Tsuboi, M, Kopperud, BT, Syrowatka, C, Grabowski, M, Voje, KL, Pelabon, C and Hansen, TF

 Measuring Complex Morphological Traits with 3D Photogrammetry: A Case Study with Deer Antlers

http://researchonline.ljmu.ac.uk/id/eprint/14390/

Article

TOOLS AND TECHNIQUES

Measuring Complex Morphological Traits with 3D Photogrammetry: A Case Study with Deer Antlers

Masahito Tsuboi[1,2] · Bjørn Tore Kopperud[3] · Christine Syrowatka[1] · Mark Grabowski[4] · Kjetil L. Voje[1] · Christophe Pélabon[5] · Thomas F. Hansen[1]

Abstract

The increasing availability of 3D-imaging technology provides new opportunities for measuring morphology. Photogrammetry enables easy 3D-data acquisition compared to conventional methods and here we assess its accuracy for measuring the size of deer antlers, a complex morphological structure. Using a proprietary photogrammetry software, we generated 3D images of antlers for 92 individuals from 29 species of cervids that vary widely in antler size and shape and used these to measure antler volume. By repeating the process, we found that the relative error averaged 8.5% of object size. Errors in converting arbitrary voxel units into real volumetric units accounted for 70% of the measurement variance and can therefore be reduced by replicating the conversion. We applied the method to clay models of known volume and found no indication of bias. The estimation was robust against variation in imaging device, distance and operator, but approximately 40 images per specimen were necessary to achieve good precision. We used the method to show that conventional measures of main-beam length are relatively poor estimators of antler volume. Using loose antlers of known weight, we also showed that the volume may be a relatively poor predictor of antler weight due to variation in bone density across species. We conclude that photogrammetry can be an efficient and accurate tool for measuring antlers, and likely many other complex morphological traits.

✉ Masahito Tsuboi
Masahito.Tsuboi@ibv.uio.no

1 Department of Biology, University of Oslo, Blindernveien 31, 0371 Oslo, Norway

2 School of Advanced Sciences, SOKENDAI, Shonan Village, Miuragun, Hayama, Kanagawa 2400193, Japan

3 Natural History Museum, University of Oslo, Blindern, PO Box 1172, 0318 Oslo, Norway

4 Research Centre in Evolutionary Anthropology and Palaeoecology, Liverpool John Moores University, Liverpool L3 3AF, UK

5 Department of Biology, Centre for Biodiversity Dynamics, NTNU, Norwegian University of Science and Technology, 7491 Trondheim, Norway

Introduction

The past decades have seen progress in morphometrics and "phenomics" due to advances in multivariate and high-dimensional statistics, computer-based reconstruction, visualization, and data storage, and in increased awareness of scaling, transformation and measurement errors (Bookstein 1991; Richtsmeier et al. 2005; Mitteroecker and Gunz 2009; Houle et al. 2010, 2011; Lawing and Polly 2010; Zelditch et al. 2012). Increasing availability of 3D-imaging technology has facilitated the study of morphological traits poorly represented by 2D data (Cardini 2014; Buser et al. 2018), but most current 3D technology, including micro-computer tomography and microscribes, are expensive, time consuming, and cumbersome to use.

Photogrammetry is a cheap and simple method to obtain 3D data from sets of 2D digital photographs (Falkingham 2012; Katz and Friess 2014). It has been successfully used to estimating body size in the field for a variety of species (Breuer et al. 2007; Waite et al. 2007; Chiari et al. 2008; de Bruyn et al. 2009; Postma et al. 2015; Beltran et al. 2018; see also

Christiansen et al. 2019) or shape of marine colonial organisms (Lavy et al. 2015; Gutierrez-Heredia et al. 2016; Roth et al. 2019). Under lab conditions, photogrammetry has proved as accurate as manual measurements in generating landmark-based 3D morphometric data from skulls of small mammals (Munoz-Munoz et al. 2016; Giacomini et al. 2019). In this study, we evaluate its utility for measuring the size of deer antlers, a morphologically complex and variable structure that is difficult to measure consistently with conventional techniques.

Cervid antlers have long been of interest to biologists (Huxley 1931; Geist 1998; Emlen 2008). They have been used in debates about orthogenesis and heterochrony (e.g. Simpson 1953; Gould 1977), allometry (Huxley 1931, 1932; Gould 1973, 1974; Lemaître et al. 2014; Ceacero 2016), sexual selection (Clutton-Brock et al. 1980; Pélabon and Joly 2000; Plard et al. 2011; Bartoszek et al. 2012; Holman and Bro-Jørgensen 2016), quantitative genetics in the wild (Kruuk et al. 2002), growth and regeneration (Moen et al. 1999; Price and Allen 2004) and to assess stress and condition (Lagesen and Folstad 1998; Pélabon and van Breukelen 1998; Kruuk et al. 2003; Mysterud et al. 2005; Mills and Peterson 2013). These studies have used a variety of antler measurements including antler span or height, length of the main-beam, the number of tines, as well as weight and volume. These measures vary widely in accuracy and biological meaning, and they are often chosen more for convenience than from principled argument as to their theoretical relevance or statistical properties.

Here we assess the accuracy of antler volume estimates obtained from photogrammetry by repeated measures and clay models of known volume. We then explore the relationship between antler length, volume, and mass across antlers of different size and shape.

Theory: Quantifying Measurement Error

Measurement error is the deviation of a measurement from the true value of the measured entity and consists of two components, bias and imprecision, corresponding to systematic and random differences between the measured and true values. These combine into an overall inaccuracy, which is most conveniently assessed as the expected squared deviation of the measurement from the true value. Formally we may quantify these entities as follows:

$$\text{Inaccuracy} = \text{E}\left[(m - x)^2\right],$$

$$\text{Imprecision} = \text{Var}[m],$$

$$\text{Bias} = \text{E}[m] - x,$$

where m is the measurement or statistics in question, x is the true value, and E and Var denote expectation and variance. These entities combine additively as

$$\text{Inaccuracy} = \text{Imprecision} + \text{Bias}^2$$

Note that a precise estimate may be inaccurate if biased, and that an unbiased estimate may be inaccurate if imprecise.

The imprecision of a measurement procedure can be evaluated by use of repeated measures. Assuming we have taken two independent repeated measurements of the same entity, the measurement variance can be estimated as half the variance of the difference between the two measurements m_1 and m_2 as:

$$\sigma_m^2 = \frac{\text{Var}[m_1 - m_2]}{2}$$

The measurement variance σ_m^2 is the entity that is normally used to correct for measurement imprecision in statistical models (e.g. Fuller 1987; Buonaccorsi 2010; Hansen and Bartoszek 2012; Morrissey 2016; Ponzi et al. 2018).

Imprecision can also be scored as a relative error

$$RE = \frac{|m_1 - m_2|}{\overline{m}}$$

where $\overline{m}$ is the average over the two repeated measurements. The expected relative error is a numerically intuitive measure of imprecision, but can not be combined additively with bias as above. Assuming normally distributed and homoscedastic errors, the expected relative error is approximately related to the measurement variance as

$$E[RE] \approx \frac{2}{\sqrt{\pi}} \frac{\sigma_m}{\overline{m}}$$

Biologists often express precision as a repeatability, also known as the intraclass correlation coefficient, which in our notation would be $1 - \sigma_m^2/\sigma^2$, where σ^2 is the total variance of the sample (e.g. Wolak et al. 2012). Because they depend on the variance of the specific sample, repeatabilities cannot be used as general measures of the imprecision of a measurement procedure. They are strictly measures of the relative impact of measurement imprecision in a specific analysis.

When measurements of different traits are taken from the same photogrammetry object, their errors may be correlated, and this needs to be taken into account in assessing the measurement variance of sums, averages or differences between traits. Just as measurement variance can be computed from repeated measures of the same specimen, the measurement covariance can be computed as

$$c_m = \frac{Cov\left[m_1 - m_2, n_1 - n_2\right]}{2},$$

where m_1 and n_1 are measures of the two traits from the first object, and m_2 and n_2 are measures from the second object assumed to be independent from the first ones.

In contrast to precision, bias can not be assessed with repeated measures. Estimation of bias, and thereby accuracy, requires information about the true value, which is rarely available. In practice bias may be studied by comparison of distinct methods of measurement, which yields the bias of one method relative to another. For example, the bias of a new method may be assessed by comparison to a known well-verified procedure.

Materials and Methods

Samples and Skull Measurements

The specimens used in this study are stored at the following museums/institutions: the American Museum of Natural History, New York, USA (AMNH), the National Museum of Natural History, Washington D.C., USA (NMNH), the Natural History Museum, London, UK (NHMUK), the Natural History Museum, Vienna, Austria (NHMW), the Swedish Museum of Natural History, Stockholm, Sweden (NHRM), and University of Oslo (UiO). We evaluated the age of specimens by the degree of tooth wear at the upper molars following Brown and Chapman (1990), and only prime-aged adult males were included.

Our main data set is based on 92 skulls across 29 species. Species were assigned to one of three antler shape categories as follows; "palmated" for *Dama dama* and *Alces alces*, "bifurcated" for *Blastocerus dichotomus*, *Elaphurus davidianus*, *Rucervus schomburgki* and *Ozotoceros bezoarticus*, and "main-beamed" for the rest (see Table S1 for the list of specimens). The main-beam lengths of left and right antlers were measured with measuring tape on all specimens by one investigator (MT) to the nearest millimeter along the outer curve from the margin of the burr to the tip of the longest tine. Repeated measures of beam length were made on a subset of 47 specimens with at least 1 day between the measurements.

To analyze the relationship between the estimated volume and physical estimates of weight we used 71 loose antlers from 14 species. These antlers were either naturally shed or sawed off at the burr (see Table S1 for the list of specimens). Each antler was weighted to the nearest gram using a digital balance.
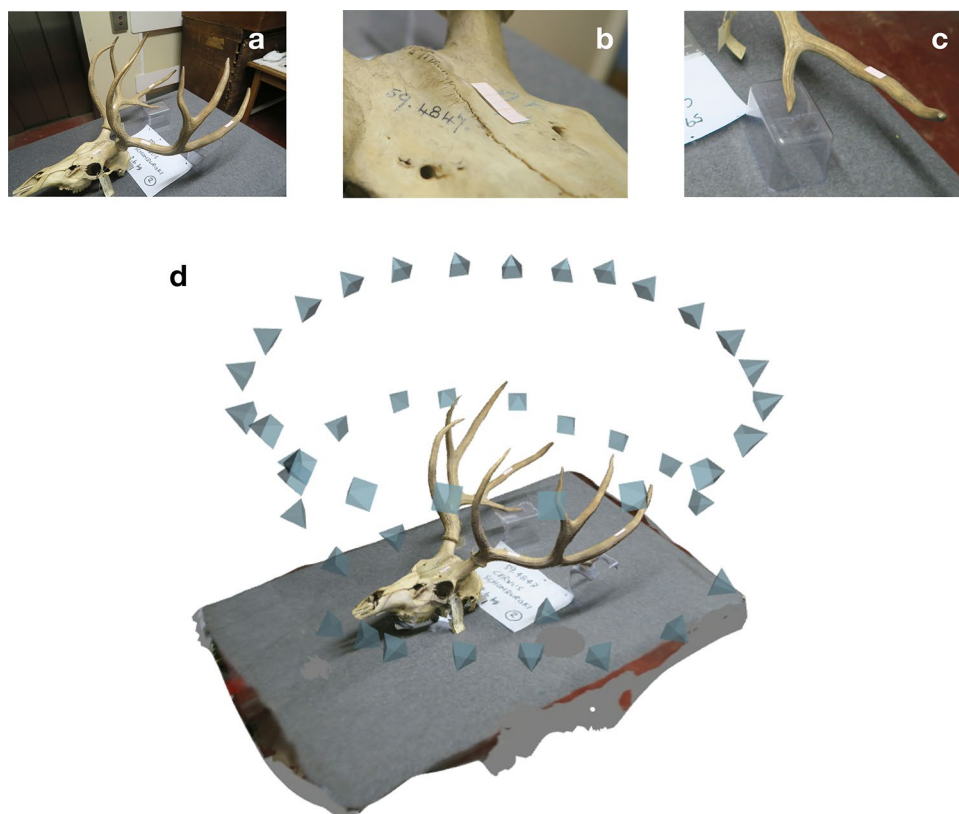
Imaging

We used the educational license of RECAP PHOTO® (Autodesk Inc. 2017) to conduct image processing. The educational license was free of charge as of 1st October 2019. For our standard protocol, two-dimensional digital images were obtained with a 20.2 megapixels Canon G7X camera equipped with 8.8–36.8 mm focal length and maximum aperture of f/1.8–2.8 lens at optical settings suitable for local light conditions and with minimum zoom. Each skull was placed at the center of a table covered by a dark grey fleece cloth to avoid damaging the skull and to minimize light reflection. We placed three 10×30 mm paper scales on the specimen: one on the skull between the two pedicels, and one each on the right and left main-beam just above the burr (Fig. 1a).

Because Beltran et al. (2018) found that a lack of images from the ventral side could bias photogrammetry estimates of body size, we elevated the antlers from the table by two transparent plastic boxes to obtain images from the ventral side of the antlers. The photographer moved around the skull while taking photographs at regular intervals: one set of photos was taken with a diagonally downward view to the skull, and another from a horizontal view. In total, approximately 50 photographs were taken per specimen (Fig. 1). The distance between the camera and the skull was kept roughly constant while ensuring that the complete skull was visible on each picture. For each specimen, two complete sets of photographs were obtained from two sessions separated by at least 1 day. During each session, we replicated every step of the protocol: placing the skull on the table and on the transparent boxes, placing the scale bars, and photographing. These repeated measurements resulted in a total of 184 photosets (i.e. two repeated sets for all 92 specimens). Three observers (CS, MG, MT) took the photographs, but since there were no evidence of an observer effect (see Supplementary Material), we pooled all data.

Each set of images was uploaded to the cloud server of RECAP PHOTO® to create 3D photogrammetry objects. Using the software, we removed all parts of the object except the antlers (i.e. we removed the background and the skull) and saved the left and the right antler as separate objects. We filled the holes created by removed parts with flat surface, and we removed isolated particles and mesh intersections with the built-in algorithm of the program. Subsequently, we scaled the object using the paper scales at the root of the main beam, and we measured the volume of each separate antler in cubic millimeters (mm^3). In 17 cases, we used the paper scales placed on the skull because the antlers were too small to support the scale. All image processing was performed by MT.

Fig. 1 Illustration of the setup of the photogrammetry protocol. **a** Photography needs to include the entire structure of interest, **b** scales are placed on the object, **c** antlers are lifted from the table using plastic boxes to enable photography from the ventral side, **d** a 3D model of the Schomburk's deer *Rucervus schomburgki* (NHMUK 59.4847), a species that is now extinct, created from 49 photographs. The pyramid icons show the location from which photographs were taken

Measuring Imprecision and Bias

We examined four sources of imprecision in our protocol: (1) taking photos, (2) image rendering, (3) image processing and (4) scaling. To evaluate the magnitude of error from each of these, we repeatedly measured 3D objects constructed from independent photosets, independent rendering, independent image processing, and independent scaling. Because volume can only be obtained after image processing and scaling are performed, we retrospectively estimated the error variance related to each of these sources by subtracting the variances in downstream steps from the overall measurement variance assuming additivity. During the process we realized that the relative error associated with rendering and image processing were substantially smaller than those due to scaling. We thus evaluated errors for a subset of n = 69 specimens for rendering and n = 21 specimens for image processing. In addition, the 17 specimens that were too small to support scale bars on their antlers were not repeatedly measured with respect to scaling.

To evaluate the influence of imaging devices and imaging conditions we examined the effect of (i) the digital camera used for photography, (ii) the number of photos taken to construct an object, and (iii) the distance between the specimen and the camera. Our control protocol used 60 photos taken by a G7X camera from a standard distance (i.e. the nearest distance from which the entire specimen could be photographed). We tested two additional photographing devices: an 8.4 megapixels Panasonic DMC-LX1 camera equipped with 34–136 mm focal length and maximum aperture of f/2.8–4.9 lens and an iPhone 6 equipped with an 8 megapixels camera with maximum aperture of f/2.2 lens. To evaluate the effect of distance from the camera to the object, we took 60 photos using the G7X camera from twice the standard distance. To investigate the effect of the number of photos, we took independent photosets with 20, 30, 40, 50, or 60 photographs by the G7X camera from both standard and twice the standard distance. Altogether, 14 alternative photographing settings were applied to a subset of samples (i.e. 9 specimens at the University of Oslo), generating a total of 126 photogrammetry objects. All photosets in this experiment were taken by MT.

To assess bias, we tested the procedure on clay models of known volume. Using modeling dough with a density of 1.2 g ml^{-1}, we constructed 10 differently-sized models of single-beamed and bifurcated antlers. With a digital balance, we prepared blocks of 100–1000 g of dough by steps of 100 g, and built antler models with them. The models were then mounted, photographed, rendered and measured by one observer (MT) with the same protocol as used for the real antlers. We used 40 photos for each clay model.

Statistics

To investigate the relationship between volume estimated from photogrammetry with manual measures of main-beam length, we used a repeated-measures model with log beam length and antler shape (palmated, bifurcated, main-beamed) as fixed predictors, and specimen as a random factor to account for the non-independence of the measurements on two antlers from the same specimen, using the model "log(antler volume) ~ log(main-beam length) + shape + (1|specimen)" in lme4, version 2.1.21 (Bates et al. 2015).

To assess the relationship between volume and weight of the loose antlers, we did simple regression on the full data set, and an ancova with species as a factor in a subset of 50 antlers from 6 species with at least five observations. All analyses were performed using R version 3.6.0 (R Core Team 2019), and uncertainties are reported as standard errors unless otherwise specified.

Results

Processing Time

Photographing a specimen took up to 10 min. After uploading the photos to the cloud server, the rendering process took between 2 to 6 h, depending on the number of photos in the set and the server status. There is no need to monitor the rendering process and multiple photosets can be processed in parallel. Manual image processing of a raw photogrammetry object took between 5 and 20 min. In total, it took up to 30 min of active work to obtain a volume estimate for a set of left and right antlers from a single specimen.

Measurement Error

The relative error for a single antler, including all sources of error, averaged $8.45 \pm 0.48\%$ (median: 7.14%), and bifurcated, main-beamed, and palmated antlers had average relative error of $9.3 \pm 1.1\%$, $8.6 \pm 0.6\%$, and $6.7 \pm 1.3\%$, respectively (Fig. 2). The measurement errors appeared homoscedastic on log scale and the overall measurement variance on this scale was $\sigma^2_m = 0.0056 \log(\text{mm}^3)^2$. Decomposition into different sources of error indicated that image processing had negligible influence (Fig. 3). The measurement error mainly arose from error in scaling (70% of the total σ^2_m), followed by error in taking photos (~ 19% of the total σ^2_m) and rendering (~ 9% of the total σ^2_m). In calculating these proportions, we excluded six outliers on the replication of rendering + scaling + processing as shown
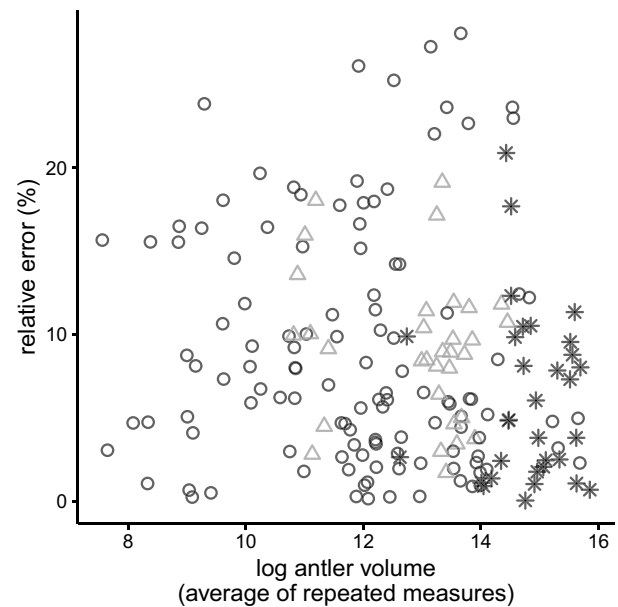

Fig. 2 Relative error of antler volume in relation to size and shape. Symbols represent different shape categories; asterisk: palmated, circle: main-beamed, triangle: bifurcated. The ANCOVA with relative error as response variable and size and shape as predictors gave intercepts (evaluated at the grand mean of log antler volume) of $9.3 \pm 1.1\%$, $8.6 \pm 0.6\%$, $6.7 \pm 1.3\%$ for bifurcated, main-beamed and palmated antlers, respectively. The common slope on log antler volume was $- 0.27 \pm 0.30$. The model explained less than 2% of the variance

in Fig. 3. These outliers were from three specimens (*Axis axis*; NHMW 20995, *Elaphurus davidianus*; NHMV 38106, *Mazama americana*; NHMW 3750) with a poor contrast between the skull and the scale bar.

The dependency of the error on scaling implies that there will be measurement covariance between any two measures taken on the same object. For example, the measurement covariance between the log volumes of the left and the right antler in our data was $0.0041 \log(\text{mm}^3)^2$ giving a correlation of 0.72 (Fig. S2).

The estimated volumes for the twenty clay models were on average $1.8 \pm 1.2\%$ smaller than the true volumes (Fig. 4). This is about the magnitude one would expect if the method was unbiased. If the measurements, *m*, are unbiased, homoscedastic and normally distributed then

$$E\left[|\overline{m} - x|\right] = \sqrt{\frac{2Var[\overline{m}]}{\pi}} = \sqrt{\frac{2}{\pi}\frac{\sigma^2_m}{n}},$$

where *x* is the true value, σ^2_m is the measurement variance, and *n* is the sample size. Assuming the measurements from the clay models have the same variance on log scale as we estimated for the real antlers, we can fit in $\sigma^2_m = 0.0056 \log(\text{mm}^3)^2$ and $n = 20$ to obtain an expected deviation of
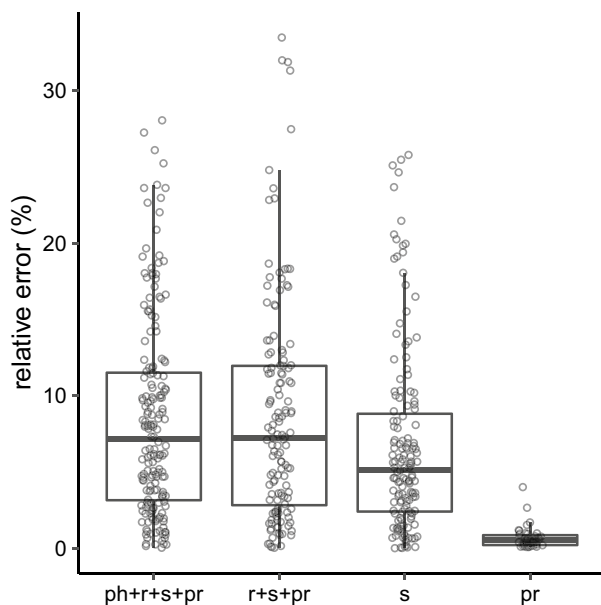
Fig. 4 Estimated bias of 3D-photogrammetry volume measurements from clay models. Proportional differences between estimated and true volumes are plotted against the true model volume. Circles show main-beamed model antlers and triangles show bifurcated model antlers. The solid line denotes no bias and the dashed lines are the means of the estimates for the two shapes

Fig. 3 Comparison of magnitudes of measurement errors (relative error, %) from different sources. Letters below each boxplot indicate the components of error included at each level as follows; ph: photoset, r: rendering, s: scaling, pr: processing. The estimated total measurement variance in log volume was $\sigma^2_m = 0.00560$ log(mm³)². Excluding six data points due to three specimens in the r+s+pr replication, which were likely due to a poor precision in scaling, the estimated measurement variance due to all sources except taking photos was 0.00451 log(mm³)². The measurement variances due to scaling and processing were 0.00392 log(mm³)² and 0.00005 log(mm³)², respectively. The boxes show the median with the 25% and 75% percentiles. Overlaid points are each repeated observation. The average relative errors of each category were: ph+r+s+pr: 8.45±0.48% ($n=184$), r+s+pr: 7.71%±0.50% (n=132, and 8.52 ± 0.63%, $n=138$, with outliers included), s: 6.73 ±0.55% ($n=150$), pr: 0.69 ±0.11%, $n=42$)

1.3% between the mean of the measurements and the true value.

The Effect of Photography Protocols

The precision of the method depends on the number of photos taken per object (Fig. 5a). With 20 pictures the rendering always failed. With 30 pictures it generally worked, but the precision was poor with a mean relative error above 10%. This decreased to 5–6% with 40, 50 or 60 pictures. The smaller relative error reported here compared to the full data set may reflect the fact that this experiment was done on a subset of antlers photographed in one location only.

Increasing the distance to the object reduced precision. At twice the distance, 60 pictures were necessary to achieve a precision similar to that obtained with 40 pictures at the original distance. Hence, optimizing distance and number of pictures is important in developing an efficient protocol. Generally, pictures should be taken at the closest distance
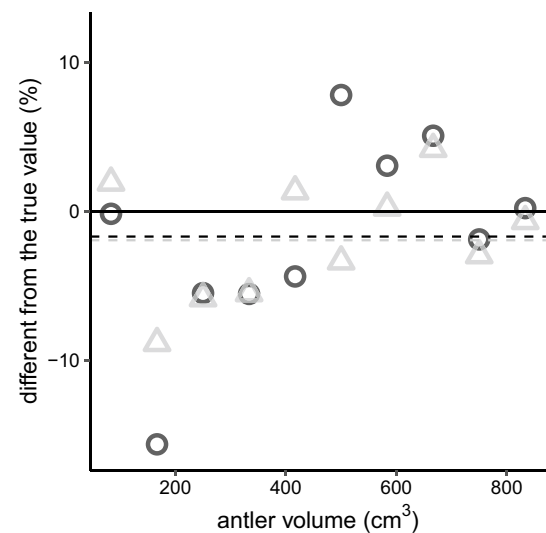
from which the entire skull can be photographed. We found no effect of the camera type on the precision of the measurements (Fig. 5b).

The Relationship Between Volume, Main-Beam Length and Weight

Main-beam length explained 92% of the variance in antler volume on a log scale across the whole sample. Since the volumes in our data set range over three orders of magnitude this can be considered a poor fit. Most of the remaining variance was explained by antler shape and an interaction between shape and size (Fig. 6). In particular, palmated antlers, which are found mostly in large-bodied species, had a larger volume for a given beam length than other shapes.

The estimated volume of an antler was a crude predictor of its weight across the sample of 71 loose antlers. The log–log regression was not isometric (slope=0.91±0.03), and log volume explained 91% of the variance in log weight. The remaining variance can be partially explained by differences among species. Restricting the analysis to the six species with more than five observations and including species as a factor explained 97% of the variance and produced a regression slope closer to isometry (slope=0.95±0.04, Fig. 7). Estimated species-specific antler densities (i.e. weight divided by volume) are reported in Table 1. The
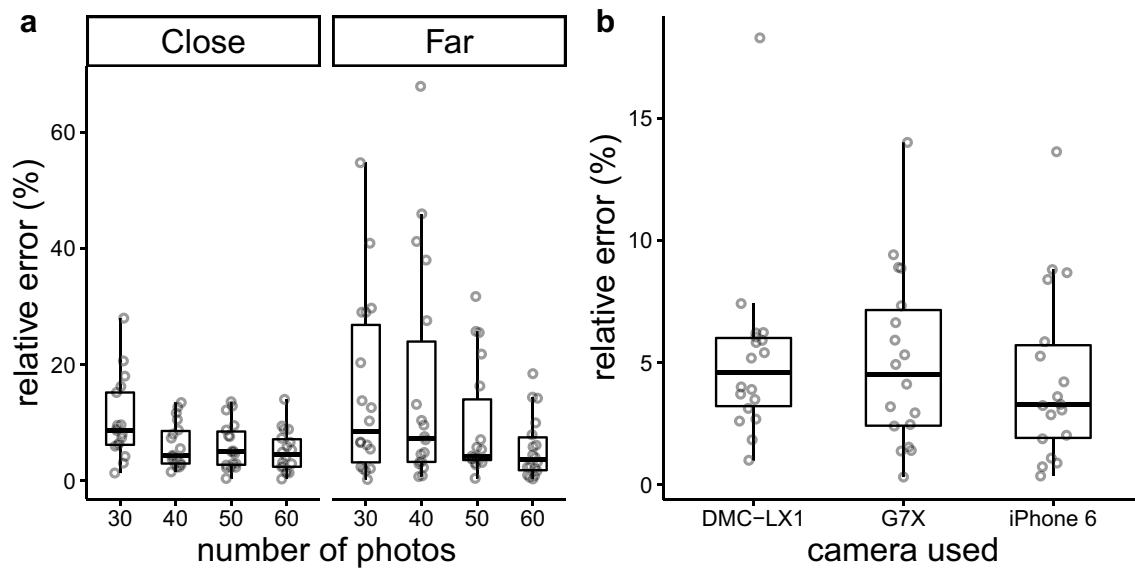
Fig. 5 Comparison of the relative error of volume estimated from different number of images taken from different distances between specimen and camera (**a**) and using different cameras (**b**). Plots under the header "Close" show results from photos taken from a standard distance (i.e. the nearest distance from which the entire specimen can be photographed), and those under "Far" show results from photos taken from twice the standard distance. Boxes as in Fig. 4
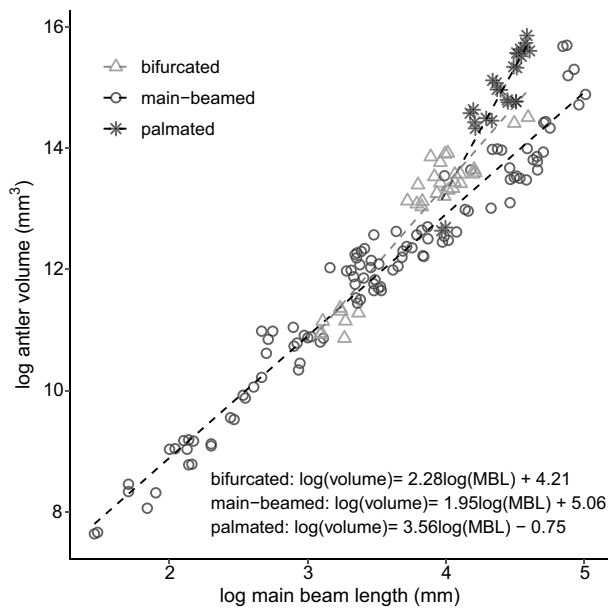

Fig. 6 Relationship between log antler volume (average of two repeated measurements) and log main-beam length (MBL). Dashed lines show regressions within each shape category. Bifurcated, main-beamed and palmated antlers are indicated by triangles, circles and asterisks, respectively
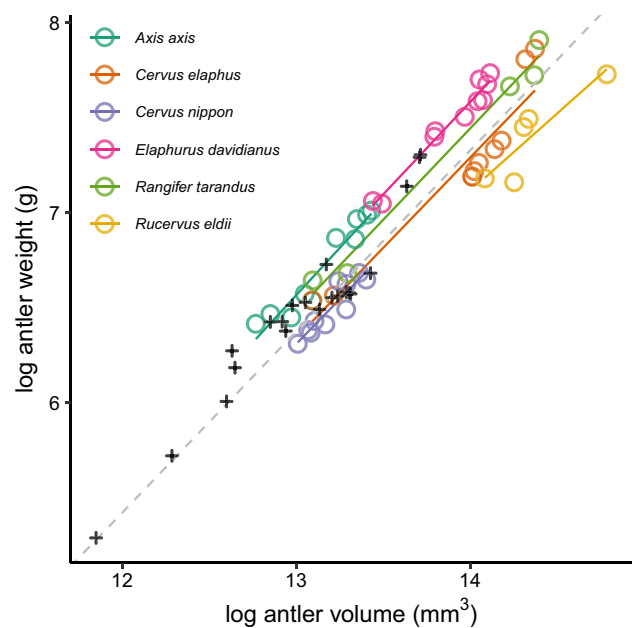

Fig. 7 Relationship between log antler weight (g) and log antler volume (mm^3). Six species with five or more individuals are represented with open circles of different colors and solid lines describe the ordinary least squares regression for each species. A common slope of the bivariate relationship estimated from the model: log(antler weight) = 0.95 (±0.039) log(antler volume) − 6.02 (±0.52) is shown in a grey dashed line. Observations with fewer than five observations per species are represented with black crosses

total variance in density of 0.065 g^2 ml^{-2} decomposes into 82% among species and 18% within species. A considerable

Table 1 Antler density of 14 deer species

Species	n	Density (g ml^{-1})	
		Mean	SD
Axis axis	10	1.61	0.110
Capreolus pygargus	4	1.36	0.160
Cervus elaphus	10	1.23	0.173
Cervus nippon	10	1.23	0.068
Dama dama	2	1.22	0.059
Dama mesopotamica	1	1.51	-
Elaphurus davidianus	10	1.64	0.073
Odocoileus hemionus	4	1.33	0.194
Odocoileus virginianus	4	1.55	0.151
Ozotoceros bezoarticus	2	1.41	0.004
Rangifer tarandus	5	1.43	0.117
Rucervus duvaucelii	2	1.64	0.009
Rucervus eldii	5	0.96	0.110
Rusa unicolor	2	1.51	0.062

Species name, sample size (*n*), mean and standard deviation (SD) are presented

amount of the among-species variance was due to the low estimated density for *Rucervus eldii*. It is possible that some of the variation in density is due to age or circumstances of collection (which was not documented for this sample), and further studies are necessary to verify and identify causes of variation in antler density among species.

Discussion

We have shown that 3D photogrammetry provides unbiased and reasonably precise measurements of the volume of deer antlers. The method appears robust to variation in object size and shape if the photogrammetry is based on at least 40 images taken at a close distance. The accuracy is comparable with 3D photogrammetry of cubic (volume or mass) measurements in other studies (Table 2). The average relative error of 8.5% is not negligible in all contexts but should provide enough precision for most studies.

The decomposition of measurement errors into different sources revealed that most of the measurement variance (70%) was due to error in the scaling of the objects and

Table 2 Summary of imprecision and bias in photogrammetry measurements in various trait and taxa

Study	Software	Taxa	Measured trait (unit)	Imprecision		Bias
				σ^2_m	Relative error	
This study	RECAP	Deer	Antler volume (mm^3)	0.0056	8.45 ± 0.48 (92)	−1.81 ± 1.20 (20)
Beltran et al. (2018)	PhotoModeller	Weddell seal	Body mass (kg)	0.0024	5.66 ± 0.32 (168)	−0.01 ± 0.74 (56)
Lavy et al. (2015)	RECAP[a]	Coral/Sponge	Colony volume (cm^3)	–	5.70 ± 0.33 (6)	−1.98 ± 3.77 (23)
			Colony area (cm^2)	–	1.35 ± 0.36 (6)	6.67 ± 5.20 (23)
Postma et al. (2015)	PhotoModeller	Primate	Body mass (kg)	–	–	32.0 ± 1.85 (7)
		Ruminant	Body mass (kg)	–	–	24.0 ± 0.75 (22)
		Carnivore	Body mass (kg)	–	–	11.8 ± 0.9 (16)
		Hindgut fermenter	Body mass (kg)	–	–	19.9 ± 0.63 (38)
		Foregut fermenter	Body mass (kg)	–	–	11.8 ± 1.13 (9)
Katz and Friess (2014)	PhotoScan	Human	Parietal area (mm^2)	–	0.32 (4)	–
		Human	Nasal area (mm^2)	–	4.48 (4)	–
de Bruyn et al. (2009)	PhotoModeller	Elephant seal	Body mass (kg)	–	–	6.59 ± 0.21 (53)
Chiari et al. (2008)	PhotoModeller	Giant tortoise	Carapace length (mm)	0.0001	0.99 ± 0.15 (24)	2.21 ± 0.55 (24)
Waite et al. (2007)	PhotoModeller	Steller sea lion	Body mass (kg)	–	–	10.0 ± 0.56 (137)
		Steller sea lion	Body length (cm)	–	–	2.5 ± 0.02 (138)
		Steller sea lion	Axillary girth (cm)	–	–	4.0 ± 0.02 (137)
		Sand seatrout	Body mass (kg)	–	–	5.6 (30)
		Guld manhaden	Body mass (kg)	–	–	3.4 (32)
		Chinese alligator	Body mass (kg)	–	–	4.8 (18)

Imprecision are calculated from repeated measurements of the same subject and expressed as measurement variance (σ^2_m) in log of examined trait values and relative error as described in the main text. Bias is calculated as the mean proportional difference between photogrammetry measurements and manual measurements. Software used, taxa, trait, mean and standard error (SE) of imprecision and bias are presented with sample size in parentheses

"–" indicates that values are either not reported or could not be calculated from available data

[a]The original article is reported using the old brand name of RECAP (123D CATCH)

another 29% was due to photosets and rendering. This indicates that the error can be reduced by replicating the scaling. For instance, estimating volume by replicating the scaling procedure and using the mean of two replicates would reduce the relative error from 8.5% to 5.5%. A further reduction to 4.3% can be achieved by replicating image rendering and scaling. Also note that using the average or sum of the left and right antler volumes would reduce the relative measurement variance, although with less than a factor of two due to the measurement covariance.

The sensitivity to scaling is further illustrated by the outliers in Fig. 3 that we removed when calculating the sources of measurement variance. These were from three skulls with unusually pale color making a poor contrast between the skull and the scale bar. To avoid such situations, we recommend using a scale bar that is easily distinguishable from its background, for instance by marking the scale with high-contrast color or by using larger scales. In general we recommend paying attention to the choice of scale bar for converting voxels to a physical scale, and particularly to the size of the scale bar relative to the size of the object.

Measurement error is rarely a problem if it can be assessed and taken into account. Even if the error in individual measurements is unknown, it is usually possible to quantify the imprecision of measurements with repeated measures. The resulting estimates of measurement variance can then be incorporated into statistical analyses with a variety of techniques (e.g. Fuller 1987; Buonaccorsi 2010; Hansen and Bartoszek 2012; Morrissey 2016; Ponzi et al. 2018). Assessment of bias requires separate treatments, but again can be accounted for if it is estimated. The photogrammetry technique described here is in fact ideal for estimation of measurement variance. As bias is likely to be small, and relatively little error variance comes from the pictures, it is easy to replicate the main error-generating parts of the process. We recommend always replicating image scaling. This will not only reduce the measurement variance by using the average of the two measures, but also generate an estimate of the measurement variance and covariance to use in statistical analyses. There is of course always a trade-off between replication and taking data from more specimens that must be assessed on a study-by-study basis, but the extra effort of repeating rendering and scaling will often be trivial compared to replicating the entire measurement procedure.

The importance of measurement error depends on the goals of the study. For comparative studies of species mean antler size, an 8.5% relative error would have little impact even if there were only a single specimen representing each species. If the goal is to study fluctuating asymmetry then an expected 8.5% relative error could rival the signal to be studied and must be incorporated into the analysis. Note also that estimates of error variance are themselves subject to estimation error (Hansen 2016), and large samples of repeated measures are generally needed for good estimates of the measurement variance. It is also important to remember that measurement bias can not be estimated from repeated measurements and requires explicit experiment as we performed with clay models.

The impact of measurement variance on a given analysis is often assessed as a repeatability (e.g. Wolak et al. 2012). As illustrated in Fig. 8, repeatability is not a general measure of precision, but only a quantification of its impact in a specific analysis. It is a "rubber scale" (sensu Houle et al. 2011), which is as dependent on the size and range of items measured as on the actual precision of the measurements. Across our whole data set the repeatability of antler volume is essentially perfect (100%), but it drops when the data spans more narrow size ranges (Figs. 8 and 9). In particular, the repeatability of fluctuating asymmetry in moose antler volume in our data set is only 68% (Fig. 9).
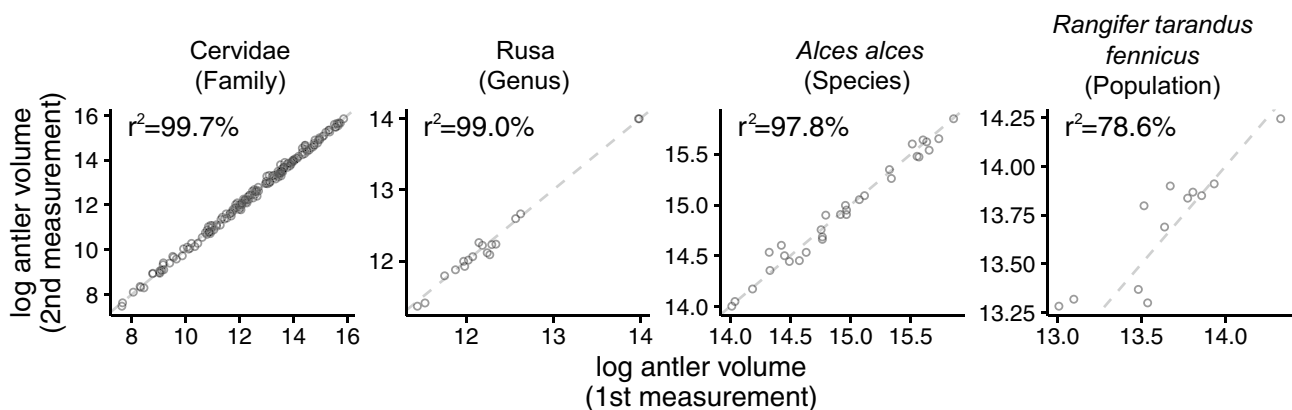

Fig. 8 Repeatability of volume measurements at different taxonomic levels. Each panel, left to right, shows repeatability of our full sample of 92 measurements, two species of the genus *Rusa*, three subspecies of moose (*Alces alces*), and within a population of reindeer. Dashed lines show one-to-one relationship and r² is the repeatability at each level
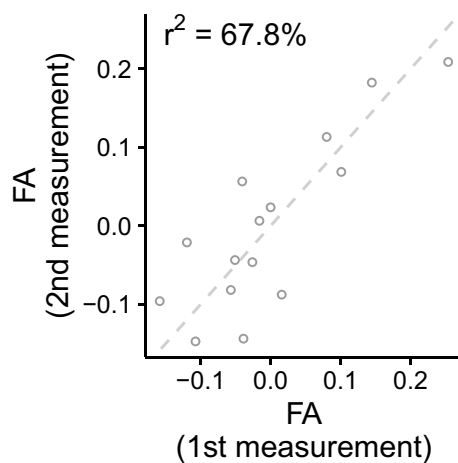
Fig. 9 Repeatability of fluctuating asymmetry (FA) in antler volume of *Alces alces*, evaluated as the proportional difference between left and right antlers where positive values are assigned to left-biased individuals. Dashed lines show one-to-one relationship and r² is the repeatability

It is noteworthy that the measurement error of antler volume appears homoscedastic on a log scale and thus heteroscedastic on an arithmetic scale (Fig. S1). This is relevant to ongoing debates on whether allometries are best studied on arithmetic or logarithmic scales (see Lemaître et al. 2015; Packard 2015; Pélabon et al. 2018 for a manifestation in relation to horns and antlers). It has been argued that biological deviations from the statistical model tend to be multiplicative and thus homoscedastic on a proportional (e.g. log) scale, while measurement errors tend to be additive and thus homoscedastic on an arithmetic scale (e.g. Riska 1991; Smith 1993; Kerkhoff and Enquist 2009; Voje et al. 2014). Our results imply that allometric analyses involving antler volumes estimated by 3D photogrammetry would be best done on a logarithmic scale also from the perspective of handling measurement error.

Main-beam length is the most common measure of antler size in the literature. We have shown that main-beam length and antler volume are not interchangeable measurements of antler size. The error in using beam length as a measure of volume mainly comes from the fact that palmated and, to a lesser extent, branched antlers have higher volumes for a given beam length. This finding prompts a reconsideration of results based on beam length. For example, Lemaître et al. (2014) reported a breakpoint in the evolutionary allometric slope of beam length at 100–120 kg of body mass, and suggested that this reflected an increased cost of producing and maintaining antlers in large species. A similar pattern has been reported in bovids (Tidière et al. 2017) and stag beetles (Huxley 1932; Knell et al. 2004). Based on our findings, an alternative explanation could be that the breakpoint results from a change in the scaling relationship with size.

Likewise, Gould's (1974) famous claim that the herculean antlers of the extinct Irish elk (*Megaloceros giganteus*) were in fact expected from the evolutionary allometry across cervine deers was based on linear measurements of antlers. Considering the heavily branched and palmated shape of the Irish elk antler, it is conceivable that linear measurements underestimates size in this species.

Although we have focused on volume, we emphasize that a general strength of photogrammetry is that most phenotypic features of antlers, including length, surface area and tine count can be measured from the objects, and that these objects or the original photosets can be revisited for further measures or replication of previous measures. As documented in Table 2, photogrammetry has already been applied to study various linear, area, or cubic (volume or mass) traits in a variety of organisms. The approach has wide potential and will be particularly useful for storing phenotypic data in an accessible and replicable form as required for proper phenomics (Houle 2001; Houle et al. 2010).

Author Contributions TFH, CP and MT conceived the ideas and designed methodology. MG, BTK, CS and MT collected data. MT analyzed data in collaboration with TFH and CP. TFH and MT led the writing of the manuscript. All authors commented critically on the manuscript and gave approval for publication.

Data Availability Data is available from the corresponding author upon request.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Bartoszek, K., Pienaar, J., Mostad, P., Andersson, S., & Hansen, T. F. (2012). A phylogenetic comparative method for studying multivariate adaptation. *Journal of Theoretical Biology, 314*, 204–215.

Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Beltran, R. S., Ruscher-Hill, B., Kirkham, A. L., & Burns, J. M. (2018). An evaluation of three-dimensional photogrammetric and morphometric techniques for estimating volume and mass in Weddell seals *Leptonychotes weddellii. PLoS ONE, 13*, e0189865.

Bookstein, F. L. (1991). *Morphometric tools for landmark data: Geometry and biology*. Cambridge: Cambridge University Press.

Breuer, T., Robbins, M. M., & Boesch, C. (2007). Using photogrammetry and color scoring to assess sexual dimorphism in wild western gorillas (*Gorilla gorilla*). *American Journal of Physical Anthropology, 134*, 369–382.

Brown, W. A. B., & Chapman, N. G. (1990). The dentition of fallow deer (*Dama dama*)—A scoring scheme to assess age from wear of the permanent molariform teeth. *Journal of Zoology, 221*, 659–682.

Buonaccorsi, J. P. (2010). *Measurement error: Models, methods, and applications*. Boca Raton: Chapman and Hall/CRC.

Buser, T. J., Sidlauskas, B. L., & Summers, A. P. (2018). 2D or not 2D? Testing the utility of 2D vs. 3D landmark data in geometric morphometrics of the sculpin subfamily Oligocottinae (Pisces; Cottoidea). *Anatomical Record, 301*, 806–818.

Cardini, A. (2014). Missing the third dimension in geometric morphometrics: How to assess if 2D images really are a good proxy for 3D structures? *Hystrix, 25*, 73–81.

Ceacero, F. (2016). Long or heavy? Physiological constraints in the evolution of antlers. *Journal of Mammalian Evolution, 23*, 209–216.

Chiari, Y., Wang, B., Rushmeier, H., & Caccone, A. (2008). Using digital images to reconstruct three-dimensional biological forms: A new tool for morphological studies. *Biological Journal of the Linnean Society, 95*, 425–436.

Christiansen, F. M., Sironi, M. J., Moore, M. D., Martino, M., Riccardi, H. A., Warick, D. J., et al. (2019). Estimating body mass of free-living whales using aerial photogrammetry and 3D volumetrics. *Methods in Ecology and Evolution, 10*, 2034–2044.

Clutton-Brock, T. H., Albon, S. D., & Harvey, P. H. (1980). Antlers, body size and breeding group size in the Cervidae. *Nature, 285*, 565–567.

de Bruyn, P. J. N., Bester, M. N., Carlini, A. R., & Oosthuizen, W. C. (2009). How to weigh an elephant seal with one finger: A simple three-dimensional photogrammetric application. *Aquatic Biology, 5*, 31–39.

Emlen, D. J. (2008). The evolution of animal weapons. *Annual Review of Ecology, Evolution, and Systematics, 39*, 387–413.

Falkingham, P. L. (2012). Acquisition of high resolution three-dimensional models using free, open-source, photogrammetric software. *Palaeontologica Electronica, 15*, 1–15.

Fuller, W. A. (1987). *Measurement error models*. Hoboken: Wiley.

Geist, V. (1998). *Deer of the world: Their evolution, behaviour, and ecology*. Mechanicsburg: Stackpole books.

Giacomini, G., Scaravelli, D., Herrel, A., Veneziano, A., Russo, D., Brown, R. P., et al. (2019). 3D photogrammetry of bat skulls:

Perspectives for macro-evolutionary analyses. *Evolutionary Biology, 46*, 249–259.

Gould, S. J. (1973). Positive allometry of antlers in Irish elk, *Megaloceros giganteus. Nature, 244*, 375–376.

Gould, S. J. (1974). The origin and function of 'bizarre' structures: Antler size and skull size in the Irish elk *Megaloceros giganteus. Evolution, 28*(2), 191–220.

Gould, S. J. (1977). *Ontogeny and phylogeny*. Cambridge: Harvard University Press.

Gutierrez-Heredia, L., Benzoni, F., Murphy, E., & Reynaud, E. G. (2016). End to end digitisation and analysis of three-dimensional coral models, from communities to corallites. *PLoS ONE, 11*, e0149641.

Hansen, T. F. (2016). On bias and precision in meta-analysis: The error in the error. *Journal of Evolutionary Biology, 29*, 1919–1921.

Hansen, T. F., & Bartoszek, K. (2012). Interpreting the evolutionary regression: The interplay between observational and biological errors in phylogenetic comparative studies. *Systematic Biology, 61*, 413–425.

Holman, L., & Bro-Jørgensen, J. (2016). Ornament complexity is correlated with sexual selection (a comment on Raia et al., "Cope's rule and the universal scaling law of ornament complexity"). *The American Naturalist, 188*, 272–275.

Houle, D. (2001). Characters as the units of evolutionary change. In G. P. Wagner (Ed.), *The character concept in evolutionary biology* (pp. 109–140). San Diego: Academic Press.

Houle, D., Govindaraju, D. R., & Omholt, S. (2010). Phenomics: The next challenge. *Nature Reviews Genetics, 11*, 855–866.

Houle, D., Pélabon, C., Wagner, G. P., & Hansen, T. F. (2011). Measurement and meaning in biology. *The Quarterly Review of Biology, 86*, 3–34.

Huxley, J. S. (1931). The relative size of antlers in deer. *Proceedings of the Zoological Society of London, 101*, 819–864.

Huxley, J. S. (1932). *Problems of relative growth*. Essex: Methuen and Company Limited.

Katz, D., & Friess, M. (2014). Technical note: 3D from standard digital photography of human crania—A preliminary assessment. *American Journal of Physical Anthropology, 154*, 152–158.

Kerkhoff, A. J., & Enquist, B. J. (2009). Multiplicative by nature: Why logarithmic transformation is necessary in allometry. *Journal of Theoretical Biology, 257*, 519–521.

Knell, R. J., Pomfret, J. C., & Tomkins, J. L. (2004). The limits of elaboration: curved allometries reveal the constraints on mandible size in stag beetles. *Proceedings of the Royal Society of London Series B: Biological Sciences, 271*, 523–528.

Kruuk, L. E. B., Slate, J., Pemberton, J. M., Brotherstone, S., Guinness, F., & Clutton-Brock, T. (2002). Antler size in red deer: Heritability and selection but no evolution. *Evolution, 56*, 1683–1695.

Kruuk, L. E. B., Slate, J., Pemberton, J. M., & Clutton-Brock, T. H. (2003). Fluctuating asymmetry in a secondary sexual trait: No associations with individual fitness, environmental stress or inbreeding, and no heritability. *Journal of Evolutionary Biology, 16*, 101–113.

Lagesen, K., & Folstad, I. (1998). Antler asymmetry and immunity in reindeer. *Behavioral Ecology and Sociobiology, 44*, 135–142.

Lavy, A., Eyal, G., Neal, B., Keren, R., Loya, Y., & Ilan, M. (2015). A quick, easy and non-intrusive method for underwater volume and surface area evaluation of benthic organisms by 3D computer modelling. *Methods in Ecology and Evolution, 6*, 521–531.

Lawing, A. M., & Polly, P. D. (2010). Geometric morphometrics: Recent applications to the study of evolution and development. *Journal of Zoology, 280*, 1–7.

Lemaître, J. F., Vanpe, C., Plard, F., & Gaillard, J. M. (2014). The allometry between secondary sexual traits and body size is nonlinear among cervids. *Biology Letters, 10*, 20130869.

Lemaître, J. F., Vanpe, C., Plard, F., Pélabon, C., & Gaillard, J. M. (2015). Response to Packard: make sure we do not throw out the biological baby with the statistical bath water when performing allometric analyses. *Biology Letters, 11*, 20150144.

Mills, K. J., & Peterson, R. O. (2013). Moose antler morphology and asymmetry on Isle Royale National Park. *Alces: A Journal Devoted to the Biology and Management of Moose, 49*, 17–28.

Mitteroecker, P., & Gunz, P. (2009). Advances in geometric morphometrics. *Evolutionary Biology, 36*, 235–247.

Moen, R. A., Pastor, J., & Cohen, Y. (1999). Antler growth and extinction of Irish elk. *Evolutionary Ecology Research, 1*, 235–249.

Morrissey, M. B. (2016). Meta-analysis of magnitudes, differences and variation in evolutionary parameters. *Journal of Evolutionary Biology, 29*, 1882–1904.

Munoz-Munoz, F., Quinto-Sanchez, M., & Gonzalez-Jose, R. (2016). Photogrammetry: A useful tool for three-dimensional morphometric analysis of small mammals. *Journal of Zoological Systematics and Evolutionary Research, 54*, 318–325.

Mysterud, A., Meisingset, E., Langvatn, R., Yoccoz, N. G., & Stenseth, N. C. (2005). Climate-dependent allocation of resources to secondary sexual traits in red deer. *Oikos, 111*, 245–252.

Packard, G. C. (2015). Allometric variation in the antlers of cervids: A comment on Lemaitre et al. *Biology Letters, 11*, 20140923.

Pélabon, C., & Joly, P. (2000). What, if anything, does visual asymmetry in fallow deer antlers reveal? *Animal Behaviour, 59*, 193–199.

Pélabon, C., Tidière, M., Lemaître, J. F., & Gaillard, J. M. (2018). Modelling allometry: Statistical and biological considerations— A reply to Packard. *Biological Journal of the Linnean Society, 125*, 664–671.

Pélabon, C., & van Breukelen, L. (1998). Asymmetry in antler size in roe deer (*Capreolus capreolus*): An index of individual and population conditions. *Oecologia, 116*, 1–8.

Plard, F., Bonenfant, C., & Gaillard, J. M. (2011). Revisiting the allometry of antlers among deer species: Male-male sexual competition as a driver. *Oikos, 120*, 601–606.

Ponzi, E., Keller, L. F., Bonnet, T., & Muff, S. (2018). Heritability, selection, and the response to selection in the presence of phenotypic measurement error: Effects, cures, and the role of repeated measurements. *Evolution, 72*, 1992–2004.

Postma, M., Tordiffe, A. S. W., Hofmeyr, M. S., Reisinger, R. R., Bester, L. C., Buss, P. E., et al. (2015). Terrestrial mammal three-dimensional photogrammetry: Multispecies mass estimation. *Ecosphere, 6*, 1–16.

Price, J., & Allen, S. (2004). Exploring the mechanisms regulating regeneration of deer antlers. *Philosophical Transactions of the Royal Society B: Biological Sciences, 359*, 809–822.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Richtsmeier, J. T., Lele, S. R., & Cole, T. M., III. (2005). Landmark morphometrics and the analysis of variation. In B. Hallgrimsson & B. K. Hall (Eds.), *Variation: A central concept in biology* (pp. 49–69). Amsterdam: Elsevier.

Riska, B. (1991). Regression models in evolutionary allometry. *The American Naturalist, 138*, 283–299.

Roth, F., Wild, C., Carvalho, S., Rädecker, N., Voolstra, C. R., Kürten, B., et al. (2019). An in situ approach for measuring biogeochemical fluxes in structurally complex benthic communities. *Methods in Ecology and Evolution, 10*, 712–725.

Simpson, G. G. (1953). *The major features of evolution*. New York: Columbia University Press.

Smith, R. J. (1993). Logarithmic transformation bias in allometry. *American Journal of Physical Anthropology, 90*, 215–228.

Tidière, M., Lemaître, J. F., Pélabon, C., Gimenez, O., & Gaillard, J. M. (2017). Evolutionary allometry reveals a shift in selection pressure on male horn size. *Journal of Evolutionary Biology, 30*, 1826–1835.

Voje, K. L., Hansen, T. F., Egset, C. K., Bolstad, G. H., & Pélabon, C. (2014). Allometric constraints and the evolution of allometry. *Evolution, 68*, 866–885.

Waite, J. N., Schrader, W. J., Mellish, J. A. E., & Horning, M. (2007). Three-dimensional photogrammetry as a tool for estimating morphometrics and body mass of Steller sea lions (*Eumetopias jubatus*). *Canadian Journal of Fisheries and Aquatic Sciences, 64*, 296–303.

Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution, 3*, 129–137.

Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. (2012). *Geometric morphometrics for biologists: A primer*. Amsterdam: Academic Press.