SALAMANDRA 55(1) | 14–26 | 15 February 2019 | ISSN 0036–3375

Fluctuating asymmetry – appearances are deceptive. Comparison of methods for assessing developmental instability in European Common Frogs (*Rana temporaria*)

Stephanie Niemeier[1,2,3], Johannes Müller[1,2] & Mark-Oliver Rödel[1,2]

[1)] Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Invalidenstr. 43, 10115 Berlin, Germany
[2)] Berlin-Brandenburg Institute of Advanced Biodiversity Research – BBIB, Altensteinstr. 34, 14195 Berlin, Germany

[3)] Corresponding author. Stephanie Niemeier, e-mail: stephanie.niemeier@mfn.berlin

Manuscript received: 13 September 2018
Accepted: 11 December 2018 by Alexander Kupfer

Abstract. Developmental instability provides a powerful monitoring tool to detect threats prior to population declines. Consequently, assessing the level of developmental instability by measuring fluctuating asymmetry (FA) of bilaterally symmetrical traits in association with environmental stress has become increasingly attractive. However, many studies failed in detecting a clear connection of FA to environmental stressors. Some of these may have suffered from large measurement error (ME) or the use of inappropriate methods. Here, we compared measurement accuracy and FA outcome from manual calliper measurements with those from non-destructive micro-3D-computed tomography (µCT) based skeletal measurements. Amphibians are assumed to be ideal models for measuring fluctuating asymmetry due to their sensitivity to environmental stress. For our study, we chose two bilateral, metric traits (femur and radio–ulna length) of the European Common Frog, *Rana temporaria.* Calliper measurements revealed meaningful FA estimates (i.e., FA exceeded ME) for radio–ulna length only. In contrast, µCT-based measurements delivered meaningful FA estimates for both traits. ME was about twice as high for calliper measurements compared to µCT-based measurements, resulting in inflated levels of FA. Using callipers, we observed higher ME for femur measurements than for radio–ulna, meaning that ME strongly depended on the respective trait. When using µCT, however, we observed comparable ME between both traits. Our study revealed that analyses of developmental instability using manual measurements should be treated with caution. For smaller vertebrates we recommend skeletal measurements with µCT as a valuable alternative due to its greater reliability, thereby allowing for multi-trait analyses with equal accuracy.

Key words. Agreement, amphibians, Bland-Altman, developmental stability, environmental stress, measurement error, Micro CT, morphology, *Rana temporaria*, traits.

Introduction

Humans are altering landscapes all over the planet, and it is becoming increasingly clear that this causes many unwelcome effects that should be avoided wherever possible. A precondition to avoid or reduce such effects is to monitor animal and plant populations facing the risk of environmental change. By timely starting necessary counteractions, one then can try to maintain populations' health prior to severe and irreversible declines. In this respect, developmental stability has been proposed as a sensitive indicator of population health (Clarke 1993, Freeman et al. 1996, Jones 1987, Lens et al. 2002b). Developmental stability, defined as the ability to develop the same phenotype irrespective of different environmental conditions (Zakharov et al. 1991), represents the ability to resist developmental accidents (Van Valen 1962) or imprecisions in developmental processes, also termed developmental noise (Waddington 1957). Decreases in developmental stability or increases of developmental noise through genetic or environmental disturbances result in developmental instability (Clarke 1995, Lens et al. 2002a, Palmer & Strobeck 1992).

The level of developmental instability is most commonly assessed by measuring the degree of fluctuating asymmetry (FA) (Mather 1953, Møller & Swaddle 1997, Palmer & Strobeck 1986). FA are small, random deviations from the symmetry of bilaterally symmetrical traits (Ludwig 1932) with symmetrically distributed right-minus-left (R-L) differences about a mean of zero (Graham et al. 2010, Palmer & Strobeck 1986). However, FA must be distinguished from two other types of asymmetry i.e., directional asymmetry and antisymmetry (Van Valen 1962). Directional asymmetry occurs when a trait within a population is consistently larger on one particular side of the body, and contrary to FA, has a mean that is significantly different from zero and is skewed to either the left or the right side. Antisymmetry occurs when a trait is usually larger on one side of the body, but the side of which is variable, having a mean of zero but a bimodal or platykurtic (i.e., broad peaked) distribution (Møller & Swaddle 1997, Palmer & Strobeck 1986, Van Valen 1962). Directional asymmetry and antisymmetry are generally considered to be inappropriate for estimating developmental instability due to their presumed heritable component (Knierim et al. 2007). Although there might be transitions from FA to the other asymmetry types (Graham et al. 1993, Lens & Van Dongen 2000), and the genetic basis of FA is still under debate (Leamy et al. 2015, Leamy & Klingenberg 2005), it is recommended to first test for directional asymmetry and antisymmetry and, if present, best avoid these traits (Graham et al. 1998, Van Dongen 2006). FA proved to have the potential to serve as a tool to detect environmental stress in all bilaterally symmetrical taxa, such as e.g. insects (Beasley et al. 2013, Schmeller et al. 2011), fish (Leary & Allendorf 1989, Valentine et al. 1973), amphibians (Costa & Nomura 2015, Söderman et al. 2007), reptiles (Lazić et al. 2013, Sarre 1996), birds (Anciães & Marini 2000, Lens & Eggermont 2008), and mammals (Marchand et al. 2003). However, there is the risk of 'false positives' when measurement error (ME) is not considered (Floate & Coghlin 2010, Hoffmann & Woods 2003). Many studies failed in detecting a clear connection of FA with environmental stressors, possibly, at least in part, because they suffered from inappropriate methods and statistics (Knierim et al. 2007, Palmer & Strobeck 2003a). Thus, increasing measurement accuracy (i.e., minimizing ME) by the application of more sensitive methods might be a solution for the inconsistency of many FA results (Beasley et al. 2013, Merilä & Björklund 1995).

In our study, we compare one commonly used manual method and a computerized approach in terms of measurement accuracy and FA outcome. We considered two bilateral metric traits (femur and radio–ulna length) of the European Common Frog, *Rana temporaria* Linnaeus, 1758, and compared levels of ME and FA from external calliper measurements with those from non-destructive micro-3D-computed tomography (µCT) based skeletal measurements. We aim at clarifying the applicability and particularly the reliability of these two methods for analysing developmental instability. In addition to our own data collection and analyses, we also summarized the currently available literature on FA in amphibians in order to get an overview of methods applied and detect possible method dependent FA outcome (see Appendix A for full summary).

Materials and methods

Samples

The correlation between environmental stress and FA is believed to be particularly pronounced in amphibians due to their physiology, semi-permeable skin, and mostly biphasic life cycles, which results in high susceptibility to environmental disturbances (Duellman & Trueb 1994, Ouellet et al. 1997, Wright & Zamudio 2002). As such, amphibians represent a very suitable model system for investigating FA variation between different methods. For our study twenty intact (no broken bones) adult ethanol-preserved specimens of *Rana temporaria* from the Berlin-Brandenburg region, Germany, were obtained from the collection of the Museum für Naturkunde Berlin (ZMB; Supplementary Table S1). Only adults were included in our study. A Snout–vent length of at least 5 cm was used as a criterion to define adults (Dittrich et al. 2018, Miaud et al. 1999).

Skeletal measurements with µCT

For µCT scanning, whole preserved frogs were removed from ethanol, wrapped in bubble wrap, and transferred to a dry plastic tube. Images were generated using a Phoenix|X-ray nanotom of the company GE Sensing & Inspection Technologies GmbH at 90 kV and 150 µA with fast scan settings for upper and lower body scans, acquiring 1000 projections per scan. Effective voxel size ranged between 19–21 µm for each scan. Volumetric reconstructions were made in Datos|x-reconstruction software (GE Sensing & Inspection Technologies GmbH, Wunstorf, Germany). The femora (from the medial condyle to the femur head) and radio–ulnae (from the olecranon process to the styloid process of the ulna) of right and left side of each individual were measured in VG Studio Max 3.0 with the distance measurement tool (Fig. 1 A–C) resulting in 40 measurements of femora and 40 measurements of radio–ulnae (20 for each side) across all 20 individuals.

External measurements with a calliper

Right and left forearm (from the flexed elbow to the base of the outer palmar tubercle) and thigh (distance from the vent to the knee) lengths (Watters et al. 2016) of the same specimens of *Rana temporaria* were measured externally with a digital calliper (INSIZE, code 1139, resolution 0.01 mm) (Fig. 1 D) resulting in 40 measurements of forearms and 40 measurements of thighs (20 for each side) across all 20 individuals. The digital calliper was zeroed after each measurement.

Fluctuating asymmetry and measurement error

Deviations from symmetry are often so small that they hardly exceed the magnitude of ME. In order to detect meaningful variations in FA between groups of interest, it is therefore essential to first assess ME for every single group and subtract it from the asymmetry mean square (Graham et al. 2010). This can be done by a two-way, mixed model ANOVA procedure (Lens et al. 2002a, Merilä & Björklund 1995, Palmer & Strobeck 1986). For this purpose, all measurements were taken twice by the same observer (SN), at different times (minimum period between repeated measures: one day) and without access to the first measurement when taking the values the second time. Thus, in total 80 measurements of each trait per method were obtained (2 per trait; 40 for each side; Supplementary Table S1). Following the protocol of Palmer and Strobeck (2003b) obvious ME outliers that were significantly greater than expected due to sampling error were identified using the Grubb's test (Grubbs & Beck 1972), which tests if an observation deviates significantly from the sample mean. However, because the difference between first and second measurement should be zero, we also compared deviations from zero. The significance level for the Grubb's test was set to p = 0.0125 after Bonferroni correction for multiple tests ($n_{Groups} = 4$). The Grubb's test led to the exclusion of one radio–ulna measured by calliper and one radio–ulna measured by µCT. Thus, the final sample size for the ME analyses was n = 40 for femora per method and n = 38 for radio–ulnae per method. All statistical analyses were performed using R (Version 3.5.0; R Core Team 2018). The significance level was set to p = 0.05 unless otherwise specified.

We then applied a two-way, mixed-model ANOVA to the repeated measurements for each trait and each method separately (R package 'lme4'; Bates et al. 2015). The main fixed factor was side (S), which had two levels (left and right). The random factor was individual (I) and the side by individual interaction (S x I) was a mixed effect. Finally, an error term (err) represented measurement error (replications within side by individual); p-values for the fixed factors were obtained by applying R package 'lmerTest' (Kuznetsova et al. 2017). Significance in the factor side indicates the presence of directional asymmetry and thus would interfere with unbiased interpretation of developmental instability (Palmer & Strobeck 1986). The mean square of the individual by side interaction is a measure of fluctuating asymmetry and antisymmetry including measurement error. To get unbiased estimates of fluctuating asymmetry we extracted the variance components (σ^2_m, $\sigma^2_{S\,x\,I}$, σ^2_I) from the random effects of the mixed model ANOVAs (Graham et al. 2010) and calculated signal (FA)-to-noise (ME) ratios (Knierim et al. 2007). The variance component for individuals (σ^2_I) is an estimate of the size variation among individuals. The variance component for the interaction effect ($\sigma^2_{S\,x\,I}$) is an estimate of the nondirectional asymmetry (fluctuating asymmetry and antisymmetry). The variance component for replicates (σ^2_{err}) is an estimate of ME. Between groups comparisons with additional tests should only be done, if within groups levels of FA exceed within group levels of ME. To estimate the contribution of ME to measured phenotypic variation and repeatability of FA results the $MS_{S\,x\,I}$ (mean squares of the side by individual interaction) and MS_{err} (mean squares of the variance of the repeated measurements [error]) from the two-way mixed model ANOVA were taken to calculate specific FA and ME indices (Palmer & Strobeck 2003a): FA excluding ME in [mm] (FA10a); ME3 expressing ME as a percentage of the total nondirectional asymmetry including ME (MS_{err} as % of $MS_{S\,x\,I}$); ME5 expressing ME as a repeatability coefficient that did not describe ME directly, but rather expressed FA variation as a proportion of the total between sides variation, which includes ME. The larger the repeatability, the

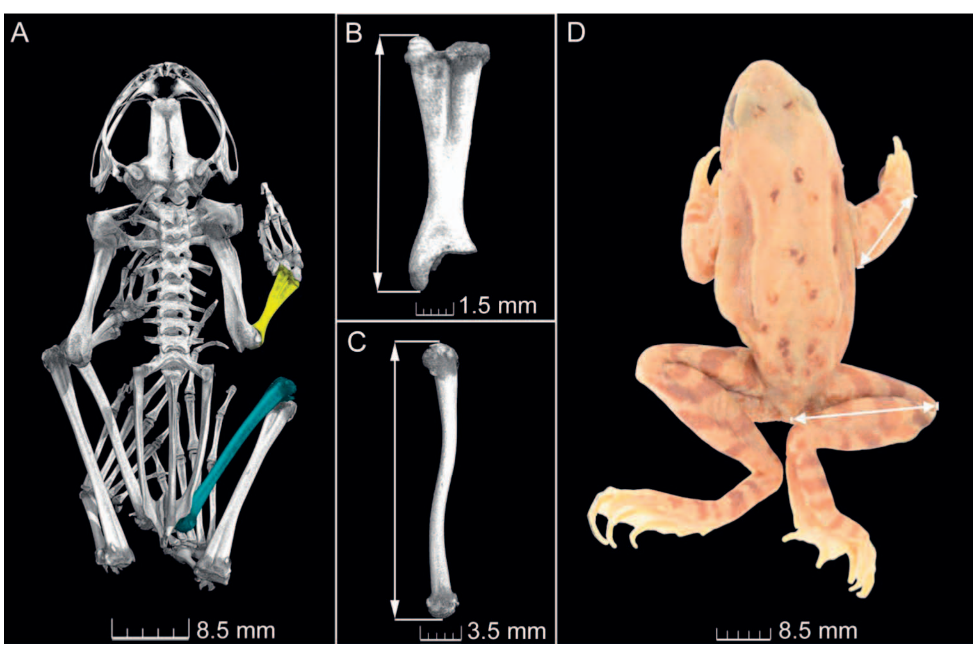

Figure 1. Traits measured for fluctuating asymmetry assessment of the European Common Frog, *Rana temporaria* (male, ZMB 87968). (A) 3D-µCT scan of entire body with (B) right radio-ulna and (C) right femur bone; D) external measures of right radio-ulna and femur.

smaller the ME is relative to FA. ME1 reporting ME in the original units of measurement as the average difference between the repeated measurements in [mm]; FA including ME in [mm] (FA4a), ME1 as a percentage of FA4a and FA1 (averaged replicates of |R-L|) mean ± SE in [mm]. From each replicated measurement the average was calculated, which was then used for further analyses. The final sample sizes for the subsequent FA analyses were n = 20 for femora per method and n = 19 for radio–ulnae per method.

To avoid misinterpretation of the FA results, trait-size dependency was tested by Spearman's rank correlation between absolute values of averaged FA replicates (averaged |R-L| = FA1) and trait size (averaged (R+L)/2) as an independent variable for each trait and each method. The absence of antisymmetry was validated by examining the frequency distributions of the averaged signed FA replicates visually for normality and by using the Anscombe-Glynn kurtosis test and D'Agostino skewness test (R package 'moments'; Komsta & Novomestky 2015). The significance level for the tests of other types of asymmetry as well as for trait-size dependency was set to p = 0.0125 after Bonferroni correction for multiple tests (n_{Groups}= 4). To determine whether the levels of ME (|M1-M2| = absolute values of Measurement 1 – Measurement 2) differed between the two methods and morphological traits, we used a mixed model ANOVA with the method by trait interaction as fixed effect and individual with random intercept and random slopes for the method by trait interaction as random effect. Post-hoc analysis was done by conducting multiple pairwise comparisons of the estimated marginal means with Tukey-adjustment (R package 'emmeans'; Lenth 2018). Absolute values of FA measurements (|R-L|) assessed with the two different methods were compared for the radio–ulna by Bland-Altman analysis (Bland & Altman 1999), adjusted for repeated measurements. Due to proportional bias hyperbolic confidence limits and prediction intervals around the line of best fit of differences on averages were constructed by correlated bivariate least square regression (Ludbrook 2010) (R package 'BivRegBLS'; Francq & Berger 2017). We set a difference of 0.1 mm between methods as acceptance interval. Finally, to assess the effect of method on mean absolute FA (averaged replicates of |R-L|) outcome we performed a mixed model ANOVA with method as fixed factor and individual as random factor. To control for measurement error we also included mean absolute ME as a fixed continuous covariate into the model. The interaction between the fixed factor and the continuous covariate was dropped for better main effect estimation because it turned out to be insignificant by preliminary model selection tests.

Results

Fluctuating asymmetry validation

Measurement errors of two radio–ulnae (one measured using a calliper, one using µCT) from two different specimens were significant outliers when compared to the mean, with one of these even if compared to zero (Grubb's critical value for n = 40 after Bonferroni correction (n_{Groups} = 4): 3.24 (Grubbs & Beck 1972); t_G (mean) for $outlier1_{(radio–ulna\_calliper)}$ = 3.46, $p_{2\text{-}tail} < 0.01$; t_G (zero) for $outlier1_{(radio–ulna\_calliper)}$ = 3.54, $p_{2\text{-}tail} < 0.01$; t_G (mean) for $outlier2_{(radio–ulna\_\mu CT)}$ = 3.59, $p_{2\text{-}tail} < 0.01$; t_G (zero) for $outlier2_{(radio–ulna\_\mu CT)}$ = 3.14, $p_{2\text{-}tail} < 0.025$). For this reason, radio–ulnae' measurements from these two specimens were excluded from subsequent analyses for the respective method. For the femora measurements, no outliers were identified.

The four single two-way mixed model ANOVAs revealed insignificance in the fixed factor side in all cases, indicating the absence of directional asymmetry (Table 1). Frequency distributions of the signed FA values (averaged replicates of (R+L)) of both traits measured by µCT appeared roughly normal, whereas the distributions of those values measured with a calliper appeared left skewed (Supplementary Fig. S1). Neither skewness nor kurtosis did approach statistical significance after Bonferroni correction (Table 2). So neither directional asymmetry nor antisymmetry was evident in these data. Insignificant Spearman's rank correlation showed that the absolute values of FA (averaged replicates of |R-L| = FA1) of both traits did not depend on trait size (averaged replicates of (R+L)/2) for either method (Table 2).

Method and trait dependent fluctuating asymmetry and measurement error

The mixed model ANOVAs revealed significance in the side by individual interaction term for both traits and both methods (Table 1). However, variance components (σ^2) of the mixed effects revealed that FA estimates ($\sigma^2_{S \times I}$) were barely higher than ME estimates (σ^2_{err}) for measurements taken by the µCT, consequently resulting in low signal (FA)-to-noise (ME) ratios. For femora measured with a calliper σ^2_{err} was about twice as high as $\sigma^2_{S \times I,}$ resulting in the lowest signal-to-noise ratio. The highest signal-to-noise ratio was achieved by radio–ulnae calliper measurements, but direct comparisons of σ^2_{err} with µCT measurements revealed a σ^2_{err} for calliper measurements about four times higher (Table 1). The descriptors derived from the ANOVA results varied between the methods (Table 3). FA excluding ME (FA10a) was higher for calliper measurements than for µCT-based measurements for both traits. The percentage of ME to the total non-directional asymmetry (ME3) was highest in calliper-measured femora as expected given the high error variance (σ^2_{err}), but was very low in calliper-measured radio–ulnae compared to µCT-based measurements. The low value of ME3 in the radio–ulnae measured by callipers was reflected in the high repeatability (ME5). The average difference between the repeated measurements (ME1) accounted for a third of FA4a (FA including ME) for the radio–ulnae measured by calliper against a half for µCT-based measurements. However, ME1 itself for radio–ulnae measured by calliper was about twice as high as the values for µCT-based measurements. Generally, de-

Table 1. Results from the two-way mixed model ANOVAs (side = fixed effect, individual = random effect, side x individual = mixed effect) on untransformed repeated measurements for two traits (femur, radio–ulna) and two methods (μCT, calliper) in *Rana temporaria*. Two ME outliers for radio-ulna (one measured by μCT, the other one by calliper) were excluded from these analyses; *** = $p < 0.001$.

Trait	Method	Source of variation	df	Expected mean squares	Variance component σ^2	Signal : noise ratio
Femur	μCT	Side (S)	1	0.0002	_	
		Individual (I)	19	49.1862***	12.2907	1.10
		Side x Individual (S x I)	19	0.0233***	0.0080	
		Measurement Error (err)	40	0.0073	0.0073	
	Calliper	Side (S)	1	0.5056	_	
		Individual (I)	19	79.7089***	19.7746	0.41
		Side x Individual (S x I)	19	0.6103***	0.1370	
		Measurement Error (err)	40	0.3364	0.3364	
Radio-ulna	μCT	Side (S)	1	0.0036	_	
		Individual (I)	18	10.4784***	2.6170	1.36
		Side x Individual (S x I)	18	0.0103***	0.0038	
		Measurement Error (err)	38	0.0028	0.0028	
	Calliper	Side (S)	1	0.1811	_	
		Individual (I)	18	15.1817***	3.7435	6.07
		Side x Individual (S x I)	18	0.2076***	0.0959	
		Measurement Error (err)	38	0.0158	0.0158	

Table 2. Results of tests for skewness (D'Agostino test), kurtosis (Anscombe-Glynn test), and trait size dependency (Spearman's rank correlation) for two traits (femur, radio–ulna) and two methods (μCT, calliper) in *Rana temporaria*. Two ME outliers for radio–ulna (one measured by μCT, the other one by calliper) were excluded from these analyses (see Supplementary Table S1). Significance level after Bonferroni correction for multiple tests ($n_{Groups} = 4$) was set as $p = 0.0125$; FA1 = |R–L| of the averaged replicate measurements; trait size = (R+L)/2 of the averaged replicate measurements.

Trait	Method	n	FA (R-L)					Corr. FA1 / trait size	
			mean ± SE [mm]	Skewness	p	Kurtosis	p	Spearman's ρ	p
Femur	μCT	20	0.003 ± 0.017	0.555	0.225	3.286	0.355	-0.314	0.178
	Calliper	20	-0.084 ± 0.082	-0.913	0.056	2.410	0.802	-0.098	0.681
Radio-ulna	μCT	19	-0.014 ± 0.012	-0.622	0.185	4.203	0.090	-0.111	0.650
	Calliper	19	0.098 ± 0.052	-1.263	0.013	4.401	0.068	0.098	0.689

Table 3. Descriptors of fluctuating asymmetry (FA) and measurement error (ME) in *Rana temporaria*, derived from the results of the two-way mixed model ANOVAs side = fixed effect, individual = random effect, side x individual = mixed effect) on untransformed repeated measurements for two traits (femur, radio–ulna) and two methods (μCT, calliper). Two ME outliers for radio–ulna (one measured by μCT, the other one by calliper) were excluded from this analysis. $FA10a = 0.798\sqrt{(MS_{S\,x\,I} - MS_{err})}$; ME3 = MS_{err} as a percentage of $MS_{S\,x\,I}$; $ME5 = (MS_{S\,x\,I} - MS_{err})/[MS_{S\,x\,I} + (2-1)MS_{err}]$; $ME1 = 0.798\sqrt{MS_{err}}$; $FA4a = 0.798\sqrt{MS_{S\,x\,I}}$; FA1 = |R – L| of the averaged replicate measurements.

Trait	Method	FA10a [mm]	ME3	Repeatability (ME5)	ME1 [mm]	FA4a [mm]	ME1 as % FA4a	FA1 mean ± SE [mm]
Femur	μCT	0.101	31.37	0.52	0.068	0.122	56.0	0.122 ± 0.019
	Calliper	0.418	55.12	0.29	0.463	0.623	74.2	0.578 ± 0.100
Radio-ulna	μCT	0.069	26.85	0.58	0.042	0.081	51.8	0.075 ± 0.015
	Calliper	0.349	7.61	0.86	0.100	0.364	27.6	0.344 ± 0.070

scriptors of FA and ME differed between traits within the calliper method, whereas results for both traits within the μCT methods were comparable.

This was statistically supported by the results of the subsequent mixed-model ANOVA on levels of measurement error (|ME| = |M1-M2|) (Fig. 2A). ME was significantly different between methods ($F = 26.28$, $p < 0.001$), traits ($F = 19.92$, $p < 0.001$). The interaction between both factors was also significant ($F = 15.66$, $p < 0.001$), meaning that both methods affect the discrepancy in measurement error between both traits to different extents. Pairwise comparisons of the estimated marginal means (emmeans ± SE) with Tukey-adjustment revealed that the |ME| for femora' calliper measurements (0.626 ± 0.102) were significantly higher than |ME| for all other measurements (radio–ulna_calliper 0.135 ± 0.035, $p = 0.001$; femur_µCT 0.090 ± 0.034, $p < 0.001$; radio–ulna_µCT 0.056 ± 0.036, $p < 0.001$), whereas |ME| for µCT-based measurements of both traits and for calliper measurements of radio–ulnae were not significantly different from each other (all $p > 0.42$) (Fig. 2A). Since the ME for femora measured with a calliper was exceptionally high we excluded that trait from further FA analyses. As revealed by the mixed-model ANOVA on fluctuating asymmetry (|FA|= averaged replicates of |R-L| = FA1), FA outcome was significantly different between methods ($F = 7.09$, $p < 0.05$). Comparison of the estimated marginal means showed significantly higher |FA| values for calliper measurements (0.316 ± 0.054) than for µCT measurements (0.103 ± 0.054) (Fig. 2B).

The correlated bivariate least square regression of the differences in absolute FA values (|R-L|) between methods on their respective averages was: Difference = 2.03-0.18 (slope SE ± 0.203, $p < 0.0001$; intercept SE ± 0.053, $p < 0.01$). There was proportional bias present, i.e. the difference in values resulting from two methods increased in proportion to the average values. This is indicated by the significant departure from zero of the slope of the least squares linear regression. For averages of the methods up to 0.12 mm differences between the two methods fell within the acceptance interval of 0.1 mm. Higher average values led to overestimation of absolute FA values by the calliper method (Fig. 3).

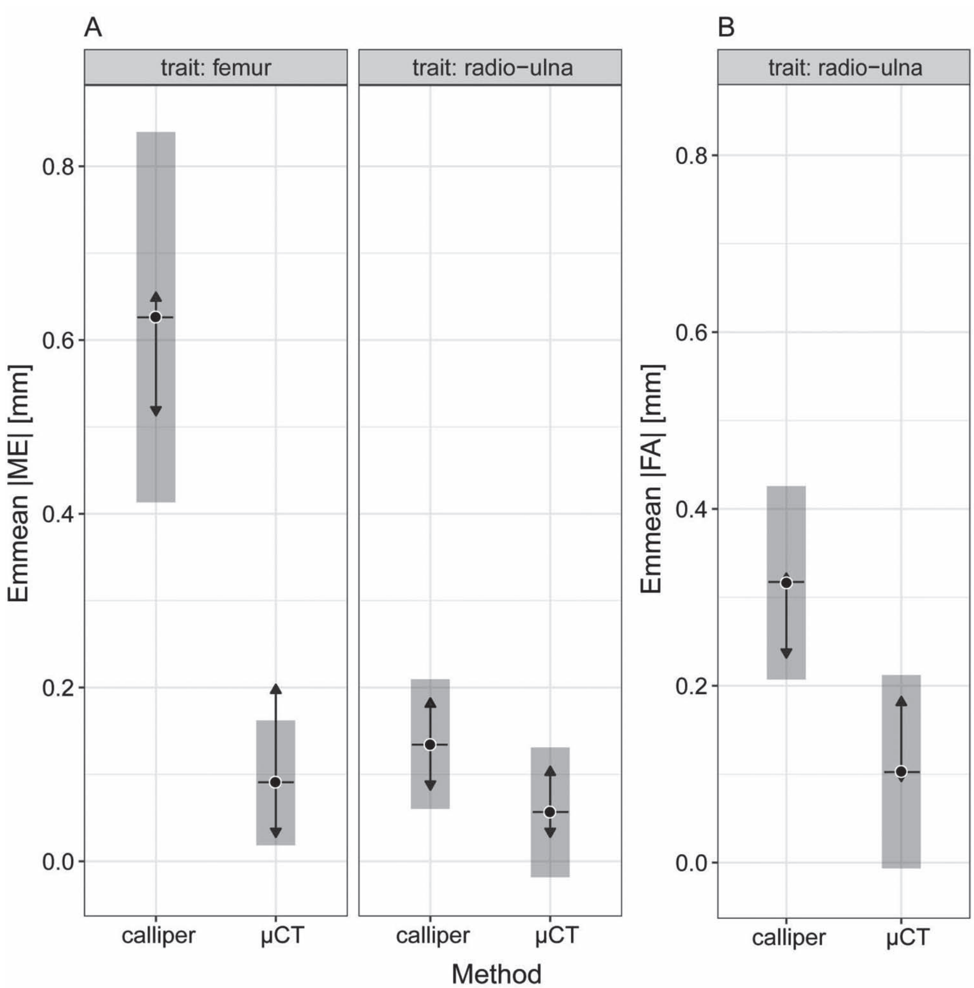

Figure 2. (A) Variations in measurement error |ME| and (B) fluctuating asymmetry |FA| for calliper and µCT measurements of femora and radio-ulnae of *Rana temporaria*; statistically significant differences ($P < 0.05$) are indicated by non-overlapping arrows, grey bars are confidence intervals for the estimated marginal means (emmeans) (mixed-model ANOVAs, for details on calculations and design of the models see text); |ME|: $n_{Femur} = 40$ for each method, $n_{Radio–ulna} = 38$ for each method. |FA|: $n_{Radio–ulna} = 19$ for each method.

Discussion

Our study emphasizes that the outcome of fluctuating asymmetry (FA) analyses is substantially influenced by the method applied. As this has been shown before, it was recommended to rely on the signal-to-noise ratios and repeatability to obtain meaningful FA results (Knierim et al. 2007). In our study, however, signal-to-noise ratio and repeatability, appeared to be adequate for calliper measurements of the radio–ulnae. However, only the direct comparison between methods revealed that ME was higher for calliper measurements. When ME is a sizeable fraction of FA, the confidence in estimates of FA is lowered, even when FA is significantly larger than ME statistically (Palmer & Strobeck 2003b). Although the difference in ME among methods for radio–ulnae was not significant, the difference in results of FA between methods clearly was. An increasing ME artificially inflates FA or obscures FA variation, resulting in false conclusions about developmental instability (Palmer 1994, Palmer & Strobeck 1986). Additionally, ME may differ among traits for several reasons and thereby again lead to the distorted impression of differences in FA among traits. For instance, ME tends to increase with decreasing size of the character for simple allometric reasons (Pankakoski et al. 1987). Likewise, imprecisely defined start- and end-points of measures on a trait or simply the nature of the trait (soft tissue vs. rigid organs) can lead to variation in ME (Van Dongen 2015, Van Nuffel et al. 2007). Furthermore, the accessibility of some traits may be lower than of others and thus impede repeatable positioning of instruments (Knierim et al. 2007). Considering the, albeit not significant, higher ME for calliper measurements the level of associated FA seemed to be inflated also for the radio–ulnae even after correction for ME, and consequently would lead to false interpretations about developmental instability. In addition the skewness in the frequency distributions of signed FA values indicated a tendency towards directional asymmetry in both traits measured with a calliper. This could reflect a bias caused by handedness (Brown & Brown 2002). Measurements on soft tissue of living or dead animals are more prone to handling bias than osteological measurements, because the pressure applied during handling and/or measuring may alter the exact position of the measuring points and consequently induce directional asymmetry (Helm & Albrecht 2000). In a recent study it has been shown that osteological µCT measurements are

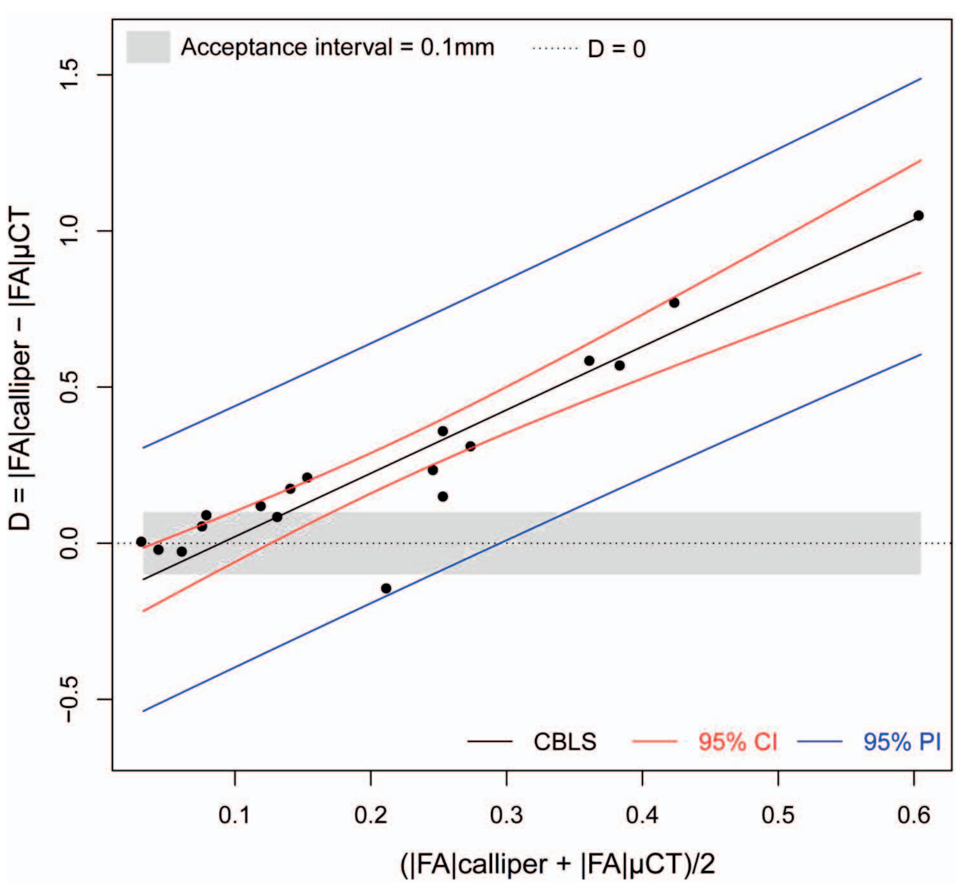

Figure 3. Bland-Altman plot illustrating differences between absolute values of fluctuating asymmetry for radio-ulnae as measured by calliper compared to µCT with hyperbolic confidence limits (Cl) and prediction intervals (PI) around the line of best fit (correlated bivariate least square regression – CBLS); acceptance interval = 0.1 mm; n = 36.

more precise than external manual measurements in detecting sexually dimorphic characters for the same reasons (Pogoda & Kupfer 2018). The difficulty of obtaining repeatable morphometric measurements of external characters is also known from taxonomic studies. Reasons vary from inter-observer ME, over preservation effects, to inconsistent descriptions of anatomical features (Bernal & Clavijo 2009, Stephens et al. 2015, Vervust et al. 2009, Watters et al. 2016).

Only for one trait, the radio–ulna, external calliper measurements could overcome the challenge of achieving a FA caused between-sides variation which exceeded the variation due to measurement error (ME). Ignoring ME of the calliper measurements would create the impression that FA levels were higher in the femora than in the radio–ulnae. Consequently, FA calliper results for femur were unreliable. However, to increase the reliability of FA as a detector of stress, measurements from single traits can also be combined by forming a composite index of FA (CFA) (Leung et al. 2000, Palmer & Strobeck 2003a). This approach is only valid if ME is comparable among traits. Contrary to the calliper method, ME among traits measured by µCT yielded very similar values and could, therefore, be used for multiple traits-analyses.

That µCT was the method with much higher accuracy was most convincingly shown by its low ME. Therefore, we set µCT as reference method for the Bland-Altman analysis, which also substantiated that the calliper measurements overestimated FA values. Furthermore, calliper measures led to proportional bias due to the high measurement values arising from this method. We assume that, in case a high accuracy method fails to detect an effect, but a low accuracy method detects one, the latter result is likely just a methodological artefact. Our sample size was relatively small, but as shown by our analyses as well as in e.g., Muñoz-Muñoz & Perpiñán (2010), still sufficient to detect reliable differences between methods in precision and outcome. This was in particular due to the fact that measurements on exactly the same individuals were opposed with both methods. When ME is relatively high, an alternative to a very sensitive method to increase accuracy, is to increase sample size and/or the number of repeated measurements on each sample. However, when ME is small, smaller sample sizes and less repeats would be sufficient to detect variation of FA between groups of interest. For that reason, sensitive methods, such as µCT, might be of particular interest in cases where the availability of individuals is limited, e.g. threatened species (Van Dongen 1999).

A further benefit of the µCT method arises from its non-destructive character. It makes it suitable for the application to valuable preserved museum specimens, thereby also avoiding measurement error that arises from external measurements. The usage of museum collections provides a great opportunity to compare levels of developmental instability before (baseline) and after an environmental impact (Lens et al. 1999, Schmeller et al. 2011). FA is a relative estimator of developmental instability because there is no standard or reference value of asymmetry that indicates stability. Any conclusions about the level of instability in a given population can only be made by the comparison with a control or reference population (Clarke 1995).To avoid unnecessary disturbance of populations µCT could even be applied to carcasses originated from road kills or predation, at least as the respective characters of interest are undamaged. A huge disadvantage of the µCT-based method, however, is its higher costs compared to the classic calliper method, it is more time-consuming, and only applicable to sacrificed or anaesthetized animals after they have been brought to the lab. Hence it cannot represent a standard technique to monitor e.g., population declines of species in the field (Alford et al. 1999). However, the probability to detect significant FA in living animals in the field seems to be low anyway due to the associated high ME, which limits the use of FA as an indicator for environmental stress and population health in studies with living animals (McCoy & Harris 2003).

In conclusion, our study shows that if fluctuating asymmetry should fulfil its goal to serve as an effective tool in the conservation of amphibians and other endangered animals, results based on calliper measurements of external traits, especially those involving soft tissue, should be treated with caution and if possible, more accurate methods, such as µCT, should be preferred.

Acknowledgements

We thank K. Mahlow for the comprehensive instruction to µCT technology and the use of the imaging software; C. Dittrich and M. Tietje for valuable discussions; F. Tillack for help with access to specimens and respective catalogue data; and one anonymous reviewer for valuable comments. As our study was based on collection material only, no permits were required. This work was funded by the German Federal Ministry of Education and Research BMBF within the Collaborative Project "Bridging in Biodiversity Science – BIBS" (funding number 01LC1501A-H).

References

Alford, R. A., K. S. Bradfield & S. J. Richards (1999): Measuring and analysing developmental instability as a tool for monitoring frog populations. – pp. 34–43 in: Campbell, A. J. (ed.): Declines and disappearances of Australian frogs – Environment Australia, Canberra, ACT.

Anciães, M. & M. Â. Marini (2000): The effects of fragmentation on fluctuating asymmetry in passerine birds of Brazilian tropical forests. – Journal of Applied Ecology, **37**: 1013–1028.

Bates, D., M. Maechler, B. Bolker & S. Walker (2015): Fitting linear mixed-effects models using lme4. – Journal of Statistical Software, **67**: 1–48. R package version 1.1-15. Available from: https://CRAN.R-project.org/package=lme4, cited 21 March 2018.

Beasley, D. E., A. Bonisoli-Alquati & T. A. Mousseau (2013): The use of fluctuating asymmetry as a measure of environmentally induced developmental instability: A meta-analysis. – Ecological Indicators, **30**: 218–226.

Bernal, M. H. & J. A. Clavijo (2009): An essay on precision in morphometric measurements in anurans: Inter-individual and temporal comparisons. – Zootaxa, **2246**: 32–44.

Bland, J. M. & D. G. Altman (1999): Measuring agreement in method comparison studies. – Statistical Methods in Medical Research, **8**: 135–160.

Brown, C. R. & M. B. Brown (2002): Ectoparasites cause increased bilateral asymmetry of naturally selected traits in a colonial bird. – Journal of Evolutionary Biology, **15**: 1067–1075.

Clarke, G. M. (1993): Fluctuating asymmetry of invertebrate populations as a biological indicator of environmental quality. – Environmental Pollution, **82**: 207–211.

Clarke, G. M. (1995): Relationships between developmental stability and fitness: Application for conservation biology. – Conservation Biology, **9**: 18–24.

Costa, R. N. & F. Nomura (2015): Measuring the impacts of Roundup Original® on fluctuating asymmetry and mortality in a neotropical tadpole. – Hydrobiologia, **765**: 85–96.

Dittrich, C., A. Rodríguez, O. Segev, S. Drakulić, H. Feldhaar, M. Vences & M.-O. Rödel (2018): Temporal migration patterns and mating tactics influence size-assortative mating in *Rana temporaria*. – Behavioral Ecology, **29**: 418–428.

Duellman, W. E. & L. Trueb (1994): Biology of amphibians. – The Johns Hopkins University Press, Baltimore, Maryland.

Floate, K. D. & P. C. Coghlin (2010): No support for fluctuating asymmetry as a biomarker of chemical residues in livestock dung. – The Canadian Entomologist, **142**: 354–368.

Francq, B. G. & M. Berger (2017): Bivregbls: Tolerance intervals and errors-in-variables regressions in method comparison studies. R package version 1.0.0. – Available from: https://CRAN.R-project.org/package=BivRegBLS, cited 28 June 2018.

Freeman, D. C., J. M. Emlen, J. H. Graham, R. L. Mara, M. Tracy & C. L. Alados (1996): Developmental instability as a bioindicator of ecosystem health – pp. 170–177 in: Barrow, J. R., E. D. Mcarthur, R. E. Sosebee & R. J. Tausch (eds): Proceedings: Shrubland ecosystem dynamics in a changing environment. 1995 May 23–25; Las Cruces, NM. Gen. Tech. Rep. Int-gtr-338. – U.S. Department of Agriculture, Forest Service, Intermountain Research Station, Ogden, Utah.

Graham, J. H., J. M. Emlen, D. C. Freeman, L. J. Leamy & J. A. Kieser (1998): Directional asymmetry and the measurement of developmental instability. – Biological Journal of the Linnean Society, **64**: 1–16.

Graham, J. H., D. C. Freeman & J. M. Emlen (1993): Antisymmetry, directional asymmetry, and dynamic morphogenesis. – Genetica, **89**: 121–137.

Graham, J. H., S. Raz, H. Hel-Or & E. Nevo (2010): Fluctuating asymmetry: Methods, theory, and applications. – Symmetry, **2**: 466–540.

Grubbs, F. E. & G. Beck (1972): Extension of sample sizes and percentage points for significance tests of outlying observations. – Technometrics, **14**: 847–854.

Helm, B. & H. Albrecht (2000): Human handedness causes directional asymmetry in avian wing length measurements. – Animal Behaviour, **60**: 899–902.

Hoffmann, A. A. & R. E. Woods (2003): Associating environmental stress with developmental stability: Problems and patterns. – pp. 387–401 in: Polak, M. (ed.): Developmental instability: Causes and consequences – Oxford University Press, Oxford, England.

Jones, J. S. (1987): An asymmetrical view of fitness. – Nature, **325**: 298–299.

Knierim, U., S. Van Dongen, B. Forkman, F. A. M. Tuyttens, M. Špinka, J. L. Campo & G. E. Weissengruber (2007): Fluctuating asymmetry as an animal welfare indicator – a review of methodology and validity. – Physiology and Behavior, **92**: 398–421.

Komsta, L. & F. Novomestky (2015): Moments: Moments, cumulants, skewness, kurtosis and related tests. R package version 0.14. – Available from: https://CRAN.R-project.org/package=moments, cited 21 March 2018.

Kuznetsova, A., P. B. Brockhoff & R. H. B. Christensen (2017): lmerTest package: Tests in linear mixed effect models. – Journal of Statistical Software, **82**: 1–26. R package version 20–36. – Available from: https://CRAN.R-project.org/package=lmerTest, cited 21 March 2018.

Lazić, M. M., A. Kaliontzopoulou, M. A. Carretero & J. Crnobrnja-Isailović (2013): Lizards from urban areas are more asymmetric: Using fluctuating asymmetry to evaluate environmental disturbance. – PloS One, **8**: e84190.

Leamy, L., C. Klingenberg, E. Sherratt, J. Wolf & J. Cheverud (2015): The genetic architecture of fluctuating asymmetry of mandible size and shape in a population of mice: Another look. – Symmetry, **7**: 146–163.

Leamy, L. J. & C. P. Klingenberg (2005): The genetics and evolution of fluctuating asymmetry. – Annual Review of Ecology Evolution and Systematics, **36**: 1–21.

Leary, R. F. & F. W. Allendorf (1989): Fluctuating asymmetry as an indicator of stress: Implications for conservation biology. – Trends in Ecology & Evolution, **4**: 214–217.

Lens, L. & H. Eggermont (2008): Fluctuating asymmetry as a putative marker of human-induced stress in avian conservation. – Bird Conservation International, **18**: S125–S143.

Lens, L. & S. Van Dongen (2000): Fluctuating and directional asymmetry in natural bird populations exposed to different levels of habitat disturbance, as revealed by mixture analysis. – Ecology Letters, **3**: 516–522.

Lens, L., S. Van Dongen, S. Kark & E. Matthysen (2002a): Fluctuating asymmetry as an indicator of fitness: Can we bridge the gap between studies? – Biological Reviews, **77**: 27–38.

Lens, L., S. Van Dongen & E. Matthysen (2002b): Fluctuating asymmetry as an early warning system in the critically endangered Taita Thrush. – Conservation Biology, **16**: 479–487.

Lens, L., S. Van Dongen, C. M. Wilder, T. M. Brooks & E. Matthysen (1999): Fluctuating asymmetry increases with habitat disturbance in seven bird species of a fragmented afrotropical forest. – Proceedings of the Royal Society B: Biological Sciences, **266**: 1241–1246.

Lenth, R. V. (2018): Emmeans: Estimated marginal means, aka Least-Squares Means. R package version 1.2.2. – Available from: https://CRAN.R-project.org/package=emmeans, cited 28 June 2018.

Leung, B., M. R. Forbes & D. Houle (2000): Fluctuating asymmetry as a bioindicator of stress: Comparing efficacy of analyses involving multiple traits. – The American Naturalist, **155**: 101–115.

LUDBROOK, J. (2010): Confidence in altman-bland plots: A critical review of the method of differences. – Clinical and Experimental Pharmacology and Physiology, **37**: 143–149.

LUDWIG, W. (1932): Das Rechts-Links Problem im Tierreich und beim Menschen. – Springer, Berlin.

MARCHAND, H., G. PAILLAT, S. MONTUIRE & A. BUTET (2003): Fluctuating asymmetry in bank vole populations (Rodentia, Arvicolinae) reflects stress caused by landscape fragmentation in the Mont-Saint-Michel Bay. – Biological Journal of the Linnean Society, **80**: 37–44.

MATHER, K. (1953): Genetical control of stability in development. – Heredity, **7**: 297–336.

MCCOY, K. A. & R. N. HARRIS (2003): Integrating developmental stability analysis and current amphibian monitoring techniques: An experimental evaluation with the salamander *Ambystoma maculatum*. – Herpetologica, **59**: 22–36.

MERILÄ, J. & M. BJÖRKLUND (1995): Fluctuating asymmetry and measurement error. – Systematic Biology, **44**: 97–101.

MIAUD, C., R. GUYÉTANT & J. ELMBERG (1999): Variations in life-history traits in the common frog *Rana temporaria* (Amphibia: Anura): A literature review and new data from the French Alps. – Journal of Zoology, **249**: 61–73.

MØLLER, A. P. & J. P. SWADDLE (1997): Asymmetry, developmental stability and evolution. – Oxford University Press, New York.

MUÑOZ-MUÑOZ, F. & D. PERPIÑÁN (2010): Measurement error in morphometric studies: Comparison between manual and computerized methods. – Annales Zoologici Fennici, **47**: 46–56.

OUELLET, M., J. BONIN, J. RODRIGUE, J.-L. DESGRANGES & S. LAIR (1997): Hindlimb deformities (ectromelia, ectrodactyly) in free-living anurans from agricultural habitats. – Journal of Wildlife Diseases, **33**: 95–104.

PALMER, A. R. (1994): Fluctuating asymmetry analyses: A primer. – pp. 335–364 in: MARKOW, T. A. (ed.): Developmental instability: Its origins and evolutionary implications – Kluwer, Dordrecht, Netherlands.

PALMER, A. R. & C. STROBECK (1986): Fluctuating asymmetry: Measurement, analysis, patterns. – Annual Review of Ecology and Systematics, **17**: 391–421.

PALMER, A. R. & C. STROBECK (1992): Fluctuating asymmetry as a measure of developmental stability: Implications of non-normal distributions and power of statistical tests. – Acta Zoologica Fennica, **191**: 57–72.

PALMER, A. R. & C. STROBECK (2003a): Fluctuating asymmetry analysis revisited. – pp. 279–319 in: POLAK, M. (ed.): Developmental instability: Causes and consequences – Oxford University Press, Oxford, England.

PALMER, A. R. & C. STROBECK (2003b): Fluctuating asymmetry analysis: A step-by-step example [electronic appendix V in: Fluctuating asymmetry analyses revisited]. – pp. 279–319 in: POLAK, M. (ed.): Developmental instability: Causes and consequences – Oxford University Press, Oxford, England.

PANKAKOSKI, E., R. A. VAISANEN & K. NURMI (1987): Variability of muskrat skulls: Measurement error, environmental modification and size allometry. – Systematic Zoology, **36**: 35.

POGODA, P. & A. KUPFER (2018): Flesh and bone: An integrative approach towards sexual size dimorphism of a terrestrial salamander (genus *Salamandrina*). – Journal of Morphology, **279**: 1468–1479.

R Core Team (2018): R: A language and environment for statistical computing. – R Foundation for Statistical Computing, Vienna, Austria. R version 3.5.0. – Available from: https://www.R-project.org/, cited 4 July 2018.

SARRE, S. (1996): Habitat fragmentation promotes fluctuating asymmetry but not morphological divergence in two geckos. – Researches on Population Ecology, **38**: 57–64.

SCHMELLER, D. S., M. DOLEK, A. GEYER, J. SETTELE & R. BRANDL (2011): The effect of conservation efforts on morphological asymmetry in a butterfly population. – Journal for Nature Conservation, **19**: 161–165.

SÖDERMAN, F., S. VAN DONGEN, S. PAKKASMAA & J. MERILÄ (2007): Environmental stress increases skeletal fluctuating asymmetry in the Moor Frog *Rana arvalis*. – Oecologia, **151**: 593–604.

STEPHENS, R. B., K. H. KARAU, C. J. YAHNKE, S. R. WENDT & R. J. ROWE (2015): Dead mice can grow – variation of standard external mammal measurements from live and three postmortem body states. – Journal of Mammalogy, **96**: 185–193.

VALENTINE, D. W., M. E. SOULÉ & P. SAMOLLOW (1973): Asymmetry analysis in fishes: A possible statistical indicator of environmental stress. – Fishery Bulletin, **71**: 357–370.

VAN DONGEN, S. (1999): Accuracy and power in fluctuating asymmetry studies: Effects of sample size and number of within-subject repeats. – Journal of Evolutionary Biology, **12**: 547–550.

VAN DONGEN, S. (2006): Fluctuating asymmetry and developmental instability in evolutionary biology: Past, present and future. – Journal of Evolutionary Biology, **19**: 1727–1743.

VAN DONGEN, S. (2015): Variation in measurement error in asymmetry studies: A new model, simulations and application. – Symmetry, **7**: 284–293.

VAN NUFFEL, A., F. A. M. TUYTTENS, S. VAN DONGEN, W. TALLOEN, E. VAN POUCKE, B. SONCK & L. LENS (2007): Fluctuating asymmetry in broiler chickens: A decision protocol for trait selection in seven measuring methods. – Poultry Science, **86**: 2555–2568.

VAN VALEN, L. (1962): A study of fluctuating asymmetry. – Evolution, **16**: 125–142.

VERVUST, B., S. VAN DONGEN & R. VAN DAMME (2009): The effect of preservation on lizard morphometrics – an experimental study. – Amphibia-Reptilia, **30**: 321–329.

WADDINGTON, C. H. (1957): The strategy of the genes. – Macmillan, New York.

WATTERS, J. L., S. T. CUMMINGS, R. L. FLANAGAN & C. D. SILER (2016): Review of morphometric measurements used in anuran species descriptions and recommendations for a standardized approach. – Zootaxa, **4072**: 477–495.

WRIGHT, A. N. & K. R. ZAMUDIO (2002): Color pattern asymmetry as a correlate of habitat disturbance in Spotted Salamanders (*Ambystoma maculatum*). – Journal of Herpetology, **36**: 129–133.

ZAKHAROV, V. M., E. PANKAKOSKI, B. I. SHEFTEL, A. PELTONEN & I. HANSKI (1991): Developmental stability and population dynamics in the common shrew, *Sorex araneus*. – The American Naturalist, **138**: 797–810.

Supplementary material

1 Supplementary Figure and 1 Supplementary Table:

Figure S1. Frequency distributions of averaged replicate measurements of signed FA (R-L) for each trait and each method.

Table S1. Raw data and background information for each *Rana temporaria* individual used in the study.

Appendix A

State of the art: Fluctuating asymmetry in amphibians

For an overview concerning FA in amphibians, we searched the 'Web of Science' (Web of Knowledge, Berlin, 10–16-2017) for articles using a minimum combination of two of the following keywords: 'fluctuating asymmetry' and 'amphibian*', 'developmental *stability' and 'amphibian*'. We also searched the reference lists of the selected articles for additional studies that met our inclusion criteria. Our search included studies that: (1) used amphibian species, (2) addressed the question: 'does a certain environmental stressor affect FA?', and (3) measured FA in bilateral metric (i.e. measurable distance) traits. In total, we found 23 publications analysing fluctuating asymmetry (FA) in bilateral metric traits of amphibians. Although it has been indicated before that measurements of morphological characters with manual methods are often imprecise (Muñoz-Muñoz & Perpiñán 2010, Van Nuffel et al. 2007), especially when applied on external instead of skeletal characters (Tuyttens et al. 2005), only three of these studies used computerized techniques on skeletal traits. Five studies used computerized methods but on external traits. Two studies used microscopes on external traits. The remaining 13 studies used calliper measurements on external traits leading to inconsistent results regarding the association of FA and environmental stress (summarized in Table Appendix A).

As apparent from the overview, six of 23 studies found a positive association between the degree of fluctuating asymmetry with environmental stress. However, nine studies did not reveal such association and the remaining eight studies found either a positive, a negative or no association depending on the investigated trait. Altogether there were 17 cases that did not detect a positive association of the degree of fluctuating asymmetry with environmental stress. In seven of these 17 cases, this was due to high measurement error (ME), directional asymmetry, kurtosis or other unfulfilled preconditions to detect FA. Six of these seven studies used external calliper measurements. In three out of the 14 remaining cases, where a positive correlation was found, ME was not assessed (or reported), thereby leaving the interpretation of the results questionable.

Our literature summary underlines the dependency of FA outcome on trait choice and highlights the inconsistency in results related to calliper measurements. Despite the weaknesses of manual measures taken from external characters, this is still the most commonly applied approach.

Table Appendix A. Summary of publications dealing with the effects of environmental stress on fluctuating asymmetry in bilateral metric traits of amphibians.

Taxon	Trait	Location	Method	Measurement	Stressor	Positive correlation of FA with stress	Reason for lack of correlation	Notes	Reference
Litoria nannotis, Litoria genimaculata	hind limbs, forelimbs	external	calliper	manual	rising temperature	yes			(Alford et al. 2007)
Pelophylax perezi	humerus, radio-ulna, metatarsal, CFA	skeletal	X-ray	computer	habitat alteration	yes			(Burghelea et al. 2013)
	tibio-fibula					no	no differences		
Physalaemus cuvieri	nostril-snout distance, eye width	external	images	computer	pesticides	yes			(Costa & Nomura 2015)
Eleutherodactylus antillensis, Eleutherodactylus coqui	femur, tibio-fibula, radio-ulna	skeletal	X-ray	computer	urbanization, habitat alteration, agriculture	no	no differences		(Delgado-Acevedo & Restrepo 2008)
Bufo americanus, Hyla chrysoscelis	eye width, eye-nostril distance, radio-ulna, tibio-fibula, calcaneum	external	images	computer	toxicant (nitrate)	no	no differences		(Earl & Whiteman 2009)
Hyla chrysoscelis	eye width, eye-nostril distance	external	images	computer	toxicant (phosphate)	no	no differences		(Earl & Whiteman 2010)
Physalaemus cuvieri	digit	external	calliper	manual	urbanization	yes			(Eisemberg & Bertoluci 2016)
	femur, tibio-fibula, radio-ulna					no	measurement error (ME), directional asymmetry (DA), kurtosis		
Bokermannohyla saxicola	eye-nostril distance	external	microscope	manual	urbanization, habitat alteration, agriculture	no	no differences		(Eterovick et al. 2015)

Taxon	Trait	Location	Method	Measurement	Stressor	Positive correlation of FA with stress	Reason for lack of correlation	Notes	Reference
Bokermannohyla saxicola	femur, tibio-fibula, radio-ulna, eye-nostril distance	external	calliper	manual	urbanization, habitat alteration, agriculture	no	ME, DA, no differences	indication of correlation of FA with heterozygosity	(Eterovick et al. 2016)
Rana pipiens	deformed radio-ulna, and normal radio-ulna, tibio-fibula, femur	external	calliper	manual	agriculture	no	DA, kurtosis, size dependence, no differences		(Gallant & Teather 2001)
	deformed femur, tibio-fibula					yes		but DA, kurtosis, and size dependence	
Bufo bufo	forearm, tibia	external	calliper	manual	agriculture	yes		no recording of ME	(Guillot et al. 2016)
	parotid gland length and width					no	no recording of ME or other preconditions, no differences		
Crinia signifera	forearm, phalanges, femur, tibio-fibula	external	microscope	manual	habitat alteration	no	negative correlation		(Lauck 2006)
Agalychnis callidryas, Dendropsophus ebraccatus	femur, tibio-fibula	external	calliper	manual	urbanization, habitat alteration, agriculture	no	negative correlation, no differences		(Matías-Ferrer & Escalante 2015)
Ambystoma maculatum	hind limbs (knee to tip of toe), forelimbs (olecranon process to tip of digit)	external	calliper	manual	low pH	no	ME, negative correlation, no differences		(McCoy & Harris 2003)
Bufo fowleri, Hyla chrysoscelis	hind limbs	external	calliper	manual	pathogen	yes		but no recording of ME	(Parris & Cornelius 2004)
Lithobates pipiens	tibio-fibula	external	calliper	manual	habitat alteration	yes			(Reeves et al. 2015)
	radio-ulna, thumb, femur, foot					no	preconditions not fulfilled		
Notophthalmus viridescens	hind limbs	external	images	computer	pathogen	yes			(Sherman et al. 2009)
Crinia signifera	hind limbs, forelimbs	external	calliper	manual	urbanization	yes			(Sievers 2017)
Rana arvalis	femur, tibio-fibula, humerus, radio-ulna	skeletal	microbalance	computer	low pH	yes			(Söderman et al. 2007)
	ilium					no	ME		
Rana clamitans	femur, tibio-fibula, foot, humerus, radio-ulna, thumb, horizontal and vertical tympanum	external	calliper	manual	pathogen	yes		higher levels of FA for individuals infected with *Ranavirus*	(St-Amour et al. 2010)
	femur, tibio-fibula, foot, humerus, radio-ulna, thumb, horizontal and vertical tympanum					no	no differences	no increase of FA levels through fungus *Batrachochytrium dendrobatidis* (*Bd*)	
Rana arvalis	thigh, crus, rostrum, eye, digit, heel tuber	external	calliper	manual	pollution, urbanization	no	neg. correlation, no differences		(Vershinin et al. 2007)
	temporal spot					yes		but only in ethanol fixed individuals	
Litoria wilcoxii/jungguy, Litoria nannotis, Litoria genimaculata, Nyctimystes dayi	tibio-fibula	external	calliper	manual	pathogen	no	no differences		(Woodhams & Alford 2005)
Ambystoma maculatum	spot area	external	images	computer	pesticides, habitat alteration	yes		no recording of ME	(Wright & Zamudio 2002)

References Appendix A

Alford, R. A., K. S. Bradfield & S. J. Richards (2007): Ecology: Global warming and amphibian losses. – Nature, **447**: E3–E4.

Burghelea, C., D. Zaharescu & A. Palanca (2013): Phenotypic indicators of developmental instability in an endemic amphibian from an altered landscape (Monegros, NE Spain). – Amphibia-Reptilia, **34**: 505–516.

Costa, R. N. & F. Nomura (2015): Measuring the impacts of Roundup Original® on fluctuating asymmetry and mortality in a neotropical tadpole. – Hydrobiologia, **765**: 85–96.

Delgado-Acevedo, J. & C. Restrepo (2008): The contribution of habitat loss to changes in body size, allometry, and bilateral asymmetry in two *Eleutherodactylus* frogs from Puerto Rico. – Conservation Biology, **22**: 773–782.

Earl, J. E. & H. H. Whiteman (2009): Effects of pulsed nitrate exposure on amphibian development. – Environmental Toxicology and Chemistry, **28**: 1331–1337.

Earl, J. E. & H. H. Whiteman (2010): Evaluation of phosphate toxicity in Cope's Gray Treefrog (*Hyla chrysoscelis*) tadpoles. – Journal of Herpetology, **44**: 201–208.

Eisemberg, C. C. & J. Bertoluci (2016): Fluctuating asymmetry in populations of the South American frog *Physalaemus cuvieri* (leptodactylidae) in areas with different degrees of disturbance. – Journal of Natural History, **50**: 1503–1511.

Eterovick, P. C., L. F. F. Bar, J. B. Souza, J. F. M. Castro, F. S. F. Leite & R. A. Alford (2015): Testing the relationship between human occupancy in the landscape and tadpole developmental stress. – PloS One, **10**: e0120172.

Eterovick, P. C., B. L. Sloss, J. A. M. Scalzo & R. A. Alford (2016): Isolated frogs in a crowded world: Effects of human-caused habitat loss on frog heterozygosity and fluctuating asymmetry. – Biological Conservation, **195**: 52–59.

Gallant, N. & K. Teather (2001): Differences in size, pigmentation, and fluctuating asymmetry in stressed and nonstressed Northern Leopard Frogs (*Rana pipiens*). – Écoscience, **8**: 430–436.

Guillot, H., A. Boissinot, F. Angelier, O. Lourdais, X. Bonnet & F. Brischoux (2016): Landscape influences the morphology of male Common Toads (*Bufo bufo*). – Agriculture, Ecosystems & Environment, **233**: 106–110.

Lauck, B. (2006): Fluctuating asymmetry of the frog *Crinia signifera* in response to logging. – Wildlife Research, **33**: 313–320.

Matías-Ferrer, N. & P. Escalante (2015): Size, body condition, and limb asymmetry in two hylid frogs at different habitat disturbance levels in Veracruz, México. – Herpetological Journal, **25**: 169–176.

McCoy, K. A. & R. N. Harris (2003): Integrating developmental stability analysis and current amphibian monitoring techniques: An experimental evaluation with the salamander *Ambystoma maculatum*. – Herpetologica, **59**: 22–36.

Muñoz-Muñoz, F. & D. Perpiñán (2010): Measurement error in morphometric studies: Comparison between manual and computerized methods. – Annales Zoologici Fennici, **47**: 46–56.

Parris, M. J. & T. O. Cornelius (2004): Fungal pathogen causes competitive and developmental stress in larval amphibian communities. – Ecology, **85**: 3385–3395.

Reeves, R. A., C. L. Pierce, K. L. Smalling, R. W. Klaver, M. W. Vandever, W. A. Battaglin & E. Muths (2015): Restored agricultural wetlands in Central Iowa: Habitat quality and amphibian response. – Wetlands, **36**: 101–110.

Sherman, E., K. Tock & C. Clarke (2009): Fluctuating asymmetry in *Ichthyophonus*-sp. Infected newts, *Notophthalmus viridescens*, from Vermont. – Applied Herpetology, **6**: 369–378.

Sievers, M. (2017): Sand quarry wetlands provide high-quality habitat for native amphibians. – Web Ecology, **17**: 19–27.

Söderman, F., S. van Dongen, S. Pakkasmaa & J. Merilä (2007): Environmental stress increases skeletal fluctuating asymmetry in the Moor Frog *Rana arvalis*. – Oecologia, **151**: 593–604.

St-Amour, V., T. W. J. Garner, A. I. Schulte-Hostedde & D. Lesbarréres (2010): Effects of two amphibian pathogens on the developmental stability of green frogs. – Conservation Biology, **24**: 788–794.

Tuyttens, F. A. M., L. Maertens, E. Van Poucke, A. Van Nuffel, S. Debeuckelaere, J. Creve & L. Lens (2005): Measuring fluctuating asymmetry in fattening rabbits: A valid indicator of performance and housing quality? – Journal of Animal Science, **83**: 2645–2652.

Van Nuffel, A., F. A. M. Tuyttens, S. Van Dongen, W. Talloen, E. Van Poucke, B. Sonck & L. Lens (2007): Fluctuating asymmetry in broiler chickens: A decision protocol for trait selection in seven measuring methods. – Poultry Science, **86**: 2555–2568.

Vershinin, V. L., E. A. Gileva & N. V. Glotov (2007): Fluctuating asymmetry of measurable parameters in *Rana arvalis*: Methodology. – Russian Journal of Ecology, **38**: 72–74.

Woodhams, D. C. & R. A. Alford (2005): Ecology of chytridiomycosis in rainforest stream frog assemblages of Tropical Queensland. – Conservation Biology, **19**: 1449–1459.

Wright, A. N. & K. R. Zamudio (2002): Color pattern asymmetry as a correlate of habitat disturbance in Spotted Salamanders (*Ambystoma maculatum*). – Journal of Herpetology, **36**: 129–133.