

# Earth observation based indication for avian species distribution models using the spectral trait concept and machine learning in an urban setting

Thilo Wellmann<sup>a,b,\*</sup>, Angela Lausch<sup>a,b</sup>, Sebastian Scheuer<sup>a</sup>, Dagmar Haase<sup>a,b</sup>

<sup>a</sup> Department of Geography, Humboldt Universität zu Berlin, Rudower Chaussee 16, D-12489 Berlin, Germany

<sup>b</sup> Department of Computational Landscape Ecology, Helmholtz Centre for Environmental Research—UFZ, Permoserstr. 15, D-04318 Leipzig, Germany

## ARTICLE INFO

### Keywords:

Remote sensing  
Spectral traits  
Species distribution model  
Random forest  
Urban birds  
Machine learning

## ABSTRACT

Birds respond strongly to vegetation structure and composition, yet typical species distribution models (SDMs) that incorporate Earth observation (EO) data use discrete land-use/cover data to model habitat suitability. Since this neglects factors of internal spatial composition and heterogeneity of EO data, we suggest a novel scheme deriving continuous indicators of vegetation heterogeneity from high-resolution EO data.

The deployed concepts encompass vegetation fractions for determining vegetation density and spectral traits for the quantification of vegetation heterogeneity. Both indicators are derived from RapidEye data, thus featuring a continuous spatial resolution of 6.5 m. Using these indicators as predictors, we model breeding bird habitats using a random forest (RF) classifier for the city of Leipzig, Germany using a single EO image.

SDMs are trained for the breeding sites of 44 urban bird species, featuring medium to very high accuracies (59–90%). Analysing similarities between the models regarding variable importance of single predictors allows species groups to be determined based on their preferences and dependencies regarding the amount of vegetation and its spatial and structural heterogeneity. When combining the SDMs, models of urban bird species richness can be derived.

The combination of high-resolution EO data paired with the RF machine learning technique creates very detailed insights into the ecology of the urban avifauna, opening up opportunities of optimising greenspace management schemes or urban development in densifying cities concerning overall bird species richness or single species under threat of local extinction.

## 1. Introduction

Modelling potential breeding sites that are species-specific can be an integral part of urban, peri-urban and non-urban biodiversity studies and conservation strategies (Guisan and Thuiller, 2005). The urban environment is especially rich in birds, often surpassing their rural surroundings in terms of biomass and diversity (Chace and Walsh, 2006). A key element of sustaining viable population sizes of single species under threat or increasing overall species richness and abundance is the identification and protection of breeding sites. A core element determining the breeding sites of birds is vegetation structure such as vegetation density and diversity (Paker et al., 2014). Earth observation datasets provide a cost-effective, reproducible and straightforward method for the analysis of such vegetation parameters.

Satellite-derived information has been widely used to predict species richness, diversity and turnover in a variety of kingdoms (Rocchini et al., 2010, 2017). While the analysis of such diversity parameters is

valuable (Rocchini et al., 2010), those analyses lack species-specific information. For multiple use cases such as species protection measures or environmental impact assessments, species distribution models (SDMs) are needed (Guisan and Thuiller, 2005). However, there is a clear lack in SDMs since existing models have two major problems regarding the characteristics of input data and modelling technique.

Regarding modelling techniques, studies often use regression (Bino et al., 2008; Warton et al., 2015). Due to the model assumptions inherent to most regression methods, problems such as collinearity between predictors, outliers or non-linear and exponential relationships may result in bad model performance (Rousseeuw and Leroy, 2005; Dormann et al., 2013). Thus, to overcome the aforementioned limitations, a more flexible machine learning approach seems favourable for SDMs. One particularly robust and well-established procedure in ecology and EO studies is thereby the RF-algorithm (Cutler et al., 2007; Belgiu and Drăgu, 2016), which is an ensemble learning method consisting of a multitude of decision trees (Breiman, 2001). RFs are able to

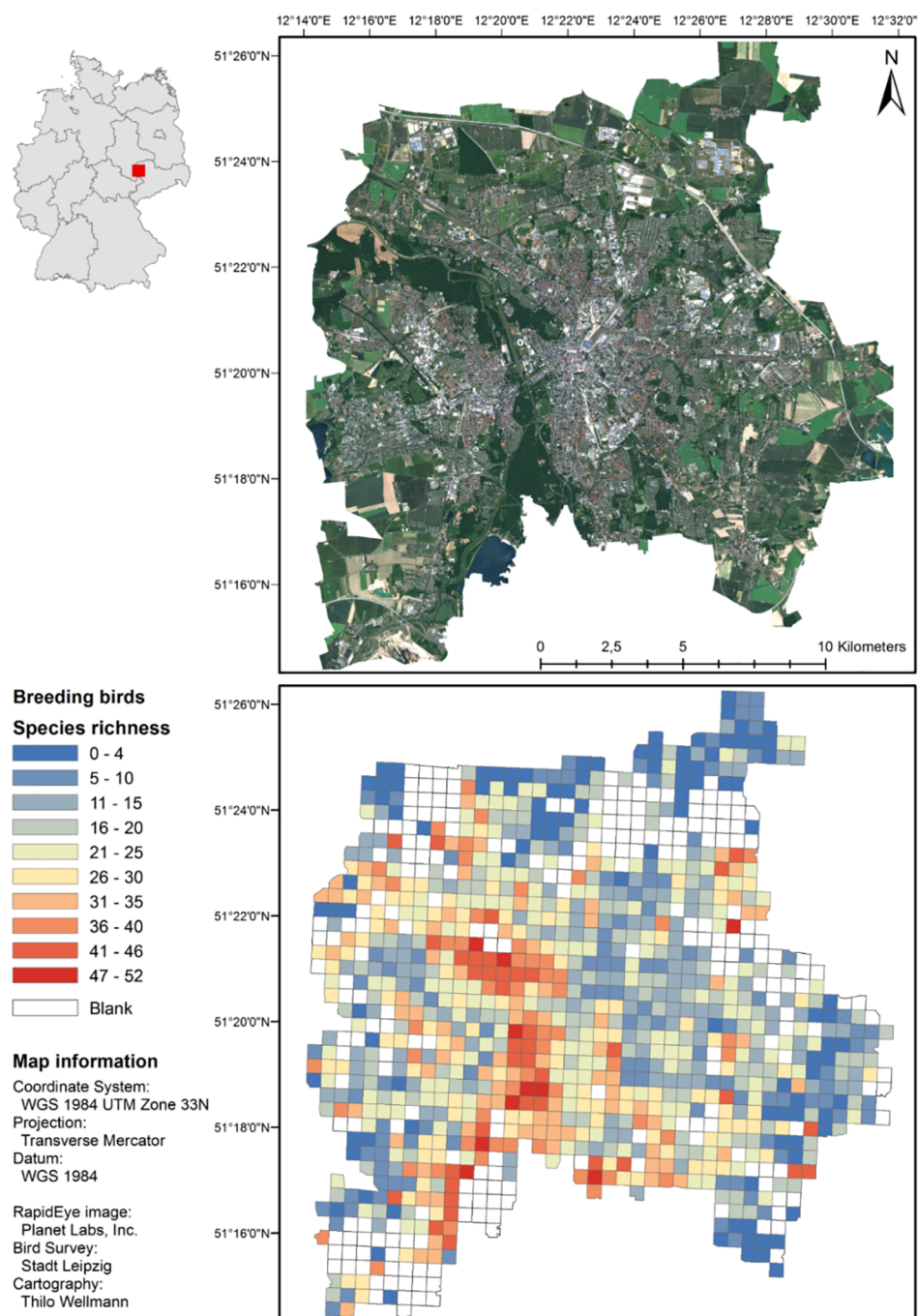
\* Corresponding author.

E-mail address: [thilo.wellmann@hu-berlin.de](mailto:thilo.wellmann@hu-berlin.de) (T. Wellmann).

<https://doi.org/10.1016/j.ecolind.2019.106029>

Received 25 June 2019; Received in revised form 19 November 2019; Accepted 17 December 2019

1470-160X/ © 2019 Elsevier Ltd. All rights reserved.



**Fig. 1.** RapidEye satellite image of the city of Leipzig with indication of the study area location in Germany and the breeding bird dataset showing bird species richness for all parts of the city but the areas that were excluded due to large-scale change in biotope composition between the acquisition of the RapidEye scene and the bird survey.

deal with highly collinear predictors that can be both quantitative (numeric) and qualitative (non-numeric) with all kinds of variable interactions, making them, therefore, often superior to regression.

Input data is often inadequate because a multitude of models use classified, discrete land-use/cover data (Falcucci et al., 2007). This implies two important pitfalls, firstly, the loss of information, namely the internal heterogeneity in a certain land-use/cover class, and secondly, the loss of transition zones between different classes through sharp boundaries (Palmer et al., 2002; Lausch et al., 2015). However, transition zones and internal heterogeneity are key factors for bird species' distribution (He et al., 2015).

Urban environments are dynamic and complex and, within them,

sites of high biodiversity can be found next to intensely managed ones (Haase et al., 2014; Knapp et al., 2017). This species richness, however, seems to be in danger as recent reports state that multiple species in Europe (Bowler et al., 2019) and also in Germany (Gedeon et al., 2004) are in rapid decline. This trend is especially apparent for bird species breeding in urban and agricultural settings, since those feature the most rapid reduction of all regarded habitat types (Gedeon et al., 2004). Since the case study area of this paper, the city of Leipzig, Germany, is characterised by a dense centre with vast parks as well as a large natural forest and fertile agrarian surroundings, it is an ideal case study for developing models for those endangered species groups and also for the large group of forest birds (Wellmann et al., 2018).

Urban ornithological studies show that even small patches of vegetation can serve as viable breeding sites (Ikin et al., 2013) and that birds respond to both vegetation composition and configuration (Chace and Walsh, 2006). Hence, for complex urban settings such as the city of Leipzig, high-resolution data is much needed. Therefore, high spatial resolution data, as provided by the RapidEye satellite fleet, seems favourable for deployment in the urban environment (Tigges et al., 2013). From such high-resolution EO data, various plant characteristics can be analysed by using the spectral traits approach (Lausch et al., 2016). This spectral trait framework, introduced by Lausch et al. (2016), builds on the traits framework (Kattge et al., 2011) by incorporating those plant traits that are detectable by EO based techniques. The spectral traits concept hence includes biochemical, biophysical, physiological, structural, phenological or functional characteristics of plants, populations and communities (Kattge et al., 2011; Lausch et al., 2016).

The spectral trait concept is a functional approach in which every plant trait corresponds to a function, that is relevant for (i) the plant and (ii) the larger ecosystem (Violle et al., 2007). Therefore, the spectral traits approach is an efficient interface linking EO data to key ecosystem characteristics, functions and services (Lausch et al., 2016), which in return could be linked to bird species breeding behaviour.

One way of analysing the spatial diversity of spectral traits in a plant community is by quantifying the composition and configuration of a plant trait related product, e.g., Normalized Difference Vegetation Index (NDVI), in space and over time (Wellmann et al., 2018). For this, texture measures by Haralick et al. (1973), such as the grey level co-occurrence matrix (GLCM), are powerful and well established methods used by St-Louis et al. (2009) for the prediction of bird species diversity.

Consequently, the combination of high-resolution satellite data paired with machine learning techniques can create novel and detailed insights into the ecology of urban birds and their habitats. Since there is no established framework for modelling bird-breeding sites based on continuous spectral EO data, this paper seeks to develop an according methodology to predict the breeding sites for urban bird species. The following research questions guide the development:

- (i) Are fractional vegetation cover and spectral plant traits meaningful indicators for the prediction of breeding sites for species in the urban environment?
- (ii) What are suitable modelling techniques?
- (iii) How accurate are SDMs solely derived from EO data?
- (iv) How do the SDMs help to predict bird species richness?

## 2. Study area

Leipzig is a dense city in Eastern Germany located at 51°20'N, 12°22'E with 560,000 inhabitants. The city houses a considerable number of natural biotopes and breeding-bird species richness is comparably high (Fig. 1). Almost 40% of all bird species breeding in Germany ( $n = 314$ ) can also be found in Leipzig ( $n = 120$ ) (StUfa, 1995; Völkl et al., 2004). Important breeding grounds are located along a north to south transect in the large remnants of the alluvial forest on the floodplains. This forest is one of the largest of its kind in Europe and features a quasi-natural structure in terms of species composition, which is dominated by ash, oak, beech, lime and sycamore trees. Next to forested areas, different types of urban greenspaces, urban building structures, and permanent agricultural systems in the surrounding may provide rich breeding grounds (StUfa, 1995) (Fig. 1).

The development of Leipzig since the German reunification in 1990 can be quickly summarized. In the observed period, Leipzig's population declined slightly leading to a stable stock in central buildings. Suburbanisation tendencies on peri-urban agricultural land in the northern outskirts lead to large-scale developments in the outskirts, mostly consisting of logistic infrastructure, industrial facilities or the exhibition grounds. These developments mostly occurred outside of

important nesting habitats for breeding birds, since the high-intensity farming during the socialist past until 1990 left few ecological niches. A second important trend in the southern outskirts of the city are the flooding of former opencast mines (Wolff et al., 2016).

Since the 1990s, the public urban green infrastructure such as the alluvial forest, parks, graveyards and allotment garden facilities did not change on a broad scale. The overall extent of forest cover did not change from the 1990s onwards and road-side trees are predominantly considerably old, which leads to small changes over a 13 year period in traits (trees younger than 20 years only make up 15.6% of total road-side trees) (Leipzig, 2018, 2019). The largest amount of old-growth and undisturbed vegetation can be found on graveyards and in the large alluvial forest, called "The Auwald." Most parts are protected under the FFH statues which therefore obliges strong protection measures. This stability is also true for parts of the private green, for instance in the Wilhelmine quarters where old grown vegetation prevailed throughout the investigation period.

In summarizing the findings above, we see that, while the structure of the central areas remained very stable, there was considerable change at the outskirts of the city. Fig. 1 shows where the change was assessed with two biotope maps, the first from 1993 and the second from 2005. All blank areas inside the city featured more than a 35% change in biotope types between the two timeframes under consideration and where subsequently disregarded in this study. This was necessary due to the time lag between the acquisition of the bird dataset and the RapidEye image.

## 3. Data and methods

To model the presence and absence of 44 breeding bird species, we propose a new methodology that only uses a single RapidEye EO data set (Fig. 2). The EO based methodology builds on fractional vegetation cover, the NDVI and a principal component analysis (PCA). We then used these products to calculate indicators of spatial heterogeneity, mainly with a grey level co-occurrence matrix (GLCM) and indicators of vegetation density. Using a random forest (RF) classifier on the aforementioned data sets, we predicted presence and absence patterns of single species, of species clusters, and of overall species richness.

### 3.1. Earth observation (EO) data and indicator calculation

#### 3.1.1. Pre-processing of RapidEye data

The RapidEye sensor features five spectral bands in the 400–850 nm range (blue, green, red, red-edge, near infra-red) with 6.5 m spatial resolution (Tigges et al., 2013). This study uses a RapidEye scene of the city of Leipzig that was acquired on June 3, 2011. The timing of the scene matches the vegetation period and the breeding patterns of the majority of the urban bird species. A bird-breeding calendar revealed that during June the largest proportion of the studied birds are actually breeding (Südbeck, 2005). The acquired RapidEye level 3A product was atmospherically corrected with ATCOR 2, assuring best transferability and interoperability between satellite scenes, geographic regions and different sensors (Richter, 2011).

Based on the atmospherically corrected RapidEye scene, a map of fractional vegetation cover was calculated (Fig. 3) (Haase et al., 2019). This dataset provides subpixel information about the share of vegetation featured in each pixel. It was used to mask out areas in the original RapidEye dataset that are not primarily vegetated. As a cut-off value, 75% was chosen, meaning that only pixels featuring at least a 75% cover in vegetation were used in the study. This threshold was found to be a good compromise between reducing spectral information of build-up origin as much as possible while not dropping too many pixels where only minor portions are not vegetated or where some soil signal is coming through, which is relevant for grasslands. Furthermore, excluded were agricultural areas that do not feature permanent crop systems in the breeding period, areas were detected using the 2011

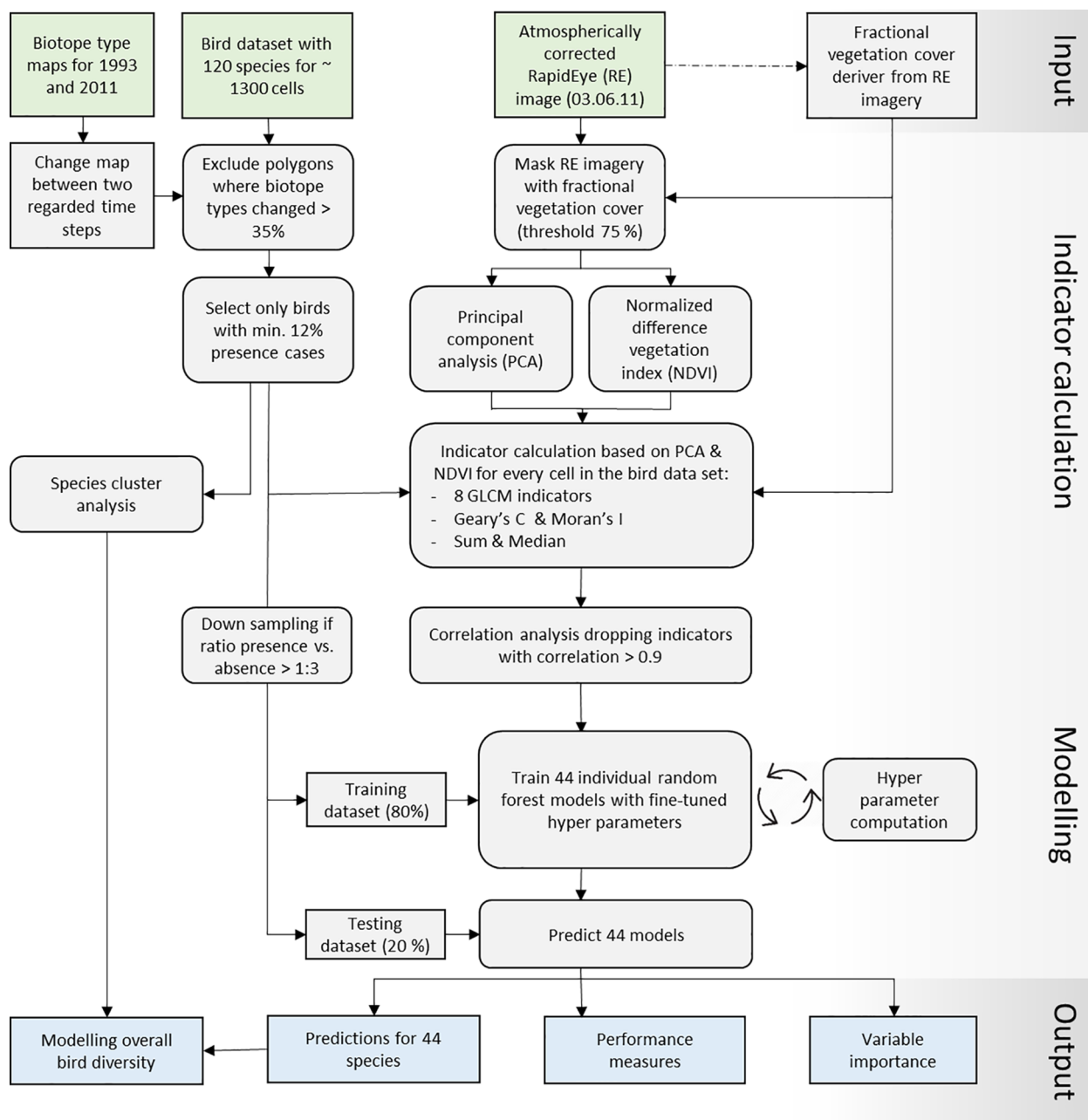


Fig. 2. Methodological overview; separated into input data, indicator calculation, modelling and the generated outputs.

biotope map.

### 3.1.2. Calculation of functional vegetation indicators

From the masked RapidEye image, a PCA and a NDVI layer were computed. A PCA is a statistical procedure that transforms collinear datasets into linearly non-correlated variables, in our case the five spectral bands of the RapidEye data set (Jolliffe, 2002). Since only pixels with vegetation were included in this procedure, the first principal component describes the diversity found in the vegetation's whole traits. The NDVI, in turn, is more specifically oriented towards the calculation of the greenness of the plants. This is very much related to the plants' capacity for performing photosynthesis and, respectively, to the traits of the plants that are associated with photosynthesis (Gamon et al., 1995).

Based on the three products (NDVI, PCA band 1 & band 2) described above, 13 different indicators that describe both state and spatial heterogeneity in the plant communities are calculated for every 500 m by

500 m cell (Table 1). The GLCM indicators by Haralick et al. (1973) assess local image texture by calculating the differences between the values of adjacent pixels values, e.g. NDVI value. Based on this frequency matrix, eight indicators were calculated and their mean and standard deviation values assessed for every cell. Furthermore, two measures of spatial autocorrelation and three summary statistics were calculated.

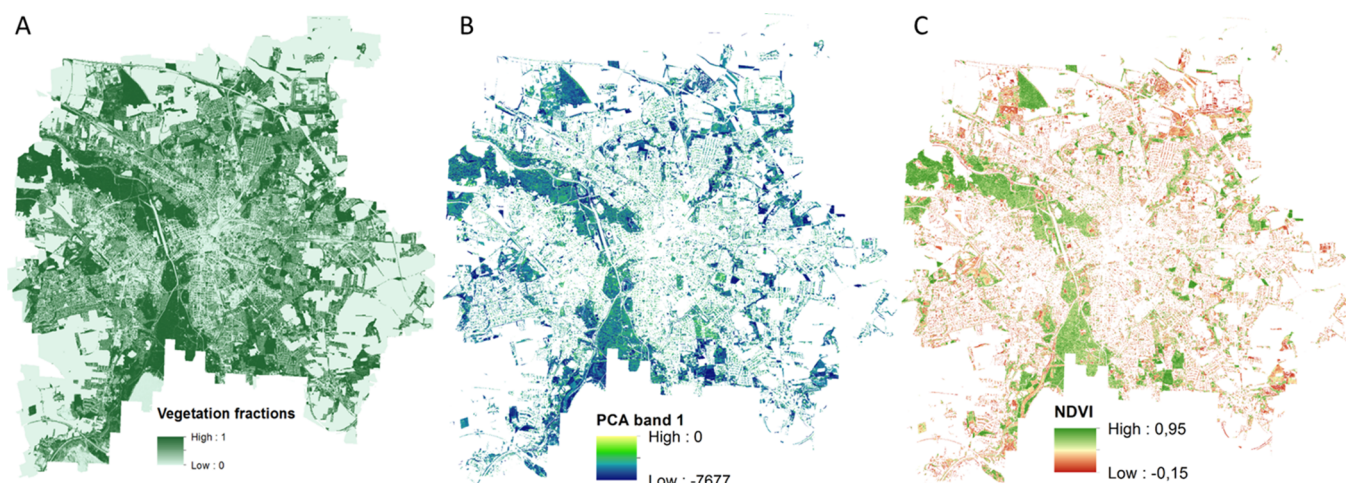
Since we derived all indicators used for modelling from the same RapidEye dataset, a correlation analysis was performed using a cut-off value of 0.9. The comparably high cut-off value was chosen because the RF algorithm is able to deal with highly collinear data (Breiman, 2001).

### 3.2. Bird species distribution data and further analysis

#### 3.2.1. The bird survey dataset and its pre-processing

The breeding bird data used in this study was collected in the city of Leipzig over three breeding periods from 1991 to 1993 between





**Fig. 3.** Earth observation indicators calculated based on the RapidEye image (June 3, 2011); (A) a fractional vegetation map, (B) the first band of the principal component analysis (PCA) and (C) the normalized difference vegetation index (NDVI).

**Table 1**

Overview of the indicator types expressing vegetation density or vegetation heterogeneity.

Type	Name	Reference
Local spatial autocorrelation	GLCM mean	(Haralick et al., 1973)
	GLCM variance	
	GLCM correlation	
	GLCM homogeneity	
	GLCM contrast	
	GLCM dissimilarity	
	GLCM entropy	
	GLCM angular second moment	
Global spatial autocorrelation	Geary's <i>C</i>	(Geary, 1954)
	Moran's <i>I</i>	
Descriptive statistics	Standard Deviation	(Datt, 1998)
	Coefficient of Variation	
	Sum	

February and July. It describes the presence or absence of 120 species within 1132 cells with a resolution of 500 m by 500 m. Each of these cells was surveyed at least 5 times per year by ornithologists to map species that were breeding. The species was marked as present in a cell if it was observed at least once. Through this scheme, very reliable presence and absence data could be generated (StUfa, 1995).

The dataset was treated to account for uncertainties in the occurrence of bird species. Firstly, only validated species sightings were included and, subsequently, all entries marked as uncertain were disregarded. Secondly, rare species had to be excluded to ensure that a sufficient amount of presence points remained for validation of the proposed RF-model. As a threshold, a presence to absence ratio greater than 10% was chosen. Due to the exclusion of the rare species, only 44 of the 120 species remained for analysis.

Finally, to account for changes in landscape composition between the acquisition time of the bird survey (1993) and the RapidEye acquisition (2011), a change analysis was conducted in order to exclude grid cells with major land use or land cover changes. This analysis is based on two biotopes, dating from 1993 and 2011. In so doing, cells with more than 35% change in biotope types were excluded from the analysis. Overall, around 200 grid cells, predominantly in the peri-urban space, were excluded.

### 3.2.2. Determination of functional species communities with a cluster analysis

A hierarchical cluster analysis was performed in order to find groups of bird species that are similar to each other in terms of their presence/

absence patterns in Leipzig, to identify coexisting species. The *hclust* (R Core Team, 2000) algorithm used in this study iteratively assigns an object to a cluster based on a distance measure. The allocation of group memberships is executed on the premise of minimizing the distance between the clusters members. Distance in this study was measured by the Jaccard distance (1), a statistical measure computing the dissimilarity between sample sets (Podani and Schmera, 2011):

$$d_j(A, B) = 1 - J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|} \quad (1)$$

where  $d_j$  is the Jaccard distance,  $J$  is the Jaccard index, and where  $A$  and  $B$  are the presence/absence points of two regarded species.

The optimal number of clusters was determined using the total within sum of squares method (TSS) (2):

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2)$$

where  $y_i$  is a single instance of the dependent variable and  $\bar{y}$  its mean. The ideal number was visually derived from an elbow plot.

### 3.3. Machine-learning based species distribution modelling techniques

This study tested five different modelling techniques in terms of their capabilities of predicting species presence/absence of birds in Leipzig (Table 2). While overall the ensemble model delivers the results with the highest accuracies, this study recommends the RF-model as the means of choice. This is because the additional complexity and computing efforts do not justify a slight increase in accuracy.

#### 3.3.1. Random forest modelling and hyperparameter refinement

RF is an algorithm capable of solving regression and classifications problems (Breiman, 2001), providing fast model training and comparably high accuracies (Mitchell, 2011). In this study, 44 random forests are grown—one for every species—to predict species presence/absence pattern with an independent test data set.

For the hyperparameter computation, we used a threefold cross validation scheme: Regarded parameters and their value ranges are *K-Features* (1–13 in steps of 1), determining the number of variables chosen at each split in a tree; *Ntree* (500–1500 in steps of 100) determining the number of trees to be grown; and *Nodesize* (1–13 in steps of 1), determining the minimum size of a terminal node (Bernard et al., 2009).

#### 3.3.2. Downsampling for bird species distribution modelling

For 32 of the 44 bird species, absence points outnumber presence

**Table 2**

Final decision matrix on choosing the modelling technique, the green shading indicates the best outcome per criterion.

Method	Result accuracies			Computation time per model in seconds	Variable importance	Categorical variables	R package name
	Mean accuracy	Mean sensitivity	Mean specificity				
Random Forest	0.78	0.70	0.77	1	yes	yes	randomForest
EV Tree	0.76	0.61	0.80	7.5	yes	yes	evtree
Neural Network	0.78	0.65	0.78	1	no	no	e1071
Support Vector Machine	0.62	0.47	0.60	140	no	no	neuralnet
Ensemble Model	0.78	0.77	0.77	150	no	no	

points. Since highly unbalanced data can cause problems in the random forest classification, a downsampling approach for the species absences was used (Chen et al., 2004). In this study, rare species were downsampled, meaning that absence points were disregarded until a ratio of 1:3 between presence and absence points was reached. This ratio was found to produce the most accurate results overall. Downsampling in this study is regarded as a very critical and important step which needs to be carefully addressed and iteratively tested.

### 3.4. Accuracy metrics

The model quality will be discussed based on three indicators; sensitivity (3) specificity (4) and overall accuracy (5) (see Table 3). Overall accuracy thereby refers to the ratio of correctly classified instances. Sensitivity denotes the ratio of all positively classified instances (in this case species presences) correctly classified. Specificity in turn builds the same ratio for species absences (Kuhn, 2008).

The quality of the diversity models will be discussed based on the mean absolute error (MAE) (6), which is a measure of difference between two continuous variables, in our case predicted species richness versus observed species richness:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i| \quad (6)$$

where  $n$  is the number of observations,  $\hat{Y}_i$  the predicted value and  $Y_i$  the

observed value.

### 3.5. Connecting remotely sensed spectral traits with bird species traits

The trait framework is an integral part of community ecology (McGill et al., 2006). As an outlook, we therefore demonstrate how the trait approach could provide for a direct modelling interface between remotely sensed spectral trait indicators and bird species traits. For this part, a dietary trait of the 44 bird species is used. The data is taken from Sibly et al. (2012).

For modelling the relationship between the proposed indicators and the selected bird species traits, we used a multiple correspondence analysis (MCA). A MCA transforms nominal categorical data into a low-dimensional feature space. It is thus a methodology similar to the PCA – but for categorical data. This way, underlying structures and correspondences of different nominal categorical variables can be detected (Greenacre and Blasius, 2006).

## 4. Results

### 4.1. Species distribution modelling

#### 4.1.1. Accuracies of computed random forest models

The dataset was split into 20% for testing and 80% for growing the random forests. The mean overall accuracy for the 44 models, based on

**Table 3**

Confusion matrix and formulas for accuracy indicators.

Reference				
Predicted	Event	No event		
Event	A	B	Sensitivity = A/(A+C)	(3)
No event	C	D	Specificity = D/(B+D)	(4)
			Overall acc. = (A+D)/(A+B+C+D)	(5)
A = True Positives; B = False Negatives; C = False Positives; D = True Negatives;				
A = True Positives; B = False Negatives; C = False Positives; D = True Negatives;				

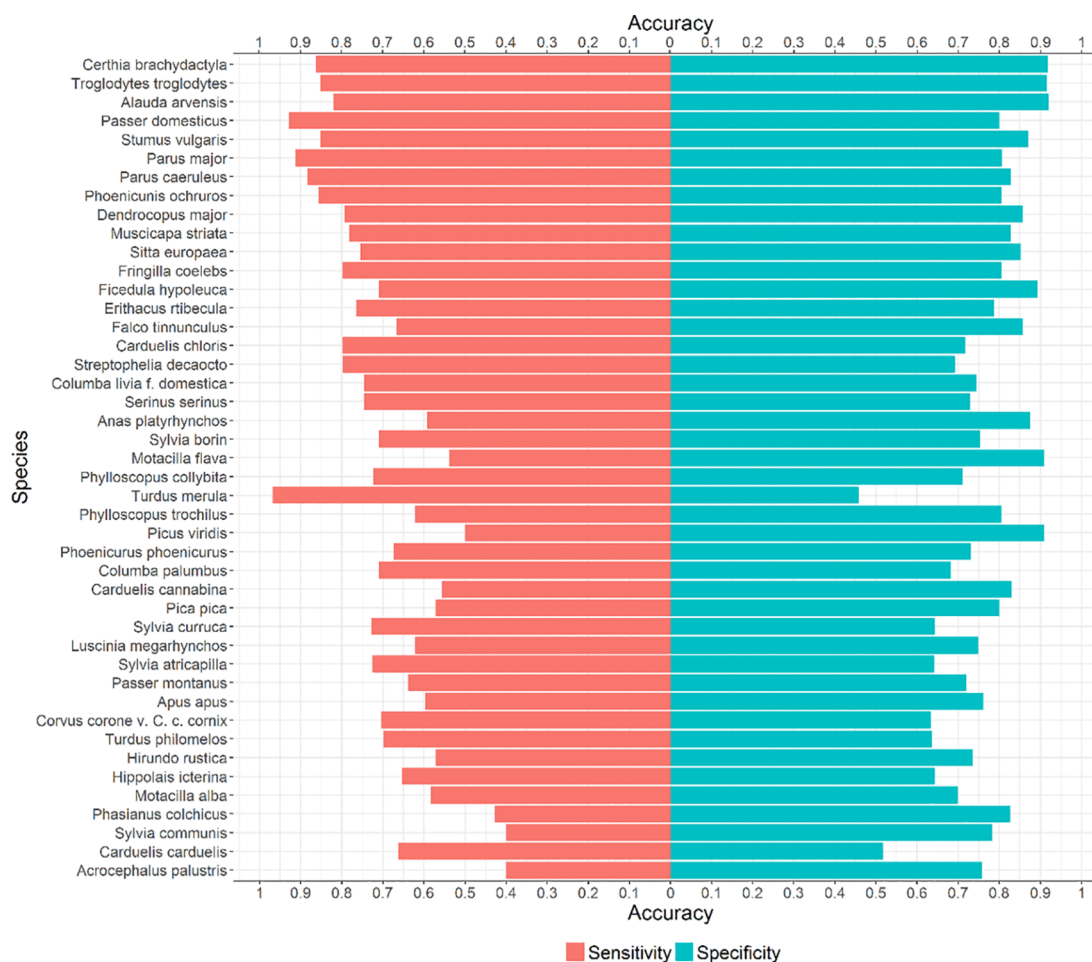


Fig. 4. Overview of the performance of the 44 models with respect to their sensitivity and specificity, ordered by their overall accuracy in the modelling process.

the testing data set, is 78%, with the best model featuring an overall accuracy of 90% and the worst 59% (Fig. 4 and Table A1 in the Appendix). The mean accuracy for predicting absences (representing species' specificity) is approximately 77%, while the prediction for presences (representing species' sensitivity) is about 70% (Table 2). Low sensitivity values can especially be found in models where down-sampling led to a strong reduction in modelling cases.

#### 4.1.2. Predictions of urban bird species distributions

The RF-models were used to predict the breeding sites of 44 urban bird species. The prediction is based on the independent testing dataset, containing 20% of the cells of the bird dataset. Fig. 5 shows the nesting patterns for the three different illustrative species introduced in Section 4.1. Overall, the breeding patterns for the illustrative species are well reproduced, with only minor misclassifications.

In Fig. 5 the sparrow is shown on the left. It becomes clear that the sparrow breeds across the whole study area except for the loam-rich riparian flood plain forest. The skylark is shown in the middle, it predominantly inhabits the agriculturally dominated peri-urban areas. Thirdly, the great spotted woodpecker is shown on the right. It breeds in forested areas, predominantly in the floodplain forest. Consequently, the three selected species demonstrate the versatility and capability of the presented RF-modelling approach to model and analyse vastly different habitats.

#### 4.1.3. Predictions of urban bird species richness

Using the TSS measure, an optimal clustering solution was found that encompassed five clusters (Fig. 6 – upper section). Based on clusters A, B and D, and a typical species for each of those, the results of the

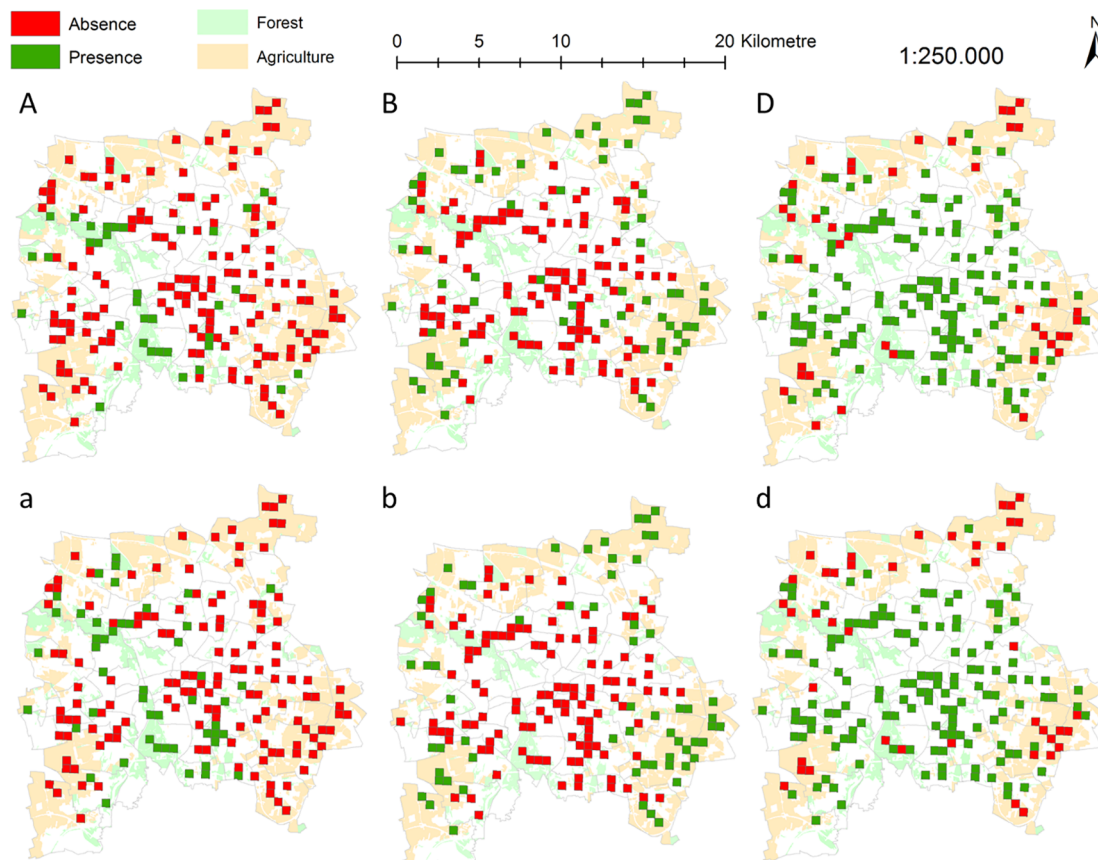
clustering will soon be illustrated. Cluster A features two species types, firstly cavity nesting birds breeding in tree holes like the great spotted woodpecker (*Dendrocopus major*) and the mallard (*Anas platyrhynchos*) breeding along the river embankments in the floodplain forest. Cluster B predominantly features ground-nesting birds, which can usually be found in agriculturally dominated surroundings on meadows and lawns like the Eurasian Skylark (*Alauda arvensis*). Cluster D features hemerophile species that have very broad geographic distributions across cities and show diverse breeding patterns. For instance, the Sparrow (*Passer domesticus*) breeds in holes in buildings but also in dense woody vegetation and is commonly found across the built-up area in Leipzig (BirdLife International, 2017).

By combining the 44 SDMs, patterns of urban bird species richness can be modelled. The most accurate results are generated for clusters A and D, which represent species in the forested areas and in the urban core respectively. This means that the model fulfils its purpose to cover the main urban bird species classes. The most inaccurate is cluster B, representing species breeding in the open landscapes in the peri-urban surrounding.

The overall bird species richness depicted in cluster F was found to feature an MAE of six. Therefore, it can be seen that species richness is underestimated in grid cells that feature comparably high numbers of present bird species and overestimated in areas where few to no birds are actually found.

#### 4.2. Variable importance of random forest models

This study introduces texture metrics as new indicators for the creation of species distribution models. Table 4 shows that the majority



**Fig. 5.** Predictions (in capital letters) and reference presence/absence data based on the testing data set (20% of all data points) for three bird species in Leipzig; the great spotted woodpecker (*Dendrocopos major*, A & a), the skylark (*Alauda arvensis*, B & b) and the sparrow (*Passer domesticus*, D & d). Naming is congruent with the belonging of the single species to the clusters presented in Fig. 6.

of species feature these metrics as their most important or second most important variable. In contrary, only five species feature the NDVI as their most important variable and thirteen a PCA. Since the majority of species feature both a classical remote sensing indicator and a texture metric as their two foremost variables, we can show the benefits both approaches can deliver when combining pixel-based and texture-based approaches.

The findings above support our previously outlined thesis, which stated that the structure in vegetation communities is of primary importance for bird habitat prediction. In contrast to this, the global spatial autocorrelation indicators which regard the whole cell rather than only in the local vegetation community, have low explanatory power as none of the species feature these measures as their most important variable.

Variable importance gives valuable insight into the functioning of the RF-model. This is shown in Fig. 7 where the interaction between the two foremost variables for two different bird species are shown. Variable relationships are highly non-linear and are indicative of the species presence/absence patterns only in specific parts of the variable range. Thus, we arrive at our second assertion: that flexible machine learning approaches are more feasible for such work and are more appropriate than many regression techniques, since they are free of prior assumptions and only work on the patterns in the data itself.

#### 4.3. Connecting the spectral trait concept to the traits of bird species

So far this study has modelled the presence/absence patterns of breeding birds, hence operating at the species level. In addition, the proposed methodology also provides the opportunity for connecting the spectral plant trait indicators directly to functional bird traits. This is

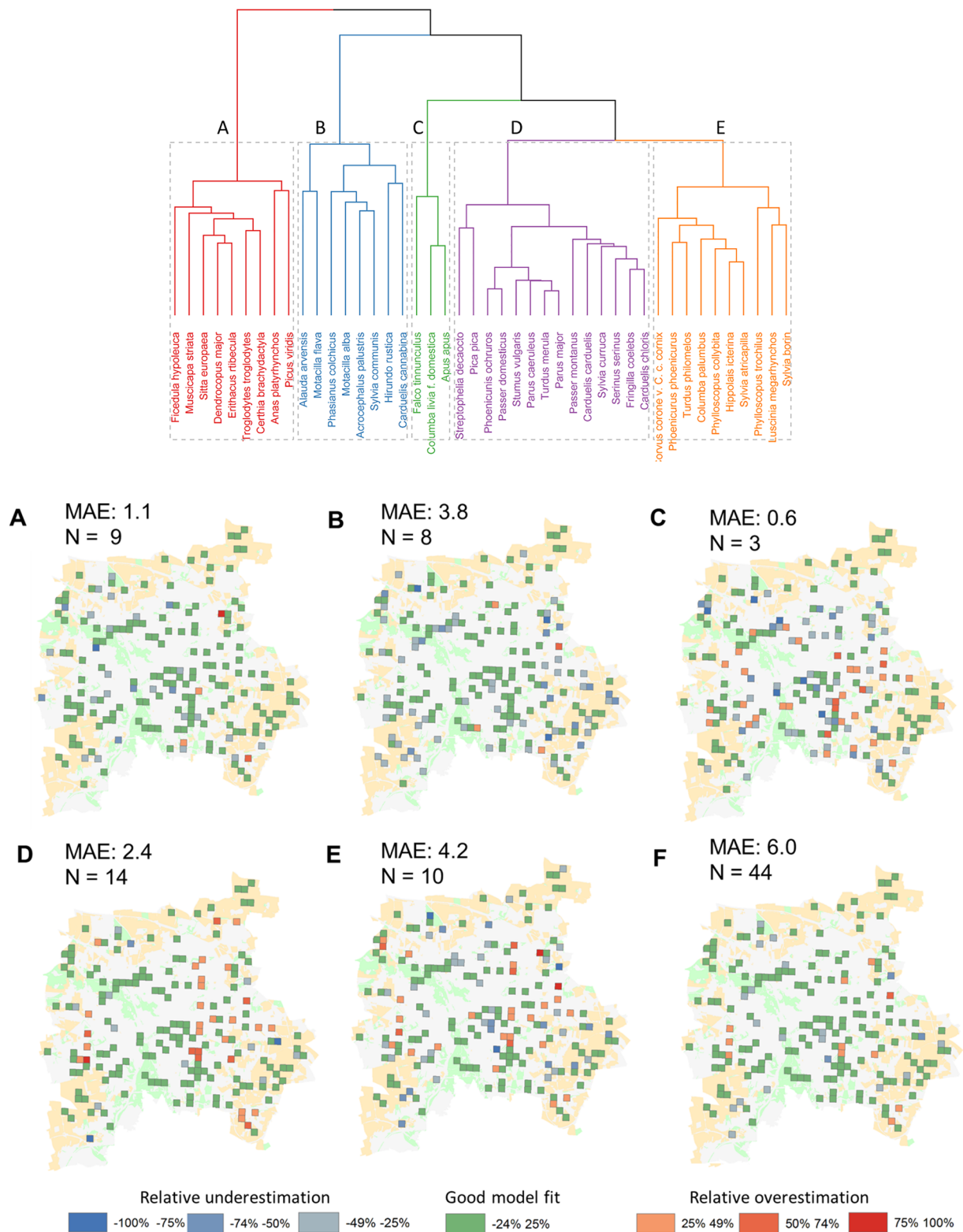
illustrated in Fig. 8, which shows that our two categories of indicators, GLCM texture metrics on the one hand and NDVI & PCA on the other, are indicative of a dietary trait.

Fig. 8 shows that our newly proposed indicators nicely correspond with a dietary species trait. This is the case if behavioural traits are closely proximate to the developed indicators in the feature space. Birds feeding on invertebrates can be associated with the GLCM texture metrics, hence the heterogeneity of local vegetation is most important for the presence of birds featuring this trait. In contrast, a diet based on plants is associated with the indicator NDVI. This means that for birds that are directly dependent on vegetation the amount of photosynthetic activity is most important for their presence/absence patterns. The first PCA band is associated with omnivores and the second PCA band with a diet based on seeds. This amplifies the finding that the pixel based approach, which describes the state of the vegetation and not its functional diversity, is most important for a diet that includes plant material.

## 5. Discussion

This study proposes a new approach to the integration of satellite-derived data for a more transferable, comparable and cost-efficient way to derive high-resolution SDMs. It does so by deriving indicators directly from continuous Earth observation data in order to reduce the deficiencies arising from pre-classified land cover/land use products. These indicators build on functional vegetation traits as crucial habitat variables for species modelling. Since previous studies predicting animal distribution patterns from space focussed on species richness and diversity (Rocchini et al., 2010), this study expands these efforts by introducing a species-specific approach.





**Fig. 6.** Models showing the relation between predicted and observed values for the five species clusters (A) to (E) and for all species (F) based on the testing data set. See Fig. 4 for the species included in each cluster.

With SDMs from EO data, new areas of environmental assessments come into reach that are highly relevant for both scientific and societal actors (Kerr and Ostrovsky, 2003). This could be species-specific conservation efforts, environmental impact assessments, or allocation of new construction developments in order to minimize environmental costs.

A key challenge for improving existing SDMs is the scarcity of spatially continuous high-resolution land-use/cover datasets, particularly in urban environments (He et al., 2015), because discrete EO data products are limited in extent and temporal and spatial resolution. Thus, they cannot show internal variability of classes or transition zones, or small linear elements (Lausch et al., 2015). This is especially

**Table 4**

Indicators covered in this study for the test for bird breeding habitat and their frequency of usage in the first two important variables in the 44 models.

Indicator type	Primary variable	Secondary variable
Texture measures	26	29
Global spatial autocorrelation	0	1
NDVI	5	4
PCA	13	10

problematic since the composition and structure of vegetation are the most important determinants for breeding sites of birds (Ikin et al., 2013; Paker et al., 2014). This study addresses this by integrating high-resolution RapidEye data that is already being used in studies of urban ecology (Wellmann et al., 2018).

This methodology will greatly benefit from the broad availability of current and future high-resolution EO datasets, for example, Planet Labs Doves (Wulder and Coops, 2014; Lausch et al., 2018). In comparison, lower resolution sensors like Landsat (30 m) or MODIS (250 m) are too coarse to discern any spatial diversity of species-specific habitat characteristics in the urban and peri-urban environment (Saveriaid et al., 2001; Goetz et al., 2007). While the need for higher spatial resolution applies to all ecosystem types, it very strongly relates to urban ecosystems where variation in the spatial dimension is very high due to the mix of built, sealed-green and blue surfaces (McPhearson et al., 2016).

Especially good SDMs were derived for birds breeding in typical urban settings, in parks, gardens or in densely built areas. Besides those hemerophile species, birds living in the alluvial forest were also modelled with high accuracy. We therefore conclude that our methodology is suitable for analysing both small- and large-scale vegetation patches of differing degrees of naturalness, opening up the possibility of transferring the methodology to areas outside of cities. In turn, the models for the open land species performed the worst. This can be partly attributed to the fact that open land species are the rarest species in the dataset. This linkage between the number of presence points and accuracy is at least partly related to the methodology as classification trees function best with larger sample sizes (Chen et al., 2004; Goetz et al., 2007). Therefore, working on larger study areas with more presence/absence data could benefit the results.

Relationships between the dependent and independent variables are highly diverse and inherently species specific. Consequently, a flexible machine learning approach that takes both presence and absence data into account would be advantageous. Overall, we tested five non-parametric machine learning techniques, in terms of their accuracy for the given task, and found that RF is most useful, but an ensemble model is most accurate. The versatility of RF-models for SDMs has been demonstrated for instance in Evans et al. (2011).

Regarding the predictors, we found that the NDVI is the second least important indicator after the measures of autocorrelation. The NDVI analyses only a single trait in the vegetation, i.e. its chlorophyll content

or degree of greenness, and is thus a poor indicator for overall functional diversity (Wang et al., 2003). However, studies that already incorporate EO data for species diversity or species richness analysis often rely on the NDVI without conducting sensitivity analysis (Seto et al., 2004; Goetz et al., 2007; Bino et al., 2008). Between the NDVI and species richness of different taxa varying indicator relationships were found (Bino et al., 2008), raising the question of reliability and transferability. In this light, a PCA seems more promising. When computed only in vegetated areas, PCA analysis represents a large variety in plant traits (Estes et al., 2010). In our case, the first two PCA bands explained 99% of the variance in the spectral backscatter. We found that the PCA-based predictors were the second most important. This suggests that more effort must be taken to analyse the degree of functional diversity in plant communities rather than analysing derivatives of photosynthesis capacity, as in the NDVI (Estes et al., 2010).

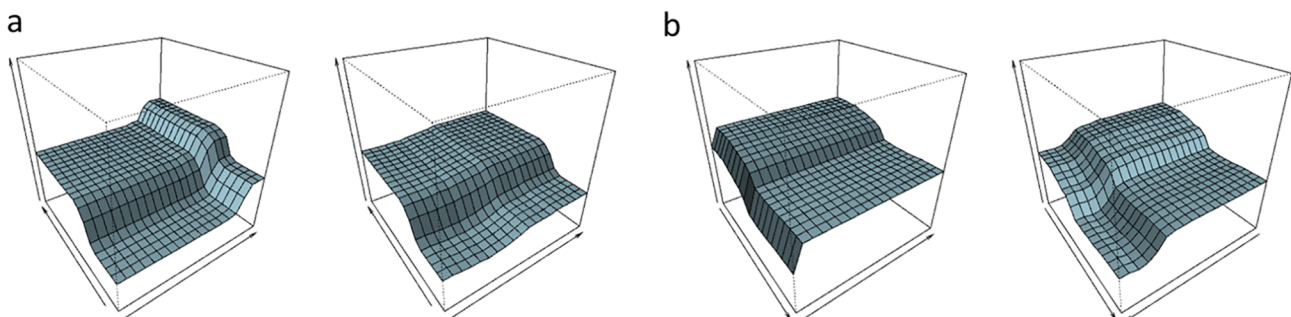
We facilitated our analysis of functional diversity by deploying texture metrics, which we find to be the most important indicator type. These indicators by Haralick et al. (1973) depict local spatial heterogeneity and are thus able to capture the diversity in functional plant characteristics. So far, this method has only been used rarely for species richness analysis (St-Louis et al., 2009; Estes et al., 2010) or in the urban context (Wellmann et al., 2018). This is unfortunate as modelling plant functional diversity from space is currently at the forefront of EO science (Jetz et al., 2016; Schneider et al., 2017; Kissling et al., 2018) but has not yet been adapted into SDM modelling.

Finally, we show that the concept of spectral traits also allows for future binding of remotely sensed characteristics to other species traits. This is a new and promising step for functional ecology and needs to be further evaluated in upcoming studies. Since there has been a large-scale diminishing of insectivorous birds across Europe recently, the modelling of birds featuring these traits is timely and needed and could help in evaluating and improving potential habitats (Bowler et al., 2019).

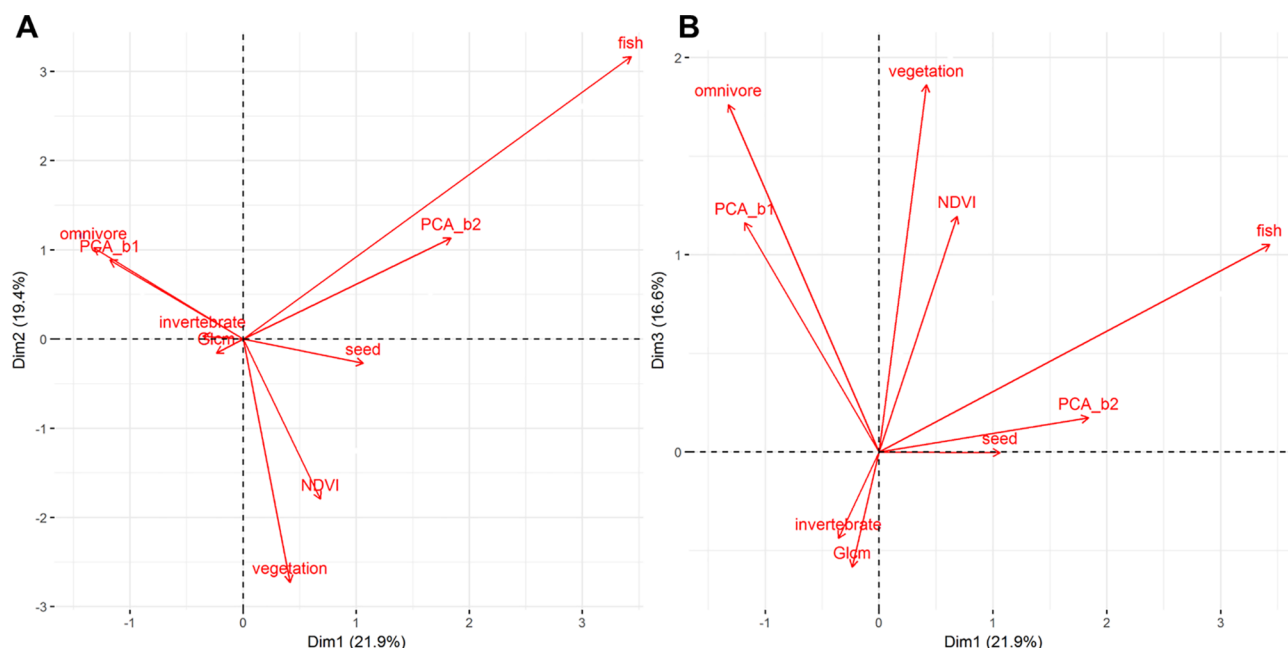
In the course of the study, datasets of vegetation and building heights (with 2 m resolution) and a soil map were added. These datasets only slightly improved model performance (1–2% on average). We therefore aimed for a less data-intensive approach by using only a single EO dataset as model input. This also means that the approach will be more transferable.

## 6. Conclusions

This study shows that satellite-derived vegetation parameters describing the composition and configuration of vegetation traits in a continuous way can play a crucial role in expanding the knowledge about species distribution patterns. Generally, results are promising and show that the usage of a single RapidEye scene paired with machine learning models can produce SDMs at high resolution and accuracy. Since the provisioning of suitable nesting grounds are key for the survival of a species, the adequate modelling of the breeding sites is very important. This is especially true because currently most cities do not



**Fig. 7.** Interactions between the two foremost important variables for (a) the Eurasian skylark (Two texture metrics) and (b) the great spotted woodpecker (PCA – Texture metric).



**Fig. 8.** Results of a Multi-Criteria Analysis (MCA) showing the relation between the proposed indicators with dietary traits of the 44 bird species for the first two dimensions in A and the first and third dimension in B, covering the major part of variability in the whole data set.

provide resources for additional bird monitoring covering the entire city.

For the preservation and extension of urban biodiversity spatially explicit data in high-resolution is necessary for well-informed land and green-space management. This study provides a window of opportunity for a better understanding of coupled human-environmental systems in the city, by exploring the effects of vegetation diversity and structure on the breeding behaviour of urban birds. This will ultimately help to adapt land management schemes or to steer urban development such that bird-breeding sites are minimally affected or might even benefit from new and/or existing constructions.

#### CRediT authorship contribution statement

**Thilo Wellmann:** Conceptualization, Methodology, Software, Writing - original draft, Data curation, Formal analysis. **Angela Lausch:** Supervision, Methodology. **Sebastian Scheuer:** Software, Writing - review & editing. **Dagmar Haase:** Supervision, Writing - review & editing.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

This research was carried out as part of the project ENABLE, funded through the 2015–2016 BiodivERsA COFUND call for research proposals, with the national funders The Swedish Research Council for Environment, Agricultural Sciences, and Spatial Planning, Swedish Environmental Protection Agency, German Aeronautics and Space Research Centre, National Science Centre (Poland), The Research Council of Norway and the Spanish Ministry of Economy and Competitiveness. We further wish to thank the Horizon 2020 innovation action CONNECTING (No 730222-2) and our colleagues Ralf Seppelt, Fernando Castillo and Manuel Wolff for fruitful discussion. T.W. receives a Humboldt research-track scholarship.

#### Data accessibility

The final modelling dataset and the fractional vegetation cover product for Leipzig can be assessed via: <http://doi.org/10.5281/zenodo.3597379>.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecolind.2019.106029>.

#### References

- Belgiu, M., Dragu, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- Bernard, S., Heutte, L., Adam, S., 2009. Influence of hyperparameters on random forest accuracy. In: *International Workshop on Multiple Classifier Systems*, pp. 171–180.
- Bino, G., Levin, N., Darawshi, S., Van Der Hal, N., Reich-Solomon, A., Kark, S., 2008. Accurate prediction of bird species richness patterns in an urban environment using Landsat-derived NDVI and spectral unmixing. *Int. J. Remote Sens.* 29 (13), 3675–3700. <https://doi.org/10.1080/01431160701772534>.
- BirdLife International, 2017. The IUCN Red List of Threatened Species. Retrieved 11 April 2018, from <http://www.iucnredlist.org>.
- Bowler, D., Heldbjerg, H., Fox, A., de Jong, M., Böhning-Gaese, K., 2019. Long-term declines of European insectivorous bird populations and potential causes. *Conserv. Biol.*
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Chace, J., Walsh, J., 2006. Urban effects on native avifauna: a review. *Landscape Urban Plann.* 74 (1), 46–69. <https://doi.org/10.1016/j.landurbplan.2004.08.007>.
- Chen, C., Liaw, A., Breiman, L., 2004. Using random forest to learn imbalanced data. *Univ. California, Berkeley* 110, 1–12.
- Cutler, D., Edwards, T., Beard, K., Cutler, A., Hess, K., Gibson, J., Lawler, J., 2007. Random forests for classification in ecology. *Ecology* 88 (11), 2783–2792. <https://doi.org/10.1890/07-0539.1>.
- Datt, B., 1998. Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a + b, and total carotenoid content in eucalyptus leaves. *Remote Sens. Environ.* 66 (2), 111–121. [https://doi.org/10.1016/S0034-4257\(98\)00046-7](https://doi.org/10.1016/S0034-4257(98)00046-7).
- Dormann, C., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36 (1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Estes, L., Reillo, P., Mwangi, A., Okin, G., Shugart, H., 2010. Remote sensing of structural complexity indices for habitat and species distribution modeling. *Remote Sens. Environ.* 114 (4), 792–804. <https://doi.org/10.1016/j.rse.2009.11.016>.
- Evans, J., Murphy, M., Holden, Z., Cushman, S., 2011. Modeling species distribution and

- change using random forest. In: Predictive species and habitat modeling in landscape ecology. Springer, pp. 139–159.
- Falcucci, A., Maiorano, L., Boitani, L., 2007. Changes in land-use/land-cover patterns in Italy and their implications for biodiversity conservation. *Landscape Ecol.* 22 (4), 617–631. <https://doi.org/10.1007/s10980-006-9056-4>.
- Gamon, J., Field, C., Goulden, M., Griffin, K., Hartley, A., Joel, G., Valentini, R., 1995. Relationships between NDVI, canopy structure, and photosynthesis in three Californian vegetation types. *Ecol. Appl.* 5 (1), 28–41. <https://doi.org/10.2307/1942049>.
- Geary, R., 1954. The contiguity ratio and statistical mapping. *Incorporated Stat.* 5 (3), 115–146.
- Gedeon, K., Mitschke, A., & Sudfeldt, C., 2004. Brutvögel in Deutschland. Stiftung Vogelmonitoring Deutschland. Retrieved from [http://www.dda-web.de/downloads/texts/adebar/adebar1\\_komplett.pdf](http://www.dda-web.de/downloads/texts/adebar/adebar1_komplett.pdf).
- Goetz, S., Steinberg, D., Dubayah, R., Blair, B., 2007. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sens. Environ.* 108 (3), 254–263. <https://doi.org/10.1016/j.rse.2006.11.016>.
- Greenacre, M., Blasius, J., 2006. Multiple Correspondence Analysis and Related Methods. Chapman and Hall/CRC.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8 (9), 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>.
- Haase, D., Jänicke, C., Wellmann, T., 2019. Front and back yard green analysis with subpixel vegetation fractions from earth observation data in a city. *Landscape Urban Plann.* <https://doi.org/10.1016/j.landurbplan.2018.10.010>.
- Haase, D., Larondelle, N., Andersson, E., Artmann, M., Borgström, S., Breuste, J., Elmqvist, T., 2014. A quantitative review of urban ecosystem service assessments: concepts, models, and implementation. *Ambio* 43 (4), 413–433. <https://doi.org/10.1007/s13280-014-0504-0>.
- Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* <https://doi.org/10.1109/TSMC.1973.4309314>.
- He, K., Bradley, B., Cord, A., Rocchini, D., Tuanmu, M.N., Schmidlein, S., Pettorelli, N., 2015. Will remote sensing shape the next generation of species distribution models? *Remote Sens. Ecol. Conserv.* 1 (1), 4–18. <https://doi.org/10.1002/rse2.7>.
- Ikin, K., Beaty, M., Lindenmayer, D., Knight, E., Fischer, J., Manning, A., 2013. Pocket parks in a compact city: how do birds respond to increasing residential density? *Landscape Ecol.* 28 (1), 45–56. <https://doi.org/10.1007/s10980-012-9811-7>.
- Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F., Asner, G., Ustin, S., 2016. Monitoring plant functional diversity from space. *Nat. Plants* 2 (3), 1–5. <https://doi.org/10.1038/nplants.2016.24>.
- Jolliffe, I., 2002. Principal Component Analysis. John Wiley Sons Ltd.
- Kattge, J., Díaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönsch, G., Wirth, C., 2011. TRY – a global database of plant traits. *Glob. Change Biol.* 17 (9), 2905–2935. <https://doi.org/10.1111/j.1365-2486.2011.02451.x>.
- Kerr, J., Ostrovsky, M., 2003. From space to species: Ecological applications for remote sensing. *Trends Ecol. Evol.* 18 (6), 299–305. [https://doi.org/10.1016/S0169-5347\(03\)00071-5](https://doi.org/10.1016/S0169-5347(03)00071-5).
- Kissling, D., Walls, R., Bowser, A., Jones, M., Kattge, J., Agosti, D., Guralnick, R., 2018. Towards global data products of Essential Biodiversity Variables on species traits. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-018-0667-3>.
- Knapp, S., Winter, M., Klotz, S., 2017. Increasing species richness but decreasing phylogenetic richness and divergence over a 320-year period of urbanization. *J. Appl. Ecol.* 1152–1160. <https://doi.org/10.1111/1365-2664.12826>.
- Kuhn, M., 2008. Caret package. *J. Stat. Softw.* 28 (5), 1–26.
- Lausch, A., Blaschke, T., Haase, D., Herzog, F., Syrbe, R.U., Tischendorf, L., Walz, U., 2015. Understanding and quantifying landscape structure – a review on relevant process characteristics, data models and landscape metrics. *Ecol. Model.* 295, 31–41. <https://doi.org/10.1016/j.ecolmodel.2014.08.018>.
- Lausch, A., Borg, E., Bumberger, J., Dietrich, P., Heurich, M., Huth, A., Schaepman, M., 2018. Understanding forest health with remote sensing, Part III: requirements for a scalable multi-source forest health monitoring network based on data science approaches. *Remote Sens.* 10 (7). <https://doi.org/10.3390/rs10071120>.
- Lausch, A., Erasmi, S., King, D., Magdon, P., Heurich, M., 2016. Understanding forest health with remote sensing – part I—a review of spectral traits, processes and remote-sensing characteristics. *Remote Sens.* 1029 (8), 1–43. <https://doi.org/10.3390/rs8121029>.
- McGill, B.J., Enquist, B.J., Weiher, E., Westoby, M., 2006. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* 21 (4), 178–185. <https://doi.org/10.1016/j.tree.2006.02.002>.
- McPhearson, T., Haase, D., Kabisch, N., Gren, Å., 2016. Advancing understanding of the complex nature of urban systems. *Ecol. Ind.* 70, 566–573. <https://doi.org/10.1016/j.ecolind.2016.03.054>.
- Mitchell, M., 2011. Bias of the random forest out-of-bag (OOB) error for certain input parameters. *Open J. Stat.* 01 (03), 205–211. <https://doi.org/10.4236/ojs.2011.13024>.
- Moran, P., 1950. Notes on continuous stochastic phenomena. *Biometrika* 37 (1), 17–23.
- Paker, Y., Yom-Tov, Y., Alon-Mozes, T., Barnea, A., 2014. The effect of plant richness and urban garden structure on bird species richness, diversity and community structure. *Landscape Urban Plann.* 122, 186–195. <https://doi.org/10.1016/j.landurbplan.2013.10.005>.
- Palmer, M., Earls, P., Hoagland, B., White, P., Wohlgemuth, T., 2002. Quantitative tools for perfecting species lists. *Environmetrics* 13 (2), 121–137. <https://doi.org/10.1002/env.516>.
- Podani, J., Schmera, D., 2011. A new conceptual and methodological framework for exploring and explaining pattern in presence – absence data. *Oikos* 120 (11), 1625–1638. <https://doi.org/10.1111/j.1600-0706.2011.19451.x>.
- R Core Team, 2000. R Language Definition. R Foundation for Statistical Computing, Vienna, Austria.
- Richter, R., 2011. Atmospheric / Topographic Correction for Satellite Imagery (ATCOR - 2/3 User Guide). DLR-German Aerospace Center, Weßling, Germany.
- Rocchini, D., Balkenhol, N., Carter, G., Foody, G., Gillespie, T., He, K.S., Neteler, M., 2010. Remotely sensed spectral heterogeneity as a proxy of species diversity: recent advances and open challenges. *Ecol. Inf.* 5 (5), 318–329. <https://doi.org/10.1016/j.ecoinf.2010.06.001>.
- Rocchini, D., Luque, S., Pettorelli, N., Bastin, L., Doktor, D., Faedi, N., Nagendra, H., 2017. Measuring  $\beta$ -diversity by remote sensing: a challenge for biodiversity monitoring. *Methods in Ecology and Evolution*.
- Rousseeuw, P., Leroy, A., 2005. Robust Regression and Outlier Detection, vol. 589 John Wiley & Sons.
- Saverail, E., Debinski, D., Kindscher, K., Jakubauskas, M., 2001. A comparison of satellite data and landscape variables in predicting bird species occurrences in the Greater Yellowstone Ecosystem, USA, (1992) 71–83. doi:10.1023/A:1008119219788.
- Schneider, F., Morsdorf, F., Schmid, B., Petchey, O., Hueni, A., Schimel, D., Schaepman, M., 2017. Mapping functional diversity from remotely sensed morphological and physiological forest traits. *Nat. Commun.* 8 (1). <https://doi.org/10.1038/s41467-017-01530-3>.
- Seto, K., Fleishman, E., Fay, J., Betrus, C., 2004. Linking spatial patterns of bird and butterfly species richness with Landsat TM derived NDVI. *Int. J. Remote Sens.* 25 (20), 4309–4324.
- Sibly, R., Witt, C., Wright, N., Venditti, C., Jetz, W., Brown, J., 2012. Energetics, lifestyle, and reproduction in birds. *Proc. Natl. Acad. Sci.* 109 (27), 10937–10941. <https://doi.org/10.1073/pnas.1206512109>.
- St-Louis, V., Pidgeon, A., Clayton, M., Locke, B., Bash, D., Radeloff, V., 2009. Satellite image texture and a vegetation index predict avian biodiversity in the Chihuahuan Desert of New Mexico. *Ecography* 32 (3), 468–480. <https://doi.org/10.1111/j.1600-0587.2008.05512.x>.
- Leipzig, Stadt, n.d. Die Stadtforstverwaltung der Stadt Leipzig. Retrieved 10 June 2019, from [https://www.leipzig.de/fileadmin/mediendatenbank/leipzig-de/Stadt/02.3\\_Dez3\\_Umwelt.Ordnung.Sport/67\\_Amt\\_fuer\\_Stadtruen\\_und\\_Gewaesser/Stadtwald\\_und\\_Auenwald/Informationen\\_zum\\_Stadtwald\\_und\\_Auenwald/Allgemeines\\_zum\\_Leipziger\\_Stadtwald.pdf](https://www.leipzig.de/fileadmin/mediendatenbank/leipzig-de/Stadt/02.3_Dez3_Umwelt.Ordnung.Sport/67_Amt_fuer_Stadtruen_und_Gewaesser/Stadtwald_und_Auenwald/Informationen_zum_Stadtwald_und_Auenwald/Allgemeines_zum_Leipziger_Stadtwald.pdf).
- Leipzig, Stadt, 2018. Baumkataster und Statistik. Retrieved 10 June 2019, from <https://www.leipzig.de/umwelt-und-verkehr/umwelt-und-naturschutz/baeume-und-baumschutz/stadtbaeume/baumkataster-und-statistik/>.
- StUfa, 1995. (Staatliches Umweltfachamt im Freistaat Sachsen) Brutvogelatlas der Stadt und des Landkreises Leipzig. Materialien zu Naturschutz und Landschaftspflege. – Leipzig.
- Südbeck, P., 2005. Methodenstandards zur Erfassung der Brutvögel Deutschlands. Max-Planck-Inst. für Ornithologie, Vogelwarte Radolfzell.
- Tigges, J., Lakes, T., Hostert, P., 2013. Urban vegetation classification: benefits of multitemporal RapidEye satellite data. *Remote Sens. Environ.* 136, 66–75. <https://doi.org/10.1016/j.rse.2013.05.001>.
- Violle, C., Navas, M., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., Garnier, E., 2007. Let the concept of trait be functional!. *Oikos* 116 (5), 882–892. <https://doi.org/10.1111/j.2007.0030-1299.15559.x>.
- Völkl, W., Blick, T., Kornacker, P., Martens, H., 2004. Quantitativer Überblick über die rezente Fauna von Deutschland. Natur Und Landschaft.
- Wang, Z., Liu, C., Huete, A., 2003. From AVHRR-NDVI to MODIS-EVI: advances in vegetation index research. *Acta Ecol. Sin.* 23 (5), 979–987.
- Warton, D.I., Shipley, B., Hastie, T., 2015. CATS regression – a model-based approach to studying trait-based community assembly. *Methods Ecol. Evol.* 6 (4), 389–398. <https://doi.org/10.1111/2041-210X.12280>.
- Wellmann, T., Haase, D., Knapp, S., Salbach, C., Selsam, P., Lausch, A., 2018. Urban land use intensity assessment: the potential of spatio-temporal spectral traits with remote sensing. *Ecol. Ind.* 85, 190–203. <https://doi.org/10.1016/j.ecolind.2017.10.029>.
- Wolff, M., Haase, A., Haase, D., Kabisch, N., 2016. The impact of urban regrowth on the built environment. *Urban Stud.* <https://doi.org/10.1177/0042098016658231>.
- Wulder, M., Coops, N., 2014. Satellites: Make Earth observations open access. *Nature* 513 (7516), 30–31. <https://doi.org/10.1038/513030a>.