Article (refereed) - postprint

Broughton, Richard K.; Clark, Jacquie A. 2018. **Understanding potential sources of bias and error in the biometric sexing of birds**.

© 2018 British Trust for Ornithology

This version available http://nora.nerc.ac.uk/519325/

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at http://nora.nerc.ac.uk/policies.html#access

This is an Accepted Manuscript of an article published by Taylor & Francis Group in *Ringing & Migration* on 05/03/2018, available online: https://doi.org/10.1080/03078698.2017.1438367

Contact CEH NORA team at
noraceh@ceh.ac.uk

The NERC and CEH trademarks and logos ('the Trademarks') are registered trademarks of NERC in the UK and other countries, and may not be used without the prior written consent of the Trademark owner.

Understanding potential sources of bias and error in the biometric sexing of birds

Richard K Broughton[a]* and Jacquie A Clark[b]

[a]Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK

[b]BTO, The Nunnery, Thetford, Norfolk, IP24 2PU, UK

* Correspondence author. Email: rbrou@ceh.ac.uk

Short title: Bias and error in sexing

Keywords: founder effect, human error, life history, Marsh Tit, ringing

ABSTRACT

Biometrics, particularly wing length, are commonly used to assign the sex of many birds in species where the plumage is similar for males and females. Virtually all species show some measurement overlap between the sexes, however, and measurement error can add further uncertainty into datasets. This can result in individuals being misclassified as the wrong sex and introducing bias into subsequent analyses, particularly if the life histories of misclassified birds differ from those of others. We used the Marsh Tit *Poecile palustris* as a case study to examine potential sources of error and bias when assigning sex based on wing length. There was no evidence for a heritability of wing lengths that could result in localised populations of atypical size via a 'founder effect', which would otherwise undermine biometric sexing. Additionally, potentially misclassified birds did not differ from others in how long they persisted in the local population, so avoiding any potential bias of misclassification in demographic analyses. Compared to Marsh Tit data collected during intensive studies, the national dataset pooled from contributors across Britain showed much greater variation in wing lengths, resulting from wider variation in the accuracy of measurement and recording. This variation in pooled data can have implications for analyses, and we discuss the importance of data quality in ringing schemes.

INTRODUCTION

For bird species where the plumage is similar in both sexes, distinguishing males from females poses a challenge for fieldworkers and analysts. The ability to separate the sexes can be crucial, however, when investigating population structure and differences in survival, dispersal or social organisation (e.g. Broughton e*t al* 2010, Broughton *et al* 2015). Sexing birds can also be important when selecting individuals for detailed study, such as radio-tracking (Holt *et al* 2012, Broughton *et al* 2014) or satellite-tagging (Hewson *et al* 2016).

A range of methods are available for sexing birds with similar plumage, but their usefulness varies with the circumstances. Molecular sexing using DNA is reliable but usually requires laboratory processing of samples. Laparoscopy can be used on larger species in the field (Richner 1989) but is invasive and requires specialist skills. Sexing by autopsy has an obvious limitation for most population studies, in that the bird must be dead, and it can also give incorrect results (Haftorn 1982). Many monomorphic species can be sexed using breeding or territorial behaviour, but individuals must be marked, e.g. with colour-rings, and later observed in the field (Broughton *et al* 2010). The presence of a brood patch (BP) or shape of the cloacal protuberance (CP) is another common field method, but is limited to the breeding season and can be ambiguous (Redfern & Clark 2001). Finally, biometrics, including wing length, have been used to sex a wide variety of species based on measurement thresholds or mathematical estimates, including raptors (Prytherch & Roberts 2015), waders (Katrínardóttir *et al* 2013), seabirds (Craik 1999, Hallgrimsson *et al* 2016) and various passerines (Ormerod *et al* 1986, Madsen 1997, Fletcher & Foster 2010, Amouret *et al* 2015).

A limitation of sexing birds based on the length of the wing alone is that almost all studies report some measurement overlap between males and females. For species where males are generally larger, this means that some small males and large females will be misclassified (Haftorn 1982, du Feu & du Feu 2014). Another problem is that measurement criteria reported from one location might not be applicable for the same species elsewhere, because of size variation related to subspecies, migratory behaviour, or measurement technique amongst ringers (Ormerod *et al* 1986, Gosler *et al* 1995, Morgan 2004, Ellrich *et al* 2010, Broughton *et al* 2016a). Nevertheless, some species, including small passerines, can have consistent wing lengths across large areas. For example, in the strictly sedentary Willow Tit *Poecile montanus* there was no difference in wing length over an 800 km range in Norway (Haftorn 1982), and similar results were found for the Marsh Tit *P. palustris* over 540 km in Britain (Broughton *et al* 2016b). These results show that, for some species at least, the same measurement threshold could be used to assign sex across large geographical areas (though see Ellrich *et al* 2010).

The Marsh Tit has been the subject of a number of studies across Europe that have used wing length to assign sex to individuals, and all reported a reliability of 92-98% when calibrated for subspecies (Amann 1980, Nilsson 1992, King & Muddeman 1995, du Feu & du Feu 2014, Broughton *et al* 2016a). The ability to sex Marsh Tits is useful, as the species has been the subject of detailed ecological study over many decades and is of conservation concern in Britain following a long-term decline in abundance (Broughton & Hinsley 2015).

Assigning sex to British Marsh Tits *P. p. dresseri* using maximum-chord wing length was first proposed by Gosler & King (1989) and later validated by King & Muddeman (1995) and Broughton *et al* (2008), who recommend a division of 62 mm or less for females and 63 mm or greater for males. Broughton *et al* (2016a) expanded this approach to give probabilities of correct sexing for birds of any given wing length in different age classes, which ranged from 63-100%.

Despite the attraction of using biometrics to sex species such as Marsh Tit, however, important questions have been raised regarding the limitations of the wing length biometric for sexing birds in general. The Marsh Tit was used as a case study by du Feu & du Feu (2014) to propose that significant error in sexing could occur due to measurement overlap, inaccurate measurement by ringers, and local variation in the size of birds due to 'founder effects' or clinal variation. Such error could result in serious bias entering analyses of biometric data if the life histories of birds that were misclassified differed from those which were sexed correctly.

Highlighting an example of three ‘small’ male Marsh Tits that would have been sexed incorrectly as females based on wing length, du Feu & du Feu (2014) calculated that these birds did not appear to persist in the local population for as long as 13 typical males. It was further noted that 29% of British Marsh Tits sexed according to BP/CP in the BTO database (Robinson 2015) would have been classified incorrectly if using the simple 62/63 mm division in wing length. This database, pooled from ringers across Britain, included 32% of apparent males with a measured wing length below the 63 mm threshold, and 26% of apparent females with a recorded wing length greater than 62 mm. Assuming that the BP/CP sexing was correct, this error rate in sexing based on wing length was substantially higher than the 2-8% reported from all of the detailed biometric studies of Marsh Tits (see above).

The implication of this large degree of apparent error in the national database is that some of the potential sources of bias suggested by du Feu & du Feu (2014) may be genuine and significant, such as human error in sexing and wing length measurement, or variation in wing length among British Marsh Tits. However, since a more recent study has established that there is no pattern of clinal or regional variation in Marsh Tit biometrics within Britain (Broughton *et al* 2016b), this suggests that the variation in sexing accuracy in the BTO database is more likely due to human factors of measurement and recording.

To better understand the limitations of using wing length to assign sex, and to test the potential sources of bias and error, we again used the Marsh Tit as a case study. A large sample of biometrics from Marsh Tits of known age and sex was available from a detailed population study to address the questions of variation in measurements and life histories among birds. Additionally, the British ringing records for Marsh Tit, held by the BTO, were used to assess variation in measurements, sexing and recording by ringers. The results are used to discuss the wider merits of sexing birds using biometrics and the importance of accurate data collection in ringing schemes.

METHODS

Marsh Tit biometrics and life histories

Between 2003 and 2015 the wing lengths (maximum chord, 1 mm precision) and life histories of 355 individual Marsh Tits of known age and sex were recorded during a population study in Cambridgeshire, eastern England. This work was centred on the 157 ha Monks Wood National Nature Reserve (52°24'N, 0°14'W) and included five other woods within a 5 km distance: Odd Quarter, Upton, Bevill's, Wennington and Holland Woods. Measurement recording was highly consistent, with 93% of wing lengths collected by a single experienced ringer (RKB) and cross-checking was performed between other ringers taking part in data collection.

Marsh Tits were ringed as nestlings or caught in baited traps throughout the year and fitted with a BTO metal ring and individual combinations of colour rings (Broughton *et al* 2010). Ageing as a juvenile (EURING code 3 or 5, du Feu *et al* 2015) or adult (codes 4 or 6) was based on the shape of the tail feathers and the presence or absence of juvenile greater coverts, which was confirmed by birds ringed as nestlings (Broughton *et al* 2008, Broughton 2010). Birds were sexed according to territorial and breeding behaviour observed in the field (Broughton *et al* 2008, 2010). Sexing was considered accurate, as DNA analysis of 55 birds all tallied with the sexing based on behaviour (Broughton *et al* 2016a).

Marsh Tit life histories and local persistence were documented during intensive fieldwork throughout the year, which was designed to cover the entire study area using playback surveys, territory mapping and nest monitoring during spring and summer, and with systematic recording and trapping at feeders from late summer to late winter (Broughton *et al* 2010, 2011).

Testing potential 'founder effects'

Although recent work has shown that there is no clinal or regional variation in wing length among British Marsh Tits (Broughton *et al* 2016b), the possibility of highly localised 'founder effects' on biometrics has yet to be tested. It was suggested by du Feu & du Feu (2014) that such effects could result from individuals of atypical size founding small populations in isolated woods, thereby producing future generations of unusually small or large birds.

We used the Cambridgeshire Marsh Tit data to test this idea, using wing lengths of 56 sexed juveniles that were ringed as nestlings and captured the following autumn/winter, and also the wing lengths of both of their apparent parents (observed tending them in the nest). This assumed that the apparent parents were also the genetic parents, which seemed reasonable because extra-pair paternity is generally low in *Poecile* tits (Orell *et al* 1997, Mennill *et al* 2004).

We used a simple multiple linear regression, performed in R version 3.0.2 (R Core Team 2013) to test for a significant relationship between wing length of the juveniles of each sex and either or both of their male and female parents. This would indicate whether short- or long-winged parents were associated with similar wing lengths among their progeny. Because some parent birds were measured as juveniles and others as adults, however, their wing lengths were standardised according to the method of Lessells & Ovenden (1989), by subtracting from each measurement the mean wing length of the appropriate age and sex class, and then dividing the result by the standard deviation of that age and sex class. The mean and standard deviation values for each age and sex class were taken from Broughton *et al*. (2016a), and were largely derived from the same population at Monks Wood.

Human error during data collection

Marsh Tit ringing data extracted from the BTO database was used to examine the consistency of ringers in recording wing length. This dataset contained 22,302 wing length measurements from 1963 to 2014, including 4,999 pairs of consecutive measurements by 720 ringers who were identified by permit number and an individual's initials recorded in the data submission. These consecutive measurements involved 2,693 individual birds where both measurements fell within a two-month period, which was to avoid undue differences due to feather abrasion which are likely to occur over a longer period (Flinks & Salewski 2012). We used these records to test the consistency of wing length recording by ringers, calculating the frequency and magnitude of differences between the initial and subsequent measurements. We also calculated the test-retest reliability by generating the Pearson correlation coefficient between consecutive measurement pairs for each bird, where a coefficient of 0 would indicate total inconsistency and a value of 1 would show perfect agreement and complete consistency of ringers' measurements.

Also within the BTO database were 839 records for 698 individual Marsh Tits which had been sexed using BP or CP. Of these, 101 birds had been captured and sexed more than once, allowing the number of discrepancies between the first and second examination to be calculated as an error rate. Any discrepancy is recognised as being a minimum value and a likely under-estimate, however, as some ringers may have omitted or amended records where this conflicted with a previous entry, thereby reducing the detectable errors among repeat captures.

Local persistence of misclassified birds

If Marsh Tits that are classified as the wrong sex using the wing length method have a lower persistence in the local population than other birds, due to lower survival or higher emigration, this could bias any demographic analyses (du Feu & du Feu 2014). The Cambridgeshire data provided a sample with which to test this question, involving 11 'short-winged' males (wing lengths of 62 mm or less) and 21 'long-winged' females (63 mm or more), which would have been sexed incorrectly using the wing length division of 62/63 mm.

As a control, these atypical birds were paired in the analyses with 11 males and 21 females of the same age class that had 'typical' wing lengths (63 mm or more for males, 62 mm or less for females), selected as the closest contemporaries of the atypical birds. Each atypical bird and its matched control bird were originally caught within a mean of 15 days of each other, with a range of 0-173 days and 84% within the same month. The persistence of these birds in the local population could then be compared in a paired test (Wilcoxon signed rank), which controlled for the effect of age, season or annual differences in survival. Persistence in the local population was defined as the number of days over which an individual was detected in the study area from first to last observation, similar to the approach adopted by du Feu & du Feu (2014) but largely based on colour-ring re-sightings rather than recaptures during ringing activities.

RESULTS

Tests of potential 'founder effects'

Regression analyses did not find any evidence of a heritability of wing lengths among Marsh Tits in the Cambridgeshire dataset. For juvenile female wing lengths, there was no significant interaction ($F_{1,18}$ = 0.01, P = 0.932, n = 21) or individual effects of the standardised wing lengths of their mothers ($F_{1,20}$ = 4.07, P = 0.058) or fathers ($F_{1,20}$ = 1.27, P = 0.274). Similarly for juvenile male wing lengths, there was no significant interaction ($F_{1,32}$ = 0.19, P = 0.666, n = 35) or individual effects of their mothers' ($F_{1,34}$ = 1.30, P = 0.263) or fathers' ($F_{1,34}$ = 0.66, P = 0.424) standardised wing lengths. This showed that short-winged juvenile males and long-winged juvenile females did not generally have parents with similarly extreme wing lengths, which undermined the potential mechanism for localised 'founder effects'.

Human error during data collection

Within the BTO database of ringing records for British Marsh Tits, 43.1% of the 4,999 consecutive wing length measurements differed from the previous value for the same bird. Of these discrepancies, only 16.2% differed by more than 1 mm above or below the previous measurement, meaning that only 7.2% of all repeated measurements differed from the original by more than 1 mm (Fig. 1). The test-retest reliability coefficient (Pearson correlation) was 0.918, indicating a very high degree of overall consistency among ringers, in that repeated measurements tended to be the same or very close to the original value.

The extremes of differences in consecutive wing length measurements varied from 10 mm below the original value to 11 mm above it, suggesting some errors resulting from incorrect reading or recording of the measurement rather than an issue with measurement technique, i.e. 'observer distraction' as reported by Morgan (1994) and Zuur *et al*. (2009). Where the same ringer had taken consecutive measurements then the proportion of discrepancies was 34.5%, which was significantly lower than the 52.5% rate of discrepancy when measurements were taken by a different ringer (count data: Chi-square = 159.2, df = 1, $P < 0.001$).

Of the 101 records of repeated sexing of Marsh Tits using BP or CP, for 6% (six records) the sex assigned to the bird had changed from the original assessment, giving an agreement of 94% using this method. Discrepancies occurred between as well as within seasons, but with no systematic pattern of e.g. spring-caught females being recorded as 'males' later in the summer. Each of the discrepancies involved a different ringer, with half disagreeing with their own initial assessment and half disagreeing with another ringer. Of all 839 records in the BTO database where sex was assigned using BP/CP, 14% fell outside of the breeding period for Marsh Tits (April-June; Broughton & Hinsley 2015) and so were unlikely to be valid. Only one of the six discrepancies fell within the non-breeding period, however, so this cannot account for the errors.

Local persistence of misclassified birds

In paired tests of the duration of observed life history for the Marsh Tits with atypical wing lengths, which would have been sexed incorrectly using the 62/63 mm division, no significant difference was found between short-winged males compared to males with other wing lengths (Wilcoxon statistic = 28.0, *P* = 0.689, n = 11), and no difference was found between long-winged females and other females in the population (Wilcoxon statistic = 142.0, *P* = 0.366, n = 21). This indicated that short-winged males and long-winged females persisted in the local population for a similar period of time as did other birds.

DISCUSSION

Potential for biased analyses

The results of this study, and other recent work (e.g. Broughton *et al* 2016a, 2016b), demonstrate that sexing of Marsh Tits using accurate wing length measurements has a strong biological basis, and this is likely to apply to other species where the sexes share only a narrow overlap in biometrics. Male and female Marsh Tits show a strongly bimodal distribution in wing length measurements that allows most individuals to be sexed with a high degree of reliability (Broughton *et al* 2016a). There is no regional or clinal variation in wing length measurements among British Marsh Tits (Broughton *et al* 2016b) and the current study found no evidence for a mechanism of highly localised 'founder effects' that could produce small populations of birds of atypical size.

The dispersal and settling ecology of Marsh Tits probably make localised founder effects quite unlikely in this species; juveniles invariably disperse out of their natal territories, and in small woods that support only a few pairs these young birds generally leave the wood completely, with any settlers tending to be immigrants that were hatched elsewhere (Broughton *et al* 2010, Wesołowski 2015). In the 13 ha Odd Quarter and 28 ha Upton Woods in Cambridgeshire, for example, which typically contained two and three breeding territories respectively, none of 88 colour-ringed nestlings subsequently bred in their natal wood, with all of the breeding recoveries (eight records) coming from other woods (pers. obs.). Such movements are virtually impossible to detect without ringing nestlings, as most dispersal occurs as soon as juveniles become independent within a few weeks of fledging (Broughton *et al* 2010).

Our tests also found no difference in local persistence (i.e. survival or emigration) between short-winged males, long-winged females and other birds in the population, as these atypical birds were recorded over similar periods of time as their contemporaries. This suggested that the small number of birds that would be sexed incorrectly on wing length would not introduce any bias into analyses of survival, as they were no more or less likely to disappear than other birds. This differs from the result reported by du Feu & du Feu (2014) for three small-winged males that were compared with 13 typical birds, but this may reflect the larger sample in the current study (32 'pairs' of birds) that controlled for any bias of age or year.

Data quality in ringing schemes

Although the Marsh Tit is an example of a species that can be sexed on wing length with a high degree of reliability (Amann 1980, Nilsson 1992, King & Muddeman 1995, du Feu & du

Feu 2014, Broughton *et al* 2016a), this requires accurate biometric data. All of the detailed studies of Marsh Tits have involved one or a small number of experienced ringers working closely together, ensuring highly consistent data. In the pooled biometrics of the BTO database, however, variation in wing length measurements are, unsurprisingly, much greater, with almost half (43%) of the repeated measurements of the same bird differing from the initial value. This appears to be due to variation or human error in measuring, reading, recording, and possibly rounding of the value, and was remarkably similar to the error rate recorded by Gosler *et al* (1995) for wing length measurements. Although discrepancies in the Marsh Tit data were more likely when a different ringer had taken the repeat measurement, more than a third of wing lengths measured by the same ringer also differed from the previous value. As the consecutive measurements were taken within two months of each other, this was not due to either an increase in wing length after the first full wing moult or to a decrease resulting from wear.

Despite this high proportion of discrepancy in consecutive wing length measurements, there was a very high statistical correlation in the test-retest analysis, showing that the great majority of second measurements were the same as or very close to the previous one. Indeed, most variation fell within 1 mm of the original measurement, which is a commonly accepted degree of tolerance among ringers in the BTO scheme (pers. obs.). For small passerines, however, such as Marsh Tits, this 2 mm range of tolerance across three possible measurements (1 mm above or below the actual wing length) introduces some uncertainty into the data by changing the probability of its categorisation (i.e. an adult wing of 63 mm is 63% likely to be female while 64 mm is more 94% likely to be male, Broughton et al. 2016a). This variation in measurement accuracy is therefore likely to be the main reason for more than a quarter of the sexed Marsh Tits in the BTO database not conforming to the wing-length pattern expected from the controlled local studies (du Feu & du Feu 2014, Robinson 2015, Broughton *et al* 2016a), rather than biological variation amongst the birds themselves.

Unlike wing measurements, where variation and error are continuous, sexing according to BP/CP is binary and either wholly correct or not, and is sometimes treated as being a generally reliable method of assigning sex (e.g. Svensson 1992, du Feu & du Feu 2014), although caution is required in its use (Svensson 1992). Detectable sexing errors of Marsh Tits using BP/CP were low (6%) in the BTO database, but the recording software used by most British ringers since 1997 (IPMR: the Integrated Population Monitoring Reporter, © 1997-2011 Mark Cubitt) queries any discrepancy in assigning sex to birds already recorded by the user. Where ringers have amended or omitted conflicting records then this will have under-estimated the number of genuine errors and the error rate will also be unknown for birds that were caught and examined just once. The 14% of sexing records based on BP/CP that fell outside of the breeding season also indicates that this method is being misapplied to some extent, as BP/CP would be ambiguous or absent during this time.

This error in sexing using BP/CP was likely to be an additional factor in the unexpected variation in wing lengths of sexed Marsh Tits in the BTO database; for example, where accurate wing length measurements were taken from some birds that were sexed incorrectly using BP/CP. The combined effect of the incorrect sexing and measurement variation in the BTO database is that a substantial proportion of Marsh Tit records could not be used in some demographic analyses, such as age and sex composition (Broughton *et al* 2016a) or racial identification (Broughton *et al* 2016b), without filtering or some attempt at applying correction algorithms.

These results from Marsh Tits in the BTO database highlight the need for adequate training and guidance in the recording and measuring of BP/CP and biometrics, as any errors in technique will likely be replicated across other species. The importance of ensuring a high degree of data quality among submissions to national databases to maximise the scientific value of the information collected cannot be overestimated.

The value of sexing birds

In view of the issues of variable data quality, a central question is the value biometrics for sexing birds at all. Whilst acknowledging that knowing the sex of a bird is very useful in demographic studies, du Feu & du Feu (2014) suggested that intensive local population studies would eventually discover the sex of an individual via behavioural observations or examination of BP/CP in the breeding season. They further suggested that it is unhelpful to assign a sex in the field based on measurement, and this could even reduce the value of the record if the sexing method is not recorded, as an analyst may discard such data because of the uncertainty as to how it was determined. However, if the ringer recorded the appropriate measurement, an analyst would then be able to investigate bias and assign sex, in which case the recording of the bird's assigned sex by the ringer is redundant.

Nevertheless, relying on sexing only during the breeding season is not without its own pitfalls; while investigating these data we have found potential error involved with sexing by BP/CP, which at a minimum of 6% was comparable to the 2-8% error rate when sexing Marsh Tits by wing length. Importantly, by only sexing birds that survived to breed, this would introduce a substantial systematic bias into demographic studies by excluding all individuals in a population that died within their first year of life. In the case of Marsh Tits this includes the great majority of juveniles (Broughton *et al* 2010), and would prevent a large range of demographic analyses, such as the estimation of juvenile sex ratios and sex-related survival or dispersal.

Aside from DNA sexing and some species where non-breeding behaviour (e.g. territoriality) may allow sex to be inferred, biometrics is often the only field-based method for assigning sex to juveniles in species where plumage is similar between males and females. This is also the case for non-breeding birds in other species studies, such as passage or winter migrants (Scebba *et al* 2015, Bozó & Heim 2016). Although this will not work for all species (Ellrich *et al* 2010), the collection and analysis of biometric data has an essential role to play in demographic studies of survival, dispersal and migration.

Conclusions

The principle of assigning sex for some bird species using the length of the wing (or other biometrics, including combinations of several measurements) is well supported in the literature. For the Marsh Tit, highly reliable estimates can be made for the proportion of males and females within a sample population, and even the proportions of adults and juveniles, based solely on accurately measured wing lengths (Broughton *et al* 2016a). Even for individuals that are misclassified, the evidence suggests that this would not significantly bias any analyses of survival or emigration by ringer-analysts or future researchers.

The numerous detailed studies of Marsh Tits show that individual ringers and groups can be highly consistent and accurate in their recording of wing length where best practice is followed, including adequate training and regular calibration and cross-checking of measurement techniques. Although perfect accuracy and replication between all ringers is impossible to attain, widespread acceptance of a measurement 'tolerance' (e.g. 1 mm for small passerines) during training and data collection may have important consequences of reduced data quality for subsequent researchers.

Minimising or eliminating such a measurement tolerance wherever possible, through improved training in measurement technique and regular re-checking within and between ringers, would ensure the greatest possible accuracy, precision and value of the data collected. Ensuring routine measurement of recaptures of recently examined birds would be a simple way of providing ringers with continual feedback and 'self-assessment' of their data quality, by highlighting any variation in repeated measurements that could then be addressed with practice or training.

The examination of BP/CP records also suggest that assessment and recording of these data could be improved, most easily through better training and awareness; the error detected in this study is concerning due to BP/CP generally being considered as a primary method of sexing birds (Svensson 1992, du Feu & du Feu 2014).

The challenge for analysts is being able to identify the reliable biometric and sexing data and filter out the errors and low-precision records that inevitably appear in pooled datasets, to allow them to undertake robust analyses (Zuur *et al*. 2009). It is vital that ringers assigning sex in their ringing data record the method they used, as recommended by du Feu & du Feu (2014). To avoid further collection of assigned sex without a method being given, the new software now being introduced by BTO for use by British & Irish ringers (Demon), has a mandatory sexing method field if a sex is assigned to an individual. To increase the value of the data collected, we suggest that this approach should also be adopted by other ringing schemes if they have not already done so. Individual ringers can further assist researchers by ensuring that they record and computerise the identity ofying the person who measured each wing length.

A wider challenge for individual ringers, ringing groups and ringing schemes in general is how to continue to foster a high degree of precision, accuracy, consistency and overall reliability in the gathering of biometric data that is of sufficient value for analysts, researchers and the ringers themselves. Although ensuring high standards in data collection is largely an individual responsibility for those handling the birds, this must be supported by the ringing schemes that govern these activities through provision of adequate training, information and guidance. Such improvements can be gained through relatively simple measures, such as regular cross-checking of measurement technique among ringers within and between groups, regular calibration of equipment, and checking of data during and after data entry. More formal efforts to improve data quality could include the design and delivery of workshops, courses or training material, and promoting access to reference literature as well as the current BTO initiative of introducing mandatory recording of sexing method.

We also suggest that, where appropriate data sets exist, further investigations of the bias and error in assigning sex are carried out.

ACKNOWLEDGEMENTS

The authors thank landowners for permitting access to Cambridgeshire study woods, particularly Natural England at Monks Wood, where Paul Bellamy, Dr Shelley Hinsley and Dr Marta Maziarz helped with data collection, and Douglas Hall, Sarah Caesar and Geoff Leach provided valuable fieldwork support. Thanks are also extended to all contributors to the BTO ringing scheme, particularly the ringers themselves, and to Dr Dave Leech for valuable comments on an initial draft of the manuscript. The BTO Ringing Scheme is funded by a partnership of the British Trust for Ornithology, the Joint Nature Conservation Committee (on behalf of: Council for Nature Conservation and the Countryside, Natural England, Natural Resources Wales and Scottish Natural Heritage), The National Parks and Wildlife Service (Ireland) and the ringers themselves. RKB was funded by the Natural Environment Research Council (NERC).

REFERENCES

Amann, F. (1980) Alters- und Geschlechtsmerkmale der Nonnenmeise *Parus palustris*. *Der Ornithologische Beobachter* **77**, 79-83.

Amouret, J., Hallgrimsson, G.T. & Pálsson, S. (2015) Sexing of Redpolls *Acanthis flammea* in Iceland from morphometrics. *Ringing & Migration* **30**, 61-66.

Bozó, L. & Heim, W. (2016) Sex-specific migration of *Phylloscopus* warblers at a stopover site in Far Eastern Russia. *Ringing & Migration* **31**, 41-46.

Broughton, R. (2010) Marsh and Willow Tits in the hand. *Ringers' Bulletin* **12**, 106.

Broughton, R.K. & Hinsley, S.A. (2015) The ecology and conservation of the marsh tit in Britain. *British Birds* **108**, 12-28.

Broughton, R.K., Hinsley, S.A., Bellamy, P.E., Carpenter, J.E. & Rothery, P. (2008) Ageing and sexing Marsh Tits *Poecile palustris* using wing length and moult. *Ringing & Migration* **24**, 88-94.

Broughton, R.K., Hill, R.A., Bellamy, P.E. & Hinsley, S.A. (2010) Dispersal, ranging and settling behaviour of Marsh Tits *Poecile palustris* in a fragmented landscape in lowland England. *Bird Study* **57**, 458-472.

Broughton, R.K., Hill, R.A., Bellamy, P.E. & Hinsley, S.A. (2011) Nest-sites, breeding failure and causes of non-breeding in a population of British Marsh Tits *Poecile palustris*. *Bird Study* **58**, 229-237.

Broughton, R.K., Bellamy, P.E., Hill, R.A. & Hinsley, S.A. (2014) Winter habitat selection by Marsh Tits *Poecile palustris* in a British woodland. *Bird Study* **61**, 404-412.

Broughton, R.K., Bellamy, P.E., Hill, R.A. & Hinsley, S.A. (2015) Winter social organisation of Marsh Tits *Poecile palustris* in Britain. *Acta Ornithologica* **50**, 11-21.

Broughton, R.K., Alker, P.J., Bellamy, P.E., Britton, S., Dadam, D., Day, J.C., Miles, M. & Hinsley, S.A. (2016a) Comparative biometrics of British Marsh Tits *Poecile palustris* and Willow Tits *P. montana*. *Ringing & Migration* **31**, 30-40.

Broughton, R.K., Burgess, M.D., Dadam, D., Hebda, G., Bellamy, P.E. & Hinsley, S.A. (2016b) Morphology, geographical variation and the subspecies of Marsh Tit *Poecile palustris* in Britain and central Europe. *Bird Study* **63**, 58-65.

Craik, J.C.A. (1999) Sexual dimorphism of Common Terns *Sterna hirundo* and Arctic Terns *S. paradisaea*. *Ringing & Migration* **19**, 311-312.

du Feu, C.R., Clark, J.A., Fiedler, W., Baillie, S.R. & Laesser, J. (2015) The EURING Exchange Code 2000 Plus. EURING, Thetford. ISBN 978-1-908581-51-8.

du Feu, C.R. & du Feu, R. (2014) No sex please, we're biased: some comments on sexing Marsh Tits *Poecile palustris* by wing length. *Ringing & Migration* **29**, 47-50.

Ellrich, H., Salewski, V. & Fiedler, W. (2010) Morphological sexing of passerines: not valid over wider geographical scales. *Journal of Ornithology* **151**, 449-458.

Fletcher, K.L. & Foster, R. (2010) Use of external biometrics to sex Carrion Crow *Corvus corone*, Rook *C. frugilegus* and Western Jackdaw *C. monedula* in northern England. *Ringing & Migration* **25**, 47-51.

Flinks, H. & Salewski, V. (2012) Quantifying the effect of feather abrasion on wing and tail lengths measurements. *Journal of Ornithology* **153**, 1053-1065.

Gosler A.G. & King, J.R. (1989) A sexually dimorphic plumage character in the Coal Tit *Parus ater* with notes on the Marsh Tit *Parus palustris*. *Ringing & Migration* **10**, 53-57.

Gosler A.G., Greenwood J.J.D., Baker J.K. & King, J.R. (1995) A comparison of wing length and primary length as size measures for small passerines: a report to the British Ringing Committee. *Ringing & Migration* **16**, 65–78.

Hallgrimsson, G.T., Helgason, H.H., Palsdottir, E.S & Palsson, S. (2016) Sexing adult and fledgling Lesser Black-backed Gulls from morphometrics. *Ringing & Migration* **31**, 68-73.

Haftorn, S. (1982). Variation in body measurements of the Willow Tit *Parus montanus*, together with a method for sexing live birds and data on the degree of shrinkage in size after skinning. *Fauna Norvegica Series C Cinclus* **5**, 16-26.

Hewson, C.M., Thorup, K., Pearce-Higgins, J.W. & Atkinson, P.W. (2016) Population decline is linked to migration route in the Common Cuckoo. *Nature Communications* 7:12296 doi: 10.1038/ncomms12296

Holt, C.A., Fraser, K.H., Bull, A.J. & Dolman, P.M. (2012) Habitat use by Nightingales in a scrub-woodland mosaic in central England. *Bird Study* **59**, 416-425.

Katrínardóttir, B., Pálsson, S., Gunnarsson, T.G. & Sigurjónsdóttir, H. (2013) Sexing Icelandic Whimbrels *Numenius phaeopus islandicus* with DNA and biometrics. *Ringing & Migration* **28**, 43-46.

King, J.R. & Muddeman, J.L. (1995) Ageing and sexing Marsh Tits *Parus palustris*. *Ringing & Migration* **16**, 172-177.

Lessells, C.M. & Ovenden, G.N. (1989) Heritability of wing length and weight in European Bee-eaters (*Merops apriaster*). *The Condor* **91**, 210-214.

Madsen, V. (1997) Sex-determination of continental European Robins *Erithacus r. rubecula*. *Bird Study* **44**, 239–244.

Mennill, D.J., Ramsay, S.M., Boag, P.T & Ratcliffe, L.M. (2004) Patterns of extrapair mating in relation to male dominance status and female nest placement in black-capped chickadees. *Behavioral Ecology* **15**, 757-765.

Morgan, J.H. (2004) Remarks on the taking and recording of biometric measurements in bird ringing. *The Ring* **26**, 71-78.

Nilsson, J-Å. 1992. Variation in wing length in relation to sex and age in Marsh Tits *Parus palustris*. *Ornis Svecica* **2**, 7-12.

Ormerod, S.J., Tyler, S.J. & Lewis, J.M.S. (1986) Biometrics, growth and sex ratios amongst Welsh Dippers *Cinclus cinclus*. *Ringing & Migration* **7**, 61-70.

Orell, M., Rytkönen, S., Launonen, V., Welling, P., Kumpulainen, K. & Bachmann, L. (1997) Low frequency extra-pair paternity in the Willow Tit *Parus montanus* as revealed by DNA fingerprinting. *Ibis* **139**, 562–566.

Prytherch, R.J & Roberts, L.F. (2015) Sexing of nestling Common Buzzards *Buteo buteo* in Britain. *Ringing & Migration* **30**, 57-59.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Redfern, C.P.F. & Clark, J.A. (2001) *Ringers' Manual*. BTO, Thetford.

Richner, H. (1989) Avian Laparoscopy as a field technique for sexing birds and an assessment of its effects on wild birds. *Journal of Field Ornithology* **60**, 137-142.

Robinson, R.A. (2015) BirdFacts: profiles of birds occurring in Britain & Ireland (BTO Research Report 407). BTO, Thetford (http://www.bto.org/birdfacts)

Scebba, S., Soprano, M. & Sorrenti, M. (2015) Sex-specific migration patterns and population trends of Skylarks *Alauda arvensis* ringed during autumn migration in southern Italy. *Ringing & Migration* **30**, 12-21.

Svensson, L. (1992) *Identification Guide to European Passerines* (4th edition). Lars Svensson, Stockholm.

Wesołowski, T. (2015) Dispersal in an extensive continuous forest habitat: Marsh Tit *Poecile palustris* in the Białowieża National Park. *Journal of Ornithology* **156**, 349-361.

Zuur, A.F., Ieno, E.N. & Elphick, C.S. (2009) A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* **1**, 3–14.

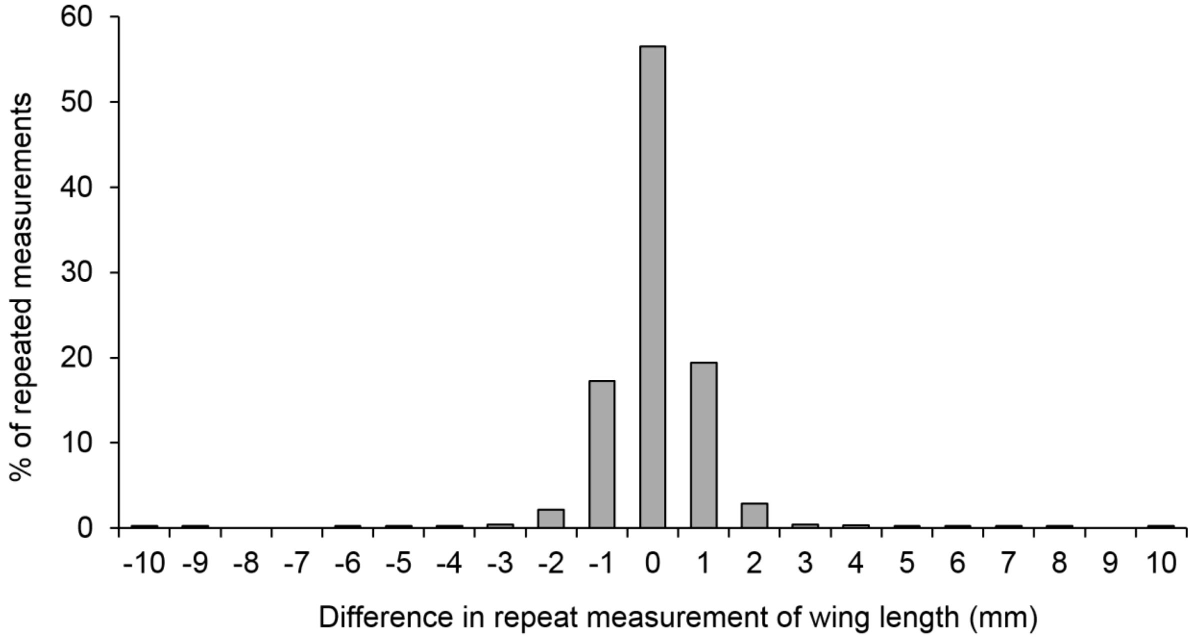

Figure 1. Distribution of 4,999 repeat measurements of Marsh Tit wing lengths, taken within two months of the initial measurement. Just over half (57%) of repeat measurements were the same as the initial value (difference = 0), but similar proportions of measurements were longer (> 0) or shorter (< 0) than the initial wing length.