6-2016

Modeling Observed Animal Performance Using the Weibull Distribution

Travis J. Hagey

Jonathan B. Puthoff

Kristen E. Crandell

Kellar Autumn

Luke J. Harmon

Modeling observed animal performance using the Weibull distribution

Travis J. Hagey[1,2,*], Jonathan B. Puthoff[3], Kristen E. Crandell[4], Kellar Autumn[3] and Luke J. Harmon[1,5]

ABSTRACT

To understand how organisms adapt, researchers must link performance and microhabitat. However, measuring performance, especially maximum performance, can sometimes be difficult. Here, we describe an improvement over previous techniques that only consider the largest observed values as maxima. Instead, we model expected performance observations via the Weibull distribution, a statistical approach that reduces the impact of rare observations. After calculating group-level weighted averages and variances by treating individuals separately to reduce pseudoreplication, our approach resulted in high statistical power despite small sample sizes. We fitted lizard adhesive performance and bite force data to the Weibull distribution and found that it closely estimated maximum performance in both cases, illustrating the generality of our approach. Using the Weibull distribution to estimate observed performance greatly improves upon previous techniques by facilitating power analyses and error estimations around robustly estimated maximum values.

KEY WORDS: Gecko, Adhesion, Lizard, Bite force, Maximum performance

INTRODUCTION

Studies of ecological morphology and evolution often link an organism's morphology, performance and ecology to suggest adaptation (Wainwright and Reilly, 1994). These studies typically assume accurate measurements of performance, which can be difficult to obtain, sometimes with a few observations contributing to a final point estimate of performance, especially when maximum performance values are of interest (Anderson et al., 2008; Garland and Losos, 1994). For the initial approval of projects by animal use and collection review boards, it is also necessary to plan for the number of individuals to be tested, the testing procedure for each individual and the resulting statistical power of the results. Power analyses can describe the number of trials and individuals needed, but require an explicit statistical distribution of results, which is difficult to estimate when quantifying maximum performance. Here, we describe a modeling approach using the Weibull distribution. In this approach, maximum performance is estimated as a parameter of the Weibull distribution. This approach is more robust than current methods of quantifying performance, which consider only a subset of maximum observations. The Weibull approach minimizes the effect of rare events and allows the use of additional statistical tools.

The Weibull distribution is valid for data that are likely to be non-normal, such as estimates of a maximum value, which would be expected to produce a skewed distribution of observations. While the Weibull distribution has been used occasionally to model behavior (Britten et al., 1992; Davis, 1996; Pugno and Lepore, 2008; Simpson and Ludlow, 1986), it is most often used in material science to predict mechanical failure. In this scenario, the Weibull distribution is modeling the likelihood of an observed event (i.e. mechanical failure) as it relates to some other factor (i.e. applied force or time), but this does not need to be the case. The Weibull distribution is highly flexible and can model many different patterns (Cornwell and Weedon, 2014; McCool, 2012; Weibull, 1951; Yang and Xie, 2003). This differs from the exponential distribution, which assumes a constant event rate and the normal distribution, which can only model non-skewed data. The Weibull distribution includes a scale parameter (λ) and a shape parameter (k, also known as the Weibull modulus). The scale parameter (λ) has the same units as the modeled data and dictates the distribution's location on the x-axis. λ is closely related to the distribution's mode when $k>1$ (Fig. 1). The dimensionless shape parameter (k) loosely controls the width and shape of the distribution for a given λ (Fig. 1). With these parameters, the Weibull distribution can model left and right skewed distributions (when $k>1$) as well as exponential distributions (when $k \leq 1$). Here, we use the λ parameter as our performance value estimate.

To investigate the use of the Weibull distribution, we quantified the adhesive performance of gecko lizards by measuring toe detachment angle (Fig. 2; Autumn et al., 2006; Hagey et al., 2014). We first compared the fit of our observations to multiple distributions. We then considered how to model multiple observations from multiple individuals to avoid pseudoreplication and produce group-level estimates. We also conducted a power analysis to identify the minimum number of trials per individual and number of individuals needed to detect differences between groups. Lastly, we explored the generality of our approach by using the Weibull distribution to analyze previously published lizard bite force data.

MATERIALS AND METHODS

To evaluate the use of the Weibull distribution, we used two empirical toe detachment datasets from the lizard *Gekko gecko* Linnaeus 1758. We used a multi-individual dataset that included 206 observations from 13 individuals with an overall average of 15.4 trials per individual (max: 40 trials per individual; min: four trials per individual). Our second dataset, a subset of the first, comprised 40 observations from a single individual. To quantify toe detachment angle, we suspended live lizards from a single rear center toe (Schulte et al., 2004; Wang et al., 2010; Zani, 2000) from a slowly inverting glass microscope slide, recording the surface angle at which the lizard spontaneously detaches (Fig. 2; Autumn

[1]Department of Biological Sciences, University of Idaho, Moscow, ID 83843, USA. [2]BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824, USA. [3]Biology Department, Lewis and Clark College, Portland, OR 97219, USA. [4]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. [5]Department of Fish Ecology and Evolution, Eawag Swiss Federal Institute of Aquatic Science and Technology, Center for Ecology, Evolution & Biogeochemistry, 6047 Kastanienbaum, Switzerland.

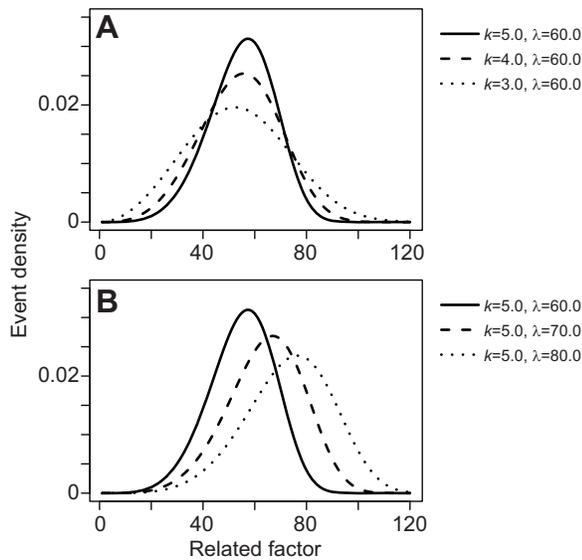*Author for correspondence (tjhagey@uidaho.edu)

Fig. 1. Weibull distribution and parameter variation. (A,B) The shape (*k*) and scale (λ) parameters of the Weibull distribution interact to dictate the shape and location of the distribution. Here, we display relevant parameter combinations illustrating the relationship between varying positive *k* values at constant λ (A) and constant positive *k* value for varying λ (B).

et al., 2006; Emerson, 1991; Hagey et al., 2014). As the glass surface inverts, the probability of detachment increases, making angle of toe detachment an excellent assay to be modeled by the Weibull distribution.

We preliminarily compared the fit of the normal, Weibull, exponential, gamma and log normal distributions qualitatively using *Q–Q* plots (Fig. 3). Previous studies have shown that the Weibull distribution can sometimes be difficult to distinguish from other distributions (Bain and Engelhardt, 1980; Fearn and Nebenzahl, 1991). We also compared the fit of these distributions to our data employing a bootstrapped Kolmogorov–Smirnov (K–S)
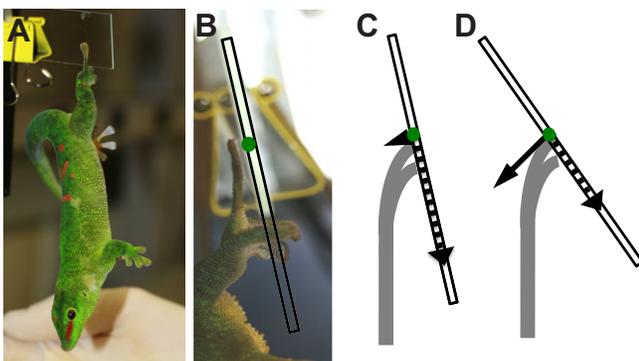

Fig. 2. Angle of toe detachment assay to quantify the adhesive performance of padded lizards. (A,B) In this assay, a lizard is suspended from a glass surface by a single rear toe via their natural adhesive properties. (C) When the surface is near vertical, the lizard is generating predominantly friction relative to the substrate (adhesive microstructures are illustrated in gray, friction as dashed arrows and adhesion as solid arrows). (D) As the substrate is slowly inverted, less friction and more adhesion relative to the substrate is generated. At the angle of toe detachment, the adhesive microstructures can no longer maintain the proper orientation with the surface and the animal falls onto a cushioned base. As a result, the angle of toe detachment quantifies the maximum amount of adhesion, relative to friction an individual is capable of generating (images modified from Hagey et al., 2014).

test with 5000 bootstraps to alleviate issues with repeated values (Table 1; Sekhon, 2011). The K–S test evaluates the probability that our observed values could have been drawn from each considered distribution. We also compared Akaike information criterion weights with a correction for small sample size (AICc; Burnham and Anderson, 2002).

To evaluate multiple observations from multiple individual lizards, we needed to determine how variable our observed individuals were and whether they could be modeled together under one distribution. We fitted the Weibull distribution to our multi-individual dataset (our null model, all observations lumped together) or allowed each individual lizard to have its own set of parameters (our alternative model). We then conducted a likelihood ratio test and used the chi-squared distribution to determine if we should reject our null model. Finally, we fitted a parametric survival regression model using the Weibull distribution with individual as a factor, predicting detachment angle using the R library 'survival' (Therneau and Grambsch, 2000).

We then calculated a weighted group average detachment angle ($\bar{x}$) using Eqn 1, where N is the number of individuals, λ_i is the individual i estimated scale parameter value and w_i is the individual i weighting:

$$\bar{x} = \frac{\sum_{i=1}^{N} (\lambda_i w_i)}{\sum_{i=1}^{N} w_i}. \tag{1}$$

We calculated the weighting (w_i) of each individual using the estimated error σ_i^2 around the scale parameter value λ, using the equation:

$$w_i = \frac{1}{\sigma_i^2}. \tag{2}$$

To calculate the variance around our weighted mean, $\sigma_{\bar{x}}^2$, incorporating within and between individual variation, we used the following equation (Bevington and Robinson, 2003):

$$\sigma_{\bar{x}}^2 = \frac{1}{\sum_{i=1}^{N} (1/\sigma_i^2)} \times \frac{1}{N-1} \times \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{\sigma_i^2}. \tag{3}$$

It is worth noting that when the individual errors (σ_i^2) are all equal, Eqn 3 simplifies to the standard error around an unweighted mean.

Using the above-described analysis approach, we simulated performance data and conducted power analyses to investigate the effect of sample size on our ability to detect differences in detachment angle between groups (Fig. 4). In order to simulate the necessary data, we first needed to obtain realistic *k* and λ values. Owing to the Weibull distribution's heteroscedastic nature, *k* is affected by changes to λ (Fig. 1; McCool, 2012). In addition, *k* is sensitive to sample size, with smaller datasets fitting distributions with larger *k* values i.e. narrower distributions. To confirm the correlated relationship between *k*, λ and sample size, we used a previously collected dataset of toe detachment observations from 53 gecko and anole species (244 individuals with an average of 9 trials per individual). This dataset was collected using similar methods to those described above, including measurements from the lab and field using captive and wild-caught specimens (Autumn et al., 2006; Hagey et al., 2014). We estimated *k* and λ for each individual and fitted these data to a linear model where *k* was predicted by λ, number of trials, and an interaction between λ and the number of trials. We found that both λ (*P*<0.0001) and number of trials (*P*<0.0001) significantly predicted *k*. The interaction term was not
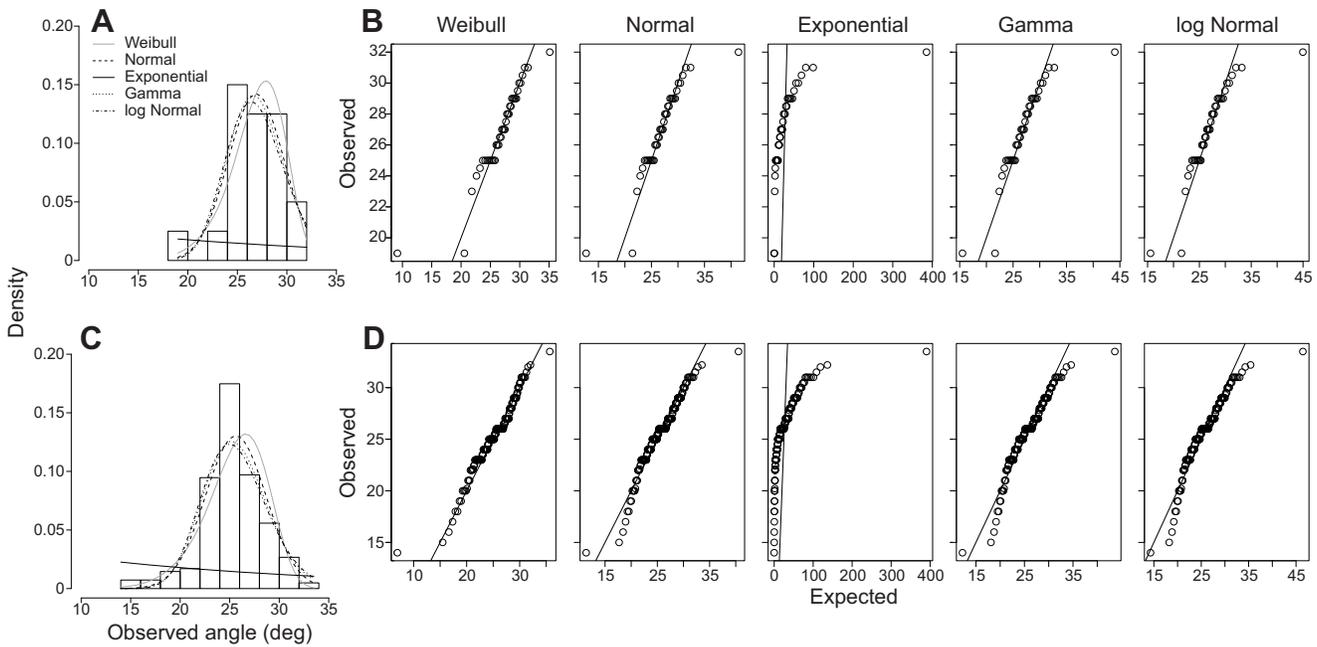
Fig. 3. Fit of empirical data to the Weibull distribution. Observations and fitted distributions from a single individual with 40 trials (A) and a larger dataset containing 206 trials from 13 individuals (C). Q–Q plots of the two toe detachment datasets (B, individual dataset; D, 13 dataset) to determine the qualitative fit of our empirical data to the Weibull distribution. The Q–Q plots also contain a line at y=x as reference.

significant and was removed from the analysis. The coefficients from this linear model were then used to estimate realistic k values for a given λ and number of trials (N):

$$k = 0.6\lambda - 0.53N + 10.2. \qquad (4)$$

We then produced multiple datasets of simulated toe detachment data using the Weibull distribution. Each dataset was composed of simulated data from two groups with different assigned λ parameter values to see if we could significantly differentiate small differences between groups. The differences between groups were one, two, three or five degrees (i.e. our effect sizes). We chose to compare groups with λ values of 15 deg vs 15 deg plus an effect size and 25 deg vs 25 deg plus an effect size based on empirical observations. We also evaluated comparisons in which groups had 20 trials per group distributed across one individual, two individuals and four

individuals or 50 trials per group distributed across two individuals, five individuals and ten individuals (Fig. 4). To statistically compare our estimated weighted averages between simulated groups, we tested whether either of the groups' means were within 1.96 standard deviations (square root of the estimated mean variance multiplied by 1.96) of the other. We then calculated the percentage of our 1000 replicate trials that produced significantly different comparisons (Fig. 4). All analyses were conducted using RStudio statistical software (v.0.98.501; https://cran.rstudio.com).

RESULTS AND DISCUSSION

We qualitatively compared our data with the normal, Weibull, exponential, gamma and log normal distributions (Fig. 3). Our Q–Q plots strongly suggest that our data differ from what would be expected under the exponential distribution because high angle observations were rare. The fitted normal, gamma and log normal distributions all yielded similar results, suggesting that low-angle observations were more common than if our data were drawn from these distributions, i.e. our data are left-skewed, which is to be expected when attempting to observe a maximum value. The Q–Q plots suggest that the Weibull distribution closely approximates our observed data. An important distinction between the Weibull and normal distributions is that the Weibull predicts rare early failure events, producing a negatively skewed distribution of observations, whereas the normal distribution does not. We also evaluated the goodness of fit for normal, exponential, Weibull, gamma and log normal distributions using a bootstrap version of the Kolmogorov–Smirnov test (Sekhon, 2011). Our multi-individual dataset was significantly different from all the distributions considered, whereas our individual dataset was not significantly different to any of the distributions considered, except the exponential distribution (Table 1). These results might be due to the fact that our multi-individual dataset represents a collection of observations from multiple individuals, each with their own performance distribution (see below). Conversely, our individual dataset might be too small

Table 1. Toe detachment angle and model selection

	K–S test P-value	log likelihood	AICc	AICc weight
Individual dataset				
Weibull	0.733	−96.3	196.9	0.81
Normal	0.461	−98.1	200.5	0.14
Exponential	0.000	−171.7	345.5	0.00
Gamma	0.443	−99.5	203.2	0.03
log Normal	0.378	−100.3	204.8	0.02
Multi-individual dataset				
Weibull	0.008	−519.8	1043.7	0.94
Normal	0.009	−522.6	1049.2	0.06
Exponential	0.000	−874.1	1750.1	0.00
Gamma	0.001	−530.5	1065.1	0.00
log Normal	0.000	−535.8	1075.7	0.00

We evaluated multiple distributions to determine which best modeled our observed toe detachment angles using Kolmogorov–Smirnov (K–S) tests and AICc weights. All model distributions evaluated had two free parameters, except the exponential, which had one.
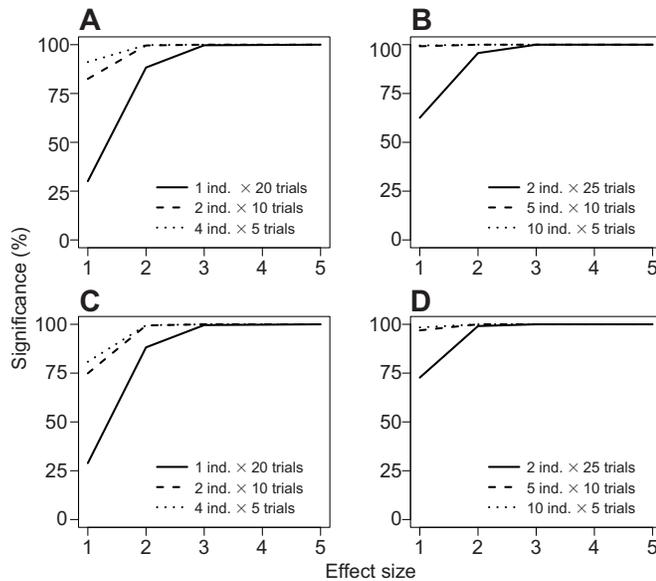
Fig. 4. Power analyses. (A,B) Results for groups with low detachment angles of 15–20 deg and 20 (A) or 50 (B) total trials. (C,D) Results for groups with high detachment angles of 25–30 deg and 20 (C) or 50 (D) total trials. The y-axes represent the percentage of 1000 replications that were significantly different.

to distinguish between distributions (see the Materials and methods). AICc weights, corrected for small sample sizes (Burnham and Anderson, 2002) provided more clear results, suggesting the Weibull distribution is the best fitting distribution considered with an AICc weight of 0.94 and 0.81 for our multi-individual and individual datasets, respectively (Table 1).

To estimate a group-level detachment angle, we needed to consider multiple observations from multiple individuals, while also limiting pseudoreplication. We found significant support to
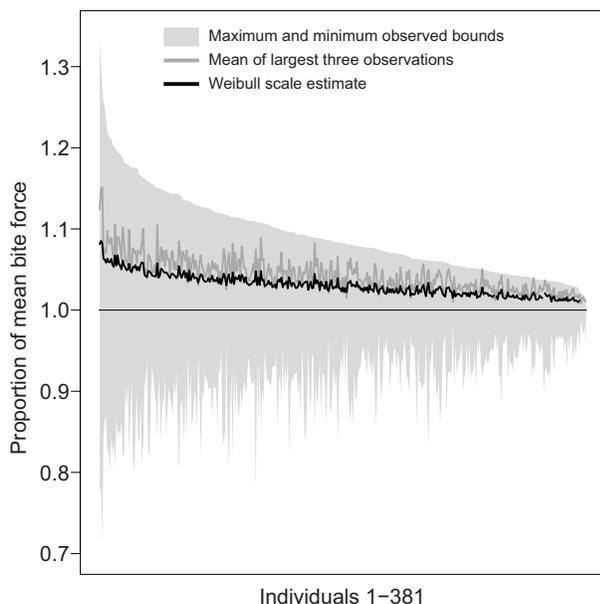

Fig. 5. Lizard bite forces. Using a compiled dataset of published lizard bite forces, we estimated each individual Weibull scale (λ) parameter (black points) and plotted these values with an individual's observed maximum and minimum observation (shaded area) and the mean of the three largest observations (gray points). All calculated values were divided by an individual's mean observation to similarly scale the dataset.

reject our null hypothesis (pooling all our observations, $D=100.3$, d.f.=24, $P<0.0001$). We also fitted a parametric survival regression model using the Weibull distribution in which individual, as a factor, significantly predicted detachment angle ($P<0.0001$), again suggesting it is better to model performance observations of each individual lizard separately (Therneau and Grambsch, 2000). By fitting each individual to a distribution and using each individual's parameter values and error to estimate weighted group averages, we prevent pseudoreplication and reduce the impact of unbalanced sampling, penalizing individuals with large error estimates due to small sample sizes or erratic observations.

We also investigated the trade-off between the number of trials per individual and the number of individuals tested using power analyses. We found abundant power to detect small differences between groups (effect sizes of one to two degrees) even with relatively small datasets, i.e. few total trials, few individuals, or few trials per individual (Fig. 4). In our simulations using 50 observations, we observed an overall increase in power, as to be expected with more data, regardless of partitioning. Fitting the Weibull distribution to datasets with five or fewer observations per individual, especially if there is little variation among the observations, can hinder the estimation of parameter estimates. Although our power analyses suggested we could detect group differences with few trials or individuals, we recommend six to ten trials per individual to assure a successful fit of the Weibull distribution. Using data from two individual lizards with 10 trials each, we consistently had enough power to detect a difference of 1 deg 80% of the time, although we recommend sampling more individuals to better capture inter-individual variation.

In addition, we did not observe differences in power between datasets considering low detachment angles (15 deg) and high detachment angles (25 deg). We believe this is due to the fact that we simulated our data using values of k that complemented λ and the number of trials per individual (see Materials and methods). We conducted similar analyses with a constant k value regardless of λ and number of trials per individual and found that power decreased when considering higher values of λ. This is because at higher λ values, with a constant k, the Weibull distribution has a larger variance (Fig. 1), reducing the power to detect differences between groups.

To investigate the generality of the ability of the Weibull distribution to model animal performance, we also compiled published bite force data from *Anolis carolinensis* (provided by A. Herrel; Irschick et al., 2005a,b). This dataset comprised 381 individuals, with 5 trials for each. We found that an individual's estimated λ values were similar to the mean of the best three trials with the added benefit of being robust to extreme values (Fig. 5).

In conclusion, using the Weibull distribution as an expected distribution of observed performance is a more robust approach compared with use of only a subset of trials. We strongly encourage researchers to investigate the fit of the Weibull distribution to their performance datasets. With the addition of estimating weighted group means, comparisons between treatment and control groups, or comparisons between species are easily facilitated. If making comparisons across species, once species means have been estimated, phylogenetic non-independence can then easily be taken into account.

References

Anderson, R. A., McBrayer, L. D. and Herrel, A. (2008). Bite force in vertebrates: opportunities and caveats for use of a nonpareil whole-animal performance measure. *Biol. J. Linn. Soc.* **93**, 709-720.

Autumn, K., Dittmore, A., Santos, D., Spenko, M. and Cutkosky, M. (2006). Frictional adhesion: a new angle on gecko attachment. *J. Exp. Biol.* **209**, 3569-3579.

Bain, L. J. and Engelhardt, M. (1980). Probability of correct selection of Weibull versus Gamma based on likelihood ratio. *Commun. Stat. A Theory Methods* **9**, 375-381.

Bevington, P. and Robinson, D. K. (2003). *Data Reduction and Error Analysis for the Physical Sciences*. Boston, MA: McGraw-Hill Education.

Britten, K. H., Shadlen, M. N., Newsome, W. T. and Movshon, J. A. (1992). The analysis of visual-motion - a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745-4765.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. New York: Springer Verlag.

Cornwell, W. K. and Weedon, J. T. (2014). Decomposition trajectories of diverse litter types: a model selection analysis. *Methods Ecol. Evol.* **5**, 173-182.

Davis, J. D. (1996). Deterministic and probabilistic control of the behavior of rats ingesting liquid diets. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **270**, R793-R800.

Emerson, S. B. (1991). The ecomorphology of Bornean tree frogs (Family Rhacophoridae). *Zool. J. Linn. Soc.* **101**, 337-357.

Fearn, D. H. and Nebenzahl, E. (1991). On the maximum likelihood ratio method of deciding between the Weibull and Gamma distributions. *Commun. Stat. Theory Methods* **20**, 579-593.

Garland, T. and Losos, J. B. (1994). Ecological morphology of locomotor performance in squamate reptiles. In *Ecological Morphology: Integrative Organismal Biology* (ed. P. C. Wainwright and S. Reilly), pp. 240-302. Chicago: University of Chicago Press.

Hagey, T. J., Puthoff, J. B., Holbrook, M., Harmon, L. J. and Autumn, K. (2014). Variation in setal micromechanics and performance of two gecko species. *Zoomorphology* **133**, 111-126.

Irschick, D. J., Carlisle, E., Elstrott, J., Ramos, M., Buckley, C., Vanhooydonck, B., Meyers, J. and Herrel, A. (2005a). A comparison of habitat use, morphology, clinging performance and escape behaviour among two divergent green anole lizard (*Anolis carolinensis*) populations. *Biol. J. Linn. Soc.* **85**, 223-234.

Irschick, D. J., Vanhooydonck, B., Herrel, A. and Meyers, J. (2005b). Intraspecific correlations among morphology, performance and habitat use within a green anole lizard (*Anolis carolinensis*) population. *Biol. J. Linn. Soc.* **85**, 211-221.

McCool, J. I. (2012). *Using the Weibull Distribution: Reliability, Modeling and Inference*. Hoboken, NJ: Wiley.

Pugno, N. M. and Lepore, E. (2008). Living tokay geckos display adhesion times following Weibull statistics. *J. Adhes.* **84**, 949-962.

Schulte, J. A., Losos, J. B., Cruz, F. B. and Nunez, H. (2004). The relationship between morphology, escape behaviour and microhabitat occupation in the lizard clade Liolaemus (Iguanidae: Tropidurinae: Liolaemini). *J. Evol. Biol.* **17**, 408-420.

Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *J. Stat. Software* **42**, 1-52.

Simpson, S. J. and Ludlow, A. R. (1986). Why locusts start to feed: a comparison of causal factors. *Anim. Behav.* **34**, 480-496.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

Wainwright, P. C. and Reilly, S. M. (1994). *Ecological Morphology*. Chicago, IL: University of Chicago Press.

Wang, Z. Y., Gu, W. H., Wu, Q., Ji, A. H. and Dai, Z. D. (2010). Morphology and reaction force of toes of geckos freely moving on ceilings and walls. *Sci. China Technol. Sci.* **53**, 1688-1693.

Weibull, W. (1951). A statistical distribution function of wide applicability. *J. Appl. Mech. Trans. Asme* **18**, 293-297.

Yang, Z. and Xie, M. (2003). Efficient estimation of the Weibull shape parameter based on a modified profile likelihood. *J. Stat. Comput. Simul.* **73**, 115-123.

Zani, P. A. (2000). The comparative evolution of lizard claw and toe morphology and clinging performance. *J. Evol. Biol.* **13**, 316-325.