REVIEW

TITLE: The Cinderella Discipline: Morphometrics and their use in botanical classification

SHORT TITLE: Morphometrics and their use

AUTHORS: Christodoulou, M.D.[1, 2]*, Clark, J.Y.[3], and Culham, A.[2]

[1] Department of Statistics, University of Oxford, Oxford, UK

[2] University of Reading Herbarium, School of Biological Sciences, University of Reading, Whiteknights, Reading, UK

[3] Department of Computer Science, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, UK

* For correspondence e-mail: maria.christodoulou@stats.ox.ac.uk

ABSTRACT

Between the 1960s and the present day, the use of morphology in plant taxonomy suffered a major decline, in part driven by the apparent superiority of DNA-based approaches to data generation. However, in recent years computer image recognition has re-kindled the interest in morphological techniques. Linear or geometric morphometric approaches have been employed to distinguish and classify a wide variety of organisms; each has strengths and weaknesses. Here we review these approaches with a focus on plant classification and present a case for the combination of morphometrics with statistical/machine learning. There is a large collection of classification techniques available for biological analysis and selecting the most appropriate one is not trivial. Performance should be evaluated using standardised metrics such as accuracy, sensitivity, and specificity. The gathering and storage of high-resolution images, combined with the processing power of desktop computers, makes morphometric approaches practical as a time- and cost-efficient way of non-destructive identification of plant samples.

Keywords: Plant taxonomy, geometric morphometrics, linear morphometrics, statistical learning, machine learning, identification, classification, neural networks.

In his keynote address during the 50th anniversary of botany MSc training at the University of Reading, Prof Vernon Heywood described a steady decline in the state of botany teaching in the UK with a resulting loss of skills in the next generation of scientists. With few institutions in the country offering training for young botanists, more and more researchers enter plant taxonomy through the field of molecular systematics, never learning the classic skills of a traditional botanist. Although great progress has been made in the development of molecular tools, increasing the insight gained from laboratory methods, what used to be the beating heart of botany - morphology - has lost some of its appeal. In our view this is because morphological data coding cannot readily be made into a clear data generation pipeline in the same way as much molecular data can. We believe this to be because morphology requires more in-depth knowledge and understanding of the organism prior to data collection than is required for DNA sequencing and that morphological variation is open-ended rather than with a fixed range of states as in DNA data. Whilst morphological data have lost favour in the construction of plant classification systems they have gained popularity in the study of evolution from variation in gross morphology of the centropogonid clade (Lobelioideae: Campanulaceae) (Lagomarsino *et al.*, 2017), speciation despite consistent floral morphology in *Myrcia* DC. (Vasconcelos *et al.*, 2019) though to detailed morphometric analysis of traits related to environment in *Vriesea* Lindl. bromeliads (Neves *et al.*, 2020).

The power of some of the more modern developments in morphometrics and statistical learning however can provide botanists with an extra toolbox to help them describe and quantify the variation that surrounds them. In this review we aim to make a case for the value of morphometrics, especially in combination with more sophisticated statistical methods, in a botanist's analytical toolbox - not to replace molecular techniques but to add to them. Morphology is often one of the most directly accessible and intuitive data sources for taxonomic research. In botanical taxonomy,

morphological characterization is the foundation of taxon description and identification, albeit often found in the formal and stylised format present in Floras and monographs. There is an opportunity for modern botanical taxonomy to explore the rapidly advancing field of morphometrics which already has some notable examples ranging from automatic leaf outline identification of *Passiflora* L. species (De Oliveira Plotze & Martinez Bruno, 2009), the tooth margin algorithm for *Tilia* L. leaf identification (Corney *et al.*, 2012), and the use of leaf venation architecture for major angiosperm clade recognition (Wilf *et al.*, 2016). Some computerised systems, starting with an existing classification of taxa, can use machine learning to handle the routine identification work, and then refer intransigent problems to a human expert (Clark, Corney, & Wilkin, 2017).

One of the principal arguments presented against morphological data is the potential for high levels of ambiguity. This ambiguity can be caused by a variety of factors such as inaccurate character definition (Assis, 2009) and difficulty in establishing homology (Schneider, Smith, & Pryer, 2009). Morphological data collection can be further complicated by plasticity of features (Perkins, Martinsen, & Falk, 2011), homoplasy (Schneider *et al.*, 2009), low numbers of characters (Giribet, 2010), and missing character states (Jenner, 2004). In some organisms such as parasites, reduced body plans can make characterisation of features even more difficult and lead to a very limited dataset (Perkins *et al.*, 2011). These concerns are neither exaggerated nor trivial and many are thoroughly discussed in the morphological literature. They do not, however, necessarily imply a lower quality of data produced by morphological work in comparison with other data sources (Jenner, 2004).

As most botanical researchers are question-driven rather than method-driven, we have structured our recommendations using general outlines on what kind of questions each combination of morphological and statistical tools can answer, with the aim of

promoting more thorough morphological investigation in botanical research. We have split this into two sections - Developmental hypotheses and Classification hypotheses. Under Developmental hypotheses we include all studies that may require the description of shape or size of a plant either to compare between treatments or to study how characters change along a particular gradient. For these we give an outline of morphometric tools available. Under Classification hypotheses we include all studies where the researcher is asking questions of taxon membership (e.g. are these two groups in the same taxon?) or questions of identification (e.g. what is the minimum set of diagnostics to accurately identify a sample?). These also require morphometric tools, such as those described under the developmental hypotheses, but can be taken further by combining them with machine learning techniques. There is a difference in terminology between the use of the word classification in biology and in computer science. Although the term is clearly defined in a taxonomic setting as the formal structure in which taxa are placed, in machine learning it means something much more general: it is the attribution of objects to a particular group. This is why identification in the machine learning context falls under classification, and therefore is included here under Classification hypotheses.

DEVELOPMENTAL HYPOTHESES

Plant growth and development studies already rely heavily on morphological measurements - size for example is often included as a proxy to an organism's developmental stage. These studies often focus on examining how the organism changes as it progresses through the various life stages. These could range from progression from seed to flower for an annual, or even development of fruit on a tree during the growing season.

Even though it is very commonly used, size itself is a complex and often unappreciated concept. Often researchers fail to explore the separation between shape and size, confounding the two and losing some of the clarity that can be obtained through their investigation. For Developmental hypotheses, we argue that the crucial point for insight is not the separation of size and shape just for the sake of it - it is for the researcher to either knowingly combine them or distinguish between the two based on the hypothesis in question. We believe that by cautiously selecting measurements that do not distinguish shape from size, a researcher can gain insight on changes in either size or shape during a developmental process based on how they use them. For example, the length of apple fruit through the growing season, plotted against time from anthesis can give insight on how size develops as time progresses (Atay, Pirlak, & Atay, 2010). The ratio between length and width for the same fruit provides an indication of the development of shape (Bollard, 1970). Both length and width independently are size metrics, but in combination they describe shape.

In the context of describing morphology, there are two mainstream methods: linear and geometric morphometrics. An essential distinction between them is that linear morphometrics do not actively separate size from shape, whereas geometric morphometrics do. We have structured the remainder of this section to describe these two techniques and have illustrated them using biological examples.

The traditional approach to morphometrics involves the measurement of distances between points deemed to be characteristic of shape and form. Measurements such as height, length, width, and diameter all fall under the general categorisation of **linear morphometrics**. These measurements are intuitive, easy to understand and to interpret, and have been in the biological toolbox for as long as the toolbox itself has existed. Linear morphometrics are quick to collect, low cost, easy to interpret, and often sufficient for biological description. Sanchez et al. (2011) compared the growth development of baobab seedlings of different origins using a variety of morphometric measurements ,such as length and diameter of roots, to establish that plants originating from drier environments grew to a smaller size even under optimal greenhouse conditions. Richardson et al. (2011) the studied fruit development patterns of kiwi from anthesis to ripening using amongst other character a collection of linear morphometrics, such as pericarp diameter. Zhang et al. (2015) performed a comparative study of the developmental patterns of Sweet cherry floral parts, using linear measurements such as pedicel length, establishing a correlation between floral morphology and environmental conditions during growth such as temperature.

For morphometric studies that require the description of very subtle shape characters, linear morphometrics may not be the most appropriate tool. The reason for this is because distance measurements, although excellent for summarising shape and size descriptions, often lack context. To correct for this, more linear measurements can be collected, creating a more complete dataset for each object. When the shape of interest is of biological form, it becomes crucial to be able to establish and quantify even the subtlest of differences. To be able to achieve this through linear morphometrics would involve an extensive collection of measurements and a generous amount of luck, as one may simply fail to measure the precise point where differences between taxa occur. Furthermore, within an evolutionary framework it is

more appropriate to view form as a whole since organisms evolve as a whole. To counter these concerns, morphometric theory progressed to what is often described as modern morphometrics, more accurately known as geometric morphometrics.

Geometric morphometrics allow the study of the shape of an organism as a whole, rather than as a collection of separate components. In contrast to linear morphometrics, by studying all the selected landmarks of a sample together, even subtle changes in geometry can be quantified and analysed using geometric morphometrics. Kendall's shape definition forms the basis of geometric morphometrics (Zelditch *et al.*, 2004). This clear separation of shape from position, orientation and size corresponds to an intuitive concept of shape. In practical terms, to achieve this separation there is a strong analytical reliance on multivariate techniques (Klingenberg & Monteiro, 2005). The way this is performed in geometric morphometrics is through the use of landmark coordinates (Van Bocxlaer & Schultheiß, 2010). A landmark is a recognizable point on the organism that, together with other landmarks, can be used to summarise the form of the organism (Zelditch *et al.*, 2004). As opposed to focusing on distance measurements, as is done in linear morphometrics, shape is summarised through the Cartesian coordinates of selected landmarks (Walker, 2000). By always analysing these coordinates together in a multidimensional space, shapes can be scaled, moved and rotated without losing any information (Goodall, 1991). Although the selection of appropriate landmarks can be difficult, this multivariate approach provides great flexibility for manipulation and statistical analysis.

After landmark selection, the recording of coordinates for all samples (a process referred to as "sample digitisation") creates the initial dataset to be used for analysis. The samples in this dataset are not, however, comparable if their coordinates have not been standardised. This is because regardless of how carefully and methodically digitisation occurred the samples are bound to not be fully aligned. Furthermore,

differences in sizes between samples will affect the position of the landmarks on the Cartesian axes, confounding shape comparison. To correct for this, the samples can be standardised using a Procrustes superimposition (Rohlf & Slice, 1990). Named after the mythical ancient Greek bandit who trimmed or stretched his victims to fit an iron bed, the process superimposes the samples using the landmarks to correct for orientation and alignment (Stegmann & Gomez, 2002). It then proceeds to stretch or shrink some samples aiming for all samples to be perfectly superimposed (Zelditch *et al.*, 2004). We have illustrated the steps of this process in Figure 1.

Selecting appropriate landmarks to summarise a shape is perhaps the most crucial aspect of geometric morphometrics. The reason for this is that if the choice of landmarks is poor, then any subsequent analysis will reflect that. Through the process of landmark selection, the overall shape of the organism in question is summarised using a small number of representative landmarks. Selecting representative landmarks is a subjective exercise that relies on in-depth knowledge and understanding of anatomy and biology of the organism in question. This is because not all landmarks are created equal. A wisely chosen landmark can summarise shape appropriately and provide adequate information for biological inference. A poorly selected landmark will at best add high levels of noise to the dataset or, at worst, result in misleading patterns.

Ideally, landmark selection requires four criteria that ensure quality: **repeatability**, **consistency of position**, **adequacy** and **homology** (Zelditch *et al.*, 2004). **Repeatability** refers to the potential of locating the selected landmark accurately on a specimen multiple times (Zelditch *et al.*, 2004). If a landmark is difficult to locate or its position is relatively vague, then samples that have no significant biological differences may be found to be different as an artefact of poor landmark choice. **Consistency of position** refers to the relative positions between landmarks (Zelditch *et al.*, 2004). If two landmarks switch relative positions between different specimens then their comparison

can lead to statistical outliers that may affect the findings and analyses. **Adequacy** refers to the number and position of landmarks used to summarise a form (Zelditch *et al.*, 2004). Although more is not always better in terms of landmark selection, including too few landmarks will not lead to a representative dataset. Even though repeatability can be quantified, and consistency of position detected, adequacy is a harder criterion to evaluate. This is because adequate coverage can be highly subjective. The concept relies on finding the golden mean between oversampling the specimen (where too many landmarks can lead to higher noise levels in the dataset) and undersampling (losing possible detectable variation between specimens).

Homology in landmark selection has both geometric and biological aspects. Two landmarks are considered homologous in two specimens if there is a degree of correspondence between them. This correspondence can be purely a geometric attribute (e.g. the tips of the Giza pyramids are geometrically homologous) or a biological attribute (e.g. the forelimbs of bats and primates). Although all four criteria are important for landmark quality, establishing homology is crucial. It is only through the use of homologous landmarks that the shapes studied are truly comparable. If the landmarks used are not homologous between the organisms in the study then there is no logical support for their comparison and the results can be highly misleading (Klingenberg, 2008). Although homology is considered one of the most crucial aspects in landmark selection, exactly how it can affect a given study depends on the nature and scope of the study itself. In general, the ability to identify homology can severely limit the quantity of potential landmark candidates.

These constraints imposed by homology increase the popularity of outline methods of analysis (Macleod, 1999). By replacing homologous landmarks with regularly spaced points along a curve, outline analysis sidesteps the issue of homology and can be used in cases where landmarks are sparse or hard to define (Macleod, 1999). Outline data

can then be analysed using Fourier harmonics (or possible variations such as Elliptical Fourier) or Eigenshape analysis (Macleod, 1999; Bonhomme & Claude, 2014). Although outline analysis is a popular and successful alternative to landmark analysis, the assumption that it bypasses homology issues may be misplaced. The reason for this is that outline methods are not completely independent of landmark correspondence assumptions (Klingenberg, 2008). That is because as with landmark methods, outline coordinates require a superimposition technique, such as Procrustes superimposition, prior to analysis (Bonhomme & Claude, 2014). This means that the outline points that are recorded are treated as actual homologous landmarks. This may appear minor, but as the superimposition process assumes a certain correspondence between points on the outline, it can result in increased levels of noise in the dataset. Furthermore, analytical approaches such as Elliptic Fourier Analysis also assume a certain degree of homology between outline points. It can therefore be argued that the principal difference between the two approaches is that in landmark analysis the homology criterion is explicit whereas in outline analysis it is implied and often ignored.

The choice between linear and geometric morphometrics for an analysis is not trivial as one technique is not necessarily superior to the other. Linear morphometrics are quick, intuitive and cost effective and often robust enough to not introduce noise in the analysis. They fail when separation of shape and size becomes important and when subtle changes in morphology are crucial - this is where geometric morphometrics excel. Selecting the appropriate method for the question in hand is always a challenging aspect of scientific discovery, although familiarity with both methods, combined with understanding of the studied organism helps when deciding which technique may provide more insightful findings. As a final point, it is not always necessary to choose one over the other, for example, Christodoulou et al. (2018) combined linear and geometric morphometrics to describe shape differences between apple cultivars with greater accuracy.

CLASSIFICATION HYPOTHESES:

Although classification in biology has a different meaning than in machine learning (a subset of statistical learning), this collection of hypotheses relies on grouping objects based on similarities between measured characters. These can include studies of morphological similarities between geographically distinct populations, segregation between species and hybrids, or revision of taxonomic limits.

Both linear and geometric morphometrics have been used for such studies. Compton and Hedderson (1997), in their taxonomic revision of the limits of *Cimicifuga foetida* L. s.l. (now *Actaea cimicifuga* L.), included 17 length variables, resulting in the detection of four geographically distinct species. Blanco-Dios (2007) used multivariate analysis of 17 linear morphometric characters to contrast the morphology of hybrid populations between *Armeria beirana* Franco and *A. pubigera* (Desf.) Boiss. with that of their progenitors, detecting clear differences between the groupings. Da Costa et al. (2009) used distance measurements for both vegetative and reproductive parts to study the variation within the *Vriesea paraibica* Wawra complex. After statistical analysis, they proceeded to recognise four species within the complex (*V. paraibica*, V. *interrogatoria* L.B.Smith, *V. eltoniana* E.Pereira & Ivo, and *V. flava* A.F. Costa, H. Luther & M.G.L. Wanderley), for which they provided a taxonomic treatment. Returning to the genus *Actaea* L., Gardner et al. (2012) used linear morphometrics to quantify the variation within *Actaea racemosa* L., establishing that between-population variation was similar to within-population variation. In a study of the *Andropogon lateralis* Nees complex, Nagahama et al. (2014) used 19 linear morphometric measurements to successfully distinguish both species and hybrids within the complex. Shipunov and Bateman (2005) used geometric morphometrics to explore the diversity of lip shapes of *Dactylorhiza* Neck. ex Nevski orchids, studying both hybridization patterns and taxonomy in Russian populations. Volkova and Shipunov (2007) used similar tools to investigate the variation between three *Nymphaea* L. species in Russia and Siberia,

finding the species delimitation to be robust. Viscosi et al. (2009) successfully used geometric morphometrics on oak leaves to distinguish between four species. Savriama et al. (2012) presented a new methodology quantifying symmetry and asymmetry of corolla shape in *Erysimum mediohispanicum* Polatschek (now *Erysimum grandiflorum* subsp. *mediohispanicum* (Polatschek) Romo), establishing symmetry to be a fundamental character for floral variation within the taxon. Finally, Fernández-Mazuecos et al. (2013) used geometric morphometrics to study the role of flower specialisation for speciation in *Linaria* Mill. subsect. *Versicolores* (Benth.) Wetst. finding corolla tube differences to correlate with divergent pollination strategies. In a comparison of leaf shape of *Anacardium microcarpum* Ducke with *A. occidentale* L. using geometric morphometric descriptors, Vieira et al. (2014) established that although the leaves do present statistically significant differences, overlap between taxa and populations prevent them from being used as unique identifiers.

Analytically, methods from statistical/machine learning can offer great insight for this type of hypothesis. There are two broad sections in statistical/machine learning: supervised learning and unsupervised learning. We are excluding deep learning methods here, as the topic is too large for an adequate description within this review and the approaches are rather different. The review on the topic by Angermueller et al. (2016) offers a good overview of the major issues. Furthermore, deep learning is primarily aimed at processing huge amounts of multivariate data (so called 'big data'), and here we are more concerned with the utilisation of relatively small datasets, often with only a few data records per taxon, which is more realistic for consideration by practising botanists.

Supervised learning focuses on using combinations of characters to circumscribe known groups (classes) and then applying this knowledge to predict the class membership of an unknown sample (Tarca *et al.*, 2007). This is essentially 'identification'

in the biological sense, if the classes represent named taxa. The classic example of supervised learning is Anderson's *Iris* dataset analysed by Fisher using Linear Discriminant Analysis (LDA) (Fisher, 1936). The original dataset contained measurements from 150 flowers belonging to three *Iris* species (50 flowers each of *I. setosa* Pall. ex Link, *I. versicolor* Thunb. and *I. virginica* L.), For each flower, length and width measurements of two tepals (one inner, and one outer tepal), as well as species, were recorded. When this dataset was analysed using LDA, discriminant functions were established for each species based on the lengths and widths of the tepals. These could then be used to establish the species of an unknown *Iris* sample using only length and width tepal measurements (provided it belonged to one of the three species). The factor that makes this example part of supervised learning is the prior knowledge of class membership, in this case *Iris* species, used for the design of the discriminant functions (Fogel, 2008).

Unsupervised learning, by contrast, has no prior knowledge of class membership, and the analysis aims to explore patterns in the data and create natural groupings (Fogel, 2008). Such groupings can then be used as justification for delimitation of traditional ranked taxa such as species. This is essentially 'classification' in the biological sense. Cluster analysis (clustering), for example, is a case of unsupervised learning. Table 1 summarises a selection of both supervised and unsupervised techniques, more extensive descriptions of which can be found in Appendix A.

Table 1 showcases botanical applications of machine learning. The combination of machine learning and morphometrics for classification has much more prominent examples outside of botany. We aim for this review to increase the uptake of these techniques in botany. In the meantime, we present some non-botanical examples here for illustration purposes. Santana et al. (2014) studied bee classification using the forewings of male members of five *Euglossa* species. This was performed by using 18

363 landmarks on the wing venation together with colour change variables, followed by
364 comparisons between classification techniques including linear discriminant analysis
365 and a modified neural network. The neural network outperformed the other classifiers,
366 with an accuracy of 87.6%. da Silva et al. (2015) used more classes than Santana et al.
367 (2014), studying 26 subspecies of *Apis mellifera* while still using the same 18 landmarks
368 on wing venation. Their focus was on the performance of feature selection and their
369 conclusion was that a Naïve Bayes classifier outperforms other classification techniques,
370 with 65% mean accuracy on cross-validation (da Silva *et al.*, 2015).
371
372 Van Bocxlaer and Schultheiß's (2010) gastropod study was one of the first in zoology to
373 combine machine learning with morphometrics, their focus was primarily on comparing
374 landmark analysis with outline analysis. For their gastropod dataset they found that
375 outline analysis outperformed landmark analysis by 3%, reaching 78% accuracy when
376 using a Support Vector Machine (SVM) classification (Van Bocxlaer & Schultheiß, 2010).
377 The high success rate of the outline analysis is likely due to the presence of three-
378 dimensional ornamentation on the shell surface. Also, the theory of outline methods for
379 biological shape analysis is not as robust as landmark analysis, as discussed briefly in
380 earlier sections.
381
382 Guisande et al. (2010) describe new software designed to identify fish species, using
383 Classification and Regression Trees (CARTs) and linear morphometrics. The structure of
384 the software is such that the user is required to make linear measurements on their
385 sample, following a certain protocol, and the measurements are then used to classify
386 the sample. This makes it similar to a multi-access key rather than a tool for automatic
387 identification. For multi-access keys, success rates can be established by testing the key
388 on the target audience and recording how successful was their navigation of the key.
389 Guisande et al. (2010) did not perform this test and only tested accuracy using samples
390 they had measured themselves.

In the field of anthropology, Velemínská et al. (2013) used semi-landmarks to study the greater sciatic notch (which is part of the pelvis bones) aiming to correctly classify the sex of the individual. Their best performing classifier was a Support Vector Machine that achieved a 92% accuracy. Instead of using a completely independent test set, the accuracy was quantified using a leave-one-out cross-validation approach on the learning set. The absence of a separate test set can lead to overestimating the accuracy of the classification as briefly discussed earlier.

The orthodontics paper by Yu et al. (2014) is based on the unusual premise of predicting attractiveness on malocclusion patients (patients with misaligned teeth). By using 101 landmarks on patient images combined with a Support Vector Machine, they achieved an accuracy of attractiveness prediction of 72%. This work is interesting because it is the only example in the literature where geometric morphometrics have been combined with the regression approaches of statistical learning, rather than the classification ones. This is because the attractiveness measure used was based on a (subjective) score from 69 orthodontics experts, therefore the prediction was a continuous measurement rather than a class.

Model evaluation

There is a large collection of classification techniques available for biological analysis and selecting the most appropriate technique is not trivial. The reason for this is that there is no single classification technique that consistently outperforms all others regardless of the dataset studied. In machine learning this concept is referred to as the "No free lunch" Theorem. Stated formally by Wolpert and Macready (1997), the theorem suggests that the performance of all classifiers is equal when the totality of possible problems is considered. This means that for every classifier available there exists a possible problem where that classifier outperforms every other classifier. In practical

terms, this makes selecting a classifier for a study harder as the only way to establish the appropriateness of the technique is after the training of the classifier. Due to this, the most common approach to classification problems is to train a variety of different classifiers and then select the one that performs best (Fogel, 2008). This strategy makes performance evaluation the focus of the classification analysis. To this extent a series of metrics have been proposed in the literature, summarised in Table 2.

All the metrics presented in Table 2 rely on describing classification success through the use of a set of samples, however selecting the set that is used is not straightforward. In most biological situations there is a limited amount of data available for study, making each individual sample valuable to the study. With a limited dataset, therefore, the decision on the appropriate "spending" of the data is not an easy one to make. This makes pilot studies that can inform power analyses (to estimate appropriate sample sizes) a crucial aspect of experimental design (McDonald, 2014).

There are three stages in machine learning that require data: training, validating and testing (Olden, Lawler, & Poff, 2008). During the first stage the classifier is primarily trained to the problem in question. If the whole dataset is used at this stage then it will have to be re-used for both validating and testing, leading to potential overfitting and unrealistically high performance metrics (Olden *et al.*, 2008). This is because the classifier would have knowledge of the full dataset at the training stage, therefore when validating occurs (which is the process that verifies that appropriate tuning parameters have been selected during training), overfitting is more likely as none of the validating samples will be new. When the classifier is then tested using known samples, the performance will appear improved due to this overfitting effect. The peril from this is that when the classifier is applied to truly unknown samples, the confidence in the resulting class could be misplaced. To avoid this, common practice involves partitioning the initial dataset to a training set (including a validation set) and a testing set. In this

447 case the testing set is used solely for establishing the final, unbiased, performance of
448 the classifier (Olden *et al.*, 2008). As this partition reduces the data available for training
449 and validating, partitioning the training dataset further may not be realistic as an
450 inappropriately small training set will create an inappropriate and untrustworthy
451 classifier.
452
453 In order to reduce overfitting during the validating process, cross-validation (CV) can
454 be used instead. In cross-validation the training dataset is partitioned, creating a
455 training set (in the strict sense) and a validation set (Olden *et al.*, 2008). Training
456 commences and is terminated when the performance with respect to the validation
457 set begins to reduce. The validation set is thus used as a dummy 'test' set. After the
458 classifier is trained and validated the two datasets are re-combined and re-partitioned
459 creating a new training and validation dataset. The learning process is repeated again
460 from the start until either a predefined number of data partitions, or all possible data
461 partitions, have been used for training. In biological applications of machine learning,
462 multifold (K-fold) cross-validation is commonly used to help avoid overfitting (Olden *et*
463 *al.*, 2008). During that process the training dataset is partitioned into K equal sets, with K-
464 1 of these recombined to create the training set and the last one used to validate. This
465 process is repeated K times for all possible (or sensible) combinations of training and
466 validation sets. More recently this technique has been slightly modified to include
467 further repetitions; for example, in M repetitions of K-fold cross-validation the process of
468 K-fold cross-validation already described is repeated M times. An example using two
469 repetitions of 5-fold cross-validation is illustrated in Figure 2.
470
471 Throughout this paper, we have explained and illustrated the many strengths of
472 morphometric study including the ability to train and evaluate a system, to conduct
473 power analysis on trial data sets to help decide on appropriate sample sizes and the
474 crucial element of reproducible measurement. Morphometric approaches can offer to

475 build strong and reproducible systems of classification and these can be combined
476 with DNA derived data to give a holistic synthesis that might improve the stability and
477 decrease the subjectivity of plant classification, especially at the species level. In short,
478 when botanists and horticulturalists catch up with other disciplines we expect to see
479 use of morphological data in the construction of more robust botanical classification
480 systems.
481

REFERENCES

Angermueller C, Pärnamaa T, Parts L, Stegle O. **2016**. Deep learning for computational biology. *Molecular Systems Biology* **12**: 878.

Assis LC. **2009**. Coherence, correspondence, and the renaissance of morphology in phylogenetic systematics. *Cladistics* **25**: 528–544.

Atay E, Pirlak L, Atay A. **2010**. Determination of fruit growth in some apple varieties. *Journal of Argicultural Sciences* **16**: 1–8.

Blanco-Dios JB. **2007**. Estudio morfométrico de una zona híbrida entre *Armeria beirana* y *A . pubigera* (Plumbaginaceae) en el noroeste de la Península Ibérica. *Anales del Jardín Botánico de Madrid* **64**: 229–235.

Van Bocxlaer B, Schultheiß R. **2010**. Comparison of morphometric techniques for shapes with few homologous landmarks based on machine-learning approaches to biological discrimination. *Paleobiology* **36**: 497–515.

Bollard EG. **1970**. The physiology and nutrition of developing fruit. In: Hulme AC, Rhodes MJ, eds. *The biochemistry of fruit and their producs: Volume I*., 387–425.

Bonhomme V, Claude J. **2014**. Momocs: Outline analysis using R. *Journal of Statistical Software* **56**: 1–24.

Christodoulou, M. D., Battey NH, Culham A. **2018**. *Can you make morphometrics work when you know the right answer? Pick and mix approaches for apple identification*.

Clark JY, Corney D, Wilkin P. 2017. Leaf-based automated species classification using image processing and neural networks. In Lestrel P, ed. *Biological Shape Analysis - Proceedings of the 4th International Symposium*: 29-56. World Scientific, Singapore.

Compton JA, Hedderson TA. **1997**. A morphometric analysis of the *Cimicifuga foetida* L. complex (Ranunculaceae). *Botanical Journal of the Linnean Society* **123**: 1–23.

Corney DPA, Tang HL, Clark JY, Hu Y, Jin J. **2012**. Automating digital leaf measurement: The tooth, the whole tooth, and nothing but the tooth. *PLoS ONE* **7**: 1–10.

Da Costa AF, Rodrigues PJFP, Wanderley MDGL. **2009**. Morphometric analysis and taxonomic revision of the *Vriesea paraibica* complex (Bromeliaceae). *Botanical*

510 *Journal of the Linnean Society* **159**: 163–181.
511 **Cuni Sanchez A, De Smedt S, Haq N, Samson R. 2011**. Variation in baobab seedling
512 morphology and its implications for selecting superior planting material. *Scientia*
513 *Horticulturae* **130**: 109–117.
514 **Fernández-Mazuecos M, Blanco-Pastor JL, Gómez JM, Vargas P. 2013**. Corolla
515 morphology influences diversification rates in bifid toadflaxes (*Linaria* sect.
516 *Versicolores*). *Annals of Botany* **112**: 1705–1722.
517 **Fisher R. 1936**. The use of multiple measurements in taxonomic problems. *Annals of*
518 *Eugenics* **7**: 179–188.
519 **Fogel GB. 2008**. Computational intelligence approaches for pattern discovery in
520 biological systems. *Briefings in Bioinformatics* **9**: 307–316.
521 **Gardner ZE, Lueck L, Erhardt EB, Craker LE. 2012**. A morphometric analysis of *Actaea*
522 *racemosa* L. (Ranunculaceae). *Journal of Medicinally Active Plants* **1**: 47–59.
523 **Giribet G. 2010**. A new dimension in combining data? The use of morphology and
524 phylogenomic data in metazoan systematics. *Acta Zoologica* **91**: 11–19.
525 **Goodall C. 1991**. Procrustes Methods in the Statistical Analysis of Shape. *Journal of the*
526 *Royal Statistical Society. Series B (Methodological)* **53**: 285–339.
527 **Guisande C, Manjarrés-Hernández A, Pelayo-Villamil P, Granado-Lorencio C, Riveiro I,**
528 **Acuña A, Prieto-Piraquive E, Janeiro E, Matías JM, Patti C, Patti B, Mazzola S, Jiménez S,**
529 **Duque V, Salmerón F. 2010**. IPez: An expert system for the taxonomic identification of
530 fishes based on machine learning techniques. *Fisheries Research* **102**: 240–247.
531 **Jenner RA. 2004**. When molecules and morphology clash: Reconciling conflicting
532 phylogenies of the Metazoa by considering secondary character loss. *Evolution and*
533 *Development* **6**: 372–378.
534 **Klingenberg CP. 2008**. Novelty and 'homology-free' morphometrics: What's in a name?
535 *Evolutionary Biology* **35**: 186–190.
536 **Klingenberg CP, Monteiro LR. 2005**. Distances and directions in multidimensional shape
537 spaces: implications for morphometric applications. *Systematic Biology* **54**: 678–688.

Lagomarsino LP, Forrestel EJ, Muchhala N, Davis CC. 2017. Repeated evolution of vertebrate pollination syndromes in a recently diverged Andean plant clade. *Evolution* **71**: 1970–1985.

Macleod N. 1999. Generalizing and extending the eigenshape method of shape space visualization and analysis. *Paleobiology* **25**: 107–138.

McDonald JH. 2014. *Handbook of Biological Statistics*. Baltimore, Maryland.: Sparky House Publishing.

Nagahama N, Anton AM, Norrmann G a. 2014. Taxon Delimitation in the *Andropogon lateralis* Complex (Poaceae) in Southern South America based on Morphometrical Analyses. *Systematic Botany* **39**: 804–813.

Neves B, Zanella CM, Kessous IM, Uribbe FP, Salgueiro F, Bered F, Antonelli A, Bacon CD, Costa AF. 2020. Drivers of bromeliad leaf and floral bract variation across a latitudinal gradient in the Atlantic Forest. *Journal of Biogeography* **47**: 261–274.

Olden JD, Lawler JJ, Poff NL. 2008. Machine learning methods without tears: a primer for ecologists. *The Quarterly Review of Biology* **83**: 171–193.

De Oliveira Plotze R, Martinez Bruno O. 2009. Automatic Leaf Structure Biometry: Computer Vision Techniques and their Applications in Plant Taxonomy. *International Journal of Pattern Recognition and Artificial Intelligence* **23**: 247–262.

Perkins SL, Martinsen ES, Falk BG. 2011. Do molecules matter more than morphology? Promises and pitfalls in parasites. *Parasitology* **138**: 1664–1674.

Richardson AC, Boldingh HL, McAtee PA, Gunaseelan K, Luo Z, Atkinson RG, David KM, Burdon JN, Schaffer RJ. 2011. Fruit development of the diploid kiwifruit, Actinidia chinensis 'Hort16A'. *BMC Plant Biology* **11**.

Rohlf F, Slice D. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Biology* **39**: 40–59.

Santana FS, Costa AHR, Truzzi FS, Silva FL, Santos SL, Francoy TM, Saraiva AM. 2014. A reference process for automating bee species identification based on wing images and digital image processing. *Ecological Informatics* **24**: 248–260.

566 **Savriama Y, Gómez JM, Perfectti F, Klingenberg CP**. **2012**. Geometric morphometrics of
567 corolla shape: Dissecting components of symmetric and asymmetric variation in
568 *Erysimum mediohispanicum* (Brassicaceae). *New Phytologist* **196**: 945–954.
569 **Schneider H, Smith AR, Pryer KM**. **2009**. Is morphology really at odds with molecules in
570 estimating fern phylogeny? *Systematic Botany* **34**: 455–475.
571 **Shipunov AB, Bateman RM**. **2005**. Geometric morphometrics as a tool for understanding
572 *Dactylorhiza* (Orchidaceae) diversity in European Russia. *Biological Journal of the*
573 *Linnean Society* **85**: 1–12.
574 **da Silva FL, Sella MLG, Francoy TM, Costa AHR**. **2015**. Evaluating classification and
575 feature selection techniques for honeybee subspecies identification using wing images.
576 *Computers and Electronics in Agriculture* **114**: 68–77.
577 **Stegmann M, Gomez DD**. **2002**. *A brief introduction to statistical shape analysis*.
578 **Tarca AL, Carey VJ, Chen X wen, Romero R, Drăghici S**. **2007**. Machine learning and its
579 applications to biology. *PLOS Computational Biology* **3**: 0953–0963.
580 **Vasconcelos TNC, Chartier M, Prenner G, Martins AC, Schönenberger J, Wingler A,**
581 **Lucas E**. **2019**. Floral uniformity through evolutionary time in a species-rich tree lineage.
582 *New Phytologist* **221**: 1597–1608.
583 **Velemínská J, Krajíček V, Dupej J, Goméz-Valdés JA, Velemínský P, Šefčáková A,**
584 **Pelikán J, Sánchez-Mejorada G, Brůžek J**. **2013**. Technical Note: Geometric
585 morphometrics and sexual dimorphism of the greater sciatic notch in adults from two
586 skeletal collections: The accuracy and reliability of sex classification. *American Journal*
587 *of Physical Anthropology* **152**: 558–565.
588 **Vieira M, Mayo SJ, de Andrade IM**. **2014**. Geometric morphometrics of leaves of
589 *Anacardium microcarpum* Ducke and *A. occidentale* L. (Anacardiaceae) from the
590 coastal region of Piauí, Brazil. *Revista Brasileira de Botanica* **37**: 315–327.
591 **Viscosi V, Fortini P, Slice DE, Loy A, Blasi C**. **2009**. Geometric morphometric analyses of
592 leaf variation in four oak species of the subgenus *Quercus* (Fagaceae). *Plant*
593 *Biosystems - An International Journal Dealing with all Aspects of Plant Biology* **143**: 575–

587.

Volkova PA, Shipunov AB. **2007**. Morphological variation of *Nymphaea* (Nymphaeaceae) in European Russia. *Nordic Journal of Botany* **25**: 329–338.

Walker J. **2000**. Ability of geometric morphometric methods to estimate a known covariance matrix. *Systematic Biology* **49**: 686–696.

Wilf P, Zhang S, Chikkerur S, Little SA, Wing SL, Serre T. **2016**. Computer vision cracks the leaf code. *Proceedings of the National Academy of Sciences*: 201524473.

Wolpert DH, Macready WG. **1997**. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* **1**: 67–82.

Yu X, Liu B, Pei Y, Xu T. **2014**. Evaluation of facial attractiveness for patients with malocclusion: *A machine-learning technique employing Procrustes*. *The Angle Orthodontist* **84**: 410–416.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL. **2004**. *Geometric morphometrics for biologists*. Oxford: Elsevier Academic Press.

Zhang L, Ampatzidis Y, Whiting MD. **2015**. Sweet cherry floral organ size varies with genotype and temperature. *Scientia Horticulturae* **182**: 156–164.