SOLSTICE:

AN ELECTRONIC JOURNAL OF GEOGRAPHY AND MATHEMATICS

VOLUME III

NUMBERS 1 (SUMMER) AND 2 (WINTER)

IMaGe

DOWNLOADED AS

MONOGRAPH #16

INSTITUTE OF MATHEMATICAL GEOGRAPHY, MONOGRAPH SERIES.
FOUNDING EDITOR, SANDRA L. ARLINGHAUS
2790 BRIARCLIFF
ANN ARBOR, MICHIGAN 48105
U.S.A.

DECEMBER, 1992

SOLSTICE

Institute of Mathematical Geography

IMAGE

IMAGINATION

Journal of the
Institute of Mathematical Geography

Summer, 1992

SOLSTICE:

AN ELECTRONIC JOURNAL OF GEOGRAPHY AND MATHEMATICS

SUMMER, 1992–11:14p.m., E.D.T., June 20

Volume III, Number 1
Institute of Mathematical Geography
Ann Arbor, Michigan

The purpose of *Solstice* is to promote interaction between geography and mathematics. Articles in which elements of one discipline are used to shed light on the other are particularly sought. Also welcome, are original contributions that are purely geographical or purely mathematical. These may be prefaced (by editor or author) with commentary suggesting directions that might lead toward the desired interaction. Individuals wishing to submit articles, either short or full- length, as well as contributions for regular features, should send them, in triplicate, directly to the Editor–in–Chief. Contributed articles will be refereed by geographers and/or mathematicians. Invited articles will be screened by suitable members of the editorial board. IMaGe is open to having authors suggest, and furnish material for, new regular features.

Send all correspondence to:

Institute of Mathematical Geography
2790 Briarcliff
Ann Arbor, MI 48105-1429
(313) 761-1231
IMaGe@UMICHUM

This document is produced using the typesetting program, TEX, of Donald Knuth and the American Mathematical Society. Notation in the electronic file is in accordance with that of Knuth's *The TEXbook*. The program is downloaded for hard copy for on The University of Michigan's Xerox 9700 laser- printing Xerox machine, using IMaGe's commercial account with that University.

Unless otherwise noted, all regular "features" are written by the Editor-in-Chief.

Upon final acceptance, authors will work with IMaGe to get manuscripts into a format well-suited to the requirements of *Solstice*. Typically, this would mean that authors would submit a clean ASCII file of the manuscript, as well as hard copy, figures, and so forth (in camera-ready form). Depending on the nature of the document and on the changing technology used to produce *Solstice*, there may be other requirements as well. Currently, the text is typeset using TEX; in that way, mathematical formulæ can be transmitted as ASCII files and downloaded faithfully and printed out. The reader inexperienced in the use of TEX should note that this is not a "what-you-see-is-what-you-get" display; however, we hope that such readers find TEX easier to learn after exposure to *Solstice*'s e-files written using TEX!

Copyright will be taken out in the name of the Institute of Mathematical Geography, and authors are required to transfer copyright to IMaGe as a condition of publication. There are no page charges; authors will be given permission to make reprints from the electronic file, or to have IMaGe make a single master reprint for a nominal fee dependent on manuscript length. Hard copy of *Solstice* is available at a cost of $15.95 per year (plus shipping and handling; hard copy is issued once yearly, in the Monograph series of the Institute of Mathematical Geography. Order directly from IMaGe. It is the desire of IMaGe to offer electronic copies to interested parties for free. Whether or not it will be feasible to continue distributing complimentary electronic files remains to be seen. Presently *Solstice* is funded by IMaGe and by a generous donation of computer time from a member of the Editorial Board. Thank you for participating in this project focusing on environmentally-sensitive publishing.

ISBN: ISSN: 1059-5325

SUMMARY OF CONTENT

Computing Areas of Regions With Discretely Defined Boundaries

Harry L. Stern
Polar Science Center
1013 N.E. 40th Street
Seattle, WA 98195

1. Introduction

It is well known that the area of a region in the plane can be computed by an appropriate integration around the boundary of the region [e.g. Hildebrand, page 306]. If the boundary is defined by a sequence of points connected by straight lines (a polygon), the parametric representation of the boundary is particularly simple, and an explicit formula for the area can be derived. Using Stokes' Theorem, this idea can be extended to derive area formulas for regions on non-planar surfaces whose boundaries are defined by a sequence of points connected by appropriate curves. In this note we present exact area formulas for regions in the plane and regions on the sphere whose boundaries are defined by such discrete sets of points.

An application of these formulas arises in computing the area of a region on a map. Suppose that the boundary of the region of interest is traced by an encoding device that records its coordinates, relative to some user-defined (x, y) system, in a computer file. Such a file may contain hundreds or thousands of coordinate pairs. If the map covers a relatively small region, the surface of the earth can be approximated locally by a plane, and the area computed directly from the (x, y) coordinate pairs. If the map covers a large region, the earth can be approximated by a sphere. The (x, y) coordinate pairs are then converted to latitude and longitude using the appropriate map projection equations, and the area on the sphere is computed.

The usual method for computing area is to divide up the two dimensional surface into a large number of small cells, and to add up the areas of those cells that lie inside the boundary of the region. This method is computationally slow, because every cell must be tested for inclusion in the region, and because high accuracy requires a small cell size. In contrast, the formulas derived here, besides being exact, are quickly evaluated on a computer because the computation is proportional to the number of boundary points. The two dimensional area calculation is reduced to a one dimensional boundary calculation.

The next section outlines the general mathematical formulation. Sections 3 and 4 give explicit results for the plane and sphere. A numerical example and concluding remarks are presented in the last section.

2. General Formulation

Stokes' theorem says

$$\iint_S (\nabla \times \mathbf{F}) \cdot \hat{\mathbf{n}} \, dA = \oint_C \mathbf{F} \cdot \frac{d\mathbf{R}}{dt} \, dt \tag{1}$$

where S is the region of a surface bounded by the curve C, $\hat{\mathbf{n}}$ is the unit outward normal on the surface, $\mathbf{R}(t)$ is a parametric representation of C, and $\mathbf{F}$ is an arbitrary vector field. We suppose that the surface is specified in some way (e.g. $x^2 + y^2 + z^2 = 1$ for the unit sphere), so that the unit outward normal $\hat{\mathbf{n}}$ can be determined (e.g. $\hat{\mathbf{n}} = x\hat{\imath} + y\hat{\jmath} + z\hat{\mathbf{k}}$ for the unit sphere). We then choose any vector field $\mathbf{F}$ such that the integrand on the left hand side of (1) is unity in S:

$$(\nabla \times \mathbf{F}) \cdot \hat{\mathbf{n}} = 1. \tag{2}$$

With F determined (though not uniquely) by equation (2), the left hand side of (1) simply reduces to the area of S, giving

$$A = \oint_C \mathbf{F} \cdot \frac{d\mathrm{R}}{dt}\, dt. \tag{3}$$

In order to evaluate the integrand on the right hand side of (3), we need a description of C. Suppose that N points on the surface are given, $\mathbf{P}_1$, $\mathbf{P}_2$, ..., $\mathbf{P}_N$, and that C is defined by connecting these points in sequence, returning to $\mathbf{P}_1$ (define $\mathbf{P}_{N+1} \equiv \mathbf{P}_1$). On each segment, from $\mathbf{P}_k$ to $\mathbf{P}_{k+1}$, let $\mathbf{R}_k(t)$ be a parametric representation of the connecting curve. There are many possible connecting curves to choose from, but the most natural choice is the geodesic, the curve of minimum length (e.g. a straight line in the plane, a great circle on the sphere). The geodesics can be found in principle from a description of the surface (for example, Weinstock pages 61-62). The collection of the N geodesics $\mathrm{R}_k(t)$ connecting the N points $\mathbf{P}_1$, $\mathbf{P}_2$, ..., $\mathbf{P}_N$, constitutes the parametric description $\mathrm{R}(t)$ of C on the right hand side of (3).

Now that we have specified how to construct the integral in (3) as a sum of integrals along the N connecting geodesics, the area formula can be written more explicitly as

$$A = \sum_{k=1}^{N} \int_0^{L_k} \mathbf{F}(s) \cdot \frac{d\mathrm{R}_k}{ds}\, ds \tag{4}$$

where s is the arc length parameter along the geodesic $\mathrm{R}_k(s)$, and L_k is the total arc length of the k-th segment. The geodesics need not necessarily be parameterized by arc length, but this is what we have used in the sections that follow.

The determination in principle of all quantities is now complete. To summarize the steps: Given a surface and a set of points $\mathbf{P}_k, k = 1, 2, \ldots, N$ that defines the boundary of a region on the surface,

(1) Find the unit outward normal on the surface, $\hat{\mathbf{n}}$;

(2) Find a vector field $\mathbf{F}$ that satisfies equation (2): $(\nabla \times \mathbf{F}) \cdot \hat{\mathbf{n}} = 1$;

(3) Find a parameterization $\mathrm{R}_k(s)$ of the geodesic from point $\mathbf{P}_k$ to $\mathbf{P}_{k+1}$;

(4) Form the integrand in equation (4) and do the integration;

(5) Sum the contributions in (4) to get the area of the region.

Some specific cases follow.

3. The Plane

In the plane $z = 0$, the unit outward normal is $\hat{\mathbf{n}} = (0, 0, 1)$ and the condition (2) on the components (F_1, F_2, F_3) of $\mathbf{F}$ is

$$\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} = 1. \tag{5}$$

We choose $F_1 = -y/2$ and $F_2 = x/2$. The geodesics $\mathbf{R}(s) = (x(s), y(s), 0)$ are straight lines, and the integral in equation (4) becomes

$$I_k = \int_0^{L_k} \frac{1}{2}\left(x\frac{dy}{ds} - y\frac{dx}{ds}\right) ds. \tag{6}$$

Let the boundary points $\mathbf{P}_k$ have coordinates (x_k, y_k). The parametric equations for the boundary segment connecting $\mathbf{P}_k$ and $\mathbf{P}_{k+1}$ (of length L_k) are

$$x(s) = x_k + \frac{s}{L_k}(x_{k+1} - x_k) \qquad y(s) = y_k + \frac{s}{L_k}(y_{k+1} - y_k). \tag{7}$$

Substituting these expressions into equation (6) with $\Delta x = x_{k+1} - x_k$ and $\Delta y = y_{k+1} - y_k$ gives

$$\begin{aligned} I_k &= \frac{1}{2}\int_0^{L_k} \left\{ \left(x_k + \frac{s\ \Delta x}{L_k}\right)\left(\frac{\Delta y}{L_k}\right) - \left(y_k + \frac{s\ \Delta y}{L_k}\right)\left(\frac{\Delta x}{L_k}\right)\right\} ds \\ &= \frac{1}{2}\int_0^{L_k} \left\{ \frac{x_k \Delta y}{L_k} - \frac{y_k \Delta x}{L_k}\right\} ds \\ &= \frac{1}{2}(x_k\Delta y - y_k \Delta x) \\ &= \frac{1}{2}(x_k y_{k+1} - y_k x_{k+1}). \end{aligned} \tag{8}$$

It follows that the area of the polygon in the plane whose vertexes are the points (x_k, y_k) is

$$A = \frac{1}{2}\sum_{k=1}^{N}(x_k y_{k+1} - y_k x_{k+1}) \tag{9}$$

where $x_{N+1} \equiv x_1$, $y_{N+1} \equiv y_1$, and the points (x_k, y_k) trace the boundary in a counter-clockwise sense. If the order of the points is reversed, the negative of the area will result.

4. The Sphere

Without loss of generality we consider the unit sphere. It will be convenient to use both rectangular and spherical coordinates. The longitude θ, measured positive eastward, and latitude ϕ, measured positive northward, are related to x, y, z via

$$x = \cos\phi\,\cos\theta \quad y = \cos\phi\,\sin\theta \quad z = \sin\phi \tag{10}$$

and the unit vectors in the θ, ϕ, and radial directions are related to the rectangular unit vectors $\hat{\imath}$, $\hat{\jmath}$, $\hat{\mathbf{k}}$ via

$$\hat{\mathbf{u}}_\theta = (-\sin\theta)\hat{\imath} + (\cos\theta)\hat{\jmath} = \frac{-y}{\sqrt{1-z^2}}\hat{\imath} + \frac{x}{\sqrt{1-z^2}}\hat{\jmath} \tag{11a}$$

$$\begin{aligned} \hat{\mathbf{u}}_\phi &= (\sin\phi\,\cos\theta)\hat{\imath} + (\sin\phi\,\sin\theta)\hat{\jmath} + (-\cos\phi)\hat{\mathbf{k}} \\ &= \frac{xz}{\sqrt{1-z^2}}\hat{\imath} + \frac{yz}{\sqrt{1-z^2}}\hat{\jmath} - \sqrt{1-z^2}\,\hat{\mathbf{k}} \end{aligned} \tag{11b}$$

$$\hat{\mathbf{u}}_r = (\cos\phi\cos\theta)\hat{\mathbf{i}} + (\cos\phi\sin\theta)\hat{\mathbf{j}} + (\sin\phi)\hat{\mathbf{k}} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}. \tag{11c}$$

The unit outward normal on the sphere is just the unit radial vector $\hat{\mathbf{u}}_r$. With the vector $\mathbf{F}$ written in terms of its spherical components $\mathbf{F} = F_\theta\hat{\mathbf{u}}_\theta + F_\phi\hat{\mathbf{u}}_\phi + F_r\hat{\mathbf{u}}_r$, the condition (2) becomes [Hildebrand]

$$(\nabla \times \mathbf{F}) \cdot \hat{\mathbf{u}}_r = \frac{1}{\cos\phi}\left[\frac{\partial}{\partial\theta}(F_\phi) - \frac{\partial}{\partial\phi}(\cos\phi\ F_\theta)\right] = 1. \tag{12}$$

This is most naturally satisfied if we take

$$\frac{\partial}{\partial\phi}(\cos\phi\ F_\theta) = -\cos\phi \qquad \frac{\partial}{\partial\theta}(F_\phi) = 0 \tag{13}$$

or

$$F_\theta = -\tan\phi + \frac{g(\theta)}{\cos\phi} \qquad F_\phi = h(\phi) \tag{14}$$

where g is an arbitrary function of θ, and h is an arbitrary function of ϕ. No radial dependence has been introduced into g and h because we are only interested in the values of $\mathbf{F}$ on the surface $r = constant$. Also, the radial component of $\mathbf{F}$, F_r, is of no consequence: any tangent vector to the sphere, $d\mathbf{R}/dt$, has no radial component, so the dot product $\mathbf{F}\cdot d\mathbf{R}/dt$ annihilates any radial contribution from $\mathbf{F}$. Therefore we take $F_r = 0$.

Now that $\mathbf{F}$ is determined (up to two arbitrary functions), we turn to the parameterization of the boundary. We suppose that N pairs of longitude/latitude coordinates are given, namely θ_k, ϕ_k for $k = 1, 2, \ldots, N$ (with $\theta_{N+1} \equiv \theta_1$ and $\phi_{N+1} \equiv \phi_1$), that form the boundary of the region when the points are connected in the given order. The boundary points will also be denoted by $\mathbf{P}_k$, and by their rectangular coordinates (x_k, y_k, z_k). We can use equation (10) to go from spherical to rectangular coordinates.

To simplify the notation a bit, let $k = 1$ and consider the great circular arc from $\mathbf{P}_1$ to $\mathbf{P}_2$. Let Δ represent the angle subtended at the center of the sphere by $\mathbf{P}_1$ and $\mathbf{P}_2$. Then Δ satisfies $\cos\Delta = \mathbf{P}_1\cdot\mathbf{P}_2$ since all the $\mathbf{P}_k$ are unit vectors. Note that Δ is also the length of the arc from $\mathbf{P}_1$ to $\mathbf{P}_2$. Let α be the arc length parameter along the great circle from $\mathbf{P}_1$ to $\mathbf{P}_2$, and let $\mathbf{R}(\alpha)$ be the position vector along the great circle. Since $\mathbf{R}(\alpha)$ lies in the plane spanned by $\mathbf{P}_1$ and $\mathbf{P}_2$, we can write

$$\mathbf{R}(\alpha) = A(\alpha)\mathbf{P}_1 + B(\alpha)\mathbf{P}_2 \tag{15}$$

where $A(\alpha)$ and $B(\alpha)$ are determined from the following two conditions:

(1) $\mathbf{R}(\alpha)$ lies on the unit sphere: $\mathbf{R}\cdot\mathbf{R} = 1$;

(2) The angle between $\mathbf{P}_1$ and $\mathbf{R}(\alpha)$ is α: $\mathbf{P}_1\cdot\mathbf{R} = \cos\alpha$. Using equation (15) for $\mathbf{R}$ and the fact that $\mathbf{P}_1\cdot\mathbf{P}_2 = \cos\Delta$, these conditions translate into

$$A^2 + B^2 + 2AB\cos\Delta = 1 \qquad A + B\cos\Delta = \cos\alpha \tag{16}$$

respectively. Solving for A and B, we find

$$\mathbf{R}(\alpha) = \frac{\sin(\Delta - \alpha)}{\sin\Delta}\mathbf{P}_1 + \frac{\sin(\alpha)}{\sin\Delta}\mathbf{P}_2. \tag{17}$$

This is the arc length parameterization for the great circle through $\mathbf{P}_1$ and $\mathbf{P}_2$.

With $\mathbf{R}(\alpha)$ determined, the next step is to compute $d\mathbf{R}/d\alpha$ and then $\mathbf{F}\cdot d\mathbf{R}/d\alpha$. Computation of $d\mathbf{R}/d\alpha$ is simple, but we want to express the result in terms of the unit vectors $\hat{\mathbf{u}}_\theta$ and $\hat{\mathbf{u}}_\phi$, to facilitate taking the dot product with $\mathbf{F}$. Toward this end, write

$$\frac{d\mathbf{R}}{d\alpha} = G(\alpha)\hat{\mathbf{u}}_\theta + H(\alpha)\hat{\mathbf{u}}_\phi \tag{18}$$

where $G(\alpha)$ and $H(\alpha)$ are determined as follows. Let $'$ denote $d/d\alpha$ and write $\mathbf{R}(\alpha) = (x(\alpha), y(\alpha), z(\alpha))$ where the functions x, y, z are given explicitly by the components of equation (17). Then the dot product of equation (18) with $\hat{\mathbf{u}}_\theta$ and $\hat{\mathbf{u}}_\phi$ gives, respectively, $G(\alpha)$ and $H(\alpha)$. Using equations (11a,b) to express $\hat{\mathbf{u}}_\theta$ and $\hat{\mathbf{u}}_\phi$ in terms of $\hat{\imath}, \hat{\jmath}, \hat{\mathbf{k}}$ we have

$$\begin{aligned} G(\alpha) &= \mathbf{R}'\cdot\hat{\mathbf{u}}_\theta \\ &= (x'\hat{\imath} + y'\hat{\jmath} + z'\hat{\mathbf{k}})\cdot\left[\frac{-y}{\sqrt{1-z^2}}\hat{\imath} + \frac{x}{\sqrt{1-z^2}}\hat{\jmath}\right] \\ &= \frac{xy' - yx'}{\sqrt{1-z^2}} \end{aligned} \tag{19}$$

and

$$\begin{aligned} H(\alpha) &= \mathbf{R}'\cdot\hat{\mathbf{u}}_\phi \\ &= (x'\hat{\imath} + y'\hat{\jmath} + z'\hat{\mathbf{k}})\cdot\left[\frac{xz}{\sqrt{1-z^2}}\hat{\imath} + \frac{yz}{\sqrt{1-z^2}}\hat{\jmath} - \sqrt{1-z^2}\,\hat{\mathbf{k}}\right] \\ &= \frac{x'xz}{\sqrt{1-z^2}} + \frac{y'yz}{\sqrt{1-z^2}} - z'\sqrt{1-z^2} \\ &= \frac{(x'x + y'y + z'z)z - z'}{\sqrt{1-z^2}} \\ &= \frac{-z'}{\sqrt{1-z^2}} \end{aligned} \tag{20}$$

where the last step follows because $(x'x + y'y + z'z)$ is the derivative of the constant $(x^2 + y^2 + z^2)/2$.

Using equations (14) for the components of $\mathbf{F}$ and converting from θ, ϕ to x, y, z gives

$$\mathbf{F} = \left[\frac{-z}{\sqrt{1-z^2}} + \frac{g(\theta)}{\sqrt{1-z^2}}\right]\hat{\mathbf{u}}_\theta + h(\phi)\hat{\mathbf{u}}_\phi. \tag{21}$$

Using the components of $d\mathbf{R}/d\alpha$ from equations (19) and (20), we have

$$\mathbf{F}\cdot\frac{d\mathbf{R}}{d\alpha} = \left[\frac{xy' - yx'}{\sqrt{1-z^2}}\right]\left[\frac{-z}{\sqrt{1-z^2}} + \frac{g(\theta)}{\sqrt{1-z^2}}\right] - \frac{z'h(\phi)}{\sqrt{1-z^2}} \tag{22}$$

This is the integrand for the segment of the boundary integral from $\mathbf{P}_1$ to $\mathbf{P}_2$. Integration is with respect to α, from $\alpha = 0$ to $\alpha = \Delta$. The variables x, y, z and their derivatives (with respect to α) x', y', z' are all functions of α, as given by the components of equation (17).

We can choose the functions g and h to simplify equation (22). Nothing is gained by retaining the last term, so we take $h \equiv 0$. This simplifies the integrand to

$$\mathbf{F} \cdot \frac{d\mathbf{R}}{d\alpha} = \frac{(xy' - yx')(g(\theta) - z)}{1 - z^2} \tag{23}$$

Notice the potential singularities at $z = \pm 1$, i.e. the North Pole and the South Pole. Writing the denominator as $1 - z^2 = (1 - z)(1 + z)$, we see that if $g \equiv 1$ we remove the singularity at $z = 1$, and if $g \equiv -1$ we remove the singularity at $z = -1$. We must not put $g = 0$, since then $\mathbf{F}$ would vanish everywhere on the equator, violating equation (2) there. This would lead to a value of zero for the areas of the northern and southern hemispheres. In the following development we take $g \equiv 1$. In case one of the $\mathbf{P}_k$ is the South Pole, g should be replaced by -1.

We can now write the first term in the area summation of equation (4) as

$$J_1 = \int_0^{\Delta} \frac{xy' - yx'}{1 + z}\, d\alpha. \tag{24}$$

Notice the similarity to the expression for the plane, equation (6). We have explicit expressions for x, y, z, x', y' from the components of equation (17) and its derivatives, namely

$$x = \frac{\sin(\Delta - \alpha)}{\sin \Delta} x_1 + \frac{\sin(\alpha)}{\sin \Delta} x_2 \tag{25a}$$

$$x' = \frac{-\cos(\Delta - \alpha)}{\sin \Delta} x_1 + \frac{\cos(\alpha)}{\sin \Delta} x_2 \tag{25b}$$

and similar equations for y, y' and z, z'. Substituting these expressions into equation (24) and using standard trigonometric identities leads to

$$J_1 = (x_1 y_2 - y_1 x_2) \int_0^{\Delta} \frac{d\alpha}{\sin \Delta + z_1 \sin(\Delta - \alpha) + z_2 \sin \alpha}. \tag{26}$$

Recalling that this is the contribution to the area summation from the segment $k = 1$ between $\mathbf{P}_1$ and $\mathbf{P}_2$, we can write the total area as

$$A = \sum_{k=1}^{N} (x_k y_{k+1} - y_k x_{k+1}) J_k \tag{27}$$

where the terms J_k are the integrals

$$J_k = \int_0^{\Delta_k} \frac{d\alpha}{\sin(\Delta_k) + z_k \sin(\Delta_k - \alpha) + z_{k+1} \sin \alpha}. \tag{28}$$

and Δ_k comes from $\cos(\Delta_k) = \mathbf{P}_k \cdot \mathbf{P}_{k+1}$. The integral can be put into a standard form and explicitly integrated with the substitution $w = e^{i\alpha}$. Under this transformation, $d\alpha = dw/(iw)$, $\sin\alpha = (w - w^{-1})/2i$, and the integral becomes

$$J_k = \int_1^{e^{i\Delta}} \frac{2\, dw}{aw^2 + 2bw + c} \tag{29}$$

where

$$a = z_{k+1} - z_k e^{-i\Delta} \quad b = i \sin\Delta \quad c = z_k e^{i\Delta} - z_{k+1} \tag{30}$$

The subscript k on Δ has been dropped to reduce notational clutter.

The value of J_k depends on the sign of the discriminant $D = b^2 - ac$, or

$$D = z_k^2 + z_{k+1}^2 - 2z_k z_{k+1} \cos\Delta - \sin^2\Delta. \tag{31}$$

The three cases are [Marsden, Appendix A]

$$J_k = \begin{cases} \frac{1}{\sqrt{D}} \ln\left[\frac{aw+b-\sqrt{D}}{aw+b+\sqrt{D}}\right] & (D > 0) \\ \frac{2}{\sqrt{-D}} \arctan\left[\frac{aw+b}{\sqrt{-D}}\right] & (D < 0) \\ \frac{-2}{aw+b} & (D = 0) \end{cases} \tag{32}$$

where the expressions must be evaluated between the upper and lower limits of $w = e^{i\Delta}$ and $w = 1$. The imaginary parts of the resulting complex expressions are zero, as they must be since the original integrand and limits are real. Algebraic simplification leads us to define

$$Q = z_k + z_{k+1} + 1 + \cos\Delta \tag{33}$$

in terms of which the expressions for J_k become

$$J_k = \begin{cases} \frac{1}{\sqrt{D}} \ln\left[\frac{Q+\sqrt{D}}{Q-\sqrt{D}}\right] & (D > 0) \\ \frac{2}{\sqrt{-D}} \arctan\left[\frac{\sqrt{-D}}{Q}\right] & (D < 0) \\ \frac{Q}{(1+z_k)(1+z_{k+1})(1+\cos\Delta)} & (D = 0) \end{cases} \tag{34}$$

This completes the determination of the terms in the area formula (27). We will now summarize the steps and put them in an algorithmic format.

Problem:

Given a sequence of (longitude,latitude) coordinates on the unit sphere, (θ_k, ϕ_k), $k = 1, 2, \ldots, N$, find the area of the region that is enclosed when the points are connected in sequence by arcs of great circles.

Solution:

(1) Set the running sum to 0 and set k to 1.

(2) Compute $\cos\Delta = \mathbf{P}_k \cdot \mathbf{P}_{k+1}$ either from $x_k x_{k+1} + y_k y_{k+1} + z_k z_{k+1}$ or from $\cos\phi_k \cos\phi_{k+1} \cos(\theta_{k+1} - \theta_k) + \sin\phi_k \sin\phi_{k+1}$. Notice that we won't ever need Δ by itself, just its cosine.

(3) Compute Q from (33): $Q = z_k + z_{k+1} + 1 + \cos\Delta$ or $Q = \sin\phi_k + \sin\phi_{k+1} + 1 + \cos\Delta$.

(4) Compute the discriminant D from (31): $D = z_k^2 + z_{k+1}^2 - 2z_k z_{k+1} \cos\Delta - \sin^2\Delta$ or $D = (\sin\phi_k + \sin\phi_{k+1})^2 - (1+\cos\Delta)(1 - \cos\Delta + 2\sin\phi_k \sin\phi_{k+1})$.

(5) Compute the integral contribution J_k in the area formula (27), using the appropriate form of equation (34).

(6) Compute the first factor in the area formula (27), $x_k y_{k+1} - y_k x_{k+1}$ or $\cos\phi_k \cos\phi_{k+1} \sin(\theta_{k+1} - \theta_k)$.

(7) Multiply together the results of steps 5 and 6 to get the k-th term in the summation of (27), and add this to the running sum.

(8) If k is less than N then increment k and go to step 2.

A computer program that implements the above algorithm is given in the appendix.

5. Numerical Example and Remarks

It is of interest in Arctic oceanography to calculate the areas of the watersheds that drain into the Arctic Ocean. The boundary of the Asian watershed that drains into the Arctic Ocean was digitized from a Mercator map of the world by tracing its circumference with an encoding device. This produced a computer file with 672 (x, y) coordinate pairs, in which the x axis coincided with the equator, the y axis coincided with the Greenwich Meridian, and the unit of length was chosen to be one degree of longitude on the equator. These (x, y) map coordinates are related to longitude θ and latitude ϕ by [Snyder]

$$x = \frac{180}{\pi}\theta \qquad y = \frac{180}{\pi}\ln\left[\arctan\left(\frac{\phi}{2} + \frac{\pi}{4}\right)\right] \tag{35}$$

where θ and ϕ are in radians. Inverting these relations and substituting the (x, y) map coordinates gives a sequence (θ_k, ϕ_k), $k = 1$ to 672, of points on the sphere that defines the boundary of the watershed.

At first a simple integration program was written in which the region lying between the minimum and maximum latitudes and longitudes of the watershed was divided into differential elements of size $\Delta\phi$ by $\Delta\theta$. The area of the watershed was calculated as $\sum \cos\phi \, \Delta\phi \, \Delta\theta$ where the summation was taken over all elements inside the watershed boundary. With each degree of latitude and longitude divided into 32 parts, this amounted to 5,918,720 elements, of which 2,516,738 were found to lie within the watershed. The program required more than 51 hours of elapsed time on a Sun workstation to arrive at the area, 1.424×10^7 km^2.

This dismal performance led to the derivation of the formulas in this work. Using the same 672 coordinates for input, the program in the appendix arrived at the same answer in about two seconds. The 5.9 million complicated comparisons in the first program were replaced by 672 iterations of simple calculations.

Of course in any real physical problem such as the one described here, there are sources of error such as uncertainty in the exact location of the boundary, inadequate representation of the boundary by too few points, and the non-sphericity of the earth. These problems can be dealt with by acquiring better maps, digitizing the boundary with more points, and modifying the formulas here to take into account the flattening of the earth at the poles, which introduces a correction on the order of three parts per thousand.

Acknowledgment

This work was supported by NASA Grant NAGW 2513. Thanks also to Erika Dade for bringing this problem to my attention and doing the original watershed calculations.

Appendix – Fortran Program

```
      program area
      implicit undefined (a-z)
c
c______________________________________________________________________
c
c     Read a sequence of (longitude,latitude) coordinates.
c     Compute the area on the unit sphere that is enclosed by connecting
c     these points in sequence with arcs of great circles.
c
c     Refer to ''Computing Areas of Regions with Discretely Defined
c     Boundaries".
c
c______________________________________________________________________
c
c     Constants.
c
      real pi, piOver180
      parameter (pi = 3.14159265358979, piOver180 = pi / 180.0)
c
c     Parameters.
c
      integer maxPoints
      parameter (maxPoints = 1000)
c
c     Mean radius of earth in kilometers.
c
      real Rearth
      parameter (Rearth = 6371.2)
c
c     Variables.
c
      integer n, k
      real sum, first, integral, cosDelta, D, Q, R
      real cosPhiK, cosPhiK1, sinPhiK, sinPhiK1
      real phi(maxPoints), theta(maxPoints)
      character*14 filename
c
c     Read number of lon/lat coordinate pairs, and
c     the name of the file containing those coordinates.
c
      read(5,*) n, filename
c
c     Read the coordinates.  Longitude is first.  Both in degrees.
c
```

```
      open(1, file=filename)
      read(1,*) (theta(k),phi(k), k=1,n)
      close(1)
c
c     Convert to radians.
c
      do 10 k=1,n
        phi(k) = phi(k) * piOver180
        theta(k) = theta(k) * piOver180
   10 continue
c
c     Make the sequence of coordinates cyclic.
c
      phi(n+1) = phi(1)
      theta(n+1) = theta(1)
c
c     Initialize for the summation.
c
      sum = 0.0
      cosPhiK1 = cos(phi(1))
      sinPhiK1 = sin(phi(1))
c
      do 20 k=1,n
c
c       Previous "k+1" values become new "k" values.
c
        cosPhiK = cosPhiK1
        sinPhiK = sinPhiK1
c
c       Get new "k+1" values.
c
        cosPhiK1 = cos(phi(k+1))
        sinPhiK1 = sin(phi(k+1))
c
c       Compute first factor in k-th term of summation.
c
        first = cosPhiK * cosPhiK1 * sin(theta(k+1)-theta(k))
c
c       Compute integral in k-th term of summation.
c       First get cosine of delta, then discriminant, then Q.
c
        cosDelta = cosPhiK * cosPhiK1 * cos(theta(k+1)-theta(k))
     .        + sinPhiK * sinPhiK1
        D = (sinPhiK + sinPhiK1)**2
     .        - (1.0+cosDelta)*(1.0-cosDelta  +2.0*sinPhiK*sinPhiK1)
        Q = sinPhiK + sinPhiK1 + 1.0 + cosDelta
```

```
c
      if (D .gt.  0.0) then
        R = sqrt (D)
        integral = alog ( (Q+R)/(Q-R) ) / R
      else if (D .lt.  0.0) then
        R = sqrt (-D)
        integral = 2.0 * atan ( R/Q ) / R
      else
        integral = Q / ((1.0+sinPhiK) *(1.0+sinPhiK1)*(1.0+cosDelta))
      endif
c
c     Accumulate sum and go on to next segment.
c
      sum = sum + first * integral
c
   20 continue
c
c   Write results and stop.
c
      write(6,90) sum, sum/(4.0*pi), sum*Rearth*Rearth
c
      stop
   90 format(1x, 'area (on unit sphere) = ', e14.6,
     .     /1x, 'area / (4*pi) = ', e14.6,
     .     /1x, 'area (km**2 on earth) = ', e14.6)
      end
```

References

(1) Francis B. Hildebrand, "Advanced Calculus for Applications", Prentice-Hall, 1976.

(2) J. E. Marsden and A. J. Tromba, "Vector Calculus", W. H. Freeman & Co., 1976.

(3) John P. Snyder, "Map Projections - A Working Manual", U. S. Geological Survey Professional Paper 1395, U. S. Government Printing Office, 1987.

(4) Robert Weinstock, "Calculus of Variations", Dover Publications, 1974.

The Quadratic World of Kinematic Waves

Sandra L. Arlinghaus, John D. Nystuen, Michael J. Woldenberg *

Kinematic waves differ from "ordinary" waves insofar as it is the kinematics–the dynamic aspects of motion other than mass and force–that are the focus. Thus, Langbein and Leopold [1968, p. 1] define a kinematic wave as "a grouping of moving objects in zones along a flow path and through which the objects pass. These concentrations may be characterized by a simple relation between the speed of the moving objects and their spacing as a result of interaction between them." Flow in a channel is characteristically expressed as a function of concentration, be that as cars per hour as a function of cars per mile or as transport in cubic feet per minute of sand in a one inch tube as a function of linear concentration of sand in pounds per square foot [Langbein and Leopold 1968; Haight 1963; Lighthill and Whitham I and II 1955]. Examples of kinematic waves are abundant in physical and urban settings alike–in realms as disparate as sand transport in a flume or car movement on an Interstate Highway [Langbein and Leopold, 1968]. When empirical data are graphed, they often trace out a parabola (or a curve close to a parabola); thus, the relationship between concentration and flow is often a quadratic one [Langbein and Leopold, 1968].

The classical analysis of the parabolic graphs of these waves rests on considering what happens to flow as a result of minor perturbations in local concentrations–techniques are based in notions from the calculus [Langbein and Leopold 1968]. Consider a concave down parabola with its maximum in the first quadrant that passes through the origin. Flow is a function of concentration; thus, concentration appears on the x-axis and flow on the y-axis. Choose two points on the curve, one with coordinates (x_1, y) and the other with coordinates (x_2, y)–the x-coordinates are different and lead to the same y-coordinate. They are placed symmetrically on the x-axis about a vertical line through the curve's maximum (Figure 1; for electronic readers only, please draw this curve and subsequent ones as per text). Assuming that x_1 is to the left of the maximum, the traditional analysis notes that at x_1, a slight increase in concentration results in a slight increase in flow; a slight decrease in concentration at x_1 results in a slight decrease in flow. The channel is relatively sparsely congested–slight changes in concentration result in directly parallel changes in flow. Further, the closer one is to the x-coordinate of the maximum, the less difference these slight changes cause. On the other hand, at x_2 (to the right of the maximum) a slight increase in concentration results in a decrease in flow, suggesting a channel which cannot easily assimilate any extra traffic. Further, a slight decrease in concentration at x_2 results in an increase in flow, again reflecting a relatively congested condition of this channel. When the horizontal line suggested by x_1 and x_2 is tangent to the parabola, at its maximum, the kinematic wave is stationary relative to the channel; thus, as the distance of horizontal lines increases away from this tangent line, there is a corresponding increase in the amount of change caused by local perturbations. The origin, as a location for x_1, represents a completely uncrowded condition, while the second intersection of the curve with the x-axis represents the most crowded position within this interval [Langbein and Leopold 1968].

The traditional analysis, based merely on considering what slight changes in x_1 and x_2 might suggest, fits well with real-world travel experience. Consider the concentration on the x-axis to be density of automobiles as vehicles per mile; on the y-axis, consider flow to be vehicles per hour. Practical evidence does suggest that an improvement in the maximum capacity of the road does result in improved transmission of flow, but only up to a point. Thus, highway systems are widened around cities and endowed with limited access to increase the number of vehicles per hour that can move from origin to destination. Beyond about 1800 vehicles per hour, this "improvement" is no longer useful [Nystuen 1992]; indeed, congestion increases and flow per hour decreases toward the point of gridlock—the ultimate disaster that can affect millions of individuals. This sort of ceaseless "improvement," to the point of disaster, of what worked well in a less congested arena, appears in a variety of contexts; when an optical cable with the capacity to serve millions is cut, disaster comes to many rather than to few, and chaos in communication becomes a real possibility [Austin 1991].

The traditional analysis also allows for computation of various other features associated with the kinematics of the phenomenon it describes. For example, the average speed of particles in the channel, or wave celerity, can be measured at any point on the curve, simply by finding the slope of the chord joining that point to the origin [Langbein and Leopold 1968]. However, when a given density leads to a certain flow, which is then used to determine the next input to create a new density level, feedback occurs. Feedback is not measured in the traditional analysis. It also fits with travel experience and indeed is the sort of process that can get chaotic. Thus, it seems plausible to consider graphical analysis of kinematic curves, based in Feigenbaum's Graphical Analysis from the mathematics of Chaos Theory, as a supplement to the traditional analysis.

Consider the following set of parabolas as Figures 2 through 7: $y = 1.5x(1-x)$; $y = 2x(1-x)$; $y = 3x(1-x)$; $y = 3.75x(1-x)$; $y = 4x(1-x)$; and, $y = 5x(1-x)$. The e-reader should draw each of these curves, noting that each parabola is of the sort described above—consider the units on the axes, ranging from 0 to less than 1.5, as percentages. Thus, 0.5 on the x-axis represents a concentration of 50%. Also include in each graph the line $y = x$. Each parabola intersects this 45-degree line in two points–one at the origin and one that is either to the left or to the right of the curve's maximum. As the coefficient of the curve increases from 1.5 to 5, the curves become successively less flat, have a higher maximum, and have a second intersection with the line $y = x$ farther to the right.

To represent geometric feedback visually on Figures 2 to 7, proceed as follows [based on material from Feigenbaum 1980; Gleick 1987; Devaney and Keen 1989]. Locate the point 0.1 on the x-axis of each figure. Draw a vertical line from that point (as a "seed" value for the graphical analysis) to the parabola. Now draw a horizontal line from the curve to the line $y = x$; next read vertically from this location to the parabola. The effect here is to use output as input; for, 0.1 was the initial input. When that value was mapped to the parabola, an output resulted— when that output was mapped horizontally to $y = x$, it was then used as input when it was next sent to the curve. Successive iteration of this process should result in the following paths from the iteration ("orbits"): Figure 2—a staircase with shallow rises; Figure 3— a staircase with sharper rises than in Figure 2; Figure 4— a tightly bounded cyclical orbit closing in on the second intersection of the line with the parabola; Figure 5—an unpredictable, bounded orbit; Figure 6—a chaotic, bounded orbit; Figure 7—an orbit that escapes to negative infinity (from a curve whose maximum is beyond the 100% concentration level). Geometrically, control over the dynamics of the orbit becomes less stable as one proceeds from Figures 2 to 7. It makes little difference which initial seed is chosen; the dynamics of the orbit are invariant with respect to these curves (parabolas). Unlike the traditional analysis, in which there is considerable variation in the measures used, with respect to a single curve, the pattern of the orbit is constant throughout each figure—as a sort of a shape-invariant. Indeed, any of these curves might be employed equally for the traditional, but not for the graphical, analysis.

What determines the extent of stability in the geometric dynamics noted in these figures are the height of the parabola and the position of the second intersection of $y = x$ with that parabola. Higher parabolas have intersection point with $y = x$ farther to the right of the curve's maximum, producing more uncontrolled feedback. This fits well with traffic observations; increase of a road's maximum capacity beyond some critical level leads to disastrous congestion. The tool of graphical analysis looks promising as a tool in analyzing real-world phenomena [Feigenbaum 1980; Gleick 1987] that follow kinematic waves as well as those that follow more complicated curves [Arlinghaus, Nystuen, and Woldenberg 1992].

* Author Woldenberg wishes to acknowledge input from M. Sonis regarding the analysis of kinematic waves—1981.

References

Arlinghaus, S. L., Nystuen, J. D., and Woldenberg, M. J. 1992 (forthcoming in July, 1992). An application of graphical analysis to semidesert soils. *Geographical Review*.

Austin, Robert F. Personal communication, 1991.

Devaney, R. L. and Keen, L. 1989. *Chaos and fractals: The Mathematics behind the Computer Graphics*. Proceedings of symposia in applied mathematics, vol. 39, American Mathematical Society, Providence, RI.

Feigenbaum, M. J. 1980. Universal behavior in non-linear systems. *Los Alamos Science*, summer: 4-27.

Gleick, J. 1987. *Chaos: Making a New Science*. New York: Penguin Books.

Haight, F. A., 1963. *Mathematical theories of traffic flow*. New York: Academic Press.

Langbein, W. B. and Leopold, L. B. 1968. *River Channel Bars and Dunes—Theory of Kinematic Waves*, USGS, Professional Paper 4222, pp. 1-20. United States Government Printing Office, Washington.

Lighthill, J. J., and Whitham, G. B., 1955. On kinematic waves. I. Flood movement in long rivers. *Royal Soc. [London] Proc.*, v. 229A, p. 281-316.

Lighthill, J. J., and Whitham, G. B., 1955. On kinematic waves II. A theory of traffic flow on long crowded roads. *Royal Soc. [London] Proc.*, v. 229A, p. 317-345.

Nystuen, J. D. Seminar on "Intelligent Vehicle Highway Systems." University of Michigan.

REVIEW of RangeMapper^{TM} (version 1.4b).

A utility for biological species range mapping, and similar mapping tasks in other fields. Price: $350

Program and manual written by **Kenelm W. Philip**. Tundra Vole Software 1590 North Becker Ridge Road Fairbanks, Alaska 99709 (907) 479-2689

Reviewed by **Yung-Jaan Lee**, Ph.D. Candidate in Urban, Technological, and Environmental Planning, The University of Michigan, Ann Arbor, MI 48109.

From the author's flyer:

"RangeMapper is a Macintosh mapping and data plotting utility. It allows rapid and accurate display of lat/long data on the user's choice of maps."

"RangeMapper Features"

"Range Mapper can bring up low-resolution maps of the world, or portions thereof, in north polar azimuthal, simple cylindrical, Mercator, orthographic, stereographic, or Lambert azimuthal equal-area projections.

Data may be plotted to maps from ASCII files of latitude, longitude, and site name in several different formats, in several sizes of open/filled circles and squares. Program-readable data files can be dumped directly from a database or spreadsheet. Lat/long coords may be read directly from the maps, and plotted points may be 'verified' by clicking on them. The Alaska map is based on the CIA World Data Bank file, and is usable down to 20-30 mile regions.

The world map is derived from the Micro World Data Bank II file. It is usable down to regions of the order of 500 miles or so in extent, which is adequate for species range mapping on small-scale maps.

Designed originally for biological species range mapping, the program has many other uses wherever data files need to be accurately plotted to maps. In conjunction with the word processor 'Nisus', RangeMapper may also be used as a visual interface to a text database, so you can open a text file on a site by clicking on that site on the displayed map.

The 'verify' feature permits rapid checking of your ASCII data files for errors. In conjunction with a DA text editor, your data files may be edited interactively from within the program—making error correction a rapid and easy job.

Points may also be placed on the maps by hand, either by eye or by reading coordinates off the map and dropping a dot at the correct coordinates.

RangeMapper can save maps to disk, print them directly to an IMageWriter or LaserWriter, or export them as PICT files to be imported into a drawing program (as MacDraw or Canvas) for enhancement and annotation.

Maps produced by RangeMapper may have a user-designed latitude/longitude grid overlaid, and a title and caption may be added. Data plotted to RangeMapper may be overlaid in up to 14 separate layers, each of which may be toggled on and off independently. Data may be plotted as dots or as connected lines."

The processing speed of this software is, to some extent, slow, especially for a small-scale map or a map with filled area. This may be due to the fact that this software involves a vast number of pixels.

Users accustomed to working with Geographic Information Systems should be aware that this software is, as it says, a mapping utility only. The spiral-bound documentation is adequate and contains samples of maps apparently made using RangeMapper; a couple of improvements seem in order.

1. On page 2, the author describes RangeMapper as needing at least 1500KB of free memory, and that the "MultiFinder partition" should be set to that value in the Get Info dialog box. This is confusing, as the user will probably select the MultiFinder icon and try to change the partition in Get Info. In fact, the user should highlight the **RangeMapper** icon, rather than the MultiFinder icon, and then go to Get Info dialog box to change the partition.
2. On page 6, the user is instructed to select the file '*MWDB3.All' under the File menu. However, there is no such file in this software. Instead, the user should select the file 'MWDB2.All' and then check the **show state/provs** under the Mapping menu in order to display the circumpolar map demonstration.
3. The printing requirements should appear early in the first part of the manual.
4. An Index at the end of the manual would be helpful.

Some other suggestions for improvement of the software are:

1. It would help to employ more of the standard Macintosh environment conventions, such as:

   a **Close** selection under the File menu;

   a **Window** sub-menu in the pull-down menu;

   the filename displayed at the top of the screen (different from the title of the map);

   a close box, zoom box, size box, and scroll bars displayed on the screen, as in a standard Macintosh window.
2. The "Menus" section could be moved to the beginning of the manual, rather than in the middle. If not, the author should describe the difference between **Map** and **Open** function in the **File** menu at the beginning.
3. After displaying a map, a selection box will automatically show up on the screen. The author should explain why this box comes up. It only later becomes apparent that it is used to link a map to adjacent regions, if available.

It may be more efficient to run this software using a Macintosh II or higher, or better, with a math co-processer because of very slow printing times. If not, users must carefully follow the recommended printing procedure to reduce the size of the output file, such as turning off "Graphics Smoothing" and checking "Precision Bitmap Alignment" (in the "Moving RangeMapper Output to Word Processors" section and the "Printing: RangeMapper" section).

In addition to the two drawing programs (Canvas 3.0 and SuperPaint 2.0), MacDraw II 1.1 and MacPaint 2.0 are capable of image size reduction. After exporting a map to

MacDraw or MacPaint, one can still copy the map to any word processor.

Those needing only a mapping program will find this sortware useful, especially if working on high latitude areas.

Note: Canvas is a trademark of Deneba Systems;

SuperPaint is a trademark of Aldus Corporation;

NISUS is a trademark of Paragon Concepts, Inc.;

Apple and LaserWriter are registered trademarks of Apple Computer, Inc.;

Macintosh is a trademark licensed to Apple Computer, Inc.; MacDraw is a trademark of Apple Computer, Inc.

FEATURES

Press Clippings

FROM SCIENCE, AAAS

Science, November 29, 1991, Vol. 254, No. 5036, copyright, the American Association for the Advancement of Science. Many thanks to Joseph Palca at *Science* for his continuing interest in online journals. The citation appeared in "Briefings" and is entitled "**Online Journals**," by Joseph Palca.

NOTE: Readers wishing to contact Richard Zander, Editor of *Flora Online*, can do so at bitnet address:

VISBMS@UBVMS

FROM SCIENCE NEWS

Math for all seasons

by Ivars Peterson

January 25, 1992, Vol. 141, No. 4. Page 61. Reprinted with permission of *Science News*.

When the American Association for the Advancement of Science announced with considerable fanfare last year the 1992 debut of *The Online Journal of Current Clinical Trials*, it was billed as the world's first peer-reviewed science journal available to subscribers electronically. What the organizers of this effort didn't know was that several such electronic journals already existed. One of these concerns the application of mathematics to geography.

Solstice: An Electronic Journal of Geography and Mathematics — published by Sandra Lach Arlinghaus of the Institute of Mathematical Geography, a small, independent research organization in Ann Arbor, Mich. — first appeared in 1990. Its two issues per year, published appropriately on the dates of the summer and winter solstices, go to about 50 individuals, wwho receive the journal free. Transmission costs for distributing the journal electronically over a computer network to all subscribers amount to less than $5 per issue, with the cost of printing passed on to the user. Libraries and other institutions that prefer printed copies pay for each issue, and those copies are generated from computer files only when needed.

"It's all very cheap, all environmentally sound," Arlinghaus says.

But getting the journal going wasn't easy, she remarks. The biggest production problem involved photographs and figures, which can't be transmitted electronically in the same, compact way as letters, numbers or even mathematical notation. At present, individuals wishing to see particular illustrations must obtain photocopies directly from the Institute of Mathematical Geography. Arlinghaus also admits that she has had trouble obtaining manuscripts for publication in this still-unconventional medium. But individuals who might initially have been skeptics "become more receptive when they see the actual product," she says.

LETTER AND RESPONSE IN SCIENCE NEWS

One from AAAS in reply to Peterson; one from IMaGe in reply to AAAS, during period from January through May, 1992.

AAG NEWSLETTER

Volume 27, Number 6, June 1992.
"Online Geographical Journals," page 10.

SOLSTICE

Institute of Mathematical Geography

IMAGE

IMAGINATION

Journal of the
Institute of Mathematical Geography

Winter, 1992

SOLSTICE:

AN ELECTRONIC JOURNAL OF GEOGRAPHY AND MATHEMATICS

WINTER, 1992

Volume III, Number 2
Institute of Mathematical Geography
Ann Arbor, Michigan

The purpose of *Solstice* is to promote interaction between geography and mathematics. Articles in which elements of one discipline are used to shed light on the other are particularly sought. Also welcome, are original contributions that are purely geographical or purely mathematical. These may be prefaced (by editor or author) with commentary suggesting directions that might lead toward the desired interaction. Individuals wishing to submit articles, either short or full- length, as well as contributions for regular features, should send them, in triplicate, directly to the Editor-in-Chief. Contributed articles will be refereed by geographers and/or mathematicians. Invited articles will be screened by suitable members of the editorial board. IMaGe is open to having authors suggest, and furnish material for, new regular features.

The opinions expressed are those of the authors, alone, and the authors alone are responsible for the accuracy of the facts in the articles.

Send all correspondence to: Institute of Mathematical Geography, 2790 Briarcliff, Ann Arbor, MI 48105-1429, (313) 761-1231, IMaGe@UMICHUM, Solstice@UMICHUM

Suggested form for citation. If standard referencing to the hardcopy in the IMaGe Monograph Series is not used (although we suggest that reference to that hardcopy be included along with reference to the e-mailed copy from which the hard copy is produced), then we suggest the following format for citation of the electronic copy. Article, author, publisher (IMaGe) - all the usual-plus a notation as to the time marked electronically, by the process of transmission, at the top of the recipients copy. Note when it was sent from Ann Arbor (date and time to the second) and when you received it (date and time to the second) and the field characters covered by the article (for example FC=21345 to FC=37462).

This document is produced using the typesetting program, TEX, of Donald Knuth and the American Mathematical Society. Notation in the electronic file is in accordance with that of Knuth's *The TEXbook*. The program is downloaded for hard copy for on The University of Michigan's Xerox 9700 laser– printing Xerox machine, using IMaGe's commercial account with that University.

Unless otherwise noted, all regular "features" are written by the Editor-in-Chief.

Upon final acceptance, authors will work with IMaGe to get manuscripts into a format well-suited to the requirements of *Solstice*. Typically, this would mean that authors would submit a clean ASCII file of the manuscript, as well as hard copy, figures, and so forth (in camera-ready form). Depending on the nature of the document and on the changing technology used to produce *Solstice*, there may be other requirements as well. Currently, the text is typeset using TEX; in that way, mathematical formulæ can be transmitted as ASCII files and downloaded faithfully and printed out. The reader inexperienced in the use of TEX should note that this is not a "what-you-see-is-what-you-get" display; however, we hope that such readers find TEX easier to learn after exposure to *Solstice*'s e-files written using TEX!

Copyright will be taken out in the name of the Institute of Mathematical Geography, and authors are required to transfer copyright to IMaGe as a condition of publication. There are no page charges; authors will be given permission to make reprints from the electronic file, or to have IMaGe make a single master reprint for a nominal fee dependent on manuscript length. Hard copy of *Solstice* is available at a cost of $15.95 per year (plus shipping and handling; hard copy is issued once yearly, in the Monograph series of the Institute of Mathematical Geography. Order directly from IMaGe. It is the desire of IMaGe to offer electronic copies to interested parties for free. Whether or not it will be feasible to continue distributing complimentary electronic files remains to be seen. Presently *Solstice* is funded by IMaGe and by a generous donation of computer time from a member of the Editorial Board. Thank you for participating in this project focusing on environmentally-sensitive publishing.

ISBN: 1-877751-54-5 ISSN: 1059-5325

SUMMARY OF CONTENT

1. A WORD OF WELCOME–FROM A TO U!

Welcome to new subscribers from Alice Springs, Australia, to Ulm, Germany and to all those in between (alphabetically or otherwise)! We hope you enjoy participating in this means of journal distribution. Instructions for downloading the typesetting have been repeated in this issue, at the end. They are specific to the TEX installation at The University of Michigan, but apparently they have been helpful in suggesting to others the sorts of commands that might be used on their own particular mainframe installation of TEX. New subscribers might wish to note that the electronic files are typeset files—the mathematical notation will print out as typeset notation. For example,

$$\Sigma_{i=1}^{n}$$

when properly downloaded, will print out a typeset summation as i goes from one to n symbol, as a centered display on the page. Complex notation is no barrier to this form of journal production.

Many thanks to the members of the Editorial Board of *Solstice*; with the publication of this issue we welcome the addition to that Board of William D. Drake, Ph.D. Engineering (Operations Research), and Professor in various departments of The University of Michigan. Bill has a brief note later in this issue of *Solstice* in which he introduces himself, some of his recent interests, and some of his students' interests.

2. PRESS CLIPPINGS—SUMMARY

Brief write-ups about *Solstice* have appeared in the following publications:

1. **Science**, "Online Journals" Briefings. [by Joseph Palca] 29 November 1991. Vol. 254.

2. **Science News**, "Math for all seasons" by Ivars Peterson, January 25, 1992, Vol. 141, No. 4.

3. **American Mathematical Monthly**, "Telegraphic Reviews" — mentioned as "one of the World's first electronic journals using TEX," September, 1992.

3A. WHAT ARE MATHEMATICAL MODELS AND WHAT SHOULD THEY BE?

Frank Harary
Distinguished Professor of Computer Science
New Mexico State University
Professor Emeritus of Mathematics
University of Michigan

Reprinted with permission from
Biometrie-Praximetrie
Vol. XII, 1-4, 1971, pp. 3-18
Published by Société Adolphe Quetelet
Belgian Region of Biometric Society
Av. de la Faculté, 22
B-5030 Gembloux, Belgium

At the time this paper was written,
Frank Harary
was also a member of the
Research Center for Group Dynamics
Institute for Social Research
The University of Michigan

No matter what the area of scientific research, whether social or physical, mathematical thinking is involved, explicitly or implicitly. At the least, the precise formulation of a problem entails some aspect of set theory and logic. Generally speaking, the working scientist uses the term 'mathematical model' for whatever branch of mathematics he may be applying to his present problem. On the other hand, the purist mathematician-logician insists strictly on the use of 'model' to mean a certain interpretation of an abstract axiom system in the real world.

We begin with a self-contained development of the concepts needed for the discussion of research processes. This leads to the distinction between the real and abstract world, and the interaction between them by interpretation and abstraction. A similar, but conceptually different bifurcation is proposed for the two levels of research: digging into the foundations versus extending the horizons of knowledge. These considerations are assembled into a comprehensive Research Schema which enables a concise analysis of scientific discovery. Classical illustrations are provided, including true stories about Newton, Darwin, Freud, and Einstein. We conclude with some subjective evaluations of acceptability of mathematical models.

1. What Are They?

We have just noted that the word 'model' has different meanings for the mathematician and the scientist. When a mathematician uses the word, he is referring to the physical or social realization of his theory. On the other hand, when a scientist speaks of a mathematical model, he means the area of mathematics which applies to his work. Thus one (following Abraham Kaplan, oral communication) could say as a mnemonic aid that a model is always the other fellow's system. Contrariwise it also appears to be customary by usage to refer to "research" as what goes on in your own domain.

In order to define a model rigorously, it is convenient to develop (as in Wilder [4] or in a more elementary presentation, Richardson [3]) several notions in the foundations of mathematics. Recall from high school geometry that Euclid's axioms are about as follows (depending on which book you read). The words "point" and "line" are undefined terms.

A_1 (Axiom 1) Every line is a collection of points.

A_2 There exist at least two points.

A_3 If u and v are points, then there exists one and only one line containing u and v.

A_4 If L is a line, then there exists a point not on L.

A_5 If L is a line, and v is a point not on L, then there exists one and only one line L' containing v which is parallel to L, i.e., $L \cap L' = \emptyset$.

Axiom 5 is the celebrated "Parallel Postulate" of Euclid.

An *axiom system* $\Sigma = (P, A)$ consists of two sets: a set P of primitives and a set A of axioms. *Primitives* are the deliberately undefined terms upon which all definitions in the system are based. *Axioms* are statements which are assumed to be true, and from which other statements called *theorems*, can be derived. Primitives and axioms serve to avoid so-called circular definitions and circular reasoning. Each axiom in the system is an assertion about the primitives.

Euclid's axiom system consists of two primitives, 'point' and 'line', and five axioms. When Euclid developed geometry, he made a distinction between axioms and postulates. Both were statements whose truth was assumed, but axioms were considered self-evident while postulates were not! This distinction eventually proved unnecessary and even undesirable, and today axiom and postulate are synonyms.

We shall denote by T or $T(\Sigma)$ the set of all theorems derivable from an axiom system Σ. Then a *mathematical system* (P, A, T) is an axiom system together with all theorems derivable from it.

An *independent axiom* A of Σ is one which cannot be derived from the remaining axioms. An *axiom system* is *independent* if every axiom is independent. In is called *consistent* if there are no two contradictory statements in $T(\Sigma)$.

One of the classical problems in 19th Century mathematics was to determine whether or not Euclid's Parallel Postulate, A_5, was independent. The consensus of opinion was that A_5 was dependent, that is, it could be derived from $A_1 - A_4$. Unsuccessful attempts to derive A_5 led to the discovery instead of non-euclidean geometry. The two types of non-euclidean geometry are now respectively called *hyperbolic geometry* (Bolyai-Lobachewski independently) in which there can be many parallels to a line through a point, and *elliptic geometry* (Riemann) in which there can be no such parallel.

An *interpretation* of an axiom system is an assignment of meanings to its primitives which makes the axioms become true statements. The results of an interpretation of Σ is called a *model* for Σ. This is the strict use of 'model' mentioned earlier.

An axiom system is called *satisfiable* if it has at least one model. Two models, M_1 and M_2 of Σ are *isomorphic* if there is a 1-1 correspondence between the elements of M_1 and those of M_2 which preserves every Σ-statement. In a *categorical* axiom system, any two models are isomorphic.

To illustrate, consider an axiom system with primitives $P = \{S, \circ\}$, where S is a set of integers, and $\circ$, is chosen as an undefined term for a binary operation denoted $a \circ b$, in order to avoid preconceived notions that a familiar symbol like $a+b$ would bring to mind. The following statements A_1 — A_4 are called *group axioms*, and any set S on which they hold under the operation $\circ$ is called a *group*.

A_1 (Closure Law) S is closed under $\circ$, that is, if a and b are in S, $a \circ b$ is in S.

A_2 (Associative Law) Operation $\circ$ is associative, that is, $a \circ (b \circ c) = (a \circ b) \circ c$ for all a, b, and c in S.

A_3 (Identity Law) There is a unique element i in S, called the identity element, such that $a \circ i = i \circ a = a$ for all a in S.

A_4 (Inverse Law) For every a in S, there is a unique element, written a^{-1} and called the *inverse* of a, such that $a \circ a^{-1} = a^{-1} \circ a = i$. Each of the four group axioms is independent, and so this is an independent axiom system. To verify that this axiom system is satisfiable, we now display a model.

One model for this system is the set $S_1 = \{1, -1\}$ under multiplication $\times$. Thus this is called a group of *order* 2, i.e., having just two elements. The identity element is 1, each element has itself as an inverse, and S is obviously closed and associative, as can be seen from the following multiplication table:

$$\begin{array}{r|rr} \times & 1 & -1 \\ \hline 1 & 1 & -1 \\ -1 & -1 & 1 \end{array}$$

Another model for this axiom system is the set $S_2 = \{0, 1\}$ under addition modulo 2. We define the sum of a and b mod 2 to be the remainder of $a+b$ after division by 2. Under this operation, we see at once from the next table that S_2 is closed and associative, 0 is the identity, and each element is again its own inverse. Thus S_2 is also a group of order 2.

$$\begin{array}{r|rr} +\ \text{mod } 2 & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array}$$

More generally, one can take S to be the set $\{0, 1, 2, \ldots, n-1\}$ and $a \circ b$ to mean $a + b \bmod n$. Then for each positive integer n, we get a distinct group of order n. Thus the above axiom system for groups is not categorical, since it has many non-isomorphic models.

These two groups, S_1 and S_2, are isomorphic since we can let operation $\times$ correspond with $+$ mod 2 and set $\{1, -1\}$ to correspond with $\{0, 1\}$. All statements derivable from the axioms still hold. That the two models are isomorphic is also shown in the fact that their tables both have the following form:

$$\begin{array}{c|cc} \circ & a & b \\ \hline a & a & b \\ b & b & a \end{array}$$

In fact, any pair of groups with two elements are isomorphic, so it is customary to speak of "the group of order two."

The study of group theory was originally motivated by properties which are possessed by the symmetries of a configuration, whether it be geometric, algebraic, architectural, physical, or chemical. It is readily verified that symmetries satisfy the four group axioms. For example, the inverse of a symmetry of a configuration is the corresponding symmetry mapping done in reverse.

2. Two Worlds: Abstract and Empirical

The realm of research activity is naturally divided into two worlds: the abstract and the empirical. The abstract world is generally regarded as the domain of the mathematician, logician, or purely theoretical physicist, while the empirical world is inhabited by experimental scientists of many varieties: physical, social, and others. [It has been established empirically that the less scientific a subject, the more likely it is that its practitioners call it a science. Outstanding examples include (in alphabetical order): divinity science, library science, military science, political science, and secretarial science.] There is a growing tendency, however, for people to live in both worlds in these interdisciplinary times.

Those who work entirely in the abstract world are engaged in deriving new theorems either from axioms or from an existing theory or coherent body of theorems. Such results are usually expressed in symbols rather than numbers, and rarely touch upon the real world.

On the other hand, the inhabitants of the empirical world "work for a living." Some live in laboratories and perform experiments in order to collect meaningful data leading to a scientific theory.

The two worlds are shown in Figure 1. The two loops, called theory building and experimentation, represent purely theoretical and purely experimental research.

Figure 1 exhibits a symmetric pair of directed links between the worlds, the first of which can be called interpretation in accordance with the use of this word in the preceding section. In a confrontation between these two worlds, the mathematician's theorems become predictions about the real world, which can be tested by the scientist. If a prediction is verified by an appropriate experiment, the scientist feels that the theorem really works, and the mathematician has found a *realization*.

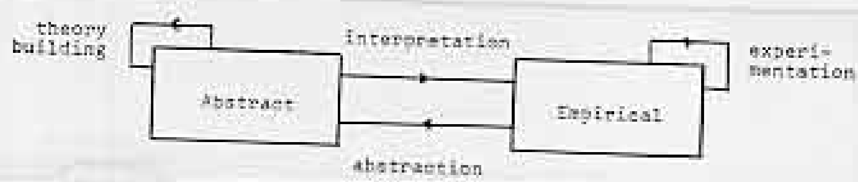

Figure 1 — Two worlds. [Two rectangles, representing the two worlds, are linked by a left arrow and a right arrow. The left rectangle is labelled "Abstract"; the right rectangle is labelled "Empirical." The right arrow, from the abstract world to the empirical world is labelled "interpretation." The left arrow, from the empirical world to the abstract world is labelled abstraction. A loop, labelled "theory building," is attached to the upper left of the abstract world. A loop, labelled "experimentation," is attached to the upper right of the empirical world. Inserted by Ed.]

If the predictions are entirely incorrect, the model cannot be used. However, in cases where the predictions are not verified, yet are "rather close" to correct, further abstraction is in order to construct a working model. This abstraction in the light of the experiment may suggest alternate hypotheses which should result in new theorems. These theorems hopefully will lead to better predictions than previously, and to a working model.

3. Two Worlds: Two Levels

Each of our two worlds may be divided into two levels. As we have indicated, the upper level of the abstract world deals with the development of mathematical systems by the derivation of theorems. We have discussed interaction between worlds at this level by means of interpretation and abstraction. In this section we shall observe that this same type of interaction can occur at the lower level.

The lower level of the abstract world deals with the foundations of mathematics, axioms, and logic. The research activities might involve trying to prove consistency or independence of an axiom system.

A rather esoteric and dramatic recent example of an important discovery at this level is given by the definitive work of Paul Cohen [1]. It is known (see Wilder [4], for example) that a $1-1$ correspondence can be constructed between the natural numbers $1, 2, \ldots$, and all the integers, $\ldots -3, -2, -1, 0, 1, 2, \ldots$, and between the integers and the rational numbers. These three sets of numbers are all said to have the same (infinite) *cardinality* which is conventionally denoted $\aleph_0$.

It is also known that there are more real numbers than integers. The real line is sometimes called the *continuum*, and so $\mathbf{c}$ is written for the number of reals. The *continuum hypothesis* states that there is no infinite set with cardinality between $\aleph_0$ and $\mathbf{c}$.

Cohen proved that the continuum hypothesis (as well as its negation) is consistent with the usual axioms of set theory. As a consequence, it is independent and can neither be proved nor disproved in that axiom system. Analogous to the development of non-euclidean geometry, two entirely different axiom systems have been created; one by assuming the continuum hypothesis, and the other by taking its negation. Cohen also proved the independence of the "axiom of choice."

On the other hand, the lower level of the empirical world also deals with foundations, but in the form of the basic laws of science. Kepler's Laws of Planetary Motion, Darwin's Law of Natural Selection, Newton's Laws of Motion, Kirchhoff's Laws of Electricity, and Einstein's Law of Special Relativity are all there.

The link between the two worlds at this lower level is quite analogous to that at the upper level. Thus interpretation of an axiom leads to a basic law about the real world, while an abstraction, a coherent set of scientific laws becomes an axiom system. The schematic representation of interaction between the two worlds is shown in Figure 2.

4. Two Levels: Derivation and Selection

Having discussed interaction between the two worlds, we shall now establish links between their upper and lower levels. The process of climbing from the lower level to the upper in the abstract world can be regarded as *derivation*. For we begin with an axiom system and then, sometimes painfully, derive progressively complicated theorems to obtain a mathematical system.

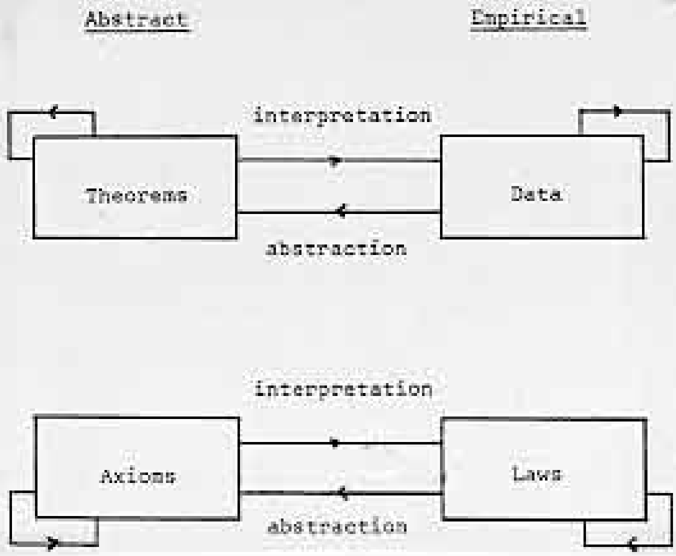

Figure 2 — Interaction between the two worlds. [There are four rectangles in this figure, arranged at the upper left, upper right, lower left, and lower right. The two uppers have a left arrow and right arrow linking them, as do the two lowers. The upper left rectangle is labelled "Theorems"; the upper right, "Data"; the lower left, "Axioms"; and, the lower right "Laws." The right arrow in each case is labelled "interpretation." The left arrow in each case is labelled "abstraction." The left hand side of the figure is labelled "Abstract"; the right, "Empirical." There is a loop attached to each of the four rectangles. Ed.]

Now consider how one goes from the upper level to the lower. From an existing body of theorems, an axiom system is to be built. To accomplish this, we select a body of particularly appropriate and fruitful theorems to use as axioms. This process of *selection* yields a small, more manageable and often more powerful system, which is conducive to the derivation of new theorems.

Selection in the empirical world involves collecting and studying vast amounts of data, and observing a pattern which may suggest a general law. Thus it is actually the *induction* process.

There appears to be no direct link in the empirical world from the lower level to the upper. Derivation does occur, and in fact uses the deduction process, but again and again we find that it takes the "long way around," as shown in Figure 3. One begins with several scientific laws (lower right), and abstracts them to formulas (lower left) from which theorems can be derived (upper left) which make predictions about the real world (upper right). It is convenient, however, to draw the link representing derivation directly as well, as we do later.

In general, innovative research is initiated in the upper level, and particularly in the upper right quadrant. This is due to the fact that the great majority of natural and fundamental questions arise from an attempt to observe or explain empirical phenomena. In fact, most research is done at the upper level, both right and left, while almost no one continuously remains at the lower level.

For example, in ancient Egypt, the discovery of geometric formulas was necessitated by the search for improved techniques in measuring and surveying. Problems in geometry were

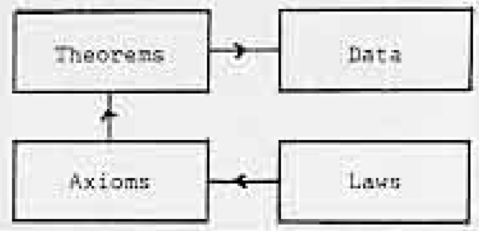

Figure 3 — Derivation in the empirical world [The four rectangles of Figure 2, are linked with three arrows from Laws to Axioms to Theorems to Data. Ed.]

solved long before Euclid organized the subject in an axiomatic formulation.

5. Research Schema

We contend that the above Research Schema represents all the types of interaction between the abstract and empirical worlds during the processes of research and discovery. Its two diagonal links are shortcuts which represent research processes that go directly to "opposite" quadrants. There do not seem to be any directly ascending diagonal links.

It is rarely but definitely possible to predict scientific laws from a body of theorems without actually working with experimental data. This is represented by the diagonal from upper left to lower right in the Research Schema. We shall see that Einstein took this route in his formulation of the theory of special relatively.

The shortcut from experimental data to axioms, skipping the formulation of laws, occasionally occurs in the social sciences when a careful analysis of data patterns produces a set of formulas that can be taken as axioms. These are then interpreted, and hopefully suggest an empirical law, without the selection process.

When considering routes between the two worlds, one must also allow for traversing loops at any quadrant one or more times. The upper right loop, for example, when traversed several times, indicates repeated efforts in observation and collection of data, before attempting to select corresponding laws.

One must also note that the most direct route is not often taken in research. This will become evident in the next section when we take a closer look at particular cases of discovery.

6. Sketches of Discovery

We shall illustrate the Research Schema with the work of several men who represent varied branches of science and mathematics. We begin with Euclid, whose work in the axiomatization and derivation of what we now call euclidean geometry is represented schematically in Figure 5. [It has been said that the ultimate recognition of a man's contribution is conferred when his name is made an adjective and not capitalized.]

Euclid: Although Euclid is the acknowledged father of geometry, his main contribution was to its organization rather than to its derivation. The early Egyptians already knew the rudiments of geometry, including a form of the pythagorean theorem, and formulas for the area and volume of many geometric figures. Thus we attribute the upper right quadrant in

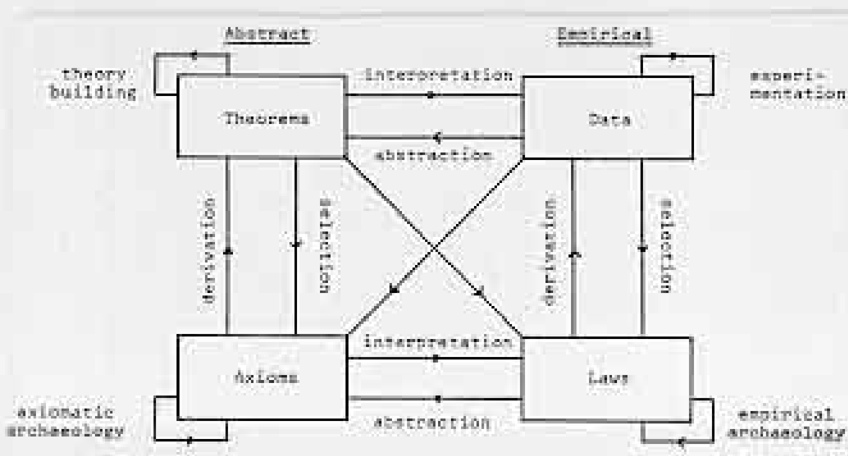

Figure 4. Research schema. [Draw Figure 2. Label the loop on "Theorems" as "theory building"; that on "Data" as "experimentation"; that on "Axioms" as "axiomatic archaeology"; and, that on "Laws" as "empirical archaeology." Add up and down vertical arrows joining the rectangles; label the downward arrow in each case as "selection"; the upward as "derivation." Draw the two diagonals – one with an arrow to suggest going from "Theorems" to "Laws" and the other from "Data" to "Axioms." Ed.]

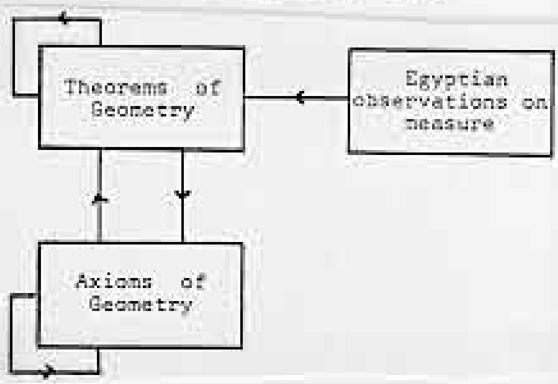

Figure 5 — Euclid's Research Schema. [Draw three rectangles: upper left, upper right, lower left. Label them, respectively, "Theorems of Geometry," "Egyptian observations on measure," "Axioms of Geometry." Add a loop to the two rectangles on the left. Join the upper left and lower left rectangles by an up arrow and a down arrow. Draw an arrow from the upper right to the upper left rectangle. Ed.]

Figure 5 to the Egyptians. The emphasis on proof, however, was introduced by the early Greeks and Euclid's contemporaries developed many of the theorems of geometry. Euclid selected the five axioms above from existing results. He then proved from these all the theorems of geometry then known and a few new ones, and presented a logical organization of the material in an exhaustive text. By today's standards, Euclid's axiomatic work is not rigorous, but it was an outstanding accomplishment for its time.

Newton: Unlike Euclid, Newton occupied every quadrant of the Research Schema. His

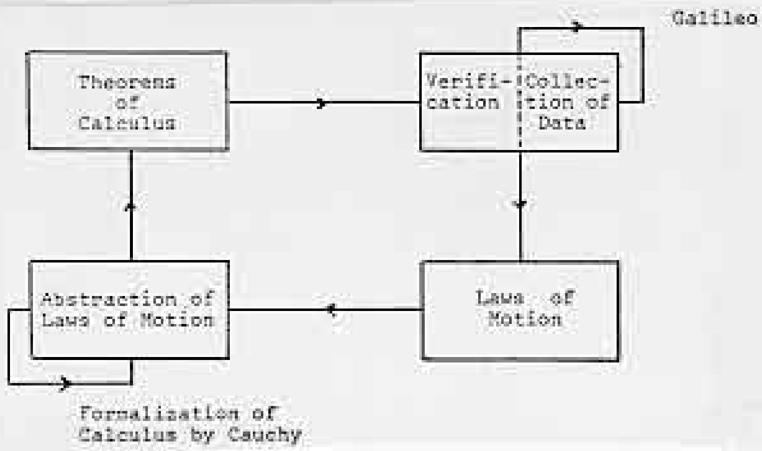

Figure 6 — Newton's Research Schema. [Draw four rectangles: upper left – "Theorems of Calculus"; upper right – left half labelled "Verification" right half labelled "Collection of Data"; lower left – Abstraction of Laws of Motion; lower right – "Laws of Motion." Join the rectangles with arrows forming a rectangular cycle oriented in a clockwise direction. Add a loop to the lower left rectangle; label the loop "Formalization of Calculus by Cauchy." Add a loop to the upper right rectangle; label the loop "Galileo." Link the "Galileo" loop to the down arrow as a dashed line separating "Verification" from "Collection of Data" in the upper right hand box. Ed.]

first work was on the upper level of the empirical world, where he experimented in chemistry and optics while still a student. Newton's most important results, however, were not derived from his own data, but from the work of those before him. His formulation of the Laws of Motion was induced from Galileo's extensive experimentation. Hence we credit the upper right loop in Newton's Research Schema to Galileo. Newton's Laws of Motion have been stated as follows:

1. Every body will continue in its state of rest or uniform motion in a straight line unless it is compelled to change that state by impressed force.
2. The rate of change of momentum is proportional to the impressed force and takes place in the line in which the force acts.
3. For every action, there is an equal and opposite reaction.

Newton left the empirical world and entered the abstract by expressing his laws symbolically as equations. His work with these resulted in the discovery of both differential and integral calculus. Others independently discovered these concepts, but it is believed that only Newton and Leibnitz (who discovered calculus independently) realized that differentiation and integration were inverse processes.

Calculus did not become mathematically precise until the next century when Cauchy introduced the necessary concepts of limit and infinite sequence. We draw a loop in the lower left quadrant of Figure 6 to represent Cauchy's work in the foundations of calculus.

This new branch of mathematics readily produced an abundant supply of theorems. The predictions which resulted were tested in the laboratory, and found to be entirely correct within the range of current measuring instruments.

Einstein: Eventually, more accurate measuring devices revealed that Newton's Laws of Motion could not explain the behavior of light on either the microscopic or astronomical level. Furthermore, the Michelson-Morley experiment proved conclusively that "ether" did

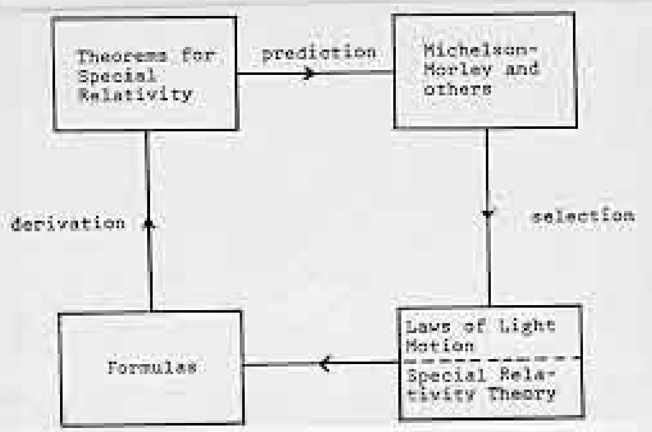

Figure 7 — Einstein's Research Schema. [Draw four rectangles. Label upper left: "Theorems for Special Relativity"; upper right – "Michelson-Morley and others"; lower left – "Formulas"; and, lower right is split (by a dashed line) - top half "Laws of Light Motion," bottom half "Special Relativity Theory." Arrows from upper left to upper right – "prediction"; from upper right to lower right – "selection"; from lower right to lower left; from lower left to upper left – "derivation." Ed.]

not exist. These discoveries led to a period of great activity in physics pioneered by Albert Einstein.

Like Newton, Einstein's major work resulted from data collected by scientists before him. Einstein was a purely theoretical physicist, and never worked in the upper right quadrant of the Research Schema himself. But he certainly stimulated an enormous number of experiments there. He proposed the following empirical axiom system as laws of light motion:

1. No physical object can travel faster than the speed of light.
2. The speed of light depends not at all on the relative positions of the source of light and the observer, or their relative speeds.
3. The mass at a velocity v of a particle equals its mass at velocity 0 divided by $\sqrt{1 - v^2/c^2}$, where c is the speed of light.

Einstein abstracted these three laws to an axiom system, from which he derived the body of theorems interpreted as the theory of special relativity. He found that in particular, his distance formulas for relativity theory were related to those of hyperbolic non-euclidean geometry; thus relativity theory provides a physical model for hyperbolic geometry. The Research Schema for this discovery is shown in Figure 7. We begin in the upper right with the Michelson-Morley experiment, and then go to the Laws of Motion of Light in the lower right, and their abstractions in the lower left. From there we go to the theorems of special relativity in the upper left, and finally to the experimental verification in the upper right where this cycle started. Einstein then went around this cycle again with his more refined theory of general relativity, which led to more precise predictions of physical measurements.

Darwin: Charles Darwin spent most of his life doing research in only one quadrant of the Research Schema, the upper right. His research career began when he became the official naturalist on the good ship Beagle, and embarked upon a five-year voyage. He made observations on all species of animals he could find, and took voluminous notes. During the remainder of his life, Darwin analyzed and classified these notes and all other available

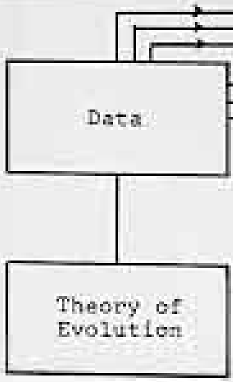

Figure 8 — Darwin's Research Schema. [Draw two rectangles, one above and one below. The top one is labelled "Data"; the bottom one is labelled "Theory of Evolution." There are three loops attached to the top one. There is a line linking the two rectangles. Ed.]

information. The climax of his work was the formulation of his Law of Natural Selection and his Theory of Evolution.

Darwin's theory asserts that all animal species have descended from a common origin. The variety of species results from "natural selection," in which those animals which are best adapted to their environment survive. Due to occasional mutations, certain animals in a species are better able to survive than others. These mutations may be passed on to their offspring who in turn will tend to survive and reproduce, eventually resulting in a new species which has been naturally selected.

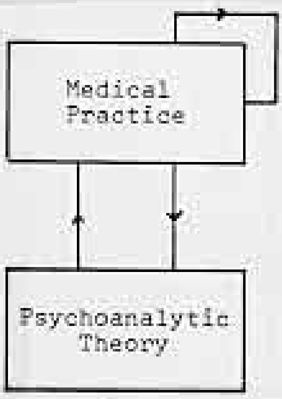

Figure 9 — Freud's Research Schema. [Draw two rectangles, one above and one below. Label the top one "Medical Practice." Label the bottom one "Psychoanalytic Theory." Join the two rectangles with an up arrow and a down arrow. There is a loop attached to the top rectangle. Ed.]

Freud: Sigmund Freud, like Darwin, stayed in the empirical world. In fact, their Research Schemata are quite alike, as seen in Figures 8 and 9. He began with a medical degree and turned from general practice to specialization. Freud (in collaboration with J. Breuer initially) did research in the treatment of "hysterical" patients who had physical symptoms for which no physical cause could be found. He inferred from the study of many cases that the symptoms could be traced back to some repressed childhood trauma, and went on to develop the concept of the subconscious together with the id, ego, and superego. First through hypnosis, and later through "free association," Freud was able to induce himself and his patients to recall these forgotten experiences, and alleviate their symptoms.

Much of the psychoanalytic theory which Freud developed is still highly controversial today, although it has made a lasting impact on the development of many modern theories in psychology.

There has been a highly publicized report of the proof of a deep and important theorem by a mathematician while boarding a bus in Paris. It may be just as true as the anecdote about Newton's finding his law of gravitational attraction when an apple fell off its tree and landed on his head. This sort of phenomenon does occur, but fortunately is not an intrinsic part of the discovery procedure. In the words of Hans Zinsser,

> It is an erroneous impression, fostered by sensational popular biography, that scientific discovery is often made by inspiration ... . This is rarely the case. Even Archimedes' sudden inspiration in the bathtub; Descartes' geometrical discoveries in his bed; Darwin's flash of lucidity on reading a passage in Malthus; Kekule's vision of the closed carbon ring came to him on top of a London bus; and Einstein's brilliant solution of the Michelson puzzle in the patent office in Bern, were not messages out of the blue. They were the final co-ordinations, by minds of genius, of innumerable accumulated facts and impressions which lesser men could grasp only in their uncorrelated isolation, which — by them — were seen in entirety and integrated into general principles. The scientist takes off from the manifold observations of predecessors, and shows his intelligence, if any, by his ability to discriminate between the important and the negligible, by selecting here and there the significant steppingstones that will lead across the difficulties to new understanding. The one who places the last stone and steps across to the *terra firma* of accomplished discovery gets all the credit. Only the initiated know and honor those whose patient integrity and devotion to exact observation have made the last step possible.

When a researcher has become sufficiently steeped in his problem, he has amassed enough meaningful data (mathematicians also accumulate data via "thought-experiments") to perceive the proper pattern and conceive the correct conjecture. This is a necessary but not sufficient step toward establishing a theorem. A proof, which is valid, must be supplied; otherwise, the assertion remains a conjecture. The two talents of conjecture and proof appear to be quite separable.

7. What Should They Be?

It is becoming more fashionable to use mathematical models as a powerful analytic device for advancing scientific research in a remarkable variety of disciplines. This usage is certainly not unwarranted, since models, when used with care and discretion, can and should be of great value in the clarification of existing problems and the formulation of important new ones. Unfortunately, it seems that models are misused all too often. The word 'model' is sometimes bandied about by people with little conception of its real meaning simply because it is *au courant*. They don't even define 'model', but use the word to suit their own purposes.

A model need not be impressively confusing in order to be valuable. In fact, one of the main contributions of a model lies in its ability to simplify a problem, and so it should be no more complicated than necessary.

Neither should a model be symbol-rich but idea-poor. Models which hide miniscule content behind a mass of symbolic formulas tend to look impressive, but add nothing. "Mystery is no criterion of knowledge." For example, a recent paper in a leading psychological journal had only one abstract idea: the number of elements in the union of two sets is the sum of the number of elements in each minus the number they have in common. Alas, the author apparently did not recognize it as the simplest special case of the Principle of Inclusion and Exclusion.

Another unfortunate use of mathematical models occurred in a published paper in sociology in which there were ten axioms and zero theorems. However, an interpretation was then given which resulted in ten "empirical theorems," one for each axiom. This 1-1 correspondence between axioms and empirical theorems simply involves the preparation of axioms which will yield desired empirical assertions.

Furthermore, an axiom system should not be constructed for the artificial purpose of deriving just one theorem which has already been verified statistically. Clearly such a model only clutters the literature and does not involve genuine derivation.

We do not wish to lay all the blame for the misuse of mathematical models on scholars in the empirical world: it occurs in the abstract world as well. The following passage by the eminent linguist Gustave Herdan [2] shows the dual roles the two worlds can play in the misuse of models.

> Without going into details, I will only mention a certain quantitative relation known to linguists as the 'Zipf law'. Mathematicians believe in it as a law, because they think that linguists have established it as a relation of linguistic facts, and linguists believe in it because they, on their part, think that mathematicians have established it to be a mathematical law. As can be shown in five minutes, it is not a law at all in the sense in which we speak of natural laws.

Loosely stated, this law of Zipf proposes a high correlation between the frequency of use of words and their brevity.

Another typical superficial use of mathematical models involves the bland assumption that the most elementary parts of an existing branch of mathematics apply unchanged to a problem in social science. Typical examples include high school algebra, coordinate geometry, matrix manipulation, graph theory, and the probabilistic theory of Markov chains. In such models, the typical procedure is to assign empirical terms to the mathematical variables by way of interpretation at the lower level. Then the existing theorems and methods of calculation are translated at the upper level into statements which are claimed to be new empirical findings.

What, then, should mathematical models be? We have suggested that they should lead to new theorems, but this is not always necessary. The precise thinking involved in the careful formulation of an axiom system will lead to an improved conceptualization of the empirical phenomena at hand. This in turn can suggest the proper variables to measure, and perhaps an approach to the measurement problem.

Sometimes an existing area of mathematics can be quite useful as a mathematical model provided it is augmented by one or more new axioms suggested by the real world. The most productive models, however, have involved derivation. For it is only after the derivation of new theorems that unexpected and far-reaching predictions can be made. From a mathematician's viewpoint, it is best if derivation leads to nontrivial theorems, which actually qualify for publication in the mathematical literature. To summarize, it is our personal and perhaps controversial contention that mathematical models will lead to significant and natural growth in both the abstract and empirical worlds.

Acknowledgment.

Research supported in part by Grant MH22743 from the National Institute of Mental Health.

References

[1] P. J. Cohen, *Set Theory and the Continuum Hypothesis*, Benjamin, New York, 1966.

[2] G. Herdan, *Sonderdruck aus Zeichen und System der Sprache*, Vol. 2, Akademie-Verlag, Berlin, 1962, p. 108.

[3] M. Richardson, *Fundamentals of Mathematics*, 3rd ed., Macmillan, New York, 1966.

[4] R. L. Wilder, *Introduction to the Foundations of Mathematics*, 2nd ed., Wiley, New York, 1965.

3B. WHERE ARE WE? COMMENTS ON THE CONCEPT OF THE "CENTER OF POPULATION"

Frank E. Barmore
Associate Professor of Physics
University of Wisconsin, La Crosse

Reprinted, with permission, from
The Wisconsin Geographer
Vol. 7, pp. 40-50, 1991
A publication of The Wisconsin Geographical Society

1. Introduction

I was recently flabbergasted when I discovered how the Bureau of the Census calculates the location of the center of population of the United States following each decennial census. I found that:

> The center of population is the point at which an imaginary, flat, weightless, and rigid map of the United Sates would balance if weights of identical value were placed on it so that each weight represented the location of one person on April 1, 1980. Located at latitude 38 degrees, 8 minutes, 13 seconds north, and longitude 90 degrees, 34 minutes, 26 seconds west, ... The computation of the center of population in 1980 was based on the 1980 population counts and the 1970 centers of population for counties. County population centers have not been determined for 1980. The center is the point whose latitude (LAT) and longitude (LONG) satisfy the equations
>
> $$\mathrm{LAT} = \frac{\Sigma w_i \times \mathrm{lat}_i}{\Sigma w_i},$$
>
> $$\mathrm{LONG} = \frac{\Sigma w_i \times \mathrm{long}_i \times \mathrm{Cos}(\mathrm{lat}_i)}{\Sigma w_i \times \mathrm{Cos}(\mathrm{lat}_i)},$$
>
> where lat_i, long_i, w_i are the latitude, longitude, and population, respectively, of the counties. (U. S. Bureau of the Census, 1983, Appendix A, p. A-5.)

The statements were surprising for a number of reasons. First, as every good introductory physical geography text book points out, a flat map of the earth's curved surface is a *distorted* map. Though distortions can be reduced by a careful choice of projection, appropriately centered, some distortion always remains and cannot be eliminated. Of course, one may define something any way one wishes, but at least it should be stated which method of projection is used to create this "flat map." Second, the formulæ given suggest that east-west distances are measured (reasonably, though arbitrarily) along parallels of latitude, which are small circles. But distances are usually measured along great circles and this would produce different results. Third, the formulæ yield results that differ from results of other accepted definitions of the center of population. Fourth, the formulæ can produce some rather peculiar results. These will be discussed below.

The purpose of this paper is to discuss some of the difficulties in the concept of the "center of population" when applied to populations spread over enough of the earth's surface that its curvature is noticeable. A more satisfactory definition of the center of population is proposed.

2. Preliminary Remarks

When considering the characteristics of a large group of anything distributed over a region it is often useful to concentrate on only the most basic characteristics, the first three moments of the distribution: 1) the population, 2) the location, and 3) the spatial dispersion of the group. This paper concentrates on the second moment, the location. There are many ways location could be specified. Almost a century ago Hayford (1902) convincingly argued that the most appropriate measure of location of an area or population is a statistic called the average (arithmetic mean). Abler, *et al.* (1971) agree: "When we ask *where* questions about distributions we almost always desire an average location which represents the entire set." When averaging the location, each location is "weighted" according to the specific characteristic of interest and the result is the average location of the weighting character. The result is a "center of mass" if the weighting character is mass, a "center of area" (or geographic center) if the weighting character is area, a "center of population" if the weighting character is population, *etc.* Sviatlovsky and Eells (1937) have discussed in some detail the use and significance of the concept of the "center" in geographical regional analysis.

Locations can be described as vectors whose magnitudes and directions are taken as the distances and directions of the items whose center is to be calculated. Then the power and convenience of vector algebra can be used to calculate the center in one, two, three or even higher dimensional spaces. If the space is "flat" (Euclidean) then the process is quite straightforward, though tedious. Also, by using vectors in a Euclidean space to represent the distances and directions of individuals in a population, there exist several interesting and useful concepts. First, when the center of the coordinate system used is at the center of population then the vector sum of the distances of all the people is zero and the sum of the squares of the distances of all the individuals is at a minimum. The minimum sum of the distances squared when measured from the average is not an accident, but rather, the result of a fundamental mathematical relation between the two quantities. Sufficiently fundamental is this relation that Warntz and Neft (1960), for example, define the mean as the place from which the sum of squares of the distances to each member of the population is minimum. Second, when the center of the coordinate system used is *not* at the center of population then the vector sum of the distances of all the individuals provides the distance and direction of the center of population from the center of the coordinate system. I will use these characteristics later.

However, the surface of the earth is not "flat," but "curved," and though finite, is without a boundary. On such a surface one can get into difficulty with the concept of average location. Where, for example, is the "center of area" (geographic center) of the earth's entire surface? If one chooses to preserve an earth surface provincialism it is not clear how one can modify the "vector representation" of distance and direction for the locations of individuals in a population. Some criteria are needed.

I suggest that any reasonable definition of "center of population" should meet at least the following standards: (1) population distributions which are symmetric about some central point should have their center of population at this central point and (2) distances should be measured as true distances, either on the surface along great circles or in three dimensions along straight lines. It would also be desirable to have any definition satisfy additional restrictions: a) it should correspond to one's common understanding of "center of population," b) it should reduce to the usual definition of "center of population" when there is no curvature and c) it should be easily extended to nonspherical surfaces — for example, the International Ellipsoid or the geoid.

In the examples and discussions which follow, all angles and great circle arc lengths are given in degrees and decimal degrees. Azimuths of places from any point are measured from North toward East. Latitudes and longitudes are designated as North or South and East or West respectively. I have ignored the differences between the shape of the earth and a sphere, as does the Bureau of the Census, when calculating centers of population (U. S. Bureau of the Census, 1973). When calculating the 1980 center of population of the United States in the various examples, I have used the original published 1980 populations (U. S. Bureau of the Census, 1983) and the unpublished 1980 centers of population for the fifty states and the District of Columbia (see Appendix A). When calculating the 1910 and 1880 centers of population of the United States I have used the published populations and centers of population of the various states and the District of Columbia (U.S. Bureau of the Census, 1913 and 1914).

3. Census Bureau Center of Population Formulæ

Imagine a circumpolar population uniformly distributed along, say, the 70th parallel of latitude north (see Fig. 1). If longitude is measured from 180° W through 0° to 180° E then the Bureau of the Census formulæ put the center of population at 70° N on the Greenwich meridian. Yet, surely the center of this population is at the North Pole. The Bureau of the Census formulæ fail to meet the suggested standard.

This failure is not due to the choice of a circumpolar population. Even at mid-latitudes the formulæ fail to meet the suggested standard. Consider a second example, a collection of eight equally populated places located on a circle of 15 degrees of arc radius. Center the circle at 38° N and 90° W. Choose the position of the first place so that its azimuth from the center point is 15 degrees and each succeeding place has its azimuth 45 degrees greater than the preceding place (see Fig. 1). The latitude and longitude of each of the eight places can be calculated with spherical trigonometry. The results are shown in Table 1. When the Bureau of the Census formulæ are used to calculate the center of population of the odd numbered places, then of the even numbered places and finally for all eight places the results differ. Specifically, for the odd numbered places one finds:

$$\text{LAT} = 37.2351\,\text{N}, \qquad \text{LONG} = 90.0146\,\text{W}.$$

For the even numbered places one finds:

$$\text{LAT} = 37.2155\,\text{N}, \qquad \text{LONG} = 89.9855\,\text{W}.$$

While for all eight places one finds:

$$\text{LAT} = 37.2253\,\text{N}, \qquad \text{LONG} = 90.0000\,\text{W}.$$

Yet, surely the different symmetric distributions centered on the same point should have their center of population at the same place and surely that place (in this example) should be 38° N and 90° W ! Once more, the Bureau of the Census formulæ do not meet the suggested standard.

For a third example, consider the simplest possible case: two equally populated places equidistant and in opposite directions from a central place. Specifically, place the center at

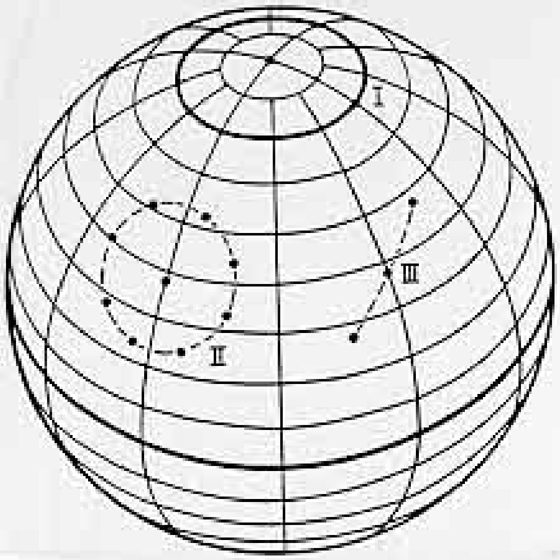

Figure 1. Example population centers and distributions used in Part 3 of the text. Figure available in hard copy only; content should be clear from text and from caption. A sphere is drawn containing parallels and meridians. Three figures are highlighted on this sphere. I: a circumpolar population distributed along the 70th parallel of latitude north. II. a symmetric population distribution centered at latitude 38° N and longitude 90° W. This population is spaced at eight locations around the perimeter of a circle. III: a simple symmetric population distribution centered at 38° N and longitude 30° W. This population is spaced at either end of a line segment centered at III. The precise locations of places in distributions II and III are listed in Tables 1 and 2, respectively.

Table 1. Locations of places, Example II

Place	N. Latitude	W. Longitude
1	52.3434	83.7050
2	44.1596	71.7938
3	32.8128	72.6948
4	24.7119	81.8100
5	23.4333	94.1868
6	29.5187	104.9264
7	40.3511	109.1501
8	50.4718	101.7316

38° N and 30° W and the two places 15 degrees of arc from the center to the northeast (Az = 45) and to the southwest (Az = 225). (See Fig. 1). The latitude and longitude of the two places can be calculated and are shown in Table 2. When the center of population is calculated for this very simple population distribution of two places using the Bureau of the Census formulæ the result is:

$$\text{LAT} = 37.2057\,\text{N}, \qquad \text{LONG} = 29.9626\,\text{W}.$$

Yet, surely the true center is at the central point: 38° N, 30° W ! And yet again, the Bureau of the Census formulæ do not meet the suggested standard.

Table 2. Locations of places, Example III

Place	N. Latitude	W. Longitude
9	47.6377	14.2401
10	26.7737	41.8289

What the Bureau of the Census formulæ calculate is not the latitude and longitude of the center of population, but rather, two different and separate statistics: 1) the average latitude of the population and 2) the longitude of the average east-west distance of the population *on a specific map projection*. This longitude is *neither* the average longitude *nor* the longitude of the center of population. The formulæ calculate the location of a place that differs from other common measures of the center, such as the median or mean location (as defined by Warntz and Neft, 1960, and used by Haggett, *et al.*, 1977). The result of the Bureau of the Census formulæ is the latitude and longitude of a place that cannot justifiably be named the "center of population" as the examples above clearly demonstrate.

For comparison with later examples and further discussion, I have calculated the 1980 center of population of the United States with the Bureau of the Census formulæ. The result is:

$$\text{LAT} = 38.1376\,\text{N}, \qquad \text{LONG} = 90.5737\,\text{W}.$$

This differs from the location published by the Bureau of the Census (latitude of 38°08′13″ or 38.1369 N and longitude of 90°34′26″ or 90.5739 W) by one to three seconds of arc. This very small difference results from my using the populations and centers of the fifty states and the District of Columbia rather than the much larger full list of populations and centers of all the individual counties or enumeration districts used by the Census Bureau. As the discussion and all the examples that follow are based on the same set of data this small difference is unimportant — the examples approximate and represent more extensive computations and their outcome well enough. The concerns in this paper are the methods used rather than the data on which they operate.

4. Census Bureau Center of Population Description

The Bureau of the Census description of the center of population does not give the map projection used. If the center of population (the "balance point" mentioned in the description) is calculated on a flat map constructed using various map projections the results vary. In order to demonstrate this I have calculated the 1980 centers of population (the balance point of the population distribution) for the U.S. using several different flat map projections. The projections used are all well known, having been developed in the 18th century, the 16th century and much earlier. For the projections chosen, descriptions given in numerous texts were sufficient for the derivation of the relevant formulæ for laying out the projections. Alternately, one may refer to detailed monographs, such as the one by Snyder (1987), for the appropriate formulæ. The results for each of the selected projections are listed in Table 3 and displayed in Fig. 2.

Table 3. The 1980 center of population of the United States when using the Bureau of the Census prose definition and various different map projections

No.	Projection and Comments	Center of Population N. Latitude	Center of Population W. Longitude
1	Cylindrical Equal-Area	37.9818	90.4237
2	Equidistant Cylindrical (Plate Carrée)	38.1376	90.4237
3	Sinusoidal (centered at 0 W)	38.1376	90.2532
4	Sinusoidal (centered at 60 W)	38.1376	90.4655
5	Sinusoidal (centered at 120 W)	38.1376	90.6778
6	Sinusoidal (centered at 180 W)	38.1376	90.8901
7	Equatorial Mercator	38.2945	90.4237
8	Transverse Mercator (centered at 90 W)	39.2344	90.6732
9	Azimuthal Equal-Area (centered at 0N,0W)	39.2583	89.0197
10	Stereographic (centered at N. Pole)	39.7137	90.4888

Note that the calculated centers depend not only on the projection chosen but also on the center and the orientation selected for the projection. The results differ as little as they do from one another because the projections chosen leave the United States in those portions of the resulting maps where the distortions are not extreme. Indeed, using the Bureau of the Census descriptive definition of the center of population, I believe, that, given sufficient time and mischievousness, one could choose projections of various orientations and center that would place the center of population *any place one wished* !

5. Agreement between Description and Formulæ

As it happens, the description and the formulæ currently given by the Bureau of the Census agree for one map projection — a normal Sinusoidal (Sanson-Flamsteed) *with the central meridian of the map the same as the meridian of the center of population*. Indeed, the Bureau of the Census formula for the longitude of the center of population can be derived by answering the following question: What must the central meridian for a normal Sinusoidal projection be in order for the center of population (that is, the balance point of the population distribution) to lie on the central meridian?

But, why is the Sinusoidal projection (and associated formulæ) preferred? If this projection holds special appeal, why isn't the latitude of the center of population determined in a similar way? One could determine the longitude *and* latitude by answering the following question: What must the center for a Sinusoidal projection be in order for the center of population (the balance point of the population distribution) to lie at the center of the map projection? The result, for 1980, is:

$$\mathrm{LAT} = 39.1825\,\mathrm{N}, \qquad \mathrm{LONG} = 90.4934\,\mathrm{W}.$$

The Bureau of the Census first calculated the "center of population" of the United States following the census of 1870 (Walker, 1874). Then, as now, the concept of a balance point was stated as underlying the computation of the center of population. The description of the calculation method (the formulæ are not displayed) indicates the method was very similar to that currently used. East-west locations were taken as the distance from the 67th meridian

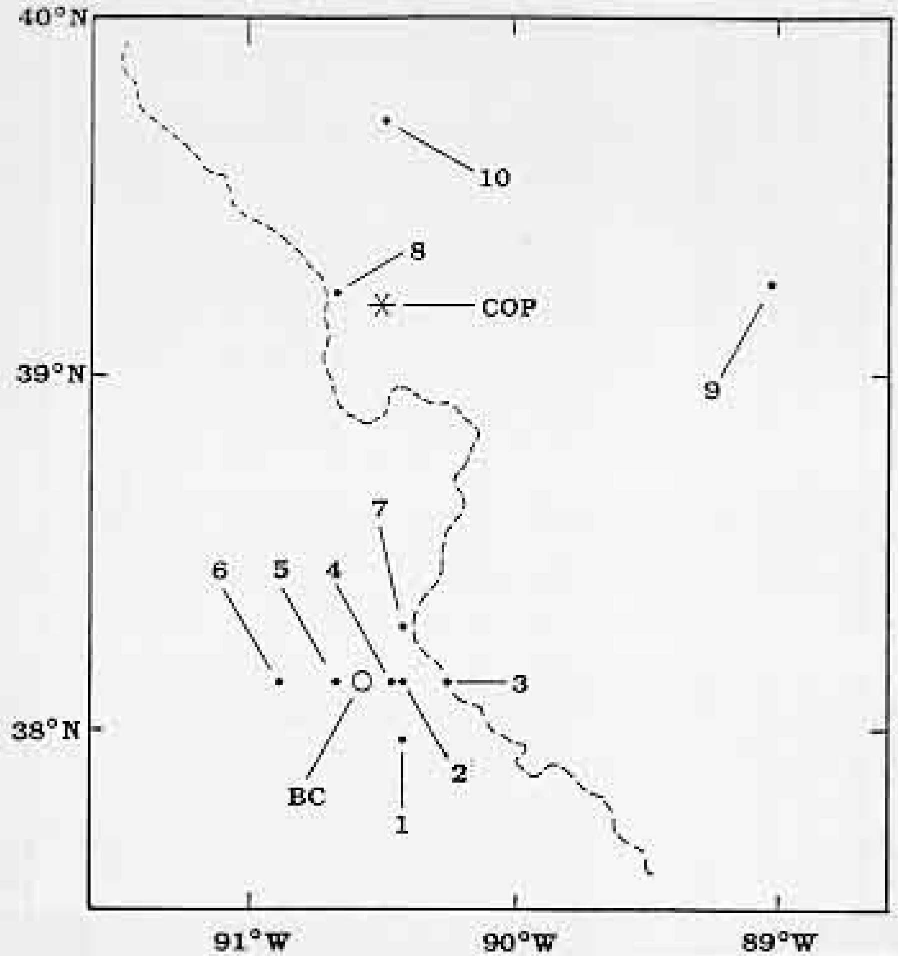

Figure 2. The "Centers of population" for 1980 for the United States calculated according to various definitions. The center shown by an asterisk (*) and labeled COP was determined by the method proposed in this paper which takes the curvature of the earth's surface into account in an appropriate manner. The place shown as an open circle (○) and labelled BC is that published by the Bureau of the Census as the location of the center of population. As discussed in the text, this location should not be called the center of population. Places shown as solid circles and numbered are those which result when the center of population is calculated on various map projections using the Bureau of the Census prose definition of the center. The prose definition does not specify which projection should be used. The numbers refer to the list of projections given in Table 3. The Illinois-Missouri boundary, shown dashed, was taken from The National Atlas (U. S. Geological Survey, 1970). This figure is available in hard copy only; its content should be clear from Table 3 when one also notes that the Census calculated "Center" lies in Missouri as do centers 1, 2, 4, 5, 6, and 7 (from Table 3); the COP Center lies in Illinois as do centers 8, 9, and 10 (from Table 3); Center 3 appears to lie on the border between Illinois and Missouri.

west, measured along parallels of latitude. North-south locations were taken as the latitude above the 24th parallel north. Thus, the calculations of the center of population for the 1879 census are equivalent to calculating the balance point on a Sinusoidal projection with its central meridian at 67° W. Since 1870, east-west distances were measured from other meridians, chosen to be near the estimated center of population. In 1910 for example, the 86th meridian was chosen (U. S. Bureau of the Census, 1913) and thus the calculations of the center of population for the 1910 census are equivalent to calculating the balance point on a Sinusoidal projection with its central meridian at 86° W.

6. Proposed Definition of the Center of Population

One could avoid some of the problems discussed above by avoiding the use of the "statistic," the average. There are other measures of the "center," such as the median. But as Hayford (1902) pointed out long ago, there are fundamental difficulties with the concept of the median for two dimensional distributions. There is no unique point that "divides" two dimensional distributions in half. Another possible measure is the "point of minimum aggregate travel" — the point for which the sum of all the distances to the various individuals would be minimum. But, as Eells (1930) demonstrates, this point has some peculiar characteristics. Also, as Court (1964) makes clear, the point of minimum aggregate travel is very difficult to find. There are now elegant, very powerful and very general techniques for solving such problems (Kirkpatrick, *et al.*, 1983). But, this is beside the point. The statistic that we want is one that reflects where people *are*, not where they might congregate with the least total travel. The appropriate statistic is the mean.

Whether calculating the mean location or the point of minimum aggregate travel, an arbitrary decision must be made: are the calculations to be done on the curved two dimensional surface of the globe (or appropriate flat map) or are they to be carried out in three dimensions? If the point of minimum aggregate travel were calculated in three dimensions the paths to be traveled and the resulting point would lie below the earth's surface. What value would there be in finding a point of least cumulative travel when the place to congregate and the paths to be traveled are inaccessible? I conclude that if one is interested in the point of minimum aggregate travel, it should be calculated on the earth's surface.

If the mean (average) location of a population is calculated in three dimensions, the resulting point is located below the surface. In the case of the United States in 1980, I find this mean location at:

$$\mathrm{LAT} = 39.1823\,\mathrm{N},$$

$$\mathrm{LONG} = 90.3477\,W \text{ and Depth} = 0.0259 \times R$$

Where R is the radius of the sphere representing the earth. Taking $R = 6371$ km, the depth is about 165 km. But calculating the average location of simple surface distributions in three dimensions can yield some peculiar results. Consider three different equatorial population distributions, each consisting of four equally populated places with longitudes as follows:

Example IV: places at 50.00 W, 15.00 W, 15.00 E, 50.00 E.
Example V: places at 50.00 W, 15.00 W, 15.00 E, 130.00 E.
Example VI: places at 62.23 W, 60.00 W, 60.00 E, 62.23 E.

In all three examples the center of population (average location) ends up at the same place — on the equator at the Greenwich meridian — and differ only in their depth below the surface, if at all. (Examples V and VI have centers at the same depth.) But, we are largely confined to the surface of the earth and from this provincial point of view the center of population in example V should be far to the east of the centers in examples IV and VI. I find it unsatisfactory for populations of such different East-West distribution to have "centers" which differ only in depth or not at all. I conclude that average locations should be calculated *on* the earth's surface.

To insist that calculation of the average location or point of minimum aggregate travel must be done in three dimensions is no more (or less) reasonable than to insist that the only proper map projection is on a globe. I leave to others the task of championing the computations of two dimensional population distribution statistics in three dimensions. I believe it is legitimate to consider the population distribution a two dimensional distribution and display its characteristics on the two dimensional surface of a globe or appropriate flat map.

I suggest that the appropriate definition of the center of population is one similar to the descriptive definition given by the Bureau of the Census but with one addition. Specifically, the center of population is the point at which an imaginary, flat, weightless, and rigid map of the United States *constructed by a specific method of projection* would balance if weights of identical value were placed on it so that each weight represented the location of one person .... It only remains for one to choose the specific type of projection. It is distance and direction which are central to any calculation of the center of any population distribution. Therefore, I suggest that the only map projection (on which to find the balance point of the population distribution) is one where distances and directions of the individuals in the population are undistorted. If one chooses to measure the distances and directions from the center of an Azimuthal Equidistant map, or on the surface of a globe, they will be undistorted. Then one can create vectors whose magnitudes and directions are the true distances and directions of the various populated places. A vector sum can be done and the result is an *estimate* of the distance and direction of the center of population. It is only an estimate because, although a map is flat, the earth's surface is not. However, it is a very good estimate, *and the closer the map projection's center is to the center of population the better is the estimate.* Whether one chooses to carry out the computation on an Azimuthal Equidistant map or on the surface of a sphere is immaterial, since the process is algebraically identical and the results are numerically identical.

Because the calculating of the center only produces an estimate, the procedure must be an iterative one, with the center of the projection in each iteration being the estimate of the center of population from the previous iteration. But the estimate is an excellent estimate, so the process converges rapidly. The iteration continues until the center of population is as close to the center of the map as one wishes. When calculating the U. S. center of population in this manner I find:

Latitude of the 1980 U. S. center of population = 39.1980 N.
Longitude of the 1980 U. S. center of population=90.4978 W.

This point lies about 125 km from the center given by the Bureau of the Census and is in Greene County, Illinois, about 14 km southwest by south of Carrollton, the county seat (see Figs. 2 and 3).

The iterative calculation is not the computational nightmare that one might imagine (see Appendix B). Even when choosing the initial starting point at latitude zero and longitude zero or at latitude 20° S and longitude 20° E, the process rapidly converges to within 0.000001 degrees of the answer in four or five iterations. But one knows that the U.S. center of population is not in or near Africa — there is no point in beginning the computations there. As the approximate center of population can be guessed, only two or three iterations are necessary to calculate the center of population to ample accuracy.

I have also tested this procedure on the three example distributions used in Part 3 above and find that in all three cases the process rapidly converges on the expected central point.

Therefore, I suggest that the proper definition of the center of population of the United States (or for any population distributed over a substantial portion of the earth's surface) is:

> The center of population is the point at which an imaginary, flat, weightless, and rigid map of the United States would balance if weights of identical value were placed on it so that each weight represented the location of one person on a specific date. The map in question is an azimuthal equidistant map whose center is at the center of population which must be calculated by successive approximation.

This suggested definition of the center of population has the following advantages: (1) The center of populations symmetric about some central point is at that central point. (2) The true distance of each place is used in the computation. (Thus, this definition satisfies the two suggested standards given in Part 2 above.) (3) The suggested definition of the center of population also satisfies two of the three additional restrictions desired and stated in Part 2 above: (a) it corresponds to one's common understanding of center of population in that it does find the balance point of a distribution — though one must be very specific about how the distribution is displayed, and (b) mathematically there is a correspondence to the usual definition of the center in the sense of the average — the vector sum of the "distances" is zero when measured from the center and the sum of the squares of the "distances" is minimum when the distances are measured from the center. In addition, when the center of the map is not at the center of population the vector sum of the "distances" points approximately to the center of population. Finally, our definition would reduce to the usual mathematical definition when there is no curvature.

It is not clear that the definition suggested can be extended to non-spherical curved surfaces and thus satisfy the additional desired restriction (c) mentioned in Part II above. I believe it would work for the center of population of the United States on an ellipsoid of revolution but there could be difficulties for non-spherical surfaces in general — the shortest distance between two points may not be uniquely defined and one may end up with several centers of population, all equally legitimate.

A widely known use of the decennial centers of population determined by the Bureau of the Census is their display on a map of the United States depicting the historic westward shift of the population. In addition to this westward shift, these centers have slowly moved south since the turn of the century. By 1980, the center determined by the Bureau of the Census was more than a degree of latitude (*ca.* 110 km) south of where it was located in 1790. In contrast, the center of population calculated by the proposed method has followed a different path, diverging from the other path, and in 1980 was located at about the same latitude as the 1790 center. (The smaller the east-west dispersion of the population, the smaller will be the difference between the center of population calculated by the proposed method and the center the Bureau of the Census calculates. Thus, one would expect the locations calculated by either method to be about the same in 1790, before extensive westward national expansion occurred.)

In order to show the increasing divergence of the two paths I have calculated the centers of population for 1910 and 1880 using the proposed method. The results are shown in Fig. 3 and labeled COP. Also shown (and labeled BC) are the locations determined and published by the Bureau of the Census for the same years. One can see that the average latitude of the population (which is what the Bureau of the Census calculates) has moved farther south than has the center of population.

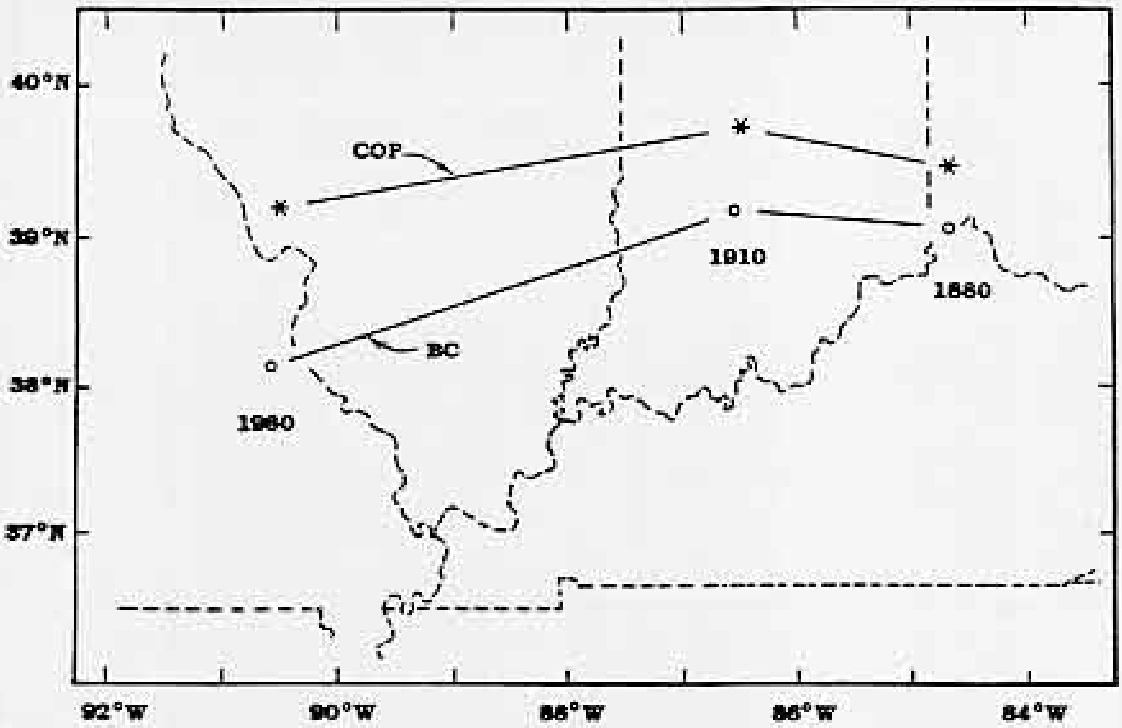

Figure 3. "Centers of population" for 1880, 1910 and 1980 for the United States calculated according to various definitions. Centers shown by asterisks (*) and labeled COP were determined by the method proposed in this paper which takes the curvature of the earth's surface into account in an appropriate manner. Places shown as open circles (○) and labeled BC are those published by the Bureau of the Census as the location of the center of population. As discussed in the text, these locations should not be called the centers of population. State boundaries, shown dashed, were taken from *The National Atlas* (U. S. Geological Survey, 1970). This map is available in hard copy only; it does not transmit electronically. Its content should be clear from the combination of the text and this caption.

7. Summary

For more than a century the Bureau of the Census has been calculating and displaying on maps a place designated as the "center of population" of the United States. The method used in this computation is equivalent to calculating the average location of the population on a Sinusoidal map projection. As indicated in the previous discussion, such a method does not adequately take into account the curvature of the earth's surface. As a result, what the Bureau of the Census calculates should not be called the "center of population." It is, rather, the location of a point that has the population's average latitude and the population's average distance (measured east-west along parallels of latitude) from an arbitrarily chosen meridian.

A different method of calculating the center of population has been proposed in this paper. Like the Bureau of the Census method of calculation, the proposed method is based on the concept of the balance point of the population distribution and thus corresponds to one's common understanding of the center. In contrast to the Bureau of the Census method, the proposed method takes into account the curvature of the earth's surface and map projection distortions in an appropriate manner and is based on measuring distances along great circles.

When calculated as proposed, the center of population's location differs substantially from that calculated by the Bureau of the Census. Not only is this true for 1980, but also for other census years, and the greater the east-west dispersion of the population, the greater will be the difference.

8. Appendix A

The unpublished 1980 population centers of the fifty states and the District of Columbia used in the various examples were obtained from the Bureau of the Census. As they are unpublished, a complete list of the center latitudes and longitudes that were used in the computations discussed in this paper is supplied below.

This is the data used as the representative example data set in "Where are we? Comments on the concept of the 'center of population' " by Frank E. Barmore, published in *The Wisconsin Geographer*, Vol. 7, pp. 40-50, (1991), a publication of the Wisconsin Geographical Society.

The table below shows the original state populations and also, State Centers of Population supplied by the Bureau of the Census. The first coordinate for the center of population is measured in degrees of longitude west of the prime meridian; the second coordinate is measured in degrees of latitude north of the equator.

Place	Population 1980	Center of population 1980	Center of population 1980
Alabama	3,893,888	86.7750	32.9923
Alaska	401,851	148.4964	61.3650
Arizona	2,718,215	111.7186	33.3245
Arkansas	2,286,435	92.4340	34.9718
California	23,667,902	119.4380	35.4746
Colorado	2,889,964	105.1809	39.4868
Connecticut	3,107,576	72.8760	41.4906
Delaware	594,338	75.5636	39.4450
D. C.	638,333	77.0088	38.9074
Florida	9,746,324	81.6735	27.7948
Georgia	5,463,105	83.8100	33.1866
Hawaii	964,691	157.6129	21.2009
Idaho	943,935	114.9358	44.2072
Illinois	11,426,518	88.4070	41.2073
Indiana	5,490,224	86.2835	40.1759
Iowa	2,913,808	93.0582	41.9858
Kansas	2,363,679	96.6379	38.4544
Kentucky	3,660,777	85.2228	37.7918
Louisiana	4,205,900	91.4656	30.7177
Maine	1,124,660	69.6408	44.4125
Maryland	4,216,975	76.7904	39.1598
Massachusetts	5,737,037	71.3844	42.2792
Michigan	9,262,078	84.1083	42.8410
Minnesota	4,075,970	93.6489	45.2543
Mississippi	2,520,638	89.6224	32.5778
Missouri	4,916,686	92.0799	38.4815
Montana	786,690	110.9159	46.8610
Nebraska	1,569,825	97.5697	41.1991
Nevada	800,493	116.7563	37.5535

New Hampshire	920,610	71.4735	43.1783
New Jersey	7,364,823	74.4172	40.4640
New Mexico	1,302,894	106.2391	34.6202
New York	17,558,072	74.7181	41.5458
North Carolina	5,881,766	79.6756	35.5676
North Dakota	652,717	99.5101	47.4277
Ohio	10,797,630	82.7006	40.5199
Oklahoma	3,025,290	96.8876	35.5880
Oregon	2,633,105	122.5648	44.6942
Pennsylvania	11,863,895	77.2024	40.4699
Rhode Island	947,154	71.4419	41.7595
South Carolina	3,121,820	81.0355	34.0472
South Dakota	690,768	99.0563	44.1116
Tennessee	4,591,120	86.4217	35.7793
Texas	14,229,191	97.4571	30.9925
Utah	1,461,037	111.8261	40.5165
Vermont	511,456	72.8055	44.0566
Virginia	5,346,818	78.0021	37.6381
Washington	4,132,156	121.5325	47.3363
West Virginia	1,949,644	80.9407	38.7202
Wisconsin	4,705,767	88.9756	43.7192
Wyoming	469,557	106.9348	42.6568

9. Appendix B

All the computations reported here were done on an Apple IIgs computer using the spread sheet in AppleWorks 3.0. Computation time for the problems varied. When calculating the center of population using the proposed method, each iteration took about 85 seconds. Computations using a more detailed list of populated places would take longer in direct proportion to the number of places used. Computers and software with enormously greater speed and capability are widely available. There is no computational reason for not using the proposed method.

10. References

Abler, Ronald, J. S. Adams and P. Gould. 1971. *Spatial Organization: The Geographer's View of the World.* Englewood Cliffs NJ, Prentice-Hall. p. 59.

Court, A. 1964. The elusive point of minimum travel. *Annals of the Association of American Geographers* 54: 400-403.

Eells, W.C. 1930. A mistaken conception of the center of population. *Journal of the American Statistical Association*, New Series No. 169 25:33-40.

Haggett, Peter, A. D. Cliff and A. Frey. 1977. *Locational Analysis in Human Geography: Vol. II, Locational Methods, 2nd Ed.* New York NY, Wiley, p. 312.

Hayford, John F. 1902. What is the center of an area, or the center of a population? *Journal of the American Statistical Association*, New Series No. 58, 8:47-58.

Kirkpatrick, S., C. D. Gelatt, Jr. and M. P. Vecchi. 1983. Optimization by simulated annealing. *Science* 220:671-680.

Snyder, J. P. 1987. *Map Projections — A Working Manual. (USGS Professional Paper 1395).* Washington DC, U. S. Government Printing Office. pp. 41, 58, 77, 91, 157, 185, 195, 247.

Sviatlovsky, E. E. and W. C. Eells. 1937. The centrographical method and regional analysis. *The Geographical Review*, 27:240-254.

U. S. Bureau of the Census. 1983. *1980 Census of Population, Vol. I*, Chapter A, Part 1 [PC80-1-A1]. Washington DC, U. S. Dept. of Commerce, Bureau of the Census. Appendix A, p. A-5 and Table 8, pp. 1-43.

U. S. Bureau of the Census. 1973. *1970 Census of Population and Housing: Procedural History* [PHC(R)-1]. Washington DC, U. S. Dept. of Commerce, Bureau of the Census. Appendix B (Computation of the 1970 U. S. center of population), pp. 3-50.

U. S. Bureau of the Census. 1914. *Statistical Atlas of the United States.* Washington DC, Government Printing Office. pp. 29-32.

U. S. Bureau of the Census. 1913. *Thirteenth Census of the United States taken in the year 1910. Vol. I: Population, 1910.* Washington DC, Government Printing Office. pp. 30, 46.

U. S. Geological Survey. 1970. *The National Atlas of the United States of America.* Washington DC, U. S. Dept. of the Interior, Geological Survey.

Walker, F. A. 1874. The progress of the nation. In *Statistical Atlas of the United States Based on the Results of the Ninth Census, 1870, with ...* . Francis A. Walker, compiler. New York NY, J. Bien, Lithographer. List of Memoirs and Discussions, Part II, pp. 5, 6.

Warntz, W. and D. Neft. 1960. Contributions to a Statistical Methodology for Areal Distributions. *Journal of Regional Science*, 2:47-66.

NOTE: The original article contains several printing errors which might cause misunderstanding. These were corrected in the reprints of the article. The *Solstice* copy was prepared from the corrected reprint. The errors in question in the original include: i. the longitudes of places in Examples IV, V and VI were incorrect and ii. near the end of part 6 of the text a date of 1790 was incorrectly given as 1970. Also, the location, about 14 km southwest by south of Carrollton, was incorrectly given as about 7 km southeast of Carrollton.

4. THE PELT OF THE EARTH: AN ESSAY ON REACTIVE DIFFUSION

Sandra L. Arlinghaus, John D. Nystuen
Founding Director, Institute of Mathematical Geography
Ann Arbor, MI
Professor of Geography and Urban Planning
The University of Michigan

Reactive diffusion (see the many references to Murray) is an idea that draws on the concept that boundary shape can influence the spatial pattern of the developing forms and processes interior to that boundary. A second idea is involved. Once a natural diffusive process has been at work, there is reaction to it, altering the shape of the underlying diffusion. Reactive diffusion is thus a dynamic process that is, to some extent, self-adjusting to change.

This sort of idea is one that has met with many expressions in the past — in the biological as well as in the geographical landscape (Arlinghaus, Nystuen, and Woldenberg, 1992). Boundary shape can determine how matter and energy travel within a closed system. Standing waves can be created in this manner, be they standing waves of translation of pigment on animal coats, producing striped animals; or standing waves of oscillation of water, producing seiches as water-stripes in reaction to lake depth and coastline shape of the containing vessel (e.g., Lake Michigan or Lake Geneva). One might even be tempted to speculate on a possible role for seiche-like stripes in the "parting of the Red Sea."

Xu, Vest, and Murray (1983) created mock animal outlines on laminar plates shaped like two-dimensional pelts — as "maps" of three-dimensional animals; when small adjustments in the outlines were made, vibrational patterns formed in a surface dust placed within these outlines created various spotted and striped patterns as a reaction to the boundary shape. Indeed, a circular drum head boundary offers one way for the roll of the drum wave of noise to interact with the boundary; using a fractal boundary for the drum head can produce a vastly different pattern of resonating pockets of drum roll (Science, 1991). The continuing work of Batty and Longley (1985 and later) in using fractal concepts to track the pattern of the urban fringe might also be (but has not yet been, to our knowledge) cast in the framework of reactive diffusion. Three-dimensional solids, covered with a coat of spots, might also have their spot patterns determined by some underlying vibrational process that causes the substance of the spots on the surface to react with the three-dimensional volume over which the surface is stretched. Thus, the calico cat and the earth might have a great deal in common when land masses, driven by tectonic rather than by biological rhythms, are seen as the calico spots on the pelt of the earth.

For example, the burn pattern created by random lightning strikes in a forest, and the reaction of firefighters to these strikes, displays a clear case of reactive diffusion and pattern formation on the earth. For, in the absence of firefighters, the random strikes start fires which coalesce to form an advancing front that may ultimately burn the entire region. When firefighters enter the scene, they work to confine the random strikes; the fire may leap the barriers they create, and when it does, the firefighters talk about it and react by moving to control the new hotspot. Ultimately, the spread of communication among firefighters, in response to the leapfrogging character of forest fires, produces a forest spotted with burnt dark patches. The reaction of firefighters to the diffusion of information about the location of fires produces characteristic, and predictable, patterns on the earth.

1. Pattern Formation: Global Views

Nystuen noted (1966) that "spatial processes depend upon the shape of the partitions created by their boundary patterns. If the boundary shape is changed the process itself is changed. In fact, the very existence of the process may depend on the boundary shape." The biologist Joseph Birdsell noted that coastline shape has affected genetic diversity in the Australian aborigine population (1950); migration patterns forced away from the concave-up portions of the northern coastline were dispersed, while those forced away from the concave-down portions were focused. With dispersal of hunting and gathering came genetic complexity; with focusing came genetic inbreeding. Arlinghaus (1977; 1986) drew on ideas from Birdsell and Nystuen in using boundary shape of a limited access arterial to suggest where new pockets of population concentration and dispersion will appear relative to the concavity of the arterial. In all of these, as with reactive diffusion, there is an adjustment of process (geographical or biological) to boundary, with implications for the spatial organization of associated human activity coming as a reaction to that adjustment.

Indeed, even in mediæval guilds, retail services clustered in pockets across the geographical landscape, as "stripes" or "spots" of commercial activity, in reaction to the diffusion of information as to type of service available (Vance, 1980). Similar urban patterns are evident in modern developing countries; and this context thus suggests, very generally, an interesting human dimension in exploring global urban change (Drake, 1993; Meadows *et al.* 1992).

In a classical urban context, one might imagine Harris and Ullman's "multiple nuclei" model recast within the replicable theoretical framework of reactive diffusion. As diffusion causes change surrounding and within the nuclei, there is a reaction, and the nuclei shift, or new nuclei spring up. The spatial evidence of reactive diffusion might be substituted for the historical evidence on which Harris and Ullman (1945) based their observational model, pulling the multiple nuclei model more in line with the earlier spatial models of Burgess and Hoyt. The multiple nuclei pattern appears as a reaction to incompatible land uses; it arises from an alternative resistance to residential and commercial land uses in which further employment centers leapfrog over existing urban neighborhoods, leading to extensive additional urban growth.

Within the Detroit metropolitan region, for example, the complex changing nature of the local political scene coupled with the increasing crime rates associated with downtown Detroit, often encapsulated quite simply in the minds of many Detroiters by the closing of the downtown Detroit Hudson's store, led to the consequent reaction of many businesses to move to the suburbs. Thus, suburban Southfield became an early hub of urban reaction in the Detroit metropolitan region — here, a new nucleus emerged. Efforts to restore the prominence of the downtown on the Detroit River are typified by the Renaissance Center — here, the old nucleus shifts toward the River banks.

Indeed, the characterization of the collapse of the central city in terms of the failure of the downtown headquarters of Hudson's department store may not be a strictly simple-minded view. Like the stars, the life of the city may take different paths — at one time a center may be a viable unit, and at another time the relative size and density of the urban area may cause inner city collapse. In a central place context, in which the "threshold" of a firm refers to the minimum number of sales which allows the firm to succeed and give an adequate return to its owners, the situation with Hudson's was simply a matter that the buying population at the center was too small to meet the threshold number. Related central place terminology involves the notions of the maximum range of a good and the minimum range of a good. The maximum range is the absolute limit on the demand of a good — beyond this limit, transportation costs reduce demand for the good to zero. The minimum range is the distance over which the firm must ship its goods to include the threshold populations. A logical consequence, all else being equal, is that the minimum range of a good is less in a densely settled region than it is in a sparsely settled one.

Thus, the common sense notion of "how can a big store like Hudson's fail in downtown Detroit?" can be translated as follows. Migration of the affluent population to the suburbs reduced the number of potential customers in the center. The minimum range therefore needed to be extended outward from downtown in order to include the threshold number of customers. But, suburban Hudson stores were already in place and also worked to attract those customers that the downtown branch now required to succeed. The three large suburban stores competed with the downtown store for these customers, won them over, and the downtown store failed. Stability in competition (Hotelling, 1929) was restored when the "empty" center was divided among the peripheral competitors, in a sort of central place (two-dimensional) Hotelling model. This sort of geometric view is a minimalist approach — a best-case scenario; when additional social (and other) issues are superimposed, acceleration along the path to collapse is more likely. When one next considers that this pattern will repeat on the periphery of these suburban stores and within the maximum ranges of the various goods, a sort of leapfrogging of circular/hexagonal trade areas occurs and suggests, once again, a conceptual context of reactive diffusion as an alternative, and addition to traditional spatial analysis.

Unlike earlier models of urban ecologists (Burgess, 1925; Hoyt, 1939), this sort of urban view of the world is not a generalization of a particular example — that is why it is important to see reactive diffusion cast in the geographical as well as the biological (or other) realms. The pattern of clusters of urban activity on a regional part of the earth's surface is one that is produced in reaction to the diffusion of urban process.

2. Pattern Formation: Local Views

Some current urban research strives to develop indices that offer an easy means for replication of experiments and that are sensitive to the role of boundary. Thus, Morrill (1991) proposed an index of segregation, modified by boundary considerations, to quantify urban spatial segregation. Wong (1992) modifies Morrill's indices by arguing that the length of the boundary separating adjacent urban areal units, as well as the shape of these adjacent units, is significant in determining segregation. Indices such as these, that already are sensitive to some boundary considerations, may offer one means to tighten the focus of application of the concept of reactive diffusion in various specific urban situations.

Often reactions to incompatible urban land uses are circumscribed by the boundary of the system of local jurisprudence. When these reactions fit reasonably well within the laws, competing commercial and residential land uses are in relative harmony. Laws, such as the apocryphal "it is illegal to tie an alligator to a parking meter" suggest a reaction to an unusual situation. When that reaction is passed as law, it diffuses to the population of the surrounding area and may disturb the sensitive balance between incompatible land uses.

Perhaps the most difficult situation of this sort is in establishing rules (legal, ethical, or otherwise) to position locally unwanted land uses ("lulus"). Human laws permit or forbid institutional boundaries that can influence how process works. Typically, a lulu, such as an adult bookstore or a toxic waste site, causes a strong local reaction around this "hotspot." This reaction is confined and suppressed by municipal authorities using the local legal system as their "hose" or "barrier" to confine the effects of the unwanted activity. As with the forest fire example, the lulu leapfrogs, and yet another hotspot of locally unwanted activity occurs.

Reactive diffusion offers an attractive conceptual context in which to examine pattern formation on the pelt of the earth: from local scenarios that mimic the forest fire example to global scenarios that examine entire closed and bounded surfaces. Beyond this essay, the next step is to use this context in specific urban or physical settings.

3. References Cited

Arlinghaus, S. L. 1986. "Concavity and human settlement patterns," *Essays on Mathematical Geography*, Monograph #3. Ann Arbor, MI: Institute of Mathematical Geography; 1977.

Arlinghaus, S. L., Nystuen, J. D., and Woldenberg, M. J. 1992. "An application of graphical analysis to semidesert soils." *Geographical Review*, American Geographical Society. July, pp. 244-252.

Batty, M. and Longley, P. 1985. "The fractal simulation of urban sructure." *Papers in Planning Research* 92, Univ. of Wales Institute of Science and Technology, Colum Drive, Cardiff, CF1 3EU.

Birdsell, J. B. 1950. "Some implications of the genetical concept of race in terms of spatial analysis," *Symposia on Quantitative Biology*, Vol. 15, Origin and Evolution of Man. Long Island, New York: The Biological Laboratory, Cold Springs Harbor.

Drake, W. D. forthcoming, 1993. Towards building a theory of population - environment dynamics: a family of transitions. In *Population - Environment Dynamics*. Ann Arbor: University of Michigan Press.

Hotelling, H. 1929. "Stability in competition," *Economic Journal*, 39: 41-57.

Hoyt, H. W. 1939. According to Hoyt (1966), Washington D. C.: Homer Hoyt Associates.

Harris, C. D. and Ullman, E. L. 1945. "The nature of cities," *Annals of the American Academy of Political and Social Science*, CCXLII, Nov. 1945, pp. 7-17.

Meadows, D. H., Meadows, D. L., and Randers, J. 1992. *Beyond the Limits*. Post Mills, VT: Chelsea Green Publishing Company.

Morrill, R. L. 1991. "On the measure of geographic segregation." *Geography Research Forum* 11, 25-36.

Nystuen, J. D. 1966. "Effects of boundary shape and the concept of local convexity." Discussion Paper 10. Michigan Inter-University Community of Mathematical Geographers (John D. Nystuen, ed.), Ann Arbor, MI. (Reprinted by the Institute of Mathematical Geography, 1986).

Park, R. E. and Burgess, E. W. 1925. *The City*. Chicago: University of Chicago Press.

Vance, J. E. 1977. *This Scene of Man: The Role and Structure of the City in the Geography of Western Civilization* . New York: Harper's College Press.

Wong, D. W. S. 1991. "Spatial indices of segregation." Preliminary version, National Meetings, Regional Science Association, 1991. Forthcoming in *Urban Geography*, 1992.

Xu, Youren; Vest, Charles M.; Murray, James D. 1983. "Holographic interferometry used to demonstrate a theory of pattern formation in animal coats." *Applied Optics* 15 Nov., Vol. 22, No. 22, pp. 3479-3483.

4. Literature of Apparent Related Interest

Bard, Jonathan B. L. 1977. "A unity underlying the different zebra striping patterns." *Journal of Zoology*, Vol. 183, part 4, pp. 527-539.

Boal, F. W. 1972. "Close together and far apart: Religious and class divisions in Belfast." *Community Forum*, Vol. 3, No. 2, pp. 3-11.

Boal, F. W. and Livingstone 1986. "Protestants in Belfast: A view from the inside." *Contemporary Review*, 248: 169-75.

Boyce, R. R., and W. A. V. Clark. 1964. "The concept of shape in geography." *Geographical Review* 54, 561-572.

Dewdney, A. K. "A home computer laboratory in which balls become gases, liquids and critical masses." Computer Recreations, *Scientific American* pp. 114-117.

Freedman, David H. 1991. "A chaotic cat takes a swipe at quantum mechanics." *Science*, Vol. 253, p. 626.

Gierer, A. 1981. "Some physical, mathematical and evolutionary aspects of biological pattern formation." *Philosophical Transactions Royal Society, London*, Series B, 295, pp. 429-440.

Hagerstrand, T. 1967. *Innovation diffusion as a spatial process.* Chicago: University of Chicago Press.

Kennedy, S., and W. Tobler. 1983. "Geographic interpolation." *Geographical Analysis* 15, 151-156.

Murray, J. D. 1981. "A pre-pattern formation mechanism for animal coat markings." *Journal of Theoretical Biology*. Vol. 88, No. 1, pp. 161-199.

Murray, J. D. 1988. How the leopard gets its spots. *Scientific American* 258:80-87.

Murray, J. D. 1981. "Introductory remarks" (to an entire volume devoted to pattern formation) *Philosophical Transactions Royal Society, London*, Series B 295, pp. 427-428.

Murray, J. D. 1989. Mathematical Biology. Springer-Verlag, Heidelberg.

Murray, J. D. 1981. "On pattern formation mechanisms for lepidopteran wing patterns and mammalian coat patterns." *Philosophical Transactions of the Royal Society, London* Series B, Vol. 295, No. 1078, pp. 473-496; Oct. 7.

Murray, J. D. and P. K. Maini. 1986. "A new approach to the generation of pattern and form in embryology." *Science Progress*, Vol. 70, No. 280, part 4, 539-553.

Nordbeck, S. 1965. "The Law of Allometric Growth." Discussion Paper #7. Michigan Inter-University Community of Mathematical Geographers (John D. Nystuen, ed.), Ann Arbor, MI. (Reprinted by Institute of Mathematical Geography, 1986).

Pool, R. 1991. Did Turing discover how the leopard got its spots? *Science* 251:627.

Shaw, L. J. and Murray, J. D. 1990. "Analysis of a model for complex skin patterns." SIAM Journal of Applied Mathematics, pp. 628-648.

Tobler, W. R. 1969. "The spectrum of U. S. 40," *Papers of the Regional Science Association*, Vol. XXIII, pp. 45-52.

Turing, A. M. 1952. "The chemical basis of morphogenesis." *Philosophical Transactions of the Royal Society, London*, Series B, 237, pp. 37-72.

White, M. J. 1983. "The measurement of spatial segregation." *American Journal of Sociology* 88, 1008-1018.

Wolpert, L. 1981. "Positional information and pattern formation." *Philosophical Transactions Royal Society of London*, Series B 295, pp. 441-450. 1981.

5. FEATURE

Meet New Solstice Board Member
William D. Drake
The University of Michigan

Professor of Resource Policy and Planning
School of Natural Resources and the Environment

Professor of Population Planning and International Health
School of Public Health, The University of Michigan

Professor of Urban, Technological and Environmental Planning
College of Architecture and Urban Planning

Bill Drake teaches courses on the Global Environment and on Population-Environment Dynamics. Much of his research portfolio is drawn from ongoing projects in the developing world. Many of these projects have been underway for several years and have focused on the problems of rural community development particularly relating to reducing child malnutrition.

Recently, he has authored articles and was co-editor for a book on population - environment dynamics. During the fall of 1992, Drake and S. Arlinghaus offered a course on the same subject which has resulted in a monograph. The focus of this course is captured in its name *Population-Environment Dynamics: Toward Building a Theory*. The effort draws upon recent work carried out as part of the University of Michigan's Population - Environment Dynamics Project. Ten graduate students and two faculty participated formally, and several other students and faculty sat in from time-to-time, with one visitor attending every session. Seminar participants came from many disciplinary backgrounds ranging from population planning, economics, engineering, biology, remote sensing, geography, natural resources, sociology, international health, business administration to mathematics. In addition to U.S. students, the course was enriched by colleagues from Mexico, Nepal, Taiwan, and Nigeria.

The monograph serves as a kind of a "time capsule"–what do students in 1992 think will be issues of great significance in the current, recently identified, need to study "global change"? Here are the sectors of that "capsule":

Dawn M. Anderson
The Historical Transition of Forest Stock Depletion in Costa Rica

Katharine A. Duderstadt
The Energy Sector of Population-Environment Dynamics in China

Eugene A. Fosnight
Population Transition and Changing Land Cover and Land Use in Senegal

Katharine Hornbarger
The Energy Crisis in India: Options for a Sound Environment

Deepak Khatry
An Analysis of the Major Sectoral Transitions in Nepal's Middle Hills
and their Relationship with Forest Degradation

Catherine MacFarlane
The Interrelationship Between the Forestry Sector

6. SAMPLE OF HOW TO DOWNLOAD THE ELECTRONIC FILE

This section shows the exact set of commands that work to download *Solstice* on The University of Michigan's Xerox 9700. Because different universities will have different installations of TeX, this is only a rough guideline which *might* be of use to the reader.

First step is to concatenate the files you received via bitnet/internet. Simply piece them together in your computer, one after another, in the order in which they are numbered, starting with the number, "1."

The files you have received are ASCII files; the concatenated file is used to form the .tex file from which the .dvi file (device independent) file is formed. The words "percent-sign" and "backslash" are written out in the example below; the user should type them symbolically.

ASSUME YOU HAVE SIGNED ON AND ARE AT THE SYSTEM PROMPT, #.

create -t.tex

percent-sign t from pc c:backslash words backslash solstice.tex to mts -t.tex char notab

[this command sends my file, solstice.tex, which I did as a WordStar (subdirectory, "words") ASCII file to the mainframe]

run *tex par=-t.tex

[there may be some underfull boxes that generally cause no problem; there should be no other "error" messages in the typesetting–the files you receive were already tested.]

run *dvixer par=-t.dvi

control *print* onesided

run *pagepr scards=-t.xer, par=paper=plain