Ann. Zool. Fennici 49: 81–92 ISSN 0003-455X (print), ISSN 1797-2450 (online)
Helsinki 20 April 2012

Observer variability in measuring animal biometrics and fluctuating asymmetry when using digital analysis of photographs

Anne E. Goodenough[1,*], Angela L. Smith[2], Hannah Stubbs[1], Rachel Williams[1] & Adam G. Hart[1]

[1] *Department of Natural and Social Sciences, University of Gloucestershire, Cheltenham, GL50 4AZ, UK (*corresponding author's e-mail: aegoodenough@glos.ac.uk)*
[2] *Gloucester Museums Service, Gloucester City Museum & Art Gallery, Gloucester, GL1 1HP, UK*

Received 26 Apr. 2011, final version received 25 Oct. 2011, accepted 27 Oct. 2011

Goodenough, A. E., Smith, A. L., Stubbs, H., Williams, R. & Hart, A. G. 2012: Observer variability in measuring animal biometrics and fluctuating asymmetry when using digital analysis of photographs. — *Ann. Zool. Fennici* 49: 81–92.

Ecological research using biometric data is only sound if biometrics themselves are accurate and not confounded by measurement error. Given concerns about the accuracy of biometrics taken directly (physical measurement of animals), digital measurement of photographs is often advocated, particularly for small or live specimens. However, there is currently limited understanding of intra- and inter-observer variability of such measurements (i.e. variability of multiple measurements of the same specimen by the same observer, and variability of multiple measurements of the same specimen by different observers, respectively). We took biometrics (two linear, two curvilinear, two angular) from moth photographs using standard image software and calculated two fluctuating asymmetry measures. Inter-observer variability was always higher than intra-observer variability. Measurement error was low for linear/curvilinear measurements (< ~4%), but high for angular variables (52%) and asymmetry measures (45%). Measurement precision correlated positively with trait size. Variability caused significant differences in mean measurements inter-specifically for half the biometrics; there were no significant intra-specific differences. We discuss the implications of our findings for research using photographically-derived biometrics and offer recommendations for reducing measurement error.

Introduction

Biometric data are used frequently in ecological research to characterise and compare individuals in taxonomic and phylogenetic studies (e.g. Nowak 2002, Smith *et al*. 2004), as well as forming an integral part of research into life-history traits, behaviour, and evolution (e.g. Ashton 2002, Kingsolver & Pfennig 2004, Molina-Borja & Rodríguez-Domínguez 2004, Mutanen & Kaitala 2006). In addition, measurements of physical traits are frequently used to quantify the condition of individuals. Such measures, often used as a proxy for fitness, are then related to other ecological variables including dominance, inbreeding coefficients, parasite load, and habitat quality. One common fitness proxy derived from biometric data is Fluctuating Asymmetry (FA), which utilises the differences in the sizes of bilateral traits (i.e. features that occur on both sides of the same individual); for example, wing length in birds (Björklund 1996), femur length in butterflies (Gage 1998), and facial variables of primates (Sefcek & King 2007). As the growth of bilateral traits is controlled by the same gene (Andersson 1994) such traits should, theoretically, be symmetrical. However, environmental stress can increase within-individual developmental instability (Debat *et al.* 2000), and thus adversely affect the precision of this developmental homeostasis (Hoffmann & Parsons 1991, Björklund 1996). In this way, FA can provide an indication of individual fitness (Parsons 1992, van Dongen 2006).

In order for research using biometrics to be rigorous, data need to be accurate and precise. Concern over intra- and inter-individual variability reducing the precision of biometric measurements (i.e. differences in repeated measurements of the same trait on the same individual by a single observer or differences in repeated measurements of the same trait on the same individual by multiple observers, respectively) has been noted for some time, following a series of small-scale studies (e.g. Nisbet *et al.* 1970, Pankakoski *et al.* 1987, Palmeirim 1998). Recent detailed research on museum bird specimens (Goodenough *et al.* 2010) demonstrated quantitatively that biometrics are indeed subject to very high levels of variability, both within and between observers. On average, the relative amount of variation in biometric measurements due to observer error rather than true (biological) variation was 13% intra-specifically, rising to 38% inter-specifically. The error in measures of asymmetry, calculated from the original biometrics, was 86% intra-specifically and 90% inter-specifically.

Because of concerns regarding the accuracy of biometrics taken directly from the individuals under study (e.g., physical measures, usually using callipers) (*see* Heathcote 1981), digital analysis of photographs to obtain measurements is often advocated. Although the technique is not always suitable, for example when traits differ in three dimensions (Faurby *et al.* 2011), it is a commonly-used and approved method where species can be well-represented in two dimensions (Loeschcke *et al.* 1999, Faurby *et al.* 2005); for example butterflies mounted with wings outstretched or flat fish such as rays. It is also often used for small species such as insects or where live individuals are involved (e.g. Hill *et al.* 2005, Sefcek & King 2007, Davis *et al.* 2008). As a result, measurements from digital images are frequently used in research using biometrics (e.g. Hassall *et al.* 2008, Gidaszewski *et al.* 2009) and in studies of asymmetry (e.g. Vilisics *et al.* 2005, Hassall & Thompson 2009). However, the extent of intra- and inter-specific variability in digitally-measured animal biometrics has not been investigated, such that the suitability of using, and indeed recommending, this approach is not clear. Work in the biomedical sciences has shown that digital image measurements can be taken consistently within- and between-observers in some situations [e.g. in measuring airway lumen (Masters *et al.* 2005), tympanic membrane perforation (Ibekwea *et al.* 2009) and major foetal variables (Perni *et al.* 2004)], but this is not always the case (e.g. Adam *et al.* 2005). Similar studies are needed in ecology to understand the baseline level of intra- and inter-observer variability in recording animal biometrics and to quantify how different types of biometrics (for example, linear, curvilinear or angular) differ in their susceptibility to such variability.

In this study, we use photographs of individual moths of different species to obtain repeated digital biometric measurements within and between observers (i.e. repeatability and reproducibility: Gosler 2004) using industry-standard image analysis software. We quantify both intra- and inter-observer variability in initial biometrics to ascertain their relative importance. This appears to be the first attempt by ecologists to quantify the relative importance of variability in digital measurements of animal biometrics, as well as any interactions between them, simultaneously. We also quantify two different asymmetry measurements, frequently used in entomological studies as proxies of fitness, to establish if these are subject to significant intra-

or inter-observer variability and whether this is higher or lower than variability in the original biometric measurements (i.e. whether variability is magnified or nullified during calculation of the fitness measures).

Methods

Photographic capture

Twenty two individual adult moths, each from a different species, were selected from the Lesley Price Entomological Collection housed at the Gloucester City Museum and Art Gallery (Gloucestershire, UK). The moths ranged in total body length from < 3 mm to > 20 mm and were drawn from 13 different genera in the families Arctiidae, Geometridae and Noctuidae. Each specimen, which had previously been mounted using a single entomological pin through the thorax, was removed from its case and remounted onto white paper with a cork backing; when viewed perpendicularly, each specimen was almost 2D in appearance (Fig. 1). One photograph of each moth was taken, using a Canon EOS 450D camera, fitted with a 18–55 mm EF Canon lens (exposure 1/30 sec at F13). All photographs were taken from the same physical distance, and at the same focal length (55 mm), to ensure comparability. The camera was positioned absolutely perpendicularly to the mounted specimen, which was located in the centre of the photograph and well away from edges of the image to avoid issues of barrelling or distortion. All resultant JPG image files were 4272 × 2848 pixels, with every 422 pixels equating to 1 cm (calibrated by measuring a known distance on graph paper, which had been photographed at the same physical distance and focal length as the specimens).

Recording biometrics

Six biometrics, all commonly used in entomological studies, were recorded for each moth using ImageJ software (Abramoff *et al*. 2004; http://rsbweb.nih.gov/ij). Two of the biometrics were linear (total body length and maximum wingspan) and were taken using the straight line tool, two were curvilinear (costal margin of the left and right forewing) and were taken using the segmented line tool, while the final two were angular (apex angle of the left and right forewing) and were taken using the angle tool (Fig. 1). All linear/curvilinear measurements were taken in pixels and converted to mm post-hoc by dividing the pixel count by 422 and multiplying the result by 10. All angular measurements were taken in degrees and were not transformed. To ensure consistency between photographs and observers, all measurements were made at 50% magnification on identical LCD screens. It is recognised that the exact method of mounting each moth could have influenced the variables that we measured (e.g. body length influenced by the angle between the prothorax and abdomen). This would be problematic if we were interested in the resultant values; for example, as part of a biometric study. However, here we are comparing between repeated measured made by the same and different recorders — and crucially on the same photographs — to check consistency in recording the same variables under exactly the same conditions, such that these issues do not apply.

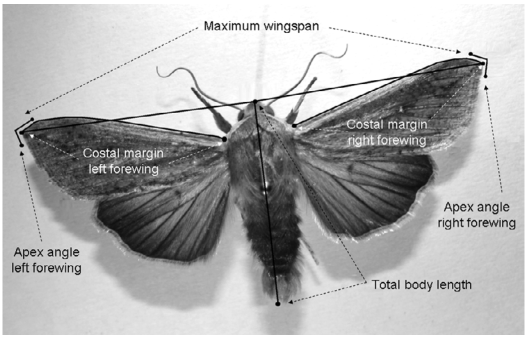
Maximum wingspan
Costal margin left forewing
Costal margin right forewing
Apex angle right forewing
Apex angle left forewing
Total body length

Fig. 1. Biometric measurements used in this study. Two measures of asymmetry were also calculated, one using the left and right costal wing margin measurements and one using the left and right apex angles. This photograph shows specimen 11 (large footman; *see* Appendix for moth-specific data) and is typical of the photographs used in this study (although this has been cropped post-processing for display purposes).

To determine inter-observer measurement variation, biometrics were recorded by five observers (each author). To ensure that each set of measurements was independent, each observer worked in isolation and values were stored electronically. Then, to determine intra-observer variation, each observer measured each moth twice more in separate recording sessions, giving a total of three sets of records for each individual moth from each observer (as per Goodenough *et al*. 2010 and also following Lougheed *et al*. 1991 and Yezerinac *et al*. 1992). Observers were unable to check their previous measurements and would have been unlikely to remember a specific measurement for a specific moth across the three recording sessions. The total number of measurements was 1980 (22 moths × 6 biometrics × 5 observers × 3 attempts per observer).

The biometrics included two bilateral traits (*viz*. costal margins and apex angle of the left and right forewings, respectively). To quantify asymmetry in these traits, the absolute difference between the sides was quantified and divided by the mean of the two bilateral measurements. This gave an FA index with trait-size correction at an individual level (known as FA2: Palmer 1994, Palmer & Strobeck 2003). There were 660 asymmetry estimates (22 moths × 2 bilateral traits × 5 observers × 3 attempts per observer).

Statistical analyses

Analysis was undertaken in Microsoft Excel and SPSS 16 for Windows. To allow for multiple analyses being undertaken on non-independent data (different biometrics of the same moth), standard Bonferroni corrections were applied to *p* values by multiplying them by six (as six biometrics were recorded per moth). In all cases, we were comparing multiple measurements by the same observers (intra-observer variability) and different observers (inter-observer variability) on the basis that these should be identical on a per-moth basis (same specimen, same photograph) and deviations from uniformity were because of variability in the measurements.

Baseline variability

To examine the relative variability of different biometrics, the coefficient of variation (CV) was calculated and converted to a percentage (CV = (standard deviation/mean) × 100). This approach is scale independent and thus not influenced by differences in trait size. Use of CV values also allowed direct comparison of digitally-recorded biometrics with those recorded, by the same research team, by manual methods (primarily callipers and stopped rulers) (Goodenough *et al*. 2010). To quantify intra-observer variation, a CV value was calculated for each biometric on a per-moth basis using the three separate measurements made by a given observer. This resulted, for each observer, in a series of 22 CV values for each biometric, which were then averaged to give a biometric-specific value. To quantify inter-observer variation, a CV value was calculated using the mean measurement of each biometric of each moth by the different observers, again resulting in 22 values for each biometric that could be averaged. The same approaches were used to assess both measures of FA. To determine whether measurement variability was related to the size of the moth, biometric-specific CV values were regressed against the size of the trait being measured.

Measurement error

To quantify the relative amount of variation in a variable that was due to measurement error rather than true biological variation, percentage measurement error (%ME) was calculated. This was done as per Bailey and Byrnes (1990), whereby ANOVA was carried out to quantify within- and among-moth components of variance (i.e. the amount of variance derived from measurement variability and biological variability, respectively) for each biometric and FA value. Measurement error was calculated from these variables as: $\%\text{ME} = 100 - [s^2_{\text{within}}/(s^2_{\text{within}} + s^2_{\text{among}}) \times 100]$.This method has been used in previous studies of biometric precision (Lougheed *et al*. 1991, Yezerinac *et al*. 1992, Goodenough *et al*. 2010).

Statistical differences resulting from intra- and inter-observer variability

We followed the method of Goodenough *et al*.

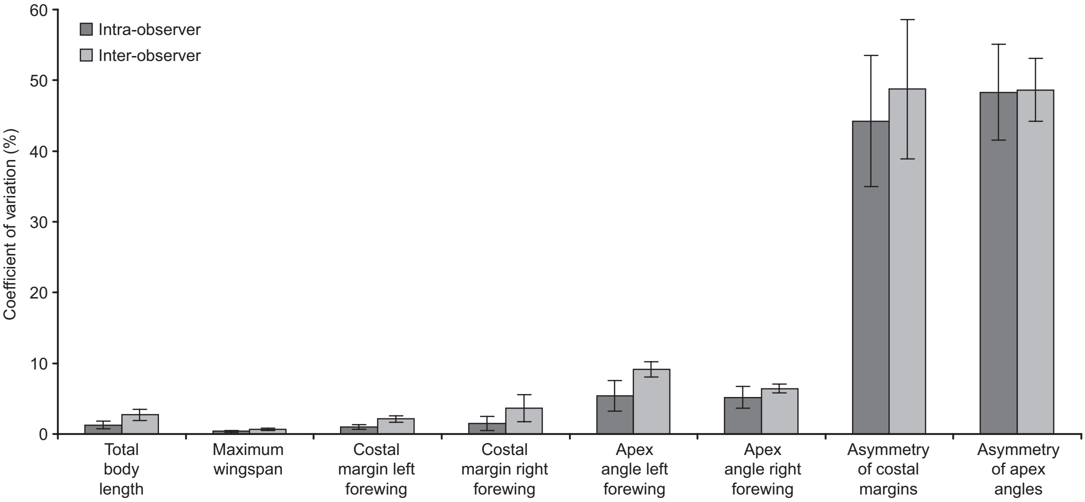

Fig. 2. Mean coefficient of variation (CV, %) values (± SE) for measurements of biometrics and asymmetry values showing intra- and inter-observer variation; *n* = 22 moths measured three times each by five different observers.

(2010) and used two-way repeated measures ANOVA (one for each biometric or FA measure), calculated using the Greenhouse-Geisser method, to quantify the relative importance of intra- and inter-observer variability. To determine the presence of significant differences in biometric measurements as a result of measurement variability, 'observer' (*n* = 5) and 'attempt' (*n* = 3) were defined as fixed factors (all measurements of any one moth should have been identical so significant deviations from uniformity indicated significant effects of variability). The interaction term was also calculated to quantify situations where observers differed in their ability to take consistent measurements. To explore significant interactions further, trends in precision of biometric and FA values between the three recording sessions were calculated by comparing, on a per-moth basis, the deviation between each record that each observer made and the mean of all measurements, from all observers, for that moth. The mean deviation for each observer × attempt combination from the grand mean was then calculated (increasing precision with learning being signified by a decrease in deviance from the grand mean between the recording sessions and *vice versa*).

Results

Baseline variability

Coefficient of variation (CV) values showed some variability in biometric measurements both between repeated measurements by the same observer (mean = 2.4%) and between measurements by different observers (mean = 4.0%) (Fig. 2) and also differed in magnitude between different moth species (Appendix). When FA values were calculated based on these biometrics, the amount of intra- and inter-observer variability increased substantially (mean = 46.2% and 48.68%, respectively), suggesting that numerous small errors in measurements are magnified during the calculation of asymmetry values (Fig. 2). There were significant relationships between moth-specific CV values and trait size, both within and between observers, for most variables (Table 1). All significant relationships were negative, with maximum variability occurring for the smallest moths.

Measurement error

As suggested by the CV values, variability in traits or FA measures that was accounted for by

Table 1. Relationships between variation in repeated measurements of moth traits (coefficient of variation) and trait size. *p* values indicating significant results are set in boldface. All significant correlations were negative (i.e. more variability in measurements of smaller traits).

Measure	Intra-specific variation			Inter-specific variation		
	$F_{1,21}$	p	r^2	$F_{1,108}$	p	r^2
Total body length	6.371	**0.020**	0.242	19.020	< **0.001**	0.150
Maximum wingspan	6.513	**0.019**	0.246	12.535	**0.001**	0.104
Costal margin left forewing	0.101	**0.005**	0.754	5.451	**0.021**	0.048
Costal margin right forewing	7.027	**0.015**	0.260	1.744	0.189	0.016

measurement error (ME), rather than true biological differences, varied between biometrics from negligible to high. Consistent trends were found, with %ME being lowest for linear measurements, moderately higher for curvilinear measurements, and substantially higher for angular measurements (Table 2). Values were always higher inter-specifically than intra-specifically, in some cases by more than an order of magnitude (Table 2). Asymmetry values ranged from 18.18% to 79.07%, with %ME in asymmetry of curvilinear traits being lower than %ME in angular measurements, and intra-specific variability being lower than inter-specific variability in both cases (Table 2). The %ME associated with asymmetry values was always higher than the %ME of the biometrics on which they were based.

Table 2. Percentage of variability in different variables accounted for by Measurement Error (ME) rather than "true" biological differences. Values were determined using the equation %ME = 100 – [$s^2_{\text{within}}/(s^2_{\text{within}} + s^2_{\text{among}})$ × 100] following ANOVA (*see* Methods). "Repeatability" can be calculated from the figures given below by subtracting at %ME value from 100.

Variable/Measure	Measurement error (%ME)	
	Intra-specific	Inter-specific
Linear		
Total body length	0.19	5.06
Maximum wingspan	0.03	0.58
Mean	0.11	2.82
Curvilinear		
Costal margin		
Left forewing	0.21	4.21
Right forewing	2.76	3.95
Mean	1.48	4.08
Angle		
Apex angle		
Left forewing	22.80	78.90
Right forewing	20.10	75.97
Mean	21.45	77.44
Asymmetry		
Asymmetry of		
Costal margins	18.18	50.00
Apex angles	33.33	79.07
Mean	25.76	64.53

Statistical differences resulting from intra- and inter-observer variability

Given the high variability and %ME rates, it was not surprising that there were significant differences between measurements when data were analysed statistically using repeated measures ANOVA (Table 3). Four out of six biometrics differed significantly between observers, with only the two linear measurements (body length and wingspan) being statistically consistent. There was no difference in observers' measurements between the three recording sessions (attempt one, two and three) for any biometric measure, suggesting that, overall, repeated measurements by the same observer were consistent. However, for costal margin measurements of each forewing, there was a significant interaction between 'observer' and 'attempt' (Table 3) suggesting that observers differed in their ability to take consistent measurements for these variables. On exploring the interaction data more, it transpired that, in both cases, one (but not the same) observer improved substantially over the course of the three recording sessions (left costal margin = observer 3 improved, right costal margin = observer 4 improved), while all other observers were consistent throughout (Fig. 3). In the case of the right costal margin, not only did observer 4 improve over the recording sessions contrary to no change for all other observers, but their mean deviance was substantially higher than for the other observers (Fig. 3). The fact that precision improved over the recording sessions for a different observer on each occasion suggests that this interaction was not the result of one atypical person, meaning it would be hard to factor out of analysis on a per-observer basis.

Both asymmetry measurements, although highly variable, did not differ significantly between observers or attempts (Table 3). Indeed it was probably the very high variability that reduced the chances of finding significant differences between groups (effectively a Type II error).

Discussion

Variability and measurement error

Research conclusions based on analyses of biometric data are only sound if the biometrics themselves are accurate and not confounded by measurement error. This study suggests that even when measurements are derived from analysis of photographs — often regarded as best practice, particularly for small study organisms or live specimens (Hill *et al*. 2005, Sefcek & King 2007, Davis *et al*. 2008) — ME can still be important.

Overall, and as expected, %ME was always lower intra-specifically than inter-specifically; and indeed variability only resulted in significant measurement differences between (and not within) observers. This is consistent with work in ecology on physical measurements of specimens (e.g. Palmeirim *et al*. 1998) and in the biomedical sciences on photographically-determined biometrics (e.g. Perni *et al*. 2004, Masters *et al*. 2005, Ibekwea *et al*. 2009). The amount of variability associated with each biometric measurement differed both between specimens (trait size) and between measurement type (linear, curvilinear, or angular).

Relationship between variability magnitude and trait size

The variability associated with measurement of the same trait was inversely related to trait size, such that most variability was associated with smaller specimens, even when variability itself was calculated to be size-independent (using the CV measure; *see* Methods). This agrees with previous studies (e.g. Pankakoski *et al*.1987, Yezerinac *et al*. 1992, Palmeirim 1998) that found significantly higher relative variability in smaller skeletal traits of small mammals and songbirds — the 'bigger is better' effect (Jamison & Ward 1993). However, other studies (e.g. Lougheed *et al*. 1991, Goodenough *et al*.

Table 3. Fully-factorial, two-way, repeated measures ANOVA results for biometrics and asymmetry values. The Greenhouse-Geisser method was used to compensate for sphericity and standard Bonferroni corrections were applied to significance values to allow for family-wise error (*see* Methods). The reason for significant interactions between observer and attempt was usually that some observers improved their ability to take precise measurements during the course of the study (i.e. between recording sessions) while others remained consistent. *p* values indicating significant results are set in boldface.

Measure	Observer		Attempt		Observer × Attempt	
	F	*p*	*F*	*p*	*F*	*p*
Total body length	1.066	0.318	1.924	0.166	0.483	0.745
Maximum wingspan	2.159	0.156	0.762	0.396	0.772	0.395
Costal margin left forewing	3.326	**0.042**	0.006	0.816	5.302	< **0.001**
Costal margin right forewing	5.126	**0.009**	1.206	0.308	2.755	**0.029**
Apex angle left forewing	9.718	< **0.001**	2.102	0.146	3.691	0.111
Apex angle right forewing	9.402	< **0.001**	0.573	0.534	2.043	0.106
Asymmetry of costal margins	1.674	0.210	1.865	0.185	1.144	0.309
Asymmetry of apex angles	0.497	0.666	0.253	0.727	1.469	0.222

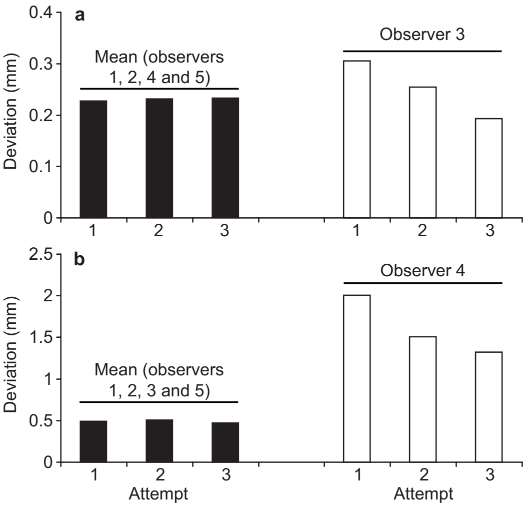

Fig. 3. Mean deviation (mm) of costal margin measurements of the (**a**) left and (**b**) right forewings relative to the grand mean. The white bars show the deviation for a single observer that improved over the course of the recording sessions, and the black bars show the deviation for all other observers combined.

2010) have not found this relationship, and have instead concluded that variability levels are independent of trait size. Given that the studies to find a relationship have all been measuring small variables (small skeletal features of small mammals and birds) or insects (this study), whereas the studies that have found no relationship have focused upon measurements of comparatively large variables (external biometrics of birds), we postulate that the relationship might only be present, or strong enough to have a significant influence, when the biometrics being measured are, on average, small (i.e. when "small" variables are very small in absolute terms, not just relative to the mean).

Differences in biometrics susceptibility to variability

Traits measureable on a photograph using a single straight line had very low variability and were never associated with significant differences within or between observers. This is probably because linear measurements are both easy to take (two clicks of a mouse in the software used, one on each terminal landmark) and, as long as terminal landmarks are obvious, objective (Bailey & Byrnes 1990, von Cramon-Taubadel *et al.* 2007). The importance of terminal landmarks being distinct is underlined by the fact that wingspan measurements were much less variable than body length measurements, both intra- and inter-specifically. This is likely because wings have clear and unambiguous wingtips, whereas the terminal posterior landmark for body length (the tip of the abdomen) is confused by the presence of posterior structures, such as genitalia and dorsal hairs, on many specimens (e.g. Fig. 1). The importance of using definitive landmarks to reduce observer variability has been noted before for physical measurements, for example for mammalian skeletal variables and avian biometrics (Palmeirim *et al.* 1998, Goodenough *et al.* 2010), and our findings suggest that this is also true for digital biometrics.

Curvilinear traits (i.e. those with a distance only measurable by summing a series of mini-linear measurements) were subject to moderately more variability. Unlike the straightforward linear measurements, where observer effort is *de facto* consistent, curvilinear measurements require creation of pseudo-landmarks for every direction change, such that the number of pseudo-landmarks (and associated mouse clicks) differs between images depending on the shape being measured (in this case, wing edges). This not only means that curvilinear measurements are more subjective, but also that effort can (and likely does) differ between observers depending how conscientiously small direction changes are plotted. These factors probably explain why: (1) curvilinear measurements are more somewhat variable than linear ones intra-specifically; and (2) why observers differ significantly (ANOVA results), although not substantially (%ME data), in measuring such traits.

The %ME estimates for all distance (linear and curvilinear) measurements made here using photographic analysis were substantially lower than those generated, by the same research team, when taking physical measurements of museum bird specimens (moth photographs = 2.4% intra-specifically and 4.0% inter-specifically; *versus* 5.1% and 7.1%, for birds, respectively: Goodenough *et al.* 2010). Indeed, all distance measurements on the moths had CV values under 5% and would thus be deemed to have low enough error rates for their use in research to be deemed appropriate (White & Folkens 2000).

Whereas linear and curvilinear variables were, overall, fairly consistent (%ME below ~5%), angular measurements were extremely variable (mean = 49%). Again, this likely resulted from a lack of definite landmarks. A few specimens had angular wingtips (*see* Fig. 1), making it comparatively easy to measure the wing angle, but most had rounded wings, which meant that decisions on where to start and finish the "angle" were more subjective. This can be a common issue where shape, rather than size, is being measured on biological specimens (Zelditch *et al*. 2004)

Use of measures of asymmetry

Numerous small errors in measuring biometrics themselves do not simply cancel one another out during calculation of FA (as they do when PC1 is calculated using PCA: Lougheed *et al*. 1991), but instead become magnified. The ME of FA values calculated here ranged from 18% to 79%, with asymmetry of distance traits being lower than asymmetry of angular traits (as expected given the difference in %ME of the underlying measurements). Although this is lower than for physical measurements (ME was 84%–91% for birds measured by hand by the same research team: Goodenough *et al*. 2010) this is still very high given that levels of FA are usually small, (around 1%–2% of mean trait size: van Dongen 2006). The potential for actual FA to be confounded with, or masked by, ME (Palmer & Strobeck 1986) is, therefore, substantial.

Conclusions and recommendations

Our findings suggest that biometrics measured from digital images on a computer tend to be less variable within and between observers than measurements taken by hand. Our results also suggest that inclusion of multi-observer data, especially for linear and curvilinear measurements, is unlikely to cause a type I error. However, studies using photographically-derived biometric measurements will have lower variability if all measurements are taken by the same observer, such that the risk of a Type II error will be reduced. The potential for residual ME can be minimised, where possible, by: (1) using linear measurements rather than curvilinear ones; (2) measuring larger traits rather than smaller ones (given the positive relationship between measurement accuracy and trait size); and (3) not using angle measurements unless the trait has a clear and unambiguous corner. As measures of asymmetry are associated with very high measurement errors, even when traits are measured from photographs rather than by hand, their use should be undertaken with extreme caution, particularly when multi-observer data are used. We recommend that studies that do use FA should calculate measures that account for %ME (such as the measurements FA10, which describes the difference between sides after ME has been factored out: Palmer 1994, Palmer & Strobeck 2003, Bechshøft *et al*. 2008).

We have shown that taking biometrics digitally using analysis of photographs is a technique that can result in consistent measurements within and between observers, and should thus continue to be recommended. However, given that traits differ in their susceptibility to error, the likely consistency of measurements should be considered when deciding which biometrics are most suitable in any given situation.

References

Abramoff, M. D., Magalhaes, P. J. & Ram, S. J. 2004: Image processing with ImageJ. — *Biophotonics International* 11: 36–42.

Adam, C. J., Izatt, M. T., Harvey, J. R. & Askin, G. N. 2005: Variability in Cobb angle measurements using reformatted computed tomography scans. — *Spine* 30: 1664–1669.

Andersson, M. B. 1994: *Sexual selection*. — Princeton University Press, Princeton.

Ashton, K. G. 2002: Patterns of within-species body size variation of birds: strong evidence for Bergmann's rule. — *Global Ecology & Biogeography* 11: 505–523.

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.

Björklund, M. 1996: The effect of male presence on nestling growth and fluctuating asymmetry in the blue tit. — *Condor* 98: 172–175.

Bland, J. M. & Altman, D. G. 1987: Statistical methods for assessing agreement between measurement. — *Biochimica Clinica* 11: 399–404.

Davis, A. K., Connell, L. L., Grosse, A. & Maerz, J. C. 2008: A fast, non-invasive method of measuring growth in tadpoles using image analysis. — *Herpetological Review* 39: 56–58.

Debat, V., Alibert, P. & David, P. 2000: Independence between developmental stability and canalisation in the skull of the house mouse. *Proceedings of the Royal Society of London B* 267: 423–430.

Faurby, S., Kjærsgaard, A., Pertoldi, C. & Loeschcke, V. 2005: The effect of maternal and grandmaternal age in benign and high temperature environments. — *Experimental Gerontology* 40: 988–996.

Faurby, S., Nielsen, K. S. K., Bussarawit, S., Intanaid, I., van Conge, N., Pertoldi, C. & Funch, P. 2011: Intraspecific shape variation in horseshoe crabs: the importance of sexual and natural selection for local adaptation. — *Journal of Experimental Marine Biology and Ecology* 407: 131–138.

Gage, M. J. G. 1998: Influences of sex, size, and symmetry on ejaculate expenditure in a moth. — *Behavioral Ecology* 9: 592–597.

Gidaszewski, N. A., Baylac, M. & Klingenberg, C. P. 2009: Evolution of sexual dimorphism of wing shape in the *Drosophila melanogaster* subgroup. — *BMC Evolutionary Biology* 9: 110–121.

Goodenough, A. E., Stafford, R., Catlin-Groves, C. L., Smith, A. L. & Hart, A. G. 2010: Variation in measurement of animal biometrics within and between observers and its influence on accurate quantification of common biometric-based condition indices. — *Annales Zoologici Fennici* 47: 323–334.

Gosler, A. G. 2004: Birds in the hand. — In: Sutherland, W. J. (ed.), *Bird ecology and conservation: a handbook of techniques*: 85–118. Oxford University Press, Oxford.

Hassall, C. & Thompson, D. J. 2009: Variation in wing spot size and asymmetry of the banded demoiselle *Calopteryx splendens* (Harris, 1780). — *Journal of the British Dragonfly Society* 25: 7–15.

Hassall, C., Thompson, D. J. & Harvey, I. F. 2008: Wings of *Coenagrion puella* vary in shape at the northern range margin (Odonata: Coenagrionidae). — *International Journal of Odonatology* 11: 35–41.

Heathcote, G. M. 1981: The magnitude and consequences of measurement error in human craniometry. — *Canadian Review of Physical Anthropology* 3: 18–40.

Hill, M. G., Mauchline, N. A., Cate, L. R. & Connolly, P. G. 2005: A technique for measuring growth rate and survival of armoured scale insects. — *New Zealand Plant Protection* 58: 288–293.

Hoffmann, A. A. & Parsons, P. A. 1991: *Evolutionary genetics and environmental stress*. — Oxford University Press, Oxford.

Ibekwea, T. S., Adeosuna, A. A. & Nwaorgua, O. G. 2009: Quantitative analysis of tympanic membrane perforation: a simple and reliable method. — *The Journal of Laryngology & Otology* 123(1): e2, doi: 10.1017/S0022215108003800.

Jamison, P. L. & Ward, R. E. 1993: Measurement size, precision, and reliability in craniofacial anthropometry: bigger is better. — *American Journal of Physical Anthropology* 90: 495–500.

Kingsolver, J. G. & Pfennig, D. W. 2004: Individual-level selection as a cause of Cope's rule of phyletic size increase. — *Evolution* 58: 1608–1612.

Loeschcke, V., Bundgaard, J. & Barker, J. S. F. 1999: Reaction norms across and genetic parameters at different temperatures for thorax and wing size traits in *Drosophila aldrichi* and *D. buzzatii*. — *Journal of Evolutionary Biology* 12: 605–623.

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.

Masters, I. B., Eastburn, M. M., Wootton, R., Ware, R. S., Francis, P. W., Zimmerman, P. V. & Chang, A. B. 2005: A new method for objective identification and measurement of airway lumen in paediatric flexible videobronchoscopy. — *Thorax* 60: 652–658.

Molina-Borja, M. & Rodríguez-Domínguez, M. A. 2004: Evolution of biometric and life-history traits in lizards (*Gallotia*) from the Canary Islands. — *Journal of Zoological Systematics and Evolutionary Research* 42: 44–53.

Mutanen, M. & Kaitala, A. 2006: Genital variation in a dimorphic moth *Selenia tetralunaria* (Lepidoptera, Geometridae). — *Biological Journal of the Linnean Society* 87: 297–307.

Nisbet, I. C., Baird, J., Howard, D. V. & Anderson, K. S. 1970: Statistical comparison of wing-length measured by four observers. — *Bird-Banding* 41: 307–308.

Nowak, R. M. 2002: The original status of wolves in eastern North America. — *Southeastern Naturalist* 1: 95–130.

Palmeirim, J. M. 1998: Analysis of skull measurements and measurers: can we use data obtained by various observers? — *Journal of Mammalogy* 79: 1021–1028.

Palmer, A. R. & Strobeck, C. 1986: Fluctuating asymmetry: measurement, analysis, patterns. — *Annual Review of Ecology and Systematics* 17: 391–421.

Palmer, A. R. & Strobeck, C. 2003: Fluctuating asymmetry analyses revisited. — In: Polak, M. (ed.), *Developmental instability (DI): causes and consequences*: 279–319. Oxford University Press, New York, NY.

Palmer, A. R. 1994: Fluctuating asymmetry analyses: a primer. — In: Markow, T. A. (ed.), *Developmental instability: its origins and evolutionary implications*: 335–364. Kluwer, Dordrecht.

Pankakoski, E., Vaisanen, R. A. & Nurmi, K. 1987: Variability of muskrat skulls: measurement error, environmental modification and size allometry. — *Systematic Zoology* 36: 35–51.

Parsons, P. A. 1992: Fluctuating asymmetry: a biological monitor of environmental and genomic stress. — *Heredity* 68: 361–364.

Perni, S. C., Chervenak, F. A., Kalish, R. B., Magherini-Rothe, S., Predanic, M., Streltzoff, J. & Skupski, D. W.

2004: Intraobserver and interobserver reproducibility of fetal biometry. — *Ultrasound in Obstetrics and Gynecology* 24: 654–658.

Sefcek, J. A. & King, J. E. 2007: Chimpanzee facial symmetry: a biometric measure of chimpanzee health. — *American Journal of Primatology* 69: 1257–1263.

Smith, F. A., Brown, J. H., Haskell, J. P., Lyons, S. K., Alroy, J., Charnov, E. L., Dayan, T., Enquist, B. J., Ernest, S. K. M., Hadly, E. A., Jablonski, D., Jones, K. E., Kaufman, D. M., Marquet, P. A., Maure, B. A., Niklas, K. J., Porter, W. P., Roy, K., Tiffney, B. & Willig, M. R. 2004: Similarity of mammalian body size across the taxonomic hierarchy and across space and time. — *American Naturalist* 163: 5.

van Dongen, S. 2006: Fluctuating asymmetry and developmental instability in evolutionary biology: past, present and future. — *Journal of Evolutionary Biology* 19: 1727–1743.

Vilisics, F., Sólymos, P. & Hornung, E. 2005: Measuring fluctuating asymmetry of the terrestrial isopod *Trachelipus rathkii* (Crustacea: Isopoda, Oniscidea). — *European Journal of Soil Biology* 41: 85–90.

White, T. D. & Folkens, P. A. 2000: *Human osteology*. — Academic Press, San Diego.

Yezerinac, S. M., Lougheed, S. C. & Handford, P. 1992: Measurement error and morphometric studies: statistical power and observer experience. — *Systematic Biology* 41: 471–482.

Zelditch, M. L., Swiderski, D. L., Sheets, H. D. & Fink, W. L. 2004: *Geometric morphometrics for biologists: a primer*. — Elsevier Academic Press, Oxford.

Appendix. Intra- and inter-specific variation in measurements on a per-month basis [shown using coefficient of variation (%)].

Species	Intra-observer variation						
	Total body length	Maximum wingspan	Costal margin left forewing	Costal margin right forewing	Apex angle left forewing	Apex angle right forewing	Mean
Broad-bordered yellow underwing	1.08	0.14	0.96	8.05	3.43	6.05	3.28
Large yellow underwing	1.22	0.20	0.74	0.77	4.11	4.02	1.84
Peppered moth	1.25	0.27	0.78	0.90	7.70	4.08	2.50
Common shark	0.97	0.30	0.70	0.59	15.34	5.03	3.82
White ermine	0.53	0.19	0.75	1.10	5.84	8.14	2.76
Garden tiger	0.63	0.15	0.96	0.77	9.56	5.44	2.92
Cream spot tiger	0.90	0.30	0.58	0.46	2.55	3.30	1.35
Large footman	0.79	0.24	0.39	0.58	1.67	8.19	1.98
Common footman	0.63	0.29	0.66	1.19	5.56	4.88	2.20
Dewdrop footman	1.17	0.21	0.77	0.84	6.71	4.17	2.31
American wainscot	0.87	0.12	0.56	0.37	3.91	2.45	1.38
Ochreous drab	0.99	0.40	1.23	1.85	4.57	3.29	2.06
Shore wainscot	0.57	0.26	0.45	0.67	5.55	3.20	1.78
Cloudy drab	1.36	0.21	1.27	0.84	5.27	3.44	2.06
Unidentified species 1	4.42	0.55	2.64	3.17	6.51	9.39	4.45
Unidentified species 2	4.27	0.88	0.60	0.77	4.56	4.09	2.53
Unidentified species 3	1.56	0.14	0.51	0.78	4.25	4.96	2.04
Unidentified species 4	0.71	0.36	1.12	1.00	3.49	4.13	1.80
Unidentified species 5	0.55	0.57	0.59	1.07	6.43	7.24	2.74
Unidentified species 6	0.74	0.37	1.47	1.95	3.36	8.21	2.68
Unidentified species 7	0.60	0.38	0.98	1.36	4.08	4.18	1.93
Unidentified species 8	0.54	0.36	0.98	1.65	2.96	4.31	1.80

continued

Appendix. Continued.

	Inter-observer variation						
	Total body length	Maximum wingspan	Costal margin left forewing	Costal margin right forewing	Apex angle left forewing	Apex angle right forewing	Mean
Broad-bordered yellow underwing	2.12	0.17	1.18	13.07	1.62	9.52	4.61
Large yellow underwing	2.95	0.40	2.50	2.36	1.99	4.64	2.47
Peppered moth	2.89	0.29	1.71	1.13	20.76	6.24	5.50
Common shark	3.43	0.24	0.95	0.93	11.55	2.99	3.35
White ermine	0.65	0.20	1.65	1.06	11.84	6.41	3.63
Garden tiger	1.69	0.48	1.09	1.12	8.31	5.60	3.05
Cream spot tiger	1.11	0.29	10.67	11.45	7.05	4.13	10.79
Large footman	1.06	0.29	0.64	0.51	6.96	5.92	2.56
Common footman	0.58	0.51	1.90	1.30	12.63	7.50	4.07
Dewdrop footman	2.04	0.06	1.10	1.11	15.53	3.76	3.93
American wainscot	5.00	0.13	1.38	1.05	3.54	4.46	2.59
Ochreous drab	1.34	0.55	1.42	1.48	8.34	7.31	3.41
Shore wainscot	0.48	0.25	1.92	0.59	9.73	8.16	3.52
Cloudy drab	1.44	0.27	1.41	0.38	8.12	3.13	2.46
Unidentified species 1	17.57	1.03	4.98	3.16	20.09	15.36	10.36
Unidentified species 2	5.06	4.19	2.13	1.32	3.02	3.40	3.19
Unidentified species 3	1.37	0.15	0.96	0.77	5.70	7.70	2.78
Unidentified species 4	3.32	0.50	2.67	0.55	9.31	8.73	4.18
Unidentified species 5	0.91	0.64	1.03	0.69	8.12	11.72	3.85
Unidentified species 6	1.24	0.82	1.57	2.05	11.09	4.21	3.50
Unidentified species 7	0.90	0.64	1.04	1.74	4.36	4.62	2.22
Unidentified species 8	0.56	0.75	1.37	0.87	9.15	4.03	2.79

This article is also available in pdf format at http://www.annzool.net/