



Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness

Jakub Z. Kosicki¹

Received: 23 May 2019 / Revised: 20 March 2020 / Published online: 29 April 2020
© The Author(s) 2020

Abstract

Species distribution modelling (SDM) is a family of statistical methods where species occurrence/density/richness are combined with environmental predictors to create predictive spatial models of species distribution. However, it often turns out that due to complex multi-level interactions between predictors and the response function, different types of models can detect different numbers of important predictors and also vary in their predictive ability. This is why we decided to explore differences in the predictive power of two most common methods, such as the Generalised Additive Model (GAM) and the Random Forest (RF) on the example of the Great Spotted Woodpecker *Dendrocopos major* and the Great Grey Shrike *Lanius excubitor*, as well as on the taxonomic and functional species richness. For each of the two bird species' densities and for two measurements of biodiversity, two sets of SDMs were generated: One based on the GAM, and the other on the RF. According to the out-of-bag, the Akaike Information Criterion (AIC) and an independent evaluation, we demonstrated that the GAM is the best method for predicting density of the Great Spotted Woodpecker and taxonomic species richness, whereas the RF has the lowest prediction error for the density of the Great Grey Shrike and functional species richness. It also becomes apparent that the GAM is responsive to taxonomic species richness and species with broad tolerance to environmental factors, i.e. the Great Spotted Woodpecker, while the RF detects more subtle relationships between density and environmental variables, rendering it more suitable for functional species richness and species with a narrow tolerance range to habitats factors, i.e. the Great Grey Shrike. Thus, effective predictive modelling of animal distribution requires considering several different analytical approaches to produce biologically realistic predictions.

Handling Editor: Bryan F. J. Manly.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10651-020-00445-5>) contains supplementary material, which is available to authorized users.

Extended author information available on the last page of the article

Keywords Niche modelling · SDM · Species density · Species Distribution Modelling · Species richness

1 Introduction

Predicting species distributions on the basis of species distribution modelling (SDM), often referred to as “predictive mapping”, has become a tool which is widely used not only in macroecological studies, research on niche evolution, interactions and co-evolution of species, but also in planning conservation areas (Warren et al. 2008; Esselman and Allan 2011; Drew et al. 2011; Fourcade et al. 2017). So, it comes as no surprise that the methodological aspect of developing models has been much discussed in the latest scientific literature (Fourcade et al. 2017). From the analytical point of view, the SDM is a group of statistical tools based on ecological niches theory, which describe non-linear relationships between species occurrence/density/richness and environmental layers in order to build statistical models which reflect processes that drive species distribution on a large geographical scale (Guisan and Zimmermann 2000; Elith and Leathwick 2009; Franklin 2010). Such models, projected in space or time, can predict not only species occurrence and density, range shifts under climate and habitat changes (Hijmans and Graham 2006), but also evaluate biodiversity surrogates and estimate invasive species distribution (Jimenez-Valverde et al. 2011). Thus, if the SDM has become a standard tool in macroecological studies (Guisan et al. 2013), it is vital to ensure the uniqueness of the process of machine learning methods both for species occurrence and other population and community parameters.

In recent years the number and complexity of statistical tools used to predict species distribution has grown exponentially (Hegel et al. 2010; Lobo et al. 2010). However, having analysed research results obtained with these tools, one comes to a conclusion that when varied analytical methods are not included in the SDM predictive process, the ecological value of results might be limited (Elith and Graham 2009; Elith et al. 2006; Segurado and Araújo 2004; Guisán and Theurillat 2000; Mattsson et al. 2013). For example, predictive occurrence of the European grayling (*Thymallus thymallus* L.) was provided by a number of different predictors coming from different models (Fukuda et al. 2013).

Therefore, we decided to compare the performance of two methods, i.e. the Generalised Additive Model (GAM) and the Random Forest (RF) to predict density of two bird species as well as two measures of bird biodiversity. From a wide spectrum of SDM methods, the GAM and the RF are considered to be among the most powerful machine learning algorithms used to predict species distribution. However, their effectiveness in projecting different levels of ecological systems, such as density and species richness, has not been fully confirmed (Araújo and New 2007; Hardy et al. 2011; Heikkinen et al. 2012; Elith et al. 2006; Wisz et al. 2008; Williams et al. 2009; Lei et al. 2011). Both algorithms can handle a large number of predictors and identify which of them are important via the out-of-bag procedure for the RF and the Relative Importance (RI) for the GAM (Breiman 2001; Hastie and Tibshirani 1990; Hastie et al. 2008). However, the output of both models can vary, because the RF creates a lot of specific models based on randomly perturbed data (Grimm et al. 2008; Lawler

et al. 2006; Mutanga et al. 2012; Prasad et al. 2006; Virkkala et al. 2010), whereas the GAM constructs one “best model” (Breiman 2001, Hastie and Tibshirani 1990; Hastie et al. 2008). So, both models can detect the influence of the same predictor, but with a varying power, which in turn may suggest that relationships between predictors and species distribution may partly reflect a spatial structure of predictors, especially climate and topography, rather than a real ecological process (Bahn and McGill 2007; Beale et al. 2008; Chapman 2010). Furthermore, in the ecological modelling approach the problem with comparing models with one another increases along with spatial complexity (Hof et al. 2012; Carrasco et al. 2014; Seppelt and Voinov 2002; Kosicki 2018). When a lot of predictors affect species density and/or species richness in different ways (positive/negative and/or linear/non-linear), and models’ output is transferred to new areas, more rigorous tests are indispensable. For that reason, an independent test dataset should be used to evaluate the quality of the GAM and the RF so as to obtain more realistic models (Fourcade 2016). Therefore, the proposed approach should indicate whether the GAM or the RF can or cannot reveal environmental drivers underlying species distribution (Petitpierre et al. 2017). Otherwise speaking, the aim of this study is to test a hypothesis that the choice of machine learning methods is crucial for ensuring the high quality of SDMs.

Consequently, spatial distribution models of farmland (Great Grey Shrike) and forest (Great Spotted Woodpecker) breeding bird species’ densities were developed along with an analysis of spatial distributions of taxonomic (expressed as the number of bird species) and functional species richness (expressed as species occurrence linked with the ecosystem’s functioning and environmental components (Hillebrand and Matthiessen 2009; Cernansky 2017)). According to a small spatial scale study, the Great Grey Shrike prefers large areas of heterogeneous farmland, where arable fields are interspersed with meadows and small forests (Antczak et al. 2004), while the Great Spotted Woodpecker favours areas dominated by large areas of all forest types (Kosiński et al. 2006, 2018). Taxonomic species richness is associated with a habitat’s arrangement, however, functional species richness is often linked with an interplay between habitat, topography and environmental conditions (Kosicki and Hromada 2018). By using two different modelling approaches, we determine whether a given procedure produces variables which are incidentally or indirectly linked to species distribution on a large spatial scale (Dormann et al. 2012) or whether the selection of relevant predictors is compatible with the species’ ecology (Petitpierre et al. 2017). Thus, a pragmatic question arises: What type of models should be used to predict density and species richness? In other words, this study aims to determine the extent to which GAM or RF models produce similar spatial predictions.

2 Materials and methods

2.1 Bird data

Bird density and the number of bird species were derived from the Common Breeding Birds Monitoring Scheme (Chylarecki and Jawińska 2007). This database is based on

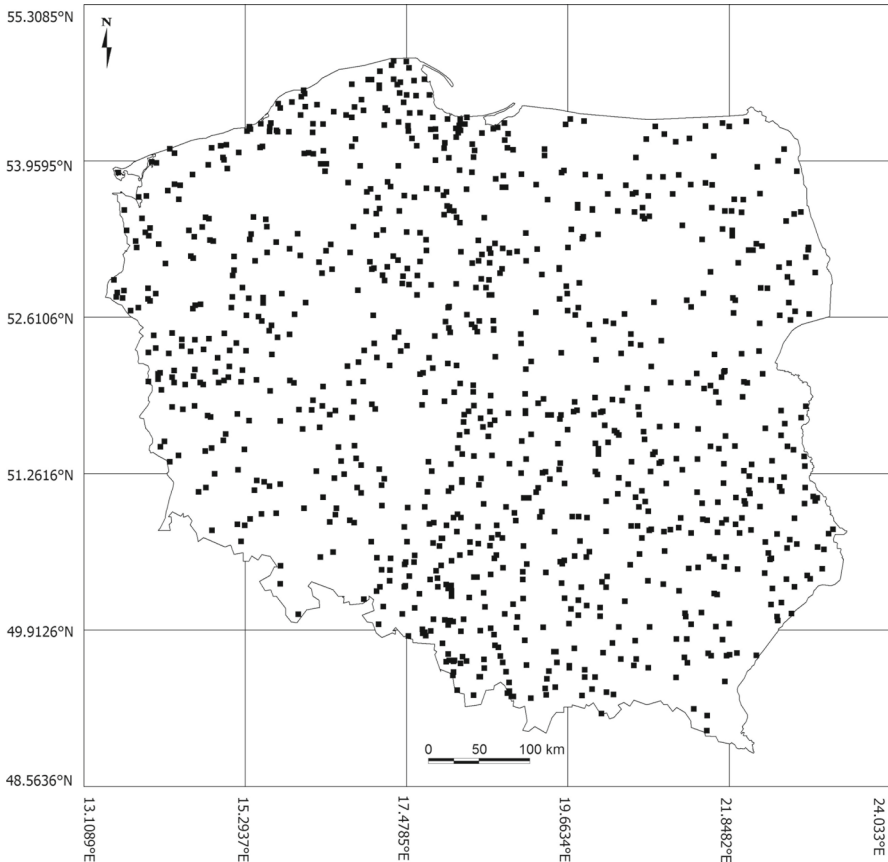


Fig. 1 Distribution of census grid cells according to the Common Breeding Birds Monitoring Scheme

field data collected in Poland in years 2000–2013 in 970 1 km² grid cells, which were randomly chosen out of 312,679 km² squares covering the whole of Poland (Fig. 1).

In each breeding season each grid cell was surveyed twice: between 10 April and 15 May, and between 16 May and 30 June. Each survey started between the dawn and 9 a.m. and lasted about 90 min. Ornithologists-volunteers noted birds perpendicular to two parallel 1-km transects along an east–west or north–south axis, and recorded them in three distance categories (<25 m, 25–100 m, >100 m). During the fourteen-year span each square was inspected on average 6.5 times (SD = 3.9). As the number of individuals in a given grid cell we used the highest recorded number of individuals during either of the two inspections in each season. Then, these values were averaged out for all the years of study.

2.2 Environmental data source

For modelling purposes we used various environmental predictors related to topography, climatic conditions, land cover, vegetation indices, forest stand structure and landscape configuration metrics (see details: Appendix A). All these data came from different GIS databases, and all of them were converted into the GRASS GIS file format (GRASS Development Team 2015) with the grid size of 1 km² (corresponding to bird survey plots) and re-projected to the co-ordinate system EPSG4284 projection (<http://spatialreference.org/ref/epsg/4284/>).

2.3 Data processing and analysis

Bird density in grid cells depended on the transect's length and the distance from the observer (Krebs 1999). Therefore, the Hayne estimator of bird density was used according to the equation below (Hayne 1949; Krebs 1999):

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

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

The measure of avian biodiversity comprised (1) Bird species richness (BSR), expressed as the number of bird species noted in a given grid cell, and (2) Functional richness (FRic), representing the amount of functional space occupied by species in grid cells. A low value of a component shows a low number of species whose niches (functional space) do not overlap, while a high value indicates a high number of species within a given functional space whose niches overlap considerably due to the fact that particular species can use the same resources in the same way without limiting one another. Functional diversity was calculated with avian niche traits (Pearman et al. 2014) based on 52 variables that described the niche of each bird species, i.e. food types (13 variables), behaviour while acquiring food (9 variables), substrate from which food was taken (9 variables), period of the day when a species foraged actively (3 variables), and nesting habitats (18 variables). All variables were binomial 0 or 1 (Pearman et al. 2014). The calculation was performed using FD library for R (Laliberté et al. 2015).

In order to avoid multicollinearity among particular predictors, the Principal Components Analysis (PCA) was employed with the Varimax normalised rotation for two (climatic and habitat) environmental datasets (Dormann et al. 2012) (Appendix B, Table S1, S2). We used two separate PCAs for two data types to enhance the interpretation of new loadings. The PCA of climate variables produced two axes (Appendix B, Table S1) which explained 83.7% of the original variation in climate variables. Habitat variables produced eleven components (Appendix B, Table S2) and explained 85.4% of the variation.

The first axis (CFIELD-CCFOR) represents a gradient from the core of non-irrigated arable land to the core of coniferous forest. The second axis (MFIELD-CCFOR) is related to a gradient from mixed arable fields and the buffer zone of

non-irrigated arable land to the core of coniferous forest. The third axis (NOVEGETAT) has the highest loadings on areas without vegetation. The fourth axis (MEAD-CFIELD) corresponds to a gradient from meadows to the core of non-irrigated arable land. The fifth axis (CDFOR-CFIELD) shows a gradient from the core of deciduous forest to the core of non-irrigated arable land. The low value of the sixth component (FRUIT) points to areas with fruit trees, while the low value of the seventh component (PEAT) reflects peats. Then, the low value of the eighth component (MARSH) reflects marshland. The ninth component (CFIELD-BBUSHE) relates to the core of non-irrigated arable land with a low value, while the high one shows the core of bushes. The tenth axis (MFOR-CBFIELD) is a gradient from mixed forest to the core of non-irrigated arable land. Finally, the low value of the eleventh axis (URB-BCFOR) shows urban areas, while a high value reflects the buffer zone of coniferous forest.

Furthermore, Pearson correlation coefficients and the Variance Inflation Factor (VIF) were applied to assess relationships between predictors which came from two separate PCA analyses, as well as other predictors not included in the PCA, i.e. topography measurements, Normalized Difference Vegetation Index (NDVI) and the shape of patches (Dormann et al. 2012) (Appendix B, Table S3). In the present study the VIF ranged from 1.1 to 22.7, so following Naimi et al. (2014) predictors with $VIF \geq 10$ were excluded from the analysis.

2.4 Generalised Additive Models (GAM)

The GAM (Hastie and Tibshirani 1990), (mgcv library in R; Wood 2013; R Development Core Team 2017) is an extension of the Generalised Linear Modelling (GLM) and provides a flexible non-linear relationship between the response variable and the predictor (Austin and Meyers 1996; Guisan et al. 2002). Response variables include (1) the Hayne estimator of the Great Grey Shrike, (2) the Hayne estimator of the Great Spotted Woodpecker, (3) bird species richness, and (4) functional species richness. The most parsimonious model was selected with the Akaike information criterion with the lowest AIC and consequently the highest Akaike weight (Burnham and Anderson 2002). All possible models using the ‘dredge’ procedure, i.e. 2^n where n = number of variables, were analysed with the MuMIn library in R (Bartoń 2013; Hastie and Tibshirani 1990).

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 (Burnham and Anderson 2002; Reino et al. 2010). Moreover, the Gamma distribution for modelling density was used, while for species richness we used the Poisson distribution. To allow some complexity in the functions, while avoiding data over-fitting, the basic dimension was defined as $k = 4$ (Santana et al. 2012). We used R^2 , expressed as percentage of the explained variance which is a quality measure of the model’s fit to the training data (Weisberg 1980). Finally, as the measure for the best model we applied: (1) evidence ratio (ER), defined as Akaike ω_1/ω_2 (Burnham and Anderson 2002), and (2) minimised generalised cross-validation (GCV) score which showed measure smoothness. Smaller values of GCV indicated better models fitting (Wood 2006).

2.5 Random Forest approach

The RF (Breiman 2001; RandomForest library in R Liaw and Wiener 2002; R Development Core Team 2017) was developed in a regression setting, in which—similarly to the GAM—the Hayne estimator of two species density and two avian biodiversity measures were treated as the model's response. This method is a machine learning technique (Breiman 2001; Elith et al. 2008) linked to classification/regression trees (CART) and bagging methods (Breiman et al. 1984; Breiman 1996; Breiman 2001). Thus, instead of constructing a single tree (CART), it constructs multiple trees formed on random samples of cases by bootstrapping technique, i.e. sampling with replacement, and each split (in each tree) from the random sample of predictors. This procedure ensures that a 1/3 of the observations is left out in the fitting process and used to calculate the prediction error, i.e. out-of-bag (OOB). We examined also the RF model's performance as Root Mean Square Errors (RMSE) and calculated the percentage of the explained variance (R^2).

The RMSE and R^2 measured the model's performance when predicting population density and/or measurement of species richness in the studied grid cells. Both measures were estimated on the basis of the training dataset. On the other hand, the OOB measured the performance of predicted densities for the whole country, and it was estimated according to the test dataset. The RF procedure also provided (1) a ranking of predictors' importance based on specified changes of mean decrease in accuracy (MSE_{OOB}), i.e. predictors were removed one by one from the model (Berk 2008; Breiman 2001), thus revealing their adequacy; and (2) partial dependence plots which showed how each predictor was related to the response variable when other predictors were held constant (Berk 2008). Then, a selection of variables (Guyon and Elisseeff 2003) was performed with the VSURF library for R (Robin et al. 2010), which made it possible to identify the smallest number of predictors that offered the best predictive power for the final model (Kohavi and John 1997). The search function commenced with all environmental variables ($n = 27$), and then progressively eliminated the least promising ones. Finally, a subset of environmental parameters with the lowest RMSE was selected (Kosicki 2017).

The RF regression required defining three parameters, such as *ntree* (number of trees), *mtry*, and the *nodesize*. The lowest RMSE was for *ntree* = 1000, *mtry* = 1/3 of variables (default value), and the *nodesize* = 5 (default value) (Vincenzi et al. 2011).

2.6 Generalised Additive Models vs. Random Forest

In order to compare two different results coming from the two methods, an independent dataset was used. Before developing GAM and RF models, 20% of the observations that had not been included in the modelling procedure were randomly chosen. For model validation, we used a correlation coefficient between predictive (from the best GAM and/or RF model) and real species densities and richness measurements (20% of observations were not included in the model fitting). Thus, the higher the value of the correlation coefficient, the higher the predictive power of the model.

Table 1 Comparison of GAM and RF predictive effectiveness

Gatunek/species richness	GAM				RF			
	ER	GCV	R ²	r (n = 175)	RMSE	OOB	R ²	r
GGS	33.3	1.03	0.23	0.30	0.24	0.44	0.69	0.95
GSW	2.12	0.97	0.55	0.94	0.53	0.36	0.27	0.59
TSR	998	0.93	0.24	0.97	0.77	0.70	0.25	0.53
FRic	16.3	27.0	0.23	0.52	0.56	0.30	0.77	0.97

The better model given in bold

ER evidence ratio, GCV general cross-validation, R² explained variance by independent variables, r correlation coefficient

2.7 Spatial autocorrelation

Residuals coming from best GAM and RF models were checked for their spatial autocorrelation by using the Moran's I statistics (Library: ape for R, Paradis 2016). The Moran's index ranged from -1 (negative spatial autocorrelation) to 1 (positive spatial autocorrelation), with non-significant values close to zero.

3 Results

3.1 Population size

The mean number of the Great Grey Shrike (number of birds in each grid cell) was 1.62 individuals/km² (95% CL 1.54–1.70, n = 970), while the mean Hayne estimator of density (\check{D}_{GGS}) was 0.36 individuals/km² (95% CL 0.29–0.43). On the other hand, the mean number of the Great Spotted Woodpecker was 0.54 individuals/km² (95% CL 0.49–0.59, n = 970), while the mean Hayne estimator of its density (\check{D}_{GSW}) was 1.66 individuals/km² (95% CL 1.60–1.72). Finally, the mean value of taxonomic species richness (BSR) was 58.40 (95% CL 57.26–59.54), while functional species richness (FRic) amounted to 0.12 (95% CL 0.11–0.12).

3.2 Models for the Great Grey Shrike (\check{D}_{GGS})

Out of all analysed Generalised Additive Models (GAM), only five gained support using information-theoretic criteria, showing AIC weights >0 (Appendix A, Table S4, model GGS2). The most parsimonious model (Table 1) included 7 environmental variables with the Relative Importance RI >0 (Appendix A, Table S5, Fig. S1). According to this model, a predictive map of spatial density distribution was created (Fig. 2a). In turn, in the Random Forest (RF) approach, the model with selected predictors based on $MSE_{OOB} \geq 14.0$ explained a higher percentage of variation than the full model (Appendix A, Table S6). The best, i.e. the selected model (Table 1), included 7 predictors (Appendix A, Table S5, Fig. S2). Similarly to the most parsimonious model

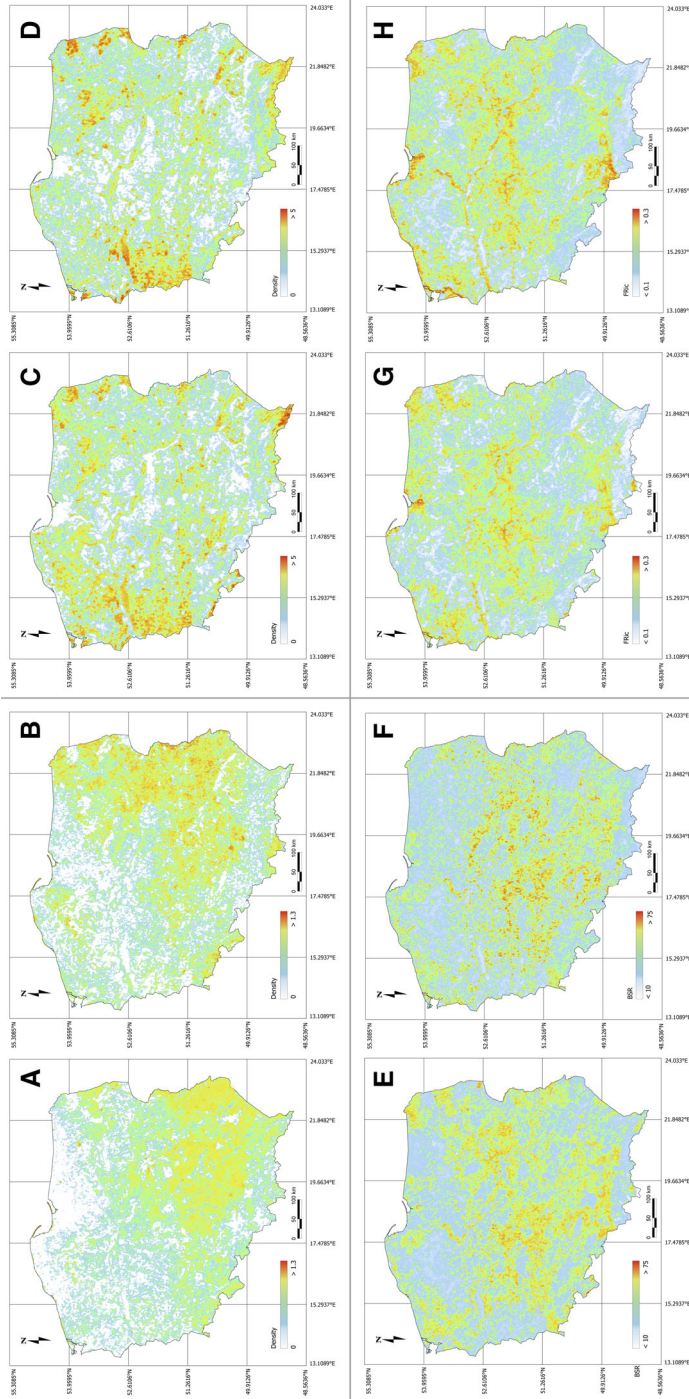


Fig. 2 Predictive maps of the Hayne Estimator of Great Grey Shirke density (a based on GAM), (b based on RF), the Hayne Estimator of Great Spotted Woodpecker (c based on GAM, d based on RF), bird species richness (e based on GAM, f based on RF), functional species richness (g based on GAM, h based on RF)

for the GAM, a predictive map of density distribution was also developed for the best RF model (Fig. 2b). According to the model evaluation procedure (Table 1, Figs. 3a, b, Fig. 4a, b), the RF model had higher predictive power than the GAM.

3.3 Models for the Great Spotted Woodpecker (\hat{D}_{GSW})

In this case five GAMs showed AIC weights >0 (Appendix A, Table S4, model GSW2). Model selection procedures resulted in 6 predictors with the Relative Importance $RI > 0$, but only 5 were included in the best-supported model (Appendix A, Table S5, Fig. S3). On the other hand, the RF model included 8 predictors based on $MSE_{OOB} \geq 22.0$, and it was better than the full model (Appendix A, Table S6, Fig. S4). Following the best RF and GAM, two predictive maps for the \hat{D}_{GSW} were developed (Fig. 2c, d). However, the model evaluation procedure showed that—contrary to the Great Grey Shrike—the GAM had higher predictive power for the Great Spotted Woodpecker's predictive density than the RF (Table 1, Figs. 3c, d, 4c, d).

3.4 Models for Bird Species Richness (BSR)

Three GAMs gained support in BSR models using information-theoretic criteria with AIC weights >0 (Appendix A, Table S4, model 2BSR). The most parsimonious model included five predictors, but six variables showed $RI > 0$ (Appendix A, Table S5, Fig. S5). In the RF approach, the selected model with 7 predictors based on $MSE_{OOB} \geq 9.0$ had a relatively higher importance than the full model (Appendix A, Table S6). According to best GAM and RF models, we created predictive maps of spatial BSR distribution (Fig. 2e, f). Having evaluated the models (Table 1), we reached a conclusion that in this case the GAM had higher predictive power than the RF (Table 1, Figs. 3e, f, 4e, f).

3.5 Models for functional species richness (FRic)

For the second measurement of species richness, i.e. FRic, four GAMs showed AIC weights >0 (Appendix A, Table S4, model FRic2). The most parsimonious model included six predictors (Appendix A, Table S5, Fig. S7), and on its basis we created a predictive map (Fig. 2g). On the other hand, the selection procedure for the RF method showed that the best model included 9 predictors (Appendix A, Table S6, Fig. S8), all of which had $MSE_{OOB} \geq 9.0$ (Appendix A, Table S5). Similarly to the best GAM, we developed a predictive map based on the best RF model (Fig. 2h). Here the evaluation procedure (Table 1, Figs. 3g, h, 4g, h) showed that the RF had higher predictive power than the GAM.

3.6 Spatial autocorrelation

In all cases the Moran's I statistics for residuals coming from each model were very low. We found statistically significant positive autocorrelation for the GAM and the

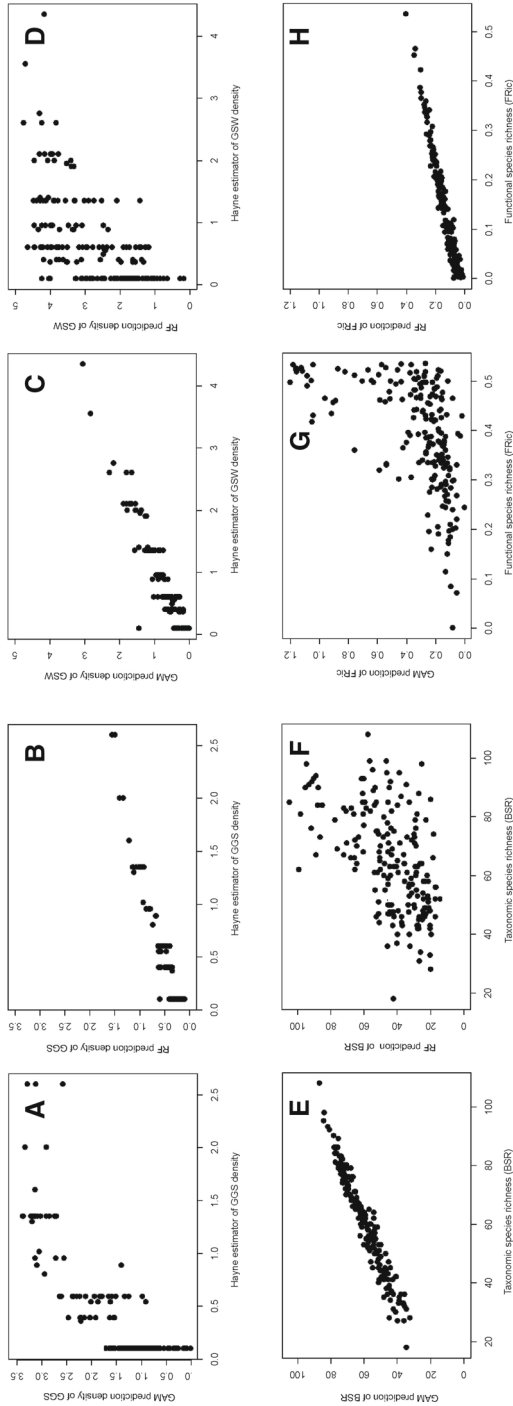


Fig. 3 Correlation between real density/species richness and predictive density/richness

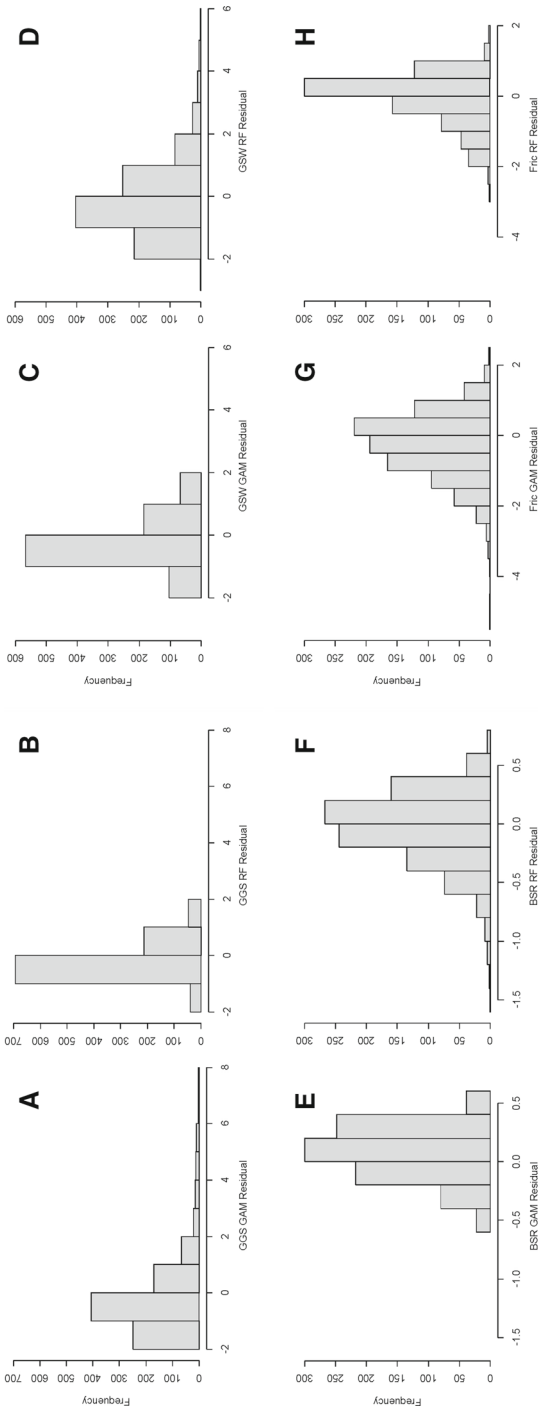


Fig. 4 Residuals' distribution for the best supported GAM and the RF. **a** GAM for Great Grey Shrike density, **b** RF for Great Grey Shrike density, **c** GAM for Great Spotted Woodpecker, **d** RF for Great Spotted Woodpecker, **e** GAM for bird species richness, **f** RF for bird species richness, **g** GAM for functional species richness, **h** RF functional species richness

RF model of the Great Grey Shrike density ($M = 0.021$, $p = 0.02$ and $M = 0.011$, $p = 0.04$, respectively) and bird species richness ($M = 0.014$, $p = 0.04$ and $M = 0.033$, $p = 0.03$, respectively). Residuals from the Great Spotted Woodpecker models indicated spatial autocorrelation only for the GAM ($M = 0.012$, $p = 0.05$), while for the RF we found no statistical significance in the Moran's I statistics ($M = 0.009$, $p = 0.06$). Finally, for the FRIc we found opposite results, i.e. there was no spatial autocorrelation in the GAM ($M = 0.007$, $p = 0.07$), while in the RF the positive spatial effect was low ($M = 0.019$, $p = 0.03$).

4 Discussion

The study supports a hypothesis that the choice of a machine learning modelling technique may influence predictive accuracy of bird density as well as species richness. Apart from the fact that in both variants of the models relationships between environmental variables and response functions were both linear and non-linear and also similar, the Random Forest (RF) detected a larger number of important predictors than the Generalised Additive Model (GAM). Consequently, in each case predictive maps were qualitatively different. This result corresponds to earlier studies based on occurrence models, where different predictive models based on machine learning methods often detected different numbers of important predictors (Pourtaghi et al. 2016). On the one hand, it was proved that the RF approach often had the lowest prediction error (Prasad et al. 2006; Cutler et al. 2007; Syphard and Franklin 2009; Vorpahl et al. 2012; Oliveira et al. 2012), but on the other, the effectiveness of the GAM had also been repeatedly confirmed (Ferguson et al. 2006; Vilchis et al. 2006; Becker et al. 2010; Mannocci and Catalogna 2014; Mannocci and Laren 2014). Our results are consistent with both of the above conclusions, because we found that the GAM was the best method for the Great Spotted Woodpecker density and bird species richness, whereas the RF had the highest predictive power for the Great Grey Shrike and functional species richness. So, we put forth an important question: What drives the differences between the two kinds of models? At first, we assumed that differences between models resulted from inherent spatial autocorrelation of species distributions, which could be driven in different ways in different models (Lichstein et al. 2002). However, the results from the Moran's I statistics obtained in this study (in all cases) were low (range – 0.040 to 0.021), while α -probabilities, i.e. the probability of the study to reject the null hypothesis, were relatively high, leading to a conclusion that spatial autocorrelation did not affect particular models. Thus, we conjectured that differences between models' predictive power resulted from different ecology of the studied species, which could be described by various methods of analysis. Here the results confirm such a speculation, because in this study the effectiveness of a given method depends on the number of predictors (environmental components) that affect density and/or species richness. The highest density of the Great Spotted Woodpecker was observed in the core of all forest types (deciduous, mixed and coniferous), while other habitat components, e.g. aspect, ndvi-jun, altitude, temperature, played only a marginal role. Correspondingly, taxonomic bird species richness was associated with the ecotone between open and forest habitats, while other environmental elements, e.g.

ndvi-jun, precipitation, shape of forest, only slightly improved the models. However, the Great Grey Shrike and functional species richness responded positively to more detailed elements of the environment. In both cases the most important components included the complicated shape of open habitats, the topographical gradient and the level of green vegetation. Both the density of the Great Grey Shrike and FRic are associated with mixed arable fields, where small forests are interspersed with small arable fields and meadows with a high level of green vegetation at the beginning of spring as well as a high roughness index. Additionally, FRic depends also on habitat and topography configuration metrics, i.e. relationships between the core and the buffer zone in the preferred habitats. Therefore, we conclude that the GAM is the most suitable method for species density and measures of species richness that require habitat requirements with sharp boundaries in the landscape (specialist species), i.e. the Great Spotted Woodpecker, and taxonomic species richness (increased efficiency as compared to the RF by 27% and 42%, respectively), while the RF method has the lowest prediction error for functional species richness and more generalist species, such as the Great Grey Shrike (increased efficiency as compared to the GAM by 34% and 30%, respectively). Hence, the results of this study confirm earlier suggestions (Oppel et al. 2012) that different analytical assumptions of models combined with species' specific ecology do influence the predictive power of machine learning methods.

Both methods are very useful for modelling many kinds of ecological phenomena, because they do not need assumptions about the distribution of predictors, and they allow for a mixed use of categorical and numerical variables. They are also capable of considering nonlinearities between variables (Becker et al. 2010; Hastie and Tibshirani 1990; Breiman 2001). Nevertheless, from the biological point of view, the biggest differences between these methods concern the assessment of predictors' importance as well as the development of partial models which are subsequently included in the final model.

The GAM produces one "best" model based on forward or backward selection independent variables, where particular predictors are linearised or included in the model as a polynomial spline (Hastie and Tibshirani 1990). The evaluation of the model and the estimation of predictors' importance are based on information-theoretic criterion, i.e. the AIC and relative importance (RI, Burnham and Anderson 2002). Although this approach can be criticised (Whittingham et al. 2005; Reino et al. 2010), an analysis of all possible models (based on a dredge procedure in *mgcv* library for R) is often used when there is not enough a priori information to develop a small set of models (Reino et al. 2010) or it can be useful in exploratory analyses or even when ecologists have vague ideas how a functional form relates explanatory variables to response variables (Pourtaghi et al. 2016). However, in this case the GAM underestimates densities of generalist species and functional species richness by underrating structures related to landscape configuration, such as topography, vegetation, geometry of patches, just as it had been previously noted for specialist species (Kosicki 2017). Therefore, a plausible conclusion is that it would be a better solution in the presented predictive perspective to average models with weights >0 rather than choose one most parsimonious models (Grenouillet et al. 2010). When models are averaged, predictors, such as elevation, ndvi-may and/or precipitation, detect in the second and third candidate models partial influence of the response function, which is then indirectly reflected in the predictive

map. However, the process of averaging many GAMs comes at a cost, because when there is a plethora of predictors an average model is usually difficult to interpret.

On the other hand, instead of the “best model” the RF constructs multiple models, formed on random samples based on the bootstrap technique (sampling with replacement) (Breiman 2001). The assessment of predictive importance is not based on the information-theoretic criterion, but on specified changes mean decrease (MSE_{OOB}) in accuracy where one predictor is removed from the model (Berk 2008; Breiman 2001), a procedure which results in a ranking of predictors. Then, by using a backward elimination search function implemented in the VSURF library for R (Ismail and Mutanga 2010; Mutanga et al. 2012), it is possible to identify the minimum number of predictors that offer the best predictive accuracy. Such an approach helps simplify the modelling framework without the risk of reducing R^2 (Guyon and Elisseeff 2003; Kohavi and John 1997). Thanks to this approach, we found only 8 environmental parameters out of total 48 predictors for the Great Grey Shrike and 11 for FRic, still achieving the best overall predictive accuracy. However, for the Great Spotted Woodpecker and taxonomic species richness, 9 and 6 independent variables, respectively, were taken into account, and they displayed moderate predictive performance when compared with the GAM.

It should be noted that we also used an additional biological evaluation of models. Standard methods, such as the AUC or the TSS (Jimenez-Valverde 2014; Lobo et al. 2008; Smith 2013) are inadequate, because this type of evaluation can be useful for occurrence models rather than density/richness models. Besides, the AUC and/or the TSS fail to show biological significance of models (Fourcade et al. 2014; Stolar and Nielsen 2015). In this study the AIC for the GAM and the OOB for the RF provide information only about a model’s fit, but not statistics that would indicate biological reality of the model (Araújo et al. 2005). Therefore, before developing the models, 20% of the observations that would be left out had been randomly chosen and treated as an independent dataset. The correlation coefficient between real densities and densities estimated according to the results from the RF and the GAM procedures safeguarded this semi-independent assessment. Although this approach is debatable (James et al. 2013) as evaluation data is only a subset of the training data (Elith et al. 2006), this subset was randomly selected from 970 grid cells which had been inspected for 13 years, thus the data can be deemed unbiased. Finally, this simple procedure intuitively indicates the predictive model’s biological ability without going into complicated algorithms (Howard et al. 2014), such as Monte-Carlo validation, which is a robust estimation of a model’s performance, but very complicated to interpret (James et al. 2013).

5 Conclusion

The study shows that the effectiveness of different Species Distribution Modelling methods depends on the ecology of the studied species. Predominantly, the Generalised Additive Model is responsive to eurybiont (generalists) i.e. species with a broad tolerance range, whereas the Random Forest detects more subtle relationships between density and environmental variables, making it more suitable for stenobiont (special-

ists) species with a low tolerance range. Our study clearly indicates that effective modelling of bird species richness, as well as taxonomic and functional species richness on a large spatial scale is contingent upon the recognition of habitat preferences on a small and meso-geographical scale. The findings contribute significantly to applied ecology by showing that the predictive power of several bird species' models depends on an ecological complexity of systems, including a complex interplay between many environmental components that can only be captured by means of diversified modelling techniques.

Acknowledgements We wish to express our gratitude to numerous observers who collected data in the field. A full list of their names can be found at <http://main5.amu.edu.pl/~kubako/tsr/>. We want to thank Przemysław Chylarecki for the access to Birds' Database, Justyna Grześkowiak for her linguistic assistance, and anonymous reviewers for valuable suggestions and improvements. The Polish Common Breeding Bird monitoring project was funded by the Royal Society for the Protection of Birds (UK), United Nations' GEF Small Grant Programme, European Commission (The Cooperation Fund), Polish Chief Inspectorate of Environmental Protection (GIOŚ) and the National Fund for Environmental Protection and Water Management (NFOŚiGW), and coordinated by the Museum and Institute of Zoology Polish Academy of Sciences and the Polish Society for the Protection of Birds (OTOP).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Antczak M, Hromada M, Grzybek J, Tryjanowski P (2004) Breeding biology of the Great Grey Shrike *Lanius excubitor* in Western Poland. *Acta Ornithol* 39:9–14
- Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends Ecol Evol* 22:42–47
- Araújo MB, Pearson R, Thuiller W, Erhard M (2005) Validation of species–climate impact models under climate change. *Glob Change Biol* 11:1504–1513
- Austin MP, Meyers JA (1996) Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *For Ecol Manag* 85:95–106
- Bahn V, McGill BJ (2007) Can niche-based distribution models outperform spatial interpolation? *Glob Ecol Biogeogr* 16:733–742
- Bartoń K (2013) MuMIn: multi-model inference. R package version 1.9.0. <http://CRAN.R-project.org/package=MuMIn>
- Beale CM, Lennon JJ, Gimona A (2008) Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proc Natl Acad Sci USA* 105:14908–14912
- Becker EA, Forney KA, Ferguson MC, Foley DG, Smith RC, Barlow J, Redfern JV (2010) Comparing California current cetacean-habitat models developed using in situ and remotely sensed sea surface temperature data. *Mar Ecol Prog Ser* 413:163–183
- Berk RA (2008) *Statistical learning from a regression perspective*. Springer, New York
- Breiman L (1996) Bagging predictors. *Mach Learn* 24:123–140
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Breiman L, Friedman J, Olshen R, Stone C (1984) *Classification and regression trees*. Wadsworth International Group, Belmont
- Burnham KP, Anderson DR (2002) *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd edn. Springer, Berlin

- Carrasco L, Mashiko M, Toquenaga Y (2014) Application of random forest algorithm for studying habitat selection of colonial herons and egrets in human-influenced landscapes. *Ecol Res* 29:483–491
- Cernansky R (2017) Biodiversity moves beyond counting species. *Nature* 546:22–24
- Chapman DS (2010) Weak climatic associations among British plant distributions. *Glob Ecol Biogeogr* 19:831–841
- Chylarecki P, Jawińska D (2007) Monitoring Pospolitych Ptaków Lęgowych. Raport z lat 2005–2006. OTOP Warszawa
- Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess TK, Gobson J, Lawler JJ (2007) Random forests for classification in ecology. *Ecology* 88:2783–2792
- Dormann CF, Elith J, Bacher S, Buchmann CM, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2012) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46
- Drew CA, Wiersma Y, Huettmann F (2011) Predictive species and habitat modeling in landscape ecology: concepts and applications. Springer, London
- Elith J, Graham CH (2009) Do they? How do they? WHY do they differ? on finding reasons for differing performances of species distribution models. *Ecography* 32:66–77
- Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Syst* 40:677–697
- Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species distributions from occurrence data. *Ecography* 29:129–151
- Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *J Anim Ecol* 77:802–813
- Esselman PC, Allan JD (2011) Application of species distribution models and conservation planning software to the design of a reserve network for the riverine fishes of northeastern Mesoamerica. *Freshw Biol* 56:71–88
- Ferguson MC, Barlow J, Reilly SB, Gerrodette T (2006) Predicting Cuvier's (*Ziphius cavirostris*) and Mesoplodon beaked whale population density from habitat characteristics in the eastern tropical Pacific Ocean. *J Cetacean Res Manag* 7(3):287–299
- Fourcade Y (2016) Comparing species distributions modelled from occurrence data and from expert-based range maps. Implication for predicting range shifts with climate change. *Ecol Inform* 36:8–14
- Fourcade Y, Engler JO, Rödder D, Secondi J (2014) Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PLoS ONE* 9:e97122
- Fourcade Y, Besnard AG, Secondi J (2017) Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Glob Ecol Biogeogr*
- Franklin J (2010) Mapping species distributions. Spatial inference and prediction. Cambridge University Press, Cambridge
- Fukuda S, De Baets B, Waegeman W, Verwaeren J, Mouton AM (2013) Habitat prediction and knowledge extraction for spawning european grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. *Environ Model Softw* 47:1–6
- GRASS Development Team (2015) Geographic Resources Analysis Support System (GRASS) programmer's manual. Open Source Geospatial Foundation. Electronic document: <http://grass.osgeo.org/programming7/>
- Grenouillet G, Buisson L, Casajus N, Lek S (2010) Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography* 34:9–17
- Grimm R, Behrens T, Märker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island-Digital soil mapping using Random Forests analysis. *Geoderma* 146:102–113
- Guisán A, Theurillat JP (2000) Equilibrium modeling of alpine plant distribution: how far can we go? *Phytocoenologia* 30:353–384
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Mod* 135:147–186
- Guisan A, Edwards TC Jr, Hastie T (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol Model* 157:89–100

- Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C (2013) Predicting species distributions for conservation decisions. *Ecol Lett* 16:1424–1435
- Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182
- Hardy SM, Lindgren M, Konakanchi H, Huettmann F (2011) Predicting the distribution and ecological niche of unexploited snow crab (*Chionoecetes opilio*) populations in Alaskan waters: a first open-access ensemble model. *Integr Comp Biol* 51:608–622
- Hastie T, Tibshirani R (1990) Generalized additive models. Chapman and Hall, London
- Hastie T, Tibshirani RJ, Friedman J (2008) The elements of statistical learning, 2nd edn. Springer, New York
- Hayne DW (1949) An examination of the strip census method for estimation animal populations. *J Wildlife Manag* 13:145–157
- Hegel TM, Cushman SA, Evans J, Huettmann F (2010) Current state of the art for statistical modelling of species distributions. In: Spatial complexity, informatics, and wildlife conservation. Springer, Japan, pp. 273–311
- Heikkinen RK, Marmion M, Luoto M (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography* 35:276–288
- Hijmans RJ, Graham CH (2006) The ability of climate envelope models to predict the effect of climate change on species distributions. *Glob Change Biol* 12:2272–2281
- Hillebrand H, Matthiessen B (2009) Biodiversity in a complex world: consolidation and progress in functional biodiversity research. *Ecol Lett* 12:1405–1419
- Hof AR, Jansson R, Nilsson C (2012) The usefulness of elevation as a predictor variable in species distribution modelling. *Ecol Model* 246:86–90
- Howard C, Stephens PA, Pearce-Higgins JW, Gregory RD, Willis SG (2014) Improving species distribution models: the value of data on abundance. *Methods Ecol Evol* 5:506–513
- Ismail R, Mutanga OA (2010) Comparison of regression tree ensembles: Predicating *Sirex noctilio* induced water stress in *Pinus patula* forest of KwaZulu-Natal, South Africa
- James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning: with applications in R. Springer
- Jimenez-Valverde A (2014) Threshold-dependence as a desirable attribute for discrimination assessment: implications for the evaluation of species distribution models. *Biodivers Conserv* 23:369–385
- Jimenez-Valverde A, Peterson AT, Soberón J, Overton JM, Aragon P, Lobo JM (2011) Use of niche models in invasive species risk assessments. *Biol Invasions* 13:2785–2797
- Kohavi R, John G (1997) Wrappers for feature subset selection. *Artif Intell* 97:273–324
- Kosicki JZ (2017) Should topographic metrics be considered when predicting species density of birds on a large geographical scale? A case of Random Forest approach. *Ecol Model* 349:76–85
- Kosicki JZ (2018) Are landscape configuration metrics worth including when predicting specialist and generalist bird species density? A case of the Generalised Additive Model approach. *Environ Model Assess* 23:193–202
- Kosicki JZ, Hromada M (2018) Cuckoo density as a predictor of functional and phylogenetic species richness in the predictive modelling approach: extension of Tryjanowski and Morelli (2015) paradigm in the analytical context. *Ecol Ind* 88:384–392
- Kosiński Z, Ksít P, Winięcki A (2006) Nest sites of Great Spotted Woodpeckers *Dendrocopos major* and Middle Spotted Woodpeckers *Dendrocopos medius* in near-natural and managed riverine forests. *Acta Ornithol* 41:21–32
- Kosiński Z, Plut M, Ulanowska A, Walczak Ł, Winięcki A, Zarębski M (2018) Do increases in the availability of standing dead trees affect the abundance, nest-site use, and niche partitioning of great spotted and middle spotted woodpeckers in riverine forests? *Biodivers Conserv* 27:123–145
- Krebs CJ (1999) Ecological methodology. Addison Wesley Longman, New York
- Laliberté E, Legendre P, Shipley B (2015) Measuring functional diversity (FD) from multiple traits and other tools for functional ecology: R Package Version1
- Lawler JJ, White D, Neilson RP, Blaustein AR (2006) Predicting climate-induced range shifts: model differences and model reliability. *Global Change Biol* 12:1568–1584
- Lei Z, Shirong L, Pengsen S, Wang T (2011) Comparative evaluation of multiple models of the effects of climate change on the potential distribution of *Pinus massoniana*. *Chin J Plant Ecol* 35:1091–1105
- Liaw A, Wiener M (2002) Classification and regression by Random Forest. *R News* 2:18–22

- Lichstein JW, Simons TR, Shriener AS, Franzreb KE (2002) Spatial autocorrelation and autoregressive models in ecology. *Ecol Monogr* 72:445–463
- Lobo JM, Jimenez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr* 17:145–151
- Lobo JM, Jimenez-Valverde A, Hortal J (2010) The uncertain nature of absences and their importance in species distribution modelling. *Ecography* 33:103–114
- Mannocci L, Catalogna M (2014) Predicting cetacean and seabird habitats across a productivity gradient in the South Pacific gyre. *Prog Oceanogr* 120:383–398
- Mannocci L, Laran S (2014) Predicting top predator habitats in the Southwest Indian Ocean. *Ecography* 37:261–278
- Mattsson BJ, Zipkin EF, Gardner B, Blank PJ, Sauer JR, Royle JA (2013) Explaining local-scale species distributions: relative contributions of spatial autocorrelation and landscape heterogeneity for an avian assemblage. *PLoS ONE* 8:e55097
- Mutanga O, Adam E, Cho MA (2012) High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *Int J Appl Earth Obs Geoinf* 18:399–406
- Naimi B, Hamm NAS, Groen TA, Skidmore AK, Toxopeus AG (2014) Where is positional uncertainty a problem for species distribution modelling? *Ecography* 37:191–203
- Oliveira S, Oehler F, San-Miguel-Ayanz J, Camia A, Pereira JMC (2012) Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *For Ecol Manag* 275:117–129
- Oppel S, Meirinho A, Ramírez I, Gardner B, O’Connell AF, Miller PI, Louzao M (2012) Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biol Conserv* 156:94–104
- Paradis E (2016) Ape library for R: analyses of phylogenetics and evolution ver: 3.5. <https://cran.r-project.org/web/packages/ape/index.html>
- Pearman PB, Lavergne S, Roquet C, Wüest R, Zimmermann NE, Thuiller W (2014) Phylogenetic patterns of climatic, habitat and trophic niches in a European avian assemblage. *Glob Ecol Biogeogr* 23:414–424
- Petitpierre B, Broennimann O, Kueffer C, Daehler C, Guisan A (2017) Selecting predictors to maximize the transferability of species distribution models: lessons from cross-continental plant invasions. *Glob Ecol Biogeogr* 26:275–287
- Pourtaghi Z, Pourghasemi HR, Aretano R (2016) Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques. *Ecol Indic* 64:72–84
- Prasad AM, Iverson RL, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9:181–199
- R Development Core Team (2017) R: a language and environment for statistical computing
- Reino L, Portod M, Morgado R, Carvalhof F, Miraf A, Bejac P (2010) Does afforestation increase bird nest predation risk in surrounding farmland? *For Ecol Manag* 260:1359–1366
- Robin G, Jean-Michel P, Tuleau-Malot Ch (2010) Variable selection using Random Forests. *Pattern Recognit Lett* 31:2225–2236
- Santana J, Porto M, Gordinho L, Reino L, Beja P (2012) Long-term responses of Mediterranean birds to forest fuel management. *J Appl Ecol* 49:632–643
- Segurado P, Araujo MB (2004) An evaluation of methods for modelling species distributions. *J Biogeogr* 31:1555–1568
- Seppelt R, Voinov A (2002) Optimization methodology for land use patterns using spatially explicit landscape models. *Ecol Model* 151:125–142
- Smith AB (2013) On evaluating species distribution models with random background sites in place of absences when test presences disproportionately sample suitable habitat. *Divers Distrib* 19:867–872
- Stolar J, Nielsen SE (2015) Accounting for spatially biased sampling effort in presence-only species distribution modelling. *Divers Distrib* 21:595–608
- Syphard AD, Franklin J (2009) Species’ traits affect the performance of species’ distribution models for plants in southern California. *J Veg Sci* 21:177–189
- Vilchis LI, Ballance LT, Fiedler PC (2006) Pelagic habitat of seabirds in the eastern tropical Pacific: effects of foraging ecology on habitat selection. *Mar Ecol Prog Ser* 315:279–292
- Vincenzi S, Zuchetta M, Franzoi P, Pellizzato M, Pranovi F, De Leo GA, Torricelli P (2011) Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy. *Ecol Model* 222:1471–1478

- Virkkala R, Marmion M, Heikkinen RK, Thuiller W, Luoto M (2010) Predicting range shifts of northern bird species: influence of modelling technique and topography. *Acta Oecol* 36:269–281
- Vorpahl P, Elsenbeer H, Märker M, Schröder B (2012) How can statistical models help to determine driving factors of landslides? *Ecol Model* 239:27–39
- Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62:2868–2883
- Weisberg S (1980) *Applied linear regression*. Wiley, New York
- Whittingham MJ, Swetnam RD, Wilson JD, Chamberlain DE, Freckleton RP (2005) Habitat selection by yellowhammers *Emberiza citrinella* on lowland farmland at two spatial scales: implications for conservation management. *J Appl Ecol* 42:270–280
- Williams JN, Seo C, Thorne J, Nelson JK, Erwin S, O'Brien JM, Schwartz MW (2009) Using species distribution models to predict new occurrences for rare plants. *Divers Distrib* 15:565–576
- Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A (2008) Effects of sample size on the performance of species distribution models. *Divers Distrib* 14:763–773
- Wood SN (2006) *Generalized additive models: an introduction with R*. Chapman and Hall/CRC, Boca Raton
- Wood S (2013) mgcv: mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. R package version 1.7-22, <http://cran.rproject.org/web/packages/mgcv>

Jakub Z. Kosicki is an associate professor at the Instytut of Environmental Biology of the Adam Mickiewicz University, Poznań, Poland. He received the Ph.D. degree from the Adam Mickiewicz University, Poznań in 2008. He has published papers in several ecological (*Journal of Animal Ecology*, *Ecological Modelling*, *Environmental modeling and assessment*, etc.) and ornithological (*Ibis*, *Journal of Ornithology*, *Bird Study*, etc.) journals. His current research interests are species distribution modeling, agent-based modeling, and analysis of population dynamics.

Affiliations

Jakub Z. Kosicki¹

✉ Jakub Z. Kosicki
kubako@amu.edu.pl

¹ Department of Avian Biology & Ecology, Faculty of Biology, Adam Mickiewicz University, Poznań, Uniwersytetu Poznańskiego str. 6, 61–614 Poznan, Poland