Shape Dimensionality Metrics for Landmark Data

F. Robin O'Keefe[1]

1. Department of Biological Sciences, Marshall University, Huntington, WV 25755

Key Words: Geometric morphometrics, effective dimensionality, latent dimensionality, dire wolf, modularity and integration, rank deficiency

Abstract

The study of modularity in geometric morphometric landmark data has focused attention on an underlying question, that of whole-shape modularity, or the pattern and strength of covariation among all landmarks. Measuring whole-shape modularity allows measurement of the dimensionality of the shape, but current methods used to measure this dimensionality are limited in application. This paper proposes a metric for measuring the "effective dimensionality", D_e, of geometric morphometric landmark data based on the Shannon entropy of the eigenvalue vector of the covariance matrix of GPA landmark data. A permutation test to establish null rank deficiency is developed to allow standardization for comparing dimensionality metrics between data sets, and a bootstrap test is employed for measures of dispersion. These novel methods are applied to a data set of 14 landmarks taken from 119 dire wolf jaws from Rancho La Brea. Comparison with the current test based on eigenvalue dispersion demonstrates that the new metric is more sensitive to detecting population differences in whole-shape modularity. The effective dimensionality metric is extended, in the dense semilandmark case, to a measure of "latent dimensionality", D_l. Latent dimensionality should be comparable among landmark spaces, whether they are homologous or not.

The use of landmark data to characterize the shape of biological structure, and to quantify shape change, has become ubiquitous since the introduction of geometric morphometrics by Bookstein (1997). All geometric morphometric techniques begin with a set of landmark data comprising spatial coordinates of homologous points on a series of specimens; the landmarks may be termed variables, and may be of two or three dimensions. Each spatial coordinate is treated as an independent variable in most subsequent analysis, so that the vector of variables **V** is:

$$\mathcal{V} = LM(x, y), or\ LM(x, y, z)$$

These landmark positions are measured on *n* samples from the group, or groups, of interest, subject to the stricture that each specimen has all of the landmarks. While methods exist to impute missing data, inclusion of a subset of specimens that lack a homologous landmark entirely is not tractable methodologically, and subsets are generally compared in the subspace of shared landmarks. As shapes become more different the intersection of these landmark spaces declines, as does the utility of the resulting subspace. Each biological shape space defined by a set of landmarks is therefore unique, and while each space is useful in itself, generalizing among shape spaces is problematic. The objective of this paper is to develop metrics that characterize shape spaces, and that are directly comparable among shape spaces regardless of the specific landmarks comprising each space.

Introduction

Analysis of landmark data generally begins with Generalized Procrustes Analysis (GPA) of the coordinate data, wherein the centroid of each specimen is translated to the origin, and the coordinates are rotated and scaled to a mean shape so that the intra-landmark variance among specimens is minimized. Given a matrix **LM** of *v* landmarks and *n* specimens, a matrix resulting from Procrustes superimposition is:

$$\mathcal{X}_{n,v} = GPA(LM_{n,v})$$

The matrix **X** contains the GPA landmark coordinates and is the same size as **LM**. This procedure also produces a vector of centroid size; this vector is interesting in itself, and can also be used to remove size-correlated shape variation from **X** by taking the residuals of its regression against v_i (Drake and Klingenberg 2010). The desirability of removing size-correlated shape variation prior to subsequent analysis is question specific, and is not considered further here (see Zelditch et al. 2012 for an excellent review of GPA).

The interlandmark distances computed from **X** have utility, but these are seldom utilized on their own. Most subsequent analyses of **X** proceed via Principal Components Analysis (PCA), a factor-analytic expression of the singular value decomposition (SVD), performed on the covariance matrix **K** computed from the variables in **X**;

$$K_{v_i v_j} = cov(v_i v_j)$$

The matrix **K** is square and symmetric and of rank v, with covariances on the off-diagonal and the variances of v_i on the diagonal. The SVD produces a matrix $\mathbf{\Lambda}$ that is equivalent to **K**, but contains zeros on the off-diagonal and the variances of eigenvectors on the diagonal:

$$\Lambda = SVD(K)$$

These variances are termed the eigenvalues, and each is associated with an eigenvector that is a linear combination of the original variables. The eigenvectors are mutually orthogonal, and the eigenvalues descend in magnitude, so that the first eigenvalue is the largest, followed by the second, etc. Both the correlation and covariance matrices can be utilized for the SVD, and the desirability of either is question-specific. The vector of eigenvalues $\mathbf{\Lambda_v}$ can be extracted from $\mathbf{\Lambda}$ for further study, and used in its raw form, or standardized by division by the trace of $\mathbf{\Lambda}$ to create a vector expressing the proportion of total variance explained by each component.

There are myriad examples in the literature of the application of principal components analysis to GPA landmark data (for example O'Keefe et al. 2014, for dire wolf crania). The usual analysis progression is to calculate the principal component scores for the first few PCs, and plot the specimens into the principal component space. Factor analytic interpretation of shape differences along PC axes is complicated compared to traditional morphometrics (Marcus 1990), where the eigenvector coefficients are readily interpretable; because each landmark coordinate is entered as its own variable the eigenvector coefficients are not interpretable as 'loadings' in the factor-analytic sense. However, various methods exist to visualize the landmark displacements along PCs, allowing biological interpretation of axes. Principal components analysis is therefore a powerful tool for dimensionality reduction, allowing the visualization and biological

interpretation of a large proportion of the variance in **K** on one or a few axes, and this has been its primary use in geometric morphometrics (Zelditch et al 2012). More recently, study of GPA covariance matrices has been extended to modularity and integration. The concept of modularity concerns the covariance structure of **K**. One attempts to identify subsets of landmarks with high intraset covariance and low covariance with other subsets (Goswami and Polly 2010), yielding powerful hypotheses of evolutionary plasticity and constraint. This concept can also be applied to a whole shape, so that a shape would be more integrated if it had fewer significant eigenvalues, and more modular if possessing more. A more integrated shape can be thought of as possessing fewer dimensions of variation, while a more modular shape would possess more dimensions.

The question of number of 'significant' eigenvalues is a general question in data analysis, and there are many techniques and rules of thumb for determining how many eigenvalues to retain. These methods are thoroughly summarized by Cangelosi and Goriely (2007) in the context of PCA of cDNA microarray data, and include the broken stick model, Cattell's SCREE test, and Bartlett's sphericity test among many others. This familiar litany of techniques used to determine the number of 'significant' eigenvalues is encountered across the many applications of PCA, but all share fundamental shortcomings. These are: 1) an integer value; 2) recourse to an *ad hoc* criterion to assess the cutoff for significance; and 3) space specificity, meaning the number of significant axes is specific to the space under study, and not comparable among spaces.

Shannon Entropy and Effective Dimensionality

The number of ‘significant’ eigenvalues may be conceptualized in other ways; in terms of matrix algebra it is equivalent to a functional rank of **K** that is less than the full rank, or to a measure of the rank deficiency of **K**. It is also equivalent to the number of dimensions present in **K**. The desirability of a metric that is continuous and theoretically grounded has been identified in several contexts, and at least two metrics based on the Shannon entropy have been proposed to characterize this dimensionality. The Shannon entropy, or statistical entropy (Shannon 1948), is a metric used in information theory to represent the information content of a system (for a thorough mathematical development see Cangelosi and Goriely 2007). Those authors develop a dimensionality metric for cDNA microarray data based on the Shannon entropy, so it is theoretically grounded, and is continuous. They term this metric the ‘information dimension’. A very similar metric was developed in a more general way by Roy and Vetterli (2010); those authors term their metric the ‘effective rank’, and use this term synonymously with ‘effective dimensionality’. The following application of the Shannon entropy to geometric morphometrics follows the Roy and Vetterli (2010) development of effective rank, and adopts their term of effective dimensionality.

Given a GPA covariance matrix **K** with singular values $\mathbf{\Lambda_V}$:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_v \geq 0$$

The Shannon entropy E_S of **K** is defined as

$$E_s = -1 * \sum_{i=1}^{v} (\frac{\lambda_i}{\sum \Lambda_V})(ln \frac{\lambda_i}{\sum \Lambda_V})$$

Roy and Vetterli define their effective rank as e raised to the power E_s; for the purposes of geometric morphometrics the 'effective dimensionality' of **K** is defined as

$$D_e = \exp\,(E_S)$$

The effective dimensionality of **K** is a continuous metric, is based on solid theoretical grounds, and may be thought of as the effective number of dimensions of the shape space represented by **K**. This metric has promise for the investigation of whole-object modularity and integration, and is examined further below.

Dense Semilandmarks and Latent Dimensionality

The effective dimensionality metric defined above should have application in the study of whole-object dimensionality of biological shape as represented by landmark data, and has the virtues of being 1) continuous rather than integer, and 2) not relying on *ad hoc* criteria for a dimensionality cutoff. However, this metric still suffers from the third limitation on dimensionality estimation, that of space specificity. For maximum utility, it is desirable that D_e should be comparable among spaces defined by different sets of landmarks. In the case of two spaces defined by the same number of landmarks, one might suppose that effective dimensionality is comparable even if the landmarks are not homologous. However this is not the case. A landmark-defined space with covariance matrix $\mathbf{K_v}$ only captures the variance of the landmarks, not of the entire shape upon which those landmarks occur. One way to represent the total covariance of the entire shape, $\mathbf{K_T}$, would be to sample the shape with an arbitrarily large number of landmarks, so that as the

magnitude of **V** increases $\mathbf{K_V}$ would approach $\mathbf{K_T}$. Thought of in this way, a typical set of geometric morphometric data would comprise

$$K_T = K_V + K_R$$

Where $\mathbf{K_R}$ is the residual covariance in the shape not captured by the landmarks in $\mathbf{K_V}$. The residual matrix $\mathbf{K_R}$ is of unknown magnitude in all mainstream applications of geometric morphometrics, and there is no guarantee that is negligible, nor that the proportion $\mathbf{K_R}/\mathbf{K_T}$ is of constant magnitude among spaces, even if $\mathbf{K_V}$s are of equivalent rank. Therefore the effective dimensionality metric defined above remains space-specific.

A recent approach to estimating residual shape variance is the employment of dense semilandmarks, where a curve or surface is first landmarked, and then densely sampled with an increasing number of semilandmarks until the coefficient of interest stabilizes (Marshall et al. 2019). This procedure can be employed here. Given a shape with variance $\mathbf{K_T}$ and landmark variance $\mathbf{K_V}$, the residual covariance $\mathbf{K_R}$ will decrease as **V** increases, so that

$$K_T \approx K_V, and\ K_R \approx 0$$

as the number of semilandmarks becomes large. The number of semilandmarks can be arbitrarily large, down to the pixel or voxel resolution of the image being digitized, but in practice semilandmarks need only be added until the coefficient of interest stabilizes. Using this procedure allows us to define the true, or latent, dimensionality of the shape

$$D_l = \lim_{v \to \infty} D_e$$

where the latent dimensionality of the shape space, D_l, is the effective dimensionality of a matrix that is semilandmarked densely enough to stabilize the coefficient. This coefficient should be useful for comparing shapes across landmark spaces, and for assessing the fidelity with which the original landmark data capture shape change.

Implementation, 1: Rank Deficiency and Permutation Scaling

The covariance matrices derived from GPA landmark coordinate data tend to be rank deficient, both because each landmark coordinate is treated as independent, and because biological shapes covary strongly. To illustrate this rank deficiency in **K**, and to implement the D_e metric developed above, a geometric morphometric data set is now introduced and analyzed. This data set concerns populations of a single species landmarked identically, so there is reasonable certainty that $\mathbf{K_R}$, while unknown, will be of similar magnitude in the populations, allowing comparisons of D_e. The data set consists of two-dimensional landmarks taken from 119 jaws of the dire wolf, *Canis dirus*, from Rancho La Brea tar pits, Los Angeles, CA. This data was published in Brannick et al. (2015). The data matrix used here is a subset of those data, consisting of 14 landmarks taken from each jaw outline (Figure 1). The data contains two subsets, one from a tar pit deposited about 19,000 years ago (Pit 13, n = 36), and the other from a pit deposited about five thousand years later (Pit 61/67, n = 83). Matrices $\mathbf{X_{13}}$ and $\mathbf{X_{6167}}$ comprise v = 28, and n= 39 and 83, respectively, of GPA coordinate data. Covariance matrices extracted from these data are $\mathbf{K_{13}}$ and $\mathbf{K_{6167}}$. The reader is referred to Brannick et al. (2015) for a thorough

analysis of shape change in these data; only the dimensionality aspects of are discussed here. All subsequent analyses were performed in R (R Core Team 2014).

Scree plots for $\mathbf{K_{13}}$ and $\mathbf{K_{6167}}$ of *Canis dirus* jaws are presented in Figure 2. This plot has two obvious features; the first is that both subsets are rank deficient, with tails of near-zero eigenvalues at higher ranks. The $\mathbf{\Lambda_{13}}$ vector also decays more quickly than $\mathbf{\Lambda_{6167}}$, suggesting that $\mathbf{K_{13}}$ is more rank deficient than $\mathbf{K_{6167}}$. Calculation of the effective dimensionality for the two matrices confirms these impressions (Table 1). Naïve interpretation of the effective dimensionalities from the two pits suggests that Pit 13 jaws have lower dimensionality, and hence are more integrated, than the Pit 61/67 jaws. This difference can be shown to be significant when an appropriate measure of dispersion is calculated for D_e (see below). However, the rank deficiency of **K** is a function of two factors: 1) the covariance structure of the landmarks, which is of biological interest; and 2) the magnitude of variance in **K**, which is proportional to sample size, redundancy in **V**, and other factors comprising variance of each variable v. In order for effective dimensionalities to be comparable, D_e needs to be scaled to a null estimate of D_e derived from a matrix that preserves the variances, $\sigma^2 v_i$, in **X**, but removes covariation. This can be achieved by permuting the columns of **X**, generating $\boldsymbol{\pi}\mathbf{X}$, thence a permuted covariance matrix $\boldsymbol{\pi}\mathbf{K}$, and calculating the D_e parameter iteratively. A mean value of the permutations (1000 were performed in this paper) yields

$$\pi D_e = \mu(\exp \pi E_s)$$

This permuted dimensionality is reported in Table 1, and depicted in Figure 2; the standard deviation of the permutations is also reported as a measure of dispersion on the

estimates. This value may be thought of as the maximum possible dimensionality of a covariance matrix with the input variances in **X**. These values are in intuitive agreement with the plots in Figure 2; both permuted D_e measures correspond to eigenvalue ranks that are approaching zero. As expected, the permuted values of D_e show that $\mathbf{K_{13}}$ and $\mathbf{K_{6167}}$ are both highly rank deficient, but that the group with smaller sample size is more rank deficient. To accommodate these differences in the input variance of data subsets, D_e can be divided by the permuted value,

$$\widehat{D}_e = \frac{D_e}{\pi D_e}$$

yielding a scaled dimensionality measure that varies from zero, in the case of total covariation, to one, in the case of complete lack of covariation. The scaled value $\widehat{D}_e$ is reported in Table 1 for $\mathbf{K_{13}}$, $\mathbf{K_{6167}}$, and $\mathbf{K_{all}}$. Importantly, this permutation procedure should be performed on the covariance matrix, not the correlation matrix. In a correlation matrix the variable variances are normalized to one, so information about relative variance magnitude is lost. This normalization will affect the eigenstructure.

The above argument also applies to latent dimensionality, D_l, and can be defined in a similar manner:

$$\widehat{D}_l = \frac{D_l}{\pi D_l}$$

This metric, $\widehat{D}_l$, is comparable among shape spaces of arbitrary landmark identity and number. It should have broad utility for the comparison of the spaces inhabited by biological shapes, homologous or not.

Implementation, 2: Bootstrapping for Dispersion, Hypothesis Testing, and Comparison with Eigenvalue Dispersion

Generating a measure of dispersion for $\widehat{D}_e$ is straightforward using a frequentist approach. Bootstrapping the rows in **X**, and then calculating the effective dimensionality iteratively, allows calculation of a bootstrap mean and standard deviation. These values are also reported in Table 1 for the *Canis dirus* data. This bootstrap dispersion measure should be normally distributed given the Central Limit Theorem, allowing the use of parametric tests. A Welch's T test was performed on the two *Canis dirus* subsets, and the effective dimensionality of Pit 13 wolves is slightly higher than that of 61/67. This difference is marginally significant (details report in Table 1 caption). These results show that effective dimensionality of a population can change significantly in a short time period, and should be of biological relevance.

The currently accepted procedure for studying whole-shape integration in a geometric morphometric context is to compute the eigenvalue dispersions of the groups of interest (Goswami and Polly 2010). This test is thought to be not very sensitive, and is susceptible to Type II error due to the presence of many near-zero eigenvalues in matrices that are rank deficient (as it typical in morphometrics). The use of the correlation matrix is thought to generalize this approach; however, the use of the correlation matrix on rank-deficient data will inflate variables of negligible importance, adding noise to the eigenstructure. The use of the correlation matrix is also not possible in a dense semi-landmark context, because the sum $\sum \Lambda_V$ will not converge on a finite value as **V** becomes arbitrarily large. This sum should converge using the covariance matrix, except in the special case of a fractal shape.

The eigenvalue dispersion, *ED*, was calculated for the matrices $\mathbf{K_{13}}$, $\mathbf{K_{6167}}$, and $\mathbf{K_{all}}$. The first procedure followed here is that recommended by Goswami and Polly (2010), where the eigenvalues are standardized by division by the mean eigenvalue, and the standard deviation is then calculated:

$$ED = sd\left(\frac{\Lambda_V}{\mu\Lambda_V}\right)$$

however the covariance matrix was used here instead of the correlation matrix, for the reasons enumerated above. These values are reported in Table 2. A permutation test was also performed, in order to establish a null expectation of *ED* for the case of no covariation:

$$\widehat{ED} = \frac{ED}{\pi ED}$$

This value is πED, and *ED* was divided by this value to yield a scaled version of *ED* that is analogous to $\widehat{D}_e$ derived above. Lastly, this value was bootstrapped to yield a mean and measure of dispersion for $\widehat{ED}$. These values are also reported in Table 2.

The results of the eigenvalue dispersion tests mirror those of the effective dimensionality tests. In these tests more integrated, lower-dimension matrices have higher values, while more modular, high dimension matrices have lower values. A niave reading of the *ED* statistic would suggest that Pit 13 wolves are more integrated, as with D_e; however, after standardization via division by the permuted null expectation, the 61/67 wolves are seen to be more integrated than those from Pit 13. Again this agrees with the D_e calculations, and highlights the importance of treating for rank degeneracy when assessing measures of whole-shape modularity. The bootstrap

values of *ED* are also always higher than the actual values, even when *n* is large, suggesting that **K** remains rank deficient. In summary the *ED* results are very similar to the D_e results, with the *ED* results showing greater statistical significance. The meaning of this relative sensitivity is unknown.

Discussion

Both the effective dimensionality test and test of eigenvalue dispersion demonstrate a similar pattern: dimensionality of the two pit samples is similar, but that 61/67 wolf jaws are slightly more integrated than those from 13. This is only apparent if both metrics are scaled appropriately; a permuted null value of both metrics was used here. The significance of this finding is marginal for the D_e statistic, and strong for the *ED* statistic. The sample size difference is large, and the Pit 13 sample is small enough to heavily impact all measures of effective rank. The difference in significance levels between the two metrics indicates that a modeling study is necessary to establish the sensitivities and biases of the two metrics. Examination of the scree plot in Figure 2 reveals that the Pit 13 eigenvalue distribution is odd; the first two principle components are almost equal in magnitude. This is unusual in a geometric morphometric data set, particularly one still containing the size vector; the Pit 61/67 eigenvalue scree plot is much more typical. This suggests that the small sample size in Pit 13 may be influencing the dimensionality metrics. A simulation study is necessary to assess the impact of sample size on dimensionality metrics.

There is evidence that the D_e metric is more sensitive than the *ED* metric. A Welch's T-test was run between the Pit 61/67 data set and the pooled data. The *ED* metric was not significant for the addition of the Pit 13 data, while the D_e metric was highly significant for the

same comparison. This indicates that the D_e metric is probably more sensitive to rank differences than *ED*, particularly at higher effective ranks, in agreement with previous suggestions of insensitivity in *ED* (Goswami and Polly 2010). It is also noted that the *ED* metric is not useable in a dense semilandmark context, as the number of zero eigenvalues will become large due to increasing rank deficiency, so for the calculation of D_l the D_e metric is required.

Examination of the bootstrap mean estimates of effective dimensionality, $\mu BS\widehat{D}_e$, shows that these estimates are always lower than the true $\widehat{D}_e$. This is not surprising in the case of Pit 13, where n = 36. The bootstrap resampling will result in matrices with fewer unique rows than 36, reducing statistical power in a sample size that is already small, heightening rank deficiency. However the bootstrap mean estimate for Pit 61/67 is also low, where n = 83, and the bootstrap mean is still low for the combined data set where n = 119 (Table 2). For these two larger samples, the effective rank is probably not sample limited, but variable limited. The value of πD_e indicates this may be true; the maximal effective rank for Pit 13 is 16.85, that for 61/67 is 21.04, and that for the pooled matrix is 22.143. All of these values are less than the nominal rank of 28 derived from the 14 sets of landmark coordinates, demonstrating that **X** is rank deficient even with large n, and the 14 sampled landmarks are probably not sufficient to characterize D_l, even with a sample of 119. This result emphasizes the importance of the latent dimensionality coefficient D_l, where the number of landmarks is made arbitrarily large.

The bootstrap results on $\widehat{D}_e$ for $\mathbf{K_{13}}$,$\mathbf{K_{6167}}$, and $\mathbf{K_{ALL}}$ demonstrates that the bootstrap mean is converging as sample size increases (Table 2). This difference is almost two effective ranks for the small data set, about one for the large one, and less than one for the pooled data set. The maximum effective rank for the *Canis dirus* jaws is only 22 when n = 119; this is less than the 28 nominal input variables and will constrain the effective rank of **K** even at large sample size. The

next logical step is to increase the effective rank by sampling more landmarks; a sampling density of 50 landmarks would yield 100 input variables for 119 samples, or almost a square matrix. Computation of the D_e of such a matrix would relax the possible rank limitation imposed by landmark number in the analyses presented here. This argument extends naturally to the calculation of D_l. Latent dimensionality requires a continuous landmark or surface for dense semilandmarking, so the data used here are not helpful. Compilation of a large-*n* study of *Canis dirus* jaw outlines is proceeding.

Acknowledgements

A. Brannick, M. DeBlois, J. Meachen, D. Polly, A. Farrell, G. Takeuchi, J. Harris, and E. Lindsey at the Tar Pit Museum. Funding sources: NSF EAR-SGP 1757236, Project SABER; Marshall University Drinko Research Fellowship to FRO.

Literature Cited

Bookstein, F.L., 1997. *Morphometric tools for landmark data: geometry and biology*. Cambridge University Press.

Brannick, A.L., Meachen, J.A. and O'Keefe, F.R., 2015. Microevolution of jaw shape in the dire wolf, Canis dirus, at Rancho La Brea. *Sci. Ser. Nat. Hist. Mus. Los Angeles County*, *42*, pp.23-32.

Cangelosi, R. and Goriely, A., 2007. Component retention in principal component analysis with application to cDNA microarray data. *Biology direct*, *2*(1), p.2.

Drake, A.G. and Klingenberg, C.P., 2010. Large-scale diversification of skull shape in domestic dogs: disparity and modularity. *The American Naturalist*, *175*(3), pp.289-301.

Marcus, L.F., 1990, May. Traditional morphometrics. In *Proceedings of the Michigan morphometrics workshop* (Vol. 2, pp. 77-122). Ann Arbor^ eMichigan Michigan: The University of Michigan Museum of Zoology.

Marshall, A.F., Bardua, C., Gower, D.J., Wilkinson, M., Sherratt, E. and Goswami, A., 2019. High-density three-dimensional morphometric analyses support conserved static (intraspecific) modularity in caecilian (Amphibia: Gymnophiona) crania. *Biological Journal of the Linnean Society*, *126*(4), pp.721-742.

O'Keefe, F.R., Binder, W.J., Frost, S.R., Sadlier, R.W. and Van Valkenburgh, B., 2014. Cranial morphometrics of the dire wolf, Canis dirus, at Rancho La Brea: temporal variability and its links to nutrient stress and climate. *Palaeontologia Electronica*.

Shannon, Claude E. "A mathematical theory of communication." *The Bell system technical journal* 27, no. 3 (1948): 379-423.

R Core Team, 2014. R: A language and environment for statistical computing.

Zelditch, M.L., Swiderski, D.L. and Sheets, H.D., 2012. *Geometric morphometrics for biologists: a primer*. academic press.

Table 1

Data Set	n	D_e	μBSD_e, +/-	πD_e, +/-	$\widehat{D}_e$	$\mu BS\widehat{D}_e$, +/-
Pit 13	36	10.40	8.54 (0.58)	16.85 (0.36)	0.617	0.507 (0.034)
Pit 61/67	83	11.36	10.35 (0.67)	21.04 (0.225)	0.540	0.493 (0.032)
Combined	119	12.343	11.48 (0.47)	22.143 (0.168)	0.557	0.518 (.022)

Effective dimensionality metrics for the dire wolf data sets. Effective dimensionality is D_e, followed by the bootstrap mean and standard deviation (μBSD_e, +/-), the permuted mean and standard deviation (πD_e, +/-), the scaled effective dimensionality ($\widehat{D}_e$), and the bootstrap mean and standard deviation of the scaled effective dimensionality ($\mu BS\widehat{D}_e$, +/-). A Welch's T test between the Pit 61/67 and Combined data sets for $\mu BS\widehat{D}_e$ was highly significant using effective dimensionality. This demonstrates the metric is sensitive enough to detect the addition of the Pit 13 data.

Table 2

Data Set	*n*	ED	$BSED$, +/-	πED, +/-	$\widehat{ED}$	$\mu BS\widehat{ED}$, +/-
Pit 13	36	0.311	0.359 (.022)	0.206 (0.006)	1.508	1.738 (0.104)
Pit 61/67	83	0.301	0.325 (.024)	0.155 (0.003)	1.943	2.105 (0.159)
Combined	119	0.274	0.294 (0.14)	0.142 (0.003)	1.933	2.070 (0.098)

Eigenvalue dispersion metrics for the dire wolf data sets. Eigenvalue dispersion is *ED*, followed by the bootstrap mean and standard deviation ($BSED$, +/-), the permuted mean and standard deviation (πED, +/-), the scaled effective dimensionality ($\widehat{ED}$), and the bootstrap mean and standard deviation of the scaled eigenvalue dispersion ($\mu BS\widehat{ED}$, +/-).A Welch's T test between the Pit 61/67 and Combined data sets for $\mu BS\widehat{ED}$ was not significant using eigenvalue dispersion. This demonstrates that eigenvalue dispersion is not sensitive enough to detect the addition of the Pit 13 data.

Figure Captions

Figure 1. Line drawing of a left dentary of the dire wolf, *Canis dirus*, with the location of the fourteen landmarks used in this study.

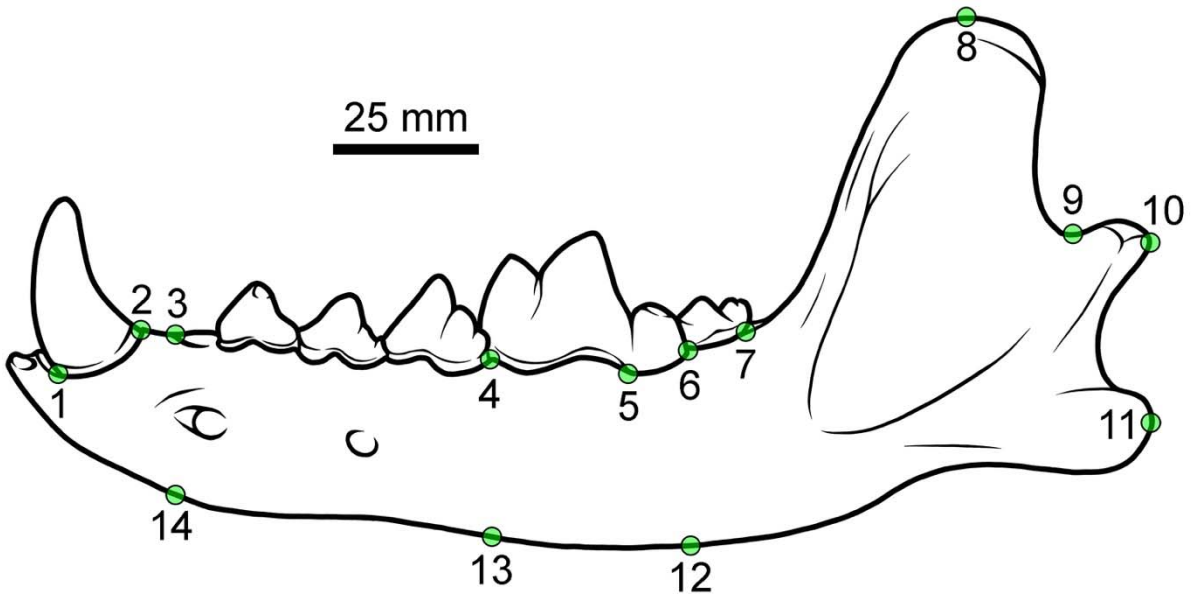
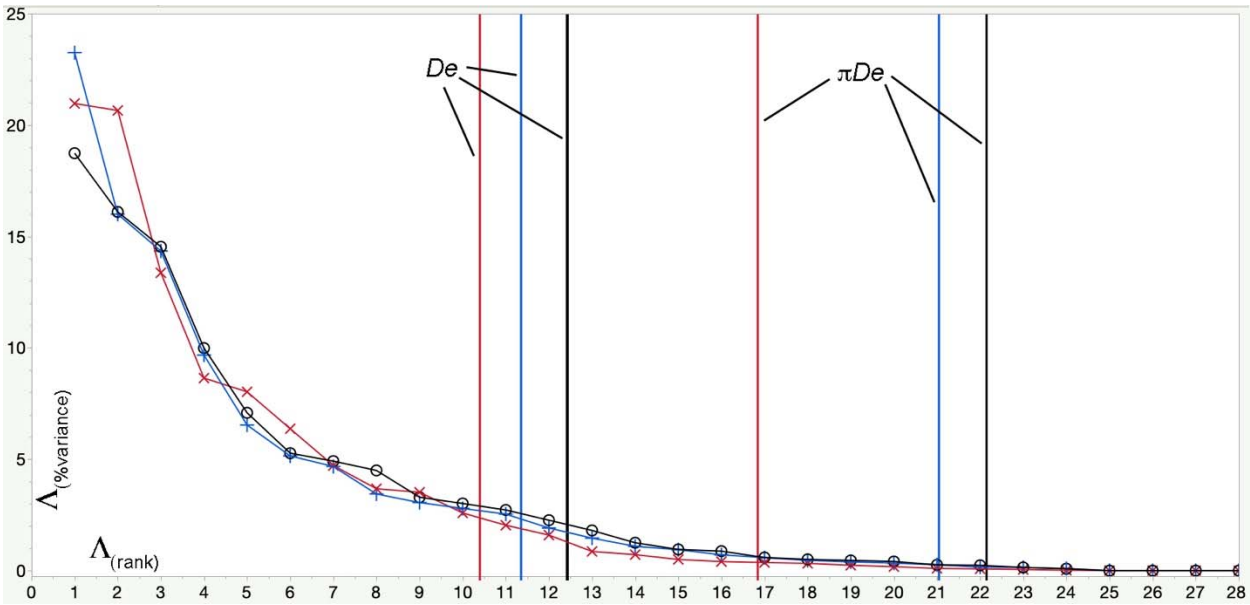
Figure 2. Annotated scree plot of the eigenvalue vectors $\mathbf{\Lambda_V}$, for matrices $\mathbf{K_{13}}$,$\mathbf{K_{6167}}$, and $\mathbf{K_{ALL}}$**.** Pit 13 jaws are in red, and depicted by X's; Pit 61/67 jaws are in blue and are depicted by crosses, and the pooled data are black circles. The effective dimensionality of each matrix is indicated in the same color, as is the permuted maximum dimensionality; Pit 13 is smallest for both measures, while the pooled data is highest for both. Note that the Pit 13 data is very rank deficient compared to the larger data sets.