ACTA STEREOL 1983; 2: 458-468
STEREOLOGY AND MORPHOMETRY IN PATHOLOGY
Proc 2nd Symp Morphometry in Morphol Diagnosis
Collan Y. et al., eds.
Kuopio University Press, Kuopio 1983

REMARKS ON ANALYTIC MORPHOMETRY IN BIOLOGY: PROCEDURE AND SOFTWARE ILLUSTRATION

Vittorio Pesce Delfino, Rosalia Ricco*

Antropologia-Istituto di Zoologia e Anatomia
Comparata, *Istituto di Anatomia e Istologia
Patologica Universita' 70124 Bari, Italy

ABSTRACT

This article deals with the problem of image processing of biological figures defined by closed or open curves and illustrates a computerized analytical method for describing the curve by a non-dimensional approach.

The valuable characteristics of the method (easy use, inexpensive programs, simple hardware, very exact results) are emphasized. The operations to which images are submitted, are also explained:

- The figure is digitized into a set of points with known coordinates.
- Upper degree polynomial equations are calculated with the least square method to obtain the best fit function curve available for direct comparison with the empirical curve.
- Fourier harmonic analysis is applied to characterize the set of the harmonics of the empirical scattered curve and the function curve. The values of sum of differences between the sets are calculated.

An Apple II Europlus computer is used. A mixed input system (TV camera and/or digitized table) is recommended. The software "SAM" (shape analytical morphometry), has been programmed by the authors.

METHOD

Classical morphometry currently utilizes measurements and derives from these the necessary parameters. Most of these parameters, however, do not represent shape whose description should be obtained independently from dimensions (Pesce Delfino et al. 1983, Ricco et al. 1983).

When we consider a bidimensional biological shape, generally we refer to a closed or open curve (the closed one circumscribes an area): the aim is to find parameters describing the curve and its related surface so that comparisons between images could become possible.

The main problem is segmentation of the image which results in a set of points of known coordinates. From these it is easy to calculate areas, lengths, indexes but for an analytic approach, it is necessary to find a parametrized description.

For this purpose a personal computer (Apple II Europlus) with 48 kbyte RAM and some peripherals (2 disc drives, a CRT colour display, a graphic printer, a b/w TV camera with a zoom lens interfaced with an AD converter, a digitizer table and a plotter) are employed; a Dimmer controlled spot light set is also necessary.

For classical methods of densitometric analysis we refer to traditional text-books and recent papers (Oja and Collan 1983).

Two input systems are available: a magnetized graphic table for manual tracing of contour and a TV camera for densitometric scanning of the image; the latter automatically finds points coordinates in which the scanning reads transition between different brightness-levels related to a cut-off value or to a densitometric slope (run-length-coding) and fills an array.

The use of TV camera, that is absolutely necessary for classification of grey levels of investigated images has a limitation: it requires much time for scanning of complicated contours.

We propose a combined method (relatively rapid, easy to use, not expensive and very exact) of 2 steps: a fast scanning of the figure which after filtering and contour enhancement is transferred to graphic table software: The digitized image is virtually on graphic table where the user draws its contours and, at the same time, follows the outline with a cursor visible on the screen.

Photographs (Figs 1,6) or histological images (by a TV camera connected with the microscope) are suitable for analysis: image position and its enlargement are easily settled. If dimensional bonds are requested, normalization (for dimensions) and standardization (for position) are performed. The latter is done easily, the former is more difficult and should be performed when comparing different figures step by step; but it is unnecessary when we are interested in classification only. Standardization should be performed before normalization and should be verified after normalization.

Parameters for position standardization (coordinates of the starting point, direction of the outline, coordinates of the barycenter) should be positioned on the same line, horizontal or vertical, and on the same side of the beginning point. The best parameter for normalization is the number of points in which perimeter is subdivided that should be the same for all figures.

Particularly for differences existing in executing normalization if the input system is the graphic tablet or the TV camera, it is necessary to refer, by software, to the set of data obtained from tablet to the standard of TV camera emulating TV camera scanning.

In classical morphometry irregular figures are frequently approximated to an ideal circle but this causes a loss of specific informations. So it is necessary to find other methods. The circle is well defined by coordinates of centre and radius length. For irregular curves corresponding descriptive parameters should be found. The first problem is definition of a characteristic point of the surface. The barycenter may be chosen because of its properties: it is a unique point in the studied figure, it is unchanged after translation, rotation or enlargement of the figure. Moreover if we arbitrarily consider the examined figure made of uniform material, the barycenter becomes characteristic for shape.

The second problem is to find analytic descriptions of the shape: descriptive polynomials and spectral analysis. Among numerous possibilities, polynomial splines, normalized polynomials, upper degree polynomials, we suggest the last that can be written as

$$y = b_k x^k + b_{k-1}x^{k-1} + \ldots\ldots + b_3 x^3 + b_2 x^2 + b_1 x^1 + b_0$$

where y is the dependent variable to interpolate, x is the given independent variable, k is the polynomials degree, b_1 - b_k are the coefficients of the equations, b_0 is the constant that has less significance than the other coefficients for the description of the curve and is related to the coordinate system (Curthbert 1980).

The coefficients of the equation are calculated with an interpolation/regression algorithm by the least square method that finds the curve with the minimum value of the sum of square deviation and of variance; then fitness is checked by estimating variance at the same time with the increasing polynomial degree; the procedure stops when a degree of polynomial is found to which a minimum variance value corresponds: this is the best fit equation. It should have a number of coefficients rather small in comparison to the number of the points of the curve (Figs. 2,3,7,8,11,12).

Naturally it is possible to apply the least square method using x-coordinates as an independent variable and y-coordinates as a dependent variable and vice versa. This would be of little use because equations so obtained for different structures are not related in any way to each other. If a closed curve is studied, only a slight winding curve, crossing the examined figure, is found.

The main goal is the study of convexity of different parts of profiles (Chassery 1982, Otsu 1982).

So we propose the following solution: x and y values are separately considered as dependent variables compared with arbitrarily selected series of positive integers from 1 to the number corresponding to the numbers of points considered.

So the closed curve (for example of 100 points) is cut off at the level from the first to the last point, half of the curve (from 1 to 50 points) is fixed in its position, the other half (from 51 to 100 points) is overturned 180 degrees around two turning points which are the last point of the first half of the curve and the first point of the second half. The opened curve is submitted to the least square method and the best fit curve is looked for (Figs. 2,3,7,8).

So two curves, one with x and the other with y as dependent variables, are obtained, both available for use (obviously the selected curve should always be the same when used for comparison); but necessarily both should be used when reconstruction as closed curves of x,y coordinates of function curve together scattered curve series is operated.

If normalization and standardization have been carried out, the number of the points, the position of the beginning point, the arrangement of values will be the same for all the figures. Moreover the position of the barycenters, the curves and its related equations will be directly comparable just in a shape match with numerical indexes of the differences.

Without previous normalization and standardization the polynomial equations give a shape factor index, variance values increasing proportionally with the complexity of the contour.

The results are related to the match between the empirical scattered curve and the function curve obtained by solving the equation of the series of independent variables corresponding to the same number of the points of the scattered curve. Because the equations have the same set of independent variables it will be possible to use variance values to make a classification (Figs 4,9).

This condition is like the traditional comparison between the examined contour and an ideal circle; but there is a difference: now comparison is with empirical scattered curve and function curve, the latter connected to the former by the least square algorithm.

Moreover the selection, dimensional or spatial criteria, of a circle to compare with the scattered curve becomes unnecessary.

There is only a rule to respect: at the input device, the beginning point and the direction of the tracing should be constantly the same.

Error between the empirical scattered curve and function curve is then calculated as total error, absolute total error, total square error, mean error, absolute mean error, square mean error and square root of the square mean error (Spiegel 1976). For previously normalized and standardized curve, these parameters are calculated also between different curves.

The difference of length of the two curves is estimated and finally the series are compared as standard error of a linear regression analysis: some parameters give indications about entity and direction of its crossings; the most synthetic indicator is the square root of the square mean error. It gives distance between scattered curve and function curve; if normalization and standardization were performed, between different curves also: so the first analytic indicator of shape is obtained.

Fourier analysis finds the harmonics of different frequencies and arranges them in order of importance in determining the final function: the number of the harmonics of evaluable amplitude is proportional to the irregularity of the contour (Churchill 1963, Spiegel 1974, Tai and Chiang 1982).

The typical function can be written as: $y = a_0 + a_1 \sin x + b_1 \cos x + a_2 \sin 2x + b_2 \cos 2x \ldots a_k \sin kx + b_k \cos kx$ where k is the maximum order of the harmonics, a and b are the coefficients that evaluate the amplitude of the contributing functions.

Fourier harmonic analysis gives the sine-cosine coefficients of the Fourier series, the amplitude of the requested number of the harmonics and the sum of the differences between the set of the harmonics of the scattered curve and the one of the function curve: distance between the curves is proportional to this difference and to the irregularity of the scattered curve: differences are calculated between curves of different figures if previously submitted to normalization and standardization; so the second shape parameter indicator is obtained (Figs 5,10).

SOFTWARE

The software utilized includes two packages:

I. Original graphic tablet software (Apple Computer Inc.) modified after conversion in disc operating system 3.3 to obtain an array with x-y coordinate values for each point of the profile. Such array is then memorized in a text-file and for closed figure, barycenter position is calculated and used for rectangular/polar conversion.

II. Original software by the authors: SAM (Biological Analytical Morphometry) that includes 3 groups of routines: The first is dedicated to the procedure of dimensional normalization and position standardization by a visible tunable coordinate system and by bright points on the analogic monitor, both corresponding to the coordinate system on digitized image and TV image.

A procedure seeks for the beginning/ending point and performs an adjustment of tablet set of data to the TV standard; there are also routines for image brightness regulation, and for the classic procedure of filtering, gray-tone balancing and contours enhancement. At the end a routine transfers digitized TV image to the software of graphic tablet whose function "slide" allows further changes of digitized image position.

The second group of routines performs densitometric analysis for 64 grey levels, draws extinction histogram, makes quantitative determinations and graphic restitutions, by grey-classes, in false colour, by three-dimensional graphics and by isodensity curves.

The third group of routines is dedicated to analytical elaboration of the textfile, obtained from TV camera or from both TV camera and the tablet.

The first routine of this group makes the splitting of the original array containing x and y values of each point, in two subarrays, one containing the x-value, the other the y-value, in the same order of the original array.

These two new textfiles are saved and submitted to the least square method starting from a second degree equation and automatically increasing by step one, and simultaneously making variance analysis; the computation stops when for a further increase of the equation degree, the variance stops its progressive diminishing; the best fit function is found and the coefficients of the corresponding Kth equation are given. These coefficients should be without rounding off or truncation (Figs. 2,3,7,8,11,12).

Now a routine gives the "distance" between different curves by evaluators previously indicated and draws a differential graphic.

Then the software discharges the intermediate system of service variables and reconstitutes an array by merging x-y values calculated by the equation and plots together the original scattered series of points and the function series of points as two closed curves with their barycenters (Figs. 4,9).

So we can visually compare the two curves and exclude errors in the intermediate steps. Errors if present, cause a topographic incompatibility of the curves. Moreover the software calculates dimensional parameters as length and surface, its indexes and cord/arch ratios.

At the end two routines are dedicated to Fourier harmonic analysis; the first calculates the sine-cosine coefficients, the amplitude and the distribution of the harmonics; the second gives the sum of the differences and the differences between the set of the harmonics of the scattered curve and the function curve and gives the results in tabular and graphical forms (Figs 5,10).

The time needed by the mixed input with TV camera and tablet for analyzing one image is, without normalization, in the order of one hour and half, about two hours if normalization is performed.

The presented sofware offers a lot of possibilities more advanced than commercial program packages but the time requested for those operations is considerable. This is strictly related to the high level language used for the software in the current version.

REFERENCES

Chassery JM: Discrete convexity; definition, parametrization and compatibility. IEEE 82CH 1801-0. Proceed. 6th Int. Conf. on Pattern Recognition. Munich, Germany 1982; 645

Churchill RV: Fourier series and boundary value problem. McGraw-Hill, New York, 1983

Curthbert D, Hood F: Fitting equations to data. Wiley J. and Sons Publishers 1980

Hall EL: Computer image processing and recognition. Acad. Press, New York 1979

Oja E, Collan Y: Basic principles of image analysis by a computer. Stereology and morphometry in pathology. Proc 2nd Symp Morphometry in Morphol Diagnosis. Collan Y et al. eds. Pp.48 - 59. Kuopio University Press, Kuopio 1984. Also published in Acta Stereologica 1983; 2: 250-261

Otsu N: Optimal linear and non linear solutions for least-square discriminant feature extraction. IEEE 82CH 1801-0. Proceed. 6th Int. Conf. on Pattern Recognition. Munich, Germany 1982; 557

Pesce Delfino V, DeLucia E, Scattarella V: Descrizione analitica e computerizzata di profili e sezioni del cranio a mezzo di equazioni polinomiali: descrizione della procedura. Boll. Soc. Ital. Biol. Sper. in press

Pesce Delfino V, DeLucia E, Scattarella V: Equazioini descrittive del profilo frontofaciale di Australopithecus Africanus STS5. Boll. Soc. Ital. Biol. Sper. in press

Ricco R, Bufo P, Pesce Delfino V: Analytic morhometry in diagnostic pathology: upper degree polynomials and Fourier harmonic analysis in infertility. Stereology and morphometry in pathology. Proc 2nd Symp Morphometry in Morphol Diagnosis. Collan Y et al. eds. Pp. 250 - 255. Kuopio University Press, Kuopio 1984. Also published in Acta Stereologica 1983; 2: 452 - 457

Spiegel MR: Fourier analysis. McGraw-Hill, New York, 1976

Tai HT, Chiang SH: Application of Fourier shape descriptors to classification of fine particles. IEEE 82CH 1801-0. Proceed. 6th Int. Conf. on Pattern Recognition. Munich, Germany 1982, 748

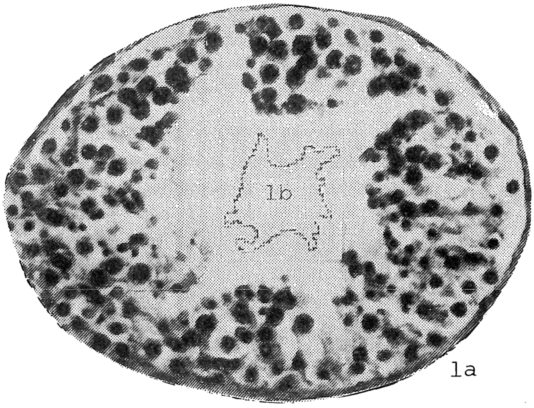


Fig.1a. Section of a seminiferous tubulus with complicated lumen contour (low maturation arrest, H&E stain, x 150). Fig. 1b. Inner profile of tubular epithelium: the contour is digitized in 271 points.

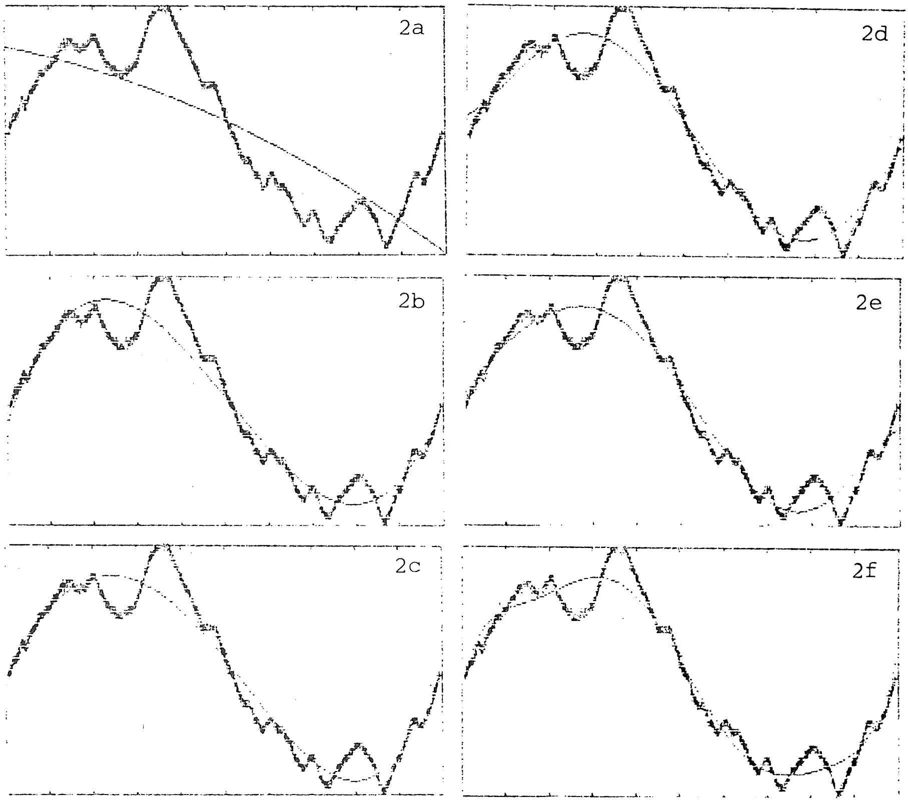
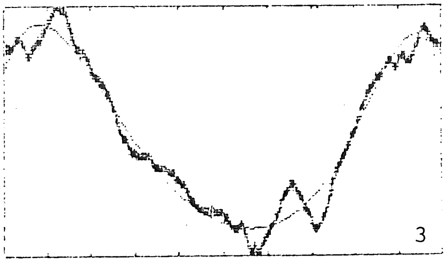


Fig. 2. Example of processing a closed curve. Polynomial function curves of the abscissa values of every point in which profile is subdivided (tubulus in Fig.1) a) 2nd degree polynomial (variance 70.9); b) 3rd degree polynomial (variance 14.4); c) 4th degree polynomial (variance 13.1); d) 5th degree polynomial (variance 11.8); 6th degree polynomial (variance 11.5); f) 7th degree polynomial (variance 8.4). Best-fit equation: $y = 152.230307 + 1.4770948x - 0.05547209x + 1.09838467x - 1.12526573x + 5.93704366x - 1.546325l3x + 1.58191434x$.



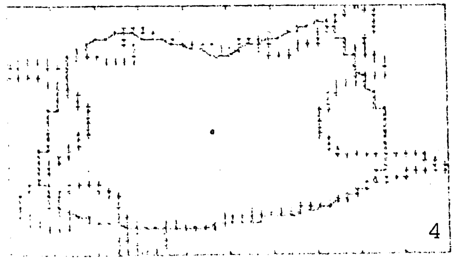
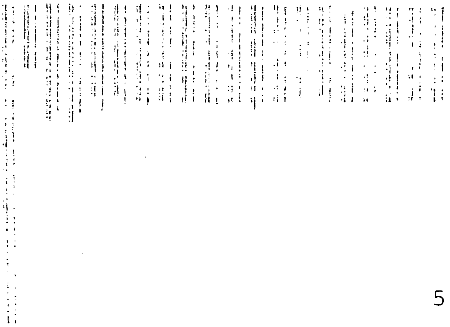


Fig. 3. Best-fit equation of the ordinate values of every point of the profile (Fig. 1 tubulus). y = 90.6349739 + 0.864901244x - 0.0246350707x + 1.86841129x - 2.43899021x - 3.79206075x + 2.02365561x - 3.07241808x (variance 6.18).

Fig. 4. Reconstruction as a closed figure of the empirical scattered curve (plus signed) and of the best-fit function curve in x and in y (dot signed) (Fig.1 tubulus). Topographic correspondence and coincidence of the barycenters of the figures are clearly visible. The scale factors of graphic restitution are responsible of some distorsion (surface defined by scattered curve 984; perimeter 275; surface defined by function curve 1048, perimeter 149).



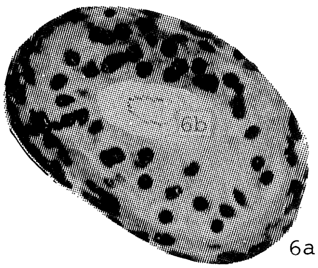
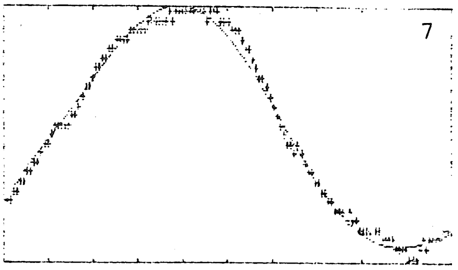


Fig. 5. Comparison of the set of the harmonics (Fourier analysis) of the scattered curve (double bars) and of function curve (single bars) (tubulus Fig. 1). Sum of the differences 26.4; maximum difference in the 5th harmonic 4.51.

Fig. 6a. Histologic aspect of a seminiferous tubulus section with slight irregular lumen contour (high level maturation arrest).

Fig. 6b. Inner profile of the tubular epithelium: the contour is digitized in 128 points.



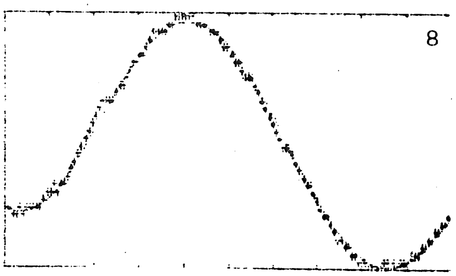
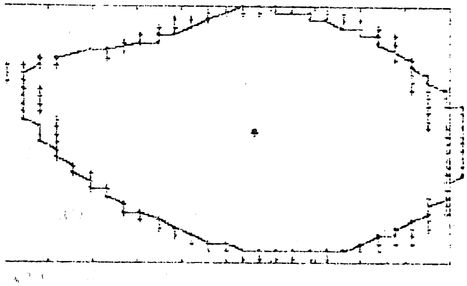
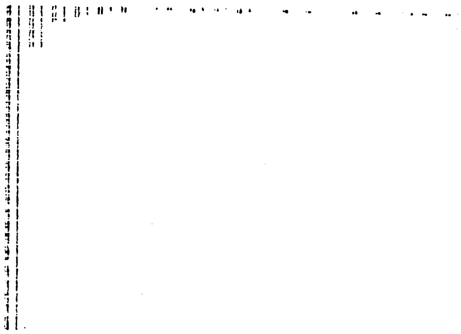


Fig. 7. Best-fit function (Fig. 6 tubulus) of abscisse values set. 7th degree polynomial (variance 0.58); $y = 137.234265 + 0.922442201x - 0.0432019958x + 1.95194144x - 4.23811307x + 4.28354952x - 2.01899305x + 3.6104286x$.

Fig. 8. Best-fit function (tubulus Fig. 6) of ordinate values. 7th degree polynomial (variance 0.31); $y = 88.3731434 + 0.617905419x - 2.23763815x - 6.76294943x + 7.54624713x + 3.64964835x - 7.23301636x + 2.26889216x$.



9



10

Fig. 9. Reconstruction as a closed figure of the scattered curve (plus signed) and of the bestfit (in x and in y) polynomial (dot-signed) of the Fig. 6 tubulus. Barycenters have the same position.

Surface defined by the scattered curve 601, perimeter 133; surface defined by function curve 600, perimeter 101.

Fig. 10. Comparison between the set of the harmonics (Fourier analysis) of the scattered curve (double bars) and the function curve (single bars) of the Fig. 6 tubulus. Sum of the differences 9.72; max. difference in the 6th harmonic 2.

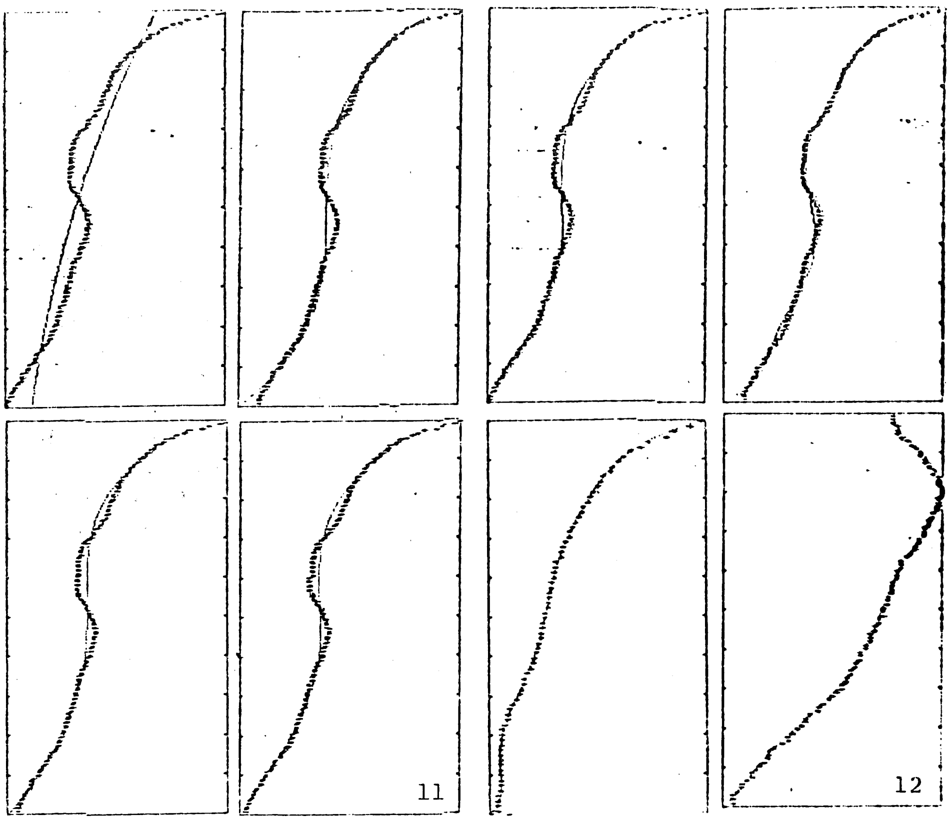


Fig. 11. Example of processing an open curve. Total fronto-facial profile (vertex-prosthion) in left lateral norma of the Australopithecus Africanus (Plesianthropus Transvaalensis). Function curves of increasing degree polynomials from 2nd to 7th. 2nd degree polynomial variance 260; 7th degree polynomial variance 5.35; scattered curve length 260; function curve length 257.

Fig. 12. Example of processing an open curve. The profile of Fig. 11 has been subdivided in 2 parts separately processed to increase analysis resolution. Frontal part (vertex-glabella): bestfit polynomial 7th degree (variance 0.41). Facial part (glabella-prosthion) bestfit polynomial 8th degree (variance 0.22). For equations in Figs. 11 and 12 see Pesce Delfino et al. 1983.