

Morphological classification of conspecific birds from closely situated breeding areas – A case study of the Common Nightingale

DÁVID KOVÁTS^{1*} & ANDREA HARNOS²



Dávid Kováts & Andrea Harnos 2015. Morphological classification of conspecific birds from closely situated breeding areas – A case study of the Common Nightingale. – Ornis Hungarica 23(2): 20–30.

Abstract In this paper, a complex morphological comparison of four Common Nightingale groups (*Luscinia megarhynchos*) is demonstrated. In total, 121 territorial nightingales were mist-netted and measured individually on four study areas called ‘Bódva’, ‘Felső-Tisza’, ‘Szatmár-Bereg’ and ‘Bátorliget’ in the North-Eastern part of Hungary in 2006–2013. To distinguish groups by morphology, *Classification and Regression Trees* (CART), *Random Forest* (RF) and *Linear Discriminant Analysis* (LDA) methods were used. Comparison of the four studied Common Nightingale groups shows substantial morphological differences in the length of the second, the third and the fourth primaries (P_2 , P_3 , P_4), in bill length (BL) and bill width (BW), while other characteristics showed greater similarities. Based on the results of all the applied classification methods, birds originated from Szatmár-Bereg were clearly distinguishable from the others. The differences in morphology can be explained by interspecific competition or phenotypic plasticity resulting from the change of ecological, environmental parameters. Our case study highlights the advantageous differences of the classification methods to distinguish groups with similar morphology and to choose important variables for classification. In conclusion, broad application of the classification methods RF and CART is highly recommended in comparative ecological studies.

Keywords: *Luscinia megarhynchos*, classification and regression trees, random forest, linear discriminant analysis, morphological classification

Összefoglalás A tanulmányban négy fülemüle (*Luscinia megarhynchos*) állomány komplex morfológiai összehasonlítását mutatjuk be, az ökológiai vizsgálatokban csak kevésbé ismert klasszifikációs módszerekkel. Magyarország északkeleti részén, négy populáció (‘Bódva’, ‘Felső-Tisza’, ‘Szatmár-Bereg’ és ‘Bátorliget’) összesen 121 egyedéről vettünk fel biometriai adatokat 2006 és 2013 között. Az állományok morfológiai elkülönítésére a véletlen erdők (RF), döntési fák (CART) és a lineáris diszkriminancia analízis (LDA) módszereket alkalmaztuk. A négy fülemüle állomány a második, harmadik és negyedik kézevezők hosszában (P_2 , P_3 , P_4), a csőrösszabban (BL) és csőr szélesség (BW) tekintetében mutatott különbséget. Az általunk használt klasszifikációs módszerekkel a szatmár-beregi fülemüle csoport a többi állománytól határozottan elkülöníthető. A morfológiai különállóságot a kompetíciós verseny vagy a fenotípusos plaszticitás magyarázhatja, amelyek gyakori oka az ökológiai, környezeti paraméterek minőségi megváltozása. Esettanulmányunk elsősorban a különböző klasszifikációs módszerek alkalmazhatóságára hívja fel a figyelmet, amellyel az egymáshoz hasonló morfológiájú populációk elkülönítésére, valamint az elkülönítésében szerepet játszó változók meghatározására is lehetőség nyílik. Eredményeink alapján a CART és RF modellek széleskörű használatát javasoljuk a hasonló jellegű ökológiai vizsgálatokban.

Kulcsszavak: *Luscinia megarhynchos*, klasszifikációs és regressziós fák, véletlen erdők, lineáris diszkriminancia elemzés, morfológiai klasszifikáció

¹Department of Evolutionary Zoology and Human Biology, University of Debrecen, 4032 Debrecen, Egyetem tér 1., Hungary, e-mail: david.kovats@gmail.com

²Department of Biomathematics and Informatics, Faculty of Veterinary Sciences, Szent István University, 1078 Budapest, István utca 2., Hungary

* corresponding author

Introduction

Despite the fact that most of the European passerines are identifiable by their morphological traits (Cramp 1988, Sibley & Monroe 1990, Svensson 1992, Watson 2005), there may be groups that are hidden in morphologically cryptic assemblages, which has impeded their discovery (e.g. Illera *et al.* 2008), especially under field circumstances.

In particular, on species of some partially or totally overlapping groups, little or non-significant differences can be observed in their morphological traits (e.g. Päckert *et al.* 2007). If there is gene flow between sister groups breeding next to each other, considerable differences may arise in their morphological or phylogenetical characteristics

(e.g. Leisler *et al.* 1997) and in their vocalization (e.g. Lara *et al.* 2012).

Linear discriminant analysis (LDA) and logistic regression (LR) are often used in ecological studies for classification. The usage of these traditional methods is restricted due to their conditions of applicability and their special purpose. For example, the binary outcome in case of LR, and – in case of LDA – the assumptions that the variables should be linearly related and come from a multivariate normal distribution with homogeneous within-group variance. Real ecological variables are often nonlinearly related and have complex interactions and nonhomogeneous dispersion (Cutler *et al.* 2007). Novel classification methods like Classification and Regression

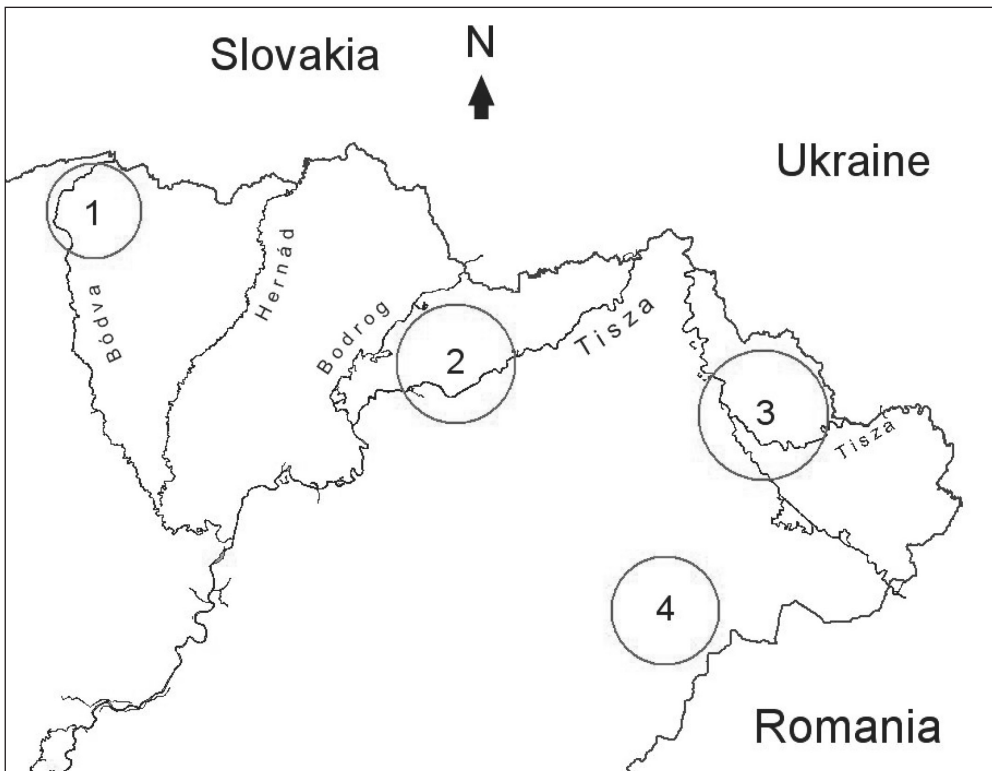


Figure 1. Location of the study areas (1. Bódva, 2. Felső-Tisza, 3. Szatmár-Bereg, 4. Bátorliget)

1. ábra A vizsgálati területek elhelyezkedése (1. Bódva, 2. Felső-Tisza, 3. Szatmár-Bereg, 4. Bátorliget)

Trees and Random Forest methods (De'ath & Fabricius 2000, Strobl *et al.* 2007) can be used in such cases.

The Common Nightingale (*Luscinia megarhynchos* Brehm, 1831) is a widely distributed species via the Palearctic (Cramp 1988). However, morphological differences among their populations have been reported so far only from Eastern Europe (e.g. Reifová *et al.* 2011, Kováts *et al.* 2013).

Here, we report how relevant variables can be selected effectively to distinguish particular groups of Common Nightingales caught in four closely situated study sites in the North-Eastern part of Hungary.

Materials and Methods

Study area

Four study areas were investigated which are situated along the river flood plain areas of Bódva (48°27'N, 20°43'E), on the river-sides of Felső-Tisza (48°10'N, 21°42'E), in Szatmár-Bereg (48°07'N, 22°30'E) and in Bátorliget (47°45'N, 22°26'E) in the north-easter region of Hungary (*Figure 1*). All of these field sites are covered by a rich vegetation of seminatural hardwood and softwood riparian forests with thickets, stands of patchy undergrowth, small ponds and sinuous backwaters. The proportion of the seminatural forests is relatively high in Felső-Tisza and in Bátorliget (approx. 32% and 21%), while it is lower (approx. 4% and 2% only) in Bódva and in Szatmár-Bereg (unpublished data).

Biometric measurements

All measurements were collected in 2006–2013. Birds were captured during the peak

of the breeding season, from early May to late June, using Ecotone® mist-nets of different sizes and tape luring (Busse 2000). All birds were measured in millimetres according to Svensson's (1992) recommendations, in the following order: maximal wing length [MWL], wing index (difference in length between the first primary and wing tip) [W^{lx}], distance between the wing point and the first secondary [I/II], relative length of the first primary [rLP_1], length of the second-, third- and fourth primaries [P_2 , P_3 , P_4], tail length [TL], bill length (measured to skull) [BL], bill width (measured on the frontal margin of nostrils) [BW] and tarsus length [TL]. In addition, wing-pointedness [W^p] and wing-symmetry [W^{sym}] indexes were calculated following Hołynski (1965) and Busse (1967, 2000). Wing characteristics were measured with a plastic ruler to the closest 1.0 mm, while for the measurements taken on the bill and tarsus a metal calliper with the accuracy of 0.1 mm was used. Furthermore, body mass was also recorded to the closest 0.1 g using a 60-g Pesola spring scale. Only the data of males were analysed.

The phylogenetic patterns of all individuals measured in this study had been previously identified and published elsewhere (Ács & Kováts 2013). Direct PCR sequencing of the mitochondrial DNA (cytochrome oxidase subunit I – COI gene) confirmed that all haplotypes of these four groups are genetically homogeneous with low sequence heterogeneity, thus, they are closely related.

Statistical analyses

In total, 121 individuals were measured in the four study areas (Bódva: $n=33$, Felső-Tisza: $n=38$, Szatmár-Bereg: $n=23$,

Bátorliget: $n=27$). For morphological classification, Classification and Regression Trees, Random Forests and Linear Discriminant Analyses were simultaneously used. All the calculations were done using the statistical software R 2.15.2. (R Core Team 2010). The level of significance was set to 0.05.

Methods for morphological classification of the breeding nightingale groups

Classification and Regression Trees (CART)

CART is ideally suited for the analysis of complex ecological data (De'ath & Fabricius 2000, Berggren & Low 2006, Cutler *et al.* 2007). It recognizes relevant variables automatically and clearly illustrates their interrelationships and interactions as well.

CARTs are built using a simple nonparametric regression approach. Both numerical and categorical variables (covariates) can be used to build a tree. The general rule is to split the observations into two parts based on a predictor variable (root), then to split the subset further based on another or the same variable. Then we repeat the process recursively until a stopping criterion is reached. So the space spanned by all predictor variables is recursively partitioned into a set of rectangular areas, where the observations with similar response values are grouped and a constant value (a simple statistics like the mean, or proportion in case of a categorical response) of the response variable is predicted. The aim at each cutting is to reach maximum homogeneity in the response variable within the groups.

In case of the 'ctree' function of the 'party' package (Hothorn *et al.* 2006a) used in this paper, the cutting points are determined

and tested by an association measure between the response and the covariates. If there is no more significant relationship between the response and covariates the recursion stops. It can also be stopped by other criteria like the size of the tree, or the number of observations at the cutting points or on the leaves. The level of significance can be adjusted too. The trees are represented graphically and make it easy to understand the relationship of explanatory variables. Important variables are tested near the root, while less important ones right next to the leaves of the CART. It is possible that some of the explanatory variables do not appear in the tree, as not influencing the response variable or explaining the same as one or more other variables strongly correlated with them. These are considered to be irrelevant. In addition, trees are invariant to monotonic transformations of the independent numeric variables; moreover, they recognize interactions. If a classification or regression tree is built, it is able to classify new data. Every new observation can be connected to the relevant leaves of the trees.

Random Forests (RF)

Random Forest (Breiman 2001, Prinzie & Van-Del Poel 2008) is a recursive classification method, which is particularly effective when the sample size (n) is small but the number of the predictor variables (p) is large even in the presence of complex interactions (Strobl *et al.* 2007, 2008, 2009).

RF is an ensemble of Classification and Regression Trees. Each tree is built using a random subset of the data and a random subset of the variables chosen at each cutting point of the tree. Due to this methodology, RFs are more suitable to investigate the role of the variables like simpler me-

thods using all of the variables simultaneously, thus the variables with greater impact can suppress the role of those having smaller impact (Strobl *et al.* 2007, 2008). Furthermore, the classification performance of an ensemble of trees can be much better.

Using this method, the importance of variables can be estimated by measuring the difference in prediction accuracy (i.e. the number of observations classified correctly) before and after permuting a variable, averaged over all trees. If accuracy changes only a little, then the variable has no important role, otherwise its role is substantial. Conditional permutation importance considering the correlation between the variables can be calculated yielding an unbiased estimate of the importance.

This variable importance measure can be applied for variable selection. In this study, we used it for selecting those variables that play important roles in separating the groups living in different areas. We used the 'cforest' function of the R 'party' package (Hothorn *et al.* 2006b, Strobl *et al.* 2007, 2008) for the analysis.

Linear Discriminant Analysis (LDA)

The purpose of discriminant analysis is to create a function which is able to classify the observations into two or more groups

(the value of the function changes substantially group by group). Linear discriminant function is a linear combination of the predictor variables. More functions (number of groups minus one) can be created (Manly 1986). LDA was fitted by the 'lda' function of the 'MASS' package (Venables & Ripley 2002). The classifications were carried out by all three methods. The efficacy of the methods can be investigated by comparing the observed and predicted groups and by calculating the prediction accuracy, which is simply the ratio in percentage of the well classified observations.

Results

Based on the results of Random Forest, P_2 , P_3 , and P_4 , followed by BL, BW and TL were the most important variables to distinguish the four Common Nightingale groups. According to the fitted RF, the prediction accuracy was 81.7%, considered to be a good prediction (Table 1). The variables in the order of importance are described on Figure 2.

In contrast, the prediction accuracy of CART was lower, reaching only 57.4% (Table 2). However, as in RF, P_2 and BL were also considered as important morphological variables by CART, while body mass (BM) –

Predicted group	Observed group			
	Bátorliget	Bódva	Felső-Tisza	Szatmár-Bereg
Bátorliget	22	2	2	1
Bódva	3	19	2	1
Felső-Tisza	2	7	33	1
Szatmár-Bereg	0	0	0	20
Prediction accuracy: 81.7%				

Table 1. Contingency table of observed and predicted groups by random forest (RF)
1. táblázat A véletlen erdők (RF) által prediktált és megfigyelt populációk kontingencia-táblája

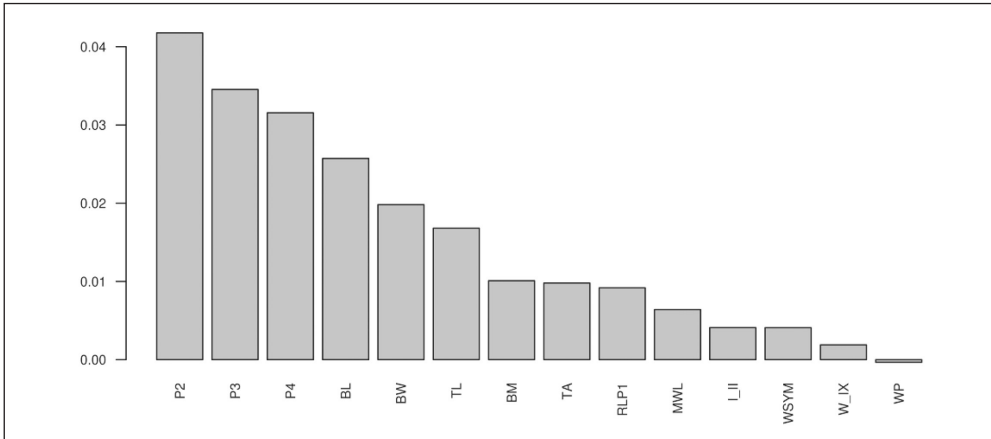


Figure 2. The morphological variables in order of importance estimated by Random Forest (RF)

2. ábra A morfológiai célváltozók a véletlen erdők (RF) által meghatározott fontossági sorrendben

Observed group				
Predicted group	Bátorliget	Bódva	Felső-Tisza	Szatmár-Bereg
Bátorliget	12	4	3	6
Bódva	7	16	6	4
Felső-Tisza	8	1	28	3
Szatmár-Bereg	0	0	0	10
Prediction accuracy: 57.4%				

Table 2. Contingency table of observed and predicted groups by classification and regression trees (CART)

2. táblázat A döntési fa (CART) által prediktált és megfigyelt populációk kontingencia-táblája

similarly to the RF's prediction – was classified farther from the root (Figure 3).

Describing the data on the plot of the first two discriminant axes of LDA, 'Szatmár-Bereg' group is clearly distinguished (Figure 4). Coefficients of the discriminant functions are given in Table 3. Considering the correlations of the discriminant functions with the original variables, LD1 is highly correlated with P_2 , P_3 and P_4 , while LD2 with BL and BW (Table 4). These results imply that the four nightingale groups differ somewhat from each other in the length of primaries and bill sizes, while other morphological traits showed greater similarities between the study areas.

In case of LDA, prediction accuracy was 64.3% (Table 5), which was slightly better than in case of CART, but worse than in case of RF.

Discussion

Morphological differences of the four nightingale groups

In zoology, the analysis of morphological characteristics is particularly important, since the individual taxa and populations can only be identified and distinguished by certain key traits under field conditions. Based

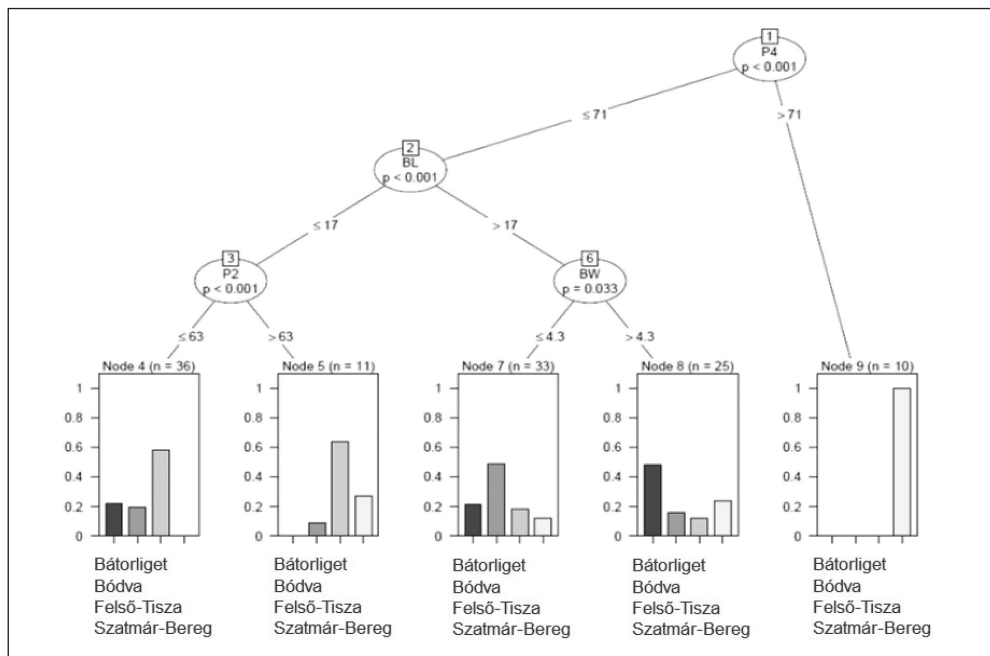


Figure 3. The most important morphological variables given by CART and their interactions. Common Nightingales with narrower bills are distributed in the surroundings of Bódva and Felső-Tisza, while other individuals with wider ones occur in Bátorliget and Szatmár-Bereg. For each inner node, the Bonferroni-adjusted P-values and the ranges on the branches are given, while the fraction of each area is displayed on the leaves

3. ábra Döntési fa (CART) által lényegesnek ítélt és kijelölt morfológiai célváltozók, valamint azok interakciói. A keskenyebb csőrű madarak inkább a Bódva és Felső-Tisza környékén, míg a szélesebb csőrűek Bátorliget és Szatmár-Bereg területén terjedtek el. Minden belső csomópont esetén láthatóak a Bonferroni módszerrel korrigált p-értékek és az ágakon a változók értéktartományai, a leveleken pedig az egyes területek aránya

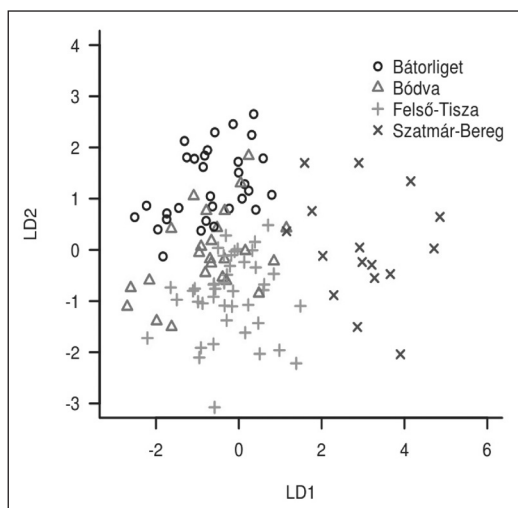


Figure 4. The data described on the first (LD1) and second (LD2) discriminant axes. Marked separation of the individuals from Szatmár-Bereg is clearly visible, whereas those of Bódva and Felső-Tisza greatly overlap

4. ábra Az adatok az első és a második diszkriminancia tengely mentén (LD1 és LD2) való eloszlása. A szatmár-beregi állomány elkülönülése jól megfigyelhető, míg a bódvai, felső-tiszai állományok egymással jelentősen átfednek

Variable	LD1	LD2	LD3
P ₂	0.1072	-0.1882	0.2886
P ₃	0.2652	0.1577	0.3383
P ₄	0.1186	0.1307	-0.6078
BL	0.3518	0.4602	0.6926
BW	0.7542	2.3300	-2.4910
TL	-0.0874	-0.2436	-0.1034
BM	-0.2831	0.2639	0.2189
rLP ₁	0.1695	-0.2287	0.2974
MWL	-0.2484	0.1171	-0.2654
I/II	0.0848	-0.3620	0.3236
W ^{sym}	0.0434	0.0234	-0.2389
W ^x	0.1022	-0.0220	0.0020
Rate of explained variance (%)	67.8	19.7	12.5

Table 3. Coefficients of the three linear discriminant functions (LD1-3)

3. táblázat A lineáris diszkriminancia függvények (LD1-3) együtthatói

Variable	LD1	LD2	LD3
BM	-0.2116	0.4398	0.167
MWL	0.2121	-0.0066	-0.369
W ^x	0.3583	-0.0776	-0.151
P ₃	0.8260	-0.0740	-0.179
rLP ₁	0.0415	-0.1572	0.419
I/II	0.3246	-0.3886	0.171
TL	0.0351	-0.3777	-0.545
W ^p	0.2272	-0.0533	0.120
W ^{sym}	0.2348	0.0773	-0.116
P ₂	0.8158	-0.0273	-0.201
P ₄	0.8239	-0.0479	-0.247
BL	0.0390	0.6675	0.323
BW	-0.0379	0.6304	-0.275

Table 4. Correlation coefficients of the linear discriminant functions and the measured variables. The strongest correlation coefficients are highlighted in bold

4. táblázat A lineáris diszkriminancia függvények és a mért változók korrelációs együtthatói. A legnagyobb korrelációs együtthatókat kiemeltük

Predicted group	Observed group			
	Bátorliget	Bódva	Felső-Tisza	Szatmár-Bereg
Bátorliget	19	6	2	4
Bódva	2	13	9	1
Felső-Tisza	6	9	26	2
Szatmár-Bereg	0	0	0	16
Prediction accuracy: 64.3%				

Table 5. Contingency table of observed and predicted groups by linear discriminant analysis (LDA)

5. táblázat A lineáris diszkriminancia analízis (LDA) által prediktált és megfigyelt populációk kontingencia-táblája

on the correlations of the measured variables and the linear discriminant functions, the four studied nightingale groups can be distinguished on the basis of the length of the primaries (P_2 , P_3 , P_4), bill length (BL) and bill width (BW). RF also found these variables as the most important ones in distinguishing the groups.

Our CART trees clearly indicate that birds the Szatmár-Bereg area can be distinguished from the others solely on the basis of the 4th primary length (P_4) (Figure 3). This kind of marked difference between groups with closely situated habitats can possibly be attributed to three alternative, but mutually non-exclusive reasons. First, phenotypic plasticity in site-faithful species can adapt their morphological characteristics to local environment (Grant 1999, Tellería & Carbonell 1999, Agrawal 2001, Pfennig *et al.* 2010, Reifová *et al.* 2011). Second, due to the site fidelity of the Common Nightingale (Sorjonen 1986, Lille 1988, Amrhein *et al.* 2004) certain characteristics can be inherited within a group consisting of phylogenetically homogenous individuals (Ács & Kováts 2013). Third, the formerly regular breeding of the closely related Thrush Nightingale *L. luscinia* in this area could have resulted in hybridization at this locality (Ács & Kováts 2013).

Our results imply that the birds of Szatmár-Bereg comprise a unique group, highlighting the importance of morphological differences. Concerning bill width (BW) the CART suggests that birds with narrower bills are distributed in the areas of Bódva and Felső-Tisza, while others with wider ones occur in Bátorliget and Szatmár-Bereg (Figure 3, 4). These differences are probably due to nutritional stress as Reifová *et al.* (2011) assumed in Poland. This can manifest in available foods, especially in the

breeding season when several kinds of insects are the most important food sources, corresponding to the particular habitat.

Discussion of the statistical methodology

Traditional classification methods like discriminant analyses or logistic regression are often used in ecological studies. Classification and Regression Trees, as well as Random Forests used by us are not as widespread. Both methods have the advantage that the response and predictor variables don't have to be linearly related and practically can have any kind of distribution. CART is able to discover the interactions between the variables and the resulting decision tree illustrates the connection system between the predictor and response variables. RF builds an ensemble of the trees from the random subsets of the observations and the variables. Using this technique one can solve classification problems where the number of observations is small compared to the number of variables. Usually, the prediction accuracy is greater with RF than in the case of building only one tree. The disadvantage of this method is that the result cannot be described as expressively as in the case of CART. A further advantage of RF is that it is very well suited for variable selection, since it quantifies the importance of each variable. Moreover, the method used by us is able to give an unbiased estimate of the importance, which is essential in case of highly correlated variables.

In our study, RF proved to be the best, considering the prediction accuracy of the three methods (81.7%). LDA and CART yielded roughly similar results (64.3% and 57.4%). RF and CART discovered more or less the same important explanatory variables, while CART fit only two of the three

highly correlated (P_2 , P_3 , P_4) variables to the tree, but this is the consequence of the CART methodology.

The simultaneous application of the three methods appears to be the most appropriate approach. RF performs very well in variable selection, and it reaches the highest prediction accuracy, while CART gives more information in describing the connection system of the variables.

LDA enables us to calculate the discriminant scores using the discriminant functions. With these scores, further analyses can be carried out (usually substantially less latent variables are generated) and the clustering of the observations can be well described graphically.

References

- Agrawal, A. A. 2001. Phenotypic plasticity in the interactions and evolution of species. – *Science* 294: 321–326. DOI: 10.1126/science.1060701
- Amrhein, V., Kunc, H. P. & Naguib, M. 2004. Seasonal patterns of singing activity vary with time of day in the Nightingale (*Luscinia megarhynchos*). – *The Auk* 121: 110–117. DOI: 10.1642/0004-8038(2004)121[0110:SPOSAV] 2.0.CO;2
- Ács, Z. & Kováts, D. 2013. Phylogenetic patterns of a nightingale population in a contact zone of *Luscinia megarhynchos* and *L. luscinia* in Hungary. – *North-Western Journal of Zoology* 9(2): 365–373.
- Berggren, Å. & Low, M. 2006. Sexual dichromatism in North Island Robins (*Petroica longipes*) is weakened by delayed plumage maturation in males and females. – *Emu* 106(3): 203–209. DOI: 10.1071/MU05057
- Breiman, L. 2001. Random Forests. – *Machine Learning* 45(1): 5–32. DOI: 10.1023/A:1010933404324
- Busse, P. 1967. Application of the numerical indexes of the wing-shape. – *Notatki Ornitologiczne* 8(1): 1–8.
- Busse, P. 2000. Bird Station Manual. SE European Bird Migration Network. – University of Gdańsk, pp. 264
- Cramp, S. 1988. The Birds of the Western Palearctic. Vol. 5. – Oxford University Press, Oxford, pp. 1063
- Cutler, D. R., Edwards, C. T., Beard, H. K., Cutler, A., Hess, T. K., Gibson, C. J. & Lawler, C. J. 2007. Random Forest for classification in ecology. – *Ecology* 88(11): 2783–2792. DOI: 10.1890/07-0539.1
- De'ath, G. & Fabricius, E. K. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. – *Ecology* 81: 3178–3192. DOI: 10.1890/0012-9658(2000)081[5B3178:CAR-TAP]5D2.0.CO;2
- Grant, P. R. 1999. The ecology and evolution of Darwin's Finches. – Princeton University Press, Princeton, pp. 492
- Holynski, R. 1965. Methods for the analysis of the wing shape of birds. – *Notatki Ornitologiczne* 6: 21–25.
- Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A. & Van Der Laan, M. 2006a Survival Ensembles. – *Biostatistics* 7(3): 355–373. DOI: 10.1093/biostatistics/kxj011
- Hothorn, T., Hornik, K. & Zeileis, A. 2006b Unbiased recursive partitioning: A conditional inference framework. – *Journal of Computational and Graphical Statistics* 15(3): 651–674. DOI: 10.1198/106186006X133933
- Illera, J. C., Richardson, D. S., Helm, B., Atienza, J. C. & Emerson, B. C. 2008. Phylogenetic relationships, biogeography and speciation in the avian genus *Saxicola*. – *Molecular Phylogenetics and Evolution* 48: 1145–1154. DOI: 10.1016/j.ympev.2008.05.016
- Kováts, D., Végvári, Zs. & Varga, Z. 2013. Morphological patterns of a nightingale population in a contact zone of *Luscinia megarhynchos* and *L. luscinia*. – *Acta Zoologica Academiae Scientiarum Hungaricae* 59(2): 157–170.

In conclusion, above we intended to demonstrate that by using the combination of these methods, even differences between geographically very closely located groups can be detected reliably.

Acknowledgements

We thank Helga Urbán, Beatrix Ferencz, Eszter Zsófia Kovács and Imre Mihalik for their kind help in the field. This work was supported by the National Scientific Research Fund of Hungary (OTKA grant no. 108571) and by a Research Faculty Grant (2014) of the Faculty of Veterinary Science at Szent István University.

- Lara, C. E., Cuervo, A. M., Valderrama, S. V., Calderón-F, D. & Cadena, C. D. 2012. A new species of wren (Troglodytidae: *Thryophilus*) from the dry Cauca River canyon, Northwestern Colombia. – *The Auk* 129(3): 537–550. DOI: 10.1525/auk.2012.12028
- Leisler, B., Heidrich, P., Schulze-Hagen, K. & Wink, M. 1997. Taxonomy and phylogeny of reed warblers (genus *Acrocephalus*) based on mtDNA sequences and morphology. – *Journal of Ornithology* 138: 469–496. DOI: 10.1007/BF01651381
- Lille, R. 1988. Species-specific song and mixed singing of Nightingale and Thrush Nightingale (*Luscinia megarhynchos*, *L. luscinia*). – *Journal of Ornithology* 129: 133–159.
- Manly, B. 1986. *Multivariate Statistical Methods: a Primer*. – Chapman and Hall, New York, pp. 159
- Päckert, M., Martens, J., Tietze, T. D., Dietzen, C., Wink, M. & Kvist, L. 2007. Calibration of a molecular clock in tits (Paridae) – Do nucleotide substitution rates of mitochondrial genes deviate from the 2% rule? – *Molecular Phylogeny and Evolution* 44(1): 1–14. DOI: 10.1016/j.ympev.2007.03.006
- Pfennig, D. W., Wund, M. A., Snell-Rood, E. C., Cruickshank, T., Schlichting, C. D. & Moczek, A. P. 2010. Phenotypic plasticity's impacts on diversification and speciation. – *Trends in Ecology and Evolution* 25: 459–467. DOI: 10.1016/j.tree.2010.05.006
- Prinzie, A. & Van-Den Poel, D. 2008. Random Forests for multiclass classification: Random MultiNominal Logit. – *Expert Systems with Applications* 34(3): 1721–1732. DOI: 10.1016/j.eswa.2007.01.029
- R Core Team 2010. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Sibley, C. G. & Monroe, B. L. 1990. *Distribution and Taxonomy of Birds of the World*. – Yale University Press, New Haven, pp. 1111
- Sorjonen, J. 1986. Mixed singing and interspecific territoriality: consequences of secondary contact of two ecologically and morphologically similar nightingale species in Europe. – *Ornis Scandinavica* 17: 53–67.
- Strobl, C., Boulesteix, A-L., Zeileis, A. & Hothorn, T. 2007. Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. – *BMC Bioinformatics* 8: 25. pp. 21, DOI: 10.1186/1471-2105-8-25
- Strobl, C., Anne-Laure, B., Thomas, K., Thomas, A. & Achim, Z. 2008. Conditional variable importance for random forests. – *BMC Bioinformatics* 9: 307. pp. 11, DOI: 10.1186/1471-2105-9-307
- Strobl, C., Malley, J. & Tutz, G. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. – *Psychol Methods* 14(4): 323–348. DOI: 10.1037/a0016973
- Svensson, L. 1992. *Identification guide to European Passerines*. 4th ed. – Stockholm, pp. 368
- Tellería, J. L. & Carbonell, R. 1999. Morphometric variation of five Iberian Blackcap *Sylvia atricapilla* populations. – *Journal of Avian Biology* 30: 63–71. DOI: 10.2307/3677244
- Reifová, R., Reif, J., Antczak, M. & Nachman, M. W. 2011. Ecological character displacement in the face of gene flow: Evidence from two species of nightingales. – *BMC Evolutionary Biology* 11: 138. pp. 11, DOI: 10.1186/1471-2148-11-138
- Venables, W. N. & Ripley, B. D. 2002. *Modern Applied Statistics with S*. 4th ed. – Springer, New York, pp. 495
- Watson, D. M. 2005. Diagnosable versus distinct: Evaluating species limits in birds. – *Bioscience* 55: 60–68. DOI: 10.1641/0006-3568(2005)05

