele.12189>
- Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2), 147–186. doi: 10.1016/S0304-3800(00)00354-9
- Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., ... Townshend, J. R. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, 1(2), e1500052. doi: 10.1126/sciadv.1500052
- Haskell, J. P., Ritchie, M. E., & Olff, H. (2002). Fractal geometry predicts varying body size scaling relationships for mammal and bird home ranges. *Nature*, 418(6897), 527–530. doi: 10.1038/nature00840
- Honeck, E., Sanguet, A., Schlaepfer, M. A., Wyler, N., & Lehmann, A. (2020). Methods for identifying green infrastructure. *SN Applied Sciences*, 2(11), 1916. doi: 10.1007/s42452-020-03575-4
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 523–549. doi: 10.1146/annurev-ecolsys-112414-054400
- IPBES, Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services*. Zenodo. doi: 10.5281/zenodo.3553579
- Keller, V., Herrando, S., Voříšek, P., Franch, M., Kipson, M., Milanese, P., Martí, D., Anton, M., Klvaňová, A., Kalyakin, M.V., Bauer, H.-G. & Foppen, R.P.B. (2020). *European Breeding Bird Atlas 2: Distribution, Abundance and Change*. European Bird Census Council & Lynx Edicions, Barcelona.
- Kéry, M., Gardner, B., & Monnerat, C. (2010). Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, 37(10), 1851–1862. doi: <https://doi.org/10.1111/j.1365-2699.2010.02345.x>
- Loss, S. R., Will, T., & Marra, P. P. (2013). The impact of free-ranging domestic cats on wildlife of the United States. *Nature Communications*, 4(1), 1396. doi: 10.1038/ncomms2380
- Maggini, R., Lehmann, A., Zbinden, N., Zimmermann, N. E., Bolliger, J., Schröder, B., ... Jenni, L. (2014). Assessing species vulnerability to climate and land use change: The case of the Swiss breeding birds. *Diversity and Distributions*, 20(6), 708–719. doi: <https://doi.org/10.1111/ddi.12207>
- Maxwell, S. L., Fuller, R. A., Brooks, T. M., & Watson, J. E. M. (2016). Biodiversity: The ravages of guns, nets and bulldozers. *Nature News*, 536(7615), 143. doi: 10.1038/536143a
- McPherson, S. C., Brown, M., & Downs, C. T. (2019). Home Range of a Large Forest Eagle in a Suburban Landscape: Crowned Eagles (*Stephanoaetus coronatus*) in the Durban Metropolitan Open Space System, South Africa. *Journal of Raptor Research*, 53(2), 180–188. doi: 10.3356/JRR-17-83

- Meillère, A., Brischoux, F., & Angelier, F. (2015). Impact of chronic noise exposure on antipredator behavior: An experiment in breeding house sparrows. *Behavioral Ecology*, 26(2), 569–577. doi: 10.1093/beheco/aru232
- Mullarney, K., Parmentier, J. L., Svensson, L., Zetterström, D., Lesaffre, G., & Grant, P. J. (1999). *Le guide ornitho*. Delachaux et Niestlé, Paris.
- Newbold, T., Hudson, L. N., Hill, S. L. L., Contu, S., Lysenko, I., Senior, R. A., ... Purvis, A. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*, 520(7545), 45–50. doi: 10.1038/nature14324
- Niittynen, P., & Luoto, M. (2018). The importance of snow in species distribution models of arctic vegetation. *Ecography*, 41(6), 1024–1037. doi: <https://doi.org/10.1111/ecog.03348>
- Pollock, L. J., O'Connor, L. M. J., Mokany, K., Rosauer, D. F., Talluto, M. V., & Thuiller, W. (2020). Protecting Biodiversity (in All Its Complexity): New Models and Methods. *Trends in Ecology & Evolution*. doi: 10.1016/j.tree.2020.08.015
- R Core Team (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing
- Ramel, C. (2018). Integrating economy with conservation planning and ecosystem services. How to value biodiversity? Master thesis. Master in Behaviour, Evolution, Conservation (BEC), University of Lausanne, Lausanne, Switzerland.
- Schaub, M., Kéry, M., Birrer, S., Rudin, M., & Jenni, L. (2011). Habitat-density associations are not geographically transferable in Swiss farmland birds. *Ecography*, 34(4), 693–704. doi: 10.1111/j.1600-0587.2011.06584.x
- Scherrer, D., Christe, P., & Guisan, A. (2019). Modelling bat distributions and diversity in a mountain landscape using focal predictors in ensemble of small models. *Diversity and Distributions*, 25(5), 770–782. doi: 10.1111/ddi.12893
- Schoener, T. W. (1968). Sizes of Feeding Territories among Birds. *Ecology*, 49(1), 123–141. doi: 10.2307/1933567
- Storchová, L., & Hořák, D. (2018). Life-history characteristics of European birds. *Global Ecology and Biogeography*, 27(4), 400–406. doi: 10.1111/geb.12709
- Tehrani, N. A., Naimi, B., & Jaboyedoff, M. (2020). Toward community predictions: Multi-scale modelling of mountain breeding birds' habitat suitability, landscape preferences, and environmental drivers. *Ecology and Evolution*, 10(12), 5544–5557. doi: 10.1002/ece3.6295
- Vicente, J. R., Gonçalves, J., Honrado, J. P., Randin, C. F., Pottier, J., Broennimann, O., ... Guisan, A. (2014). A framework for assessing the scale of influence of environmental factors on ecological patterns. *Ecological Complexity*, 20, 151–156. doi: 10.1016/j.ecocom.2014.10.005

Appendix

Supplementary figures

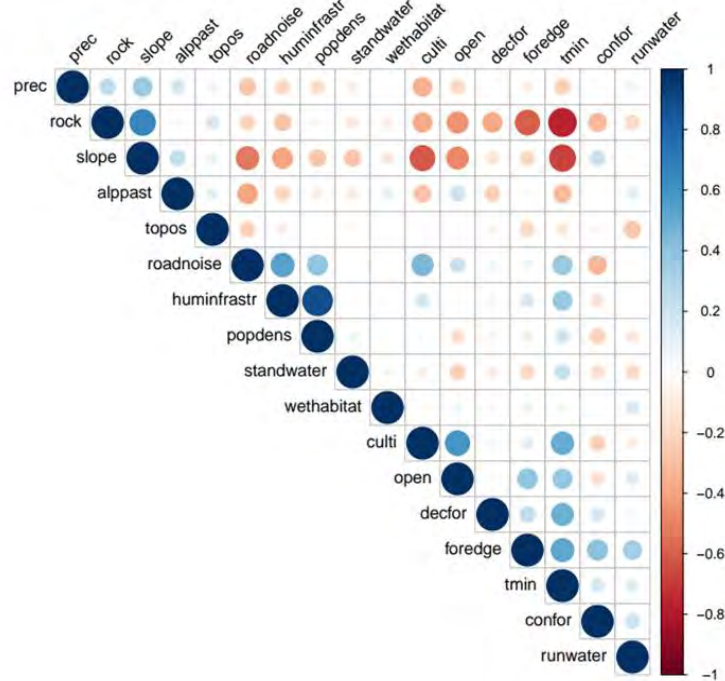


Figure S1 Pairwise correlations plot of the environmental variables used in the models

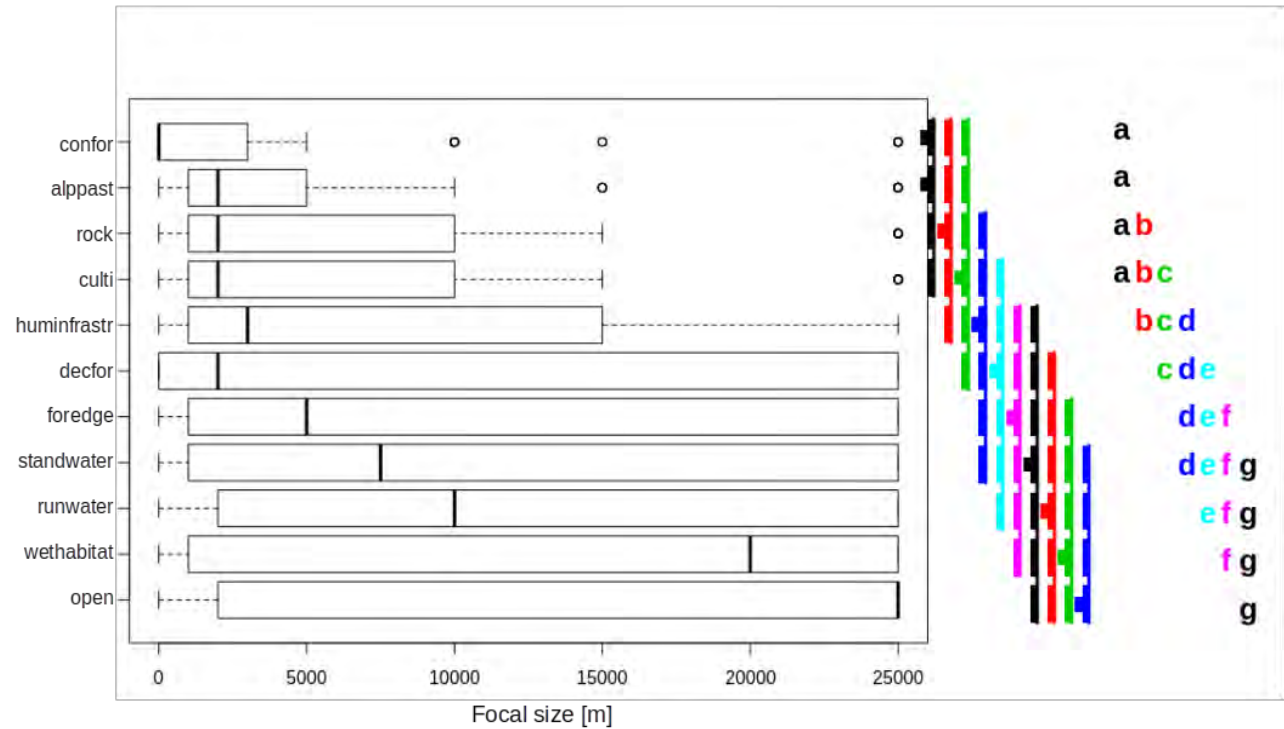


Figure S2 Boxplot of the optimal focal sizes of each focal variable for the 170 species, as found with the univariate models. Results of multiple paired comparisons between variables are shown on the right with both letter-based representation and grouping T. Variables in a same group are not significantly different.

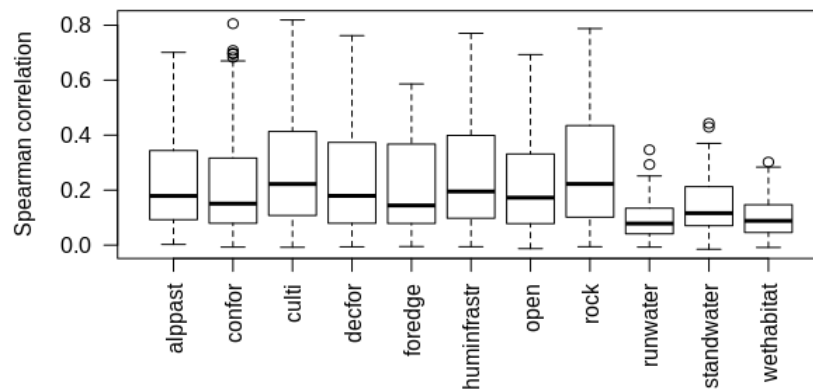


Figure S3 Boxplot of the model accuracy (measured by the Spearman correlation between the observations and the predictions from the univariate model) for each landuse variable, for the 170 species.

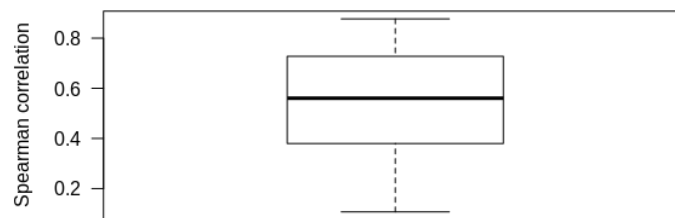


Figure S4 Boxplot of the model accuracy (measured by the Spearman correlation between the observations and the predictions) of the multivariate model, for the 170 species.

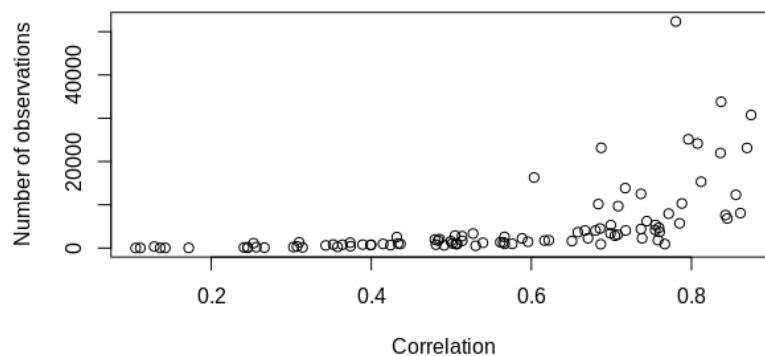


Figure S5 Total number of counts per species, in function of the accuracy of the multivariate models (measured by the correlation between observations and predictions). Each point represents one species.

Tables

These two tables are found in a separate Excel file "Sup_tables.xlsx" on Moodle

Table S1 Optimal focal scales of each variable and for each species, determined with a) the Spearman correlation between observations and predictions from the univariate models and b) the adjusted explained deviance (expl D2) of the same models. (Excel file, sheet 1)

Table S2 Model accuracy of the multivariate model measured by the correlation between the observations and the predictions, listed for the 170 species. (Excel file, sheet 2)

R code

```
##### PART 1 : Creation and selection of environmental variables

circularWindow<- function(radius,resolution){
  npix<-radius/resolution*2+1
  r<-raster(nrows=npix,ncols=npix,xmn=0,xmx=npix*resolution,ymn=0,ymx=npix*resolution)
  values(r)<-c(rep(NA,(npix*npix/2)-0.5),1,rep(NA,(npix*npix/2)-0.5))
  suppressWarnings({crs(r)<-ProjLV03<-CRS('+init=epsg:21781')}) # CH1903 / LV03
  rd<-distance(r)
  mw<-matrix(values(rd),npix,npix)
  mw<-mw<radius
  #image(mw)
  return(mw)
}

#devtools::install_github("CRAN/rgdal")
library(raster)
library(corrplot)

#####
# proportions of landuse/landcover in increasing focal windows
#OFS GeoStats data

lc<-raster("/nas/unilgis/20_Europe/Switzerland/All/Landcover/OFS/Land-
Cover/NOLC_2004/AREA_NOLC04_27_130918_LC09_27.TIF")
lc<-extend(lc, extent(485000,834000,75000,296000))
lu<-raster("/nas/unilgis/20_Europe/Switzerland/All/Land-
cover/OFS/LandUse/NOLU_2004/AREA_NOLU04_46_130918_LU09_46.TIF")
lu<-extend(lu, extent(485000,834000,75000,296000))

radius<-c(1000,2000,3000,5000,10000,15000,25000)
mask<-aggregate(lu,10)

# alpine pasture, OFS NOLU04 242 alpage pâturé
x0<-lu==242
alppast<-x0<-flip(aggregate(flip(x0,1),10),1)
alppast
x0<-lu==242
alppast<-x0<-aggregate(x0,10)
alppast
names(alppast)<-"f0"
for (r in radius){
  xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
  names(xrad)<-paste0("f",r)
  alppast<-stack(alppast,xrad)
}
alppast<-mask(alppast,x0)
save(alppast,file="alppast.rData")
writeRaster(alppast,file="alppast.tif",overwrite=TRUE)

# cultivated areas, OFS NOLU04 201 Arboriculture / 202 Viticulture / 203 Horticulture / 221 Terres arables au
sens large

x0<-lu==201|lu==202|lu==203|lu==221
culti<-x0<-aggregate(x0,10)
names(culti)<-"f0"
for (r in radius){
```

```

xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
names(xrad)<-paste0("f",r)
culti<-stack(culti,xrad)
}
culti<-mask(culti,x0)
save(culti,file="culti.rData")
writeRaster(culti,file="culti.tif",overwrite=TRUE)

# coniferus forest, NFI data

forest<-raster("//nas/ecospat/common/50_data/GeoData_other/EnvData/20_Europe/Switzerland/All/forest/Forest-
MixRateNFI/NFI")
forestcon<-reclassify((100-forest)/100,t(matrix(c(NA,0))))
x0<-resample(forestcon,mask)
confor<-x0<-mask(x0,mask)
names(confor)<- "f0"
for (r in radius){
  xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
  names(xrad)<-paste0("f",r)
  confor<-stack(confor,xrad)
}
confor<-mask(confor,x0)
save(confor,file="confor.rData")
writeRaster(confor,file="confor.tif",overwrite=TRUE)

# deciduous forest, NFI data

forest<-raster("//nas/ecospat/common/50_data/GeoData_other/EnvData/20_Europe/Switzerland/All/forest/Forest-
MixRateNFI/NFI")
forestdec<-reclassify(forest/100,t(matrix(c(NA,0))))
x0<-resample(forestdec,mask)
x0<-mask(x0,mask)
decfor<-x0
names(decfor)<- "f0"
for (r in radius){
  xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
  names(xrad)<-paste0("f",r)
  decfor<-stack(decfor,xrad)
}
decfor<-mask(decfor,x0)
save(decfor,file="decfor.rData")
writeRaster(decfor,file="decfor.tif",overwrite=TRUE)

# running waters, TLM3D

driver<-raster("X:/common/50_data/GeoData_other/EnvData/20_Europe/Switzerland/All/hydrology/distrivers")
runwater<-driver<26
runwater<-x0<-resample(runwater,mask)
names(runwater)<- "f0"
for (r in radius){
  xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
  names(xrad)<-paste0("f",r)
  runwater<-stack(runwater,xrad)
}
runwater<-mask(runwater,x0)
save(runwater,file="runwater.rData")
writeRaster(runwater,file="runwater.tif",overwrite=TRUE)

# standing waters, TLM3D

dlakes<-raster("X:/common/50_data/GeoData_other/EnvData/20_Europe/Switzerland/All/hydrology/distlakes")
standwater<-dlakes<26
standwater<-x0<-resample(standwater,mask)
names(standwater)<- "f0"
for (r in radius){
  xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
  names(xrad)<-paste0("f",r)
  standwater<-stack(standwater,xrad)
}
standwater<-mask(standwater,x0)
save(standwater,file="standwater.rData")
writeRaster(standwater,file="standwater.tif",overwrite=TRUE)

# humid habitats, TLM3D

```

```

dhumid<-raster("X:/common/50_data/GeoData_other/EnvData/20_Europe/Switzerland/All/hydrology/disthumid")
wethabitat<-dhumid<26
wethabitat<-x0<-resample(wethabitat,mask)
names(wethabitat)<-"f0"
for (r in radius){
  xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
  names(xrad)<-paste0("f",r)
  wethabitat<-stack(wethabitat,xrad)
}
wethabitat<-mask(wethabitat,x0)
save(wethabitat,file="wethabitat.rData")
writeRaster(wethabitat,file="wethabitat.tif",overwrite=TRUE)

# bare ground, TLM3D rochers, pierriers

drocks<-raster("X:/common/50_data/GeoData_other/EnvData/20_Europe/Switzerland/All/landcover/TLM3D/distrocks")
rock<-drocks<26
rock<-x0<-resample(rock,mask)
names(rock)<-"f0"
for (r in radius){
  xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
  names(xrad)<-paste0("f",r)
  rock<-stack(rock,xrad)
}
rock<-mask(rock,x0)
save(rock,file="rock.rData")
writeRaster(rock,file="rock.tif",overwrite=TRUE)

# forest edge, TLM3D haies,arbres isolé, lisière de forêts

distedges<-raster("X:/common/50_data/GeoData_other/EnvData/20_Europe/Switzerland/All/forest/DistHedges.tif")
foredge<-reclassify(distedges,matrix(c(0, 24, 0, 24, 49, 1, 49, 13000, 0), ncol=3, byrow=TRUE))
foredge<-x0<-resample(foredge,mask)
names(foredge)<-"f0"
for (r in radius){
  xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
  names(xrad)<-paste0("f",r)
  foredge<-stack(foredge,xrad)
}
foredge<-mask(foredge,x0)
save(foredge,file="foredge.rData")
writeRaster(foredge,file="foredge.tif",overwrite=TRUE)

# habitat & infrastructure, OFS GeoStats NOLC04 11 Surfaces compactées 12 Bâtiments 13 Serres 14 Structures
des cultures en planches 15 Gazon 16 Arbres sur terrains aménagés 17 Petites structures mixtes

x0<-lc>=15&lc<=17
huminfrastr<-x0<-resample(x0,mask)
names(huminfrastr)<-"f0"
for (r in radius){
  xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
  names(xrad)<-paste0("f",r)
  huminfrastr<-stack(huminfrastr,xrad)
}
huminfrastr<-mask(huminfrastr,x0)
save(huminfrastr,file="huminfrastr.rData")
writeRaster(huminfrastr,file="huminfrastr.tif",overwrite=TRUE)

# open, OFS GeoStats NOLC04 21 Végétation herbacée

x0<-lc==21
open<-x0<-resample(x0,mask)
names(open)<-"f0"
for (r in radius){
  xrad<-focal(x0,circularWindow(r,1000),'mean',na.rm=TRUE,pad=TRUE)
  names(xrad)<-paste0("f",r)
  open<-stack(open,xrad)
}
open<-mask(open,x0)
save(open,file="open.rData")
writeRaster(open,file="open.tif",overwrite=TRUE)

#####
#climatic
#Mean of daily minimal temperature between May and August—averaged from 1981 to 2010

```

```

l<-list.files("//nas/CHCLIM25/Tmin/monthly/1981_2010_average",full.names = TRUE)
tmin<-stack()
for (i in 5:8){
  load(l[i])
  names(X)<-paste0("tmin",i)
  tmin<-stack(tmin,X)
}
tmin<-calc(tmin,mean)/10
tmin<-resample(tmin,mask)

# preci Sum of daily precipitation between May and August--averaged from 1981 to 2010

l<-list.files("//nas/CHCLIM25/Prec/monthly/1981_2010_average",full.names = TRUE)
prec<-stack()
for (i in 5:8){
  load(l[i])
  names(X)<-paste0("tmin",i)
  prec<-stack(prec,X)
}
prec<-calc(prec,sum)/10
prec<-resample(prec,mask)

clim<-stack(tmin,prec)
save(clim,file="clim.rData")
writeRaster(clim,file="clim.tif",overwrite=TRUE)

#####
# topographic

# slope inferred from dem at a 25m

slope<-raster("//nas/unilgis/20_Europe/Switzerland/All/Climate/wsl_bioclim/wsl_topo/slp25")
slope<-resample(slope,mask)
slope<-mask(slope,mask)

#topographic position from wsl data at 25m
topos<-raster("//nas/unilgis/20_Europe/Switzerland/All/Climate/wsl_bioclim/wsl_topo/topos")
topos<-resample(topo,mask)
topos<-mask(topo,mask)

topo<-stack(slope,topos)
save(topo,file="topo.rData")
writeRaster(topo,file="topo.tif",overwrite=TRUE)

#####
# Human influence

# Daytime traffic noise

roadnoise<-raster("//nas/ecospat/common/50_data/GeoData_other/EnvData/20_Europe/Switzerland/All/transporta-
tion/BAFU_Daytime_road_traffic_noise/STRASSENLAERM_Tag/STRASSENLAERM_Tag.tif")
roadnoise<-resample(roadnoise,mask)

# population density at 100m

popdens<-raster("//nas/unilgis/20_Europe/Switzerland/All/Human/OFS/ofS_recense-
ment_2013/StatPop13_B13BTOT.tif")
popdens<-resample(popdens,mask)

human<-stack(roadnoise,popdens)
human<-mask(human,mask)
names(human)<-c("roadnoise","popdens")

save(human,file="human.rData")
writeRaster(human,file="human.tif",overwrite=TRUE)
#popdens is a proxi for buildingsED Euclidian distance between the closest building and the centre of the
cell
#roadnoise is a proxi fors roadsED Euclidian distance between the closest road and the centre of the cell
#wate is a proxi for waterED Euclidian distance between the closest water source (stream or lake) and the
centre of the cell
#foreedge is a proxi for forestED Euclidian distance between the closest forest and the centre of the cell
#foreedge is a proxi for canopy Variance of canopy height at a 100 m resolution, calculating from a 1 m reso-
lution raster masked by forest

```

```

#bare is a proxy for ndvi Normalized difference vegetation index at 100 m resolution. Aggregate from 10 m
resolution
#bare is a proxy for ndvi_focal Mean of normalized difference vegetation index at different focal scale

S<-stack(alppast[[1]],culti[[1]],confor[[1]],decfor[[1]],runwater[[1]],standwater[[1]],wethabitat[[1]],
rock[[1]],foredge[[1]],huminfrastr[[1]],open[[1]],clim,topo,human)
names(S)<-c("alppast","culti","confor","decfor","runwater","standwater","wethabitat",
"rock","foredge","huminfrastr","open","tmin","prec","slope","topos","roadnoise","popdens")

cor<-layerStats(S,'pearson',na.rm = T)
corrplot(cor$`pearson correlation coefficient`, type = "upper", order = "hclust", tl.col = "black", tl.srt =
45)

##### PART 2 : Univariate and multivariate models
library(rgdal)
library(raster)
library(sp)
library(corrplot)
library(gtools)
library(modEvA)
library(glmnet)
library(fmsb)
library(multcomp)

`%notin%` <- Negate(`%in%`)

ProjCH1903plus<-CRS('+init=epsg:2056') # CH1903+ / LV95
ProjCH1903<-CRS('+init=epsg:21781') # CH1903 / LV03

##### Raster avec les variables focales

list.files("//nas/ecospat/projects-unil/BIRDS/Aurelien/data/env",pattern = ".rData")

files<-c("alppast.rData","confor.rData","culti.rData","decfor.rData","foredge.rData",
"huminfrastr.rData","open.rData","rock.rData","runwater.rData","standwater.rData",
"wethabitat.rData","clim.rData","human.rData","topo.rData")
files<-paste0("//nas/ecospat/projects-unil/BIRDS/Aurelien/data/env/",files)

for (i in 1: length(files)) load(files[i])

#####

envlist=c("alppast", "confor", "culti", "decfor", "foredge", "huminfrastr",
"open", "rock", "runwater", "standwater", "wethabitat", "tmin",
"prec", "roadnoise", "popdens",
"slope", "topos")

envtot=stack(alppast,confor, culti, decfor, foredge, huminfrastr,
open, rock, runwater, standwater, wethabitat, clim, human, topo)

focalsize=c(0, 1000, 2000, 3000, 5000,10000,15000,25000)

##### liste espèces oiseaux

sp<-read.delim("//nas/ecospat/projects-unil/BIRDS/Aurelien/data/sp/Data_sp_delim.txt",h=TRUE)
sp.list<-unique(sp$species_name)

x<-as.numeric(sp[,5]) # sp[,5] = x_centroid
y<-as.numeric(sp[,6])
xyCH1903plus<-cbind(x,y)
lonlat <-project(xyCH1903plus, ProjCH1903plus@projargs, inv=T)
xyCH1903<-project(lonlat, ProjCH1903@projargs, inv=F)
envCH=extract(envtot, xyCH1903)

xySuisse<-coordinates(envtot)

#remove NAs

ligne2=apply(is.na(envCH[,]),1,sum)==0

spcor=cbind(sp[,1:2], sp[,7:8], xyCH1903)
spcorCH=spcor[ligne2,]
spcorCHpres = spcorCH[spcorCH$Avg_count>0,]

##### Boucle pour les espèces avec i

```

```

maxfocal=data.frame(fsize=0, maxdev=0, env=0, sname=0)
maxfocalcor=data.frame(fsize=0, maxcor=0, env=0, sname=0)

glmmulticor=data.frame(cor=0, sname=0)

varcoef=data.frame(matrix(nrow=170,ncol=34)) # Les coef du cv.glmnet
names(varcoef)<-c(envlist,paste0(envlist,2))
row.names(varcoef)=as.character(sp.list)

##### Pour choisir pour quelle espèce faire des plots :

# P.ex. c(1,4,8) pour choisir la 1e, 4e et 8e espèces de la liste (sur 170)
# if (i %in% spselect) { plot }

# indice i des groupes avec +2 car pour datastar
# (avec max et min donc décale de 2 les lignes)
# Donc pour les vrai i, on enlève 2

spgroup=c(rapaces, eau, common, altitude)-2
spgroup=sort(spgroup)

spselect=100

##### Boucle for i

for (i in 1:length(sp.list)){

for (i in 158:170){

# i=14,38,55,75,91,113,120,136,157 Erreur, boucle for interrompu à cv.glmnet
# + i=27

##### GLM univarié

sp.name<-sp.list[i]
rows<-which(spcorCH$species_name==sp.name)
data<-spcorCH[rows,]
xy=data[,5:6]
rep<-data$Avg_count

devtot = data.frame(f0=0, f1000=0,f2000=0, f3000=0, f5000=0, f10000=0, f15000=0, f25000=0 )
cortot = data.frame(f0=0, f1000=0,f2000=0, f3000=0, f5000=0, f10000=0, f15000=0, f25000=0 )

envmulti<-data.frame(matrix(nrow=nrow(xy),ncol=17))
names(envmulti)<-c("alppast","confor","culti","decfor","foredge","huminfrastr","open","rock","runwater",
"standwater","wethabitat","tmin","prec","roadnoise","popdens","slope","topos")
envmulti[,12:17]<-extract(envtot[[89:94]],xy)

envpredict<-data.frame(matrix(nrow=nrow(xySuisse),ncol=17))
names(envpredict)<-c("alppast","confor","culti","decfor","foredge","huminfrastr","open","rock","runwater",
"standwater","wethabitat","tmin","prec","roadnoise","popdens","slope","topos")
envpredict[,12:17]<-extract(envtot[[89:94]],xySuisse)

##### Boucle pour les variables
for (j in 1:11) {

env<-extract(envtot[[(8*(j-1)+1) : (8*(j-1)+8)]], xy)
envSuisse<-extract(envtot[[(8*(j-1)+1) : (8*(j-1)+8)]], xySuisse)

##### Boucle pour les variables
for (f in 1:8) {

explD2<-c()
glmcor<-c()
for (iter in 1:100){

cal<- sample(1:length(rep),round(length(rep)*0.8,0))
eval<-which(1:length(rep)%notin%cal)
caldata<-as.data.frame(cbind(env[cal,f],I(env[cal,f]^2))); names(caldata)<-c("x","x2")
evaldata<-as.data.frame(cbind(env[eval,f],I(env[eval,f]^2))); names(evaldata)<-c("x","x2")

glm1<-glm(rep[cal]~x+x2,family = "poisson",data=caldata)
pred.eval<-predict.glm(glm1,newdata=evaldata,type="response")
explD2[iter]<-Dsquared(glm1,adjust = T)

```

```

glmcor[iter]=cor(rep[eval], pred.eval ,method="spearman")

devtot[j,f]= mean(explD2)
cortot[j,f]= mean(na.omit(glmcor))

} # fin boucle iter
}# fin boucle f

focalcol<-which.max(devtot[j,])[1]
maxdev<-max(devtot[j,])[1]
focalcolcor<-which.max(cortot[j,])[1]
maxcor<-max(cortot[j,])[1]

maxfocal[((i-1)*11)+j,1]=focalsize[focalcol]
maxfocal[((i-1)*11)+j,2]=maxdev
maxfocal[((i-1)*11)+j,3]=envlist[j]
maxfocal[((i-1)*11)+j,4]=as.character(sp.list[i])

maxfocalcor[((i-1)*11)+j,1]=focalsize[focalcolcor]
maxfocalcor[((i-1)*11)+j,2]=maxcor
maxfocalcor[((i-1)*11)+j,3]=envlist[j]
maxfocalcor[((i-1)*11)+j,4]=as.character(sp.list[i])

# Pour le glm multi, on stocke pour chaque variable la couche focal max
envmulti[,j]=env[, focalcolcor]
envpredict[,j]=envSuisse[, focalcolcor]

##### PLOT

if (i %in% spselect){
  print(i)

  ##### plot explD2 RED

  titrej=paste0(envlist[j], as.character(i))
  pdf(file = titrej)

  plot(focalsize, as.numeric(devtot[j,]),
       type="b", main=c(envlist[j], "-", as.character(sp.list[i])),
       xlab="focal size [m]", ylab="explD2")

  points(as.numeric(maxfocal[((i-1)*11)+j,1]),maxfocal[((i-1)*11)+j,2] ,col="red", pch=19)
  abline(h=max(as.numeric(devtot[j,])), col="red")

  ##### plot corrélation BLUE !

  plot(focalsize, as.numeric(cortot[j,]),
       type="b", main=c(envlist[j], "-", as.character(sp.list[i])),
       xlab="focal size [m]", ylab="cor spearman")

  points(as.numeric(maxfocalcor[((i-1)*11)+j,1]),maxfocalcor[((i-1)*11)+j,2] ,col="blue", pch=19)
  abline(h=max(as.numeric(cortot[j,])), col="blue")

  ##### plot raster de envtot avec couche focal max avec cor

  plot(envtot[[8*(j-1) + which(focalsize==as.numeric(maxfocalcor[((i-1)*11)+j,1])]]],
       main=c(envlist[j], "- f", maxfocalcor[((i-1)*11)+j,1], as.character(sp.list[i]))

  points(spcorCHpres$x[spcorCHpres$species_name==as.character(sp.list[i])],
        spcorCHpres$y[spcorCHpres$species_name==as.character(sp.list[i])],
        col=heat.colors(length(unique(spcorCHpres$Avg_count[spcorCHpres$species_name==as.character(sp.list[i])]))))

  plot(env[,which(focalsize==as.numeric(maxfocalcor[((i-1)*11)+j,1])]],
       data$Avg_count,
       main=c(envlist[j], maxfocalcor[((i-1)*11)+j,1], as.character(sp.list[i])),
       xlab=c("Proportion of ",envlist[j] ), ylab="Average counts")
  # Pour voir s'il y a un optimum biologique selon les valeurs de env
  # avec la couche de focal max
  # p.ex. un % de pâturage avec plus de count

  dev.off() # Pour finir le pdf
} # fin if plot

```

```

cat(j)

} # fin boucle j

##### PLOT
if (i %in% spselect){
  print(i)
  hist(spcorCHpres$Avg_count[spcorCHpres$species_name==as.character(sp.list[i])],
       breaks=length(spcorCHpres$Avg_count[spcorCHpres$species_name==as.character(sp.list[i])]),
       main=c("Repartition of counts - ", as.character(sp.list[i])),
       xlab= "Average counts")
}

##### GLM multivarié

# On fait un glm multi par espèce donc dans la boucle i
# Il contient les 17 variables (11 sélectionnées par la boucle f et j) et
# les 6 autres non focales.
# Avec le lasso, on garde les variables avec coef>0

##### lasso regression cv.glmnet

#dataset quadratic
envmulti2<-as.matrix(cbind(envmulti,envmulti^2))
colnames(envmulti2)<-c(names(envmulti),paste0(names(envmulti),2))

envpredict2<-as.matrix(cbind(envpredict,envpredict^2))
colnames(envpredict2)<-c(names(envpredict),paste0(names(envpredict),2))

#envmulti2<-scale(envmulti2) #scaling

myglm=cv.glmnet(envmulti2,rep,alpha=1, family = "poisson",type.measure="mse")

d2<-myglm$glmnet.fit$dev.ratio[which(myglm$lambda == myglm$lambda.1se)]

coef<-coef(myglm,s="lambda.1se")[-1,]
plot(myglm) #mse
plot(myglm$glmnet.fit,xvar = "lambda") #coef
abline(v=log(myglm$lambda.1se),lty=2)

if (i %in% spgroup) {
  titrecoef=paste0("coefbis", as.character(i))

  png(filename =titrecoef , width = 550, height = 300,
       pointsize=13)
  barplot(coef,las=2,main=paste0("D2 = ",round(d2,2)),
         col=col,cex.axis=1,cex=0.5, cex.names=0.8)
  dev.off()
}

predglmnet<-predict(myglm,newx=envmulti2,type = "response",s="lambda.1se",)

glmmulticor[i,1]=cor(rep, predglmnet, method="spearman")
glmmulticor[i,2]=as.character(sp.list[i])

varcoef[i,]=coef # Pour chaque espèce, on garde les 34 coef des variables
# comme sur le barplot

##### Prédictions pour toute la Suisse

# rep contient les counts aux points xy, et on veut étendre ça à
# tous les points du raster envtot, grâce à predict avec glm lasso.

# sum(!is.na(envpredict[,1]))

if (i %in% spselect){

  bird<-predict(myglm, envpredict2)
  bird[bird<0]<-0

  mapbird=rasterFromXYZ(cbind(xySuisse, bird))
  plot(mapbird)
  summary(bird)
  # plot(mapbird, zlim=c(0,3))
  points(xy[rep>0,],cex=rep[rep>0]/max(rep))
}

```

```

# cex entre 0 et 1, car divisé par max(rep)

  title(sp.name)
}
cat(i)
} # fin boucle i

##### nbreobs et correlation

nbreobs=data.frame(matrix(nrow=170,ncol=1))
row.names(nbreobs)=as.character(sp.list)

for (i in 1:170){

  sp.name<-sp.list[i]
  rows<-which(spcorCHpres$species_name==sp.name)
  obs<-spcorCHpres[rows,"Avg_count"]

  nbreobs[i,1]=sum(obs)
}

corNA=which(!is.na(glmmulticor[,1]))

cor(glmmulticor[corNA,1], nbreobs[corNA,1])
plot(glmmulticor[corNA,1], nbreobs[corNA,1],
      xlab="Correlation", ylab="Number of observations")

which.max(glmmulticor[,1])
# i=96 Merle noir
glmmulticor[96,1]
# 0.8745

# Groupe bird common
# seuil
nbreobs[nbreobs>10000]
common2=sp.list[nbreobs>10000]

##### PLOT

hist(maxfocalcor$fsize[maxfocalcor$env=="alppast"], breaks=20,
      xlab="Focal size [m]", main=paste0("Frequency of optimal focal size for ", envlist[1]))

### BOXPLOT avec maxfocalcor

boxplot(datastar[3:172,], breaks=20, notch=T,boxwex=0.5,
        ylab="Focal size [m]",
        main="Optimal focal size ", las=2)

##### Stat. significatif les différences de boxplot

multcompBoxplot(fsize ~ env, data = maxfocalcor, horizontal=F,
                decreasing = F)

#####
boxplot(datacorstar[3:172,], breaks=20,
        ylab="Spearman correlation",
        las=2)

##### GROUPE d'oiseaux

# Rapaces de jour i+2 (3,6,13,21,44,47,48,49,106,107)

rapaces=2+c(3,6,13,21,44,47,48,49,106,107)
boxplot(datastar[rapaces,],
        ylab="Focal size [m]",
        main="Optimal focal size - Rapaces ", las=2)
boxplot(datacorstar[rapaces,],
        ylab="Spearman correlation",
        las=2)
# Oiseaux d'eau
eau=2+c(23,41,54,56,65,67,68,76)

boxplot(datastar[eau,],
        ylab="Focal size [m]",
        main="Optimal focal size - Water birds ", las=2)

```

```

boxplot(datacorstar[eau,],
        ylab="Spearman correlation",
        las=2)
# Common birds
# Groupe bird common
# seuil
nbreobs[nbreobs>10000]
common2=sp.list[nbreobs>20000]

common=2+c(50,96,101,104,109,133,148,167)

boxplot(datastar[common,],
        ylab="Focal size [m]",
        main="Optimal focal size - Common birds ", las=2)
boxplot(datacorstar[common,],
        ylab="Spearman correlation",
        las=2)
# Altitude bird
altitude=2+c(1,7,29,40,74,86,114,161)

boxplot(datastar[altitude,],
        ylab="Focal size [m]",
        main="Optimal focal size - Altitudinal birds ", las=2)
boxplot(datacorstar[altitude,],
        ylab="Spearman correlation",
        las=2)

##### BOXPLOT varcoef

boxplot(varcoef[,c(1:11)], las=2, outline=T, notch=T, boxwex=0.5,
        ylab="Variable coefficient")

# Avec les 4 groupes :

boxplot(varcoef[(rapaces-2),1:34], las=2, outline=F,
        main = "Rapaces", ylab="Variable coefficient")
boxplot(varcoef[(eau-2),1:34], las=2, outline=F,
        main="eau", ylab="Variable coefficient")
boxplot(varcoef[(common-2),1:34], las=2, outline=F,
        main="common", ylab="Variable coefficient")
boxplot(varcoef[(altitude-2),1:34], las=2, outline=F,
        main="altitude", ylab="Variable coefficient")
boxplot(varcoef[1,1:34], las=2) # C'est la même chose que coef1
boxplot(varcoef[,1:17], las=2, outline=T)
sum(abs(varcoef$alppast), na.rm = T)
sum(abs(varcoef$tmn2), na.rm = T)
sum(abs(varcoef$wethabitat2), na.rm = T)

##### varcoefsum avec barplot

varcoefsum=data.frame(matrix(nrow=1,ncol=34))
names(varcoefsum)<-c(envlist,paste0(envlist,2))

for (i in 1:34){
  varcoefsum[1,i]=sum(abs(varcoef[,i]), na.rm = T)
}
barplot(as.numeric(varcoefsum), names.arg=names(varcoefsum) ,
        ylab="Variable coefficient", las=2)
barplot(as.numeric(varcoefsum[1,c(1:27,29:34)]), names.arg=names(varcoefsum[1,c(1:27,29:34)]),
        ylab="Variable coefficient", las=2)
barplot(as.numeric(datastar[98,]), names.arg=names(varcoefsum[1:11]),
        las=2)
barplot(as.numeric(datacorstar[98,]), names.arg=names(varcoefsum[1:11]),
        las=2)
summary(abs(varcoef$wethabitat2), na.rm = T)
summary(varcoef$wethabitat2, na.rm = T)

##### Varcoefmax avec hist

varcoefmax=data.frame(matrix(nrow=170,ncol=1))
row.names(varcoefmax)=as.character(sp.list)

# if avec deux tests logiques séparés par un "OU"
# pour enlever les NA des espèces i=14, ...
# et aussi quand il n'y a que des 0 pour tous les coef (ça arrive souvent)

```

```

for (j in 1:170){
  if (is.na(varcoef[j,1]) == T | sum(varcoef[j,]) == 0) {
    varcoefmax[j,1]=NA
  }
  else {
    varcoefmax[j,1]=which.max(na.omit(abs(varcoef[j,])))
  }
}
hist(as.numeric(na.omit(varcoefmax[,1])), breaks=50,
      axes=F, xlab=NULL , main=NULL, labels=F)
axis(1, labels =names(varcoefsum[1,1:28]), las=2, at=seq(1,28,1) )
sort(unique(as.numeric(na.omit(varcoefmax[,1]))))
axis(2, at=seq(0,20,2) )
apply(envmulti2, 2, max)
boxplot(envmulti2)

##### BOXPLOT deviance avec maxfocal

boxplot(maxfocal[,1] ~ maxfocal$env, xlab = NULL, notch=T, boxwex=0.5,
        ylab="Focal size [m]",
        main="Optimal focal size with deviance ", las=2)
boxplot(maxfocal[1:110,1] ~ maxfocal$spname[1:110], xlab = NULL,
        ylab="Focal size [m]",
        main="Optimal focal size with deviance ", las=2)
boxplot(maxfocal[,2] ~ maxfocal$env, xlab = NULL,
        ylab="Deviance explD2",
        main="Deviance ", las=2)

# GLM glmmulticor
summary(glmmulticor)
boxplot(glmmulticor[,1], ylab="Spearman correlation", las=2)

##### PLOT Spider/Radar des 11 variables focales

# Pour utiliser stars, il faut modifier le format des données (cf. ex USJudge)
# les données proviennent de maxfocalcor. On change pour que les lignes
# deviennent les colonnes

# Création d'un nouveau dataframe pour pouvoir utiliser stars et radarchart

# DATASTAR focal size
# 172 lignes car 1ere et 2e lignes pour max et min (25000 et 0)
datastar=data.frame(matrix(nrow=172, ncol=11))
names(datastar)<-envlist[1:11]
row.names(datastar)=c("max", "min", as.character(sp.list))

# !!! Pour accéder aux espèces, mettre i+2, à cause des 2 premières lignes
# qui décalent tout

# DATACORSTAR max correlation
# 172 lignes car 1ere et 2e lignes pour max et min (1 et 0)
datacorstar=data.frame(matrix(nrow=172, ncol=11))
names(datacorstar)<-envlist[1:11]
row.names(datacorstar)=c("max", "min", as.character(sp.list))

# DATASTAR focal size
datastar[1,]=rep.int(25000,11)
datastar[2,]=rep.int(0,11)

for (i in 1:170){
  datastar[i+2,]=as.numeric(maxfocalcor$fsize[((i-1)*11)+1:(i*11)])

  # datastar[1,]=as.numeric(maxfocalcor$fsize[1:11])
  # datastar[2,]=as.numeric(maxfocalcor$fsize[12:22])
  # datastar[3,]=as.numeric(maxfocalcor$fsize[23:33])
  # datastar[4,]=as.numeric(maxfocalcor$fsize[34:44])
}

# DATACORSTAR max correlation

datacorstar[1,]=rep.int(1,11)
datacorstar[2,]=rep.int(0,11)

for (i in 1:170){

```

```

    datacorstar[i+2,]=as.numeric(maxfocalcor$maxcor[((i-1)*11)+1:(i*11)])
  }

# Avec fmsb package : radarchart
# DATASTAR focal size

radarchart(datastar[1:4,], maxmin = F, axistype = 1, seg = 7,
            caxislabels = as.character(focalsize), axislabcol = "black",
            pdensity = 70)

# Set graphic colors
library(RColorBrewer)
coul <- brewer.pal(3, "BuPu")
colors_border <- coul[2]
library(scales)
colors_in <- alpha(coul[2],0.1)
colors_in2 <- alpha("red",0.1)

# na.omit(datastar[1:170,]) à cause de ligne 91 et 157
radarchart(na.omit(datastar[1:172,]), maxmin = T, axistype = 1, seg = 5,
            caxislabels = c("0", "5000", "10000", "15000", "20000", "25000"),
            axislabcol = "black",
            pfc=colors_in, pcol = colors_border, pty = 32)

# DATACORSTAR max correlation

radarchart(na.omit(datacorstar[1:172,]), maxmin = T, axistype = 4, seg = 5,
            axislabcol = "black",
            pfc=colors_in2, pcol = colors_border, pty = 32)

# Merle i=96
radarchart(na.omit(datacorstar[c(1,2,98),]), maxmin = T, axistype = 4, seg = 5,
            axislabcol = "black",
            pfc=colors_in2, pcol = "blue", pty = 32)

radarchart(na.omit(datastar[c(1,2,98),]), maxmin = T, axistype = 1, seg = 5,
            caxislabels = c("0", "5000", "10000", "15000", "20000", "25000"),
            axislabcol = "black",
            pfc=colors_in, pcol = "blue", pty = 32)

# i=71 et 133
# Le + petit et le + grand range size
radarchart(datastar[c(1,2,71,133),], maxmin = T, axistype = 1, seg = 5,
            caxislabels = c("0", "5000", "10000", "15000", "20000", "25000"),
            axislabcol = "black",
            pfc=c(coul[2], colors_in2), pcol = colors_border, pty = 32)

##### Par groupe
radarchart(datastar[c(1,2,rapaces),], maxmin = T, axistype = 1, seg = 5,
            caxislabels = c("0", "5000", "10000", "15000", "20000", "25000"),
            axislabcol = "black",
            pfc=color.in , pcol = colors_border, pty = 32)
radarchart(datastar[c(1,2,eau ),], maxmin = T, axistype = 1, seg = 5,
            caxislabels = c("0", "5000", "10000", "15000", "20000", "25000"),
            axislabcol = "black",
            pfc=color.in , pcol = colors_border, pty = 32)
radarchart(datastar[c(1,2,common),], maxmin = T, axistype = 1, seg = 5,
            caxislabels = c("0", "5000", "10000", "15000", "20000", "25000"),
            axislabcol = "black",
            pfc=color.in , pcol = colors_border, pty = 32)
radarchart(datastar[c(1,2,altitude),], maxmin = T, axistype = 1, seg = 5,
            caxislabels = c("0", "5000", "10000", "15000", "20000", "25000"),
            axislabcol = "black",
            pfc=color.in , pcol = colors_border, pty = 32)

##### Corrélation entre les focal size trouvées avec explD2 ou avec cor

ensemble=data.frame(fsize_explD2=maxfocal[,1], fsize_cor=maxfocalcor[,1],
                    env=maxfocal[,3], spname=maxfocal[,4])

cor(maxfocal[,1],maxfocalcor[,1], method = "spearman")
cor(na.omit(maxfocal[,1]), na.omit(maxfocalcor[,1]), method = "spearman")
cor(na.omit(maxfocal[,2]), na.omit(maxfocalcor[,2]))

##### Pour exporter le data.frame des résultats dans un autre format

```

```
write.csv(maxfocal, file = "maxfocal.csv")

listei=data.frame(ID=0)
for (i in 1:170){
  listei[(11*(i-1)+1) : (11*(i-1)+11), 1]=rep(i, 11)
}
write.csv(listei, file = "listei.csv")
```