Problems of shape characterization and comparison in bryophytes

JERZY SZWEYKOWSKI¹, MIROSLAW KRZYŚKO²

¹Department of Genetics, Adam Mickiewicz University, Dąbrowskiego 165, 60-594 Poznań, Poland.

²Department of Mathematics, Adam Mickiewicz University, Matejki 48/49, 60-769 Poznań, Poland

(Received: May 16, 1990. Accepted: November 20, 1990)

Abstract

Shapes of leaves of two liverwort species, viz. *Lophozia ventricosa* and *Lophozia silvicola*, were characterized by means of three different methods: polar coordinates, conics, and Fourier series. Results were then used to compare the leaf shapes of the two taxa, and differentiate between their leaves. Characterization by Fourier series gave the most reliable results.

Key words: shape characterization, shape comparison, polar coordinates, conics, Fourier series, bryophytes

INTRODUCTION

Taxonomists paid always great attention to differences in shape of organisms and their parts. Bryology is no exception in this respect — the shape of leaves, e.g., is one of the most important characters in distinguishing taxa, mainly of lower ranks. However, this basic character — shape — is usually described rather superficially or even vaguely, as e.g. leaves circular, oblong, and so on. In this manner, the judgements of shapes have always a substantial deal of subjectivity in them. To avoid this, taxonomists used many methods — some of them rather sophisticated — of measuring the shape. As an example, a method of measuring the shape of leaves in *Orthocaulis* (a genus of hepatics) can be cited (S z w e y - k o w s k i 1960, Fig. 1); the senior author used this method to distinguish Holarctic species of this genus even on the basis of a single leaf.

However, measurements of this and similar kinds characterize the shape in an indirect way. In these methods much of the important information concerning shape is lost. Recently, trends are observed to get more information in this respect. One of the simplest ways to achieve this, is to put the outline — drawn on a paper sheet or photographed — of the organ to be characterized on the

Cartesian coordinate system and to get the values of coordinates of points lying on the outline (Fig. 5). These values can be then treated as particular characters. The great convenience of this method is that one can put any number of points on the outline, i.e. one can characterize the shape as detailed as needed.

Fig. 1. A leaf of a hepatic species *Orthocaulis binsteadii* with measurements indicated (after Szweykowski 1960).

The other, more exact attitude to the problem of characterizing the shape, is to treat the outline of the shape as a curve and then to fit a mathematical formula to that curve. In this way the shape is characterized by this formula. Two methods are used to achieve this. First, we can use conics — the special kind of curves formed when a plane cuts through a cone (Fig. 2). According to mathematical theory, every curve can be expressed by a conic polynomial; we will not give here the technical details, they can be found in the literature (e.g. Bookstein 1978). The other method is to use the Fourier analysis to find a periodic trigonometric function characterizing the curve of shape (e.g. Christopher and Waters 1974, Gerdol 1987). The Fourier polynomial can be then fitted to the real shape with an amazing precision, as the results of fitting a Fourier series to the rather complicated shape of leaves in *Orthocaulis atlanticus* show (Fig. 3). The Fourier polynomial is of course infinite (see formula (2) in Appendix) and each of the double terms contains additional information of shape. The successive approximations using one, three, five and so on, terms show, how the shape improves gradually, and when we use the 23 terms, there is practically no difference between the curve based on mathematical formula and the real shape. Thus, there are methods allowing the mathematical characterization of even rather complicated or bizarre shapes. The problem that emerges, is how to compare shapes characterized in this way. Some basic problems that are encountered when applying these methods to bryology will be discussed below.

Fig. 2. Diagram showing the conics (heavy lines)

Fig. 3. Approximations of a real leaf shape using successive numbers of Fourier polynomial terms (n , see Appendix, formula 2). The real shape of the measured leaf is shown in hollow circles

At the beginning, 12 samples of *Lophozia ventricosa* s.l. from various sites in Poland were taken. Type of oil bodies in living plants was determined and then one leaf from five different stems of each sample was detached. Thus, every sample was characterized by five leaves. The leaves were then drawn with a camera lucida on a sheet of millimeter paper and the leaf outlines were used as "shapes".

As it is well known, Buch (1933) discovered that the European *Lophozia ventricosa* was polymorphic in regard to its oil bodies: some plants had a distinct

nucleus in nearly each oil body, whereas oil bodies of the other plants were composed of small droplets without a distinct nucleus. B u c h (1933) described the plants with nucleated oil bodies as a new species, namely *Lophozia silvicola*. He stated also that this plant showed narrower leaves that did *Liphozia ventricosa* s. str. We are not going to discuss the taxonomic value of *L. silvicola* here — our problem was rather to see whether the difference in oil bodies correlate with differences in the shape of leaves.

First of all, the lengths and widths of each leaf were measured and a scatter diagram constructed (Fig. 4). As it can be seen, there is no doubt that according to these two charcters both forms differ, although the amplitudes of variation overlap.

Fig. 4. Scatter diagram of 30 leaves of *Lophozia ventricosa* s. str. (full circles) and 30 leaves of *L. silvicola* (hollow circles)

The next step was to find the coordinates of points lying on the leaf outlines. We did it in the following way: the drawings were put on the Cartesian coordinate system in such a way that both lower ends of the leaf blade touched the abscissa, and the ventral side of each leaf touched with one point the

ordinate axis. Then, the ventral lower end of the leaf outline was chosen as the starting point, and from this point lengths of lines connecting it with the leaf's outline were measured. Every next line was led at the angle of 5° to the preceding one. Thus, we got the polar coordinates of each point of the leaf outline. The number of such points was about 30, and we can say that we characterized the shape of each particular leaf by 30 independent characters (Fig. 5).

Fig. 5. Technique of measuring leaf shape for the Fourier analysis

Then, three kinds of analysis were performed on vectors of the coordinate values. First — a discriminant analysis on the rough data, second — a conic was fitted to each leaf and the polynomials were then subjected to a discriminant analysis, and thirdly — Fourier analysis was performed, the Fourier polynomials were subjected to discriminant analysis (see Appendix).

The results are interesting. Some values that show the information content in the first four discriminant axes are given in the Table 1.

On the basis of the discriminant variables, a program of identification of singular leaves was constructed (see Appendix). To have a leaf told whether it

Table 1

Percent information in non-zero eigen-values for discriminant analysis of crude data (second column), transformed to conics (third column) and transformed to Fourier polynomials (fourth column), per cents

	Crude data	Conics	Fourier series
I	38.37	27.65	31.55
II	31.99	22.51	14.35
III	11.88	17.27	13.04
IV	4.62	11.24	8.90

belongs to *L. silvicola* or to *L. ventricosa* s. str. all that should be done, it is to measure the leaf exactly in the same way as it was done earlier. It was also possible to calculate the probabilities of correct classification (Table 2).

Table 2

Probability of correct classification of a single leaf of two species of hepatics, *Lophozia ventricosa* s. str. and *Lophozia silvicola* (per cents)

Species	Crude data	Conics	Fourier series
<i>L. ventricosa</i>	60.00	71.43	90.00
<i>L. silvicola</i>	66.67	65.52	83.33

Thus, for the purpose of identification, the Fourier method gives the most reliable results. Takings e.g. measurements of 5 leaves from one sample, one might be reasonably sure that the identification is correct (the probability of misidentification is about 10^{-4}).

A rather unpleasant thing about the methods discussed is that they hardly can be visualized graphically. E.g. when we take the first two discriminant axes from, say a Fourier polynomial, the information content is close to 46%. In another words it means that over 50 per cent of information has been lost. In order to use e.g. 90% per cent information, we should take into account not less than nine discriminant axes. It means that we need a nine-dimensional space to place every leaf. This can be of course calculated, but is not possible to visualize graphically in a simple way.

These rather sophisticated methods are rather laborious and we think that they should be used in special cases only. One of such cases is a situation that is rather frequently met in contemporary bryology. Namely, working with cytological, ultrastructural or biochemical methods, bryologists discover differences between individuals belonging to one classical species. These differences are sometimes substantial, the variants show not infrequently different geographic areas and have — no doubt — character of biological species. However, the basic differences can be scored only on living plants. The method of characterizing shapes, we discussed, can be used in such situations to find the correlated differences — if any — in shape. Then the formal taxonomy can start.

APPENDIX

It is assumed that the shape of the leaf is well described by a function $R(\theta)$, where R is the radius that joins the centre of gravity with the edge of the leaf and θ is the polar angle measured from a fixed line. The description of the shape of the

leaf using the function $R(\theta)$ is possible when any radius from the centre of gravity crosses the edge of the leaf exactly at one point.

The Fourier series of $R(\theta)$ is of the form:

$$R(\theta) = R_0 + \sum_{n=1}^{\infty} R_n \cos(n\theta - \psi_n),$$

where R_0 is the average value of the radius, n is the number of given harmonic, R_n is the amplitude of the n^{th} harmonic and ψ_n is the n^{th} phase angle.

The analysis of the shape of the leaf begins with the evaluation of the centre of gravity from the Cartesian coordinates. The presented method localized the centre of gravity only with some numerical accuracy. When the centre of gravity is known, then the edge points of the leaf are transformed to polar coordinates and the Fourier series coefficients for any harmonic are evaluated. The coefficients from the set of characteristics based on which and using a specified method of discriminant analysis we distinguish between different species or varieties of plants.

Let the leaf be represented by a counter-clockwise ordered set of L points (x_i, y_i) ; $i = 1, 2, \dots, k$ lying on the edge of the leaf (see Fig. 7). The first point (x_1, y_1) can be selected arbitrarily. The number of points representing the edge of the leaf must be at least twice as high as the number of harmonics considered in the Fourier series. We treat the given leaf as a flat homogeneous region. The centre of gravity is found using the trapezium method. The equation of the segment that joins (x_i, y_i) with (x_{i+1}, y_{i+1}) is given by:

$$y_i(x) = \frac{x_i - y_{i+1}}{x_i - x_{i+1}} x + \frac{x_i y_{i+1} - x_{i+1} y_i}{x_i - x_{i+1}}$$

The area of the i^{th} trapezium bounded by that segment and the axis is equal to:

$$A_i = \frac{1}{2} (y_i + y_{i+1}) (x_i + x_{i+1})$$

The static moment in relation to the y -axis for the i^{th} trapezium is equal to:

$$Mx_i = \int_{y=y_i}^{y_{i+1}} \int_{x=0}^{x_i(y)} x dx dy = -\frac{1}{6} (x_{i+1}^2 + x_i x_{i+1} + x_i^2) (y_i - y_{i+1}).$$

The static moment in relation to the x -axis for the i^{th} trapezium is equal to:

$$My_i = \int_{x=x_i}^{x_{i+1}} \int_{y=0}^{y_i(x)} y dy dx = -\frac{1}{6} (y_{i+1}^2 + y_i y_{i+1} + y_i^2) (x_i - x_{i+1}).$$

Hence the coordinates of the centre of gravity of the leaf are equal to:

$$\bar{x} = \frac{\sum_{i=1}^L Mx_i}{\sum_{i=1}^L A_i}, \quad \bar{y} = \frac{\sum_{i=1}^L My_i}{\sum_{i=1}^L A_i}.$$

It is assumed in the above formulas that the point (x_{L+1}, y_{L+1}) is the same as (x_1, y_1) .

Fig. 6. 30 Leaves of *Lophozia ventricosa* s. str. (full circles) and 30 leaves of *L. silvicola* (hollow circles) on the plane of two first discriminant axes (I. and II.). Per cent of information carried by the discriminant axes are also given

The Cartesian coordinates (x_i, y_i) of the edge points of the leaf are transformed to the polar system (R_i, φ_i) with the centre in $(\bar{x}, \bar{y})$ and the φ -axis that covers with the line passing through (x_i, y_i) and $(\bar{x}, \bar{y})$:

$$R_i = \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}, \quad \Theta_i = \arctg(y_i - \bar{y}) / (x_i - \bar{x}),$$

$$\varphi_i = \Theta_i - \Theta_1, \quad 0 = \varphi_1 < \varphi_2 < \dots < \varphi_L < 2\pi, \quad i = 1, 2, \dots, L.$$

In order to simplify the computations the Fourier series (1) can be transformed to:

$$R(\Theta) = R_o = \sum_{n=1}^{\infty} A_n \cos n\Theta + \sum_{n=1}^{\infty} B_n \sin n\Theta, \quad (2)$$

where

$$A_n = R_n \cos \psi_n, \quad B_n = R_n \sin \psi_n, \quad n = 1, 2, \dots$$

Fig. 7. Diagram showing the method of finding the centre of gravity of the shape. For details, see Appendix

Using the orthogonal properties of sine and cosine functions the Fourier coefficients have the form:

$$R_o = \frac{1}{2\pi} \int_0^{2\pi} R(\Theta) d\Theta = \frac{1}{2\pi} \sum_{i=1}^L \int_{\Theta_i}^{\Theta_{i+1}} R_i(\Theta) d\Theta, \quad (3a)$$

$$A_n = \frac{1}{\pi} \int_0^{2\pi} R(\Theta) \cos n\Theta d\Theta = \frac{1}{\pi} \sum_{i=1}^L \int_{\Theta_i}^{\Theta_{i+1}} R_i(\Theta) \cos n\Theta d\Theta, \quad (3b)$$

$$B_n = \frac{1}{\pi} \int_0^{2\pi} R(\Theta) \sin n\Theta d\Theta = \frac{1}{\pi} \sum_{i=1}^L \int_{\Theta_i}^{\Theta_{i+1}} R_i(\Theta) \sin n\Theta d\Theta, \quad (3c)$$

$n = 1, 2, \dots$,

where $R_i(\Theta)$ is the equation of the edge in polar coordinates between (R_i, φ_i) and (R_{i+1}, φ_{i+1}) , $i = 1, 2, \dots, L$.

Assuming that $R_i(\Theta)$ is linear in Θ we have:

$$R_i(\Theta) = \frac{R_{i+1}-R_i}{\Theta_{i+1}-\Theta_i} \Theta + \frac{\Theta_{i+1}R_i-\Theta_iR_{i+1}}{\Theta_{i+1}-\Theta_i}, \quad i = 1, 2, \dots, L. \quad (4)$$

Substituting (4) into (3a-3c) we get:

$$R_o = \frac{1}{4\pi} \sum_{i=1}^L (R_{i+1}+R_i)(\varphi_{i+1}-\varphi_i) \quad (5a)$$

$$A_n = \frac{1}{\pi} \sum_{i=1}^L \left[\frac{(R_{i+1}-R_i)(\cos n\varphi_{i+1}-\cos n\varphi_i)}{n^2(\varphi_{i+1}-\varphi_i)} + \frac{R_{i+1}\sin n\varphi_{i+1}-R_i\sin n\varphi_i}{n} \right] \quad (5b)$$

$$B_n = \frac{1}{\pi} \sum_{i=1}^L \left[\frac{(R_{i+1}-R_i)(\sin n\varphi_{i+1}-\sin n\varphi_i)}{n^2(\varphi_{i+1}-\varphi_i)} - \frac{R_{i+1}\cos n\varphi_{i+1}-R_i\cos n\varphi_i}{n} \right], \quad n = 1, 2, \dots \quad (5c)$$

After the evaluation of the coefficients (5a-5c) we standardize the size of the leaf dividing the Fourier series by the average radius R_o :

$$r(\varphi) = 1 + \sum_{n=1}^{\infty} a_n \cos n\varphi + \sum_{n=1}^{\infty} b_n \sin n\varphi, \quad (6)$$

$$\text{where } r(\varphi) = \frac{R(\varphi)}{R_o}, \quad a_n = \frac{A_n}{R_o}, \quad b_n = \frac{B_n}{R_o}, \quad n = 1, 2, \dots$$

For p terms of the series we evaluate the quantity

$$\sigma_p^2 = (L-1)^{-1} \sum_{n=1}^{\infty} d_p^2(i),$$

$$\text{where } d_p(i) = R_i/R_o - \sum_{n=1}^p (a_n \cos n\varphi_i + b_n \sin n\varphi_i).$$

If $\sigma_p < \varepsilon$ denotes the accuracy of the measurements of the points (x_i, y_i) , $i = 1, \dots, L$, then we assume that p normalized amplitudes equal to

$$r_n = R_n/R_o = \sqrt{a_n^2 + b_n^2}, \quad n = 1, 2, \dots, p$$

describe the given shape of the leaf in the leaf in the satisfying way. The set of p characteristics formed in such a way is used to distinguish between different species or varieties of plants with the methods of discriminant analysis.

We assume that the observed random vector $x^i = (x_{i1}, \dots, x_{ip})'$ has the distribution $N(\mu_i)' \Sigma$ in the i^{th} population, $i = 1, 2, \dots, k$. The classifying function connected with the i^{th} population has the form (see Krzyśko (1982)):

$$g_i(x) = -\frac{1}{2} (x - \mu_i)' \Sigma^{-1} (x - \mu_i) + \ln q_i,$$

where $q_i > 0$ ($q_1 + q_2 + \dots + q_k = 1$) is the prior probability that the observation x belongs to the i^{th} population, $i = 1, 2, \dots, k$.

The classification rule is as follows:

the observation x is classified to the i^{th} population if

$$g_i(x) = \max_{1 \leq j \leq k} g_j(x), \quad i = 1, 2, \dots, k.$$

The most efficient estimator based on k samples of sizes $N_1, N_2, \dots, N_k$ has the form (see Krzyśko 1985):

$$g_i(x) = -\frac{1}{2} \frac{N-k-p-1}{N-k} (x - x_i)' (S^{-1} (x - x_i)) + \frac{1}{2} \frac{p}{N_i} + \ln q_i, \quad (7)$$

where $N = N_1 + N_2 + \dots + N_k$, $N-k-p-1 > 0$, x_i and S are unbiased estimators of the parameters μ_i and Σ , $i = 1, 2, \dots, k$.

Acknowledgements

Thanks are due to Mr. Leszek Mierzyński, M. Sc., for performing the calculations; to Miss Renata Słodowa, M. Sc., who measured the leaves, and to Miss Katarzyna Buczkowska, M. Sc., for making figures 1-6.

REFERENCES

- Bookstein F. L., 1978. The measurement of biological shape and shape change. Springer V., Berlin-Heidelberg, 191 pp. (Lecture Notes in Biomathematics vol. 24).
- Buch H., 1933. *Lophozia silvicola* Buch sp. nov. Ann. Bryol. 6: 125-129.
- Christopher R. A., Waters J. A., 1974. Fourier series as a quantitative descriptor of miospore shape. J. Paleont. 48: 697-709.
- Gerdol R., 1987. Discrimination of *Sphagnum* sect. *Acutifolia* based on a quantitative expression of steamleaf forms. Lindbergia 13 (3): 150-154.
- Krzyśko M., 1982. Analiza dyskryminacyjna (Discriminant analysis, in Polish), Wydawnictwo Naukowe UAM, Poznań.
- Krzyśko M., 1985. Estimation associated with linear and quadratic discriminant functions. In: Mathematical statistics and applications. W. Grossman, I. Vincze, W. Wertz (eds.), Akademiai Kiado, Budapest, Vol. B: 185-199.
- Szweykowski J., 1960. *Orthocaulis binsteadii* (Kaal.) Buch — a new liverwort for Central Europe. Fragm. Flor. Geobot. 6 (3): 399-405.

*Charakterystyka kształtu oraz metody jego porównania na przykładzie mszaków***Streszczenie**

Do scharakteryzowania kształtu liści dwóch gatunków wątrobowców (*Lophozia ventricosa* i *L. silvicola*) użyto trzech różnych metod: wyznaczenia współrzędnych biegunowych, dopasowania krzywych stożkowych oraz analizy Fouriera. Ta ostatnia dała najlepsze wyniki zarówno pod względem dopasowania kształtu krzywej do rzeczywistego obrysu liścia, jak i możliwości odróżnienia liści należących do tych dwóch taksonów.