An investigation into the possibilities of sex and age determination of Eurasian woodcock (Scolopax rusticola L.) based on biometric parameters, using conditional inference trees and minimal important differences

Authors: Bende, Attila, László, Richárd, Faragó, Sándor, and Fekete, István

RESEARCH PAPER

An investigation into the possibilities of sex and age determination of Eurasian woodcock (*Scolopax rusticola* L.) based on biometric parameters, using conditional inference trees and minimal important differences

Attila BENDE[1]*, Richárd LÁSZLÓ[1], Sándor FARAGÓ[1] and István FEKETE[2]

[1] Institute of Wildlife Biology and Management, University of Sopron, Sopron, Hungary;
   e-mail: bende.attila@uni-sopron.hu, farago.sandor@uni-sopron.hu, richard.laszlo@uni-soporn.hu

[2] Amity Institute of Psychology & Allied Sciences, Amity University Kolkata, Kolkata, India,
   e-mail: ifekete@kol.amity.edu

Abstract. Morphometric characteristics of Eurasian woodcock collected during spring hunting (March) in Hungary between 2010 and 2014 were investigated to evaluate the accuracy of methods for determining the sex of live birds. We analysed the size dimorphism of biometric traits by sex, age, and sex and age, with sex determination (n = 13,226) performed by destructive methods and age determination based on wing examination (n = 8,905). Using the minimal important differences (MID) method, we demonstrated that, during spring migration, adult females have significantly greater mass and bill length than juvenile females and adult males, as well as a significant difference in body length compared to juvenile females. No biologically relevant differences were demonstrated between the sexes or age classes for other morphometric parameters. Conditional inference trees were applied to test whether body size parameters could be used to separate the age and sex of individuals. Based on posterior probabilities (55.4%), we suggest that biometric parameters no longer provide a sufficiently reliable method to separate age classes during the spring migration. Separation of sexes showed the best results for adult birds, with bill length (85.4%) and body mass (85.2%) proving the best predictors. The inclusion of additional morphometric variables (tarsus, tail, body and wing length) in the model did not increase the reliability of sex segregation, confirming the results obtained using MID, i.e. that there is no statistically verifiable biologically relevant difference between adult male and female birds for these parameters. A methodological innovation in this study was using MIDs for comparisons to determine biological thresholds for differences, the procedure helping to exclude Type I errors and determine biological significance.

Key words: morphometrics, body size by gender, body size by age, biologically relevant significance

Introduction

As there are only slight differences between the sexes in Eurasian woodcock (*Scolopax rusticola*), it can be difficult to separate the sexes based on appearance traits such as plumage colouration and markings or leg colour alone (Clausager 1973, Cramp & Simmons 1983, Ferrand & Gossman 2009). Nevertheless, several attempts have been made to separate the sexes based on biometric and/or morphological parameters. Clausager (1973) was the first to point out the possibility of using the quotient of central tail feathers and bill length for separating the sexes. Subsequently, several studies (MacCabe & Brackbill 1973, Artmann & Schroeder 1976, Rockford & Wilson 1982) attempted to determine sex based on the size of individual body parts (e.g. bill, tail, wing measurement or body weight), though none of these allowed the sexes to be distinguished with sufficient reliability. According to Glutz von Blotzheim et al. (1977), a woodcock with a bill longer than 77 mm and a tarsus longer than 38 mm was most likely to be a female; however, no information was given on the reliability of the method. One of the most widely known and cited formulas for the separation of woodcock sexes based on morphological characteristics is that developed by Stronach et al. (1974), based on the formula ($I = [0.2952X] − [0.1566Y]$), where X is the length of the bill (in mm), and Y is the length of the tail (in mm). In this case, if the value of I is > 8.364, then the bird is a female (75% correct), and if the value of I is < 8.364, then the bird is a male (72% correct). The probability of error was 28% if birds that were not yet adults were included in the analysis. Birds of < 12 months of age may be excluded when examining the tips and proximal edges of outer primaries (ragged outline in first years; smooth on older birds, at least until April) and the terminal lighter bar on primary coverts (broader and browner on young birds). However, when all birds that had not yet undergone full moulting were excluded, only 2-5% accuracy was achieved (Shorten 1975). Considering the above criteria, it can be concluded that the applicability of the method is severely limited. To address this problem, the present paper aims to provide a morphological basis for sex determination by employing contemporary, biologically pertinent and statistically advanced methods to a large sample.

Ferrand & Gossmann (2009) obtained even worse results in a similar study. Their results showed that males, on average, have shorter bills and longer tails than females. However, the authors pointed out that there was so much overlap between the data that it was impossible to determine the sex for most birds reliably. Based on their data, a bill length > 80 mm represented a female, and a tail length > 88 mm represented a male. Further, adult birds with a tail/bill ratio ≥ 1.20 were males and females if the ratio was ≤ 1.10, while juvenile birds with a tail/bill ratio ≥ 1.20 were males and females if the ratio was ≤ 1.00. As the overlap was high, the method was only 45% accurate for adults and 25% for juveniles (Ferrand & Gossmann 2009).

Detailed statistical studies based on differences in morphometric data for other charadriiform species, such as linear models or discriminant and principal component analysis, have also not provided definite results (Remisiewicz & Wennerberg 2006, Schroeder et al. 2008). According to Hoodless (1994), the difference in body weight between sexes during the laying phase of the nesting period could be suitable for sexing some woodcock; however, the method has not proved sufficiently reliable, even during this narrow time interval (Aradis et al. 2015). Furthermore, Aradis et al. (2015) reported that discriminant function analysis applied to a set of woodcock morphometric traits failed to achieve 80% confidence in the case of juveniles and 79.6% and 77.1% for adult females and males, respectively.

Between 1983 and 1999, Faragó et al. (2000) conducted a study that drew conclusions from 1,008 birds collected during the spring hunt in Hungary. However, some biometric parameters examined were only available in sample sizes below 100. Hence, their results cannot be considered representative due to the low number of annual observations. The year 2009 marked a turning point in woodcock research when spring woodcock hunting in Hungary was put at risk due to the enforcement of the EU Birds Directive (79/409 EEC). As a condition for an exception from the Directive, the Hungarian Hunters National Association launched the Hungarian woodcock monitoring programme in 2009. In 2010, the Institute of Game Management and Vertebrate Zoology of the University of West Hungary joined the monitoring with a biometric testing module. As more than 5,000 data providers contributed data collected according to a standard protocol between 2010 and 2014, this 'new' national woodcock monitoring programme provided an unprecedented opportunity for a time-series analysis of woodcock migration based on a large sample size (n = 13,471). In a statistical analysis of this large biometric dataset, we seek to answer whether age and sex determination based on biometric traits is possible in live woodcock and, if so, how reliably these parameters indicate age and sex.

Material and Methods

Since the spring of 2010, woodcock bag monitoring, coordinated by the Hungarian Hunters National Association, has formed the basis for a nationwide, large-sample, age- and sex-differentiated study of woodcock biometrics. Biometric data were collected in March each year between 2010 and 2014 from all 19 counties of Hungary, the monitoring program targeting up to 5,600 bagged woodcock per year (for the number of birds collected per county, see Table 1). For each sample, the person responsible recorded the place where the bird was bagged (municipality and recording code), the exact time of sampling (month, day, hour and minute), and the sex of the bird. For age determination purposes, each hunter was required to send in at least 25%, and from 2011, 40%, of wings from the woodcocks he had killed, stretched and prepared. Age determination was carried out according to the widely used methodology for woodcock, based on the state and degree of moulting and the characteristic features (moulted or unmoulted) of each feather group (Glutz von Blotzheim et al. 1977, Cramp & Simmons 1983, Ferrand & Gossmann 2009). The birds were separated into 'juvenile' and 'adult' age groups, with no further detailed classification applied (Bende 2021). The recording of biometric parameters (body weight, bill length, body length, wing length, tail length and tarsus length) and the choice of instruments used for measurement were in accordance with conventional ornithological methods. Body weight (1 g accuracy) was measured using a balance scale or a letter scale, while length measurements were obtained using a standard ruler (tail length), tape measure (wing length) or calliper (bill and tarsus length) (Faragó et al. 2000). All data were sent to the Institute of Wildlife Biology and Management of the University of Sopron on standard sampling forms, together with the wing samples.

Statistical analysis

Statistical analyses were conducted in R Studio version 4.3.1 (2020), built on the R platform version 4.2.3 (R Development Core Team 2022). We first performed descriptive statistics, such as minimum, maximum, median, range, and standard deviation of the sample mean (SD) and valid cases (n), with SD values > +4 or < –4 from the mean excluded from further analysis. The cleaned dataset contained 13,471 individuals ordered in rows, with each observation being an individual.

We provide a rigorous justification for employing parametric statistical measures such as mean and SD. The central limit theorem posits that, given a sufficiently large sample size, the sampling distribution of the mean for any independent, identically distributed random variable will be approximately normally distributed, irrespective of the original distribution of the variable (Efron & Tibshirani 1993, Lumley et al. 2002, Hoekstra et al. 2014). Hence, for large samples, the mean and SD can serve as robust parameters for statistical inference without the necessity of assuming a specific (e.g. normal or t-) distribution, particularly when one considers large samples (e.g. $n \geq 10,000$). Additionally, large samples tend to mitigate the influence of extreme values, which further justifies the use of SD (Rousseeuw & Croux 1993, Wilcox 2012, Kwak & Kim 2017). We excluded extreme values using the ±4 SD criterion mentioned earlier. Therefore, any limitation of using SD to assume a normal or t-distribution does not hold empirical ground when large sample sizes are involved.

Given the high number of observations and the fact that the sampling region was counterbalanced, our dataset is statistically representative of the bird population in Hungary. Specifically, each of the 19 counties within Hungary is represented in our sample in a manner commensurate with its proportion

Table 1. Summary statistics for six biometric body measurements for female and male woodcock during the spring migration in Hungary (mean ± SD; min-max = range). All measurements are given in mm, except weight in g. 'n' = valid cases, the entire dataset containing 13,471 observations.

Body measurements	Female				Male			
	Min-Max	Median	SD	n	Min-Max	Median	SD	n
Body weight	208-419	315.5	29.2	2,222	202-420	311.5	26.2	10,955
Body length	268-410	339.1	15.7	2,217	266-410	339.2	15.3	10,929
Wing length	131-280	201.2	17.9	2,218	120-285	202.5	18.5	10,929
Tail length	48-121	84.3	8.8	2,231	48.0-120.0	85.5	8.2	10,995
Tarsus length	22.2-55.0	38.2	3.6	2,218	22.1-55.2	38.0	3.6	10,949
Bill length	57.5-88.3	73.0	4.0	2,231	57.0-88.5	72.4	3.6	10,995

in Hungary's overall migratory bird population. To ensure the statistical representativeness of the findings, the current investigation utilised the dataset furnished by Szemethy et al. (2014). Their research posited an estimated migratory population range of 1.48 to 6.89 million woodcock transiting through Hungary during the spring season, which aligns temporally with the period under scrutiny in the present study. For rigour, the upper population estimate of 6.9 million woodcock was adopted as the stringent criterion for population size. Using this parameter, the minimum requisite sample size was computed using the frequentist framework (e.g. Rosner 2015). Employing a highly rigorous confidence level of 98%, an estimated population proportion of 0.5 (as this value maximises the required sample size) and an exceptionally stringent margin of error of 1.1% (which signifies that the true population parameter is anticipated to lie within ±1.1% of the observed value), the calculated minimum sample size for achieving statistical representativeness was determined to be 11,199 individuals or more, i.e. 11,199 or more individuals are needed to have a confidence level of 98% that the real value is within ±1.1% of the measured value.

The six biometric variables examined were treated as numeric variables, while age, sex, sampling year (2010-2014), sampling month (first or second half of March) and county were treated as factor variables. All measurements of numeric variables are given in mm, except weight, which is in g. When investigating interactions between time (year and month in our study) and other variables, treating time as a factor may provide more precise estimates of these effects (Kutner et al. 2005). Second, when months or years represent distinct periods showing cyclic or seasonal trends, as in our study, treating them as factors may capture these differences effectively (Box et al. 2015). In other words, when the relationship between, for instance, 'year' and the dependent variable (i.e. the biometric parameters) is non-linear, it becomes imperative to treat 'year' as a categorical factor. This approach can be critical in capturing effects such as biological changes with abrupt impacts not captured by a linear term. When 'year' is treated as a continuous variable, the assumption is made that the gap between each year has an identical impact on the dependent variable. However, this assumption might be flawed; hence, treating 'year' and 'month' as factor variables mitigates this issue.

Treating the variable 'year' as a factor variable with four levels (representing the four consecutive years in

our study) can be further justified as advantageous. For instance, *post hoc* tests can be conducted to compare the different years with each other. These tests can offer valuable insights into which years are statistically different from each other (Hsu 1996). Finally, decision-tree models often benefit from categorical variables due to their decision-tree foundation, hence improving predictive power (Breiman 2001).

Given the large sample size of more than 10,000 individuals in the present study and the relatively high number of biometric predictors, traditional statistical procedures could lead to Type I errors, aligning with research highlighting the issue of 'p-hacking' or inflation of Type I error rates in large samples (Ioannidis 2005, Benjamini et al. 2006, Button et al. 2013). Hence, large samples may detect statistically significant but trivial effects, especially when multiple predictors are involved, thereby increasing the risk of false positives (Maxwell et al. 2008).

For pairwise comparisons (six biometric parameters grouped by sex, age and sexes by age groups), we considered that, given the large sample size, the likelihood of Type I error was very high (Sullivan & Feinn 2012, Lin et al. 2013), causing even a biologically irrelevant and negligible difference to become statistically significant. As such, *post hoc* tests would lead to Type I errors, as mentioned earlier. There are multiple solutions to combat this issue, such as bootstrapping or measurement of the Bayes factor; however, to avoid this issue, we computed estimates of minimal important difference (MID) (e.g. see Jaeschke et al. 1989, Norman et al. 2004).

We opted for this method because we aimed to determine threshold values for each biometric parameter, which the previously mentioned statistical procedures do not perform. Our second motivation for using MIDs was that, over past decades, there has been a shift from statistical significance to practical significance, or practical relevance, in the interpretation of study results (e.g. Terwee et al. 2011). Specifically, we employed the SD criterion (Crosby et al. 2003, Engel et al. 2018, Revicki et al. 2008). In the present study, MID is a measure for the smallest difference in a biological parameter that is biologically relevant, significant, meaningful, or considered biologically important. In this way, we can detect results that are the product of Type I errors and, crucially, unravel biologically meaningful/significant differences. MID can be conceived as

a cut-off point or threshold value for a biologically significant difference, with the latter being above statistical significance.

While Copay et al. (2007) and Norman et al. (2004) both suggest an SD criterion of 0.5, Farivar et al. (2004) and Eton et al. (2004) suggest an SD of 0.3 (i.e. ⅓ SD). To keep the MID as low as possible, we adopted the most liberal 0.2 SD criterion in the literature (e.g. Samsa et al. 1999, Mouelhi et al. 2020), equivalent to a small effect, allowing us to detect even minimally biologically relevant differences. To our knowledge, MIDs have not been employed in ornithology research, mainly in human medicine or studies outside the natural sciences (e.g. Fekete et al. 2018).

We first computed the pooled SD from the two independent groups (e.g. female and male or adult and juvenile), in line with recommendations on the computation of MID suggested by Watt et al. (2021) for every comparison. Next, we compared the MIDs to the estimated differences (δ) derived by Tukey HSD *post hoc* tests (Tukey multiple comparisons of means). For the computation of the pooled SD, we employed the formula in Cohen (1988), i.e. $\sigma_{pooled} = \sqrt{[(\sigma1_2) + (\sigma2_2)]/2}$, where the two symbols represent the SDs of two independent groups, e.g. male and female. The estimated difference, designated as δ, was computed using the Tukey HSD *post hoc* test (i.e. the estimated mean difference), which is considered more reliable than Tukey HSD *P*-values, which are subject to *P*-inflation (Type I error). For this study, *P*-values < 0.05 were considered statistically significant. The MID was computed using the formula MID = 0.2 * σ_{pooled} (see Watt et al. 2021 for the pooled SD). In summary, *P*-values may show Type I errors; hence, MIDs should be taken as a benchmark to interpret the meaningfulness (i.e. biological relevance) of the estimated differences.

Subsequently, we evaluated whether sex and age could be ascertained through the biometric parameters present in the dataset. Theoretically, multivariate methods could prove informative in such cases, and indeed, logistic regression on multiple traits has proved helpful in predicting sex in previous studies on Charadriiformes, with Hallgrimsson et al. (2008), for example, successfully applying general linear models (GLM) on purple sandpipers (*Calidris maritima*), and Katrínardóttir et al. (2013) on Eurasian whimbrels (*Numenius phaeopus*). However, Hallgrimsson et al. (2008) only used a sample of 222 adult birds, and Katrínardóttir et al. (2013) used an even smaller sample of 50 whimbrels. In contrast,

large datasets like ours, with many predictors, could lead to models that are too complex, capturing noise rather than the underlying data structure, which is a form of overfitting (Babyak 2004, Harrell 2015). Such overfit models, in turn, lack 'generalisability' and could result in misleading conclusions (Harrell et al. 1996).

A secondary issue concerning the use of logistic regression with a set of six predictors arises from model saturation, as outlined by Hosmer et al. (2013, section 9.2). According to these authors, a saturated model incorporates all conceivable main effects and interactive terms among the independent variables. Hosmer et al. (2013) further asserted that such saturated models are inherently unsuitable for hypothesis testing due to their inherent capacity to fit the data perfectly.

To avoid these problems with logistic regression, we employed conditional inference tree models (henceforth ctree; Hothorn et al. 2006), a feature selection decision-tree approach (Hothorn et al. 2006 or Levshina 2020). This statistical technique models the distribution of an outcome variable using a set of independent variables (predictors), which, in our case, are the biometric parameters. Ctrees can explain the outcome variable via the combination of these predictors. Ctrees have already been employed in Hungarian ornithological research (e.g. Vili et al. 2013, Kováts & Harnos 2015); however, they have not been applied yet on datasets with large sample sizes, a methodological novelty of our study given the high likelihood of Type I errors in such large samples (Sullivan & Feinn 2012, Lin et al. 2013). In large samples, ctrees yield more accurate and consistent estimates of predictor importance than logistic regression, converging towards true population parameters (Bühlmann & Yu 2003, Couronné et al. 2018).

As missing values in the outcome variable are not allowed in ctrees, we only analysed those observations with a valid value. This served as our single exclusion criterion in the ctree analysis. We also opted for using this non-parametric statistical framework as it can predict the outcome variable via a multi-hierarchy of numerous independent variables, unlike other traditional statistical procedures such as ANOVA, and because we obtain cut-off values, also known as splits or threshold values, on the significant predictors, again, unlike other traditional statistical approaches. For example, a cut-off of 184 g on body weight indicates that the sample can be partitioned into two subsamples with a cut-off of 184 g. Independent

variables not appearing in the ctree do not improve the model's accuracy in the presence of the rest of the significant independent variables. Most importantly, this statistical approach can be used without additional cross-validation (Hothorn et al. 2015). Given this latter condition, the large sample size and the statistical representativeness of our dataset, ctree models also serve as predictive models for the bird population in Hungary, highlighting a further novelty of our study. A further advantage of ctrees is that they can explain and/or predict the outcome variable without overfitting the model (Hothorn et al. 2006, 2015). Here, we built ctrees using the standard options, but increased the minimum criterion from 0.95 to 0.99 to avoid overfitting (Levshina 2020, p. 623), then applied Bonferroni-correction was applied to reduce Type I error.

In the tree representation, the classification of observations starts at the topmost node, also called node 1, which shows the strongest association with the outcome variable. The nodes at the bottom of the ctree are termed terminal nodes and display the predictions based on the model, also called posterior class probabilities or conditioned frequencies

(Hothorn et al. 2015). The total number of observations on the 'routes' is represented by 'n' at the bottom of every node of the ctree. To conduct the ctree analysis, we used the 'party' R-package with the ctree function (Hothorn et al. 2006), confining the ctree analysis to the adult sub-population, as the juvenile sub-population had yet to attain their terminal biometric parameters.

Results

Comparative analysis of body size

Sex determination was undertaken on 13,226 specimens, of which 10,995 were male, 2,231 female and 8,905 unknown (Table 1). When analysed by sex and age, two-way ANOVA for all biometric parameters except tarsus length indicated significant differences between mean values for each age group (Table 2). Given the large number of samples, deviations were not accepted unconditionally; instead, MID was used to undertake a differential analysis of the biometric parameters. Tukey HDS results showed significant differences between male and female woodcock for body weight, wing length and bill length; however, the δ values for these parameters were less than the

Table 2. Results of two-way ANOVA on six biometric parameters examined separately (main effects of sex and age and their interaction as independent variables in each model, with the biometric parameter as the dependent measure). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; n.s. = not significant; 'n' = valid cases, the entire dataset containing 13,471 observations; 'df' = degrees of freedom for the main effects.

Body measurements	n	df	Sex F	Sex P	Age F	Age P	Sex × Age F	Sex × Age P
Body weight	13,400	1	21.014	***	45.913	***	4.369	**
Body length	13,353	1	0.923	n.s.	23.287	***	2.072	n.s.
Wing length	13,345	1	6.765	**	12.931	***	0.776	n.s.
Tail length	13,351	1	7.397	***	15.799	***	0.382	n.s.
Tarsus length	13,381	1	2.975	n.s.	1.549	n.s.	1.562	n.s.
Bill length	13,385	1	28.066	***	17.759	***	2.714	*

Table 3. Comparative analysis of male and female body measurements during the spring migration of woodcock in Hungary. δ refers to the estimated absolute difference between the means computed by Tukey HSD *post hoc* comparisons. MID is computed by taking 0.2 of the pooled SD, relying on the 0.2 * SD criterion (e.g. Mouelhi et al. 2020). MIDs are rounded to two decimal places. *P*-values are derived from Tukey HSD *post hoc* tests.

Body measurements	Female Mean (± SD)	Female Range	Male Mean (± SD)	Male Range	δ	P (Tukey HSD)	MID (0.2 SD)
Body weight	315.5 ± 29.2	208-419	311.5 ± 26.2	202-420	3.98	$P < 0.001$	5.55
Body length	339.1 ± 15.7	268-410	339.2 ± 15.3	266-410	0.09	$P = 0.97$	3.10
Wing length	201.2 ± 17.9	131-280	202.5 ± 18.5	120-185	2.0	$P < 0.001$	3.64
Tail length	84.3 ± 8.8	48-121	85.0 ± 8.2	48-120	0.99	$P = 0.06$	1.70
Tarsus length	38.2 ± 3.6	22.2-55.0	38.0 ± 3.6	22.1-52.2	0.11	$P = 0.32$	0.72
Bill length	73.0 ± 4.0	57.5-88.3	72.4 ± 3.6	57.0-88.5	0.63	$P < 0.001$	0.76

Table 4. Comparative analysis of adult female and adult male body sizes during the spring migration of woodcock in Hungary. All measurements are given in mm, except weight in g. Differences in measurements between sexes by age group tested using Tukey HSD *post hoc* tests. δ indicates estimated differences from Tukey HSD *post hoc* tests.

Body measurements	Adult female		Adult male		Δ	P (Tukey HSD)	MID (0.2 SD)
	Mean (± SD)	Range	Mean (± SD)	Range			
Body weight	320.0 ± 29.8	208-419	312.9 ± 25.7	214-420	7.04	*P* < 0.001	5.57
Body length	341.1 ± 15.0	273-405	340.0 ± 15.1	266-285	1.11	*P* = 0.68	3.01
Wing length	202.5 ± 19.0	131-280	203.4 ± 18.9	120-285	0.85	*P* = 0.96	3.79
Tail length	84.8 ± 8.5	49-112	85.4 ± 8.6	48-120	0.53	*P* = 0.81	1.71
Tarsus length	38.3 ± 3.6	23.2-52.4	38.0 ± 3.5	22.3-55.0	0.22	*P* = 0.83	0.71
Bill length	73.4 ± 4.1	59-88	72.5 ± 3.6	57.0-88.5	0.95	*P* < 0.001	0.77

Table 5. Comparative analysis of juvenile female and juvenile male body sizes during the spring migration of woodcock in Hungary. All measurements are given in mm, except weight in g. Differences in measurements between sexes by age group tested using Tukey HSD *post hoc* tests. δ indicates estimated differences from Tukey HSD *post hoc* tests.

Body measurements	Juvenile female		Juvenile male		Δ	P (Tukey HSD)	MID (0.2 SD)
	Mean (± SD)	Range	Mean (± SD)	Range			
Body weight	310.7 ± 28.2	219-410	308.7 ± 26.2	207-412	1.96	*P* = 0.70	5.44
Body length	336.7 ± 15.9	270-410	338.1 ± 15.7	266-410	1.38	*P* = 0.44	3.16
Wing length	199.5 ± 17.1	148-280	201.6 ± 18.0	125-282	2.42	*P* = 0.04	3.51
Tail length	83.7 ± 8.6	48-120	84.4 ± 8.2	48-115	0.70	*P* = 0.53	1.68
Tarsus length	38.0 ± 3.7	22.2-51.4	38.0 ± 3.6	22.1-55.2	0.04	*P* = 0.99	0.73
Bill length	72.5 ± 4.0	57.5-86	72.2 ± 3.6	57-87	0.29	*P* = 0.58	0.76

Table 6. Comparative analysis of adult female and juvenile male body sizes during the spring migration of woodcock in Hungary. All measurements are given in mm, except weight in g. Differences in measurements between sexes by age group tested using Tukey HSD *post hoc* tests. δ indicates estimated differences from Tukey HSD *post hoc* tests.

Body measurements	Adult female		Juvenile male		Δ	P (Tukey HSD)	MID (0.2 SD)
	Mean (± SD)	Range	Mean (± SD)	Range			
Body weight	320.0 ± 29.8	208-419	310.7 ± 28.2	219-410	9.40	*P* < 0.001	5.80
Body length	341.1 ± 15.0	273-405	336.7 ± 15.9	270-410	4.33	*P* < 0.001	3.09
Wing length	202.5 ± 19.0	131-280	199.5 ± 17.1	148-280	3.30	*P* = 0.02	3.61
Tail length	84.8 ± 8.5	49-112	83.7 ± 8.6	48-120	1.10	*P* = 0.23	1.71
Tarsus length	38.3 ± 3.6	23.2-52.4	38.0 ± 3.7	22.2-51.4	0.26	*P* = 0.91	0.73
Bill length	73.4 ± 4.1	59-88	72.5 ± 4.0	57.5-86	0.98	*P* < 0.001	0.81

MID values, suggesting that, while the results were statistically significant, they were not biologically relevant (Table 3).

For the adult age group, while we again recorded a significant difference in body weight and bill length between males and females (Table 4; *P* < 0.001 and δ > MID), significant differences in other body size parameters between adult males and females did not reach the biologically relevant threshold (i.e. δ < MID). While no significant differences were observed for any biometric parameter between juvenile males and females (Table 5), we recorded significant differences in body weight, body length, and bill length between adult and juvenile females, which proved to be biologically relevant (Table 6).

Comparative analysis by sex and by sex and age showed that the differences observed could only be confirmed at the age group level and only for three biometric variables. Even small differences between biometric parameters of juvenile and adult birds, typically below the level of biologically relevant significance, were enough to mask differences between
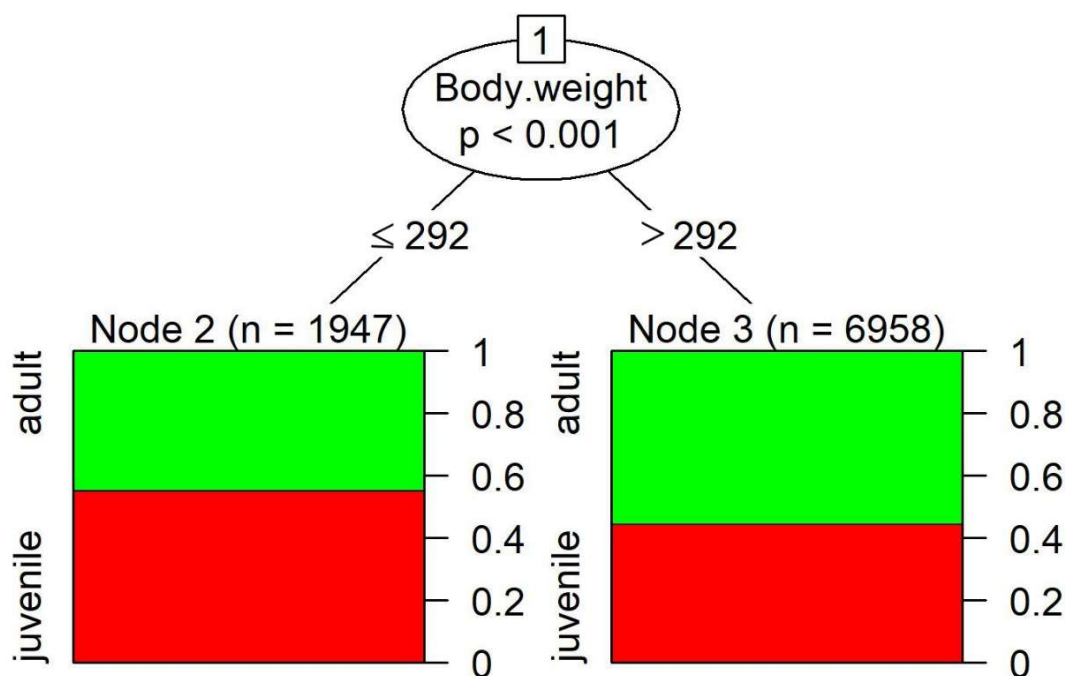
Fig. 1. Conditional inference tree analysis predicting age ('adult' and 'juvenile'). After the removal of missing age values, the analysis contained 8,905 observations. Body weight (node 1 at the top of the ctree), with a cut-off of 292 g, is the only statistically significant predictor of age among the biometric variables of body weight, body length, wing length, tail length, bill length and tarsus length entered in the ctree model ($P < 0.001$, Bonferroni-corrected). Nodes 2 and 3 are bar plots representing the posterior probabilities (predictions), where red represents the probability of the animal being 'juvenile', while green represents 'adult'. The green areas (nodes 2 and 3) represent the probability of the animal being an adult (0.554 and 0.446, respectively), while the red areas (nodes 2 and 3) represent the probability of the animal being a juvenile (0.148 and 0.264, respectively). The red areas can also be conceived of as prediction 'error'. The y-axis represents the posterior probabilities for the two classes of age.

the sexes, and thus only age-class differences are detected. In addition to the two biometric parameters above, average body weight and bill length in adult females were always significantly different, as was body length compared to juvenile females.

Predicting age in the entire sample, using a conditional inference tree based on six potential biometric explanatory variables

While investigating the extent to which small morphological differences allowed separation of sexes and age classes, six biometric predictors were entered into the ctree model, i.e. body weight, body length, wing length, tail length, bill length and tarsus length, with age at two levels, 'adult' and 'juvenile', serving as binary dependent variables. Since ctrees do not allow missing values on the outcome measure (i.e. age in the present analysis), we removed those cases where age was missing. After removing such cases, 8,905 observations remained in the ctree analysis (Fig. 1).

The predictions, i.e. the posterior probabilities of being 'adult' or 'juvenile', from the ctree analysis indicated body weight as the only statistically significant predictor (node 1; Fig. 1), this displaying the strongest association with age ($P < 0.001$, Bonferroni-corrected) in the presence of the other biometric variables. In other words, adding more variables to the model from the set of variables entered did not improve the predictive accuracy of the ctree model. The cut-off for body weight in predicting age yielded a value of 292 g (Body weight ≤ 292 g; criterion = 1, statistic = 82.916), i.e. birds weighing > 292 g (6,958 observations in our sample) were statistically more likely to be adults than juveniles, with a posterior probability of 55.4% (see Fig. 1). Node 2, which contained 1,947 observations of both adult and juvenile birds, indicated a 44.6% posterior probability of being an adult (Fig. 1). Crucially, the body weight of these 1,947 observations was ≤ 292 g. The other 'branch' (node 3, Fig. 1) could be interpreted by the same logic. While our results demonstrate that, of all the biometric parameters for age determination included in the analysis, only body weight showed a small but significant difference between juvenile and adult age groups. Nevertheless, this variation lacked enough empirical weight to be a reliable discriminator between the two age groups (Fig. 1).

Predicting sex in adult woodcock using a conditional inference tree based on six potential biometric explanatory variables

Of the 13,471 total observations, 73 were removed as sex evaluations were missing, giving 13,398 observations for the ctree analysis. A comparative analysis by sex and sex plus age indicated that differences could only be confirmed at the age group level and only for three biometric variables. For adult females, body weight and bill length always showed a significant difference between averages, and a significant difference was also recorded for body length when compared to juvenile females. The adult sub-sample in the ctree analysis contained 4,712 observations, of which 3,944 were male and 768 female. Ctree analysis was used to explain and predict the distribution of sexes as the outcome variable using the same set of biometric variables as in the age analysis, i.e. body weight, body length, wing length, tail length, bill length and tarsus length, but with sex serving as the binary dependent variable.

As preliminary studies have shown a significant difference between some biometric parameters of juvenile and adult birds, we decided to investigate the possibility of sex separation in adult birds only as small morphometric differences between age classes could bias morphometric differences between the sexes. Once again, the ctree analysis revealed that only body weight separated females from males (Fig. 2; $P < 0.001$, Bonferroni-corrected, criterion = 1, statistic = 44.901), with no other statistically significant biometric predictors of sex in the model. Adding factor variables for month (two levels) and year of sampling had no effect on the model outcomes.

Ctree node 2 comprised 4,098 observations and indicated an 85.2% posterior probability of a bird being a male, while node 3 comprised 614 observations and indicated a 73.6% posterior probability of being a male (Fig. 2). The posterior probabilities for being a female were 14.8% if body weight was ≤ 343 g (node 2) and 26.4% if body weight was > 343 g (node 3; Fig. 2).

Predicting sex in the adult sample using a conditional inference tree with bill length and tail length as potential biometric explanatory variables

While we had 4,712 observations in the ctree analysis (adult subsample), there was a class imbalance on the distribution of the sexes, with 3,944 males and 768 females. As in the previous analysis, we aimed to explain and predict sex as the outcome variable using
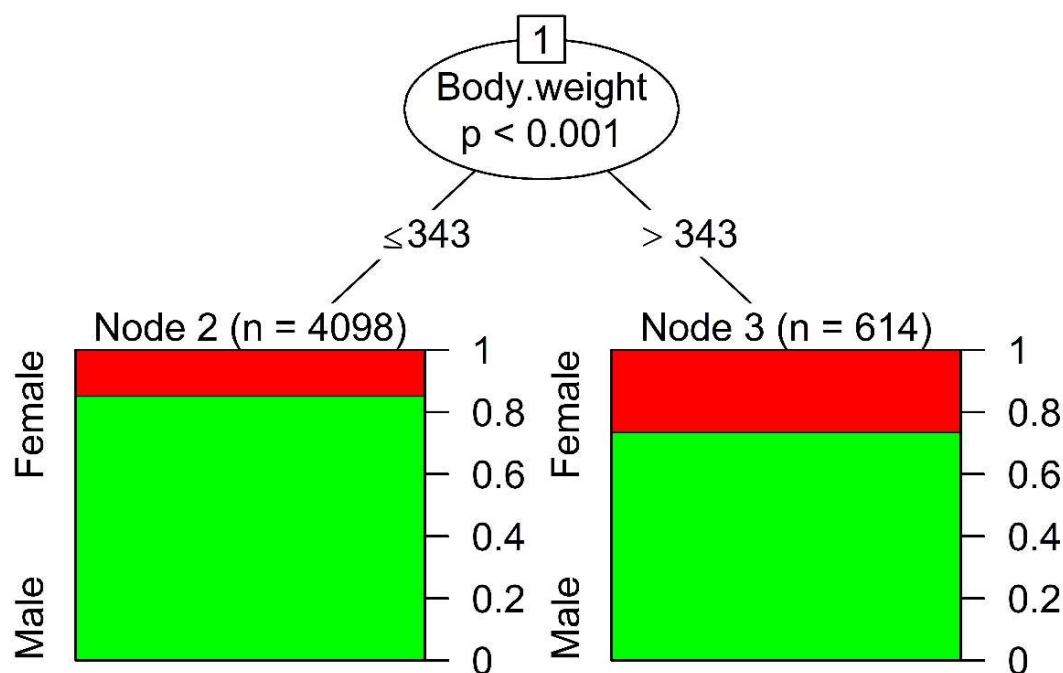

Fig. 2. Conditional inference tree analysis predicting sex in the adult subsample. Total number of observations in the analysis = 4,712. Body weight (node 1), with a cut-off of 343 g, is the only statistically significant predictor of sex among the biometric variables of body weight, body length, wing length, tail length, bill length and tarsus length entered in the ctree model ($P < 0.001$, Bonferroni-corrected for the variable body weight). The green areas (nodes 2 and 3) represent the probability of the animal being a male (0.852 and 0.736, respectively), while the red areas (nodes 2 and 3) represent the probability of the animal being a female (0.148 and 0.264, respectively). The red areas can also be conceived of as prediction ´error´. The y-axis represents the posterior probabilities for the two sex classes.
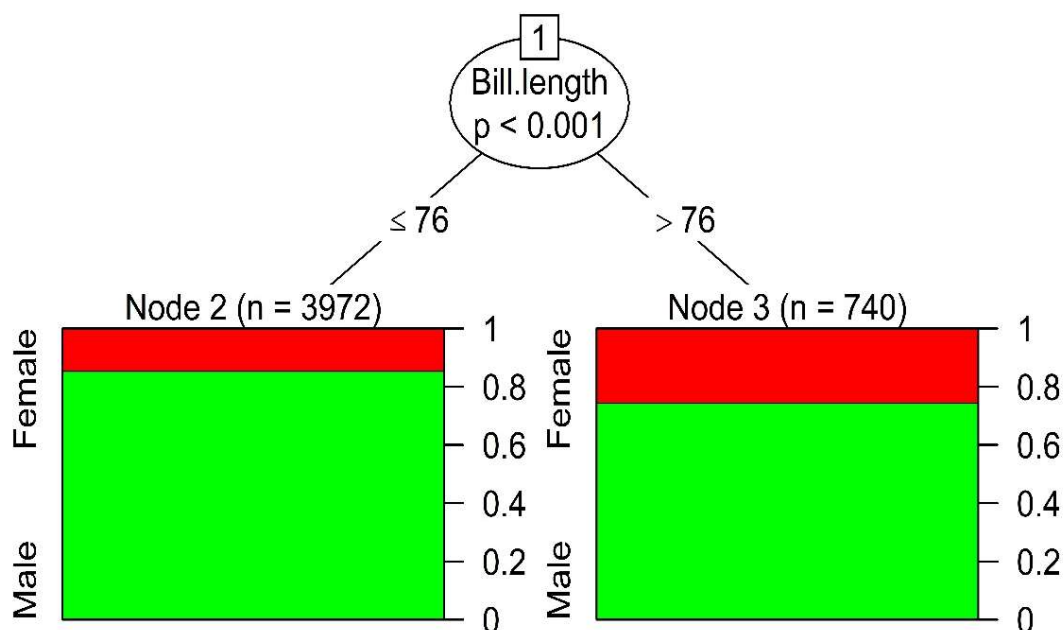
Fig. 3. Conditional inference tree analysis predicting sex in the adult subsample (total number of observations = 4,712). Bill length (node 1), with a cut-off of 76 mm, is the only statistically significant predictor of sex (P < 0.001, Bonferroni-corrected) if bill length and tail length are entered in the model. Tail length did not improve the accuracy of the model. The green areas (nodes 2 and 3) represent the probabilities of the animal being a male (0.854 and 0.745, respectively), while the red areas (nodes 2 and 3) represent the probability of the animal being a female (0.146 and 0.255, respectively). The red areas can also be conceived of as prediction 'error'. The y-axis represents the posterior probabilities for the two sex classes.

the same set of biometric variables as in the previous analyses. Two potential biometric predictors were fed into the model, i.e. tail length and bill length, with sex (two levels, 'female' and 'male') serving as the binary dependent variable. The ctree analysis revealed a significant difference between female and male bill length only (P < 0.001, Bonferroni-corrected, criterion = 1, statistic = 41.796; Fig. 3), tail length having no effect on the model's accuracy. Moreover, tail length proved not to be a significant predictor of sex (Fig. 3). Node 2 of the ctree analysis comprised 3,972 observations, indicating an 85.4% posterior probability of a bird being male, while node 3 comprised 740 observations and indicated a 74.5% posterior probability of being a male (Fig. 3). The posterior probability of being a female was 14.6% if bill length was ≤ 76 mm (node 2) and 25.5% if bill length was > 76 mm (node 3; Fig. 3).

Discussion

Aradis et al. (2015), who compared a small number of woodcocks (n = 259) during the overwintering period to explore the extent of variation between sexes and age classes, found that while several morphometric traits differed noticeably between sexes (wing, bill, tarsus length) and age classes (wing), no significant differences were observed between sexes, ages or

their interaction (orthogonal contrasts). Using the same morphometric traits, we examined 13,226 samples from the March hunting bag in Hungary and found that only age-differentiated analyses demonstrated biologically significant differences. The results of *post hoc* tests showed that adult female body weight and bill length were significantly higher than those for both juvenile females and male age groups. In previous Hungarian investigations (1996-1999), significant mass differences could not be consistently confirmed for smaller samples of between 78 and 364 birds (Faragó et al. 2000). In 1999, however, Faragó et al. (2000) recorded a significant difference in body length in favour of females older than one year compared to younger females (P < 0.01). For younger birds, the same authors found no significant differences in morphometric characteristics between the sexes (Faragó et al. 2000). In the study of Aradis et al. (2015), no significant differences in body weight were observed between sexes or age groups in wintering areas in Italy. Nevertheless, other studies suggest that differences in weight between the sexes may be due to the start of egg growth (e.g. Hoodless 1994). Our study detected an initial follicle production stage during destructive sex determination, indicating that egg formation had not yet begun; hence, this did not affect the sex differences. We obtained the same results for bill length as Aradis et al. (2015).

Application of MID confirmed biologically relevant morphometric differences in body weight, bill length and body length parameters indicated by the predictor variables defined in the ctree analysis, the morphometric parameters with highest variance, selected according to the decision rules they define, and the results of the segregation into the groups they divide. The results were consistent despite the *post hoc* tests examining differences in means between the two groups and the statistical significance of this difference (i.e. they focus on group-level comparison). At the same time, ctrees capture non-linear relationships between variables that *post hoc* tests fail to indicate due to linear assumptions.

In the case of woodcock, there is not enough sexual dimorphism to separate the sexes through visual inspection (Cramp & Simmons 1983); hence morphometric parameters have typically been used to separate the sexes in this species (Stronach et al. 1974, Rochford & Wilson 1982, Hoodless 1994). The first results on the identification of sexes based on morphometric differences were published by Stronach et al. (1974), who, based on an equation, reported 75% reliability for female identification and 72% for male identification. Using a linear model with empirical multipliers calculated from bill and tail length from our data, we were able to determine sex with relatively low confidence, the model reliability for adult birds being 59.0% (n = 4,702) and that for juveniles 58.4% (n = 4,121). Glutz von Blotzheim et al. (1977), using a simpler approach for biometric sex identification, stated that if a woodcock's beak was > 77 mm long and the tarsus > 38 mm long, the specimen was typically female. However, our results show no statistically verifiable difference in tarsus length between adult males and females ($P = 0.83$; $\delta = 0.22 < MID = 0.71$). Ferrand & Gossmann (2009) found that male bills were, on average, shorter (male bill length > 80 mm) than those of females but that the rectrices were longer (male tail length > 88 mm). Based on the results of our large-sample investigation, we found that adult females had significantly longer bill lengths than adult males ($P < 0.001$; $\delta = 0.95 > MID = 0.77$), but no significant difference in bill length between juvenile birds ($P = 0.58$; $\delta = 0.29 < MID = 0.76$). In addition, attempts were made to separate individual sexes based on the ratio of morphometric parameters (tail length/bill length ≤ 1.20 = female); however, even when restricted to the adult age group, reliability for sex determination based on morphometric parameters was no better than 45%. In comparison, the model developed and applied by Aradis et al. (2015) was applicable with a confidence level of 77.1% for adult male birds and 79.6% for females. Our validated ctree model and MID results produced the same conclusion, i.e. no morphometric variable or combination of variables could predict age with high confidence. Instead, body weight was the best predictor in the total sample of known age (n = 8,905), with a separation point at 292 g. For birds > 292 g (n = 6,958), the model predicted age with 55.4% confidence and 44.6% confidence for birds weighing < 292 g (n = 1,947).

To separate the sexes, the ctree analysis was performed on a dataset restricted to adult birds (n = 4,712) while also taking account of MID results. In this study, several biometric parameters could indicate sex with high confidence, with bill length found to be the strongest predictor, the sexes separating at a cut-off value of 76 mm. Our results further indicated that if the bill length was ≤ 76 mm (n = 3,972), the model had an 85.4% probability of correctly predicting sex, and if the bill length was > 76 mm (n = 740), the model had 74.5% reliability. In addition to bill length, body weight proved a strong predictor, separating the sample with a cut-off value at 343 g. In our sample, if body weight was > 343 g (n = 4,098), the model predicted sex with 85.2% confidence, while confidence was 73.6% for birds of ≤ 343 g (n = 614). However, while body weight was a significant predictor, its contribution to enhancing the model's predictive power was not substantial.

Even when using a large number of samples, we could not achieve a result of more than 85% confidence in age estimation, even when using the best morphometric predictor variables, despite the novelty of the statistical procedures used in this ornithological application. On the other hand, our results confirmed that there is statistically verifiable and biologically relevant morphometric variation in woodcock. However, the extent of this variation is not sufficient to separate the sexes with adequate reliability. In ornithological work (e.g. ringing, telemetry transmitters), knowledge of the bird's sex is highly desirable; however, morphometric characteristics do not allow us to determine this with sufficient reliability using any of the methods presently available. This finding suggests that using semi-invasive techniques may still be relevant in ornithology, e.g. DNA analysis of blood and feather samples, as these allow sex segregation with 100% confidence (Bende et al. 2023).

We consider the present study important due to its methodological novelty in using MIDs, which helped

us determine thresholds/cut-off values of biological significance for estimated differences beyond mere statistical significance; this method has been underemployed in ornithology to date. Furthermore, MIDs allowed us to rule out Type I errors during the analysis. Future ornithological research should incorporate MIDs to determine meaningful differences in large samples.

Acknowledgements

Author Contributions

S. Faragó planned and organised the national-scale research and acquired funding. A. Bende and R. László compiled the database and carried out age determination based on the wing samples. I. Fekete conceptualised and executed the statistical analysis. A. Bende and I. Fekete wrote the paper. All authors approved the final version of the manuscript.

Literature

Aradis A., Landucci G., Tagliavia M. & Bultrini M. 2015: Sex determination of Eurasian woodcock *Scolopax rusticola*: a molecular and morphological approach. *Avocetta 39: 83–89.*

Artmann J.W. & Schroeder L.D. 1976: A technique for sexing woodcock by wing measurement. *J. Wildl. Manage. 40: 572–574.*

Babyak M.A. 2004: What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom. Med. 66: 411–421.*

Bende A. 2021: Spring migration dynamics, age and sex ratio, and breeding biology of the woodcock (*Scolopax rusticola* L.) in Hungary. *PhD thesis, University of Sopron, Sopron, Hungary. (in Hungarian with English abstract)*

Bende A., Pálinkás-Bodzsár N., Boa L. & László R. 2023: Sex determination of Eurasian woodcock (*Scolopax rusticola* L.) by genetic and imaging diagnostic methods. *Biodiversity & Environment 15: 29–35.*

Benjamini Y., Krieger A.M. & Yekutieli D. 2006: Controlling the false discovery rate in large-scale multiple testing. *J. R. Stat. Soc. B: Stat. Methodol. 68: 405–416.*

Box G.E.P., Jenkins G.M., Reinsel G.C. & Ljung G.M. 2015: Time series analysis: forecasting and control, 5th ed. *John Wiley & Sons, Hoboken, New Jersey, USA.*

Breiman L. 2001: Random forests. *Mach. Learn. 45: 5–32.*

Button K.S., Ioannidis J.P., Mokrysz C. et al. 2013: Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci. 14: 365–376.*

Bühlmann P. & Yu B. 2003: Boosting with the L2 loss: regression and classification. *J. Am. Stat. Assoc. 98: 324–339.*

Clausager I. 1973: Age and sex determination of the woodcock, *Scolopax rusticola*. *Dan. Rev. Game Biol. 8: 3–18.*

Cohen J. 1988: Statistical power analysis for the behavioral sciences, 2nd ed. *Lawrence Erlbaum Associates, Hillsdale, New Jersey, USA.*

Copay A.G., Subach B.R., Glassman S.D. et al. 2007: Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J. 7: 541–546.*

Couronné R., Probst P. & Boulesteix A.L. 2018: Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics 19: 270.*

Cramp S. & Simmons K.E.L. 1983: Handbook of the birds of Europe, the Middle East and North America: the birds of the Western Palearctic. Waders to gulls, vol. 3. *Oxford University Press, Oxford, UK.*

Crosby R.D., Kolotkin L.R. & Williams G.R. 2003: Defining clinically meaningful change in health-related quality of life. *J. Clin. Epidemiol. 56: 395–407.*

Efron B. & Tibshirani R.J. 1993: An introduction to the bootstrap. *Chapman & Hall, New York, USA.*

Engel L.D., Beaton E. & Touma Z. 2018: Minimal clinically important difference: a review of outcome measure score interpretation. *Rheum. Dis. Clin. N. Am. 44: 177–188.*

Eton D.T., Cella D., Yost K.J. et al. 2004: A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *J. Clin. Epidemiol. 57: 898–910.*

Faragó S., László R. & Sándor Gy. 2000: Body dimensions, sex and age relationships of the woodcock *(Scolopax rusticola)* in Hungary between 1990-1999. *Hungarian Waterfowl Publications 6: 409–461. (in Hungarian with English abstract)*

Farivar S.S., Liu H. & Hays R.D. 2004: Half standard deviation estimate of the minimally important difference in HRQOL scores? *Expert Rev. Pharmacoeconomics Outcomes Res. 4: 515–523.*

Fekete I., Schulz P. & Ruigendijk E. 2018: Exhaustivity in single bare wh-questions: a differential-analysis of exhaustivity. *Glossa 3: 96.*

Ferrand Y. & Gossmann F. 2009: Ageing and sexing series 5: ageing and sexing the Eurasian woodcock *Scolopax rusticola*. *Wader Study Group Bull. 116: 75–79.*

Glutz von Blotzheim U.N., Bauer K.M. & Bezzel R. 1977: Handbuch der Vögel Mitteleuropas, vol. 7. *AULA Verlag, Wiesbaden, Germany.*

Hallgrimsson G.T., Palsson S. & Summers R.W. 2008: Bill length: a reliable method for sexing purple sandpipers. *J. Field Ornithol. 79: 87–92.*

Harrell F.E., Jr. 2015: Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis, 2nd ed. *Springer, New York, USA.*

Harrell F.E., Jr., Lee K.L. & Mark D.B. 1996: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med. 15: 361–387.*

Hoekstra R., Morey R.D., Rouder J.N. & Wagenmakers E.J. 2014: Robust misinterpretation of confidence intervals. *Psychon. Bull. Rev. 21: 1157–1164.*

Hoodless A.N. 1994: Aspects of the ecology of the European woodcock *Scolopax rusticola* L. *PhD thesis, Durham University, UK.*

Hosmer D.W., Jr., Lemeshow S. & Sturdivant R.X. 2013: Applied logistic regression, 4th ed. *John Wiley & Sons, Hoboken, New Jersey, USA.*

Hothorn T., Hornik K., Strobl C. & Zeileis A. 2015: "Party: a laboratory for recursive partytioning." Version 1.0-20. *http://cran.r-project.org/web/packages/party/*

Hothorn T., Hornik K. & Zeileis A. 2006: Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Stat. 15: 651–674.*

Hsu J.C. 1996: Multiple comparisons: theory and methods. *Chapman & Hall, London, UK.*

Ioannidis J.P. 2005: Why most published research findings are false. *PLOS Med. 2: e124.*

Jaeschke R., Singer J. & Guyatt G.H. 1989: Measurement of health nature: ascertaining the minimal clinically important difference. *Control. Clin. Trials 10: 407–415.*

Katrínardóttir B., Pálsson S., Gunnarsson T.G. & Sigurjónsdóttir H. 2013: Sexing Icelandic whimbrels numenius *Phaseolus islandicus* with DNA and biometrics. *Ringing Migr. 28: 43–46.*

Kováts D. & Harnos A. 2015: Morphological classification of conspecific birds from closely situated breeding areas – a case study of the common nightingale. *Ornis Hung. 23: 20–30.*

Kutner M.H., Nachtsheim C.J., Neter J. & Li W. 2005: Applied linear statistical models, 5th ed. *McGraw-Hill/Irwin, New York, USA.*

Kwak S.K. & Kim H. 2017: Statistical data analysis using SAS: intermediate statistical methods. *SAS Institute, Cary, USA.*

Levshina N. 2020: Conditional inference trees and random forests. In: Paquot M. & Gries S.T. (eds.), A practical handbook of corpus linguistics. *Springer, Cham, Germany: 611–643.*

Lin M., Lucas H.C. & Shmueli G. 2013: Too big to fail: large samples and the *P*-value problem. *Inf. Syst. Res. 24: 906–917.*

Lumley T., Diehr P., Emerson S. & Chen L. 2002: The importance of the normality assumption in large public health data sets. *Annu. Rev. Public Health 23: 151–169.*

MacCabe R.A. & Brackbill M. 1973: Problems in determining sex and age of European woodcock. *Proceeding of the 10th International Congress of Game Biology, Office National de la Chasse, Paris: 619–637.*

Maxwell S.E., Kelley K. & Rausch J.R. 2008: Sample size planning for statistical power and accuracy in parameter estimation. *Annu. Rev. Psychol. 59: 537–563.*

Mouelhi Y., Jouve E., Castelli C. & Gentile S. 2020: How is the minimal clinically important difference established in health-related quality of life instruments? Review of anchors and methods. *Health Qual. Life Outcomes 18: 136.*

Norman G.R., Sloan J.A. & Wyrwich K.W. 2004: The truly remarkable universality of half a standard deviation: confirmation through another look. *Expert Rev. Pharmacoeconomics Outcomes Res. 4: 581–585.*

R Development Core Team 2022: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.*

Rochford J.M. & Wilson H.J. 1982: Value of biometric data in the determination of age and sex in the woodcock *(Scolopax rusticola). United States Fish and Wildlife Service, Research Report 14, Pennsylvania, USA: 158–167.*

Remisiewicz M. & Wennerberg L. 2006: Differential migration strategies of the wood sandpiper (*Tringa glareola*): genetic analyses reveal sex differences in morphology and spring migration. phenology. *Ornis Fenn. 83: 1–10.*

Revicki D.A., Hays R.D., Cella D. & Sloan J. 2008: Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J. Clin. Epidemiol. 61: 102–109.*

Rosner B. 2015: Fundamentals of biostatistics, 8th ed. *Cengage Learning, Boston, USA.*

Rousseeuw P.J. & Croux C. 1993: Alternatives to the median absolute deviation. *J. Am. Stat. Assoc. 88: 1273–1283.*

Samsa G., Edelman D., Rothman M.L. et al. 1999: Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *PharmacoEconomics 15: 141–155.*

Schroeder J., Lourenço P.M., van der Velde M. et al. 2008: Sexual dimorphism in plumage and size in black-tailed godwits *Limosa limosa limosa. Ardea 96: 25–37.*

Shorten M. 1975. Woodcock research group (IWRB). *Wader Study Group Bull. 15: 12.*

Stronach B., Harrington D. & Wilhsnes N. 1974: An analysis of Irish woodcock data. *Proceedings of the 5th American Woodcock Workshop, University of Georgia, Athens, USA.*

Sullivan G.M. & Feinn R. 2012: Using effect size - or why the *P* value is not enough. *J. Grad. Med. Educ. 4: 279–282.*

Szemethy L., Schally G., Bleier N. et al. 2014: Results of Hungarian woodcock monitoring. *Rev. Agric. Rural Dev. 3: 12–19.*

Terwee C.B., Terluin B., Knol D.L. & de Vet H.C. 2011: Combining clinical relevance and statistical significance for evaluating quality of life changes in the individual patient. *J. Clin. Epidemiol. 64: 1465–1467; author reply 1467–1468.*

Vili N., Nemesházi E., Kovács S. et al. 2013: Factors affecting DNA quality in feathers used for non-invasive sampling. *J. Ornithol. 154: 587–595.*

Watt J.A., Veronik A.A., Tricco A.C. et al. 2021: Using a distribution-based approach and systematic review methods to derive minimum clinically important differences. *BMC Med. Res. Methodol. 21: 41.*

Wilcox R.R. 2012: Introduction to robust estimation and hypothesis testing, 3rd ed. *Academic Press, Amsterdam/Boston, USA.*