W.R. Tobler

Bidimensional Regression †

Since its invention by Francis Galton in 1877 regression analysis has been found useful in almost all disciplines. Comparison of geographic phenomena requires a two-dimensional extension of this technique. In this manner geographic maps can be compared with each other. Possible applications include geometric comparison of ancient and modern maps, or of "mental" maps, or for rubber-sheeting as used in Geographic Information Systems and in remote sensing. Other applications, for example, in biology for the comparison of shapes of leaves, fish, faces, or skulls after the manner of D'Arcy W Thompson are also possible, as are higher-dimensional and multivariate cases. The method implements, and puts into this new context, existing models from the field of cartography. The linear case yields an easy definition of a Pearsonian-like correlation coefficient The bidimensional case is richer in mathematical options than is the usual unidimensional version. The curvilinear case is of even greater utility. Here the regression coefficients constitute a spatially varying, but coordinate invariant, second-order tensor field defined by the matrix of partial derivatives of the transformation. This can be shown to be essentially equivalent to Tissot's Indicatrix, used in cartography to determine the properties of a map projection. In a computer implementation a nonparametric approach allows visualization of the regression by automatically plotting the pair of scatter diagrams, drawing of the displacement field, differentiable smooth interpolation of the warped coordinates and predicted image by a diagram of the principal strains, and with contour maps of the estimated local angular, areal, and total distortion.

There are many situations in which it is of interest to compute the degree of resemblance between two plane figures. Suppose we have pictures of the faces of two people, for example. Can we measure the degree of similarity of these faces? Or of two signatures? Or of two leaves? Or of two geographical maps? These questions are here approached by an analogy to regression analysis, as practiced in elementary statistics. Bidimensional regression is an extension of ordinary regression to the case in which both the independent and dependent variables are two-dimensional.

The possible applications of this technique cover such a large range that the procedures should become as well known, and as readily available, as the technique of simple regression. The student who has mastered the art of curve fitting by least squares should have no difficulty with the materials presented here. Mathematically the work can be considered an elementary application of empirical differential geometry (although this aspect is not stressed), or as a study of nonlinear transformations in two-dimensional Euclidean space. But the interest and focus are on real objects worthy of study, and the transformations are defined by these empirical consequences, rather than by abstract *a priori* properties. The biological and evolutionary speculations so imaginatively presented by D'Arcy Thompson in his classic *On Growth and Form* are probably the most famous realizations of transformations of the type under discussion.

But the proportionate studies of Albrecht Dürer must also be cited.

There is an obvious relation of the materials presented to canonical correlation, also not stressed. Nor is anything said about multivariate bidimensional regression, nor about bidimensional time series and growth patterns, nor is there reference to weighted least squares. And the treatment is deterministic, without significance tests. Extension to tridimensional regression is trivial, except for some display problems. And even these are not severe, given the observations. But not-Euclidean manifolds could be considered, as could general curvilinear coordinates; a treatment using the algebra of tensors would then be appropriate and is in fact practiced in the closely related field of continuum mechanics. Why then, given all these possibilities, is the treatment restricted to using only a few theorems from cartography? There is, of course, the professional interest in the study of ancient maps, and in the measurement of the distortions contained in "mental maps," but all of the foregoing possibilities are equally relevant there.

An objective is to stress a point of view, namely, that empirical two-dimensional transformations can be used much more effectively than other existing methods to study problems of shape change or growth. Obscuring this point with all of the rich detail that is possible is not appropriate. As an example, consider the current geometrical methods for the comparison of biological forms (skulls, crabs, etc.). Most frequently they are based on the measurement of attributes like the length and breadth of the object, and similar distances between landmarks. Each of these measures is in turn used to assign a coefficient on an attribute scale. The entire form then becomes a point in multivariate attribute space, generally assumed Euclidean in nature. The distance between points in this space is taken to be an inverse measure of the similarity of the original objects. But suppose one knows the physical distances between all landmarks on the form. Then it is possible to convert these into a list of Cartesian coordinates, and thus to specify the configuration of the form. How much simpler it is to measure the similarity between the objects by the magnitude of the transformation that is required to carry the one form onto the other. This is what bidimensional regression does. And it is based directly on the coordinates of the landmarks. And it gives more detail on where the difference between the forms it to be found. One doesn't even need matrix algebra to understand it, though it helps. An indictment of phase space methods is not intended, but most multivariate morphometrics is not thermodynamic in its use of attribute spaces. Fortunately bidimensional regression yields results that are independent of the particular coordinate system chosen, so that one does not even have to worry about this detail. As in ordinary regression it is a simple matter to define a bidimensional correlation coefficient relating two plane configurations, and also higher-order correlations. Thus all of the methods based on correlation matrices become available for the comparison of sets of forms. None of this is particularly exciting, or novel, from a mathematical point of view. But the applications need to be developed, and more widely distributed and understood. For advanced work one of course wants to consider all the stochastic and other ramifications.

THE REGRESSION ANALOGY

For unidimensional regression one has an independent variable X and a dependent variable Y and each consists of N numbers:

X Y

x_1 y_1
x_2 y_2
x_3 y_3
. .
. .
x_n y_n

Every x is associated with a particular y and we can construct a scatter diagram relating the two:

The objective is now to relate Y to X by a function Y* = f (X) in such a manner that the mapping X —> Y* is as nearly as possible, the same as the observed association X —> Y.

In particular we wish to be able to assign a y* when given a value of x when there was no observation at that location, or when there were several ys observed at that location, or when we suspect that the y measured at that location contains an error. Or we would like to know the general rate of change of y with respect to x given only discrete observations, or would like to know a most probable value for y when given an x. “As nearly as possible” is usually taken to mean that we minimize

$$\frac{1}{N}\sum_{k=1}^{N}(y_k - \hat{y}_k)^2 = \frac{1}{N}\sum_{k=1}^{N}(y_k - f(x))^2.$$

In order to find this minimum we must choose a model by deciding on a relation to be used for f(x). The degree of success is measured by the ratio

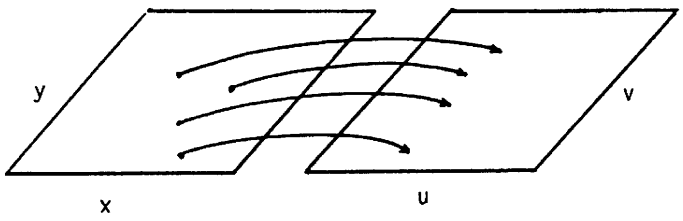
$$1 - \frac{\sum_{k=1}^{N} (y_k - \hat{y}_k)^2}{\sum_{k=1}^{N} (y_k - \bar{y})^2} .$$

In effect it indicates how much better is our estimate of y* than would be the guess using the average y-bar of the Y observations no matter what is the value of X.

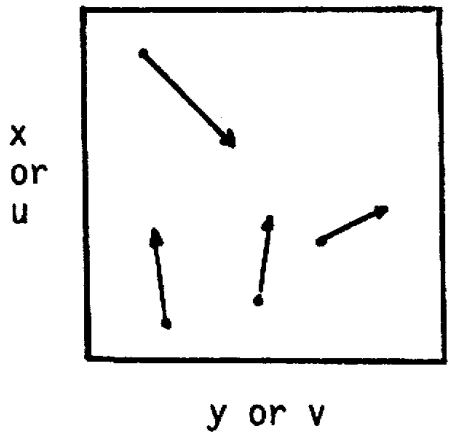
In the bidimensional case one has an independent variable Z and a dependent variable W and each consists of N pairs of numbers:

Z	W
$x_1 y_1$	$u_1 v_1$
$x_2 y_2$	$u_2 v_2$
$x_3 y_3$	$u_3 v_3$
.	.
.	.
$x_n y_n$	$u_n v_n$

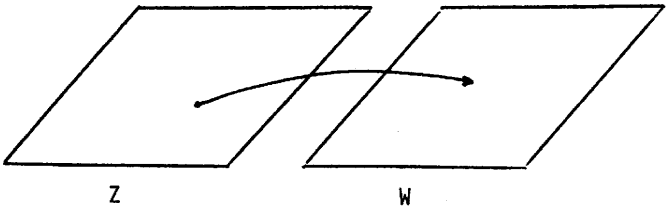
Every x, y is associated with particular u, v and we can construct a scatter diagram relating the two:



or, although misleading since the origin is arbitrary:



The objective is now to relate W to Z by a function W* = f (Z) in such a manner that the mapping Z —> W* is, as nearly as possible, the same as the observed association Z —> W.



In particular we wish to be able to assign a W* when given a value of Z when there was no observation at that location, or when there were several u, v observed at that location, or when we suspect that the u, v measured at that location contain an error. Or we would like to know the general rate of change of u, v with respect to x, y when given only discrete observations, or would like to know a most probable u, v when given an x, y.

As nearly as possible is usually taken to mean that we minimize

$$\frac{1}{N}\sum_{k=1}^{N}(W-\hat{W})^2 = \frac{1}{N}\sum_{k=1}^{N}(W-f(Z))^2.$$

In order to find this minimum we must choose a model by deciding on a relation to be used for f (Z). The degree of success is measured by the ratio

$$1 - \frac{\sum_{k=1}^{N} \left(W_k - \hat{W}_k\right)^2}{\sum_{k=1}^{N} \left(W_k - \overline{W}\right)^2} .$$

In effect it indicates how much better is our estimate u*, v* than would be the guess using the average u-bar, v-bar of the observations W.

In the bidimensional case each variable has two components. Thus the mapping Z —> W is equivalent to (x, y) —> (u, v) but would most commonly be written as two separate real functions:

$$u = f(x, y)$$
$$v = g(x, y).$$

Linear Models

As usual we begin with linear functions. There are three different mappings that can be called linear. These are, in order of the number of free parameters, the Euclidean transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \beta_{11} & -\beta_{12} \\ \beta_{12} & \beta_{11} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

the affine transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

and the projective transformation

$$u = \frac{\beta_{11}x + \beta_{12}y + \beta_{13}}{\beta_{31}x + \beta_{32}y + \beta_{33}}$$

$$v = \frac{\beta_{21}x + \beta_{22}y + \beta_{23}}{\beta_{31}x + \beta_{32}y + \beta_{33}} .$$

The geometrical meaning of linearity is that straight lines in the original are also straight lines in the image.

Derivation of the least squares Euclidean transformation (a Helmert transformation in the geodetic literature) is most easily effected using complex variables. The multivariate

bidimensional regression may in this case be written as

$$W = \alpha + \beta Z_1 + \gamma Z_2 + \delta Z_3 + \ldots$$

where $W = u + iv$, $Z_q = x_q + iy_q$, $\alpha = \alpha_1 + i\alpha_2$, $\beta = \beta_1 + i\beta_2$, etc., and $i^2 = -1$. Let $W^* = u - iv$. Then the least squares problem becomes one of finding the constants α, β, γ, and δ so that the residual

$$\epsilon^2 = \sum_{k=1}^{n} \left(W_k - \hat{W}_k\right)\left(W_k - \hat{W}_k\right)^*$$

becomes a minimum. The normal equations are derived in the usual fashion, and are given for three independent bidimensional variables and one dependent bidimensional variable in Figure 1.

$$\begin{bmatrix}
N & 0 & \Sigma x_1 & \Sigma y_1 & \Sigma x_2 & \Sigma y_2 & \Sigma x_3 & \Sigma y_3 \\
0 & N & \Sigma y_1 & -\Sigma x_1 & \Sigma y_2 & -\Sigma x_2 & \Sigma y_3 & -\Sigma x_3 \\
\Sigma x_1 & \Sigma y_1 & \Sigma(x_1^2+y_1^2) & 0 & \Sigma(x_1x_2+y_1y_2) & \Sigma(x_1y_2-x_2y_1) & \Sigma(x_1x_3+y_1y_3) & \Sigma(x_1y_3-x_3y_1) \\
\Sigma y_1 & -\Sigma x_1 & 0 & \Sigma(x_1^2+y_1^2) & \Sigma(x_2y_1-x_1y_2) & \Sigma(x_1x_2+y_1y_2) & \Sigma(x_3y_1-x_1y_3) & \Sigma(y_1y_3+x_1x_3) \\
\Sigma x_2 & \Sigma y_2 & \Sigma(x_1x_2+y_1y_2) & \Sigma(x_2y_1-x_1y_2) & \Sigma(x_2^2+y_2^2) & 0 & \Sigma(x_2x_3+y_2y_3) & \Sigma(x_2y_3-x_3y_2) \\
\Sigma y_2 & -\Sigma x_2 & \Sigma(x_1y_2-x_2y_1) & \Sigma(x_1x_2+y_1y_2) & 0 & \Sigma(x_2^2+y_2^2) & \Sigma(x_3y_2-x_2y_3) & \Sigma(x_2x_3+y_2y_3) \\
\Sigma x_3 & \Sigma y_3 & \Sigma(x_1y_3+y_1y_3) & \Sigma(x_3y_1-x_1y_3) & \Sigma(x_2x_3+y_2y_3) & \Sigma(x_3y_2-x_2y_3) & \Sigma(x_3^2+y_3^2) & 0 \\
\Sigma y_3 & -\Sigma x_3 & \Sigma(x_1y_3-x_3y_1) & \Sigma(x_1x_3+y_1y_3) & \Sigma(x_2y_3-x_3y_2) & \Sigma(x_2x_3+y_2y_3) & 0 & \Sigma(x_3^2+y_3^2)
\end{bmatrix}
\cdot
\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \gamma_1 \\ \gamma_2 \\ \delta_1 \\ \delta_2 \end{bmatrix}
=
\begin{bmatrix}
\Sigma u \\ \Sigma v \\ \Sigma(ux_1+vy_1) \\ \Sigma(uy_1-vx_1) \\ \Sigma(ux_2+vy_2) \\ \Sigma(uy_2-ux_2) \\ \Sigma(ux_3+vy_3) \\ \Sigma(uy_3-vx_3)
\end{bmatrix}$$

FIG. 1. Normal Equations for Euclidean Bidimensional Regression. Univariate case is unshaded. Trivariate case includes shading. u, v is the dependent variable and x, y is the independent variable. α, β, γ, and δ are parameters to be determined. The summations are to be taken over all of the N observations.

One might let W represent coordinate land marks on a child's face, and Z_1 and Z_2 represent landmarks on the parents' faces. We thus have a way of calibrating the geometry of an offspring as a weighted linear combination of the geometry of the parents. Galton would be delighted! Or W, Z_1, Z_2, Z_3, could be geometrical positions at four distinct times, that is, one could postulate that an old map has three precedents and it is desired to calibrate the relation by computing the α, β, γ, and δ. Thus $W = \Sigma^3_{q=0} CqZq$ represents a bidimensional linear time series model.

In the present instance we are concerned only with the univariate bidimensional case. The normal equations are given by the first four (unshaded) rows of Figure 1, in real form. The Euclidean transformation is given by this model. As is well known, it consists of a translation to bring the mean locations into coincidence, a rotation to principal axes by an angle θ about this location, and a uniform change of scale, all of which bring the dependent and independent variables into the best coincidence possible using a rigid motion. The manual equivalent is superimpositioning with enlargement or reduction of one image. Under some circumstances it may be desirable to delete the change of scale, or of origin, and then slightly different normal equations must be derived. The constant "slope" of the regression is $(\beta^2_{11} + \beta^2_{12})^{1/2}$, and the correlation coefficient is as was given above. The slope tells one how much change is expected in W when given a specific change in Z.

The least squares affine model would normally be used only in circumstances that call for an unequal change of scale in two different directions. The average regression "slope" is still ($\beta_{11}\beta_{22}$

$- \beta_{12}\beta_{21})^{1/2}$ but this is misleading because the rate of change of W with changes in Z is now different in different directions. One can no longer ask for the account of change in W given a change in Z but must ask what is the change in W when given a change in Z in a particular direction. The situation is still fairly simple, however, because changes in W do not depend on where Z is located, only on the direction.

NONLINEAR MODELS

The importance of the affine model is mostly in its relation to the more complicated curvilinear transformation. Specifically, any differentiable transformation Z —>W can be approximated locally by a translationless affine transformation

$$du = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy$$

$$dv = \frac{\partial v}{\partial x}\,dx + \frac{\partial v}{\partial y}\,dy\,.$$

Thus the matrix $\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}$ of the linear case becomes the matrix

$$\begin{pmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{pmatrix}$$

of the nonlinear case, or, in words, every nonlinear transformation is treated as a conjunction of infinitely many local transformations. The eigenvectors and eigenvalues of the coefficient matrix become important in analyzing the transformation. The regression “slope” depends on where Z is located and on the direction dx, dy of change. The theorems of Tissot (1881) are now very useful. The main results are that any infinitesimal circle in the original Z becomes an infinitesimal ellipse in the image W. That is, the radius of any infinitesimal circle is stretched by different amounts in different directions, and these amounts depend on where the circle is in the original, but there is a maximum stretching and a minimum stretching, and a smooth variation between these two. Importantly the maximum and minimum occur at right angles to each other, both in the original and in the image. We can thus construct a particular system of orthogonal coordinates in the original and these then correspond to a set of orthogonal coordinates in the image. And there is only one pair of orthogonal coordinates having this relation between the image W and the original Z. The subroutine STRECH computes the principal axes of Tissot’s indicatrix and the program can draw these axes, or draw the ellipses, or both, appropriately

scaled, on either the original or the image.

If the transformation is Euclidean the ellipses will all be of zero eccentricity, that is, circles and of constant size. If it is affine they will be of constant shape, size, and orientation. If it is projective they will be ellipses, whose shape, size, and orientation varies systematically away from the viewing point. If the transformation is curvilinear and equal area the eccentricity and orientation of the ellipses will change from place to place, but their area does not. If it is curvilinear and conformal the circles will go into circles, but their sizes will change from place to place. In the most general case the ellipses are everywhere different in shape, size, and orientation. Thus we can infer properties of the transformation by examination of the ellipses, or, of course, from the eigenfunctions used to compute the ellipses. Especially important are local length changes (stretchings of the radii of the infinitesimal circle), and deformation of angles and areas. In the curvilinear case the eigenfunctions depend on the partial derivatives of the transformation. Thus we require that the model chosen to fit the observational data must be differentiable.

FITTING CURVILINEAR TRANSFORMATIONS

The mapping (x, y) —> (u, v) can be approximated by a large variety of curvilinear models. The fitting presented here uses u* + iv* = f (x, y) + i g (x, y), where $i^2 = -1$ and the circumflex denotes an estimate. The functions f and g are to be chosen so that the difference between the estimates and the actual observations is as small as possible. By separating the real and imaginary parts of the equation the mapping can be written as two bivariate real functions u* = f (x, y) and v* = g (x, y). These functions, f and g, are not independent. For the mapping to be one-to-one, it must satisfy

$$\frac{\partial \hat{u}}{\partial x}\frac{\partial \hat{v}}{\partial y} - \frac{\partial \hat{u}}{\partial y}\frac{\partial \hat{v}}{\partial x} = J > 0.$$

Thus, when the mapping is formulated in least squares terms we minimize

$$\sum_{k=1}^{K}\left[(u_k - \hat{u}_k)^2 + (v_k - \hat{v}_k)^2\right]$$

$$= \sum_{k=1}^{K}\left[(u_k - f(x_k, y_k))^2 + (v_k - g(x_k, y_k))^2\right],$$

subject to J > 0. In this statement, use has been made of the partial derivatives of the mapping. It is therefore already implicit that some continuity and differentiability properties are required. And the bidimensional regression coefficients, as strain tensors, are computed as the eigenfunctions of the matrix of partial derivatives of the mapping. A necessary stipulation therefore is that the mapping and its derivatives be smooth, in a sense to be defined. One would

now like to simplify the problem further by invoking some theoretical insight in order to select from the set of all possible curvilinear models. Lacking such insight we can still reject some classes of models in order to simplify the selection problem.

An elementary curvilinear model might use the pair of bivariate algebraic polynomials:

$$\hat{u} = \sum_{p=0}^{P} \sum_{q=0}^{Q} A_{pq} X^p Y^q, \qquad \hat{v} = \sum_{p=0}^{P} \sum_{q=0}^{Q} B_{pq} X^p Y^q .$$

The main advantage of this model is its simplicity and generality. The model is differentiable, and can always fit the observations exactly, given a sufficiently high order. The disadvantages are equally well known, and include unrealistic oscillations in areas lacking data, and a difficulty in providing a substantive interpretation for the numerical coefficients. A minor, but practical, problem is the numerical instability, largely due to rounding errors, encountered in estimating the coefficients of high-order models. All of these difficulties occur in one-dimensional data and are severely compounded in higher dimensions. In the cases of concern here the observations are randomly scattered in a plane and thus the difficulties are compounded even further. There is also no coupling between the estimates for u* and v* in this model, thus there is no assurance that J > 0. Similar comments would apply to trigonometric polynomials, exponentials, Walsh functions, or other global models. Thus such models are rejected. Clearly there are some situations in which low-order polynomials, for example, bivariate quadratics or cubics, would provide useful approximations, and they have been used for this purpose by Sneath (1967). But they are not adequate for the present purpose.

Another apparently attractive model is of the form u* + iv* = f (x + iy), where f is an analytic function. Thus the complex polynomial

$$\hat{u} + i\hat{v} = \sum_{k=0}^{K} (a_k + ib_k)(x + iy)^k$$

is an obvious choice and gives an appealingly smooth mapping. There is now a strong relation between u* and v*. It is in fact too strong, and defines a conformal mapping. From a scientific point of view it is more reasonable to let the data speak for themselves. Thus one would like to test the hypothesis that the data are reasonably suggestive of a conformal relation. There is no value in an *a priori* assumption of conformality. As with real polynomials there are occasionally instances in which a quadratic or cubic complex polynomial is a plausible choice. But it again does not generally give an adequate fit to the data.

INTERPOLATION

Having thus chosen to use a local, rather than an *a priori* global, model, there still remain many possible curvilinear fittings. The choices are now to be made from the set of empirical double bivariate interpolations, from observations randomly scattered in a two space, with a restraint on the Jacobian of the mapping function and a smoothness requirement on the derivatives. The literature on interpolation of two-dimensional scalar fields is very large, and has been reviewed repeatedly, for example, Cram (1970), Lawson (1978), Schumaker (1976), Schut (1976). Much less is available on interpolating bivariate mappings.

One approach to the scalar interpolation problem is to assign blended patches of simple mathematical functions, fitted on a local basis. The best of these spline surfaces, for scattered spatial data, is described by Knudson and Nagy (1974). It would serve rather well for the present purpose and includes provision for observational variances when these are known (or can be estimated), but was not available in time to be used here. A stochastic approach, incorporating a spatial extension of the statistical interpolation scheme, is advocated for a similar purpose in geodesy and photogrammetry, under the name of collocation, and in geology as Kriging and also in the theory of regionalized variables. After assuming a particular covariance or variogram strncture one can use this theory to construct an "optimal" interpolation. Optimal here means minimization of the mean square interpolation error, contingent upon the assumed covariance structure. An attractive feature of the method is that it filters the observations and gives an estimate of the "noise" at each observation point, and of the interpolation error field. The main difficulty lies in the inadequacies of the procedures for estimating the covariance structure from the scattered observations, especially when the data are not spatially homogeneous, or stationary in the statistical sense. The alternate approach actually adopted here is also a least squares model and is optimal with respect to smoothness rather than to a covariance structure.

Another possible tack is to construct a triangulation, with the observations at the vertices of the triangles. A smooth surface, with smooth derivatives, can be constructed on such a basis. The main difficulty is that an upper bound on the number of topologically possible triangulations is (Brown 1965)

$$\frac{2(2B-3)!(4I+2B-5)!}{(B-1)!(B-3)!I!(3n+2B-3)!}$$

where B is the number of points in the convex hull and I is the number of interior points. Additional criteria are therefore required in order to pick a particular triangulation from this large set of possible triangulations. Several "optimal" criteria have been proposed, some of which have recently been shown to be equivalent (Lawson 1972). Unfortunately one of the natural criteria leads to an N-P complete problem. And the sensitivity of interpolation results to alternate triangulations has not been studied. The triangulation-based interpolation schemes, as published, also do not allow any variance in the observations. This limitation is not inherent in the method, but appears in the implementations. All interpolation schemes, including those based on a triangulation, can invariably be improved by additional substantive knowledge, but this is usually particularistic by field.

The method finally adopted begins by positioning a square lattice of equally spaced mesh points over the region of concern. The rectangular frame for this mesh extends approximately 5 percent beyond the area actually covered by the observations. This is a standard approach. The problem is now to assign values to the nodes of the mesh. Once this has been done for both u* = f (x, y) and v* = g (x, y) one then uses linear interpolation to assign values at points interior to each square of the mesh. Then the mapping function (x, y) —> (u*, v*) is defined everywhere within a frame slightly larger than the convex hull containing the observations. If the estimates u* and v* are determined independently of each other, the Jacobian determinant may pass through zero.

There is virtually no discussion of methods for choosing an appropriate size for the mesh in any of the literature reviewed. But the sampling theorem suggests that this must be smaller than one-half the size of the smallest feature to be detected. An estimate of this smallest feature is

obtainable by considering the highest degree of trigonometric polynomial that could be calibrated from the data. This degree cannot exceed the number, n, of observations. In K dimensions the effective number of observations is the K^{th} root of n. Thus an upper bound on the size of the mesh interval should be the K^{th} root of the volume of space dominated by each observation. In two dimensions this becomes, approximately, $\Delta y = \Delta x \leq (\text{area}/n)^{1/2}$, which has units of length per effective observation. It is not efficient to stray too far from this value but the spatial arrangement of the observations may exhibit clustering so that the exact limit is too lax. A lattice spacing that is two-thirds of the foregoing limit has been found effective when the observations lie within a square region. It is always possible to arrange that the data lie on an orthogonal but irregularly spaced n-by-n mesh, with observations at n nodes, leaving n^2 - n nodes for which estimates must be assigned, but this is practicable only for small n, small being defined by the size of one's computer. Since this is not the method used here, a more complete description of this alternative is omitted. The 2/3 $(\text{area}/n)^{1/2}$ rule is used in most of the present examples.

The assignment of estimates u* (or v*) to the nodes of the mesh can still be done in a large variety of ways, including the collocation and splining techniques cited above, and others cited in the literature. The particular method chosen assigns values by solving a partial differential equation, tuned to the observations in a manner to be described shortly. This particular choice was made in order to obtain a smoothly differentiable mapping. Once values are assigned to the nodes of the mesh, these numbers, along with bivariate linear interpolation, define the mapping function. No explicit equation exists, nor is one calibrated by estimating coefficients. The only parameters in this curvilinear model are the mesh size, the numerical quantities at the nodes, and the boundary assignment.

The lattice-tuning procedure consists of two parts. The first part constrains the estimates to agree, as nearly as possible, with the observations. The second part specifies the smoothness function. These two parts are applied cyclically in an iterative fashion. To understand how this tuning works, consider a single observation contained inside of a cell of the mesh. Assume some (any) numerical values at the four nodes of this cell. Use these four values to compute (by linear interpolation) an estimate at the same location as the observation. The numerical value of this estimate will, with probability one, differ from the numerical value of the actual observation. Suppose that the estimate is too low. Then it can be improved by changing (raising) one or more of the four values at the nodes of the surrounding mesh. In particular they can all four be chosen so that the linear re-interpolation will yield a value that agrees exactly with the observed value. But there are an infinite number of ways in which this can be done. Thus there is freedom to choose the values at the nodes of the mesh to satisfy some further objective. The criterion chosen is to desire that the numerical value assigned to each node should, as nearly as possible, be equal to a weighted average of the values assigned to the neighboring nodes. This target is initially somewhat elusive since the numeric value assigned to the nodes changes during the computations to accommodate both the observations and the neighboring values. But the process is stable and converges to a fixed numerical value at each node.

Nodes that are associated with cells lacking observational points must be treated differently. Estimates can be made at these locations since the connectivity of the lattice points to each other is known. It is this lack of knowledge and lack of uniqueness of adjacencies that makes interpolation from randomly scattered observations difficult. Thus the nodes tuned to the observation act as constraints on the unconstrained ones. This may be visualized as follows. After an initial assignment has been made for each mesh point, replace this number by the value

that it would receive by interpolation from its immediate neighbor. Do this in some regular sequence so that after an interpolation is made at one point this interpolated value is used to interpolate the value at the neighboring point. In this manner every point affects its neighbors, and is affected by them. This bootstrap operation pulls along all of the values to which it is applied. After passing over the array in this manner several times, a stable situation will have been reached in which the interpolated value at a point is sensibly the same as the value obtained on the previous pass. Neighboring nodes will then have values that are similar, and it is time to stop. The actual observations in this process are fixed points that affect the neighboring nodes, but that are not affected by them.

When there is more than one observation in a cell of the mesh, a perfect fit to the data cannot normally be achieved. To choose a mesh sufficiently fine so that this never occurs is usually impractical. The residual error must be balanced against computational cost. The initial guess from which one starts the iterations is not critical. Nor is it crucial whether linear or curvilinear interpolation is used within the mesh, recognizing of course that any method of interpolation is a hypothesis about nature that may or may not be true. The choice of the target function more fundamentally affects the solution.

A smooth function, intuitively, is one that has few oscillations (neighboring points have similar values) or is one that has a small rate of change in all directions (its partial derivatives are small). Thus it is natural to minimize the sum of the squares of the partial derivatives

$$\int\int \left(\frac{\partial u^2}{\partial x} + \frac{\partial u^2}{\partial y} + \frac{\partial v^2}{\partial x} + \frac{\partial v^2}{\partial y} \right) dx\, dy$$

and the solution (Kantorovich and Krylov 1958, p. 246 et seq.) to this least squares problem is a pair of Laplace Equations

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} .$$

The well-known finite difference approximation is to require that the value at any mesh point equal the average of its neighbors,

$$u_{ij} = \tfrac{1}{4}\left(u_{i-1j} + u_{i+1j} + u_{ij-1} + u_{ij+1} \right) .$$

An even stronger requirement is that some derivative be the average of its neighbors

$$\frac{\partial u}{\partial x_{ij}} = \frac{1}{4}\left(\frac{\partial u}{\partial x_{i-1j}} + \frac{\partial u}{\partial x_{i+1j}} + \frac{\partial u}{\partial x_{ij-1}} + \frac{\partial u}{\partial x_{ij+1}} \right) .$$

In this equation one substitutes the finite approximation to the derivative

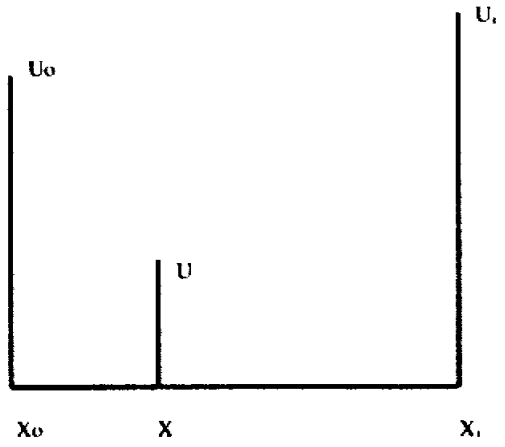
$$\frac{\partial u}{\partial x_{ij}} \simeq \frac{1}{2} \cdot \frac{1}{\Delta x} \cdot (u_{ij+1} - u_{ij-1}),$$

and, after some tedious but simple algebra and setting $\Delta x \equiv 1$, obtains as the condition that

$$\begin{aligned} 20u_{ij} = {} & 8(u_{i+1j} + u_{i-1j} + u_{ij-1} + u_{ij+1}) \\ & - 2(u_{i+1j+1} + u_{i-1j+1} + u_{i+1j-1} + u_{i-1j-1}) \\ & - (u_{i+2j} + u_{ij+2} + u_{i-2j} + u_{ij-2}). \end{aligned}$$

This is a finite difference approximation to the biharmonic equation, and can also be derived as a minimization of the linearized version of the curvature function (Aleksandrov, Kolmogorov, and Lavent'ev 1969; Briggs 1974; Weinstock 1974). In the present context this equation defines the target that we would like to satisfy at every mesh point.

Consider first a one-dimensional version in which an observation lies along the coordinate line between two mesh points Δx units apart:



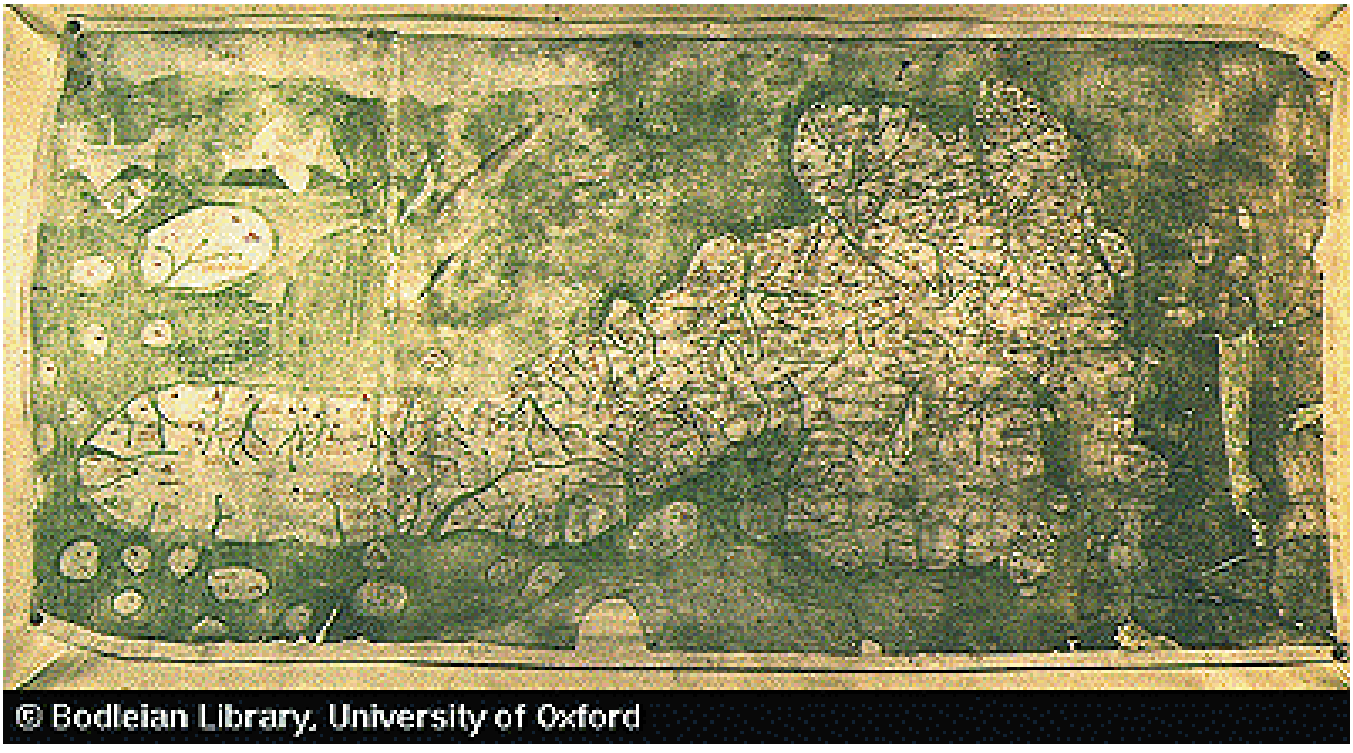
Let u_0 and u_1 be the desired target values, obtained by averaging from neighboring mesh points. Linear interpolation from u_0 to u_1 does not pass through the observation u at x. Now draw a straight line through u to intersect the coordinate lines x_0 and x_1 at u'_0 and u'_1 (not shown). The line through u can be tilted up and down on this point, somewhat like a see-saw, and u'_0 and u'_1 can be varied. Let $\Delta_0 = u_0 - u'_0$ and $\Delta_1 = u_1 - u'_1$. These two variables are related to each other; as one goes up the other goes down when we pivot about u. A little algebra shows that the relation is linear, $\Delta_0 = A + B\Delta_1$. One would now like to come as close to the target as possible, while passing through u. This is achieved by minimizing the sum $\Delta^2_0 + \Delta^2_1 = \varepsilon^2$, that is, setting $\partial\varepsilon^2/\partial\Delta_1 = 0$. The rest is a straightforward matter of calculus. The two-dimensional situation is essentially the same but requires linear interpolation from the four surrounding mesh points, and minimization of the sum of the squared departures from the four target values. The algebra is simple but dreary; it took me forty hours to derive the equations used. The values obtained on

each pass influence the targets assigned to the neighboring points, and convergence to smooth derivatives and a good fit to the data are assured, when using an appropriate mesh size. The iterations are applied to obtain $u^* = f(x, y)$ and to obtain $v^* = g(x, y)$ separately. Southwell (1956) gives equations in a "two-diagram technique" in which the finite differences for u^* and v^* are coupled to each other in a biharmonic solution. His relation between u^* and v^* is derived from the compatibility condition for elastic strain. When programmed, his technique yielded solutions that did not differ from those obtained by the current method. Both allow smooth folds in the mapping (J passes through zero smoothly) and the result is not always monotonic. A linearly monotonic mapping can be defined as one that leaves unchanged the orientation of all triangles when converted from the original plane to the image plane. In the present context some of the displacement vectors cross each other and a monotonic mapping is incompatible with data fidelity. Thus a choice must be made between these two objectives. Fain (1975) presents an interesting procedure that avoids multivaluedness in his mappings.

The curvilinear target-fitting procedure is described above as the solution of an elliptic partial differential equation. In order to solve such an equation it is necessary to specify boundary conditions. In the present instance a Neumann condition is used, with $\partial u^*/\partial x = \beta$ along the left and right edges of the frame, and $\partial u^*/\partial y = \gamma$ along the bottom and top edges. These values are obtained from the linear least squares plane $u = \alpha + \beta x + \gamma y$, fit to the entire set of observations. Similar normal derivatives are used in estimating v^*. In retrospect this is not a particularly felicitous choice. A better procedure might be to require that $\partial u^*/\partial n = 0$ along the boundary of the convex hull of the observations. In this respect the method of Bookstein (1977) is an improvement. Alternately a specification might be made from the outer two convex hulls of the data.

EXAMPLES

The first example uses measurements made on a large facsimile of a fourteenth-century map (Parsons 1958). A crude reproduction of the original is given in Figure 2. The coordinate values were obtained as a student project in 1966, with modern latitudes and longitudes from atlases (Table 1). In this particular case many additional towns can be identified, and landmarks along the outline could have been included. A manual interpolation of the terrestrial graticule has been



constructed by J. Reyer, and is included here for comparison. (Figure 3). The Orkneys should obviously be considered an inset, and not interpreted as continuously connected to the main body of the map.

TABLE 1

Spherical and Plane Coordinates for a Selection of Points on the Gough Map

Latitude*	Longitude*	X Coord.**	Y Coord.**	Town
58.43	—3.12	6.7	38.1	Wick
57.17	—2.08	8.9	32.5	Aberdeen
56.47	—3.03	8.8	28.5	Dundee
56.47	—3.33	7.2	28.2	Scone
55.88	—4.17	5.1	26.0	Glasgow
56.43	—3.50	5.2	28.3	Dunkeld
55.88	—3.00	7.7	24.1	Newbattle
55.37	—4.55	4.5	23.8	Ayr
55.28	— 1.72	9.9	21.8	Alnwick
55.42	—2.77	7.0	21.2	Hawick
54.83	—2.12	9.5	19.5	Hexham
54.70	—2.43	9.0	18.7	Alston
54.53	—3.53	5.9	18.9	Workington
53.41	—3.00	7.7	13.6	Liverpool
53.18	—4.53	4.2	14.4	Aberifraw
52.83	—4.30	3.0	11.5	Criccieth
52.40	—0.87	12.8	9.8	Market Harborough
51.68	— 1.03	13.6	7.1	Thame
52.08	—2.22	10.2	7.9	Worcester
51.45	—2.57	10.1	4.7	Bristol
51.68	0.37	17.2	7.8	Chemsford
51.10	1.22	19.9	5.7	Dover
50.78	0.30	16.5	3.4	Shoreham-by-Sea
50.63	— 1.30	14.1	2.1	Newport
50.62	—3.50	7.5	1.8	Exeter
51.37	—3.13	8.3	4.9	Cardiff
50.28	—4.13	5.5	0.5	Plymouth
50.05	—5.45	0.9	0.5	Penzance
51.82	—4.83	3.1	4.8	Haverfordwest
52.60	1.08	18.3	12.0	Norwich
52.92	0.05	14.6	12.6	Boston
53.38	— 1.45	11.0	12.9	Sheffield
52.27	—2.68	8.2	8.5	Ludlow
53.62	0.03	14.4	15.5	Patrington
53.83	— 1.55	10.9	14.4	Leeds
57.87	—4.08	7.7	37.3	Dornoch
57.45	—4.28	9.0	33.7	Inverness

* In decimal degrees; **in inches.

Coastal Outline Longitude, Latitude for 14 points: —3.57, 50.0; 0.87, 50.8; 1.09. 52.6; — 1.55. 56.1; —0.81. 57.6; —2.36. 57.8; —1.86, 58.6; —3.10, 58.6; —3.72. 56.8; — 1.86, 53.5; — 2.85, 53.4; —3.29. 51,6; —2.36. 51.3; —3.57, 50.0

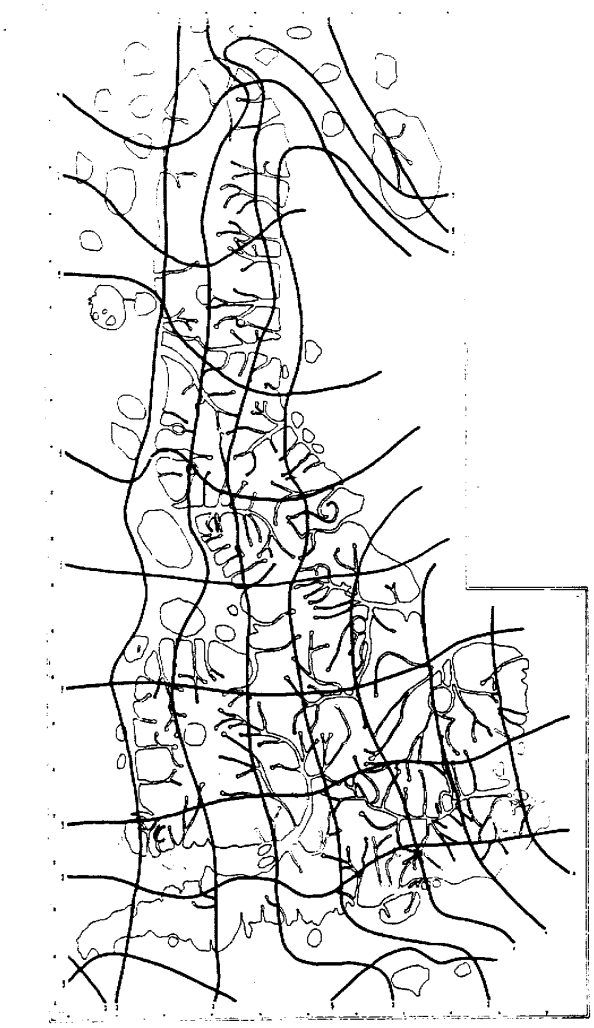


FIG. 3. Gough Map. Interpolation by J. Reyer, University of Michigan, 1966

In a computer analysis one can convert the latitudes and longitudes to kilometers on a map projection,

$$x = \Delta\lambda \cos \phi_0 - \Delta\lambda \, \Delta\phi \sin \phi_0$$

$$y = \Delta\phi + \tfrac{1}{2} \Delta\lambda \, \Delta\lambda \cos \phi_0 \sin \phi_0$$

with $\Delta\varphi = \varphi - \varphi_0$, $\Delta\lambda = \lambda - \lambda_0$, $\varphi_0 =$ 53.68°N, $\lambda_0 =$ 2.46°W, based on a sphere of 6,385-kilometer radius. The least squares estimate of the relation between the coordinates is

then

$$\begin{pmatrix} u \\ v \end{pmatrix}_{est} = \begin{pmatrix} 0.0410342 & -0.000325386 \\ 0.000325386 & 0.0410342 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 9.23143 \\ 15.5383 \end{pmatrix}$$

with a root mean square error of 39.644 kilometers for a correlation of r^2 = 96.12%. The scale change is from kilometers to inches, 0.041034, and there is a minor rotation of 0.04543 degrees. If the map projection conversion is not applied, relation between the coordinates is

$$\begin{pmatrix} u \\ y \end{pmatrix}_{est} = \begin{pmatrix} 4.15092 & 0.237396 \\ -0.237396 & 4.15092 \end{pmatrix} \begin{pmatrix} \phi \\ \lambda \end{pmatrix} + \begin{pmatrix} 6.78802 \\ -207.811 \end{pmatrix}$$

with an r^2 of 84.96% and a RMSE of 0.8061 degrees. The scale change is 4.1577, and the rotation is - 3.27 degrees. Using a lattice of eleven rows by ten columns, the nonlinear fitting procedure estimates the thirty-seven observations with a residual interpolation RMSE of 0.1477 degrees; this is an 82 percent reduction from the original RMSE. A conservative estimate of the amount of "explanation" is 96.64 percent. In terms of the deviation-from-the-mean-explained the r^2 = 99.71 percent. The RMSE could be reduced further by using a finer lattice.

The residual difference between the images is shown in the first computer drawing (Figure 4), made after the sets of data have been scaled and rotated to give the best L^2 fit, and positioned so that their averages coincide. Another way of illustrating the relation between the two sets of coordinates is via "before" and "after" triangulations, shown next in Figure 5. The result of the nonlinear fitting designed to describe the differences between the images is given in Figure 6, in the form of interpolation vectors. Figure 7 shows the nonlinear transformation by a warped grid with the interpolated coastal outline superimposed. The strain tensor is shown for a 23-by-20 lattice.

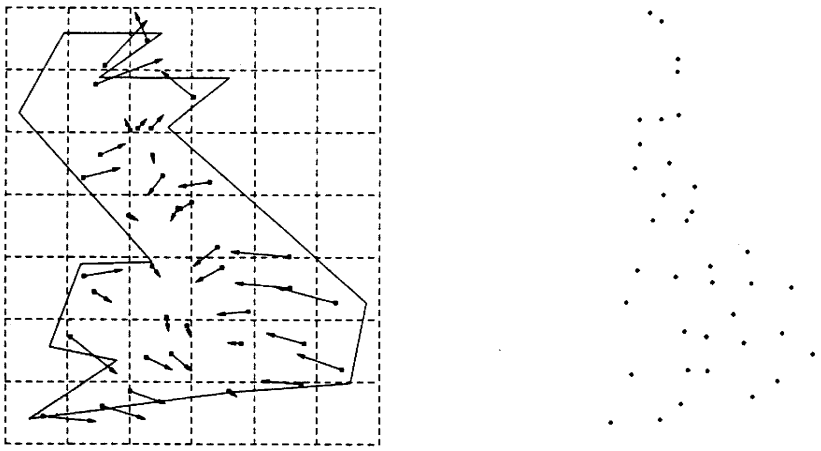


FIG. 4. Gough Map: Difference between the Images after a Euclidean Fit. Left: Original Latitude/Longitude. Right: The Target Image after a Euclidean Fit

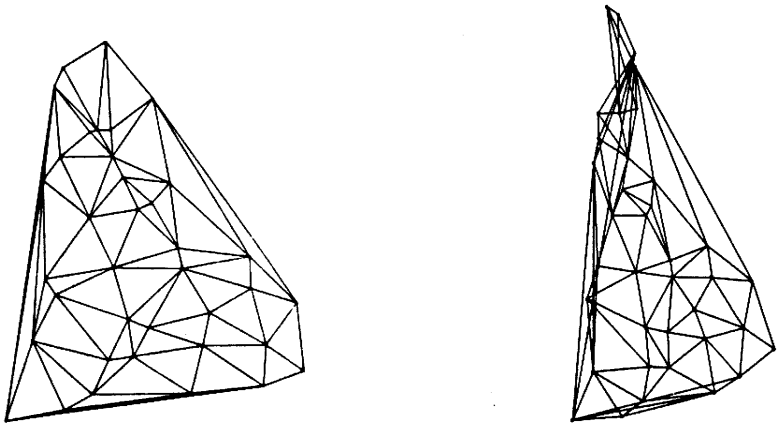


FIG. 5. Gough Map. Left: The "Before" Triangulation, Original. Right: The "After" Triangulation, Target

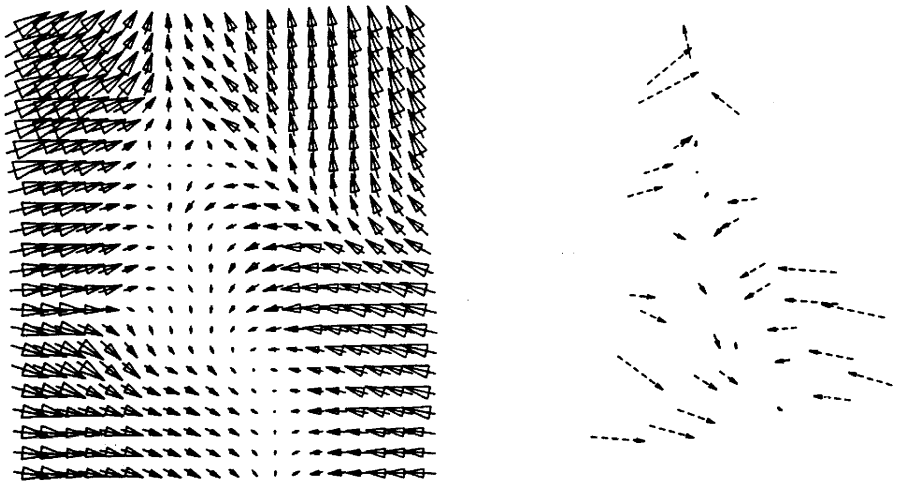


FIG. 6. Gough Map. Left: Interpolated Difference as a Vector Field (half-scale for clarity). Right: Interpolated Displacements

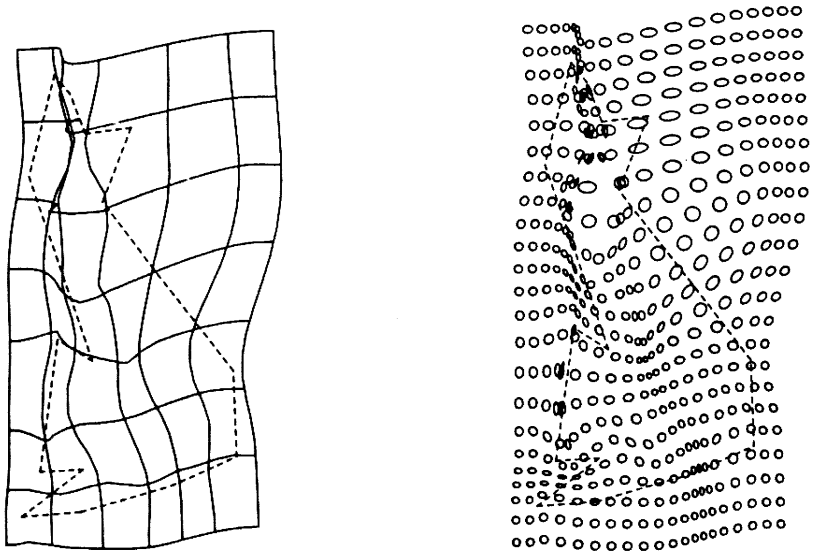


FIG. 7. Gough Map. Left: Warped Grid of Latitude and Longitude; The Transformed Image. Right: Tissot's Indicatrix; Image, Strain Tensor

The analysis has quickly found two misplaced towns, Dornock and Inverness. These cause the overlap warp in the northern part of the map. Recomputing without them is illustrated in Figures 8 and 9, which follow. They are distinguished by N = 35. There are now only five violations of the piecewise linear monotonicity as defined by the triangulation (instead of nine), and the Jacobian determinant is now everywhere positive. The interpolation Tables 2 and 3 give the functions. On the basis of these data, several measures of the total distortion are available. The maximum value of stretching, as measured by Tissot's indicatrix, is 1.4435; the minimum is 0.6975, with the average values of the indicatrix axes being a = 1.2289, b = 1.0829. The sum of the squares of the partial derivatives, taken over all of the 110 lattice points is 195.853, in units of length squared; dividing by 110 yields 1.7805 and the square root of this value is 1.3343, in tolerable agreement with the value obtained from Tissot's indicatrix. The triangulation requires 144.982 units of length (degrees) for its total connections before the transformation and only 134.984 afterwards. Thus the lengths are only 93 percent of their original values

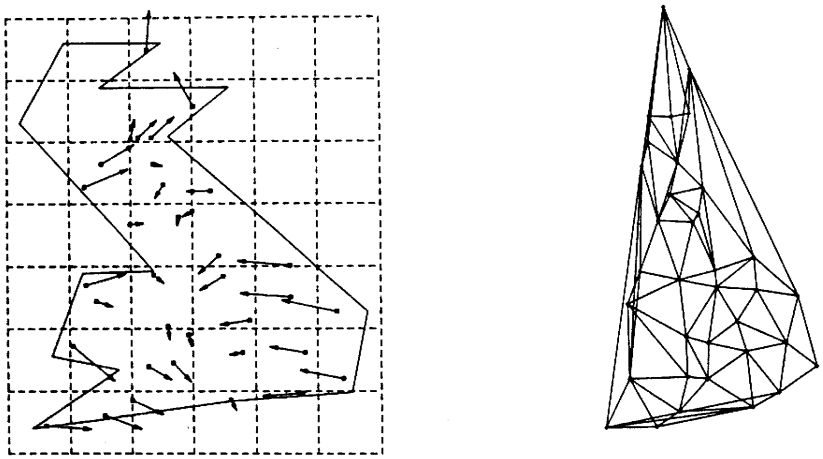


FIG. 8. Gough Map. Left: Difference between the Images after a Euclidean Fit, Original, Latitude and Longitude. Right: The "After" Triangulation, Target. $N = 35$.

(neglecting the spherical shape of the earth). This result, which indicates an overall shrinking, is a direct contradiction to the values obtained from Tissot's indicatrix. The reason is clearly because the triangulation is based directly on the observations; the partial derivatives are estimates and are evaluated at each of the lattice points, 64 percent of which are extrapolations outside of Great Britain, where there are no observations. It is obvious which is the better measure in this instance. The interpolation tables shown have a maximum distance error of 0.4 degrees, and a RMSE of 0.1452 degrees.

TABLE 2

Interpolation Table $\hat{U} = f(x, y)$ for Variable Number 1

−5.412	−4.697	−4.016	−3.371	−2.746	−2.114	−1.420	−0.708	0.008	0.723
−5.413	−4.697	−4.016	−3.371	−2.746	−2.116	−1.424	−0.713	0.001	0.717
−5.349	−4.634	−3.942	−3.306	−2.732	−2.249	−1.486	−0.750	−0.027	0.688
−5.249	−4.534	−3.790	−3.171	−2.588	−2.112	−1.464	−0.768	−0.055	0.661
−5.202	−4.487	−3.671	−3.182	−2.649	−2.087	−1.456	−0.797	−0.091	0.625
−5.361	−4.645	−3.951	−3.289	−2.608	−2.036	−1.469	−0.883	−0.127	0.589
−5.436	−4.721	−4.098	−3.331	−2.499	−2.008	−1.455	−0.884	−0.061	0.655
−5.339	−4.623	−3.953	−3.273	−2.539	−1.845	−1.251	−0.616	0.176	0.891
−5.227	−4.511	−3.721	−2.992	−2.236	−1.529	−0.976	−0.442	0.411	1.127
−5.356	−4.641	−3.761	−2.904	−2.089	−1.250	−0.801	−0.424	0.358	1.074
−5.366	−4.659	−3.757	−2.904	−2.089	−1.250	−0.801	−0.424	0.358	1.074

Calculated from the data. Thirty-five observations.
$N = 110$ X MIN = −6.688 Y MIN = 49.230 MESH SIZE = 1.014 Z MAX = 1.127 Z MIN = −5.436 ZBAR = −2.246 ZSIG = 1.937

TABLE 3

Mid-fourteenth Century Gough Map Interpolation Table $\hat{V} = g(x, y)$ for Variable Number 2

60.627	60.500	60.440	60.348	60.367	60.462	60.542	60.618	60.659	60.659
59.372	59.350	59.290	59.222	59.240	59.331	59.404	59.472	59.509	59.509
58.241	58.198	58.134	58.059	58.081	58.199	58.267	58.330	58.362	58.362
57.015	56.971	56.889	56.720	56.624	56.829	57.042	57.135	57.176	57.176
55.802	55.774	55.712	55.426	55.214	55.365	55.719	55.866	55.932	55.932
54.632	54.582	54.486	54.288	54.110	54.092	54.381	54.562	54.696	54.696
53.357	53.324	53.427	53.252	53.052	52.918	53.213	53.419	53.580	53.585
51.925	51.841	51.730	51.954	51.989	52.081	52.248	52.417	52.572	52.555
50.869	50.780	50.568	50.839	50.942	51.051	51.205	51.318	51.502	51.486
50.244	50.239	50.018	49.953	49.918	49.936	50.055	50.150	50.315	50.377
49.118	59.152	48.884	48.813	48.781	48.784	48.936	49.069	49.248	49.248

Calculated from the data. Thirty-five observations.
$N = 110$ X MIN = −6.688 Y MIN = 49.230 MESH SIZE = 1.014 Z MAX = 60.659 Z MIN = 48.781 ZBAR = 54.619 ZSIG = 3.695

The second example uses data collected by E. Franckowiak (1973) in the city of Toledo, Ohio (Figure 10). The displacement vectors are the difference between the true location of stores, and their estimated locations, averaged over 345 people. The warped grid, principal axes and distortion tensor are also shown. The principal axes form an orthogonal grid and this bears some resemblance to the arterial street pattern in Toledo. Monotonicity violations are again observed in the original data, and the final curvilinear correlation is only 70 percent.

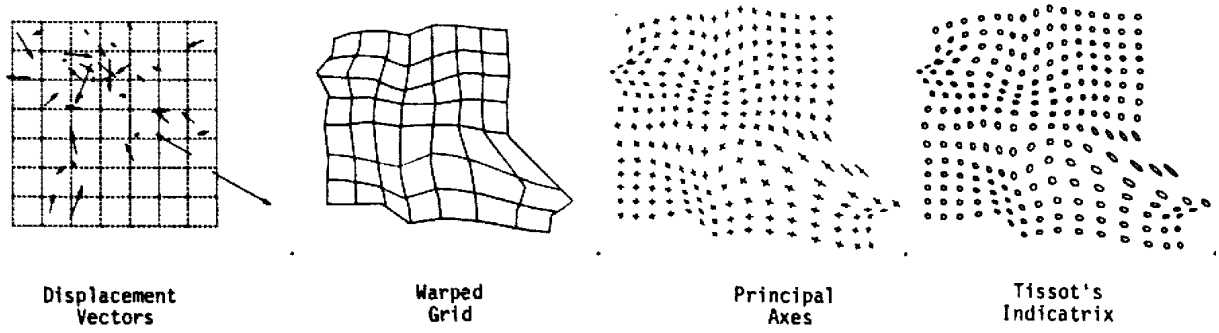


FIG. 10. Perceptions of Toledo. Data from Franckowiak (1973)

The growth of tobacco leaf has been studied carefully *by* Avery (1933) and his data were further examined by Richards and Kavanagh (1945). For comparability these data have again been used, with results (some of which are new) in pictorial form (Figure 11). The change in form due to growth is illustrated by displacements and by a warped grid. The growth of this leaf can almost be represented exactly by an affine transformation, $r^2 = 92\%$, and thus seems very simple, if the assumption that it is a mathematical continuum without biological cells can be tolerated. Clearly such an assumption can only be invoked after the leaf reaches some minimum size. The deformed rectangular coordinates seem to serve nicely, and Avery actually drew such a system on a growing leaf. But it has been suggested (Tobler 1963) that Avery should have used a coordinate system that corresponds more closely to the actual geometry of the growing leaf. One coordinate should be parallel to the edge of the leaf and the other coordinate should be at right angles to this and lie along the lines of growth. The lines that remain orthogonal after an arbitrary continuous transformation are the axes of Tissot's indicatrix, and these form the natural coordinate system. The program has drawn these axes on both the grown leaf and on the small leaf. The length of the axes in each instance is what they would be if they were of unit size in the opposite image, and thus they show the relative amount of linear exaggeration in the principal

Displacement Vectors Warped Grid Principal Axes Tissot's indicatrix

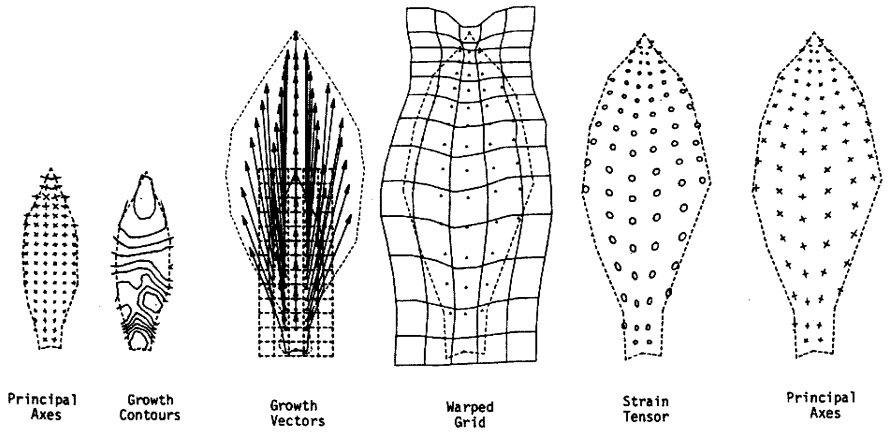


FIG. 11. Tobacco Leaf Growth. Data from Avery (1933)

Fig. 11. Tobacco Leaf Growth. Data from Avery (1933)

directions at each position on the leaf. The only possible postulate is that the lines of growth lie along the major axis of the indicatrix, and that this would hold for other biological forms, and in three dimensions as well. Diagrammed as contours is the sum of the squares of the partial derivatives of the transformation at each point of the leaf. The volume under this surface can be computed to assign a single numerical value to the entire transformation and this can be taken as one representation of the total amount of effort required to change the one form into the other. The natural question to ask is whether this number can be related to the biological or chemical energy requirements for the growth of the leaf. These contours may be compared to those previously derived in the literature. This is such a simple transformation that the curvilinear fitting is essentially perfect.

Another biological example of bidimensional regression, using data employed by Sneath (1967) is shown in Figure 12. The biological significance of the illustration is beyond my provenance. A linear interpolation of outline skulls intermediate between Australopithecus and Homo is also shown.

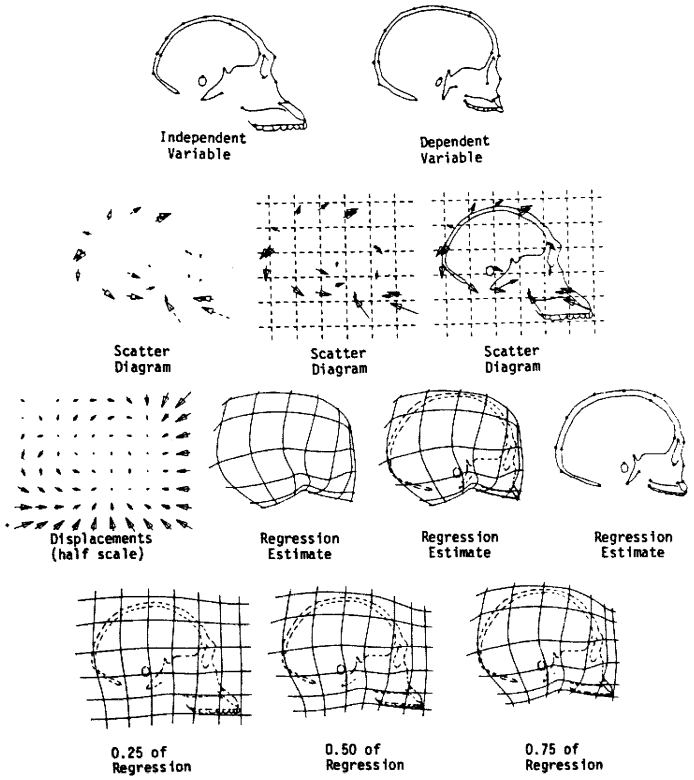


FIG. 12. Outline Skills Intermediate between Australopithecus and Homo. Data from Sneath (1967)

Finally Figure 13 presents a comparison of two sketches drawn and kindly lent by Leo Goodman. It is not difficult to identify several corresponding points, and size differences are easily eliminated. But the two people are seen at different angles and from different perspectives. It will be noticed that the predicted face, shown as dashed lines along with the warped grid, is not identical with the target face; the r^2 = 91%. Even at the data points there is some slight discrepancy; look at the left end of the left eyebrow, and at the ears. And there are lines transformed from the original that were not in the target, for example, at the right edge of the mouth. In part these differences stem from deficiencies in the interpolation procedure. In part they are due to inadequate quantities of identifiable point pairs. Some are due to the continuous nature of the transformation relative to the different perspectives, that is, topological differences. Judicious use of interactive computer graphics would allow some of these difficulties to be overcome very easily.

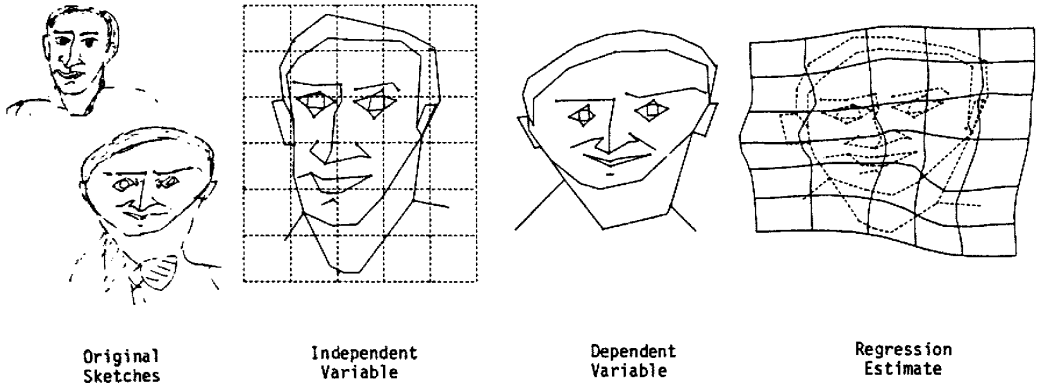


FIG. 13. Comparison of Faces. Data courtesy of L. Goodman

APPENDIX

A computer program has been prepared that performs an empirical transformation regressing an independent plane configuration against a similar dependent configuration as described in the text and illustrated in Figure 14. The program constructs the mapping by estimating two bivariate tables, with arguments on a square lattice interpolated from the irregularly arranged original observations. The tables are constructed so that the transformation has smooth derivatives and so as to agree, as nearly as possible, with the empirical data. The curvilinear regression coefficients are represented by a spatially varying, but coordinate invariant, second-order tensor field. The program has an option to draw this field. The bidimensional regression can automatically be illustrated by a warped coordinate grid, by comparison of the original configuration and the regression estimate, by a field of displacement vectors, by a diagram of the principal strains, and by contour maps. Intermediate configurations and extrapolations are also available.

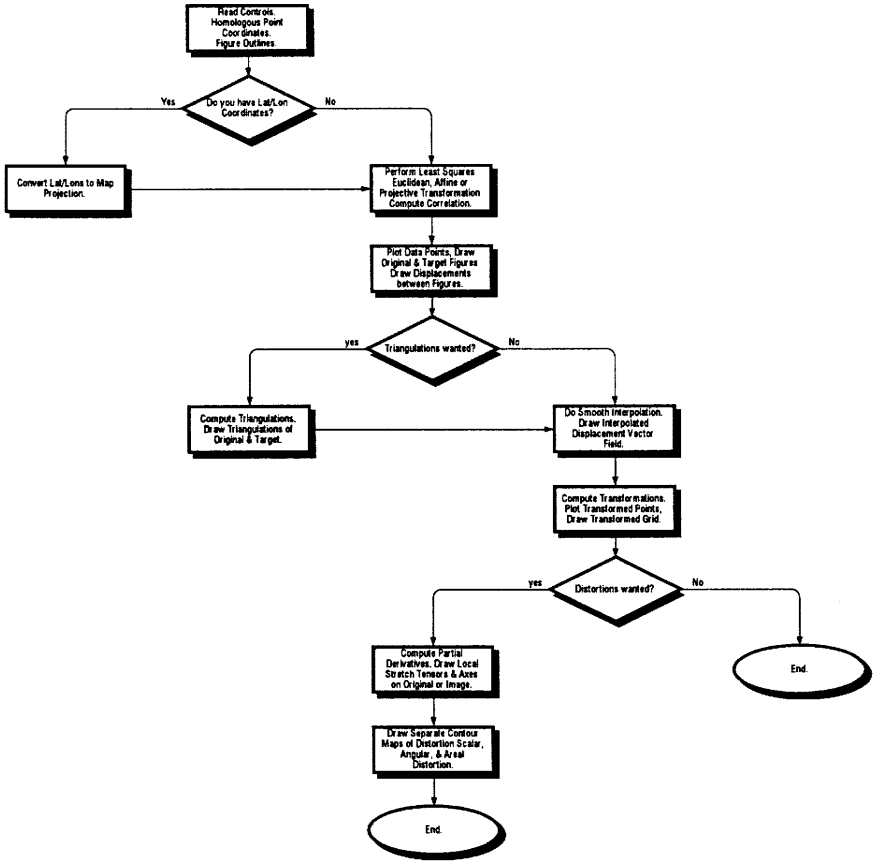


FIG. 14. Simplified Bidimensional Regression Program Flow

LITERATURE CITED

Aleksandrov, A., A. Kolmogorov, and M. Lavrent'ev (1969). *Mathematics: Its Content, Methods, and Meaning*, vol. 2, ch. 7. Cambridge: MIT Press.

Appleby, R., and C. Jones (1978). "The Analogue Videoreshaper - A New Tool for Palaeontologists." *Paleontology* 19, 565-586.

Avery, G. (1933). "Structure and Development of the Tobacco Leaf." *American Journal of Botany* 20. 565 et seq.

Baum, D. (1977). "Cognitive Maps: The Mental Representation of Geographic Distance." Ph.D. thesis in psychology, University of Michigan.

Bernstein, R. (1976). "Digital Image Processing of Earth Observation Sensor Data." *IBM Journal of Research and Development* 20, January, 40-57.

Blackith, R., and R. Reyment (1971). *Multivariate Morphometrics.* New York: Academic Press.

Bookstein, F. (1977). "The Measurement of Biological Shape and Shape Change." PhD. thesis, University of Michigan.

_____ (1977). "The Study of Shape Transformation after D'Arcy Thompson." *Mathematical Bioscience* 34, 177-219.

Briggs, 1. (1974). "Machine Contouring Using Minimum Curvature." *Geophysics* 39, 39-48.

Brown, W. (1965). "Historical Note on a Recurrent Combinatorial Problem." *American*

Mathematics Monthly 72, 973-77.
Bunge, W. (1966). *Theoretical Geography,* revised ed. Lund, Gleerup.
Cottafava, G., and C. Le Moli (1969). "Automatic Contour Map." *Communications, Association of Computing Machinery* 12, 386-91.
Cram, I. (1970). "Computer Interpolation and Contouring of Two-Dimensional Data: A Review." *Geoexploration* 8, 71-86.
Dooley, J. (1976). "Two-Dimensional Interpolation of Irregularly Spaced Data Using Polynomial Splines.' *Physics of the Earth and Planetary Interiors* 12, 180-87.
Dürer, A. (1973). *The Human Figure,* Robert Strauss, ed. New York: Dover.
Ellison, T. (1954). "On the Correlation of Vectors." *Quarterly Journal of the Royal Meteorological Society* 80. 93-96.
Erickson, H. (1966). "Relative Elemental Rates and Anisotropy of Growth in Area: A Computer Program. *Journal of Experimental Botany* 17, 390 et seq.
Fain, V. (1975). *Algoritmicheskoy Modelirovaniy Fonnoobrazoniya.* Moskva: Nauk.
Forrest, D. (1974). *Francis Galton: The Life and Work of a Victorian Genius.* New York: Taplinger.
Franckowiak, E. (1973). "Location Perception and the Hierarchical Structure of Retail Centers." Ph.D. thesis, University of Michigan.
Gold, C., T. Charters, and J. Ramsden (1977). "Automated Contour Mapping Using Triangular Element Data." *Computer Graphics* 11, 170-75.
Grafarend, E. (1975). "Geodetic Prediction Concepts.: In *Methoden and Verfahren der matematischen Physik* 13 (Mathematical Geodesy, part 2), pp. 161-200. Zürich: Bibliographisches Institut.
Green, G., S. Outcalt, and W. Tobler (1977). "Reconstruction and Parsing of Topographic Data." Manuscript.
Grenander, U. (1976). *Pattern Synthesis.* New York: Springer.
Guillaume, A. (1977). *Analyse des Variables Regionalisees.* Paris: Dion.
Helmert, F. (1880). Mathematischen and Physicalischen Theorien der Höheren Geodasie. Leipzig: Teubner.
Huxley, J. (1932). *Problems of Relative Growth.* London: Methuen.
Kantorovich, L., and V. Krvlov (1958). *Approximate Methods of Higher Analysis.* The Hague: Noordhoff lnterscience.
Knudson, W., and D, Nagy (1974). "Discrete Data Smoothing by Spline Interpolation with Application to Initial Geometry of Cable Nets." *Computer Methods in Applied Mechanics and Engineering,* 4, 321-48.
Kraus, K. (1972). "Film Deformation Correction with Least Squares Interpolation." *Photogrammetric Engineering* May, 487-93.
Krige, D. (1966). "Two-dimensional Weighted Moving Average Trend Surfaces for Ore Evaluation," In *Proceedings,* Symposium on Mathematical Statistics and Computer Applications in Ore Valuation, pp. 13-38. Johannesburg.
Lancaster, P., and K. Salkaulkas (1977). *A Survey of Curve and Surface Fitting.* Manuscript, University of Calgary.
Lawson, C. (1972). "Transforming Triangulations." *Discrete Mathematics* 3, 365-72.
Lawson, C. L. (1978). "Software for C Surface Interpolation." In *Mathematical Software Ill,* edited by J. R. Rice. Academic Press.
Lawson, C., and R. Hanson (1974). *Solving Least Squares Problems.* Englewood Cliffs, N.J.:

Prentice Hall.
Liggett, J. (1974). *The Human Face.* New York: Hem and Day.
Mansfield, L. (1972). "Optimal Approximation and Error Bounds in Spaces of Bivariate Functions." *Journal of Approximation Theory* 5, 77-96.
Masuyama, M. (1939). "Correlation between Tensor Quantities." *Proceedings,* Physico-Mathematical Society of Japan, 3d series, 221, 638-47.
Matheron, G. (1971). *The Theory of Regionalized Variables and Its Applications.* Cahiers du Centre de Morphologie Mathematique de Fontainhleau, 5. Ecole National Superieure des Mines de Paris.
Moritz, H. (1970). "Eine allgemeine Theorie der Verarbeitung von Schwermessungen nach Kleinsten Quadraten." *Heft Nr 67A* Deutsche Geodatische Kommission, Munchen.
___ (1975). "Elements of Collocation." In *Methoden and Verfahren der Matematischen Physik* 13 (Mathematical Geodesy, part 2), pp. 125-59. Zürich: Bibliographisches Institut.
Parsons, E. (1958). *The Map of Great Britain circa A. D. 1360 Known as the Gough Map*. Oxford: University Press, 38 pp.
Pearson, K. (1924). *The Life, Letters, and Labours and Francis Galton,* vol. 2, Cambridge: University Press.
Richards, 0., and A. Kavanagh (1945). "The Analysis of Growing Form." In Essays *on Growth and Form,* edited by P. Medawar and W. Clark. Oxford: Clarendon.
Schumaker, L. (1976). "Fitting Surfaces to Scattered Data." In *Approximation Theory II,* edited by G. Lorentz et al., pp. 203-68. Academic Press.
Schut, G. (1976). "Review of Interpolation Methods for Digital Terrain Models." *The Canadian Surveyor* 30, 389-412.
Shamos, M. (1977). *Computational Geometry.* New York: Springer.
Southwell, R. (1956). *Relaxation Methods in Theoretical Physics,* pp. 312-14. Oxford: Clarendon Press.
Sneath, P. (1967). "Trend-Surface Analysis of Transformation Grids," *Journal of Zoology* 151, 65-122.
Steiner, D. (1974). "Digital Geometric Picture Correction Using a Piecewise Zero-Order Transformation." *Remote Sensing Environment* 3, 261-83.
Steiner, D., and M. Kirby (1977). "Geometrical Referencing of Landsat Images by Affine Transformation and Overlaying of Map Data." *Photogrammetria* 33, 41-75.
Stevens, P. (1974). *Patterns in Nature.* Boston: Little-Brown.
Swain, C. (1976). "A Fortran IV Program for Interpolating Irregularly Spaced Data Using the Difference Equations for Minimum Curvature." *Computers and Geosciences* 1, 231-40.
Thompson, D'Arcy (1917). *On Growth and Form.* Cambridge: University Press.
Tissot, M. (1881). *Mémoire sur la répresention des surfaces.* Paris: Gauthier.
Tobler, W. (1963). "D'Arcy W. Thompson and the Analysis of Growth and Form." *Papers,* Michigan Academy of Science, Arts, and Literature 48, 385-90.
_____ (1965). "Computation of the Correspondence of Geographical Patterns." *Papers,* Regional Science Association 15, 131-39.
_____ (1966). "Medieval Distortions: The Projections of Ancient Maps." *Annals,* Association of American Geographers 56, 351-60.
_____ (1974). "Local Map Projections." *The American Cartographer* 1, 51-62.
_____ (1975). "Linear Operators Applied to Areal Data," In *Display and Analysis of Spatial Data,* edited by J. Davis and M. MacCullogh. New York: J. Wiley.

_____(1976). “The Geometry of Mental Maps.” In *Spatial Choice and Spatial Behavior,* edited by R. Golledge and G. Rushton, pp. 69-81. Columbus: Ohio State University.
_____ (1977). “Numerical Approaches to Map Projections.” In *Contributions to Theoretical Cartography,* edited by I. Kretschmer, pp. 51-64. Vienna: Deuticke.
_____ (1978). “The Comparison of Plane Forms.” *Geographical Analysis* 10.
Walker, G., and C. Kowalski (1972). “A New Approach to the Analysis of Craniofacial Morphology and Growth.” *American Journal of Orthodontia* 61, 221 et seq.
Weinstock, R. (1974). *Calculus of Variations,* pp. 228-60. New York: Dover.

Editorial Preface to Tobler’s “Bidimensional Regression”

The preceding article is a portion of a 1977 discussion paper by the same title, published by the Department of Geography, University of California at Santa Barbara, that was accompanied by a bidimensional regression computer program. Both the paper and the program were authored by Professor Tobler. The text published here reproduces the theoretical and applications sections of the discussion paper with minor changes in the abstract and in the sentences that make direct reference to the computer program.

I have been aware of Professor Tobler’s work in this area since before his discussion paper was published. Some time ago, I asked him to submit the paper for possible publication in *Geographical Analysis,* and he kindly obliged. During the review process, I became convinced that it was useful to publish segments of the 1977 text with minimal modifications. Some pertinent additional references that materialized during the course of the review process and others forwarded to me by Dr. Tobler are appended to this note.

Professor Tobler kindly agreed to provide at cost, for IBM-compatible PCs, a disk with the bidimensional regression computer program and with full use instructions. Interested parties should address their requests to Dr. Waldo Tobler, Department of Geography, University of California, Santa Barbara, CA 93106-4060.

I am confident that this paper will call attention to the bidimensional regressions that in my perception are especially well suited to a wide class of geographical analyses, and I am looking forward to receiving submissions to *Geographical Analysis* that pursue research themes in this general area

EMJLIO CASETTI

REFERENCES

Anderson, S. (1993). *Morphing Magic.* Cannel, MD.: Sams. 275 pp.
Bookstein, F. (1991). *Morphometric Tools for Landmark Data.* Cambridge: University Press, 435 pp.
Brown, L. (1992). “A Survey of Image Registration Techniques.” *ACM Computing Surveys* 24, 325-76.
Crimson, W. (1983). “An Implementation of a Computational Theory of Visual Surface Interpolation.” *Computer Vision, Graphics, and Image Processing* 22, 39-69.
Hall, V. (1993). “Morphing in 2-D and 3-D.” *Dr. Dobb’s Journal* 18, 18-26, 92-94, 132.
Mekenkamp, P., and 0. Koop (1986). “Nauwkeurigheits analyse van oudekaarten met behuip van de computer.” *Caert Thresoor* 5, 45-52.
Ruprecht, D., and H. Mueller 91993). “Free Form Deformation with Scattered Data Interpolation Methods.’ *Geometric Modelling* (Computing Supplementum 8), edited by C. Farm, H. Hagen,

and H. Noltemeier, pp. 267-8 1. Vienna: Springer.
Tobler, W., and S. Kennedy (1985). “Smooth Multidimensional Interpolation.” *Geographical Analysis* 17, 251-57.
Watson, D. (1992). Contouring: A Guide to the Analysis and Display of Spatial Data. New York: Pergamon. 321 pp.
Wolberg, C. (1990). *Digital Image Warping.* Los Alamitos: IEEE Computer Society. 318 pp.

†Geographical Analysis, Vol. 26, No. 3 (July 1994) © 1994 Ohio State University Press

Typing kudos are due to T. Everett. E. Gustafson assisted in the preparation of the figures. Computing facilities were made available by the University of Michigan. NSF Grant 50C77-00978 provided partial support. Numerous helpful comments were provided by F. Bookstein.

W R. Tobler is professor of geography at the University of California at Santa Barbara.

See also

T.Nakaya, “Statistical Inference in Bidimensional Regression Models”, *Geographical Analysis*, 29, 2 (April 1997): 169-186.

K.K.Schmid, D.B.Marx, A.Samal, "Weighted Bidimensional Regression", Geographical Analysis, 43,1 (1-13-2011): 1-7.

Online version of Bidimensional Regression at:
www.spatial-modelling.info/Darcy-2-module-de-comparison