

This is the accepted manuscript version of the contribution published as:

Wellmann, T., Lausch, A., Scheuer, S., Haase, D. (2020):

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

Ecol. Indic. **111** , art. 106029

The publisher's version is available at:

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

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

Thilo Wellmann ^{1,2,*}, Angela Lausch ^{1,2}, Sebastian Scheuer ¹, Dagmar Haase ^{1,2}

¹ Department of Geography, Humboldt Universität zu Berlin, Rudower Chaussee 16, D-12489 Berlin

² Department of Computational Landscape Ecology, Helmholtz Centre for Environmental Research—UFZ, Permoserstr. 15, D-04318 Leipzig

* thilo.wellmann@hu-berlin.de

4

5 This is the accepted manuscript.

6 For the published original paper see; DOI: <https://doi.org/10.1016/j.ecolind.2019.106029>

7 This manuscript version is made available under the CC-BY-NC-ND 4.0 license

8 <http://creativecommons.org/licenses/by-nc-nd/4.0/>

9

10

11 Abstract

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

17 The deployed concepts encompass vegetation fractions for determining vegetation density
18 and spectral traits for the quantification of vegetation heterogeneity. Both indicators are

19 derived from RapidEye data, thus featuring a continuous spatial resolution of 6.5 m. Using
20 these indicators as predictors, we model breeding bird habitats using a random forest (RF)
21 classifier for the city of Leipzig, Germany using a single EO image.

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

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

32

33 **Keywords:**

34 Remote sensing, spectral traits, species distribution model, random forest, urban birds,
35 machine learning

36

37 1 Introduction

38 Modelling potential breeding sites that are species-specific can be an integral part of urban,
39 peri-urban and non-urban biodiversity studies and conservation strategies (Guisan & Thuiller,
40 2005). The urban environment is especially rich in birds, often surpassing their rural
41 surroundings in terms of biomass and diversity (Chace & Walsh, 2006). A key element of
42 sustaining viable population sizes of single species under threat or increasing overall species
43 richness and abundance is the identification and protection of breeding sites. A core element
44 determining the breeding sites of birds is vegetation structure such as vegetation density and
45 diversity (Paker et al., 2014). Earth observation datasets provide a cost-effective, reproducible
46 and straightforward method for the analysis of such vegetation parameters.

47 Satellite-derived information has been widely used to predict species richness, diversity and
48 turnover in a variety of kingdoms (Rocchini et al., 2010, 2017). While the analysis of such
49 diversity parameters is valuable (Rocchini et al., 2010), those analyses lack species-specific
50 information. For multiple use cases such as species protection measures or environmental
51 impact assessments, species distribution models (SDMs) are needed (Guisan & Thuiller, 2005).
52 However, there is a clear lack in SDMs since existing models have two major problems
53 regarding the characteristics of input data and modelling technique.

54 Regarding modelling techniques, studies often use regression (Bino et al., 2008; Warton et al.,
55 2015). Due to the model assumptions inherent to most regression methods, problems such as
56 collinearity between predictors, outliers or non-linear and exponential relationships may
57 result in bad model performance (Rousseeuw & Leroy, 2005; Dormann et al., 2013). Thus, to
58 overcome the aforementioned limitations, a more flexible machine learning approach seems
59 favourable for SDMs. One particularly robust and well-established procedure in ecology and
60 EO studies is thereby the RF-algorithm (Cutler et al., 2007; Belgiu & Drăgu, 2016), which is an
61 ensemble learning method consisting of a multitude of decision trees (Breiman, 2001). RFs are
62 able to deal with highly collinear predictors that can be both quantitative (numeric) and
63 qualitative (non-numeric) with all kinds of variable interactions, making them, therefore, often
64 superior to regression.

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

71 Urban environments are dynamic and complex and, within them, sites of high biodiversity can
72 be found next to intensely managed ones (Haase et al., 2014; Knapp et al., 2017). This species
73 richness, however, seems to be in danger as recent reports state that multiple species in
74 Europe (Bowler et al., 2019) and also in Germany (Gedeon et al., 2004) are in rapid decline.
75 This trend is especially apparent for bird species breeding in urban and agricultural settings,
76 since those feature the most rapid reduction of all regarded habitat types (Gedeon et al.,
77 2004). Since the case study area of this paper, the city of Leipzig, Germany, is characterised by
78 a dense centre with vast parks as well as a large natural forest and fertile agrarian
79 surroundings, it is an ideal case study for developing models for those endangered species
80 groups and also for the large group of forest birds (Wellmann et al., 2018).

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

93 The spectral trait concept is a functional approach in which every plant trait corresponds to a
94 function, that is relevant for (i) the plant and (ii) the larger ecosystem (Violle et al., 2007).

95 Therefore, the spectral traits approach is an efficient interface linking EO data to key
96 ecosystem characteristics, functions and services (Lausch et al., 2016), which in return could
97 be linked to bird species breeding behaviour.

98 One way of analysing the spatial diversity of spectral traits in a plant community is by
99 quantifying the composition and configuration of a plant trait related product, e.g.,
100 Normalized Difference Vegetation Index (NDVI), in space and over time (Wellmann et al.,
101 2018). For this, texture measures by Haralick et al. (1973), such as the grey level co-occurrence
102 matrix (GLCM), are powerful and well established methods used by St-Louis (2009) for the
103 prediction of bird species diversity.

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

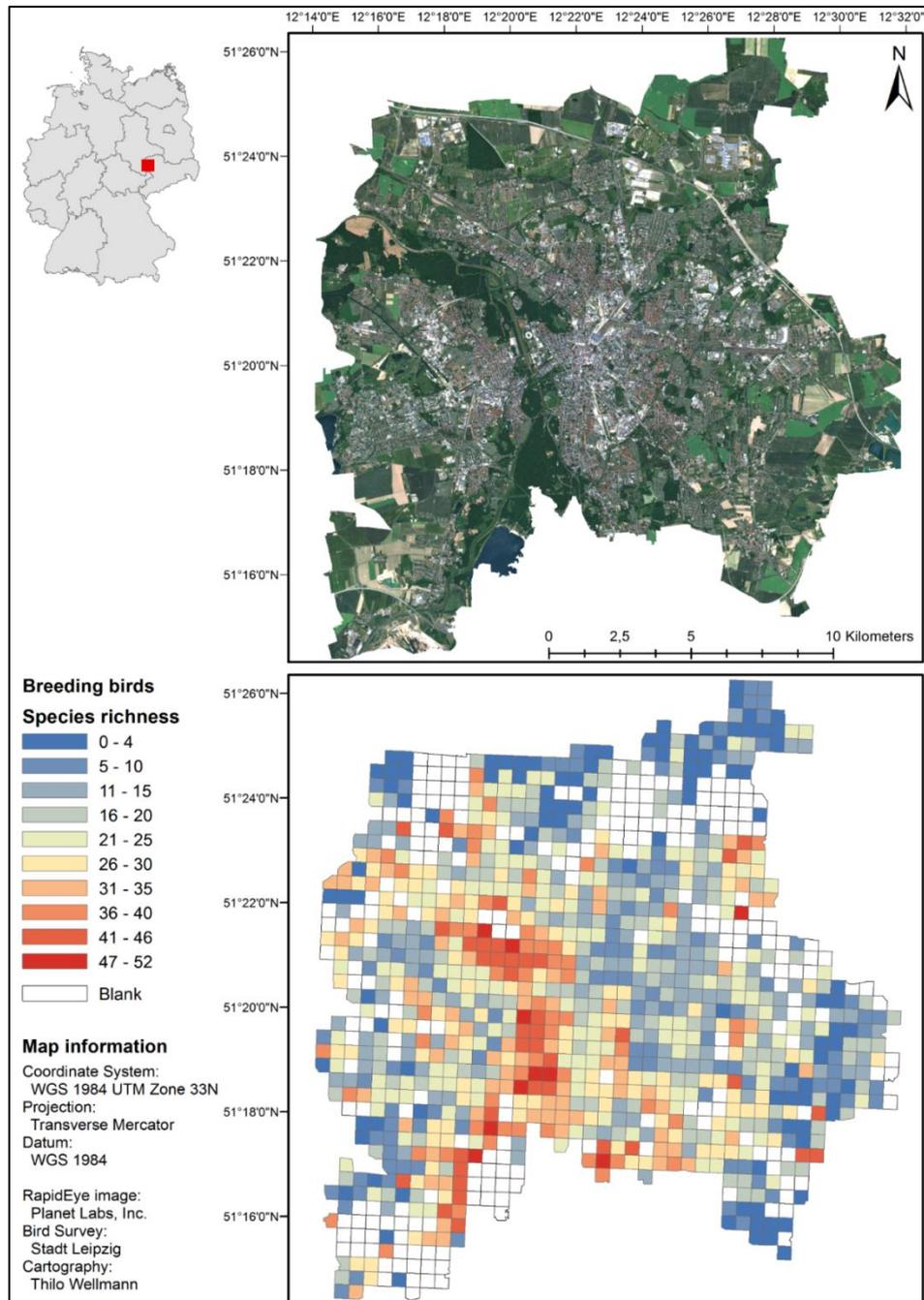
- 110 (i) Are fractional vegetation cover and spectral plant traits meaningful indicators for the
111 prediction of breeding sites for species in the urban environment?
- 112 (ii) What are suitable modelling techniques?
- 113 (iii) How accurate are SDMs solely derived from EO data?
- 114 (iv) How do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

115

116 2 Study area

117 Leipzig is a dense city in Eastern Germany located at 51°20'N, 12°22'E with 560,000
118 inhabitants. The city houses a considerable number of natural biotopes and breeding-bird
119 species richness is comparably high (Figure 1). Almost 40% of all bird species breeding in
120 Germany (n = 314) can also be found in Leipzig (n= 120) (StUfa, 1995; Völkl et al., 2004).
121 Important breeding grounds are located along a north to south transect in the large remnants
122 of the alluvial forest on the floodplains. This forest is one of the largest of its kind in Europe

123 and features a quasi-natural structure in terms of species composition, which is dominated by
 124 ash, oak, beech, lime and sycamore trees. Next to forested areas, different types of urban
 125 greenspaces, urban building structures, and permanent agricultural systems in the
 126 surrounding may provide rich breeding grounds (StUfa, 1995) (Figure 1).



127
 128 Figure 1. RapidEye satellite image of the city of Leipzig with indication of the study area location in
 129 Germany and the breeding bird dataset showing bird species richness for all parts of the city but the
 130 areas that were excluded due to large-scale change in biotope composition between the acquisition
 131 of the RapidEye scene and the bird survey

132

133 The development of Leipzig since the German reunification in 1990 can be quickly
134 summarized. In the observed period, Leipzig's population declined slightly leading to a stable
135 stock in central buildings. Suburbanisation tendencies on peri-urban agricultural land in the
136 northern outskirts lead to large-scale developments in the outskirts, mostly consisting of
137 logistic infrastructure, industrial facilities or the exhibition grounds. These developments
138 mostly occurred outside of important nesting habitats for breeding birds, since the high-
139 intensity farming during the socialist past until 1990 left few ecological niches. A second
140 important trend in the southern outskirts of the city are the flooding of former opencast mines
141 (Wolff et al., 2016).

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

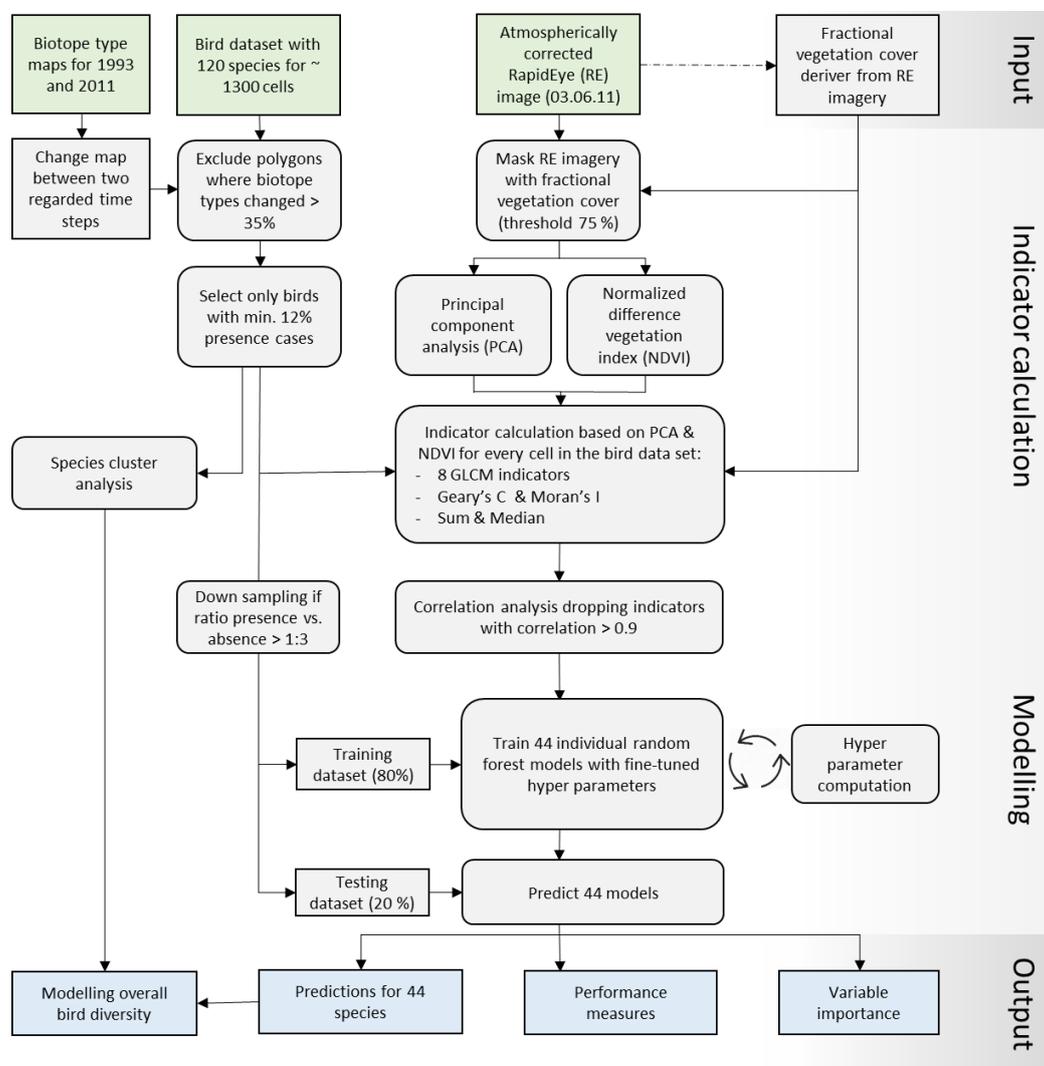
152 In summarizing the findings above, we see that, while the structure of the central areas
153 remained very stable, there was considerable change at the outskirts of the city. Figure 1
154 shows where the change was assessed with two biotope maps, the first from 1993 and the
155 second from 2005.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

159

160 3 Data and Methods

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

168



169

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

172

173 3.1 Earth observation (EO) data and indicator calculation

174 *3.1.1 Pre-processing of RapidEye data*

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

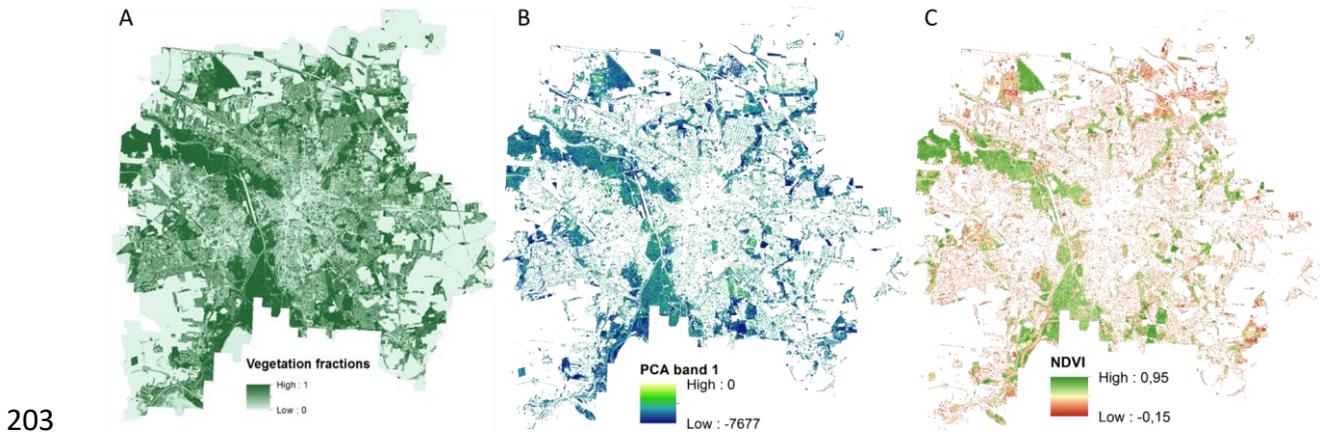
183 Based on the atmospherically corrected RapidEye scene, a map of fractional vegetation cover
184 was calculated (Figure 3) (Haase et al., 2019). This dataset provides subpixel information about
185 the share of vegetation featured in each pixel. It was used to mask out areas in the original
186 RapidEye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

193

194 *3.1.2 Calculation of functional vegetation indicators*

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

201 capacity for performing photosynthesis and, respectively, to the traits of the plants that are
202 associated with photosynthesis (Gamon et al., 1995).



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

207

208 Based on the three products (NDVI, PCA band 1 & band 2) described above, 13 different
209 indicators that describe both state and spatial heterogeneity in the plant communities are
210 calculated for every 500 m by 500m cell (Table 1). The GLCM indicators by Haralick et al. (1973)
211 assess local image texture by calculating the differences between the values of adjacent pixels
212 values, e.g. NDVI value. Based on this frequency matrix, eight indicators were calculated and
213 their mean and standard deviation values assessed for every cell. Furthermore, two measures
214 of spatial autocorrelation and three summary statistics were calculated.

215

216 Table 1. Overview of the indicator types expressing vegetation density or vegetation heterogeneity.
 217

Type	Name	Reference
Local spatial autocorrelation	GLCM mean	
	GLCM variance	
	GLCM correlation	
	GLCM homogeneity	(Haralick et al., 1973)
	GLCM contrast	
	GLCM dissimilarity	
	GLCM entropy	
Global spatial autocorrelation	GLCM angular second	
	Geary's <i>C</i>	(Geary, 1954)
	Moran's <i>I</i>	(Moran, 1950)
Descriptive statistics	Standard Deviation	
	Coefficient of Variation	(Datt, 1998)
	Sum	

218

219 Since we derived all indicators used for modelling from the same RapidEye dataset, a
 220 correlation analysis was performed using a cut-off value of 0.9. The comparably high cut-off
 221 value was chosen because the RF algorithm is able to deal with highly collinear data (Breiman,
 222 2001).

223

224 3.2 Bird species distribution data and further analysis

225 3.2.1 The bird survey dataset and its pre-processing

226 The breeding bird data used in this study was collected in the city of Leipzig over three
 227 breeding periods from 1991 to 1993 between February and July. It describes the presence or
 228 absence of 120 species within 1132 cells with a resolution of 500m by 500 m. Each of these
 229 cells was surveyed at least 5 times per year by ornithologists to map species that were
 230 breeding. The species was marked as present in a cell if it was observed at least once. Through
 231 this scheme, very reliable presence and absence data could be generated (StUfa, 1995).

232 The dataset was treated to account for uncertainties in the occurrence of bird species. Firstly,
 233 only validated species sightings were included and, subsequently, all entries marked as

234 uncertain were disregarded. Secondly, rare species had to be excluded to ensure that a
235 sufficient amount of presence points remained for validation of the proposed RF-model. As a
236 threshold, a presence to absence ratio greater than 10% was chosen. Due to the exclusion of
237 the rare species, only 44 of the 120 species remained for analysis.

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

244

245 *3.2.2 Determination of functional species communities with a cluster analysis*

246 A hierarchical cluster analysis was performed in order to find groups of bird species that are
247 similar to each other in terms of their presence/absence patterns in Leipzig, to identify
248 coexisting species. The HCLUST (R Core Team, 2000) algorithm used in this study iteratively
249 assigns an object to a cluster based on a distance measure. The allocation of group
250 memberships is executed on the premise of minimizing the distance between the clusters
251 members. Distance in this study was measured by the Jaccard distance (1), a statistical
252 measure computing the dissimilarity between sample sets (Podani & Schmera, 2011):

$$253 \quad d_j(A, B) = 1 - J(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}, \quad (1)$$

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

256

257

258

259

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

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

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

265 3.3 Machine-learning based species distribution modelling techniques

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

271 Table 2. Final decision matrix on choosing the modelling technique, the green shading indicates the
 272 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	

274

275 3.3.1 Random forest modelling and hyperparameter refinement

276 RF is an algorithm capable of solving regression and classifications problems (Breiman, 2001),
 277 providing fast model training and comparably high accuracies (Mitchell, 2011). In this study,
 278 44 random forests are grown—one for every species—to predict species presence/absence
 279 pattern with an independent test data set.

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

285 3.3.2 Downsampling for bird species distribution modelling

286 For 32 of the 44 bird species, absence points outnumber presence points. Since highly
 287 unbalanced data can cause problems in the random forest classification, a downsampling
 288 approach for the species absences was used (Chen et al., 2004). In this study, rare species
 289 were downsampled, meaning that absence points were disregarded until a ratio of 1:3
 290 between presence and absence points was reached. This ratio was found to produce the most
 291 accurate results overall. Downsampling in this study is regarded as a very critical and
 292 important step which needs to be carefully addressed and iteratively tested.

293

294 3.4 Accuracy metrics

295 Table 3. Confusion matrix and formulas for accuracy indicators

296

		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)	

297 A = True Positives; B = False Negatives; C = False Positives; D= True Negatives;

298

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

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

$$307 \quad MAE = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i| \quad (6)$$

308 Where n is the number of observations, \hat{Y}_i the predicted value and Y_i the observed value.

309 3.5 Connecting remotely sensed spectral traits with bird species traits

310 The trait framework is an integral part of community ecology (McGill et al., 2006). As an
311 outlook, we therefore demonstrate how the trait approach could provide for a direct
312 modelling interface between remotely sensed spectral trait indicators and bird species traits.
313 For this part, a dietary trait of the 44 bird species is used. The data is taken from Sibly et al.
314 (2012).

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

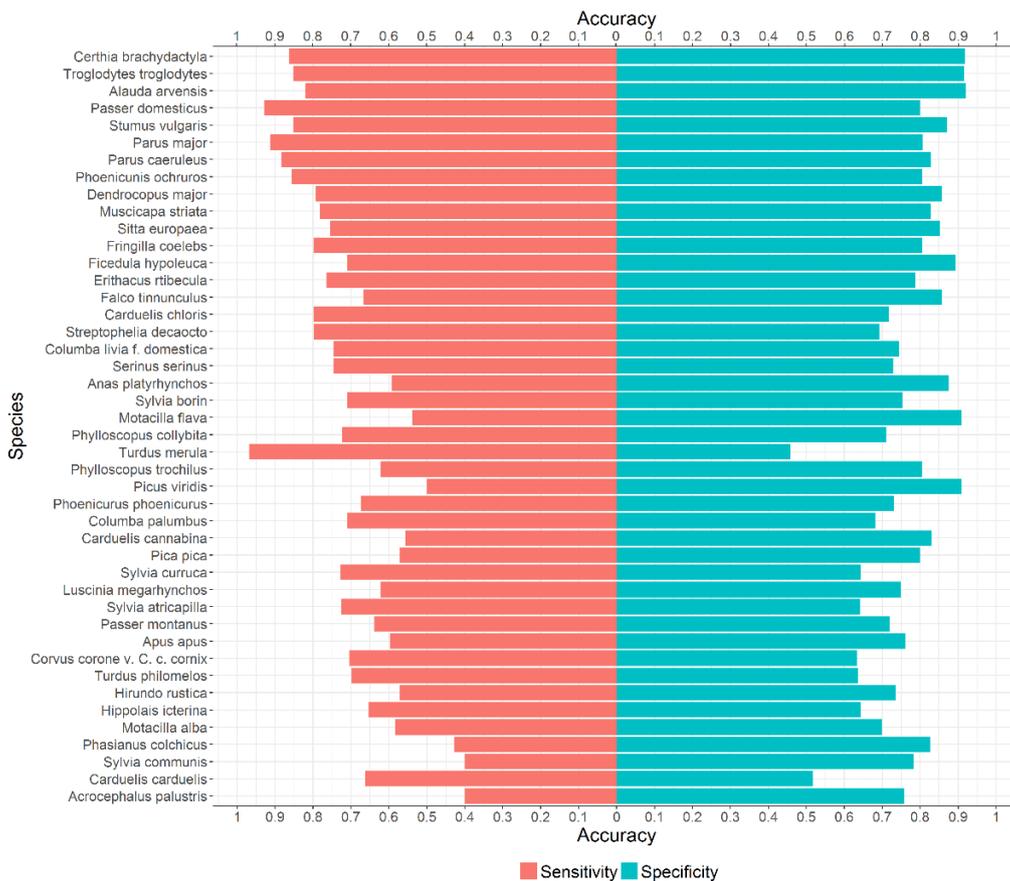
320

321 4 Results

322 4.1 Species distribution modelling

323 4.1.1 Accuracies of computed random forest models

324 The dataset was split into 20% for testing and 80% for growing the random forests. The mean
 325 overall accuracy for the 44 models, based on the testing data set, is 78%, with the best model
 326 featuring an overall accuracy of 90% and the worst 59% (Figure 4 and Table A1 in the
 327 Appendix). The mean accuracy for predicting absences (representing species' specificity) is
 328 approximately 77%, while the prediction for presences (representing species' sensitivity) is
 329 about 70% (Table 2). Low sensitivity values can especially be found in models where
 330 downsampling led to a strong reduction in modelling cases.



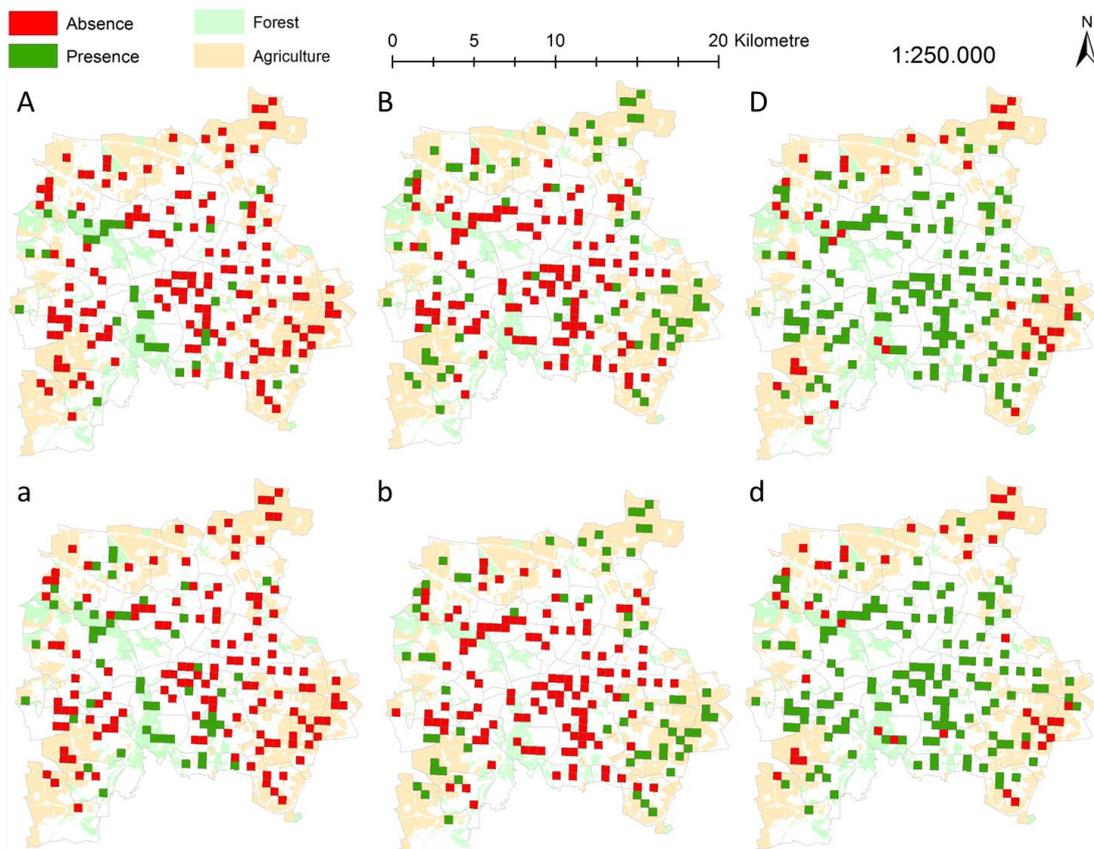
331
 332 Figure 4. Overview of the performance of the 44 models with respect to their sensitivity and
 333 specificity, ordered by their overall accuracy in the modelling process.

334

335 4.1.2 Predictions of urban bird species distributions

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

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



348
349 Figure 5 Predictions (in capital letters) and reference presence/absence data based on the testing
350 data set (20% of all data points) for three bird species in Leipzig; the great spotted woodpecker
351 (*Dendrocopos major*, A & a), the skylark (*Alauda arvensis*, B & b) and the sparrow (*Passer*

352 domesticus, D & d). Naming is congruent with the belonging of the single species to the clusters
353 presented in Figure 6.

354

355 4.1.3 Predictions of urban bird species richness

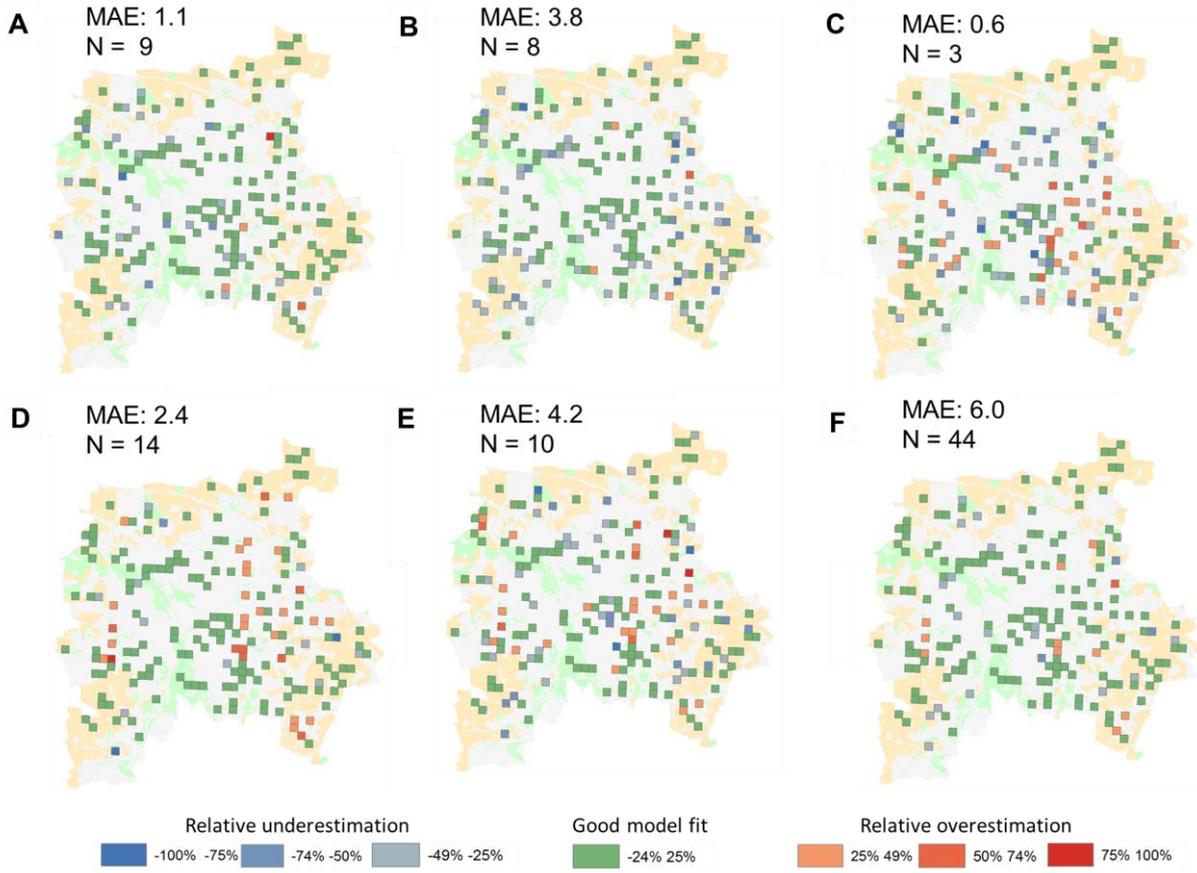
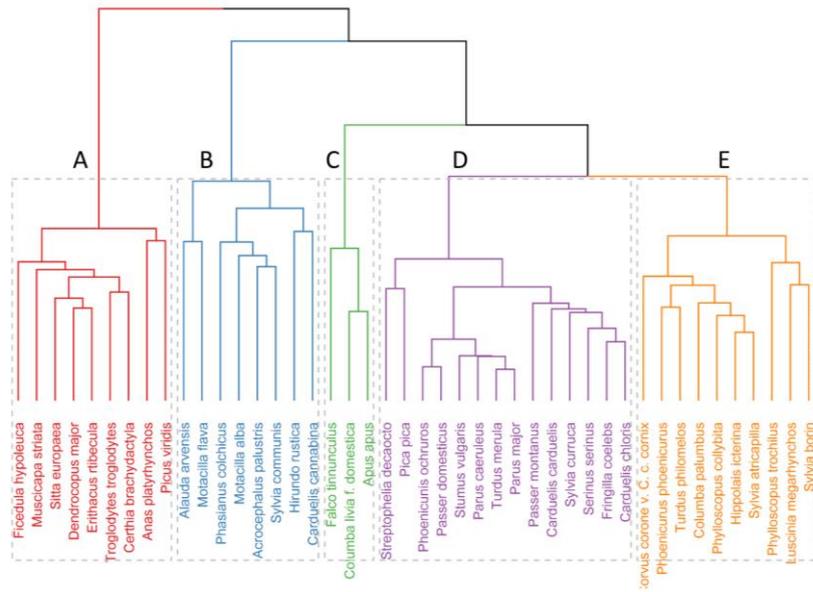
356 Using the TSS measure, an optimal clustering solution was found that encompassed five
357 clusters (Figure 6 – upper section). Based on clusters A, B and D, and a typical species for each
358 of those, the results of the clustering will soon be illustrated. Cluster A features two species
359 types, firstly cavity nesting birds breeding in tree holes like the great spotted woodpecker
360 (*Dendrocopus major*) and the mallard (*Anas platyrhynchos*) breeding along the river
361 embankments in the floodplain forest. Cluster B predominantly features ground-nesting birds,
362 which can usually be found in agriculturally dominated surroundings on meadows and lawns
363 like the Eurasian Skylark (*Alauda arvensis*). Cluster D features hemerophile species that have
364 very broad geographic distributions across cities and show diverse breeding patterns. For
365 instance, the Sparrow (*Passer domesticus*) breeds in holes in buildings but also in dense woody
366 vegetation and is commonly found across the built-up area in Leipzig (BirdLife International,
367 2017).

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

373 The overall bird species richness depicted in cluster F was found to feature an MAE of six.

374 Therefore, it can be seen that species richness is underestimated in grid cells that feature
375 comparably high numbers of present bird species and overestimated in areas where few to
376 no birds are actually found.

377



378

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

382 4.2 Variable importance of random forest models

383 This study introduces texture metrics as new indicators for the creation of species distribution
384 models. Table 4 shows that the majority of species feature these metrics as their most
385 important or second most important variable.). In contrary, only five species feature the NDVI
386 as their most important variable and thirteen a PCA. Since the majority of species feature both
387 a classical remote sensing indicator and a texture metric as their two foremost variables, we
388 can show the benefits both approaches can deliver when combining pixel-based and texture-
389 based approaches.

390

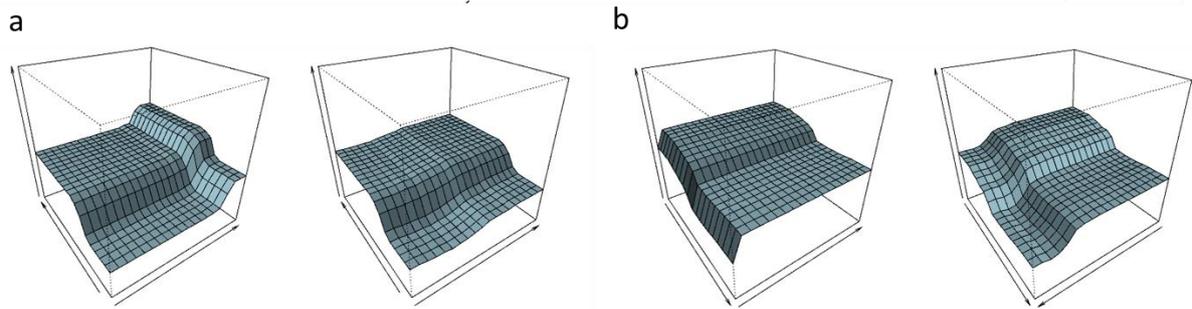
391 Table 4. Indicators covered in this study for the test for bird breeding habitat and their frequency of
392 usage in the first two important variables in the 44 models.

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

393

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

399



400

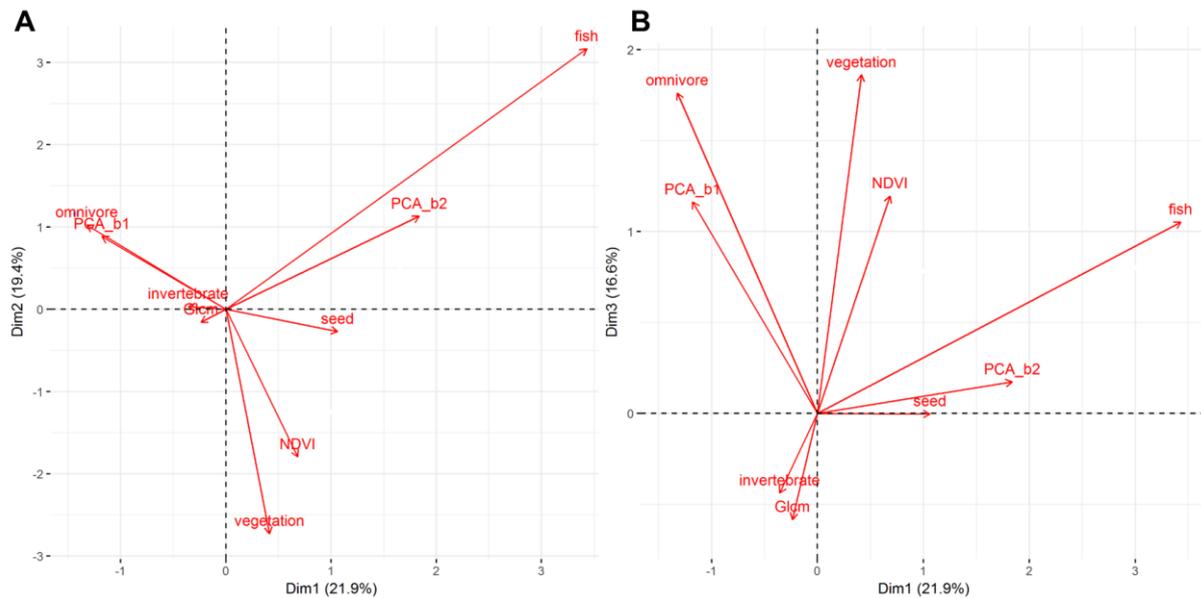
401 Figure 7. Interactions between the two foremost important variables for (a) the Eurasian skylark (Two
 402 texture metrics) and (b) the great spotted woodpecker (PCA – Texture metric)

403

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

411 4.3 Connecting the spectral trait concept to the traits of bird species

412 So far this study has modelled the presence/absence patterns of breeding birds, hence
 413 operating at the species level. In addition, the proposed methodology also provides the
 414 opportunity for connecting the spectral plant trait indicators directly to functional bird traits.
 415 This is illustrated in Figure 8, which shows that our two categories of indicators, GLCM texture
 416 metrics on the one hand and NDVI & PCA on the other, are indicative of a dietary trait.



417

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

421

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

433

434 5 Discussion

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

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

447 A key challenge for improving existing SDMs is the scarcity of spatially continuous high-
448 resolution land-use/cover datasets, particularly in urban environments (He et al., 2015),
449 because discrete EO data products are limited in extent and temporal and spatial resolution.
450 Thus, they cannot show internal variability of classes or transition zones, or small linear
451 elements (Lausch et al., 2015). This is especially problematic since the composition and
452 structure of vegetation are the most important determinants for breeding sites of birds (Ikin
453 et al., 2013; Paker et al., 2014). This study addresses this by integrating high-resolution
454 RapidEye data that is already being used in studies of urban ecology (Wellmann et al., 2018).

455 This methodology will greatly benefit from the broad availability of current and future high-
456 resolution EO datasets, for example, Planet Labs Doves (Wulder & Coops, 2014; Lausch et al.,
457 2018). In comparison, lower resolution sensors like Landsat (30m) or MODIS (250m) are too
458 coarse to discern any spatial diversity of species-specific habitat characteristics in the urban
459 and peri-urban environment (Saveraid et al., 2001; Goetz et al., 2007). While the need for
460 higher spatial resolution applies to all ecosystem types, it very strongly relates to urban
461 ecosystems where variation in the spatial dimension is very high due to the mix of built,
462 sealed-green and blue surfaces (McPhearson et al., 2016).

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

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

480 Regarding the predictors, we found that the NDVI is the second least important indicator after
481 the measures of autocorrelation. The NDVI analyses only a single trait in the vegetation, i.e.
482 its chlorophyll content or degree of greenness, and is thus a poor indicator for overall
483 functional diversity (Wang et al., 2003). However, studies that already incorporate EO data for
484 species diversity or species richness analysis often rely on the NDVI without conducting
485 sensitivity analysis (Seto et al., 2004; Goetz et al., 2007; Bino et al., 2008). Between the NDVI
486 and species richness of different taxa varying indicator relationships were found (Bino et al.,
487 2008), raising the question of reliability and transferability. In this light, a PCA seems more
488 promising. When computed only in vegetated areas, PCA analysis represents a large variety in
489 plant traits (Estes et al., 2010). In our case, the first two PCA bands explained 99% of the
490 variance in the spectral backscatter. We found that the PCA-based predictors were the second
491 most important. This suggests that more effort must be taken to analyse the degree of
492 functional diversity in plant communities rather than analysing derivatives of photosynthesis
493 capacity, as in the NDVI (Estes et al., 2010).

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

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

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

512

513 6 Conclusions

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

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

529 7 Acknowledgements

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

538

539 8 Data Accessibility

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

542 9 References

- 543 Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future
544 directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31.
545 doi:10.1016/j.isprsjprs.2016.01.011
- 546 Bernard, S., Heutte, L., & Adam, S. (2009). Influence of hyperparameters on random forest accuracy.
547 In *International Workshop on Multiple Classifier Systems* (pp. 171–180).

- 548 Bino, G., Levin, N., Darawshi, S., Van Der Hal, N., Reich-Solomon, A., & Kark, S. (2008). Accurate
549 prediction of bird species richness patterns in an urban environment using Landsat-derived
550 NDVI and spectral unmixing. *International Journal of Remote Sensing*, 29(13), 3675–3700.
551 doi:10.1080/01431160701772534
- 552 BirdLife International. (2017). The IUCN Red List of Threatened Species. Retrieved 11 April 2018, from
553 <http://www.iucnredlist.org>
- 554 Bowler, D., Heldbjerg, H., Fox, A., de Jong, M., & Böhning-Gaese, K. (2019). Long-term declines of
555 European insectivorous bird populations and potential causes. *Conservation Biology*.
- 556 Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324
- 557 Chace, J., & Walsh, J. (2006). Urban effects on native avifauna: A review. *Landscape and Urban*
558 *Planning*, 74(1), 46–69. doi:10.1016/j.landurbplan.2004.08.007
- 559 Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of*
560 *California, Berkeley*, 110, 1–12.
- 561 Cutler, D., Edwards, T., Beard, K., Cutler, A., Hess, K., Gibson, J., & Lawler, J. (2007). Random Forests
562 for Classification in Ecology. *Ecology*, 88(11), 2783–2792. doi:10.1890/07-0539.1
- 563 Datt, B. (1998). Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+b, and total carotenoid
564 content in eucalyptus leaves. *Remote Sensing of Environment*, 66(2), 111–121.
565 doi:10.1016/S0034-4257(98)00046-7
- 566 Dormann, C., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Lautenbach, S. (2013).
567 Collinearity: a review of methods to deal with it and a simulation study evaluating their
568 performance. *Ecography*, 36(1), 27–46. doi:10.1111/j.1600-0587.2012.07348.x
- 569 Estes, L., Reillo, P., Mwangi, A., Okin, G., & Shugart, H. (2010). Remote sensing of structural
570 complexity indices for habitat and species distribution modeling. *Remote Sensing of*
571 *Environment*, 114(4), 792–804. doi:10.1016/j.rse.2009.11.016
- 572 Evans, J., Murphy, M., Holden, Z., & Cushman, S. (2011). Modeling species distribution and change
573 using random forest. In *Predictive species and habitat modeling in landscape ecology* (pp. 139–
574 159). Springer.
- 575 Falcucci, A., Maiorano, L., & Boitani, L. (2007). Changes in land-use/land-cover patterns in Italy and
576 their implications for biodiversity conservation. *Landscape Ecology*, 22(4), 617–631.
577 doi:10.1007/s10980-006-9056-4
- 578 Gamon, J., Field, C., Goulden, M., Griffin, K., Hartley, A., Joel, G., ... Valentini, R. (1995). Relationships
579 between NDVI, canopy structure, and photosynthesis in three Californian vegetation types.
580 *Ecological Applications*, 5(1), 28–41. doi:10.2307/1942049
- 581 Geary, R. (1954). The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5(3),
582 115–146.
- 583 Gedeon, K., Mitschke, A., & Sudfeldt, C. (2004). *Brutvögel in Deutschland*. Stiftung Vogelmonitoring
584 Deutschland. Retrieved from <http://www.dda->
585 [web.de/downloads/texts/adebar/adebar1_komplett.pdf](http://www.dda-web.de/downloads/texts/adebar/adebar1_komplett.pdf)
- 586 Goetz, S., Steinberg, D., Dubayah, R., & Blair, B. (2007). Laser remote sensing of canopy habitat
587 heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA.

- 588 *Remote Sensing of Environment*, 108(3), 254–263. doi:10.1016/j.rse.2006.11.016
- 589 Greenacre, M., & Blasius, J. (2006). *Multiple correspondence analysis and related methods*. Chapman
590 and Hall/CRC.
- 591 Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat
592 models. *Ecology Letters*, 8(9), 993–1009. doi:10.1111/j.1461-0248.2005.00792.x
- 593 Haase, D., Jänicke, C., & Wellmann, T. (2019). Front and back yard green analysis with subpixel
594 vegetation fractions from earth observation data in a city. *Landscape and Urban Planning*.
595 doi:10.1016/j.landurbplan.2018.10.010
- 596 Haase, D., Larondelle, N., Andersson, E., Artmann, M., Borgström, S., Breuste, J., ... Elmqvist, T.
597 (2014). A quantitative review of urban ecosystem service assessments: Concepts, models, and
598 implementation. *Ambio*, 43(4), 413–433. doi:10.1007/s13280-014-0504-0
- 599 Haralick, R., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE*
600 *Transactions on Systems, Man and Cybernetics*. doi:10.1109/TSMC.1973.4309314
- 601 He, K., Bradley, B., Cord, A., Rocchini, D., Tuanmu, M. N., Schmidtlein, S., ... Pettorelli, N. (2015). Will
602 remote sensing shape the next generation of species distribution models? *Remote Sensing in*
603 *Ecology and Conservation*, 1(1), 4–18. doi:10.1002/rse2.7
- 604 Ikin, K., Beaty, M., Lindenmayer, D., Knight, E., Fischer, J., & Manning, A. (2013). Pocket parks in a
605 compact city: How do birds respond to increasing residential density? *Landscape Ecology*, 28(1),
606 45–56. doi:10.1007/s10980-012-9811-7
- 607 Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F., Asner, G., ... Ustin, S. (2016).
608 Monitoring plant functional diversity from space. *Nature Plants*, 2(3), 1–5.
609 doi:10.1038/nplants.2016.24
- 610 Jolliffe, I. (2002). *Principal component analysis*. John Wiley & Sons, Ltd.
- 611 Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., ... Wirth, C. (2011). TRY - a global
612 database of plant traits. *Global Change Biology*, 17(9), 2905–2935. doi:10.1111/j.1365-
613 2486.2011.02451.x
- 614 Kerr, J., & Ostrovsky, M. (2003). From space to species: Ecological applications for remote sensing.
615 *Trends in Ecology and Evolution*, 18(6), 299–305. doi:10.1016/S0169-5347(03)00071-5
- 616 Kissling, D., Walls, R., Bowser, A., Jones, M., Kattge, J., Agosti, D., ... Guralnick, R. (2018). Towards
617 global data products of Essential Biodiversity Variables on species traits. *Nature Ecology and*
618 *Evolution*. doi:10.1038/s41559-018-0667-3
- 619 Knapp, S., Winter, M., & Klotz, S. (2017). Increasing species richness but decreasing phylogenetic
620 richness and divergence over a 320-year period of urbanization. *Journal of Applied Ecology*,
621 1152–1160. doi:10.1111/1365-2664.12826
- 622 Kuhn, M. (2008). Caret package. *Journal of Statistical Software*, 28(5), 1–26.
- 623 Lausch, A., Blaschke, T., Haase, D., Herzog, F., Syrbe, R. U., Tischendorf, L., & Walz, U. (2015).
624 Understanding and quantifying landscape structure - A review on relevant process
625 characteristics, data models and landscape metrics. *Ecological Modelling*, 295, 31–41.
626 doi:10.1016/j.ecolmodel.2014.08.018

- 627 Lausch, A., Borg, E., Bumberger, J., Dietrich, P., Heurich, M., Huth, A., ... Schaepman, M. (2018).
628 Understanding forest health with remote sensing, Part III: Requirements for a scalable multi-
629 source forest health monitoring network based on data science approaches. *Remote Sensing*,
630 10(7). doi:10.3390/rs10071120
- 631 Lausch, A., Erasmi, S., King, D., Magdon, P., & Heurich, M. (2016). Understanding Forest Health with
632 Remote Sensing -Part I — A Review of Spectral Traits , Processes and Remote-Sensing
633 Characteristics. *Remote Sensing*, 1029(8), 1–43. doi:10.3390/rs8121029
- 634 McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from
635 functional traits. *Trends in Ecology and Evolution*, 21(4), 178–185.
636 doi:10.1016/j.tree.2006.02.002
- 637 McPhearson, T., Haase, D., Kabisch, N., & Gren, Å. (2016). Advancing understanding of the complex
638 nature of urban systems. *Ecological Indicators*, 70, 566–573. doi:10.1016/j.ecolind.2016.03.054
- 639 Mitchell, M. (2011). Bias of the Random Forest Out-of-Bag (OOB) Error for Certain Input Parameters.
640 *Open Journal of Statistics*, 01(03), 205–211. doi:10.4236/ojs.2011.13024
- 641 Moran, P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1), 17–23.
- 642 Paker, Y., Yom-Tov, Y., Alon-Mozes, T., & Barnea, A. (2014). The effect of plant richness and urban
643 garden structure on bird species richness, diversity and community structure. *Landscape and
644 Urban Planning*, 122, 186–195. doi:10.1016/j.landurbplan.2013.10.005
- 645 Palmer, M., Earls, P., Hoagland, B., White, P., & Wohlgemuth, T. (2002). Quantitative tools for
646 perfecting species lists. *Environmetrics*, 13(2), 121–137. doi:10.1002/env.516
- 647 Podani, J., & Schmera, D. (2011). A new conceptual and methodological framework for exploring and
648 explaining pattern in presence - absence data. *Oikos*, 120(11), 1625–1638. doi:10.1111/j.1600-
649 0706.2011.19451.x
- 650 R Core Team. (2000). R language definition. *Vienna, Austria: R Foundation for Statistical Computing*.
- 651 Richter, R. (2011). Atmospheric / Topographic Correction for Satellite Imagery (ATCOR - 2/3 User
652 Guide). *DLR-German Aerospace Center, Wessling, Germany*, (January).
- 653 Rocchini, D., Balkenhol, N., Carter, G., Foody, G., Gillespie, T., He, K. S., ... Neteler, M. (2010).
654 Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and
655 open challenges. *Ecological Informatics*, 5(5), 318–329. doi:10.1016/j.ecoinf.2010.06.001
- 656 Rocchini, D., Luque, S., Pettorelli, N., Bastin, L., Doktor, D., Faedi, N., ... Nagendra, H. (2017).
657 Measuring β -diversity by remote sensing: a challenge for biodiversity monitoring. *Methods in
658 Ecology and Evolution*, (November).
- 659 Rousseeuw, P., & Leroy, A. (2005). *Robust regression and outlier detection* (Vol. 589). John Wiley &
660 sons.
- 661 Saveraid, E., Debinski, D., Kindscher, K., & Jakubauskas, M. (2001). A comparison of satellite data and
662 landscape variables in predicting bird species occurrences in the Greater Yellowstone
663 Ecosystem, USA, (1992), 71–83. doi:10.1023/A:1008119219788
- 664 Schneider, F., Morsdorf, F., Schmid, B., Petchey, O., Hueni, A., Schimel, D., & Schaepman, M. (2017).
665 Mapping functional diversity from remotely sensed morphological and physiological forest
666 traits. *Nature Communications*, 8(1). doi:10.1038/s41467-017-01530-3

- 667 Seto, K., Fleishman, E., Fay, J., & Betrus, C. (2004). Linking spatial patterns of bird and butterfly
668 species richness with Landsat TM derived NDVI. *International Journal of Remote Sensing*,
669 25(20), 4309–4324.
- 670 Sibly, R., Witt, C., Wright, N., Venditti, C., Jetz, W., & Brown, J. (2012). Energetics, lifestyle, and
671 reproduction in birds. *Proceedings of the National Academy of Sciences*, 109(27), 10937–10941.
672 doi:10.1073/pnas.1206512109
- 673 St-Louis, V., Pidgeon, A., Clayton, M., Locke, B., Bash, D., & Radeloff, V. (2009). Satellite image texture
674 and a vegetation index predict avian biodiversity in the Chihuahuan Desert of New Mexico.
675 *Ecography*, 32(3), 468–480. doi:10.1111/j.1600-0587.2008.05512.x
- 676 Stadt Leipzig. (n.d.). Die Stadtforstverwaltung der Stadt Leipzig. Retrieved 10 June 2019, from
677 <https://www.leipzig.de/fileadmin/mediendatenbank/leipzig->
678 [de/Stadt/02.3_Deiz3_Umwelt_Ordnung_Sport/67_Amt_fuer_Stadtgruen_und_Gewaesser/Stadt-](https://www.leipzig.de/fileadmin/mediendatenbank/leipzig-de/Stadt/02.3_Deiz3_Umwelt_Ordnung_Sport/67_Amt_fuer_Stadtgruen_und_Gewaesser/Stadt-)
679 [wald_und_Auenwald/Informationen_zum_Stadtwald_und_Auenwald/Allgemeines_zum_Leipzi-](https://www.leipzig.de/fileadmin/mediendatenbank/leipzig-de/Stadt/02.3_Deiz3_Umwelt_Ordnung_Sport/67_Amt_fuer_Stadtgruen_und_Gewaesser/Stadt-wald_und_Auenwald/Informationen_zum_Stadtwald_und_Auenwald/Allgemeines_zum_Leipzi-)
680 [ger_Stadtwald.pdf](https://www.leipzig.de/fileadmin/mediendatenbank/leipzig-de/Stadt/02.3_Deiz3_Umwelt_Ordnung_Sport/67_Amt_fuer_Stadtgruen_und_Gewaesser/Stadt-wald_und_Auenwald/Informationen_zum_Stadtwald_und_Auenwald/Allgemeines_zum_Leipzi-ger_Stadtwald.pdf)
- 681 Stadt Leipzig. (2018). Baumkataster und Statistik. Retrieved 10 June 2019, from
682 <https://www.leipzig.de/umwelt-und-verkehr/umwelt-und-naturschutz/baeume-und->
683 [baumschutz/stadtbaeume/baumkataster-und-statistik/](https://www.leipzig.de/umwelt-und-verkehr/umwelt-und-naturschutz/baeume-und-baumschutz/stadtbaeume/baumkataster-und-statistik/)
- 684 StUfa. (1995). (Staatliches Umweltfachamt im Freistaat Sachsen) Brutvogelatlas der Stadt und des
685 Landkreises Leipzig. Materialien zu Naturschutz und Landschaftspflege. - Leipzig.
- 686 Südbeck, P. (2005). *Methodenstandards zur Erfassung der Brutvögel Deutschlands*. Max-Planck-Inst.
687 für Ornithologie, Vogelwarte Radolfzell.
- 688 Tigges, J., Lakes, T., & Hostert, P. (2013). Urban vegetation classification: Benefits of multitemporal
689 RapidEye satellite data. *Remote Sensing of Environment*, 136, 66–75.
690 doi:10.1016/j.rse.2013.05.001
- 691 Violle, C., Navas, M., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the
692 concept of trait be functional! *Oikos*, 116(5), 882–892. doi:10.1111/j.2007.0030-1299.15559.x
- 693 Völkl, W., Blick, T., Kornacker, P., & Martens, H. (2004). Quantitativer Überblick über die rezente
694 Fauna von Deutschland. *Natur Und Landschaft*.
- 695 Wang, Z., Liu, C., & Huete, A. (2003). From AVHRR-NDVI to MODIS-EVI: Advances in vegetation index
696 research. *Acta Ecologica Sinica*, 23(5), 979–987.
- 697 Warton, D. I., Shipley, B., & Hastie, T. (2015). CATS regression - a model-based approach to studying
698 trait-based community assembly. *Methods in Ecology and Evolution*, 6(4), 389–398.
699 doi:10.1111/2041-210X.12280
- 700 Wellmann, T., Haase, D., Knapp, S., Salbach, C., Selsam, P., & Lausch, A. (2018). Urban land use
701 intensity assessment: The potential of spatio-temporal spectral traits with remote sensing.
702 *Ecological Indicators*, 85, 190–203. doi:10.1016/j.ecolind.2017.10.029
- 703 Wolff, M., Haase, A., Haase, D., & Kabisch, N. (2016). The impact of urban regrowth on the built
704 environment. *Urban Studies*. doi:10.1177/0042098016658231
- 705 Wulder, M., & Coops, N. (2014). Satellites: Make Earth observations open access. *Nature*, 513(7516),
706 30–31. doi:10.1038/513030a

