



## 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](mailto: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.

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

2

3 Richard K Broughton<sup>a\*</sup> and Jacquie A Clark<sup>b</sup>

4 <sup>a</sup>Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK

5 <sup>b</sup>BTO, The Nunnery, Thetford, Norfolk, IP24 2PU, UK

6

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

8 Short title: Bias and error in sexing

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

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29 ABSTRACT

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

46

47

48

49

50

51

52

53

54

55

56

57 INTRODUCTION

58

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

65 A range of methods are available for sexing birds with similar plumage, but their  
 66 usefulness varies with the circumstances. Molecular sexing using DNA is reliable but usually  
 67 requires laboratory processing of samples. Laparoscopy can be used on larger species in  
 68 the field (Richner 1989) but is invasive and requires specialist skills. Sexing by autopsy has  
 69 an obvious limitation for most population studies, in that the bird must be dead, and it can  
 70 also give incorrect results (Haftorn 1982). Many monomorphic species can be sexed using  
 71 breeding or territorial behaviour, but individuals must be marked, e.g. with colour-rings, and  
 72 later observed in the field (Broughton *et al* 2010). The presence of a brood patch (BP) or  
 73 shape of the cloacal protuberance (CP) is another common field method, but is limited to the  
 74 breeding season and can be ambiguous (Redfern & Clark 2001). Finally, biometrics,  
 75 including wing length, have been used to sex a wide variety of species based on  
 76 measurement thresholds or mathematical estimates, including raptors (Prytherch & Roberts  
 77 2015), waders (Katrínardóttir *et al* 2013), seabirds (Craik 1999, Hallgrímsson *et al* 2016) and  
 78 various passerines (Ormerod *et al* 1986, Madsen 1997, Fletcher & Foster 2010, Amouret *et*  
 79 *al* 2015).

80 A limitation of sexing birds based on the length of the wing alone is that almost all  
 81 studies report some measurement overlap between males and females. For species where  
 82 males are generally larger, this means that some small males and large females will be  
 83 misclassified (Haftorn 1982, du Feu & du Feu 2014). Another problem is that measurement  
 84 criteria reported from one location might not be applicable for the same species elsewhere,

85 because of size variation related to subspecies, migratory behaviour, or measurement  
 86 technique amongst ringers (Ormerod *et al* 1986, Gosler *et al* 1995, Morgan 2004, Ellrich *et*  
 87 *al* 2010, Broughton *et al* 2016a). Nevertheless, some species, including small passerines,  
 88 can have consistent wing lengths across large areas. For example, in the strictly sedentary  
 89 Willow Tit *Poecile montanus* there was no difference in wing length over an 800 km range in  
 90 Norway (Haftorn 1982), and similar results were found for the Marsh Tit *P. palustris* over 540  
 91 km in Britain (Broughton *et al* 2016b). These results show that, for some species at least, the  
 92 same measurement threshold could be used to assign sex across large geographical areas  
 93 (though see Ellrich *et al* 2010).

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

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

106 Despite the attraction of using biometrics to sex species such as Marsh Tit, however,  
 107 important questions have been raised regarding the limitations of the wing length biometric  
 108 for sexing birds in general. The Marsh Tit was used as a case study by du Feu & du Feu  
 109 (2014) to propose that significant error in sexing could occur due to measurement overlap,  
 110 inaccurate measurement by ringers, and local variation in the size of birds due to 'founder  
 111 effects' or clinal variation. Such error could result in serious bias entering analyses of

112 biometric data if the life histories of birds that were misclassified differed from those which  
 113 were sexed correctly.

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

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

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

139

140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167

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).

168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195

**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**



196 Marsh Tit ringing data extracted from the BTO database was used to examine the  
197 consistency of ringers in recording wing length. This dataset contained 22,302 wing length  
198 measurements from 1963 to 2014, including 4,999 pairs of consecutive measurements by  
199 720 ringers who were identified by permit number and an individual's initials recorded in the  
200 data submission. These consecutive measurements involved 2,693 individual birds where  
201 both measurements fell within a two-month period, which was to avoid undue differences  
202 due to feather abrasion which are likely to occur over a longer period (Flinks & Salewski  
203 2012). We used these records to test the consistency of wing length recording by ringers,  
204 calculating the frequency and magnitude of differences between the initial and subsequent  
205 measurements. We also calculated the test-retest reliability by generating the Pearson  
206 correlation coefficient between consecutive measurement pairs for each bird, where a  
207 coefficient of 0 would indicate total inconsistency and a value of 1 would show perfect  
208 agreement and complete consistency of ringers' measurements.

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

216

217

### 218 **Local persistence of misclassified birds**

219

220 If Marsh Tits that are classified as the wrong sex using the wing length method have a lower  
221 persistence in the local population than other birds, due to lower survival or higher  
222 emigration, this could bias any demographic analyses (du Feu & du Feu 2014). The  
223 Cambridgeshire data provided a sample with which to test this question, involving 11 'short-

224 winged' males (wing lengths of 62 mm or less) and 21 'long-winged' females (63 mm or  
 225 more), which would have been sexed incorrectly using the wing length division of 62/63 mm.

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

237

238

239 **RESULTS**

240

241

242 **Tests of potential 'founder effects'**

243

244 Regression analyses did not find any evidence of a heritability of wing lengths among Marsh  
 245 Tits in the Cambridgeshire dataset. For juvenile female wing lengths, there was no  
 246 significant interaction ( $F_{1,18} = 0.01$ ,  $P = 0.932$ ,  $n = 21$ ) or individual effects of the  
 247 standardised wing lengths of their mothers ( $F_{1,20} = 4.07$ ,  $P = 0.058$ ) or fathers ( $F_{1,20} = 1.27$ ,  $P$   
 248  $= 0.274$ ). Similarly for juvenile male wing lengths, there was no significant interaction ( $F_{1,32} =$   
 249  $0.19$ ,  $P = 0.666$ ,  $n = 35$ ) or individual effects of their mothers' ( $F_{1,34} = 1.30$ ,  $P = 0.263$ ) or  
 250 fathers' ( $F_{1,34} = 0.66$ ,  $P = 0.424$ ) standardised wing lengths. This showed that short-winged  
 251 juvenile males and long-winged juvenile females did not generally have parents with similarly

252 extreme wing lengths, which undermined the potential mechanism for localised ‘founder  
253 effects’.

254

255

256 **Human error during data collection**

257

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

265 The extremes of differences in consecutive wing length measurements varied from  
266 10 mm below the original value to 11 mm above it, suggesting some errors resulting from  
267 incorrect reading or recording of the measurement rather than an issue with measurement  
268 technique, i.e. ‘observer distraction’ as reported by Morgan (1994) and Zuur *et al.* (2009).  
269 Where the same ringer had taken consecutive measurements then the proportion of  
270 discrepancies was 34.5%, which was significantly lower than the 52.5% rate of discrepancy  
271 when measurements were taken by a different ringer (count data: Chi-square = 159.2, df =  
272 1,  $P < 0.001$ ).

273 Of the 101 records of repeated sexing of Marsh Tits using BP or CP, for 6% (six  
274 records) the sex assigned to the bird had changed from the original assessment, giving an  
275 agreement of 94% using this method. Discrepancies occurred between as well as within  
276 seasons, but with no systematic pattern of e.g. spring-caught females being recorded as  
277 ‘males’ later in the summer. Each of the discrepancies involved a different ringer, with half  
278 disagreeing with their own initial assessment and half disagreeing with another ringer. Of all  
279 839 records in the BTO database where sex was assigned using BP/CP, 14% fell outside of

280 the breeding period for Marsh Tits (April-June; Broughton & Hinsley 2015) and so were  
 281 unlikely to be valid. Only one of the six discrepancies fell within the non-breeding period,  
 282 however, so this cannot account for the errors.

283

284

285 **Local persistence of misclassified birds**

286

287 In paired tests of the duration of observed life history for the Marsh Tits with atypical wing  
 288 lengths, which would have been sexed incorrectly using the 62/63 mm division, no significant  
 289 difference was found between short-winged males compared to males with other wing  
 290 lengths (Wilcoxon statistic = 28.0,  $P = 0.689$ ,  $n = 11$ ), and no difference was found between  
 291 long-winged females and other females in the population (Wilcoxon statistic = 142.0,  $P =$   
 292  $0.366$ ,  $n = 21$ ). This indicated that short-winged males and long-winged females persisted in  
 293 the local population for a similar period of time as did other birds.

294

295

296 **DISCUSSION**

297

298

299 **Potential for biased analyses**

300

301 The results of this study, and other recent work (e.g. Broughton *et al* 2016a, 2016b),  
 302 demonstrate that sexing of Marsh Tits using accurate wing length measurements has a  
 303 strong biological basis, and this is likely to apply to other species where the sexes share only  
 304 a narrow overlap in biometrics. Male and female Marsh Tits show a strongly bimodal  
 305 distribution in wing length measurements that allows most individuals to be sexed with a high  
 306 degree of reliability (Broughton *et al* 2016a). There is no regional or clinal variation in wing  
 307 length measurements among British Marsh Tits (Broughton *et al* 2016b) and the current

308 study found no evidence for a mechanism of highly localised ‘founder effects’ that could  
 309 produce small populations of birds of atypical size.

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

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

330

331

332 **Data quality in ringing schemes**

333

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

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

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

363 Unlike wing measurements, where variation and error are continuous, sexing  
 364 according to BP/CP is binary and either wholly correct or not, and is sometimes treated as  
 365 being a generally reliable method of assigning sex (e.g. Svensson 1992, du Feu & du Feu  
 366 2014), although caution is required in its use (Svensson 1992). Detectable sexing errors of  
 367 Marsh Tits using BP/CP were low (6%) in the BTO database, but the recording software  
 368 used by most British ringers since 1997 (IPMR: the Integrated Population Monitoring  
 369 Reporter, © 1997-2011 Mark Cubitt) queries any discrepancy in assigning sex to birds  
 370 already recorded by the user. Where ringers have amended or omitted conflicting records  
 371 then this will have under-estimated the number of genuine errors and the error rate will also  
 372 be unknown for birds that were caught and examined just once. The 14% of sexing records  
 373 based on BP/CP that fell outside of the breeding season also indicates that this method is  
 374 being misapplied to some extent, as BP/CP would be ambiguous or absent during this time.

375 This error in sexing using BP/CP was likely to be an additional factor in the  
 376 unexpected variation in wing lengths of sexed Marsh Tits in the BTO database; for example,  
 377 where accurate wing length measurements were taken from some birds that were sexed  
 378 incorrectly using BP/CP. The combined effect of the incorrect sexing and measurement  
 379 variation in the BTO database is that a substantial proportion of Marsh Tit records could not  
 380 be used in some demographic analyses, such as age and sex composition (Broughton *et al*  
 381 2016a) or racial identification (Broughton *et al* 2016b), without filtering or some attempt at  
 382 applying correction algorithms.

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

388

389

390 **The value of sexing birds**

391

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

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

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

418



419

420 **Conclusions**

421

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

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

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

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

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

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

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

472

473

474 ACKNOWLEDGEMENTS

475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502

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.

- 503 **Broughton, R.K. & Hinsley, S.A.** (2015) The ecology and conservation of the marsh tit in  
504 Britain. *British Birds* **108**, 12-28.
- 505
- 506 **Broughton, R.K., Hinsley, S.A., Bellamy, P.E., Carpenter, J.E. & Rothery, P.** (2008)  
507 Ageing and sexing Marsh Tits *Poecile palustris* using wing length and moult. *Ringling &*  
508 *Migration* **24**, 88-94.
- 509
- 510 **Broughton, R.K., Hill, R.A., Bellamy, P.E. & Hinsley, S.A.** (2010) Dispersal, ranging and  
511 settling behaviour of Marsh Tits *Poecile palustris* in a fragmented landscape in lowland  
512 England. *Bird Study* **57**, 458-472.
- 513
- 514 **Broughton, R.K., Hill, R.A., Bellamy, P.E. & Hinsley, S.A.** (2011) Nest-sites, breeding  
515 failure and causes of non-breeding in a population of British Marsh Tits *Poecile palustris*.  
516 *Bird Study* **58**, 229-237.
- 517
- 518 **Broughton, R.K., Bellamy, P.E., Hill, R.A. & Hinsley, S.A.** (2014) Winter habitat selection  
519 by Marsh Tits *Poecile palustris* in a British woodland. *Bird Study* **61**, 404-412.
- 520
- 521 **Broughton, R.K., Bellamy, P.E., Hill, R.A. & Hinsley, S.A.** (2015) Winter social  
522 organisation of Marsh Tits *Poecile palustris* in Britain. *Acta Ornithologica* **50**, 11-21.
- 523
- 524 **Broughton, R.K., Alker, P.J., Bellamy, P.E., Britton, S., Dadam, D., Day, J.C., Miles, M.**  
525 **& Hinsley, S.A.** (2016a) Comparative biometrics of British Marsh Tits *Poecile palustris* and  
526 Willow Tits *P. montana*. *Ringling & Migration* **31**, 30-40.
- 527
- 528 **Broughton, R.K., Burgess, M.D., Dadam, D., Hebda, G., Bellamy, P.E. & Hinsley, S.A.**  
529 (2016b) Morphology, geographical variation and the subspecies of Marsh Tit *Poecile*  
530 *palustris* in Britain and central Europe. *Bird Study* **63**, 58-65.

531

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

534

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

537

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

540

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

543

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

547

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

550

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

553

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

557

- 558 **Hallgrímsson, G.T., Helgason, H.H., Pálsdóttir, E.S & Pálsson, S.** (2016) Sexing adult  
559 and fledgling Lesser Black-backed Gulls from morphometrics. *Ringling & Migration* **31**, 68-73.  
560
- 561 **Haftorn, S.** (1982). Variation in body measurements of the Willow Tit *Parus montanus*,  
562 together with a method for sexing live birds and data on the degree of shrinkage in size after  
563 skinning. *Fauna Norvegica Series C Cinclus* **5**, 16-26.  
564
- 565 **Hewson, C.M., Thorup, K., Pearce-Higgins, J.W. & Atkinson, P.W.** (2016) Population  
566 decline is linked to migration route in the Common Cuckoo. *Nature Communications* **7**:12296  
567 doi: 10.1038/ncomms12296  
568
- 569 **Holt, C.A., Fraser, K.H., Bull, A.J. & Dolman, P.M.** (2012) Habitat use by Nightingales in a  
570 scrub-woodland mosaic in central England. *Bird Study* **59**, 416-425.  
571
- 572 **Katrínardóttir, B., Pálsson, S., Gunnarsson, T.G. & Sigurjónsdóttir, H.** (2013)  
573 Sexing Icelandic Whimbrels *Numenius phaeopus islandicus* with DNA and biometrics.  
574 *Ringling & Migration* **28**, 43-46.  
575
- 576 **King, J.R. & Muddeman, J.L.** (1995) Ageing and sexing Marsh Tits *Parus palustris*. *Ringling*  
577 *& Migration* **16**, 172-177.  
578
- 579 **Lessells, C.M. & Ovenden, G.N.** (1989) Heritability of wing length and weight in European  
580 Bee-eaters (*Merops apriaster*). *The Condor* **91**, 210-214.  
581
- 582 **Madsen, V.** (1997) Sex-determination of continental European Robins *Erithacus r. rubecula*.  
583 *Bird Study* **44**, 239–244.  
584

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

588

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

591

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

594

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

597

598 **Orell, M., Rytkönen, S., Launonen, V., Welling, P., Kumpulainen, K. & Bachmann, L.**  
599 (1997) Low frequency extra-pair paternity in the Willow Tit *Parus montanus* as revealed by  
600 DNA fingerprinting. *Ibis* **139**, 562–566.

601

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

604

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

607

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

609

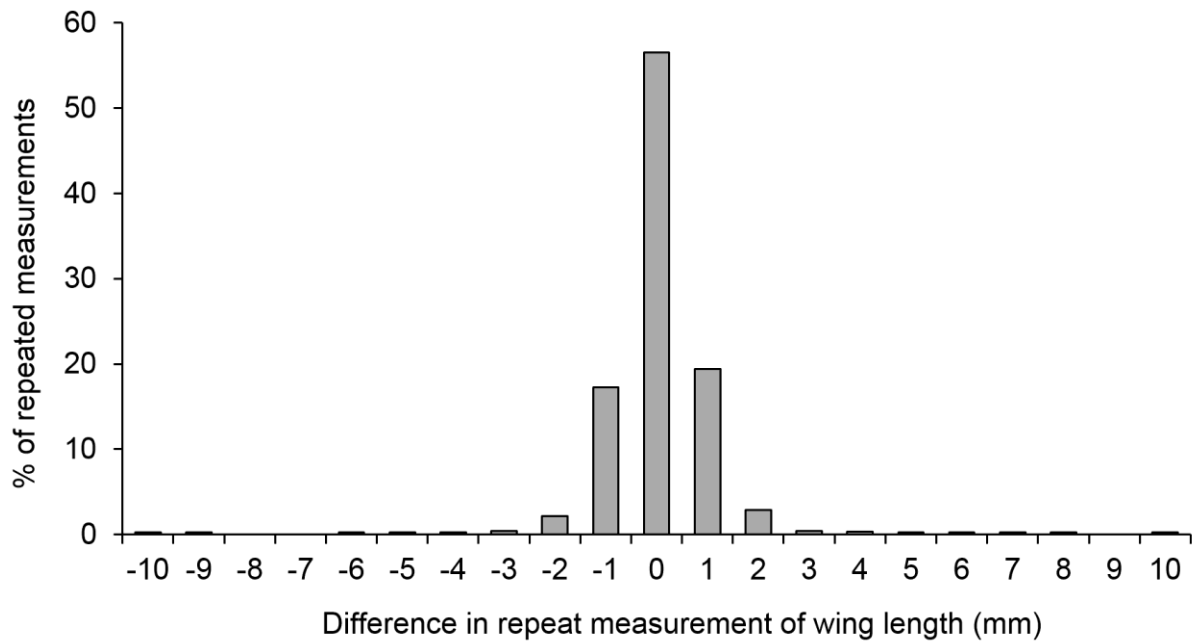
610 **Richner, H.** (1989) Avian Laparoscopy as a field technique for sexing birds and an  
611 assessment of its effects on wild birds. *Journal of Field Ornithology* **60**, 137-142.

612

- 613 **Robinson, R.A.** (2015) BirdFacts: profiles of birds occurring in Britain & Ireland (BTO  
614 Research Report 407). BTO, Thetford (<http://www.bto.org/birdfacts>)  
615
- 616 **Scebba, S., Soprano, M. & Sorrenti, M.** (2015) Sex-specific migration patterns and  
617 population trends of Skylarks *Alauda arvensis* ringed during autumn migration in southern  
618 Italy. *Ringing & Migration* **30**, 12-21.  
619
- 620 **Svensson, L.** (1992) *Identification Guide to European Passerines* (4<sup>th</sup> edition). Lars  
621 Svensson, Stockholm.  
622
- 623 **Wesołowski, T.** (2015) Dispersal in an extensive continuous forest habitat: Marsh Tit  
624 *Poecile palustris* in the Białowieża National Park. *Journal of Ornithology* **156**, 349-361.  
625
- 626 **Zuur, A.F., Ieno, E.N. & Elphick, C.S.** (2009) A protocol for data exploration to avoid  
627 common statistical problems. *Methods in Ecology and Evolution* **1**, 3–14.



628



629

630

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