Are Landscape Configuration Metrics Worth Including When Predicting Specialist and Generalist Bird Species Density? A Case of the Generalised Additive Model Approach

Jakub Z. Kosicki¹

Received: 14 March 2017 / Accepted: 31 July 2017 / Published online: 10 August 2017
© The Author(s) 2017. This article is an open access publication

Abstract Due to the complexity of factors that influence species density on a large geographical scale, the effectiveness of the species distribution model (SDM) is still debatable. That is why the buffer zone (the area within 100 m from the outside edge of the patch), the core, i.e. (patches excluding the 100 m buffer zone from the patch's edge) and patch shape are explored in this study to see how they affect the density of habitat specialist and generalist bird species. Two sets of generalised additive models were generated separately for each of the four bird species: One set of models contained landscape configuration metrics as an additional predictor variable, and the other did not. The results showed that models including the core, the buffer zone and the shape of patches turned out to be definitely better than models without them. Specialist species, the Corn bunting and the Wood nuthatch, are more likely to occur in the core of the preferred patches, and they choose those of a simple shape; while generalist species, the Whinchat and the Tree pipit, are more probable to be present in the buffer zone of a more complicated shape. Thus, the results clearly show that specific landscape configuration models can improve the predictive power of SDMs and can be used as an effective tool for predicting species density and functional bird diversity (specialist and generalist). Furthermore, from the applied ecology perspective, detailed landscape

configuration metrics can be considered as a surrogate of elusive habitat conditions.

Keywords Species distribution model · Landscape configuration metrics · Shape · Core · Buffer zone · Predictive map

1 Introduction

Species distribution models (SDMs) are a group of multiple statistical tools that have been widely used to predict species occurrence, density and richness [1, 2]. These methods are based on an ecological niche concept where environmental factors are combined with species occurrence/density in order to create predictive maps of species distribution [3]. What is most important from the ecological point of view is that these methods show tendencies how environmental estimates affect species occurrence or density [4]. The modelling framework typically comprises such predictors as land-use typologies, climate estimation, the level of green vegetation and biotic interactions [5–8]. However, when specific landscape metrics are not taken into account in the modelling approach as additional predictor variables, SDMs effectiveness may be decreased. Recent studies conducted on a small spatial scale showed that patterns of species distribution were not only the result of a complex interplay between environmental estimates and landscape heterogeneity but also more detailed landscape configuration metrics, i.e. patch shape and the proportion between the core and the buffer zone of preferred habitats, which also played an important role [9–11]. Yet, this kind of data are rarely used in SDMs, especially in species density modelling [12–14]. Therefore, the usual approach was extended so as to analyse factors shaping the distribution of four bird species, using not only topography, climate, vegetation and

Electronic supplementary material The online version of this article (doi:10.1007/s10666-017-9575-1) contains supplementary material, which is available to authorized users.

✉ Jakub Z. Kosicki
kubako@amu.edu.pl

¹ Department of Avian Biology and Ecology, Faculty of Biology, Adam Mickiewicz University, Poznań, Umultowska 89, 61–614 Poznań, Poland

land use types but also landscape configuration metrics, i.e. the shape of the core and the buffer zone, and consider them as additional predictors.

This study analyses spatial distribution patterns of two farmland (Corn bunting *Emberiza calandra*, Whinchat *Saxicola rubetra*) and two forest (Tree pipit *Anthus trivialis* and Wood nuthatch *Sitta europaea*) breeding bird species of different ecological requirements [15]. According to a small spatial scale study, the Corn bunting prefers large areas of homogeneous crop fields [16, 17], while the Whinchat favours areas dominated by pastures interspersed with small arable fields [18]. By contrast, the Tree pipit is associated with edges of rare deciduous and coniferous forests, even though its woodland availability is not a primary factor influencing the species occurrence [19], whereas the Wood nuthatch occurs in the core of large areas of deciduous and mixed forests, and is negatively associated with landscape heterogeneity [20]. Thus, despite extensive knowledge about these species' habitat selection, the impact of landscape configuration metrics on a large geographical scale remains unknown. Most recent studies [16, 20, 21] did not address this issue, as they did not evaluate models with and without detailed information on the spatial structure of patches.

Importantly, in order to define more effective strategies for biodiversity conservation on a trans-national scale, it is essential to understand how different large-scale landscape aspects influence species density [8]. That is why conservationists pay a lot of attention to develop tools that enhance remote sensing data interpretation. Today, landscape metrics, such as the patch area, shape and connectivity are often used as surrogates of biodiversity [22–24]. This approach is obvious because landscape metrics can be relatively easily measured on a large scale for any place for which satellite or aerial images are available [25], but ecological sense of such biodiversity surrogates is still debatable [26]. The most severe criticism concerns the methodological approach where indicators are defined on the basis of a simple positive correlation between potential surrogates and target species, while the context of environmental preferences of species is often overlooked [27]. Bearing the above in mind, two kinds of models for species density were developed in this research: the first without landscape configuration metrics and the second with the core, buffer zone, and shape of patches as additional variables. By comparing both models, i.e. with and without landscape metrics, potential surrogates could be evaluated. So this analytical approach should indicate not only the relationship between environmental components and species diversity but also show whether landscape metrics reflect bird species density or not.

The aim of this study is to determine the extent to which land cover typologies, climate, topography and NDVI are able

to predict distribution of the four studied bird species as compared to patch structure metrics data.

2 Material and Methods

2.1 Bird Data

The Corn bunting, Whinchat, Tree pipit and Wood nuthatch density data were derived from the Common Breeding Birds Monitoring Scheme [28] and collected in Poland in years 2006–2013 in 876 1 km² grid cells (see Appendix A, Fig. S1). Survey plot squares had been chosen at random out of 311,664 1 km² squares covering the whole territory of Poland. In each breeding season each plot was surveyed twice. The first visit took place between 10 April and 15 May, and the second between 16 May and 30 June. The bird census within each square consisted of two parallel 1 km transects along either an east-west or north-south axis. Each transect was divided into five 200 m sections, in which birds were noted within three distance categories (< 25 m, 25–100 m, > 100 m). Birds were noted perpendicular to the transect line. Each survey started between the dawn and 9 am and lasted for about 90 min. Surveys were carried out by volunteers, but regrettably many squares were not regularly monitored. During the 8-year period, each square was inspected on average ($\pm$ SD) in 5.0 ± 2.3 breeding seasons.

2.2 Environmental Data

For modelling purposes, several environmental variables related to topography, climatic conditions, land cover, lights at night, vegetation indices and landscape metrics were used (Table 1, see also Appendix A). These data were converted into GRASS GIS file format [29] with the grid size of 1 km² and re-projected to the coordinate system EPSG4284 projection (<http://spatialreference.org/ref/epsg/4284/>).

As landscape structure metrics, the core, the buffer zone and the patch shape were obtained from the Corine Land Cover (CLC) and the Joint Research Centre (Table 1) using QGIS with LecoS library [30]. The core and the buffer zone are expressed as the percentage of the area in all grid cells (1 × 1 km). The buffer zone is defined as an area within 100 m from the outside edge of the patch, while the core shows patches excluding the 100 m buffer zone from the patch's edge. The patch shape index is expressed as a mean patch perimeter in each grid cell divided by the minimum perimeter possible for a maximal patch, i.e. the index equals 1 when the patch is square and increases without limit as the patch shape becomes more irregular.

Table 1 Environmental predictors used for the model design

Variable code	Description
Predictors extracted from PCA (see Table S2)	
PREC	had the highest loadings with precipitation variables
TEMP	had the highest loadings with temperature variables
FARMFOR	represents a habitat gradient from non-irrigated arable fields to three types of forest (coniferous, deciduous and mixed)
URBCFOR	represents the habitat gradient from urban areas to coniferous forest
PASDMFOR	represents the habitat gradient from pastures to deciduous and mixed forest
PASFMIX	represents the habitat gradient from pastures to mixed agricultural area interspersed with coniferous forest
DFORWEFM	represents the habitat gradient from deciduous forest to mixed agricultural area surrounded by wetlands and water bodies
SCRUWFOR	represents the habitat gradient scrub and fruit trees plantations to mixed forest localised in wet areas
Predictors extracted directly from remote sensing data	
LONGITUDE	geographic coordinate that specifies the east-west position
LATITUDE	geographic coordinate that specifies the south-north position
ALTITUDE	above sea level (a.s.l.)
DENIW	differences between maximal and minimal altitudes per grid cells
NDVIMAR	mean value of green vegetation level in March
NDVIAPR	mean value of green vegetation level in April
NDVIMAY	mean value of green vegetation level in May
NDVIJUN	mean value of green vegetation level in Jun
NDVICV	low value describes areas where the level of vegetation is similar in each year
Patch structure metrics	
The prefix: BUFFER indicates the buffer zone (defined as an area within 100 m from the outside edge of the patch), CORE - shows patches excluding the 100 m buffer zone from the patch's edge, SHPAE - patch shape index is expressed as a mean patch perimeter in each grid cell divided by the minimum perimeter possible for a maximal patch	
BUFFERNIAF	non-irrigated arable land
CORENIAF	
SHAPENIAF	
BUFFRFRUIT	fruit trees and berry plantations
COREFRUIT	
SHAPEFRUIT	
BUFFERPAST	pastures
COREPAST	
SHAPEPAST	
BUFFERMAF	mixed cultivation patterns (as a aggregation: complex cultivation patterns and land principally occupied by agriculture
COREMAF	
SHAPEMAF	
BUFFERMFOR	mixed forest
COREMFOR	
SHAPEMFOR	
BUFFERSCRUB	scrub (as a aggregation: moors and heathland and transitional woodland-shrub)
CORESCRUB	
SHAPESCRUB	

Table 1 (continued)

Variable code	Description
BUFFERWETL	inland wetlands (as a aggregation: inland marshes and peat bogs)
COREWETL	
SHAPEWETL	
BUFFERLAKES	water bodies
CORELAKES	
SHAPELAKES	
BUFFERDFOR	deciduous forest
COREDFOR	
SHAPEDFOR	
BUFFERCFOR	mixed forest
CORECFOR	
SHAPECFOR	

2.3 Data Processing and Analysis

The species density in grid cells was expressed as the Hayne estimator, calculated according to the equation below [6, 31–34]:

$$\hat{D}_H = \frac{n}{2L} \left(\frac{1}{n} \sum \frac{1}{r_i} \right)$$

where

H —Hayne's estimator of density, n —number of animals seen, L —length of transect, and r_i —sighting distance to each animal.

The estimator was necessary because the number of birds in each grid cells depended on the transect's length and the distance from the observer [35, 36].

In order to avoid multicollinearity among environmental variables, the principal components' analysis (PCA) was performed with the Varimax normalised rotation, separately for each of the two environmental datasets, i.e. climate and habitat [37]. Principal components' axes with eigenvalues > 1 were retained as predictor variables in the analyses.

The PCA of climate variables produced two axes, which explained 83.7% of the original variation in climate variables (see: Table 1 and Appendix A, Table S1A).

Habitat variables derived from the Corine Land Cover (CLC), lights at night representing urban area and the type of forest the Joint Research Centre produced seven components and explained 78.0% of the variation (see: Table 1 and Appendix A, Table S1B). The seventh component (PC7) had a low value of the explained variance and was not included in the analysis.

Pearson correlation coefficients were used to assess relationships between predictors which came from two separate PCA analyses, NDVI variable, geographical coordination, topography and patch structure metrics (Appendix A, Table S2).

The generalised additive model (GAM) was employed to fit resource selection functions [6], and the Hayne estimator of species' density was used as the response variable. Variables

extracted by PCAs, geographical variables (longitude, latitude, altitude and denivelation), NDVI from March to April, NDVICV and patch structure metrics (buffer zone, core and patch shape) were used as predictors. Two sets of GAMs were developed for each species: (1) models based on topographic, climatic, general land cover types and vegetation indices without patch structure metrics; (2) models based on both datasets, i.e. those from model 1 and patch structure metrics (buffer zone, core and patch shape). The most parsimonious models were selected using the Akaike information criterion (mgcv library in R; [38]) with the lowest AIC and, consequently, the highest Akaike weight [39]. All possible models were analysed (2^n , where n = number of variables), using MuMIn library in R [40, 41]. The probability of including a variable in the best parsimonious model was estimated as the relative importance (RI) by summing the Akaike weights of all candidate models in which the variable was included [39, 42]. As a measure of the best model, the evidence ratio was used [39]. In order to allow some complexity in the functions and avoid data over-fitting, the basic dimension was defined as $k = 4$ [43]. Then, the Gaussian distribution of errors and the identity link function were applied. As a measure of deviance reduction, D^2 coefficient equivalent to R^2 was used, which is well known from least squares estimation. Finally, on the basis of each species' best model predicted distributions were established. The correlation coefficient between the predicted versus the observed densities (log-transformed) was used as a measure of error prediction [35, 44].

2.4 Cross-validation

In order to increase the predictive power of models, an independent evaluation was carried out. Before developing GAMs, 20% of the observations that had not been used in the GAM procedure were randomly chosen. As a cross-validation test, we used the correlation coefficient between the predictive (from the best GAM model) and real density expressed as the Hayne estimator of species density (20% of observations were not included in model fitting). The higher the value of the correlation coefficient, the bigger the predictive power of the model. So, D^2 was estimated in the training dataset, and it measured the model's performance when predicting a population's density in a grid cell; while the "correlation coefficient" was estimated according to the test dataset, and it measured the model's performance when forecasting densities for the whole country.

3 Results

3.1 Population Size

Breeding populations of the Corn bunting, Whinchat, Tree pipit and Wood nuthatch were recorded respectively in 63.1,

70.5, 56.8 and 52.1% of grid cells. The mean density of the Corn bunting expressed as the Hayne estimator equalled 3.33 (95% CL, 3.09–3.58) individuals/1 km², while the mean Hayne estimator on occupied plots was 5.30 (95% CL, 5.01–5.58) individuals/1 km². The mean density of the second farmland bird species, the Whinchat, expressed also as the Hayne estimator amounted to 2.95 (95% CL, 2.75–3.15) individuals/1 km², while the mean Hayne estimator on occupied plots was 3.54 (95% CL, 3.30–3.78) individuals/1 km². In turn, the mean Hayne estimator of the Tree pipit density was 2.22 (95% CL, 2.02–2.41), while on occupied plots it was 3.90 (95% CL, 3.65–4.15). Finally, the mean Hayne estimator of the Wood nuthatch density amounted to 1.47 (95% CL, 1.34–1.59), while on occupied plots it was 2.82 (95% CL, 2.65–2.98).

3.2 Habitat Use by Corn Bunting and Predictive Map of $\hat{D}_{CB}$

Out of all analysed models, only four gained support using information-theoretic criteria, showing AIC weights > 0 (Appendix A, Table S3, model 2A). Model selection procedures allowed to identify 6 predictors with the Relative Importance RI > 0 , but only 4 were included in the best-supported model (Table 2). D^2 -coefficient of the best model was 35.7%, and it was definitely better for describing the variation of $\hat{D}_{CB}$ than the second model (evidence ratio = 1.85). The correlation coefficient between the predictive density based on the most parsimonious model and the real density based on the test dataset ($n = 175$) was 0.49, $p < 0.0001$, (Fig. 1a).

In turn, adding landscape configuration metrics as additional predictors to the modelling framework made ten GAMs for $\hat{D}_{CB}$ gain AIC weights > 0 (Appendix A, Table S3, model 2B). D^2 -coefficient of the best model was 68.6%, and it was better than the second model (evidence ratio = 3.52) in the candidate set. Model selection procedures resulted in 16 predictors with the Relative Importance RI > 0 , but only 9 were included in the best-supported model (Table 2). The correlation coefficient between the predictive density based on the best model and the real density based on the test dataset ($n = 175$) was 0.79 $p < 0.0001$ (Fig. 1b). Having used this model, a predictive map of $\hat{D}_{CB}$ was designed (Fig. 2a).

3.3 Habitat Use by Whinchat and Predictive Map of $\hat{D}_w$

Based on remotely sensed data without landscape metrics (the same dataset as in $\hat{D}_{CB}$) and the fit to the Hayne estimator of the Whinchat's density ($\hat{D}_w$), four GAMs showed AIC weights > 0 (Appendix A, Table S3, model 2C). Model selection procedures brought about 8 predictors with the Relative Importance RI > 0 , and all of them were included in the best-supported model (Table 2). D^2 -coefficient of the model was 33.2%, and it was slightly better for describing the variation of

Table 2 Predictors including in the best supported GAMs

Species	Without landscape configuration metrics				With landscape configuration metrics			
	Predictor	F	P	RI	Predictor	F	P	RI
Corn bunting	FARMFOR	11.85	< 0.0001	1	FARMFOR	11.19	< 0.0001	1
	LATITUDE	2.83	< 0.0003	1	BUFFERNIAF	2.25	< 0.0001	1
	TEMP	1.70	< 0.0001	1	CORENIAF	2.11	< 0.0001	1
	NDVICV	1.19	< 0.0001	0.622	BUFFERPAST	2.47	< 0.0001	1
					COREPAST	2.40	< 0.0001	1
					SHAPENIAF	3.12	< 0.0001	1
					NDVICV	1.92	< 0.01	0.609
					LATITUDE	1.89	< 0.007	0.529
					TEMP	1.19	< 0.001	0.490
Whinchat	PASDMFOR	8.95	< 0.0007	1	BUFFERMAF	2.90	< 0.0001	1
	PASTFMIX	2.39	< 0.0001	1	BUFFERPAST	2.54	< 0.0001	1
	DFORWEFM	5.13	< 0.0003	1	BUFFERNIAF	2.11	< 0.0001	1
	SCRUFOR	1.39	< 0.002	1	COREPAST	1.23	< 0.0001	1
	LONGITUDE	1.55	< 0.0001	0.949	CORENIAF	1.44	< 0.002	1
	LATITUDE	3.13	< 0.0001	0.949	SHAPEPAST	4.56	< 0.01	1
	NDVICV	1.01	< 0.04	0.634	SHAPEMAF	1.89	< 0.008	1
	NDVIAPR	1.52	< 0.01	0.633	PASDMFOR	7.15	< 0.003	0.940
					DFORWEFM	4.43	< 0.005	0.937
					LONGITUDE	1.22	< 0.001	0.931
Tree pipit	FARMFOR	7.33	< 0.0001	1	LATITUDE	4.47	< 0.001	0.911
	URBCFOR	4.57	< 0.0001	1	BUFFERCFOR	2.75	< 0.0001	1
	PASTFMIX	3.41	< 0.0001	0.961	BUFFERMFOR	2.59	< 0.0001	1
	NDVIMAR	1.93	< 0.02	0.919	BUFFERDFOR	2.79	< 0.0001	1
	NDVICV	1.01	< 0.03	0.552	URBCFOR	6.39	< 0.0001	1
					SHAPECFOR	1.59	< 0.01	1
					COREMFOR	1.56	< 0.02	1
					COREDFOR	1.01	< 0.01	1
					COREDFOR	1.98	< 0.0001	1
					COREMFOR	8.18	< 0.0001	1
Wood nuthatch	FARMFOR	4.17	< 0.0001	1	SHAPEMFOR	1.08	< 0.008	1
	PASTDMFOR	3.59	< 0.0003	1	BUFFERMFOR	1.11	< 0.01	1
	PASTFMIX	3.62	< 0.0001	1	NDVIJUN	1.96	< 0.02	0.971
	NDVIJUN	1.76	< 0.0001	0.993	FARMFOR	6.99	< 0.0001	0.685
	NDVICV	2.39	< 0.0001	0.984	PREC	4.53	< 0.001	0.685
	PREC	4.26	< 0.0001	0.841	NDVICV	3.38	< 0.0001	0.367

$\check{D}_w$ than the second model (evidence ratio = 1.006). The correlation coefficient between the predictive density based on the best supported model and the real density based on the test dataset ($n = 175$) was 0.48, $p < 0.0001$ (Fig. 1c).

In case of GAMs for $\check{D}_w$, where landscape configuration metrics were used as additional predictors, nine models showed AIC weights > 0 (Appendix A, Table S3, model 2D). Model selection procedures produced 13 predictors with the Relative Importance RI > 0 , but only 11 were included in the best-supported model (Table 2). D^2 -coefficient of the model was 78.4% and it was definitely better than the second

model in the candidate set (evidence ratio 37.95). The correlation coefficient between the predictive density based on the best supported model and the real density based on the test dataset ($n = 175$) was 0.84, $p < 0.0001$ (Fig. 1d). Having used this model, a predictive map of $\check{D}_w$ was created (Fig. 2b).

3.4 Habitat Use by Tree Pipit and Predictive Map of $\check{D}_{TP}$

Out of all GAMs for the Hayne estimator of the Tree pipit density ($\check{D}_{TP}$) without landscape metrics, only four models gained support using information-theoretic criteria (Appendix A, Table S3,

Fig. 1 Correlations between Hayne estimator of observed density and predictive density. **a** Model for Corn bunting density without landscape configuration metrics. **b** Model with core, buffer zone and shape of patches. **c** Model for Whinchat density without landscape configuration metrics. **d** Model with core, buffer zone and shape of patches. **e** Model for Tree pipit density without landscape configuration metrics. **f** Model with core, buffer zone and shape of patches. **g** Model for Wood nuthatch density without landscape configuration metrics. **h** Model with core, buffer zone and shape of patches

model 2E). Following model selection procedures, 6 predictors with the Relative Importance $RI > 0$ were identified, of which 5 were included in the best-supported model (Table 2). D^2 -coefficient of this model was 44.6%, and it was better for describing the variation of $\check{D}_{TP}$ than the second model (evidence ratio = 1.50). The correlation coefficient between the predictive density based on the best supported model and the real density based on the test dataset ($n = 175$) was 0.66, $p < 0.0001$, (Fig. 1e).

By adding landscape metrics as additional predictors to the GAMs, eight models gained AIC weights > 0 (Appendix A, Table S3, model 2F). D^2 -coefficient of the most parsimonious model was 51.2%, and it was slightly better than the second model (evidence ratio = 1.01) in the candidate set. Model selection procedures resulted in identifying 12 predictors with the Relative Importance $RI > 0$, but only 7 were included in the best-supported model (Table 2). The correlation coefficient between the predictive density based on the best

parsimonious model and the real density based on the test dataset ($n = 175$) was 0.80 $p < 0.0001$ (Fig. 1f). Having used this model, a predictive map of $\check{D}_{TW}$ was drawn (Fig. 2c).

3.5 Habitat Use by Wood Nuthatch and Predictive Map of $\check{D}_{WN}$

Four GAMs for the Hayne estimator of the Wood nuthatch density ($\check{D}_{WN}$) without landscape metrics showed AIC weights > 0 (Appendix A, Table S3, model 2G). Model selection procedures produced 7 predictors with the Relative Importance $RI > 0$, but only 6 were included in the best-supported model (Table 2). D^2 -coefficient of this model was 29.5%, and it was definitely better for describing the variation of $\check{D}_{WN}$ than the second model (evidence ratio = 5.88). The correlation coefficient between the predictive density based on the best parsimonious model and the real density based on the test dataset ($n = 175$) was 0.54 $p < 0.0001$ (Fig. 1g).

In case of GAMs for $\check{D}_{WN}$, where landscape metrics were added to the modelling framework, four models also demonstrated AIC weights > 0 (Appendix A, Table S3, model 2H). Model selection procedures identified 9 predictors with the Relative Importance $RI > 0$, and all of them were included in the best-supported model (Table 2). D^2 -coefficient of the best model was 76.2%, and it was better than the second model in the candidate set (evidence ratio = 1.13). The correlation coefficient between the predictive density based on the best supported model and the real density based on the test dataset ($n = 175$) was 0.84 $p < 0.0001$ (Fig. 1h). Having used this model, a predictive map of $\check{D}_{TW}$ was designed (Fig. 2d).

4 Discussion

4.1 GAMs for Bird Species Density With and Without Landscape Configuration Metrics

The study confirms a pattern that models with landscape configuration metrics have higher prediction power than models without them. This result is promising because it clearly shows the way in which geophysical factors combined with landscape metrics affect species density on a large geographical scale.

The first group of GAMs without landscape metrics showed that all analysed species responded positively to the area of preferred habitats. Thus, the result is not surprising since these preferences had been described before [16–18, 20, 45]. However, it should be remarked that using only an area of patches as a predictor may lead to a conclusion that high densities of species depend on habitats' participation, whereas the structure of a given patch is irrelevant. Obviously, in some situations, species occurrence depends only on the area of patches (e.g. White stork *Ciconia ciconia*, [46]), but it is not

Fig. 2 Predictive maps of the Hayne Estimator of: **a** the Corn bunting density, **b** Whinchat density, **c** Tree pipit density and **d** Wood nuthatch density

always the case [47, 48]. A study conducted on a small spatial scale suggested that habitat specialist species with narrow niches were more likely to be found in the core of patches with a simple shape, while generalist species tended to occur in the buffer zone of a more complicated shape [49, 50]. Therefore, SDMs that use only the area of patches may not provide complete information on species' specific habitat preferences [7, 24, 33]. Indeed, this study confirmed and generalised the pattern on a large spatial scale, because when landscape configuration metrics were used as additional predictors, models turned out to be definitely better than models without them. Consequently, the distribution of farmland and forest bird species was best

described by a combination of features measured by land-use typologies, climate, topography, NDVI, core, buffer zone and patch shape, and the structure of patches played a great role in predicting distributions of the four species. The Corn bunting is a specialist of intensively used farmland where it prefers large areas of monoculture crops, but—what is important in this context—it favours a simple shape of patches and definitely avoids the buffer zone. A similar pattern is also followed by the Wood nuthatch which is a specialist of large areas of deciduous and mixed forests. It chooses the core and, like the Corn bunting, avoids patches of a complicated shape. A contrary preference is displayed by the Whinchat and the Tree pipit, which are

generalists, and more frequently observed in the buffer zone of a mixed arable land and coniferous forest, respectively, but at the same time both prefer more complicated patch shapes.

The results also extend a pattern in which a higher spectral index, here NDVI, corresponds to higher species density [51]. It shows that an inter-seasonal variation of green vegetation also plays an important role. A low level of green vegetation variation is featured by both coniferous forest (except extreme situations green vegetation is constant throughout a year, [33]) and intensively used farmland wherein repeatability of green vegetation conditions is provided by multi-annual fertilisation [46]. On the other hand, a high level of green vegetation variation is characteristic to pastures, especially located near water bodies, and deciduous forest where NDVI depends on floods and the rate of leaf development in a given year. So the level of inter-seasonal green vegetation variation can be considered as an indicator of a habitat's condition. Nonetheless, a more specific study should be carried out in this context.

The observed geographical gradient cannot be explained by the climate, distribution of habitat, patch structure metrics and the NDVI, as these factors are controlled by their inclusion in the best model. However, it can be presumed that other aspects of habitat and climate not measured in these models are likely to explain the pattern. For instance, the most important predictor for the Meadow pipit's *Anthus pratensis* density is the distribution of habitat-specific plant species while a high abundance of the Crested lark *Galerida cristata* is affected by pigs' density as a measure of the level of farming [21, 32]. Furthermore, other study also indicated that coexisting species [7] as well as a quantified co-evolutionary process of a bird parasite (i.e. the cuckoo, [14]) constitute one of the most important predictors of functional species richness.

4.2 Modelling Approach

The Generalised Additive Model method used in this study is often used as a tool for predictive spatial distribution modelling [6, 32, 33, 35, 41]. However, despite the advantages of GAMs described by [41], this procedure—being a tool for creating a predictive spatial density pattern—requires evaluation based on an independent dataset [52]. Before developing the models, 20% of observations were chosen randomly and treated as an independent dataset. The correlation coefficient between real densities and densities estimated according to results from the GAM procedure safeguarded this independent assessment. This simple and rarely chosen procedure intuitively indicated the predictive model's ability without going into complicated algorithms [53].

It should also be mentioned that the necessity of using independent variables (no spatial correlation) as predictors may limit the use of landscape configuration metrics as explanatory variables [54]. It was discovered that a relatively high correlation coefficient between altitude and the shape

of all patches seemed to prove that such a relationship may stem from geomorphology, where slope or roughness affected a more complicated patch shape. According to a non-stationary phenomenon if Poland's territory was divided into smaller geographic sub-regions based on topographic zones, it could turn out that in some places the shape of patches has a high correlation with elevation, especially in highlands, while in others it does not (in lowlands). Therefore, it can be suspected that on a large geographical scale the observed patterns in the spread of populations result from stochastic and deterministic relationships between the habitat, geometry of patches, climate and topography.

5 Conclusion

Our study that follows a new approach to ecological modelling [55] shows that the abundance of all studied species overlaps with the distribution of preferred habitats but mainly with patch structure metrics, meaning that these variables are among key factors influencing species densities (all RI = 1). The findings constitute an important contribution to applied ecology and the ecological modelling approach by showing that the predictive power of several bird species depends on an ecological complexity of systems, including the interplay between many environmental components. Importantly, this approach revealed new surrogates for specialist and generalist bird species, so on a large spatial scale geometry of patches can be considered as one of the most important environmental components reflecting bird density.

Acknowledgements I wish to express my gratitude to the numerous observers who collected data in the field. The full list of their names can be found at <http://main5.amu.edu.pl/~kubako/tsr/>. I would like to thank Przemysław Chylarecki for the access to Birds' Database and Justyna Grześkowiak for her linguistic assistance.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Franklin, J. (2010). *Mapping species distributions. Spatial inference and prediction*. Cambridge: Cambridge University Press.
2. Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modeling*, 135, 147–186.
3. Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677–697.
4. Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I., Regan, T. J., Brotons, L., McDonald-

- Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M. R., Possingham, H. P., & Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, 16, 1424–1435.
5. Bradley, B. A. (2016). Predicting abundance with presence-only models. *Landscape Ecology*, 31, 19–30.
6. Kosicki, J. Z., Zduniak, P., Ostrowska, M., & Hromada, M. (2016). Are predators negative or positive predictors of farmland bird species community on a large geographical scale? *Ecological Indicators*, 62, 259–270.
7. Morelli, F., & Tryjanowski, P. (2014). Associations between species can influence the goodness of fit of species distribution models: the case of two passerine birds. *Ecological Complexity*, 20, 208–212.
8. Tryjanowski, P., Hartel, T., Bldi, A., Szymański, P., Tobolka, M., Herzon, I., Goławski, A., Konvička, M., Hromada, M., Jerzak, L., Kujawa, K., Lenda, M., Orłowski, G., Panek, M., Skórka, P., Sparks, T. H., Tworek, S., Wuczyński, A., & Żmihorski, M. (2011). Conservation of farmland birds faces different challenges in Western and Central-Eastern Europe. *Acta Ornithologica*, 46, 1–12.
9. Brown, J. L., Collopy, M. W., & Smallwood, J. A. (2014). Habitat fragmentation reduces occupancy of nest boxes by an open-country raptor. *Bird Conservation International*, 24, 364–378.
10. Song, W., & Kim, E. (2015). Landscape factors affecting the distribution of the great tit in fragmented urban forests of Seoul, South Korea. *Landscape and Ecological Engineering*. doi:10.1007/s11355-015-0280-4.
11. Zapponi, L., Luiselli, L., Cento, M., Catorci, A., & Bologna, M. A. (2014). Disentangling patch and landscape constraints of nested assemblages. *Basic and Applied Ecology*, 15, 712–719.
12. Angelieri, C. C. S., Adams-Hosking, C., Paschoaletto, K. M., De Barros Ferraz, M., De Souza, M. P., & McAlpine, C. A. (2016). Using species distribution models to predict potential landscape restoration effects on puma conservation. *PloS One*, 11, e0145232.
13. Duerr, A. E., Miller, T. A., Cornell Duerr, K., Lanzone, M. J., Fesnock, A., & Katzner, T. E. (2015). Landscape-scale distribution and density of raptor populations wintering in anthropogenic-dominated desert landscapes. *Biodiversity and Conservation*, 24, 2365–2381.
14. Tryjanowski, P., & Morelli, F. (2015). Presence of cuckoo reliably indicates high bird diversity: a case study in a farmland area. *Ecological Indicators*, 55, 52–58.
15. Hagemajier, W. J. M., & Blair, M. J. (1997). *The EBCC atlas of European breeding birds: Their distribution and abundance*. London: T & AD Poyser.
16. Altwischer, A., Buschewski, U., Ehrke, C., Fröhlich, J., Gärtner, A., Giese, P., Günter, F., Heitmann, N., Hestermann, M., Hoffmann, H., Kleinschmidt, F., Kniepkamp, B., Linke, W., Mayland-Quellhorst, T., Pape, J., Peterson, T., Schendel, V., Schwieger, S., Wadenstorfer, A., & Fischer, K. (2015). Habitat preferences of male corn buntings *emberiza calandra* in North-Eastern Germany. *Acta Ornithologica*, 50, 1–10.
17. Morelli, F., & Girardello, (2014). Buntings (Emberizidae) as indicators of HNV of farmlands: a case of study in Central Italy. *Ethology Ecology and Evolution*, 26, 405–412.
18. Fischer, K., Busch, R., Fahl, G., Kunz, M., & Knopf, M. (2013). Habitat preferences and breeding success of Whinchats (*Saxicola rubetra*) in the Westerwald mountain range. *Journal of Ornithology*, 154, 339–349.
19. Burgess, M. D., Bellamy, P. E., Gillings, S., Noble, D. G., Grice, P. V., & Conway, G. J. (2015). The impact of changing habitat availability on population trends of woodland birds associated with early successional plantation woodland. *Bird Study*, 62, 39–55.
20. Alderman, J., McCollin, D., Hinsley, S. A., Bellamy, P. E., Picton, P., & Crockett, R. (2005). Modelling the effects of dispersal and landscape configuration on population distribution and viability in fragmented habitat. *Landscape Ecology*, 20, 857–870.
21. Kuczyński, L., & Chylarecki, P. (2012). *Atlas of common breeding birds in Poland: distribution, habitat preferences and population trends*. Warszawa: GIOŚ (in Polish).
22. Faith, D. P. (2003). Environmental diversity (ED) as surrogate information for species-level biodiversity. *Ecography*, 26, 374–379.
23. Hortal, J., Araújo, M. B., & Lobo, J. M. (2009). Testing the effectiveness of discrete and continuous environmental diversity as a surrogate for species diversity. *Ecological Indicators*, 9, 138–149.
24. Morelli, F., Pruscini, F., Santolini, R., Perna, P., Benedetti, Y., & Sisti, D. (2013). Landscape heterogeneity metrics as indicators of bird diversity: determining the optimal spatial scales in different landscapes. *Ecological Indicators*, 34, 372–379.
25. Banks-Leite, C., Ewers, R. M., Kapos, V., Martensen, A. C., & Metzger, J. P. (2011). Comparing species and measures of landscape structure as indicators of conservation importance. *Journal of Applied Ecology*, 48, 706–714.
26. Lindenmayer, D. B., Manning, A. D., Smith, P. L., Possingham, H. P., Fischer, J., Oliver, I., & McCarthy, M. A. (2002). The focal-species approach and landscape restoration: a critique. *Conservation Biology*, 16, 338–345.
27. Carroll, C., Noss, R. F., & Paquet, P. C. (2001). Carnivores as focal species for conservation planning in the Rocky Mountain region. *Ecological Applications*, 11, 961–980.
28. Chylarecki, P., & Jawińska, D. (2007). *Monitoring Pospolitych Ptaków Lęgowych. Raport z lat 2005–2006*. Warszawa: OTOP (in Polish).
29. Neteler, M., & Mitasova, H. (2008). *Open source GIS: A GRASS GIS approach* (3rd ed.). New York: Springer.
30. Jung, M. (2016). LecoS—a python plugin for automated landscape ecology analysis. *Ecological Informatics*, 31, 18–21.
31. Hayne, D. W. (1949). An examination of the strip census method for estimation animal populations. *Journal of Wildlife Management*, 13, 145–157.
32. Kosicki, J. Z., & Chylarecki, P. (2013). Predictive mapping of meadow pipit density using integrated remote sensing data with atlas of vascular plants dataset. *Bird Study*, 60, 500–508.
33. Kosicki, J. Z., Stachura, K., Ostrowska, M., & Rybska, E. (2015). Complex species distribution models of Goldcrests and Firecrests densities in Poland: Are remote sensing-based predictors sufficient? *Ecological Research*, 30, 525–638.
34. Krebs, C. J. (1999). *Ecological methodology*. Boston: Addison Wesley, Longman.
35. Kosicki J.Z., & Chylarecki P. (2014). The Hooded Crow *Corvus cornix* density as a predictor of wetland bird species richness on large geographical scale in Poland. *Ecological Indicators*, 38, 50–60.
36. Pesenti, E., & Zimmermann, F. (2013). Density estimations of the Eurasian lynx (*Lynx lynx*) in the Swiss Alps. *Journal of Mammalogy*, 94, 73–81.
37. Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge: Cambridge University Press.
38. Wood, S. (2016). mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation. R package version 1.7–22. <http://cran.rproject.org/web/packages/mgcv>.
39. Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach* (2nd ed.). Berlin: Springer-Verlag.
40. Bartoń, K. (2016). MuMIn: multi-model inference. R package version 1.9.0. <http://CRAN.R-project.org/package=MuMIn>.
41. Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. London: Chapman and Hall.

42. Reino, L., Portod, M., Morgado, R., Carvalhof, F., Miraf, A., & Bejac, P. (2010). Does afforestation increase bird nest predation risk in surrounding farmland? *Forest Ecology and Management*, 260, 1359–1366.
43. Santana, J., Porto, M., Gordinho, L., Reino, L., & Beja, P. (2012). Long-term responses of Mediterranean birds to forest fuel management. *Journal of Applied Ecology*, 49, 632–643.
44. Hastie, T., Tibshirani, R. J., & Friedman, J. (2008). *The elements of statistical learning* (2nd ed.). New York: Springer.
45. Cramp, S. (1998). *The complete birds of western palearctic on CD-ROM*. Oxford: Oxford Univeristy Press.
46. Tryjanowski, P., Jerzak, L., & Radkiewicz, J. (2005). Effect of water level and livestock on the productivity and numbers of breeding White Storks. *Waterbirds*, 28, 378–382.
47. Forman, R. T. T. Ž., & Godron, M. (1986). *Landscape ecology*. New York: Wiley.
48. Wiens, J. A. (1997). Metapopulation dynamics and landscape ecology. In I. A. Hanski & M. E. Gilpin (Eds.), *Metapopulation biology: ecology, genetics, and evolution* (pp. 43–67). San Diego: Academic.
49. Clavel, J., Julliard, R., & Devictor, V. (2011). Worldwide decline of specialist species: toward a global functional homogenization? *Frontiers in Ecology*, 9, 222–228.
50. Devictor, V., Julliard, R., & Jiguet, F. (2008). Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. *Oikos*, 117, 507–514.
51. Sicurella, B., Musitelli, F., Rubolini, D., Saino, N., & Ambrosini, R. (2016). Environmental conditions at arrival to the wintering grounds and during spring migration affect population dynamics of barn swallows *Hirundo rustica* breeding in Northern Italy. *Population Ecology*, 58, 135–145.
52. Smith, A. B., Santos, M. J., Koo, M. S., Rowe, K. M., Rowe, K. C., Patton, J. L., Perrine, D. J., Beissinger, R. S., & Moritz, C. (2013). Evaluation of species distribution models by resampling of sites surveyed a century ago by Joseph Grinnell. *Ecography*, 36, 1017–1031.
53. Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D., & Willis, S. G. (2014). Improving species distribution models: the value of data on abundance. *Methods in Ecology and Evolution*, 5, 506–513.
54. Guisán, A., & Theurillat, J. P. (2000). Equilibrium modeling of alpine plant distribution: how far can we go? *Phytocoenologia*, 30, 353–384.
55. Morelli, F., Møller, A. P., Nelson, E., Benedetti, Y., Tichit, M., Šimová, P., Jerzak, L., Moretti, M., & Tryjanowski, P. (2017). Cuckoo as indicator of high functional diversity of bird communities: a new paradigm for biodiversity surrogacy. *Ecological Indicators*, 72, 565–573.