RANDOM MEASUREMENT ERROR AND SPECIMEN SHRINKAGE IN SHORT-TAILED SHEARWATERS *PUFFINUS TENUIROSTRIS*

STEPHEN L. TOTTERMAN

179 Reedy Creek Road, Empire Vale, NSW 2478, Australia (stephen@totterman.net.au)

Received 21 July 2015, accepted 27 October 2015

SUMMARY

TOTTERMAN, S.L. 2016. Random measurement error and specimen shrinkage in Short-tailed Shearwaters *Puffinus tenuirostris*. *Marine Ornithology* 44: 11–20.

External biometrics have many applications in ornithology, and study skins are a major source of these measurements. However, measurements can be imprecise, and skins tend to shrink when they dry — two problems rarely investigated for petrels (family: Procellariidae). This study examined measurement error and shrinkage for 15 biometrics, using Short-tailed Shearwater *Puffinus tenuirostris* as the subject species. Random measurement error, defined as the variability of repeated measurements of a particular character taken on the same individual relative to its variability among individuals in a particular group, ranged from 0.3% for head plus bill length in dry specimens to 36% for tarsus width in freshly dead birds. Shrinkage of skin specimens stabilised within 2–5 months after preparation. Average fresh-dry shrinkage ranged from 0.2% for head plus bill to 12% for tarsus height. A new method was used to estimate shrinkage variability among individuals. "Shrinkage variation," defined as the proportion of unexplained variance ($1 - r^2$) in the correlation between paired fresh and dry measurements of a particular character after correcting for measurement error, ranged from 0% for wing chord to 33% for bill base width. More robust biometrics from this study were measurements of large, inflexible characters with well-defined measurement "landmarks."

Key words: biometrics, morphometrics, petrel, shearwater, measurement error, precision, shrinkage

INTRODUCTION

The popularity of external biometrics and appearance in bird studies is partly due to the availability of round skin specimens in natural history collections (Jenkinson & Wood 1985, Watson 2005). For petrels, breeding colonies may be unknown or difficult to access, and skin specimens may be the only resource for biometric data (e.g. Imber & Tennyson 2001). Researchers commonly perform such measurements on live birds as well.

External biometrics have many applications for petrel research, including studying ecological adaptation (e.g. Spear & Ainley 1998) and geographical variation (e.g. Ainley 1980), developing taxonomic hypotheses (e.g. Imber & Tennyson 2001), biometric sexing (e.g. Carey 2011), measuring body condition (e.g. Meathrel *et al.* 1993) and measuring growth (e.g. Pettit *et al.* 1984). However, measurements can be imprecise, and skins tend to shrink when they dry. With modern statistical methods being used to interrogate biometrics and detect fine differences in size and shape (e.g. Bretagnolle & Shirihai 2010), greater awareness of these two problems is needed.

Statistical analyses of biometrics commonly assume that the observed total variance among individuals reflects underlying phenotypic variance. However, measurements are affected by instrument precision, measuring conditions, flexibility of characters and human inconsistency (Yezerinac *et al.* 1992). Measurements tend to fluctuate randomly around the true values, such that some measured values will be higher than the true values, some will be lower, and the mean of these fluctuations will be zero. Thus, some proportion of the observed total variance is due to this random measurement error (ME):

$$ME = \frac{s_W^2}{s_W^2 + s_A^2}$$

where s_W^2 is the variance of repeated measurements on the same individual and s_A^2 is the variance among individuals (Bailey & Byrnes 1990). The denominator ($s_W^2 + s_A^2$) is the total variance. It is important to recognise that s_A^2 is a property of the particular group being examined, and measurement error will increase when s_A^2 is low (Yezerinac *et al.* 1992). Measurement errors in this study were relative to variance in size within a sample of female and male Short-tailed Shearwaters *Puffinus tenuirostris* (i.e. sexual size dimorphism contributed to s_A^2). Other studies might be concerned with geographical variation, where size differences between localities and within each sex can be smaller than differences between sexes (e.g. Einoder *et al.* 2008), or with variation between different species, where size variance can be large.

Random measurement error increases total variance, reduces statistical power and dilutes trends and patterns in the data (Bailey & Byrnes 1990, Hutcheon *et al.* 2010). Measurement error tends to increase for characters that are small, flexible and lack well-defined measurement landmarks, and tends to decrease with observer experience (Yezerinac *et al.* 1992). For birds, Lougheed *et al.* (1991) noted that measurement error was low for most biometrics (e.g. 2% for wing chord in freshly dead American Coots *Fulica americana*), but that there were always some biometrics for which measurement errors were high in their datasets (e.g. 16% for bill width in this species). A few petrel studies have taken repeated measures of external biometrics; however, two of those did not report measurement error (Granadeiro 1993, Genovart *et al.* 2003). Bourgeois *et al.* (2007) reported a maximum measurement error of 21% without specifying which of the 11 biometrics this result referred to.

Shrinkage of bird skins during the process of drying is well known and results in systematic errors when measurements of skins are compared with measurements of live or freshly dead birds. Only wing length shrinkage has been widely investigated (summarised in Winker 1993) and differences found have generally been small, e.g. a 1% decrease in mean wing chord for Atlantic Puffins *Fratercula arctica* (Harris 1980). Kinsky & Harper (1968) reported much larger 6%–12% mean shrinkage for bill width in three prion species (*Pachyptila* spp.). Differences in morphology between taxa affect shrinkage; as a result, estimates from one taxon may have limited applicability outside that group (Winker 1993). A further problem is variable shrinkage among individuals (Kinsky & Harper 1968). When shrinkage variation is large, average shrinkage corrections can be applied to sample means, but not to individual specimens (Harris 1980). The only published shrinkage study for petrels that I am aware of is Kinsky & Harper (1968).

The objectives of this study were to evaluate measurement error and shrinkage for a petrel species. After highlighting the strengths and weaknesses for 15 biometrics measured, I select six informative and more robust biometrics.

METHODS

The Short-tailed Shearwater breeds during the Austral summer, with birds arriving from wintering areas in the North Pacific Ocean beginning in September. Large numbers of returning birds "wreck" along the east Australian coast in some years (Marchant & Higgins 1990).

Thirty beach-wrecked Short-tailed Shearwaters were collected in October and November 2013 in northern New South Wales, Australia. Dead birds were collected only when very fresh: with shiny eyes, soft skin with natural colours, and supple toes. Wing and tail feathers were new and fully grown. These specimens were measured when freshly dead and they were then frozen in double polythene bags at about -15 °C.

Biometrics

Fifteen external biometrics were recorded for each bird, including many of those recommended for petrels (Powlesland & Imber 1988, Camphuysen & van Franeker 2007) as well as some less common measurements (references given below). Thirteen of these are illustrated in Fig. 1. A stainless steel vernier caliper (resolution 0.1 mm) was used to measure exposed culmen length (from the outermost curved edge of the bill to the anterior edge of feathering on the forehead), total culmen length (to the intersection of the upper mandible and the cranium), nalospi (from the outermost curved edge of the bill to the anterior edge of the left nostril), bill depth at the base of the exposed culmen, minimum bill depth between the nostrils and maxillary unguis (e.g. Bull *et al.* 2005), bill depth at gonys (hereafter "gonys depth"), bill width at the base of the exposed culmen, bill width at the intersection of the latericorn and the maxillary and mandibular ungues (e.g. Thalmann *et al.* 2007; hereafter "unguis width"), mid-toe and claw length, tarsometatarsus (hereafter "tarsus") length, tarsus height and tarsus width at the middle of the tarsus (e.g. Guicking *et al.* 2004). A modified "head caliper," with a 10 × 14 mm flat pad on the inner jaw, was used to measure head plus bill length. Tail length was measured by gently pushing the outer jaw of the vernier caliper between the central rectrices. Maximum flattened wing chord was measured with a stainless steel wing ruler (resolution 1 mm).

Biometrics were measured twice on each occasion for evaluation of measurement errors. Repeat measurements were non-consecutive and independent. I measured several birds before the repeat pass, so I could not remember measurements from the first pass when repeating measurements.

I performed all of the measurements and used the same measuring techniques, same calipers and wing ruler throughout the study. The same wing (right) and leg (left) were measured on each occasion.

Round skin preparation

Round skins were prepared in April–May 2014. Each bird was measured after thawing and before skinning. Specimens were sexed (17 females, 13 males) and approximately aged (10 immatures, 20 adults) by examination of their sexual organs.

During preparation, I avoided damaging any characters that are measured to obtain the 15 biometrics listed above. I did not cut away any bone from the skull and left the supraoccipital lobe complete. I did not cut away any of the bones from the lower jaw. I did not strip any of the secondary or primary feathers from their natural attachments. To achieve a natural folded wing for these long-winged birds, I tied the wings at the humerus. For drying, the bill was closed with two small zip-ties.

Skins were air-dried for approximately two weeks and then stored in plastic boxes with silica gel desiccant to control humidity (about 35 g per specimen, relative humidity about 50%–70%) and naphthalene to control insects. The silica gel was refreshed every two to three months. Skins were measured at around two weeks, five weeks, eight weeks, 19 weeks and 40 weeks after preparation.

For comparison with Short-tailed Shearwater shrinkage, I used three Wedge-tailed Shearwater *P. pacificus* specimens (collected January 2013; prepared July 2013) and three Fluttering Shearwater *P. gavia* specimens (collected January 2013, January and February 2014; prepared July 2013 and April 2014, respectively) in my collection. These were measured, prepared and stored as described above, but sample sizes were too small to include these data in the main analysis.

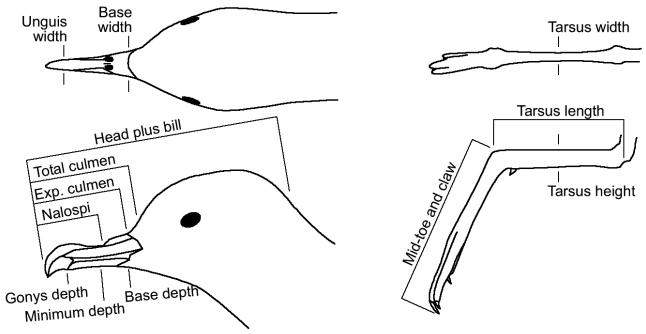

Fig. 1. Head, bill and leg biometrics measured among Short-tailed Shearwaters in this study. Note that bill length measurements are to the outermost curved edge of the bill and not to the pointed tip.

Statistical analyses

Three different linear shrinkage models for paired fresh and dry measurements of a particular character can be used. The intercept-only model (Fig. 2a) is equivalent to a paired *t*-test and predicts constant absolute shrinkage (i.e. difference in means). The slope-only model (Fig. 2b) predicts size-proportional shrinkage (constant relative shrinkage). The slope-and-intercept model has two parameters (Fig. 2c). All of these models can provide a satisfactory fit over a narrow range of sizes. For wider ranges, the intercept-only model is unsatisfactory because large and small birds are not expected to shrink by the same absolute amount. A practical problem for the slope-and-intercept model is that reversals in shrinkage can be predicted for intercept > 0 and slope < 1 or intercept < 0 and slope > 1. A statistical problem is that the ordinary least-squares regression intercept and slope are biased when there is measurement error in the *x* variable (Hutcheon *et al.* 2010). This leaves the slope-only linear model as an appropriate starting point for estimation of shrinkage.

Several prior shrinkage studies have confused the above linear models, applying the paired *t*-test for statistical evaluation of shrinkage and then summarising results as percent shrinkage computed from the mean difference. Other studies have computed mean percent shrinkage from individual shrinkage results. Using scatterplots (Fig. 2), it appears that shrinkage estimation is a problem of prediction, and ordinary least-squares regression is appropriate. For the slope-only model (Fig. 2b), average shrinkage is calculated from the slope (shrinkage = 1 − slope), and the slope is tested for significant shrinkage (slope ≠ 1).

Variable shrinkage among individuals and random measurement errors will increase noise in fresh-dry scatterplots and decrease correlation coefficients. Correlation coefficients can be corrected for measurement error (Spearman 1904), and any remaining weakness (1 − *r*) can then be interpreted as variable shrinkage.

Statistical analyses were performed with R version 3.0.1 (R Core Team 2013). Short-tailed Shearwater biometrics were screened for outliers by plotting shrinkage ratios (relative to fresh size) over time and looking for points that were far outside the general scatter. Outliers were replaced with values from other measuring occasions or remeasured for completely dry skins. There were four outliers corrected out of a total of 900 (30 × 15 × 2) fresh measurements, two outliers in thawed measurements and three in the final dry measurements. A few skins had defects or damage: lost feathers at the base of the exposed culmen (3 specimens), imperfect bill closure (4), curled toes (5) and loose rectrices (1). Affected specimens were excluded from the relevant analyses.

Analysis of variance (ANOVA) was used to partition variance for each biometric into variance among and within individuals. ANOVA assumptions were checked with linear-model diagnostic plots. Measurement errors and associated confidence intervals were calculated following McGraw & Wong (1996). Measurement error is the complement of intra-class correlation (ME = 1 − ICC).

Means of repeated measures were used to reduce measurement error (Bailey & Byrnes 1990) for regression and correlation analyses. Shrinkage was estimated by ordinary least-squares regression, and slope-only models were assumed. For slope-only models, the intercept is anchored at zero, and the slope is not biased by measurement error in the *x* variable. I verified this with simulated data. Ordinary least-squares estimates from slope-only models were therefore suitable for statistical inference. Ordinary least-squares assumptions were checked with linear-model diagnostic plots.

Linear correlations were quantified using Pearson's product-moment correlation. Fresh-dry correlations were corrected for measurement error attenuation with Spearman's (1904) formula:

$$r_{xyc} = \frac{r_{xy}}{\sqrt{r_{xx}\, r_{yy}}}$$

where r_{xy} is Pearson's product-moment correlation, r_{xyc} is corrected, r_{xx} is the intra-class correlation for the *x* variable and r_{yy} is the intra-class correlation for the *y* variable. Intra-class correlations for means of repeated measures were calculated following McGraw & Wong (1996). Corrected correlations > 1 were set equal to one. Shrinkage variation (SV) was defined as the unexplained variance for a corrected fresh-dry correlation:

$$SV = 1 - r_{xyc}^2 .$$

RESULTS

Some drying and shrinkage of specimens occurred after 6–8 months in the freezer (Fig. 3). Fresh-thawed average shrinkage detected in slope-only models ranged from 0% for head plus bill to 4% for bill base depth. Total culmen increased by 2%. Fresh-thawed shrinkage was significant (shrinkage ≠ 0, $P < 0.05$) except for nalospi, bill

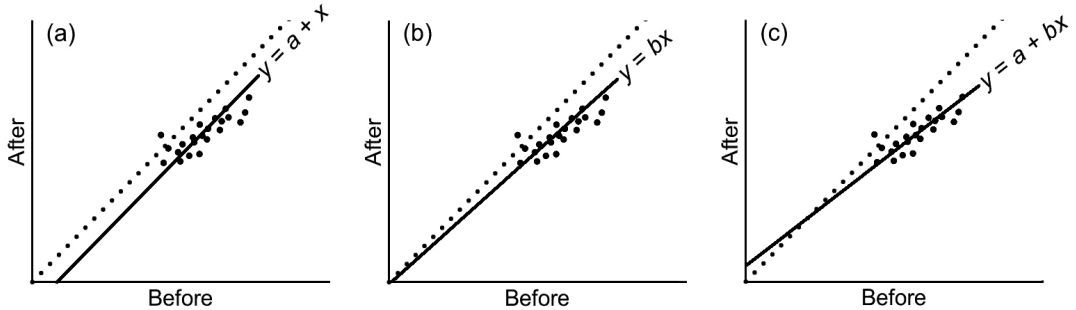

Fig. 2. Illustration of three different linear shrinkage models (solid lines): (a) intercept-only model (constant shrinkage), (b) slope-only model (size proportional shrinkage), (c) slope and intercept model (illustrated with intercept > 0 and slope < 1). Lines of agreement are dotted.

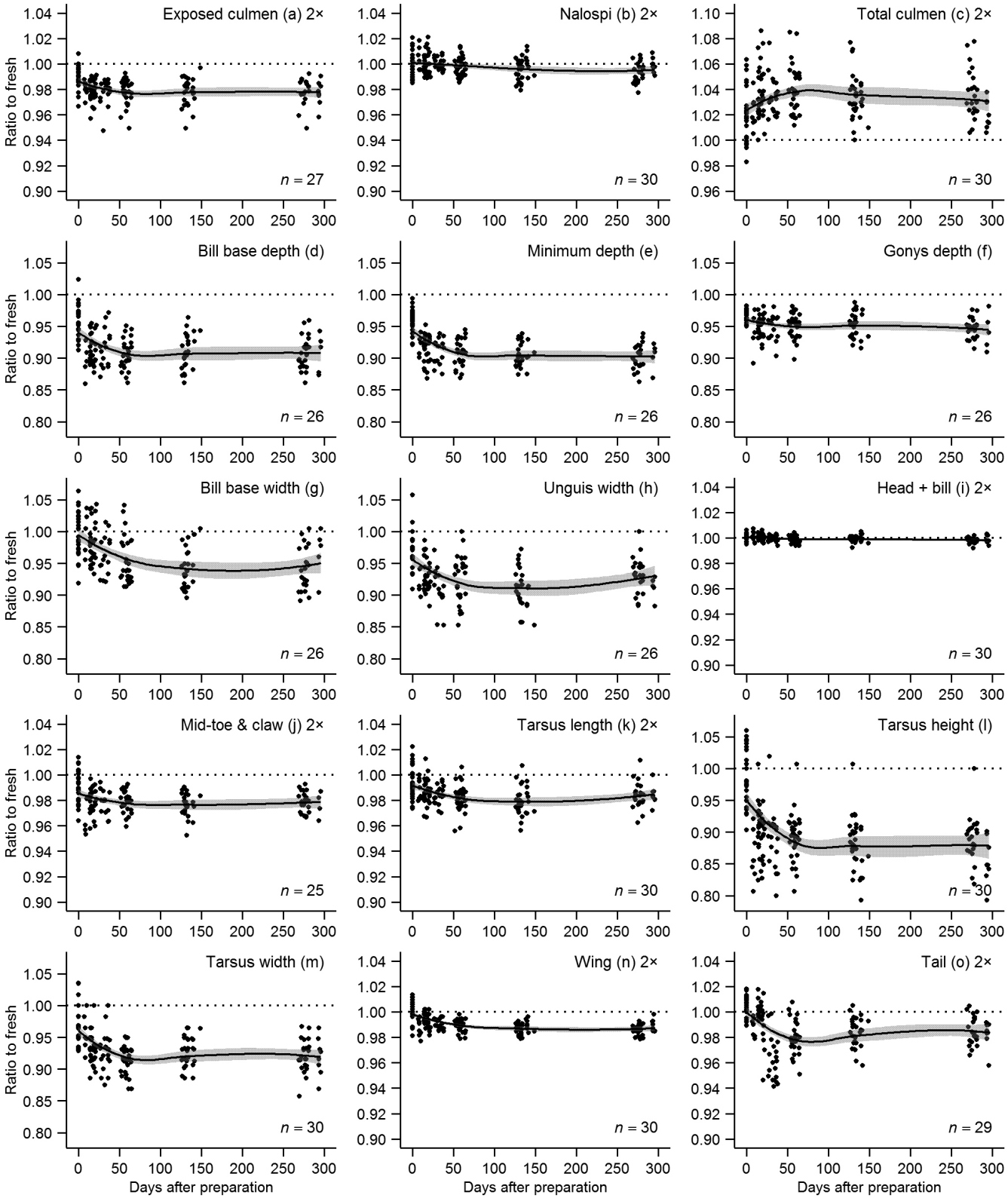

Fig. 3. Shrinkage ratios relative to fresh size over time for 15 Short-tailed Shearwater biometrics. Results at zero days are for thawed specimens. Expanded scales (2×) are used for biometrics with smaller shrinkage and less variation. Dotted lines indicate zero change (ratio = 1). Curved lines are "loess" curves (local polynomial regression fitting) with 95% confidence bands to assist detection of trends. Sample sizes vary because a few round skins had defects or damage for particular biometrics (exposed culmen, bill depths, bill widths, mid-toe and claw, tail).

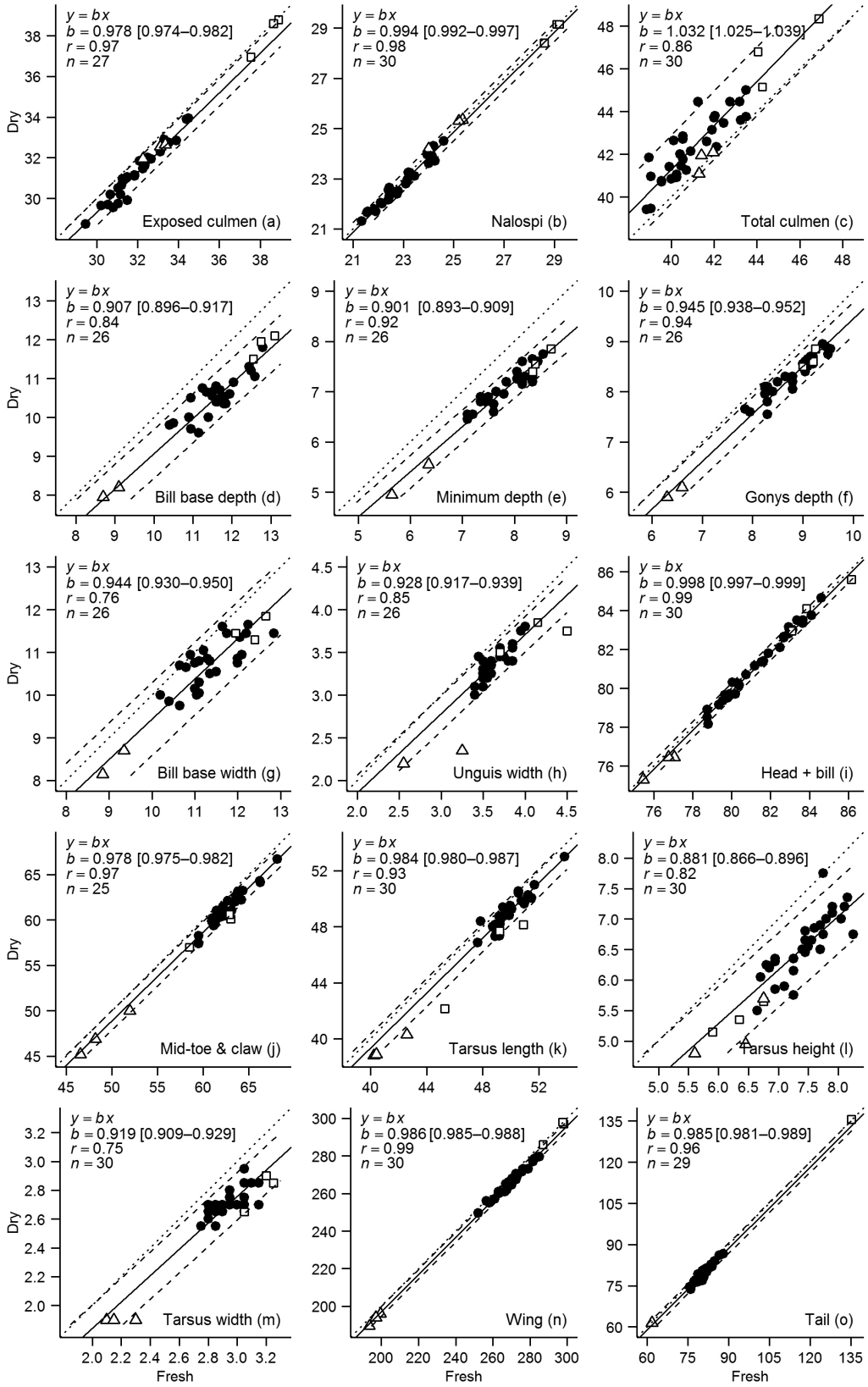

Fig. 4. Fresh-dry shrinkage models (solid lines) fitted to 15 biometrics for Short-tailed Shearwater (solid circles). Fresh and dry measurements on axes are in millimetres. Ordinary least squares regression slopes are reported with 95% confidence intervals in square brackets. Average shrinkage is calculated from the slopes (shrinkage = 1 − slope). Correlation results are Pearson product-moment correlations (see Fig. 6). Sample sizes are for Short-tailed Shearwaters. Lines of agreement are dotted. Dashed lines show 95% prediction intervals. Supplementary data for three Wedge-tailed Shearwaters (open squares) and three Fluttering Shearwaters (open triangles) are shown for comparison. Sample sizes vary because a few round skins had defects or damage for particular biometrics (exposed culmen, bill depths, bill widths, mid-toe and claw, tail).

base width, head plus bill, tarsus height and wing.

Mean biometrics stabilised within 2–5 mo after skin preparation (Fig. 3). This indicates that shrinkage was largely complete, and there was minimal systematic observer error between measuring occasions. Fresh-dry average shrinkage from slope-only models ranged from 0.2% for head plus bill to 12% for tarsus height (Fig. 4). Total culmen increased by 3% as the skin retracted around the skull. Shrinkage was large for soft parts (12% for tarsus height, 9% for bill base depth), parts of the bill not supported by bone (6% for gonys depth, 6% for bill base width) and small characters (10% for minimum bill depth, 7% for unguis width, 8% for tarsus width). Supplementary measurements from three Wedge-tailed Shearwater specimens (measured 13 months after preparation) and three Fluttering Shearwaters (measured 11 months after preparation) agreed with predictions from the Short-tailed Shearwater shrinkage models (Fig. 4).

Fresh measurement errors were large (>10%; Lougheed *et al.* 1990) for small characters (32% for unguis width, 36% for tarsus width), soft parts (11% for bill base depth, 21% for tarsus height) and for biometrics that lacked well-defined measurement landmarks (15% for total culmen, 20% for bill base width; Fig. 5). Measurement error decreased for the mean of two repeat measurements, although four fresh biometrics still had large measurement errors: bill base width (11%), unguis width (19%), tarsus height (12%) and tarsus width (22%). Measurement errors for round skins averaged 0.4 times smaller than those for freshly dead birds (Fig. 5). Dry measurement error was large only for tarsus width (26%). Measurement error decreased with increasing mean size for both fresh biometrics (Spearman's rank-based correlation: n = 15, $r_S = -0.66$, $P = 0.01$) and dry (n = 15, $r_S = -0.76$, $P < 0.01$).

Correlations between paired fresh and dry measurements were very strong ($r > 0.9$) except for those of total culmen, bill base depth, bill base width, unguis width, tarsus height and tarsus width (Fig. 6). After correcting correlations for measurement error attenuation, three biometrics had shrinkage variation >19% ($r_c < 0.9$): bill base depth, bill base width and tarsus height (Fig. 6). Shrinkage variation increased with increasing average shrinkage (n = 15, $r_S = 0.78$, $P < 0.001$).

Correlations between fresh biometrics were strong ($r > 0.7$) for bill lengths and head plus bill, for bill depths and for mid-toe and claw and tarsus lengths (Table 1). The dataset can be simplified by discarding two of the three bill lengths, two of the three bill depths

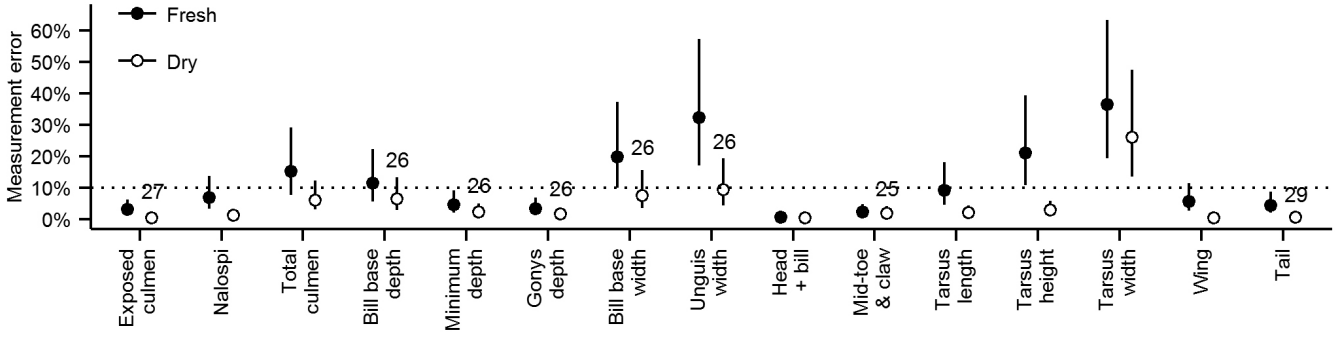

Fig. 5. Random measurement error for 15 fresh (solid circles) and 15 dry (open circles) Short-tailed Shearwater biometrics. Lines show 95% confidence intervals. Measurement error > 10% (dotted line) was considered to be large by Lougheed *et al.* (1991). Sample sizes were 30 except where indicated above the intervals. Dry sample sizes vary because a few round skins had defects or damage for particular biometrics (exposed culmen, bill depths, bill widths, mid-toe and claw, tail).

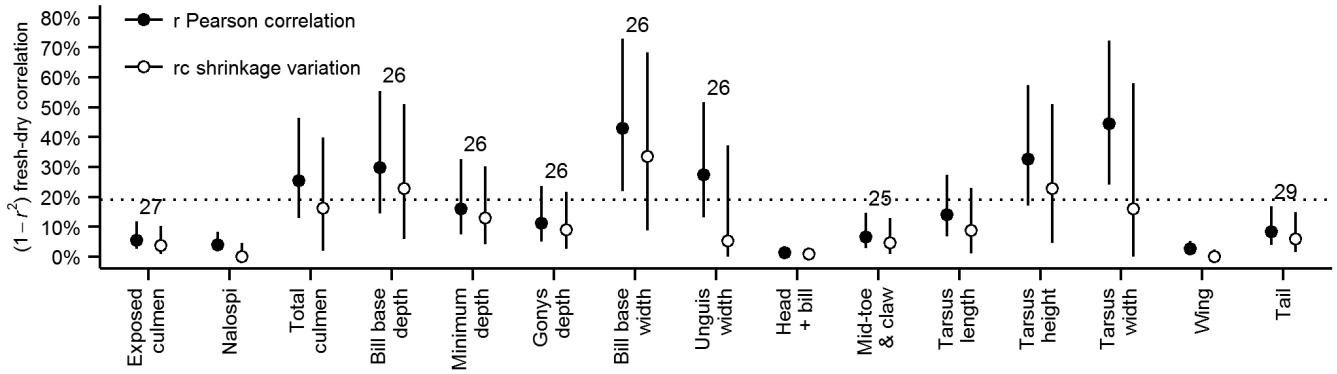

Fig. 6. Fresh-dry correlations for 15 Short-tailed Shearwater biometrics before (solid circles) and after (open circles) correction for measurement error attenuation. Unexplained variance ($1 - r^2$) is presented. Lines show 95% confidence intervals. Unexplained variance for corrected correlations was interpreted as shrinkage variation. Shrinkage variation > 19% ($r_C < 0.9$) was considered to be large (dotted line). Sample sizes were 30 except where indicated above the intervals. Sample sizes vary because a few round skins had defects or damage for particular biometrics (exposed culmen, bill depths, bill widths, mid-toe and claw, tail).

and one of mid-toe and claw or tarsus length without any substantial loss of information compared with the complete dataset.

DISCUSSION

This study examined random measurement error and specimen shrinkage for 15 external biometrics, using Short-tailed Shearwater *Puffinus tenuirostris* as the subject species. Although most of these biometrics had low measurement error and shrinkage, six troublesome biometrics were identified: total culmen, bill base depth, bill base width, unguis width, and tarsus height and width (Table 2). Two of these, bill base depth and width, are commonly measured for petrels (e.g. Powlesland & Imber 1988, Camphuysen & Franeker 2007). Six more robust biometrics identified were nalospi, gonys depth, head plus bill, tarsus length, wing and tail (Table 2).

Measurement error

Measurement error decreased from fresh to dry specimens. Skins were measured under laboratory conditions; the dry birds were rigid and self-supporting, and soft parts had hardened. By contrast, fresh birds were measured under field conditions; the limp, dead birds required support while measuring, soft parts were flexible and the feathers were often damp. Even larger measurement errors are expected for live birds, which tend to struggle in the hand.

Measurement error increased for small characters, as small measurements are sensitive to errors in caliper positioning and instrument precision. For example, a 0.1 mm error is only 0.1% of mean head plus bill length but 2.7% of mean fresh unguis width in the Short-tailed Shearwaters of this study. Negative correlations found between measurement error and mean size in this study ($r_S = -0.66$ for fresh biometrics and $r_S = -0.76$ for dry biometrics) agreed with $r = -0.42$ in Yezerinac *et al.* (1992), who measured skeletal characters for seven passerine species. Lougheed *et al.* (1991) found no correlation between measurement error and size ($r = -0.02$); however, their result was based on dissimilar biometrics (skeletal and external characters) for two dissimilar birds (a sparrow and a coot).

Measurement error also increased for soft parts and for biometrics that lacked well-defined measurement landmarks. Total culmen was difficult to measure because the junction of the bill and cranium is obscured by feathers. Bill base width was difficult because the bill is tapered near the gape. Bill base depth measurement error of 11% and bill base width measurement error of 20% for fresh Short-tailed Shearwaters in this study were similar to 7% and 16%, respectively, for freshly dead American Coots *Fulica americana* in Lougheed *et al.* (1991).

Biometrics should be selected because they are informative and not simply because they can be measured precisely (Bailey & Byrnes 1990). For example, bill base depth is informative for sexual size dimorphism in shearwaters (Bull *et al.* 2005). Researchers can counteract measurement error by increasing sample sizes and statistical power or by averaging repeated measurements to increase precision (Bailey & Byrnes 1990). Only measurements that are known to have high measurement error should be repeated (Yezerinac *et al.* 1992). Repeats are particularly helpful to detect gross measurement errors that should be repeated again or otherwise corrected (Lougheed *et al.* 1991).

Shrinkage

Shrinkage stabilised within 2–5 months after skin preparation, and further shrinkage is not expected or unlikely to be substantial. Previous studies have reported that shrinkage of bird skins stabilised within 2 months (Harris 1980, Knox 1980), 6 months (Green 1980) and 12 months (Ewins 1985). Engelmoer *et al.* (1983) noted that

TABLE 1
Linear correlations between 15 fresh biometrics for Short-tailed Shearwaters (n = 30)[a]

Biometric	Nalospi	Total culmen	Bill depth	Minimum depth	Gonys depth	Bill width	Unguis width	Head plus bill	Mid-toe and claw	Tarsus length	Tarsus height	Tarsus width	Wing	Tail
Exposed culmen	**0.80**	**0.77**	0.42	-	-	0.65	0.47	0.70	-	0.50	-	-	-	-
Nalospi		**0.78**	-	-	-	0.53	0.47	**0.86**	0.51	0.51	-	-	0.56	-
Total culmen			-	-	-	-	0.52	**0.85**	0.52	0.49	-	-	0.49	-
Bill base depth				**0.83**	**0.75**	0.46	0.63	0.49	-	-	0.54	-	-	-
Minimum depth					**0.82**	0.46	0.60	0.46	-	-	0.58	-	-	-
Gonys depth						0.54	0.54	0.46	-	-	0.64	0.47	-	-
Bill base width							0.42	0.42	-	0.43	0.57	0.57	-	-
Unguis width								0.52	0.40	0.40	0.44	-	0.45	-
Head plus bill									0.47	-	-	-	0.52	-
Mid-toe and claw										**0.74**	-	-	0.62	-
Tarsus length											-	-	0.51	-
Tarsus height												-	-	-
Tarsus width													-	-
Wing														0.50

[a] Strong correlations (in bold; $r > 0.7$) suggest redundancy. Weak correlations ($r < 0.4$) are not presented.

shrinkage for shorebird skins that were several decades old was the same as for new skins. While long-term shrinkage may not be a problem, methods of preparation (e.g. Winker 1993) and wear and damage for old specimens can result in differences when compared with newer skins. Some shrinkage also occurred while the birds were frozen, and it should not be assumed that measurements of thawed birds are equivalent to those for freshly dead birds (also see Bjordal 1983).

Comparisons of measurements between dry and live or freshly dead birds are flawed when shrinkage is large relative to the size differences of interest (Winker 1993). For example, fresh-dry shrinkage was large relative to sexual size dimorphism (SSD) in Short-tailed Shearwaters for five biometrics. Bill base depth shrinkage of 9% was greater than 6% SSD for Short-tailed Shearwater round skins in Bull *et al.* (2005). Minimum bill depth shrinkage of 10% was greater than 7% SSD for live Short-tailed Shearwaters in Carey (2011) and 6% SSD for round skins in Bull *et al.* (2005). Kinsky & Harper (1968) similarly reported large 6%–12% bill width shrinkage for three prion species. Culmen shrinkage of 2% was less than 3% SSD in Carey (2011), but equal to 2% SSD in Bull *et al.* (2005). Tarsus length shrinkage of 2% was equal to 2% SSD in Carey (2011) and greater than 0.3% SSD in Bull *et al.* (2005). Mid-toe and claw shrinkage of 2% was greater than 1% SSD in Bull *et al.* (2005). Correction for shrinkage is needed before applying biometric classification criteria for sex, subspecies or species that are based on museum specimens to live or freshly dead birds (Winker 1993).

Multivariate biometrics (e.g. ratios of different biometrics, principal components) can also be affected by shrinkage, because variable shrinkage among different characters will distort shapes. For example, 9% bill base depth and 6% bill base width shrinkage were large relative to 2% exposed culmen shrinkage, and round skins had more slender bills than freshly dead Short-tailed Shearwaters.

Size-proportional shrinkage models were supported by supplementary data that I obtained from three Wedge-tailed Shearwaters and three Fluttering Shearwaters, indicating that general shrinkage models could be developed for groups of petrels that share similar morphology. However, Engelmoer *et al.* (1983) reported that relative wing length shrinkage was not constant, increasing with size for 13 shorebird species. A scatterplot of individual measurements for all species together would have been helpful for interpreting those data.

Shrinkage variation

Several earlier studies have promoted shrinkage corrections for skin biometrics, yet few have considered variable shrinkage among individuals. Shrinkage variation increased strongly with average shrinkage, and was large for bill base depth, bill base width and tarsus height. Kinsky & Harper (1968) similarly reported highly variable bill width shrinkage among individual prions.

Harris (1980) reported highly variable shrinkage among individual Atlantic Puffins for wing and bill lengths. However, shrinkage variation was small for wing, exposed culmen and nalospi in this study. The two exceptional shrinkage results from 38 wings in Harris (1980) could have been gross measurement errors. Neither Kinsky & Harper (1968) nor Harris (1980) considered that random measurement error increases variance in individual shrinkage results. It is best to compare like with like and avoid shrinkage problems altogether. This may require efforts to collect new specimens or to measure live birds in the field (Winker 1996). Accurate shrinkage corrections are often impossible because specimen shrinkage has been measured only for a limited number of taxa.

Selecting biometrics

Biometrics research will benefit from careful selection of measurements. Informative biometrics can be identified from theory and earlier investigations. Biometrics with low measurement error and low shrinkage reduce noise in the data and increase confidence in shrinkage corrections when required. Highly correlated biometrics increase data collection costs and add little extra information.

From the 15 measurements evaluated in this study, I selected six informative and more robust biometrics (Table 2) based on the above considerations. Among the three highly correlated bill lengths, I selected nalospi rather than exposed culmen, for which skins were affected by shrinkage at the tubes and lost feathers, or total culmen, which had large measurement error. Among the three highly correlated bill depths, I selected gonys depth rather than bill base depth, which had large measurement error and shrinkage, or minimum depth, which had large shrinkage. I did not select any of the two bill widths, because these had large measurement error and shrinkage and have not proven to be informative for shearwaters. Thalmann *et al.* (2007) did use unguis width for sexing Flesh-footed Shearwaters *P. carneipes*; however, the 90% sex classification accuracy (n = 104 birds) for their discriminant function including unguis width was certainly not significantly greater than 89% accuracy without unguis width. Head plus bill is a powerful biometric because it has low measurement error and low shrinkage.

TABLE 2
Summary of the strengths (+) and weaknesses (–) of 15 biometrics for Short-tailed Shearwaters; selected informative and more robust biometrics in bold (see Discussion)

Biometric	Low measurement error	Low shrinkage	Low shrinkage variation
Culmen	+ +	+	+ +
Nalospi	+	+ +	+ +
Total culmen	–	–	–
Bill base depth	–	– –	– –
Minimum depth	+ +	– –	–
Gonys depth	+ +	–	+
Bill base width	–	–	– –
Unguis width	– –	– –	+
Head plus bill	+ +	+ +	+ +
Mid-toe and claw	+ +	+	+ +
Tarsus length	+	+	+
Tarsus height	– –	– –	– –
Tarsus width	– –	– –	–
Wing	+	+	+ +
Tail	+ +	+	+

A "head caliper" should be used to make this measurement. Leg and foot measurements were strongly correlated, and I selected tarsus length. I did not select mid-toe and claw, because curled toes are a common attribute of skin specimens (17% of 30 skins in this study). Furthermore, I did not select tarsus width or height, because these had large measurement error and shrinkage and have not proven to be informative for shearwaters. Guicking *et al.* (2004) reported significant differences in tarsus height and width among individuals from two colonies of Pink-footed Shearwater *P. creatopus*. However, there was no biological theory underlying that comparison, and the result could be due to systematic measurement error and therefore spurious. Bourgeois *et al.* (2007) did not find tarsus height informative for sexual size dimorphism in Yelkouan Shearwaters *P. yelkouan*. I did select both wing and tail, although feather measurements are affected by wear and moult (Pienkowski & Minton 1973).

The six more robust biometrics selected above are generally measurements of large, inflexible characters with well-defined measurement landmarks.

ACKNOWLEDGEMENTS

Birds were acquired and held under General Licence MWL000101288 issued by the Office of Environment and Heritage, New South Wales, Australia. This paper benefited from comments provided by Peter Pyle, Edward Soldaat, Alan Tennyson and Dave Watson. Hadley Wickham is also thanked for the ggplot2 graphics package for R (Wickham 2009).

REFERENCES

AINLEY, D.G. 1980. Geographic variation in Leach's Storm-Petrel. *Auk* 97: 837–853.

BAILEY, R.C. & BYRNES, J. 1990. A new, old method for assessing measurement error in both univariate and multivariate morphometric studies. *Systematic Zoology* 39: 124–130.

BJORDAL, H. 1983. Effects of deep freezing, freeze-drying and skinning on body dimensions of House Sparrows (*Passer domesticus*). *Cinclus* 6: 105–108.

BOURGEOIS, K., CURÉ, C., LEGRAND, J., GÓMEZ-DÍAZ, E., VIDAL, E., AUBIN, T. & MATHEVON, N. 2007. Morphological versus acoustic analysis: what is the most efficient method for sexing Yelkouan shearwaters *Puffinus yelkouan*? *Journal of Ornithology* 148: 261–269.

BRETAGNOLLE, V. & SHIRIHAI, H. 2010. A new taxon of Collared Petrel *Pterodroma brevipes* from the Banks Islands, Vanuatu. *Bulletin of the British Ornithologists' Club* 130: 286–301.

BULL, L.S., BELL, B.D. & PLEDGER, S. 2005. Patterns of size variation in the shearwater genus *Puffinus*. *Marine Ornithology* 33: 27–39.

CAMPHUYSEN, C.J. & VAN FRANEKER, J.A. 2007. Procellariidae: Petrels and shearwaters. Technical documents 4.1. In: CAMPHUYSEN, C.J., BAO, R., NIJKAMP, H. & HEUBECK, M. (Eds.) *Handbook on Oil Impact Assessment.* Version 1.0. European Oiled Wildlife Response Assistance (EUROWA). [Available online at: http://www.oiledwildlife.eu/. Accessed 6 April 2015].

CAREY, M.J. 2011. Sexual size dimorphism, within-pair comparisons and assortative mating in the short-tailed shearwater (*Puffinus tenuirostris*). *Notornis* 58: 8–16.

EINODER, L.D., PAGE, B. & GOLDSWORTHY, S.D. 2008. Sexual size dimorphism and assortative mating in the Short-tailed Shearwater *Puffinus tenuirostris*. *Marine Ornithology* 36: 167–173.

ENGELMOER, M., ROSELAAR, K, BOERE, G.C. & NIEBOER, E. 1983. Post-mortem changes in measurements of some waders. *Ringing & Migration* 4: 245–248.

EWINS, P.J. 1985. Variation of Black Guillemot wing lengths post-mortem and between measurers. *Ringing and Migration* 6: 115-117.

GENOVART, M., McMINN M. & BOWLER, D. 2003. A discriminant function for predicting sex in the Balearic shearwater. *Waterbirds* 26: 72–76.

GRANADEIRO, J.P. 1993. Variation in measurements of Cory's shearwater between populations and sexing by discriminant analysis. *Ringing and Migration* 14: 103–112.

GREEN, G.H. 1980. Decrease in wing length of skins of Ringed Plover and Dunlin. *Ringing and Migration* 3: 27–28.

GUICKING, D., FIEDLER, W., LEUTHER, C., SCHLATTER, R. & BECKER, P.H. 2004. Morphometrics of the pink-footed shearwater (*Puffinus creatopus*): influence of sex and breeding site. *Journal for Ornithology* 145: 64–68.

HARRIS, M.P. 1980. Post-mortem shrinkage of wing and bill of puffins. *Ringing and Migration* 3: 60–61.

HUTCHEON, J.A., CHIOLERO, A. & HANLEY, J.A. 2010. Random measurement error and regression dilution bias. *BMJ* 340: c2289.

IMBER, M.J. & TENNYSON, A.J.D. 2001. A new petrel species (Procellariidae) from the south-west Pacific. *Emu* 101: 123–127.

JENKINSON, M.A. & WOOD, D.S. 1985. Avian anatomical specimens: a geographic analysis of needs. *Auk* 102: 587–599.

KINSKY, F.C. & HARPER, P.C. 1968. Shrinkage of bill width in skins of some *Pachyptila* species. *Ibis* 110: 100–102.

KNOX, A. 1980. Post-mortem changes in wing-lengths and wing-formulae. *Ringing and Migration* 3: 29–31.

LOUGHEED, S.C., ARNOLD, T.W. & BAILEY, R.C. 1991. Measurement error of external and skeletal variables in birds and its effect on principal components. *Auk* 108: 432–436.

MARCHANT, S. & HIGGINS, P.J. 1990. *Handbook of Australian, New Zealand and Antarctic Birds.* Volume 1: Part A: Ratites to petrels. Melbourne, Australia: Oxford University Press.

McGRAW, K.O. & WONG, S.P. 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1: 30–46.

MEATHREL, C.E, BRADLEY, J.S., WOOLER, R.D. & SKIRA, I.J. 1993. The effect of parental condition on egg-size and reproductive success in Short-Tailed Shearwaters *Puffinus tenuirostris*. *Oecologia* 93: 162–164.

PETTIT, T.N., BYRD, G.V., WHITTOW, G.C. & SEKI, M.P. 1984. Growth of the Wedge-tailed Shearwater in the Hawaiian Islands. *Auk* 101: 103–109.

PIENKOWSKI, M.W. & MINTON, C.D.T. 1973. Wing length changes of the knot with age and time since moult. *Bird Study* 20: 63–68.

POWLESLAND, R.G. & IMBER, M.J. 1988. OSNZ Beach Patrol Scheme: information and instructions. *Notornis* 35: 143–153.

R CORE TEAM. 2013. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. [Available online at: http://www.R-project. org/. Accessed 30 July 2013].

SPEAR, L.B. & AINLEY, D.G. 1998. Morphological differences relative to ecological segregation in petrels (family: Procellaridae) of the Southern Ocean and tropical Pacific. *Auk* 115: 1017–1033.

SPEARMAN, C. 1904. The proof and measurement of association between two things. *American Journal of Psychology* 15: 72–101.

THALMANN, S., BAKER, G.B., HINDELL, M., DOUBLE, M.C. & GALES, M. 2007. Using biometric measurements to determine gender of Flesh-footed Shearwaters, and their application as a tool in long-line by-catch management and ecological field studies. *Emu* 107: 231–238

WATSON, D.M. 2005. Diagnosable versus distinct: evaluating species limits in birds. *BioScience* 55: 60–68.

WICKHAM, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

WINKER, K. 1993. Specimen shrinkage in Tennessee Warblers and "Traill's" Flycatchers. *Journal of Field Ornithology* 64: 331–336.

WINKER, K. 1996. Specimen shrinkage versus evolution: I'iwi morphology. *Conservation Biology* 10: 657–658.

YEZERINAC, S.M., LOUGHEED, S.C. & HANDFORD, P. 1992. Measurement error and morphometric studies: statistical power and observer experience. *Systematic Biology* 41: 471–482