Validation Bioscore 2.0 distribution maps for breeding birds

Christian Kampichler & Henk Sierdsema

Sovon-rapport 2016/06

Validation Bioscore 2.0 distribution maps for breeding birds

Christian Kampichler & Henk Sierdsema

This report was commisioned by Planbureau voor de Leefomgeving

Planbureau voor de Leefomgeving

Colophon

© Sovon Dutch Centre for Field Ornithology 2016

This report was commisioned by Planbureau voor de Leefomgeving

Recommended citation: Kampichler C & Sierdsema H. 2016. Validation Bioscore 2.0 distribution maps for breeding birds. Sovon-report 2016/06. Sovon Vogelonderzoek Nederland/Sovon Dutch Centre for Field Ornithology, Nijmegen.

Illustratie omslag: Henk Sierdsema (Rila National Park, Bulgaria), Martin Mollet - Saxifraga (Nuthatch)

Opmaak: John van Betteray

ISSN-nummer: 2212 5027

Sovon Vogelonderzoek Nederland/Sovon Dutch Centre for Field Ornithology Toernooiveld 1 6525 ED Nijmegen *e-mail:* info@sovon.nl *website:* www.sovon.nl

Nothing of this report may be multiplied or published by means of print, photocopy, microfilm or any other means without written consent by Sovon and/or the commissioning party.

Inhoud

1. Introduction

The aim of the Bioscore 2.0 project is to design a tool for assessing the impacts of European Community policies on biodiversity in Europe (van Hinsberg *et al.* 2014).

In the course of developing BioScore 2.0 the Planbureau voor de Leefomgeving (PBL) asked for a response to the following tasks and topics:

- 1. Determine a full multivariate model (FMM) version for all BioScore 2.0 species, that is, with all pressure factors in one single multivariate analysis. Make a validation of the FMM's, preferably on independent data if available or by a crossvalidation. Provide a possibility to examine the model for non-linearities and interactions in the relationship between predictors and predictions.
- 2. Describe provenance and quality of the distribution data that are used for model calibration and validation.
- 3. Compare the results of BioScore 2.0 of the current

situation with the FMM of the current situation.

- 4. Compare the univariate dose-response relations that were used for BioScore 2.0 with the respective partial dependence plots of the FMM for ten sample species.
- 5. For each of the four ecosystem types (urban, forest, other natural and agrarian), specify which of the BioScore 2.0 species are typical and characteristic for them (may be more than one type).
- 6. Complement the earlier delivered univariate doseresponse relationships for BioScore 2.0 for ten sample species. Display the quadratic as well as the linear relationshipes, analyse and describe the differences.
- 7. Determine the effect of cut-off value choice for ten sample species. Describe the effects of the different choices.

In this report we pursue these questions and present the results of the analyses.

2. Methods

2.1. Full multivariate models for all BioScore 2.0 species

2.1.1. Modelling

For 299 species (for an overview of all selected species see appendix 2 in Sierdsema 2014) a full multivariate model (FFM) was generated. The bird data are the same that were used for the BioScore 2.0 models (van Hinsberg *et al.* 2014) and were collated from eBird, Observado, waarneming.nl, the European commission and Bulgarian bird counts. The characteristics of these data sources as well as the processing of the data were described in the technical report on breeding birds in BioScore 2.0 (Sierdsema 2014). Sierdsema (2015) showed that including GBIF data to the FMMs can lead to a considerable overestimation of the the presence of species in certain parts of Europe. This is due to a heavy geographical bias in this data set. We thus omitted GBIF data from the analysis (as was done for the BioScore 2.0 models). No zeroes were generated additionally to the observed absences in the data.

The explanatory variables used in the FMMs were the same as described in the additional questions technical report (Sierdsema 2015).

The models were using a boosted regression tree approach (Friedman 2001) implemented in the custom-tailed R package TRIMmaps (Kampichler *et al.* 2015). This methodological approach was described in detail in the technical report on breeding birds in BioScore 2.0 (Sierdsema 2014). Modelling with boosted regression trees requires specification of a series of parameters. Species-specific optimisation of these parameters in order to get the optimal model for each single species was not feasible given the large amount of species that had to be modelled and the number of parameters that can be tuned. We thus ran all models with a uniform set of parameter choices, namely learning.rate = 0.01, step. $size = 2$ and tree.complexity $= 2$). This combination of parameters has proven to be robust in the case of automatized analysis of large numbers of species. Other parameter specifications might result in more accurate models but need higher computation times and are prone to failure of the modelling algorithm¹. Even though the parameters were chosen to ease the automatized modelling process, for some species no full multivariate models could be produced, not even when the parameters were modified. Although the

full multivariate models do not produce the best possible predictions for each species, we regard them to deliver reasonable benchmarks for the evaluation of the BioScore 2.0 maps.

2.1.2. Validation

The specific algorithm by Elith *et al.* (2008) that was used allows for a crossvalidation during the modelling process. This is particularly useful when a large number of species has to be modelled and saves a considerable amount of time compared to a post-hoc model evaluation. We applied a ten-fold crossvalidation and calculated the area under the receiveroperator curve (AUC) for the respective ten set-aside datasets that were not used for model training in the crossvalidation iterations. The mean of the ten AUC values and the deviance explained by the full model were used to evaluate the quality of each FMM.

2.1.3. Provenance and quality of the distribution data

An extended description of the bird data sources can be found in Sierdsema 2014 and Sierdsema 2015. The most important data sources used for the spatial modelling of the FMM's were eBird checklists, 10x10 km N2000 Article 12-reporting data sets for a number of countries and observado.org (currently called observation.org). The GBIF-dataset is heavily biased towards the Fennoscandinavian countries and had a profound influence on the distribution maps of European species that are also common in the north of Europe. The comparison of models with and without GBIF-data showed that the GBIF-data had limited influence on the distribution outside Fennoscandinavia and were therefore omitted from the dataset used to create the FMM-models. The general experience was that it is better to use a limited dataset with zero-observations than a large dataset with few zero-observations.

2.2. Comparison between BioScore 2.0 and full multivariate models

The predictions of BioScore 2.0 models and FMMs cannot be compared directly due to the fact that they are generated in completely different ways:

• The output of a BioScore 2.0 model is the combined result of a distribution map based on a soilclimate model, an overlay with suitable land cover

¹ A typical error message is "Error in .gbm.step : restart model with a smaller learning rate or smaller step size".

classes in the Corine Land Cover map and the stacked predictions of presence probability based on a number of univariate logistic models, each one for a determined environmental pressure variable (Sierdsema 2014). The combined score of a BioScore 2.0 is scaled between 0 and 10,000.

• An FMM processes all explanatory variables (soil, climate, land use, pressure factors) simultaneously and predicts the probability of presence in the range of (0, 1).

To tackle this problem we binarised both FMM probabilities and BioScore 2.0 score, that is, we transformed them to predictions of presence and absence.

2.2.1 Production and evaluation of binary maps Probability maps—that is, each cell in the grid contains a value in the interval (0, 1)—can easily made binary (only absences and presences shown instead of probabilities of presences) by applying a cut-off and setting all values below the cutoff to zero (absence) and all values larger than the cut-off to one (presence). Among all approaches to find an optimal cut-off, the true skill statistic (TSS) has been shown to be one of the best for ecological data (Allouche *et al.* 2006). It is frequently used in ecological studies (e.g. Zimmermann *et al.* 2009, Barbet-Massin *et al.* 2012) and implemented in ecological distribution modelling software, for example BIOMOD (Thuiller *et al.* 2009). The TSS compares the number of correct predictions minus those attributable to random guessing to that of a hypothetical set of perfect forecasts. It is defined as

 $(ad - bc)/(a + c)(b + d)$ = sensitivity + specificity - 1 where

- a = correctly predicted presences,
- b = incorrectly predicted presences,
- c = incorrectly predicted absences,
- d = correctly predicted absences,
- sensitivity = proportion of correctly predicted presences a $/$ (a + c), and
- specificity = proportion of correctly predicted absences $d / (b + d)$ (Table 2.1).

Table 2.1. Matrix of observed and predicted values for calculating the true skill statistic.

For finding the optimal cut-off value, the TSS must be maximised. This method was already applied earlier to regression models within the BioScore 2.0 project (Sierdsema 2014). At the same time, the maximised TSS serves a statistic for the quality of the binary maps. It ranges from −1 to +1, where +1 indicates perfect agreement between observation and prediction and 0 indicates random model performance. Values close to −1 indicate "perverse" modelling performance; by exchanging presences and absences the maps can be converted into maps of high quality.

For both the FMM and the BioScore 2.0 predictions the observations used for the FMMs were used for maximising the TSS. We thus a priori expect the TSS of FMM maps to be higher than that of the BioScore 2.0 maps. This is justified since the FMM predictions serve as benchmark for the BioScore 2.0 models and we evaluate the BioScore 2.0 binary maps by observing how close their TSS comes to the TSS of the FMM maps.

2.2.2. Estimation of correspondence between binary maps

The difference between binary maps based on FMMs and BioScore 2.0 models is visualised as a difference map. For each 5x5 km cell we subtract the presence value (0 for absence, 1 for presence) of the FMM map from the the presence value of the BioScore 2.0 map. Cells in the difference map with value 1 thus indicate that the BioScore 2.0 model predicted presence, whereas the FMM predicted absence; cells with value -1 indicte the BioScore 2.0 predicted absence, whereas the FMM predicted presence.

The correspondence between two binary predictions can be done by analysing their confusion matrix (Table 2.2) where the predictions of one map are in the columns, those of the other in the rows (Table 2.2). The positions in the matrix count the cells where the models agree and disagree, respectively:

- a = both models predict presence
- b = BioScore 2.0 predicts presence, FMM predicts absence
- c = BioScore 2.0 predicts absence, FMM predicts presence
- d = both models predict absence

Table 2.2. Sample confusion matrix for analysing the predictions of FMM and BioScore 2.0 models.

A simple measure of correspondence is the proportion of cells where both models agree. It is called overall accuracy and is calculated as $(a + d) / (a + b + c + d)$ d). It must taken into account, however, that overall accuracy is dependent on prevalence. High accuracies will thus be observed for rare species with large shares of absences as well as for wide-spread species with large shares of presences. If, for example, FMM predicts presence in 10 % of the cells and BioScore 2.0 predicts presence in 15 %, then the expected proportion of cells with agreement is 0.78 ($0.1 * 0.15 +$ 0.9 * 0.85) even when the predictions are randomly distributed (Reineking & Schröder 2004). Other measures that take the success as a result of random guessing into account (for example, the TSS) are not symmetric and depend on which model is represented in the column (usually the true values, see Table 2.2) and which in the rows (usually the estimated values) of the confusion matrix. Since here the FMM maps are used as a benchmark, their predictions are shown in the columns, BioScore 2.0 in the rows, and we calculated the corresponding TSS. Both overall accuracy and TSS are reported.

2.2.3. Final evaluation and expert judgement

If BioScore 2.0 can model a species satisfactorily was evaluated by combining information on model quality of FMMs and BioScore 2.0 models (as measured by their TSS, in the following called TSS.FMM and TSS.BioScore, respectively) and the concordance of FMM and BioScore 2.0 maps (also measured by the TSS. in the following called TSS.map_correspondence) according to the following reasoning:

- 1) If FMM quality is high and BioScore 2.0 model quality comes close to that of the FMM and if at the same time map correspondence is also high, then BioScore 2.0 is suitable for modelling the respective species.
- 2) There can, however, arise the situation that FMM quality is only intermediate, BioScore 2.0 model quality comes close to that of the FMM (and is thus at best intermediate as well) and map correspondence is good. In this unclear case the maps should be subjected to the visual evaluation by an expert.
- 3) When a species cannot be modelled satisfactorily by the FMM, it is also not suitable for modelling by BioScore 2.0. This is the logical consequence of the fact that all TSS.BioScore values were smaller than their TSS.FMM counterpart (see Results section).
- 4) In the cases when BioScore 2.0 quality did not come close to that of the FMM or map correspondence was low, again an expert should judge the maps by eye.

Using threshold values 0.75 and 0.5 for the separation between good/intermediate and intermediate/ poor TSS.FMM and TSS.map_correspondence values, respectively, and a multiplication factor of 0.8 for TSS.FMM to define a TSS.BioScore that is "close to" the TSS.FMM yields the following rules:

1)

- IF TSS.FMM ≥ 0.75 (= FMM quality is good)
- AND TSS.BioScore ≥ 0.8 * TSS.FMM (= BioScore comes close to FMM)
- AND TSS.map_correspondence ≥ 0.75 (= correspondence is high)
- THEN Species is well modelled by BioScore 2.0

2)

- IF 0.5 <= TSS.FMM < 0.75 (= *FMM quality is intermediate*)
- AND TSS.BioScore ≥ 0.8 * TSS.FMM (= *BioScore comes close to FMM*)
- AND TSS.map_correspondence ≥ 0.75 (= *correspondence is high*)
- THEN Situation unclear, an expert should visually check and evaluate the maps

3)

IF TSS.FMM < 0.5 (= *FMM quality is low*) THEN Species cannot be modelled

4)

- IF 0.5 <= TSS.BioScore < 0.8 * TSS.FMM (= *BioScore does not come close to FMM*)
- OR TSS.map_correspondence < 0.5 (= *correspondence is low*)
- THEN Doubtful if the species should be used for BioScore 2.0, but an expert should visually check and evaluate the maps

Next to formal interpretation of the results, all distribution maps were also evaluated by a species expert (Ruud Foppen). The combined results of the two evaluation approaches can be found in Appendix 7.

2.3. Comparison between dose-response relations and partial dependence plots

Multivariate models are not as transparent as univariate models and visualisation is difficult due to the high dimensionality of the relationships between explanatory and response variables. Partial dependence plots can give an impression of how a given explanatory variable is related to the response by varying its value across its entire range and observing the resulting value of the response variable while keeping all other explanatory variables constant at their mean values. Interactions can only be shown in

three-dimensional plots, but even in this case only two explanatory variables can be displayed at a time. Direct comparisons of the effects of single explanatory variables between FMMs and BioScore 2.0 models must thus be viewed with caution and must not be over-interpreted.

The dose-response curves come from univariate models and relate the respective explanatory variable to the probability of presence between 0 (certain absence) and 1 (certain presence). In contrast, the partial dependence plots stem from multivariate models that simultaneously take all explanatory variables into account. When all other variables are held constant at their mean values and the variable under scrutiny is varied over its range, the corresponding probabilities of presence are within a much smaller range. The less important a variable in the multivariate model is, the smaller gets this range, ending up in a horizontal line at a single value for probability of presence when the variable has no effect at all. Only when a single variable would be important in a multivariate model, the probability of presence in its partial dependence plot could range from 0 to 1. To be able to compare the shape of the curves, the y-axis of the partial dependence plots thus are not limited by 0 and 1 like in the dose-response curves but they zoom to the range of variation of the probability of presence.

Only explanatory variables with an AUC of at least 0.6 in the univariate BioScore 2.0 models were considered for comparison.

The ten following species were chosen to make the comparison:

- Great reed warbler *Acrocephalus arundinaceus*
- Aquatic warbler *Acrocephalus paludicola*
- Eurasian skylark *Alauda arvensis* (common, not very sensitive)
- European roller *Coracias garrulus* (rare, very sensitive)
- White-backed woodpecker *Dendrocopos leucotos*
- Woodlark *Lullula arborea*
- Corn bunting *Emberiza calandra*
- Red-backed shrike *Lanius collurio* (common, sensitive)
- Lesser grey shrike *Lanius minor*
- Grey partridge *Perdix perdix* (fairly common, not very sensitive)

2.4. Specification of typical and characteristic BioScore 2.0 species

For the Bioscore 2.0 output species are combined that occur in the same major habitat types. For this purpose species have to be assigned to one the four following habitat classes: Urban, Farmland, Forest and Nature. For some species, like the Coal tit, this classification is straightforward as is occurs almost only in one of those types. But for others this is much more complicated when they occur regularly in two or more types. In those cases we selected the habitat type where we expect that the largest past of European population occurs. As basis for this judgement we used the overlays that we made between the observations and the Corine Land Cover classes, showing the proportions of the observations in the various CLC-classes.

2.5. BioScore 2.0 models with only linear model relationships

The original BioScore 2.0 models were generated as univariate logistic models including a linear and a quadratic model term. The final model was determined by skipping the quadratic term and choosing the model with the lower AIC. This approach assumed that in the majority of cases the dose-response relationship would be a unimodal curve, in some cases—when the quadratic term is not significant attaining the shape of a sigmoidal (or even linear) curve (van Hinsberg *et al.* 2014).

Here, we compare the original BioScore 2.0 models with models containing only the linear term, thus assuming a sigmoidal relationship between explanatory and response variable. Like in the original analysis (Sierdsema 2014), we determine the robustness of the AUC of a given model by a ten-fold crossvalidation and used 20 bootstrap iterations—each with a random subset of 50% of the data—to determine the reliability of the dose-response relationship.

We did this analysis for the same set of ten sample species as listed in section 2.5 and thus compared 210 (10 species x the 21 pressure factors "desic_ mean_5km", "div3_1_20_mean_5km" "div3_1_50_ mean 5km", "div3 2 20 mean 5km", "div3 250 mean $5km$ ", "div3 320 mean 5km", "div3 3 50 mean 5km", "div4_1_20_mean_5km", "div4_1_50_mean_5km", "div4_2_20_mean_5km", "div4_2_50_ mean $5km$ ", "div $5\ 1\ 20\$ mean $5km$ ", "div5_1_50_mean_5km" "fma_f1_5km", "fma f2_5km", "fma_f3_5km", "fma_f4_5km", "fma_ f5_5km", "napplication_5km", "ndep_mean_5km"

and "sdep_mean_5km") pairs of unimodal and sigmoid dose-response relationships.

2.6. Effect of cut-off values on binary distribution maps

For the sample species listed in section 2.5 we determined the effect of the cut-off value used for transforming maps with presence probability into binary maps: all values larger than the cut-off are converted into 'presence', all values smaller into 'absence'.

We used three different approaches for choosing the cut-off:

1. Fixed cut-off

This is the most basic way of transforming probabilities of occurrence into presence-absence data. Using a threshold of 0.5 has been used widely used in ecology, but comparative studies showed that is the least reliable among all methods (LIU *et al.* 2005).

2. True skill statistic (TSS)

This statistics has been explained in section 2.2.1. We added another feature to the TSS, namely a weighing factor that modifies the importance of specificity compared to the importance of sensitivity. For example, when we aim to minimize the risk of displaying cells as 'present' although the species is absent, we can decrease the weight of specificity in the calculation of TSS and the distribution area will be enlarged compared to the application of TSS without weighing factor. In contrast, when we aim to minimize the risk of displaying cells as 'absence' although the species is present, we can increase the weight of specificity and the resulting distributional area will be smaller. In our analyses, we varied the weighing factor λ in the weighted TSS (= λ^* sensitivity + specificity -1) between 0.5 and 1.5.

3. Prevalence as cut-off

Prevalence is the proportion of presences in the data and this proportion can be used as cut-off. Despite its simplicity, this approach has been shown to be almost as good as the TSS (Allouche *et al.* 2006).

3. Results

3.1. Full multivariate models for all BioScore 2.0 species

For 16 species (*Chlamydotis undulata, Cursorius cursor, Cygnus bewickii, Emberiza caesia, Emberiza cineracea, Falco biarmicus, Falco eleonorae, Francolinus francolinus, Gavia immer, Lyrurus mlokosiewiczi, Melanocorypha bimaculata, Polysticta stelleri, Sitta krueperi, Sylvia rueppelli, Vanellus spinosus, Xenus cinereus*) a FMM could not be fitted with the standard parameter settings.

These species typically have a small distribution area and/or a low number of records. Reducing the number of absences in the observations and decreasing the learning rate to 0.001 resulted in successful modelling of *Cygnus bewickii* and *Falco eleonorae* but for the remaining 14 species still no FMMs could be achieved.

We combined the mean crossvalidation AUC and explained deviance to evaluate FMM quality in the following way:

According to this classification, the FMMs for 126 species could be regarded as *very good*, for 78 species as *good*, for 42 species as *moderate*, for 26 species as *poor*, and only for 13 species as *very poor* (Figure 3.1 and Appendix 1).

For the ten sample species, Figure 3.2 to 3.4 show the thirty most important explanatory variables in the respective FMMs.

Figure 3.1. Evaluation of FMM quality based on the mean AUC determined by a ten-fold crossvalidation and the explained deviance of the full model. Three "very poor" outlier species are not shown: Anser erythropus *(explained deviance = 48.8, mean AUC = 0.66),* Falco eleonorae *(explained deviance = 47.3, mean AUC = 0.71) and* Cygnus berwickii *(explained deviance = -12.3, mean AUC = 0.64).*

Figure 3.2. The thirty most important explanatory variables in the FMMs for the Great reed warbler Acrocephalus arundinaceus *(A), the Aquatic warbler* Acrocephalus paludicola *(B), the Eurasian skylark* Alauda arvensis *(C) and the European roller* Coracias garrulus *(D). Pressure variables are shown as blue symbols.*

Figure 3.3. The thirty most important explanatory variables in the FMMs for the White-backed woodpecker Dendrocopos leucotos *(A), the Woodlark* Lullula arborea *(B), the Corn bunting* Emberiza calandra *(C) and the Redbacked shrike* Lanius collurio *(D). Pressure variables are shown as blue symbols.*

Figure 3.4. The thirty most important explanatory variables in the FMMs for the Lesser grey shrike Lanius minor *(A) and the Grey partridge* Perdix perdix *(B). Pressure variables are shown as blue symbols.*

3.2. Comparison between BioScore 2.0 and full multivariate models

3.2.1. Evaluation of binary maps with the maximised TSS2

The evaluation of binary maps based on BioScore 2.0 models and FMMs produced heterogeneous results. For a few species as well BioScore 2.0 as FMM maps had high TSS values (upper right corner in Figure 3.5), for example the grey-headed chickadee *Poecile cinctus* (TSS.BioScore = 0.859, TSS.FMM = 0.975) (Figure 3.6), the rustic bunting *Emberiza rustica* $(TSS.BioScore = 0.855, TSS.FMM = 0.896)$, the rough-legged hawk *Buteo lagopus* (TSS.BioScore $= 0.852$, TSS.FMM $= 0.904$) and the Siberian jay *Perisoreus infaustus* (TSS.BioScore = 0.844, TSS. $FMM = 0.945$. Conspicously, all these species have a Scandinavian breeding range.

Various species showed moderate modelling success for both models (points close to the dashed diagonal in Figure 3.5), for example the boreal owl *Aegolius funereus* (TSS.BioScore = 0.735, TSS.FMM = 0.779), the black woodpecker *Dryocopus martius* (TSS. $BioScore = 0.644$, TSS.FMM = 0.711) and the golden eagle *Aquila chrysaetos* (TSS.BioScore = 0.635, TSS. $FMM = 0.705$).

In both former cases the BioScore 2.0 models came close to the benchmark established by the FMMs. For some other species TSS.FMM attained very high values close to one, indicating almost perfect agreement with the observations, while the corresponding TSS.BioScore values were nearly zero, indicating random relationship between observations and binary predictions (upper left corner in Figure 3.5), for example the pallid harrier *Circus macrourus* (TSS. $BioScore = -0.009$, TSS.FMM = 0.9997), western rock nuthatch *Sitta neumayer* (TSS.BioScore = 0.002, TSS.FMM = 0.959), the Barbary partridge *Alectoris barbara* (TSS.BioScore = 0.051, TSS.FMM = 0.999) and the pygmy cormorant *Microcarbo pygmeus* (TSS.BioScore = 0.004, TSS.FMM = 0.904). Noticeable, these are South or South-East European breeding species.

The wide scatter of data points in Figure 3.5 stands in vivid contrast to the results of the butterfly models where the TSS.FMM and TSS.BioScore values show a close linear relationship and the majority of BioScore 2.0 models comes very close to the FMMs (van Swaay 2016).

Tables showing the maximised TSS for each species per model type can be found in Appendix 2.

² Careful: the TSS is used as a quality measure for the evaluation of binary maps as well as a measure of correspondence between BioScore and FMM binary maps! In order not to mix them up, we use subscripts to make clear to which application a TSS refers: TSS. BioScore and TSS.FMM refer to the evaluation of binary BioScore 2.0 and FMM maps, respectively; TSS.map_correspondence refers to the correspondence of binary BioScore 2.0 and FMM maps for a given species.

Figure 3.6. Binary distribution maps of the Greyheaded chickadee Poecile cinctus *based on the BioScore 2.0 model (A) and on a full multivariate model (FMM) (B) as well as the difference map between them (C). The difference map was produced by subtracting the presence value (0 or 1) of the full multivariate model map from the presence value of the BioScore 2.0 map. Difference values of 1 thus show the areas where BioScore 2.0 predicted presence but the FMM predicted absence (red), values of -1 show the areas where BioScore 2.0 predicted absence but FMM predicted presence (blue).*

Figure 3.7. Frequency distribution of concordance between 273 binary breeding bird distribution maps based on full multivariate models and BioScore 2.0 models as evaluated by overall accuracy.

3.2.2. Correspondence between binary maps Overall accuracy showed a negatively skewed distribution with a mode higher than 0.9, indicating good correspondence between BioScore 2.0 and FMM binary maps for most of the species (Figure 3.7). For some species, however, overall accuracy is smaller than 0.7.

Not surprisingly, the species with both high TSS. FMM and TSS.BioScore values score high when the species are ordered according to decreasing overall accuracy: the above-mentioned species rough-legged hawk, grey-headed chickadee, rustic bunting and Siberian jay are 15^{th} , 17^{th} , 29^{th} and 39^{th} , respectively, in a list of 273 species (see Table App3.1 in Appendix 3 for the full list). As pointed out in section 2.2.2, overall accuracy is dependent on the prevalence of the dominant class (*presence* or *absence*). This can distinctly be seen for the species that have been mentioned as examples for the case "TSS.FMM is

Figure 3.8. Frequency distribution of concordance between 273 binary breeding bird distribution maps based on full multivariate models and BioScore 2.0 models as evaluated by the true skill statistic.

high and TSS.BioScore is close to zero", or in other words, "FMM binary map is good and BioScore 2.0 binary map is poor": the pygmy cormorant ranks $55th$, the western rock nuthatch ranks $14th$, the pallid harrier ranks 8th, and the Barbary partridge even is on 4th place! For these species with a restricted breeding area, the large numbers of 'correctly' predicted absences cause a huge overall accuracy, even if the few predicted presences are completely misplaced.

The TSS as a measure of binary map correspondence better reflects the information attained during map evaluation. It is less optimistic than overall accuracy and its frequency distribution is not biased towards higher values (Figure 3.8). The species with both high TSS.FMM and TSS.BioScore values generally scored even higher in a list with species ordered according to decreasing TSS.map_correspondence (rough-legged hawk $6th$, rustic bunting $8th$, Siberian jay 24th, grey-headed chickadee $37th$) (see Table App3.2 in Appendix 3 for the full list) than in the list ordered according to decreasing overall accuracy. In

contrast to the evaluation based on overall accuracy, the species with high TSS.FMM and a TSS.BioScore close to zero are only 215th (pygmy cormorant), 236th (pallid harrier), $268th$ (western rock nuthatch) and 270th (Barbary partridge), confirming their difference in map quality between FMMs and BioScore 2.0 models.

TSS, however, not always performs in an optimal way. The red-throated pipit *Anthus cervinus*, for example, is predicted for the north of Sweden and Finland by both modelling approaches. Inspecting the distribution maps (Figure 3.9) by eye would lead to the conclusion that, apart from some distribution details, both draw a good picture of the real situation and that they show a high degree of concordance. Overall accuracy for the species is 0.987, ranked on 6th place (Table App3.1 in Appendix 3), but TSS. map_correspondence is only 0.593, ranked on 66th place (Table App3.2 in Appendix 3), and thus overemphasizing the differences in distribution details.

Figure 3.9. Binary distribution maps of the Redthroated pipit Anthus cervinus based on the BioScore 2.0 model (A) and on a full multivariate model (FMM) (B) as well as the difference map between them (C). The difference map was produced by subtracting the presence value (0 or 1) of the full multivariate model map from the presence value of the BioScore 2.0 map. Difference values of 1 thus show the areas where BioScore 2.0 predicted presence but the FMM predicted absence (red), values of -1 show the areas where BioScore 2.0 predicted absence but FMM predicted presence (blue).

Figure 3.10. Scatterplots of map correspondence (TSS.map_correspondence) versus quality of binary maps based on full multivariate models (TSS.FMM) (A) and of binary maps based on BioScore 2.0 model (TSS.BioScore) (B).

TSS.BioScore and TSS.map_correspondence are highly correlated $(r = 0.79, p < 0.001)$ whereas TSS. map_correspondence shows no relationship with TSS.FMM (*r* = -0.11, *p* = 0.06) (Figure 3.10). We conclude that the correspondence between FMM and BioScore 2.0 maps depends only on the quality of the BioScore 2.0 maps: the better the BioScore 2.0 models, the better their correspondence with the FMM maps. An evaluation of which species are appropriate for being modelled according to the BioScore 2.0 approach can thus simply be based on the TSS.BioScore (Table App2.2 in Appendix 2).

3.2.3. Final evaluation and expert judgement Evaluating the maps based on the TSS of binary FMM and BioScore 2.0 maps and on the TSS of map correspondence according to the scheme presented in section 2.2.3, resulted in only 20 species that are satisfactorily modelled by BioScore 2.0 (Table 3.1). The vast majority of the species (86 %) fall in the category "doubtful is species should be used for BioScore 2.0" and called for an inspection of the maps by an expert.

Due to the high share of doubtful species, the expert made advantage of the opportunity and evaluated *all* species and, at the same time, not only judged the BioScore 2.0 maps but also the FMM maps. This led to some unexpected results.

For 29 species (10 %) the expert judged FMM as well as BioScore 2.0 to deliver good maps (Table 3.2). While according to TSS the FMM maps exceptionless outperformed their BioScore 2.0 counterparts (Figure 3.5), the expert evaluated 83 (32 %) BioScore 2.0 maps as good, while only 63 (24 %) FMM maps received this score. 97 (37 %) BioScore 2.0 maps were regarded as being poor or bad, but 121 (47 %) FMM maps scored evenly bad. In summary, according to expert judgement BioScore 2.0 delivered better maps than the FMMs.

It should however be stressed that the expert evaluated the binary Bioscore maps and the continuous FMM-maps. Unexpected differences in probability resulted often in a 'poor' or 'bad' score, despite the fact that the range was predicted quite well. This

*Table 3.1. Verdict over BioScore 2.0 maps based on TSS.FMM, TSS.BioScore and TSS.map_correspondence. Strict, the desired TSS.BioScore was calculated as 0.9 * TSS.FMM (see Introduction, section 2.2.3); Gentle, for the same calculation 0.8 was used instead of 0.9.*

Table 3.2. Expert evaluation of FMM binary maps and BioScore 2.0 binary maps. Frequency per modelling approach and quality class is presented, along with the marginal sums.

may also be the basis for the observed differences between the formal evaluation and the experts' evaluation: the formal evaluation tends to look more to general quality, while the expert tends to too give focus on the local deviations.

All TSS values along with expert evaluation and comments for each species are presented in Table App3.3 in Appendix 34.

3.3. Comparison between dose-response relations and partial dependence plots

3.3.1. Effect of zooming-in

As described in the methods section, dose-response curves will be displayed by zooming in to the range

of variation of the probability of presence. The effect of zooming-in and the shape of a partial dependence plot for a given variable when the y-axis ranges from 0 to 1 are shown in Figure 3.11.

3.3.2. Concordance between dose-response curves and partial dependence plots

Generally, the dose-response curves and partial dependence plots showed a low degree of concurrence. Among the many examples are the curves of div5_1_20_5km for the great reed warbler *Acrocephalus arundinaceus* (Figure 3.12): the doseresponse curve is unimodal with its mode at 0.2 while the partial dependence plots shows a marked increase at very low levels and a horizontal line for the rest of the values of the explanatory variable.

*Figure 3.11. Partial dependence plots from a full multivariate model of the Great reed warbler Acrocephalus arun*dinaceus *for the explanatory variable div3_2_20_5km with probability of presence ranging from 0 to 1 (left) and zoomed in to the actual range of probability of presence (right). Dots at y = 0 and y = 1 in the left figure represent the observations with absences and presences, respectively, for given values of the explanatory variable. Rugs at the x-axis in the right figure represent observations for given values of the explanatory variable, not distinguishing between presences and absences.*

Figure 3.12. Dose effect relations based on BioScore 2.0 univariate binomial models (left panels) and partial dependence plots from a full multivariate model (right panels) of the Great reed warbler Acrocephalus arundinaceus *for the variables div5_1_20_5km (upper panels), div3_2_20_5km (middle panels) and dive_1_20_5km (lower panels) to div4_1_20_mean_5km. The AUC (area under the receiver operating characteristic curve) for the BioScore 2.0 models and the relative variable importance for the FMM are reported. Dots at y = 0 and y = 1 in the dose-response curve panels represent the observations with absences and presences, respectively, for given values of the explanatory variable. Rugs at the x-axis in the partial dependence plot panels represent observations for given values of the explanatory variable, not distinguishing between presences and absences.*

In some cases some correspondence can be observed. The presence probability of the great reed warbler decreases with increasing values of div3 2 20 5km and div3 1 20 5km and also the corresponding partial dependence plots show the same pattern (Figure 3.12). The gradual decrease in the dose-response curves, however, is accompanied by a very steep decrease in the partial dependence plots. Tree-based models such as the boosted regression trees that have been used for making the full multivariate models are known for their capability of identifying thresholds, while the dose-response curves of binomial GLMs are forced into a smooth unimodal or sigmoidal shape. Varying div3 2 20 5km and div3_1_20_5km while averaging out all other explanatory variables, as is done in the partial dependence plots, results in only slight changes in the probability of presence which is visible on the y-axis. This is not surprising since the relative importance of these variables is only 2.8 and 0.7 %, respectively. Notice that the change in the probability of presence for the univariate dose-response relations is considerably larger when these pressure-variable vary over their range.

3.3.3. Comparison of the most important explanatory variable per species

In Figures 3.13 to 3.16 we show a selected subset of dose-response curves and partial dependence plots. We chose those pressure variables that were identified as the most important ones in the FMMs and at the same time had an AUC > 0.6 in the unimodal models used for BioScore 2.0. For the European roller (Figure 3.14A), the corn bunting *Emberiza calandra* (Figure 3.15A) and the red-backed shrike *Lanius collurio* (Figure 3.15B) some resemblance can be perceived, but for the others the shapes of dose-response curves and partial dependence plots do not show much concordance. Moreover, some of the most important pressure variables as identified by the FMMs had an AUC < 0.6 and thus had no explanatory power in the univariate models. This was the case for sdep_mean_5km for the white-backed woodpecker *Dendrocopos leucotos* and the corn bunting *Emberiza calandra* and for desic_mean_5km for the lesser grey shrike *Lanius minor*. In the case of the grey partridge *Perdix perdix* even the three most important pressure variables (ndep_mean_5km, div5_1_50_mean_5km, div5_1_20_mean_5km) had no explanatory power in the univariate models.

Figure 3.13. Dose effect relations based on BioScore 2.0 univariate binomial models (left) and partial dependence plots from a full multivariate model (right panels) of the Great reed warbler Acrocephalus arundinaceus*, the Aquatic warbler* Acrocephalus paludicola *and the Eurasian skylark* Alauda arvensis *for the most important pressure variable according to the FMMs among those with an AUC > 0.6 in the unimodal models for BioScore 2.0.*

Figure 3.14. Dose effect relations based on BioScore 2.0 univariate binomial models (left) and partial dependence plots from a full multivariate model (right panels) of the European roller Coracias garrulus *(A), the White-backed woodpecker* Dendrocopos leucotos *(B) and the Woodlark* Lullula arborea *(C) for the most important pressure variable according to the FMMs among those with an AUC > 0.6 in the unimodal models for BioScore 2.0.*

Figure 3.15. Dose effect relations based on BioScore 2.0 univariate binomial models (left) and partial dependence plots from a full multivariate model (right panels) of the Corn bunting Emberiza calandra *(A), the Red-backed shrike* Lanius collurio *(B) and the Lesser grey shrike* Lanius minor *(C) for the most important pressure variable according to the FMMs among those with an AUC > 0.6 in the unimodal models for BioScore 2.0.*

Figure 3.16. Dose effect relations based on BioScore 2.0 univariate binomial models (left) and partial dependence plots from a full multivariate model (right panel) of the grey partridge Perdix perdix *for the most important pressure variable according to the FMMs among those with an AUC > 0.6 in the unimodal models for BioScore 2.0.*

3.4. Specification of typical and characteristic BioScore 2.0 species

Appendix 7 shows the assignment of the Bioscore 2.0 bird species into the four main land use classes urban, farmland, forest and natural habitats. The assignment shows the land use type where we expect that the largest part of the European population will occur. It should be mentioned that this proportion, especially for species of open habitats, is not the same across their range: many species, like the Meadow pipit *Anthis pratensis*, that are abundant in moors and treeless open natural areas in northern Europe and mountain ranges, occur mostly in farmlands in the southern parts of their range.

3.5. BioScore 2.0 models with only linear model relationships

Sixty-one of the 210 unimodal dose-response relationships (10 sample species x 21 pressure factors, see Methods, section 2.5) had AUC-values of at least 0.6 (variables with AUC < 0.6 were regarded as having no effect on presence probability). Among the comparisons of the unimodal and sigmoid dose-response curves of these 61 variables, 33 had identical AUC for the Bioscore2 models and the models with only a linear term, 15 had a higher AUC in the Bisocore2 model and 13 higher AUC in the models with only a linear term (Appendix 5). These values are variable themselves due to the stochasticity in the 10-fold crossvalidation and can vary by values of ±0.01.

They show, however, the negligible differences in terms of predictive power.

In various cases it could be observed that the rightside part of the unimodal dose-effect curves are rather unstable. As an example, we look at the variable div4 $\,$ 1 $\,$ 20 $\,$ mean $\,$ 5km for the Great reed warbler *Acrocephalus arundinaceus* (Figure 3.17). Due to the low number of data points with div4_1_20_ mean_5km larger than 0.2, the curves based on bootstrapped subsample from the data show a considerable variability. This effect has already been described by SIERDSEMA (2014). The unimodal shape of the dose-effect relation is enforced by the quadratic term although there are no absences reported for large values of div4_1_20_mean_5km. When only a linear term is used for modelling, the dose-effect relation is represented by the right-side half of a sigmoid curve and and its variability is much lower (Figure 3.17). The AUC (the area under the receiver operating characteristic curve) of the models are almost identical because the few data points left to $div4_1_2o_mean_5km = 0.2$ have low impact on overall sensitivity and specificity. Remarkably, the Akaike information Criterion (AIC) suggested to keep the quadratic term in the model in the Bioscore modelling run although almost no difference in model performance is visible between the models with and without it.

The enforcement of a unimodal shape of the doseeffect curve can repeatedly be seen, for example also for div5_1_50_mean_5km for the Great reed

Figure 3.17. Dose effect relations based on a univariate binomial model relating the probability of presence of the Great reed warbler Acrocephalus arundinaceus *to div4_1_20_mean_5km, including a linear and a quadratic term as used in Bioscore2 (left) and including only the linear term (right). The AUC (area under the receiver operating characteristic curve) for both models are reported. Dots at y = 0 and y = 1 represent the observations with absences and presences, respectively, for given values of the explanatory variable.*

warbler (Figure 3.18). Again, above a certain value of the explanatory variable (here: 0.35) no absences were observed but the curve still bends downward. The alternative model with only the linear term is represented by the central, almost linear part of a very flat sigmoidal curve. It has an even higher AUC than the original Bioscore2 model, which makes even

more visible that the model-selection based on the AIC depends to a large extent on the stochasticity of the 10-fold crossvalidation, or in plain words: if the quadratic term is kept in the model or not seems to be coincidental.

Appendix 5 summarises the effects of keeping only the linear term for ten sample species.

Figure 3.18. Dose effect relations based on a univariate binomial model relating the probability of presence of the Great reed warbler Acrocephalus arundinaceus *to div5_1_50_mean_5km, including a linear and a quadratic term as used in Bioscore2 (left) and including only the linear term (right). The AUC (area under the receiver operating characteristic curve) for both models are reported. Dots at y = 0 and y = 1 represent the observations with absences and presences, respectively, for given values of the explanatory variable.*

3.6. Effect of cutoff values on binary distribution maps

3.6.1. Calculation of the cutoff value with the true skill statistics (TSS)

We demonstrate the effect of cutoff choice on sensitivity, specificity and distributional area for the Great reed warbler *Acrocephalus arundinaceus* when calculated by the TSS (Table 3.3, Figure 3.20). Varying the weighing factor for specificity from 0.5 to 1.5 produces cutoffs ranging from 0.06 to 0.67. The change of sensitivity and specificity for different cutoffs can clearly be seen. For example, for a weighing factor of 0.5 99.41 % of the presences are correctly classified, whereas only 49.85 % of the absences are correctly predicted. Consequently, the predicted distribution area is very large and only very few presences are not covered by this area. On the contrary, if a high weighing factor of 1.5 is chosen, almost all absences (96.54 %) are correctly predicted, but sensitivity is low (64.36 %) and a large proportion of the presences are missed. The corresponding distribution area is very small.

In practice, such extreme factor values would not be chosen. It might for many species be desirable, however, to fine-tune the cutoff, setting the weighing factor to values close to but not identical with 1.0, as shows the example of the Aquatic warbler *Acrocephalus paludicola*. This is one of the rarest passerine species in Europe, being restricted to wetlands in Poland, Belarus en Ukraine (apart from some very small populations in Hungary en Western Siberia). The cutoff based on the standard TSS

Table 3.3. The effect of the weighing factor for the true skill statistic on cutoff, sensitivity and specificity for the Great reed warbler Acrocephalus arundinaceus, based on the probabilities of presence predicted by the soilclimate distribution model.

(weighing factor 1.0, cutoff 0.001) leads to a considerable overestimation of the distribution area of the species and predicts it for a large contiguous region far into Southern Russia and even for Northern Scandinavia. Increasing the weighing factor to 1.1 raises the cutoff value to 0.036 and causes the distribution area to shrink to a realistic extent (Figure 3.19).

3.6.2 Comparison with fixed and prevalencebased cutoff

A fixed cutoff value of 0.5 does not take the relative importance of sensitivity and specificity into consideration. In the case of the Great reed warbler, a

 $Factor = 1.0$; cutoff = 0.001

Figure 3.19. The effect of the weighing factor for the true skill statistic on the predicted distribution area of the Aquatic warbler Acrocephalus paludicola*, based on the probabilities of presence predicted by the soil-climate distribution model.*

 $Factor = 0.5$; cutoff = 0.06

 $Factor = 0.6$; cutoff = 0.11

Factor = 0.7 ; cutoff = 0.15

 $Factor = 0.8$; cutoff = 0.18

 $Factor = 1.1$; cutoff = 0.39

 $Factor = 0.9$; cutoff = 0.25

Factor = 1.2 ; cutoff = 0.48

Factor = 1.4 ; cutoff = 0.63

Factor = 1.5 ; cutoff = 0.67

Figure 3.20. The effect of the weighing factor for the true skill statistic on the cutoff value and the predicted distribution area of the Great reed warbler Acrocephalus arundinaceus*, based on the probabilities of presence predicted by the soil-climate distribution model.*

Factor = 1.0 ; cutoff = 0.32

Factor = 1.3 ; cutoff = 0.57

28

Figure 3.21. Comparison of binary distribution maps of the Great reed warbler Acrocephalus arundinaceus produced with a fixed cutoff at 0.5 (left), the cutoff as calculated by the TSS (centre) and prevalence used as cutoff (right), based on the probabilities of presence predicted by the soil-climate distribution model.

cutoff of 0.5 would be attained with the TSS with the weighing factor set between 1.2 and 1.3. For this species, the fixed value implicitly gives high weight to the correct prediction of absences, resulting in a comparably small distribution area (Figure 3.21). Prevalence as a cutoff (0.33), on the other hand, is very close to the standard TSS cutoff with even

weight for sensitivity and specificity (0.32) and yields an almost identical distribution map.

Generally, cutoffs based on the prevalence lie closely to the TSS cutoffs with weighing factors set at 0.9, 1.0 or 1.1 (see Appendix 6 for a complete table of cutoff values for 10 sample species).

References

Allouche O., Tsoar A. & Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43:1223-1232.

BARBET-MASSIN M., THUILLER W. & JIGUET F. 2012. The fate of European breeding birds under climate, land-use and dispersal scenarios. Global Change Biology 18:881-890.

Elith J., Leathwick J.r. & Hastie T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77:802-813.

Friedman J.h. 2001. Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics 29:1189-1232.

Kampichler C., Hallmann C. & Sierdsema H. 2015. TRIMmaps: an R package for the analysis of species abundance and distribution data – Extended manual. Sovon Vogelonderzoek Nederland, Nijmegen.

Legendre P. 2005. Species Associations: The Kendall Coefficient of Concordance Revisited. Journal of Agricultural, Biological and Environmental Statistics 10:226–245.

Liu C., Berry P. M., Dawson T. P. & Pearson R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28:385-393.

Reineking R. & Schröder B. 2004. Gütemaße für Habitatmodelle. *In:* DORMANN C. F., BLASCHKE T., Lausch A., Schröder B. & Söndgerath D. (*Eds.*) Habitatmodelle – Methodik, Anwendung, utzen. UFZ-Berichte 9/2004.

SIERDSEMA H. 2014. Technical report breeding birds in BioScore 2.0. Sovon-report 2014/52. Sovon Vogelonderzoek Nederland, Nijmegen.

SIERDSEMA H. 2015. Technical report breeding birds in BioScore 2.0: additional questions. Sovon-report 2015/34. Sovon Vogelonderzoek Nederland, Nijmegen.

Thuiller W., Lafourcade B., Engler R. & Araújo M.B. 2009. BIOMOD – a platform for ensemble forecasting of species distributions . Ecography 32:369-373

van Hinsberg A., Hendriks M., Hennekens S., Sierdsema H., van Swaay C., Rondinini C., SANTINI L., DELBAERE B., KNOL O., WIERTZ J. 2014. BioScore $2.0 - A$ tool to assess the impacts of European Community policies on Europe's biodiversity. First Draft.

Zimmermann N. E., Yoccoz N. G., Edwards T. C. Jr., Meier E.S., Thuiller W., Guisan A., SCHMATZ D.R. & PEARMAN P.B. 2009. Climatic extremes improve predictions of spatial patterns of tree species. PNAS 106, Suppl. 2:10723-19728.

Appendices

Appendix 1: Evaluation of full multivariate models

Species with very good FMM quality (126):

Accipiter brevipes Acrocephalus paludicola Aegypius monachus Alectoris barbara Alectoris chukar Alectoris graeca Anthus cervinus Apus caffer Aquila adalberti Aquila heliaca Ardea purpurea Ardeola ralloides Aythya nyroca Buteo lagopus Buteo rufinus Calandrella rufescens Calcarius lapponicus Calidris maritima Calidris minuta Caprimulgus ruficollis Cecropis daurica Charadrius alexandrinus Charadrius morinellus Chersophilus duponti Chlidonias hybrida Ciconia ciconia Ciconia nigra Clanga clanga Clanga pomarina Clangula hyemalis Cygnus cygnus Dendrocopos leucotos Dendrocopos syriacus Elanus caeruleus Emberiza melanocephala Emberiza pusilla Emberiza rustica Eremophila alpestris Falco cherrug Falco naumanni Falco rusticolus Falco vespertinus Ficedula albicollis Ficedula parva Ficedula semitorquata Fulica cristata Galerida cristata Galerida theklae Gallinago media

Gavia arctica Gelochelidon nilotica Glareola pratincola Glaucidium passerinum Grus grus Gypaetus barbatus Gyps fulvus Haliaeetus albicilla Hieraaetus pennatus Himantopus himantopus Hippolais icterina Hippolais olivetorum Hydrobates pelagicus Ichthyaetus audouinii Lagopus lagopus Lagopus muta Lanius meridionalis Lanius minor Lanius nubicus Lanius senator Larus canus Locustella fluviatilis Loxia leucoptera Loxia pytyopsittacus Loxia scotica Luscinia luscinia Lyrurus tetrix Marmaronetta angustirostris Melanitta fusca Melanitta nigra Melanocorypha calandra Merops apiaster Microcarbo pygmeus Monticola solitarius Motacilla citreola Neophron percnopterus Nucifraga caryocatactes Numenius phaeopus Oenanthe hispanica Oenanthe isabellina Oenanthe leucura Oenanthe pleschanka Otis tarda Oxyura leucocephala Passer hispaniolensis Pelecanus crispus Pelecanus onocrotalus Perisoreus infaustus Phoenicopterus roseus Phylloscopus borealis Platalea leucorodia Plectrophenax nivalis

Plegadis falcinellus Poecile cinctus Poecile lugubris Porphyrio porphyrio Prunella collaris Pterocles alchata Pterocles orientalis Pyrrhocorax graculus Pyrrhocorax pyrrhocorax Recurvirostra avosetta Sitta neumayer Stercorarius longicaudus Strix nebulosa Strix uralensis Sturnus unicolor Sylvia cantillans Sylvia crassirostris Sylvia melanocephala Sylvia nisoria Sylvia sarda Sylvia undata Tetrao urogallus Tetrastes bonasia Tetrax tetrax Tringa glareola

Species with good FMM quality (78):

Acrocephalus melanopogon Aegolius funereus Alectoris rufa Anas penelope Anthus campestris Anthus pratensis Anthus trivialis Aquila chrysaetos Aquila fasciata Ardea alba Botaurus stellaris Branta leucopsis Burhinus oedicnemus Calandrella brachydactyla Calidris alpina Chlidonias niger Chroicocephalus genei Circaetus gallicus Clamator glandarius Coracias garrulus Corvus frugilegus Crex crex Dendrocopos medius Dryocopus martius

Egretta garzetta Emberiza calandra Emberiza cia Emberiza cirlus Emberiza citrinella Emberiza hortulana Gallinago gallinago Gavia stellata Haematopus ostralegus Hippolais polyglotta Hydrocoloeus minutus Hydroprogne caspia Ichthyaetus melanocephalus Lanius collurio Lanius excubitor Limosa lapponica Limosa limosa Locustella luscinioides Locustella naevia Lophophanes cristatus Luscinia megarhynchos Luscinia svecica Mergellus albellus Mergus merganser Milvus migrans Monticola saxatilis Numenius arquata Nycticorax nycticorax Oriolus oriolus Otus scops Pandion haliaetus Phalaropus lobatus Phylloscopus bonelli Phylloscopus sibilatrix Phylloscopus trochilus Picoides tridactylus Picus canus Pluvialis apricaria Podiceps auritus Poecile montanus Regulus ignicapilla Regulus regulus Saxicola rubetra Scolopax rusticola Sterna paradisaea Sternula albifrons Streptopelia turtur Surnia ulula Sylvia curruca Sylvia hortensis Tadorna ferruginea Thalasseus sandvicensis Upupa epops Uria aalge

Species with moderate FMM quality (41):

Acrocephalus arundinaceus Anas acuta Anas querquedula Anas strepera Athene noctua Aythya ferina Aythya fuligula Bubo bubo Bucanetes githagineus Circus aeruginosus Circus macrourus Circus pygargus Corvus corax Coturnix coturnix Emberiza schoeniclus Falco columbarius Ixobrychus minutus Jynx torquilla Loxia curvirostra Lullula arborea Milvus milvus Motacilla flava Netta rufina Oenanthe oenanthe Panurus biarmicus Passer montanus Perdix perdix Philomachus pugnax Phoenicurus phoenicurus Poecile palustris Porzana parva Porzana pusilla Prunella modularis Saxicola torquatus Sitta europaea Sterna dougallii Sterna hirundo Sturnus vulgaris Sylvia borin Tringa totanus Vanellus vanellus

Species with poor FMM quality (26):

Accipiter gentilis Acrocephalus scirpaceus Alauda arvensis Anas clypeata Asio flammeus Caprimulgus europaeus Circus cyaneus Coloeus monedula

Columba oenas Dendrocopos major Dendrocopos minor Falco peregrinus Fulica atra Hirundo rustica Linaria cannabina Motacilla cinerea Muscicapa striata Pernis apivorus Phylloscopus collybita Picus viridis Podiceps cristatus Porzana porzana Saxicola rubicola Strix aluco Sylvia communis Tyto alba

Species with very poor FMM quality (10):

Alcedo atthis Columba palumbus Falco subbuteo Falco tinnunculus Glareola nordmanni Motacilla alba Sylvia atricapilla Tachybaptus ruficollis Turdus viscivorus Vanellus gregarius
Appendix 2: Maximised True Skill Statistics (TSS) of binary maps **Appendix 2: Maximised True Skill Statistics (TSS) of binary maps**

Table App2.1: TSS of binary maps based on BioScore 2.0 models (TSSBioScore) and on full multivariate models (TSSFMM), ordered according to decreasing TSSFMM.

Sovon-report 2016/06

Validation Bioscore 2.0 distribution maps for breeding birds

Table App2.2: TSS of binary maps based on BioScore 2.0 models (TSS_{BioScore}) and on full multivariate models (TSS_{FMM}), ordered according to decreasing TSS_{BioScore}.

Validation Bioscore 2.0 distribution maps for breeding birds

Validation Bioscore 2.0 distribution maps for breeding birds

Appendix 3: Correspondence of BioScore 2.0 and FMM binary maps **Appendix 3: Correspondence of BioScore 2.0 and FMM binary maps**

Table App3.1: Correspondence between binary maps based on BioScore 2.0 models and full multivariate models as measured by overall accuracy ordered, ordered according to decreasing overall accuracy.

Table Appendix 3.2: Correspondence between binary maps based on BioScore 2.0 models and full multivariate models as measured by the TSS, ordered according to decreasing TSS.

Appendix 4: Integrative table of map evaluation Appendix 4: Integrative table of map evaluation

Based on true skill statistics and on expert judgement. TSS FMM, true kill statistic of the binary FMM map; TSS Bioscore, true kill statistic of the binary BioScore 2.0 map; TSS comp, true skill statistic of map concordance between binary FMM and BioScore 2.0 maps; TSS Eval, evaluation based on TSS FMM, TSS Bioscore and TSS comp: 1 - "Species is well modelled by BioScore 2.0", 2 - "Situation unclear, an expert should visually check and evaluate the maps", 3 - "Species cannot be modelled", 4 - "Doubtful if the species should be used for BioScore 2.0, but an expert should visually check and evaluate the maps".

Appendix 5: Comparison of AUC and dose-effect relation between univariate Appendix 5: Comparison of AUC and dose-effect relation between univariate Bioscore2 Bioscore2 pressure models and alternative univariate models with only a linear term pressure models and alternative univariate models with only a linear term

Comparison of AUC and summarizing description of dose-effect relation between univariate Bioscore2 pressure models and alternative univariate models with only a linear term. Ten sample species are analysed. Only variables where at least in one of the two models an AUC of 0.60 or higher was attained are shown. The direction of doseeffect relations is described as from the left to the right; 'ascending' thus means that presence probability increases with increasing values of the explanatory variable. Used abbreviations: AUC-B, AUC of Bioscore2 model; AUC-L, AUC of model with only linear term; DRC-B, dose-response curve of the Bisocore2 model; DCR-L, dose-response curve of the model with only linear term.

Appendix 6. Effect of cutoff values on binary distribution maps Appendix 6. Effect of cutoff values on binary distribution maps

Cut-off values for transforming maps with presence probability into binary maps for ten sample species. Cut-offs were determined according to three methods: calculating cut-off with the the True Skill Statistic (TSS), using a species' prevalence as cut-off and choosing a fixed cut-off. Weighing factor, factor for weighing the relative importance of sensitivity and specificity in the calculation of the cut-off with the TSS; Cut-off, the cut-off caluclated with the TSS; Sensitivity and Specificity, the resulting sensitivity and specificity for a determined cut-off; Prevalence, the prevalence for a given species; Fixed, the fixed cut-off. The values of Prevalence and Fixed are written in the lines with the values of Cut-Off closest to Prevalence and Fixed.

Appendix 7. Assignment of species to the four main land use classes Appendix 7. Assignment of species to the four main land use classes

In opdracht van:

Planbureau voor de Leefomgeving

Sovon Vogelonderzoek Nederland

Postbus 6521 6503 GA Nijmegen Toernooiveld 1 6525 ED Nijmegen T (024) 7 410 410

E info@sovon.nl I www.sovon.nl

