NCGIA

National Center for Geographic Information and Analysis

Spatial Analysis on the Sphere:

A Review

by

Robert G. Raskin, Visiting Fellow
NCGIA
University of California, Santa Barbara

Technical Report 94-7

October 1994

Simonett Center for Spatial Analysis
University of California
35 10 Phelps Hall
Santa Barbara, CA 93106-4060
Office (805) 893-8224
Fax (805) 893-8617
ncgia@ncgia.ucsb.edu

State University of New York
301 Wilkeson Quad, Box 610023
Buffalo NY 14261-0001
Office (716) 645-2545
Fax (716) 645-5957
ncgia@ubvms.cc.buffalo.edu

University of Maine
348 Boardman Hall
Orono ME 04469-5711
Office (207) 581-2149
Fax (207) 581-2206
ncgia@spatial.maine.edu

ABSTRACT

Methods of analysis on a spherical earth are reviewed. The techniques are relevant to both point data and data from field variables (functions that are continuous and everywheredefined on the sphere). The background mathematics and statistics applicable to spherical surfaces are presented. The probability density function (pdf) for nearest neighbor distance on a sphere generated by a Poisson process is derived. The use of spherical analysis in conjunction with global geographic information systems is detailed. This review is relevant to researchers in geography, global change and related fields who study processes at global scales.

TABLE OF CONTENTS

1. Why use spherical analysis?

Over 100 years have past since the original publication of Flatland (Abbot, 1884), a description of a culture cognizant of only two-dimensions, and their inabilities to grasp their missing third dimension. The relatively recent sequel Sphereland (Burger, 1965) describes the Flatlanders' attempts to grasp the spherical nature of their planet. That geographical analysis is typically performed on a two-dimensional plane is a testimonial to the success of map projections in allowing us to neglect the curvature of the earth and work with data that appears to originate on a plane, much as a Flatlander would do. This success is due in part to the limited geographical domains that characterize most research pursuits.

This situation is likely to change, as there is currently great interest in global-scale phenomena, for which the curvature of the earth cannot be neglected in analytical formulations. In a sense, analysis at the global level represents the "final frontier" for geographers, as little analysis of this type has been carried out to date. Many analytic techniques are available for this task, but their descriptions are widely dispersed in the literature of many fields: geology, astronomy, meteorology, statistics, operations research, and computer science, among others.

This report provides a review of spatial analysis techniques in geography and earth systems science that take into account the spherical geometry of the earth. It is oriented to both human and physical geographers and should be of particular interest to global change researchers who analyze global-scale databases. The primary objectives are to describe the intrinsic characteristics of spherical data and the differences between planar and spherical analysis techniques. Working examples of these techniques in geography and other fields are summarized and evaluated, and the underlying mathematics clarified.

Spatial analysis can be defined as the formal analysis of spatial patterns and orientations, with tools that include: point pattern analysis, interpolation, spectral analysis, search techniques, location analysis, optimization, estimation, and forecasting. Examples of very large-scale applications of spatial analysis include:

1. Find the cost minimizing location for a global distribution center.
2. Find the area of lowest concentration of some element, e.g. ozone.
3. Find the mean international migration paths during a specified time period.
4. Find the cross-correlation of two variables defined on the sphere.
5. Determine whether an observed change in a global data set is statistically significant.
6. Find the optimal satellite path given a set of constraints.
7. Find the center of population of the former Soviet Union
8. Identify the continental drift over a designated time epoch.
9. Interpolate precipitation data over a region with sparsely situated weather stations

Most geographical analysis tools have been derived for use on a two-dimensional plane. Although the surface of the earth is two-dimensional to a first approximation, conventional analysis techniques designed for the plane may not be applicable, as a result of the convergence of the meridians at the poles. Three conceptual modeling approaches are used to apply and adapt spatial analysis techniques to the sphere (Jupp and Mardia, 1989):

1. Projection: The sphere is projected onto a plane using a conventional map projection. Planar analysis is performed on the transformed data in the two-dimensional space R^2, and the results mapped back onto the sphere.
2. Intrinsic: The sphere surface is considered an intrinsic space in its own right, with analysis performed in the non-Euclidean space S^2.
3. Embedding: The sphere is considered a constrained subset of the three-dimensional space R^3. Three-dimensional spatial analysis is performed, with a constraint imposed to limit solutions to the sphere surface.

The projection approach is used implicitly in conventional studies that ignore the curvature of the earth. Wilmott et al. (1985) and Turner (1986) demonstrate some of the difficulties met in performing analysis on projected data, as no projection can simultaneously preserve both distances and areas. Nevertheless, certain planar properties involving distance can be adapted to the sphere using an equidistant mapping of the plane to the sphere, a strategy referred to as *wrapping* (Jupp and Mardia, 1989). The second approach deals directly with the spherical domain, but requires a considerable amount of trigonometric calculations. Such calculations are more computationally intensive than addition or multiplication, a factor that should be taken into account in large real-time applications. In the embedding approach, there is a larger storage overhead associated with the extra (third) dimension, but computational time is reduced relative to the intrinsic approach. The embedding approach will be used extensively in this review.

The amount of existing literature that is relevant to the present subject is quite varied. Several textbooks and review articles are devoted entirely to *statistics* on the sphere (Watson, 1983; Fisher et al., 1987, Upton and Fingleton 1989, Chapter 10; Jupp and Mardia, 1989). Statistics of *functions* on the sphere is the topic of the review article by Kaula (1967). Non-statistical aspects of spherical analysis are much more scarce, with few applications specific to the earth surface as the sphere. A monograph describing geographic information systems (GIS) for global-scale data (Mounsey, 1988) does not address analysis in great detail, although brief reviews of spherical analysis in geography are found in Goodchild (1988) and Tobler (1993). Thus, the reader will find numerous opportunities to apply and extend the material presented here.

The organization of this review is as follows. Section 2 summarizes the pertinent mathematics of spherical surfaces. The treatment includes material from spherical geometry, trigonometry, calculus, and topology. Virtually all the underlying mathematics for the techniques discussed later are included in this section. Section 3 introduces spherical probability distributions that have been found to be useful.

Our subject spans two types of spherical analysis, corresponding roughly to analysis of point data and "field" data. Examples of point data include locations of international airports, weather radar towers, and large-scale supply centers. Point processes on a two-dimensional plane have been well studied and are the subject of several books (Getis and Boots, 1978; Okabe et al., 1992). However, study of *spherical* point data has been largely confined to the statistical literature, and appears under the names "directional," "orientation," or "vector" statistics (Fisher, 1987; Watson, 1983). This terminology arises because a point on the sphere can be associated with a unit vector originating at the sphere center. For our purposes, a data point is a geographic location on the earth surface, without any associated "direction." Point data analysis is the subject of Section 4.

Analysis of functions on the sphere is the subject of Section 5. The primary emphasis is upon analysis of fields, or functions that are continuous and everywhere defined on the sphere surface; examples include elevation and temperature. Section 6 concludes with a discussion of relevant issues that are likely to be important in the coming years, including potential applications to global change.

The emphasis of this review is on the differences between planar and spherical analysis. It is specific to analysis on the surface of a spherical earth, and thus, the ellipsoidal nature of the earth and its vertical dimension are not considered.

2. Spherical mathematics

This section contains the underlying mathematics that is used in the remainder of the review. Included are sections on spherical geometry, trigonometry, and calculus. Also included are sections on tessellations of the sphere and the spherical harmonic transform.

2.1 Points

A point on the surface of a sphere can be referenced using either spherical or Cartesian

coordinates. In spherical coordinates, a point is represented as (φ, λ), where φ is latitude and λ is longitude, each in radians. The sphere consists of the set of points:

$$S^2 = \{ (\varphi, \lambda) \in \Re^2 \mid -\pi/2 \le \varphi \le \pi/2, \quad 0 \le \lambda \le 2\pi \} \quad . \tag{2.1.1}$$

In Cartesian coordinates, nondimensional units are usually used, known as the direction cosines. In this form, all points at the surface are unit vectors relative to the sphere center:

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = (x,y,z)' \tag{2.1.2}$$

where the symbol $'$ denotes transpose, with

$$S^2 = \{ \mathbf{x} \in \Re^3 \mid |\mathbf{x}| = 1 \} \quad . \tag{2.1.3}$$

Spherical coordinates and direction cosines are related through:

$$x = \cos\varphi \cos\lambda, \qquad y = \cos\varphi \sin\lambda, \qquad z = \sin\varphi \tag{2.1.4a}$$

$$\varphi = \sin^{-1} z, \qquad \lambda = \tan^{-1}(y/x) \quad . \tag{2.1.4b}$$

The Jacobian of the transformation (2.1.4a) is

$$J(\varphi, \lambda) = \cos\varphi \quad . \tag{2.1.4c}$$

The direction cosines $\mathbf{x}$ are converted to dimensional units $\mathbf{x}^*$ through:

$$\mathbf{x}^* = r_S \mathbf{x} \tag{2.1.5}$$

where r_S is the radius of the sphere, assumed to be a fixed constant (typically, $r_S = 6371$ km).

Two points $\mathbf{x}_1$ and $\mathbf{x}_2$ are *antipodal* if $\mathbf{x}_1 = -\mathbf{x}_2$. Antipodal points lie on either end of an axis directed through the center of the sphere. Two points are *orthogonal* if

$$\mathbf{x_1} \cdot \mathbf{x_2} = 0 \tag{2.1.6}$$

where

$$\begin{aligned} \mathbf{x_1} \cdot \mathbf{x_2} &= (x_1x_2 + y_1y_2 + z_1z_2) \\ &= \cos\theta \end{aligned} \tag{2.1.7}$$

where θ is the angle between $\mathbf{x_1}$ and $\mathbf{x_2}$.

The spherical midpoint of a pair of points $\mathbf{x_1}$ and $\mathbf{x_2}$ is the point:

$$\mathbf{x}_{\text{mid}} = \frac{\mathbf{x}_1 + \mathbf{x}_2}{|\mathbf{x}_1 + \mathbf{x}_2|} \tag{2.1.8}$$

where for an arbitrary vector $\mathbf{y}$,

$$|\mathbf{y}| = (\mathbf{y} \cdot \mathbf{y})^{1/2} \quad . \tag{2.1.9}$$

The midpoint in (2.1.8) is undefined if $\mathbf{x_1}$ and $\mathbf{x_2}$ are antipodal.

2.1.1 Rotations

The coordinates of a point relative to a rotated pole can be obtained by multiplication with a *rotation matrix* **R**:

$$\mathbf{x''} = \mathbf{Rx} \quad . \tag{2.1.10}$$

A rotation matrix is a special case of an *orthogonal* matrix, or a matrix that satisfies

$$\mathbf{A'A} = \mathbf{I} \tag{2.1.11}$$

and det $\mathbf{A} = \pm 1$. A rotation matrix corresponds to the case:

$$\det \mathbf{R} = 1, \tag{2.1.12}$$

while a *reflection matrix* corresponds to det $\mathbf{A} = -1$. A reflection matrix reverses the orientation of a set of points about a plane in addition to performing a rotation, but has little physical relevance.

To place the point $\mathbf{x}_0 = (x_0, y_0, z_0)'$ with spherical coordinates (φ_0, λ_0) at the north pole, the rotation matrix has the form:

$$\mathbf{R} = \begin{bmatrix} \cos\lambda_0 \sin\varphi_0 & \sin\lambda_0 \sin\varphi_0 & -\cos\varphi_0 \\ -\sin\lambda_0 & \cos\lambda_0 & 0 \\ \cos\lambda_0 \cos\varphi_0 & \sin\lambda_0 \cos\varphi_0 & \sin\varphi_0 \end{bmatrix} \tag{2.1.13a}$$

$$= \begin{bmatrix} x_0 z_0 / w_0 & y_0 z_0 / w_0 & -w_0 \\ -y_0 / w_0 & x_0 / w_0 & 0 \\ x_0 & y_0 & z_0 \end{bmatrix} \tag{2.1.13b}$$

where

$$w_0 = (1 - z_0^2)^{1/2} \quad . \tag{2.1.14}$$

A more general form of (2.1.13a) can be used to align a particular meridian ψ_0 with the prime meridian in the rotated coordinate system:

$$\mathbf{R} = \begin{bmatrix} \cos\lambda_0 \sin\varphi_0 \cos\psi_0 - \sin\lambda_0 \sin\psi_0 & \sin\lambda_0 \sin\varphi_0 \cos\psi_0 + \cos\lambda_0 \sin\psi_0 & -\cos\varphi_0 \cos\psi_0 \\ -\cos\lambda_0 \sin\varphi_0 \sin\psi_0 - \sin\lambda_0 \cos\psi_0 & -\sin\lambda_0 \sin\varphi_0 \sin\psi_0 + \cos\lambda_0 \cos\psi_0 & \cos\varphi_0 \sin\psi_0 \\ \cos\lambda_0 \cos\varphi_0 & \sin\lambda_0 \cos\varphi_0 & \sin\varphi_0 \end{bmatrix} \tag{2.1.15}$$

In general, it may be difficult to identify ψ_0 a priori, as it is specific to the rotated coordinates.

However, $\psi_0+\pi$ represents the longitude of the old north pole in the new coordinates; thus ψ_0 can be selected on this basis.

2.2 Distance

The shortest distance between two points $\mathbf{x}_1$ and $\mathbf{x}_2$ is known as the great circle (or arc or geodesic) distance. This distance is denoted $s(\mathbf{x}_1, \mathbf{x}_2)$ or simply s, and is expressed in dimensional units as:

$$s = r_s\theta \tag{2.2.1a}$$

$$= r_s \cos^{-1}(\mathbf{x}_1 \bullet \mathbf{x}_2) \tag{2.2.1b}$$

$$= r_s \cos^{-1}[\sin\varphi_1 \sin\varphi_2 + \cos\varphi_1 \cos\varphi_2 \cos(\lambda_1-\lambda_2)] \tag{2.2.1c}$$

where θ is the angle between $\mathbf{x}_1$ and $\mathbf{x}_2$, and the range of $\cos^{-1}$ is taken to be $[0,\pi]$. Equation (2.2.1c) follows from (2.2.1b) using the trigonometric identity:

$$\mathbf{x}_1 \bullet \mathbf{x}_2 = \sin\varphi_1 \sin\varphi_2 + \cos\varphi_1 \cos\varphi_2 \cos(\lambda_1-\lambda_2)\ . \tag{2.2.2}$$

Comparing (2.2.1b) and (2.2.1c), distance computations are much simpler using direction cosines, as the number of trigonometric calculations is greatly reduced.

Distances on the sphere satisfy the *triangle inequality*; that is, for three arbitrary points $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$:

$$s(\mathbf{x}_1,\mathbf{x}_2) \le s(\mathbf{x}_1,\mathbf{x}_3) + s(\mathbf{x}_3,\mathbf{x}_2)\ . \tag{2.2.3}$$

Equality in (2.2.3) results only if the three points are collinear. The inequality can be used to show that the distance between two points is greater than or equal to the x-, y-, or z- distances alone:

$$s(\mathbf{x}_1,\mathbf{x}_2) \ge r_s\,|x_1-x_2| \tag{2.2.4a}$$

$$s(\mathbf{x}_1,\mathbf{x}_2) \ge r_s\,|y_1-y_2| \tag{2.2.4b}$$

$$s(\mathbf{x}_1,\mathbf{x}_2) \ge r_s\,|z_1-z_2| \tag{2.2.4c}$$

but less than or equal to the sum of the three quantities:

$$s(\mathbf{x}_1,\mathbf{x}_2) \le r_s\,(\,|x_1-x_2| + |y_1-y_2| + |z_1-z_2|\,)\ . \tag{2.2.5}$$

In spherical coordinates, a parallel of latitude is not a great circle arc, and the corresponding inequalities are:

$$s(\mathbf{p}_1,\mathbf{p}_2) \le r_s\,(\,|\Delta\varphi| + |\Delta\lambda| \cos[\,\max(|\varphi_1|, |\varphi_2|)\,]\,) \tag{2.2.6a}$$

$$s(\mathbf{p}_1,\mathbf{p}_2) \ge r_s\,|\Delta\varphi| \tag{2.2.6b}$$

$$s(\mathbf{p}_1,\mathbf{p}_2) \ge r_s\,|\Delta\lambda| \cos[\,\max(|\varphi_1|,|\varphi_2|)]\ . \tag{2.2.6c}$$

where

$$\mathbf{p}_1=(\varphi_1,\lambda_1) \quad \mathbf{p}_2=(\varphi_2,\lambda_2) \tag{2.2.6d}$$

For short distances it may be possible to obtain an acceptable approximation to (2.2.1c)

using the first few terms of the series expansions:

$$\sin\theta = \theta - \theta^3/3! + \theta^5/5! + \ldots$$
$$\cos\theta = 1 - \theta^2/2! + \theta^4/4! - \ldots \qquad (2.2.7a)$$

where

$$\theta! = \theta(\theta-1)(\theta-2)\ldots 1 \quad . \qquad (2.2.7b)$$

In some applications, the small angle approximations

$$\theta \approx \sin\theta \qquad (2.2.8)$$
$$\theta^2/2 \approx 1-\cos\theta \qquad (2.2.9)$$

are used. When reading the literature on spherical analysis, one should be alert to formulas derived using (2.2.8) or (2.2.9) to simplify calculations, and to use them with caution over large distances.

Unlike the case on the plane, geodesic distance is not a convex function. A function g is convex on a set if for all points $\mathbf{x}_1$ and $\mathbf{x}_2$ in that set,

$$g[w\mathbf{x}_1 + (1-w)\mathbf{x}_2] \leq wg(\mathbf{x}_1) + (1-w)\, g(\mathbf{x}_2) \qquad (2.2.10)$$

for all $0 \leq w \leq 1$. In words, a function is convex if its value at a linear combination of two points is less than or equal to the linear combination of functional values at the points. Nonconvexity creates difficulties for optimization schemes, because it implies that local maxima may not be global maxima. To see that spherical distance is not convex, let s represent geodetic distance from the north pole, let w= 1/2, and let $\mathbf{x}_1$ and $\mathbf{x}_2$ be a pair of points at latitude $\varphi = -\pi/4$, separated in longitude by $\Delta\lambda = \pi$. The right hand side of (2.2.10) is the distance from the pole to either of the points, that is $3\pi/4\ r_S$. The left-hand side is the pole-to-pole distance πr_S. While distance is not a convex function over the entire sphere, it is convex on a spherical cap (see Section 2.4.2) of spherical radius of $\pi/4\ r_S$ or smaller (Drezner and Wesolowsky, 1978).

2.3 Circles

2.3.1 Great circles

A great circle plays the same role on a sphere as a line on the plane. A great circle is defined as the intersection of the sphere surface with a plane passing through the sphere center. Geodesic distance always follows a segment of a great circle.

Points $\mathbf{x}$ on a great circle satisfy the general equation:

$$\mathbf{x}\bullet\mathbf{c} = 0 \qquad (2.3.1)$$

where $\mathbf{c} \in S^2$. The point $\mathbf{c}$ and its antipode $-\mathbf{c}$ define the axis or normal vectors of the circle, and are poles of a coordinate system with both $\mathbf{x}_1$ and $\mathbf{x}_2$ on the equator. The equation of the great circle containing any two distinct points $\mathbf{x}_1$ and $\mathbf{x}_2$ is:

$$\mathbf{c} = \frac{\mathbf{x}_1 \times \mathbf{x}_2}{|\mathbf{x}_1 \times \mathbf{x}_2|} \qquad (2.3.2)$$

where the *cross product* is defined as the vector:

$$\mathbf{x}_1 \times \mathbf{x}_2 = \det \begin{bmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix} \tag{2.3.3}$$

$$= (y_1z_2 - y_2z_1, z_1x_2 - z_2x_1, x_1y_2 - x_2y_1) \quad .$$

The cross product is a location interior to the sphere (unless $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal) with magnitude:

$$|\mathbf{x}_1 \times \mathbf{x}_2| = \sin\theta \quad . \tag{2.3.4}$$

The perpendicular bisector of a pair of points is the great circle that is orthogonal to both $\mathbf{x}_1 \times \mathbf{x}_2$ and $(\mathbf{x}_1 + \mathbf{x}_2)/2$. Thus,

$$\mathbf{c}_b = \pm \frac{(\mathbf{x}_1 \times \mathbf{x}_2) \times (\mathbf{x}_1 + \mathbf{x}_2)}{|(\mathbf{x}_1 \times \mathbf{x}_2) \times (\mathbf{x}_1 + \mathbf{x}_2)|} \tag{2.3.5}$$

The distance of a point $\mathbf{x}$ from the great circle with axis $\mathbf{c}$, is

$$s = r_s \cos^{-1} \frac{\mathbf{x} \times \mathbf{c}}{|\mathbf{x} \times \mathbf{c}|} \tag{2.3.6}$$

Two distinct great circles cross at exactly two points: a point and its antipode. This is different from the planar case where parallel lines exist that never cross. The angle θ between two circles is that of their respective axes $\mathbf{c}_1$ and $\mathbf{c}_2$:

$$\theta = \cos^{-1} (\mathbf{c}_1 \bullet \mathbf{c}_2) \quad . \tag{2.3.7}$$

An alternative means of representing a great circle is parametrically, relative to any two orthogonal points $\mathbf{x}_1$ and $\mathbf{x}_2$ on the circle. The great circle consists the collection of points satisfying:

$$\mathbf{x}(t) = (\cos t)\mathbf{x}_1 + (\sin t)\mathbf{x}_2 \qquad (0 \le t \le 2\pi) \tag{2.3.8}$$

For some applications, it is desirable to define a *directed* great circle. A great circle is directed when one of the two possible normal vectors $\mathbf{c}$ and $-\mathbf{c}$ is specified. This is equivalent to assigning an arrow to one of the two possible directions of traversal along the circle. In the standard right-hand coordinate system, looking in the direction of the normal vector, this arrow is directed clockwise.

A directed great circle divides the sphere into two hemispheres that can be labeled "northern" and "southern". For the directed great circle associated with $\mathbf{x}_1 \times \mathbf{x}_2$, the hemisphere of an arbitrary point $\mathbf{x}_3$ can be determined from the sign of the quantity:

$$b = \det(\mathbf{x}_1 \ \ \mathbf{x}_2 \ \ \mathbf{x}_3) = (\mathbf{x}_1 \times \mathbf{x}_2) \bullet \ \mathbf{x}_3 \tag{2.3.9}$$

For $b>0$, $\mathbf{x}_3$ lies in the northern hemisphere and for $b<0$, $\mathbf{x}_3$ lies in the southern hemisphere. For $b=0$, $\mathbf{x}_3$ lies on the equator, and the three points are collinear.

Because $\mathbf{c}$ is itself a point in S^2, there is a duality between antipodal axis points and their associated great circle. A similar duality exists between a single point and its associated directed great circle. For example, three points lie on a common great circle if and only if $b=0$ in (2.3.9); thus by duality, three great circles meet at a point (and its antipode) if and only if

$$\det(\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3) = 0. \tag{2.3.10}$$

Connecting any two nonantipodal points are exactly two great circle segments. The shorter segment is known as the *minor arc* and the longer segment is the *major arc*. An infinite number of arcs connect antipodal points, each with arc length πr_S. While distance on the sphere satisfies

$$s \le \pi r_S \tag{2.3.11}$$

an arc may extend over the entire sphere; hence arc length $\tilde{s}$ satisfies

$$\tilde{s} \le 2\pi r_S \quad . \tag{2.3.12}$$

A great circle arc can be represented parametrically as in (2.3.7) using a continuous subset of the parameter space t. In this representation, two points are separated by the distance

$$s = r_S\, |\mathbf{x}(t_2) - \mathbf{x}(t_1)| \tag{2.3.13}$$

if

$$|t_2 - t_1| < \pi \quad . \tag{2.3.14}$$

2.3.2 Small circles

Although the $\varphi=0$ parallel of latitude is a great circle, other parallels are examples of *small circles.* A small circle is defined as the intersection of the sphere surface with a plane *not* passing through the sphere center. Its general equation is:

$$\mathbf{c}\bullet\mathbf{x} = \sin\theta \tag{2.3.15}$$

where θ is the angle between any element of the circle and the north pole $\mathbf{c}$. This circle has circumference

$$\tilde{s} = 2\pi\, r_S \sin\theta\ . \tag{2.3.16}$$

Very little mention is made of small circles in texts of spherical geometry and trigonometry, however, national and state borders are often small circle segments.

Connecting three arbitrary points $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$, no two of which are antipodal, is a unique small circle. As this circle is normal to the plane containing the three points, the axis is found from

$$\mathbf{c} = \pm\frac{(\mathbf{x}_1 - \mathbf{x}_3)\times(\mathbf{x}_2 - \mathbf{x}_3)}{\left|(\mathbf{x}_1 - \mathbf{x}_3)\times(\mathbf{x}_2 - \mathbf{x}_3)\right|} \tag{2.3.17}$$

The parameter θ is obtained by inserting (2.3.17) into (2.3.15).

Just as a great circle splits a sphere into two hemispheres, a small circle divides a sphere into an "interior" region and "exterior" region. The location of a point $\mathbf{x}4$ relative to this circle can be ascertained using:

$$b = \det(\mathbf{x}_2 - \mathbf{x}_1 \;\; \mathbf{x}_3 - \mathbf{x}_1 \;\; \mathbf{x}_4 - \mathbf{x}_1) = (\mathbf{x}_2 - \mathbf{x}_1) \times (\mathbf{x}_3 - \mathbf{x}_1) \bullet (\mathbf{x}_4 - \mathbf{x}_1) \qquad (2.3.18)$$

Then, $b > 0$ if $\mathbf{x}4$ is interior, $b < 0$ if $\mathbf{x}4$ is exterior, and $b = 0$ if the four points lie on a common small circle.

2.4 Regions

2.4.1 Spherical triangles

A *spherical triangle* has sides that are great circle arcs. On a sphere, the specification of three noncolinear points uniquely define a pair of triangles, rather than a single one, because an "inside" and "outside" triangle are produced. In classical spherical trigonometry, it is assumed that all angles are less than π (180°), and thus the subject is limited to analysis of *proper* spherical triangles (corresponding to the "inside" case). This restriction to proper triangles is also used in the discussion that follows, however, *improper* spherical triangles may occur in actual GIS applications.

The area S of a spherical triangle in dimensional units is

$$S = r_s^2 \left[\sum_{i=1}^{3} \alpha_i - \pi\right] \qquad (2.4.1)$$

where α_i are the interior angles of the triangle in radians. The sum of the three angles is always greater than π (180°) and less than 3π (540°) (for a planar triangle, $\Sigma\alpha_i = \pi$). The angle $\angle \mathbf{x}_1\mathbf{x}_2\mathbf{x}_3$ of a spherical triangle can be calculated from the formula:

$$\angle \mathbf{x}_1\mathbf{x}_2\mathbf{x}_3 = \cos^{-1}\left[\frac{\mathbf{x}_2 \times \mathbf{x}_1}{|\mathbf{x}_2 \times \mathbf{x}_1|} \bullet \frac{\mathbf{x}_2 \times \mathbf{x}_3}{|\mathbf{x}_2 \times \mathbf{x}_3|}\right] . \qquad (2.4.2)$$

This angle can also be found in the spherical coordinate representation, by rotating coordinates to place $(\varphi 2, \lambda 2)$ at a pole. The angle is then equal to the difference in longitudes: $|\lambda 1 - \lambda 3|$ in the new coordinate system. A FORTRAN subroutine for this calculation is given in Bevis and Cambareri (1987).

On a sphere of fixed radius, the specification of the three angles uniquely defines the triangle up to an arbitrary rotation. Thus, there is no concept on the sphere of *similar* triangles, or triangles with identical angles but proportionally shorter or longer sides. An *equilateral* spherical triangle has sides of equal length; an example is a quarter hemisphere.

The sum of the angular lengths of the *sides* of a spherical triangle is less than 2π (360°). Any particular side is both less than π and less than the sum of the two other angles. The spherical law of sines states that sines of angles are proportional to the sines of their opposite sides. Denoting the three angles by α, β, and γ, and their opposite sides as a, b, and c, respectively:

$$\frac{\sin \alpha}{\sin a} = \frac{\sin \beta}{\sin b} = \frac{\sin \gamma}{\sin c} \qquad . \qquad (2.4.3)$$

Use of (2.4.3) to solve for an unknown side or angle produces ambiguous results because the arc

sin function is multivalued. This indeterminacy can be resolved by noting that the relative magnitude of the angles is the same as that of their respective opposite sides.

The spherical law of cosines states that the cosine of a side equals the product of the cosines of the other two sides plus the product of their sines, multiplied by the cosine of the included angle:

$$\cos c = \cos a \cos b + \sin a \sin b \cos \gamma \tag{2.4.4}$$

Using the small angle approximations (2.2.7), this relation reduces to the planar form:

$$c^2 = a^2 + b^2 - 2ab \cos \gamma \tag{2.4.5}$$

By duality, we can interchange the words *side* and *angle* in (2.4.4) to obtain an analogous formula for the cosine of any angle:

$$\cos \gamma = \cos \alpha \cos \beta + \sin \alpha \sin \beta \cos c \tag{2.4.6}$$

A triangle containing a right angle is known as a *right spherical triangle*; such a triangle may contain more than one right angle. A useful rule relating sides and angles in right spherical triangles is Napier's rules of circular elements. To apply these rules, consider a right spherical triangle with the right angle excluded, leaving the ordered sequence of five elements:

side, angle, side, angle, side .

Considering this list as a cyclical sequence, Napier's rule states that the sine of the middle element equals both: (i) the product of the tangents of its two adjacent elements and (ii) the product of the cosines of the other two (non-adjacent) elements

Rules relating sides and angles of oblique spherical triangles (triangles with no right angle) can be found in texts on spherical trigonometry (e.g. Morgan, 1951).

2.4.2 Spherical caps

The surface area contained within a pair of latitude bands is:

$$S = 2\pi r_s^2 \, |\sin \varphi_1 - \sin \varphi_2| \quad . \tag{2.4.7}$$

A special case of (2.4.7) is the *spherical cap* of spherical radius s, defined as the set of points with arc distance less than or equal to s from a specified point. Setting $\varphi_1 = \pi/2$ in (2.4.7), and switching to colatitude, a spherical cap has surface area:

$$S = 2\pi r_s^2 \, (1 - \cos (s/r_s)) \tag{2.4.8a}$$

$$= \pi s^2 \, \mathrm{sinc}^2 (s/2r_s) \qquad (0 \le s \le \pi r_s) \tag{2.4.8b}$$

where the sinc function is:

$$\mathrm{sinc}\, \theta = \sin \theta / \theta \tag{2.4.9}$$

The circumference of this cap is:

$$\tilde{s} = dS/ds = 2\pi r_s \sin (s/r_s) \tag{2.4.10a}$$

$$= 2\pi s \operatorname{sinc}(s/r_S) . \qquad (2.4.10b)$$

The sinc function approaches 1 for small θ (where $\sin\theta \approx \theta$), and a planar approximation may be justifiable in such cases. For a 1000 km arc length, neglection of this spherical correction term induces only a 0.2% error in the area calculation; however, for a 5000 km arc length (representing one-eighth the circumference of the earth), the error is 5%. The corresponding planar approximation for P induces an error that is roughly twice as large. This is because for small values of $\theta = r/r_S$, the ratio of errors is:

$$\frac{1-\operatorname{sinc}\theta}{1-\operatorname{sinc}^2(\theta/2)} \approx \frac{1-[1-\theta^2/6+O(\theta^4)]}{1-[1-\theta^2/12+O(\theta^4)]} \approx 2 \qquad (2.4.11)$$

where $O(\theta^4)$ refers to terms of the fourth or higher power of θ.

2.4.3 Other regions

A *spherical polygon* is defined as a polygon whose sides are great circle segments. It can be used to approximate any shape to any desired degree of accuracy using a sufficiently large number of sides. The area of a spherical polygon with n vertices is:

$$S = r_s^2 \left[\sum_{i=1}^{n} \alpha_i - \pi(n-2)\right] . \qquad (2.4.12)$$

Kimerling (1984) provides an algorithm for carrying out this computation by accumulating triangular areas.

A subset of the sphere is *convex* if it contains the arcs connecting all non-antipodal pairs of points in the set. Examples of convex sets on the sphere include a point, a great circle, a spherical cap of spherical radius less than or equal to $\pi r_S/2$, and the entire sphere. An example of a nonconvex set is a spherical cap of radius $3\pi r_S/4$ or extending in latitude from the north pole to $\varphi = -\pi/4$ (45° S).

The *spherical convex hull* of a set D is the intersection of all convex sets that contain D, that is, the smallest convex set containing D. A subset of the sphere that lies entirely within a hemisphere is referred to *hemispheric*. For a set of points that are hemispheric, the spherical convex hull is also hemispheric. For points not hemispheric, the spherical convex hull is the entire sphere.

2.5 Tessellation of the sphere

A *tessellation* of the sphere is a spatial decomposition that is exhaustive and mutually exclusive (except perhaps along boundaries). Tessellations can be classified as either polyhedral or empirical.

While it is desirable to subdivide the earth into regions of equal area and shape, this can be done only by projection of the sphere onto one of the polyhedral solids. There are five regular polyhedral (or Platonic) solids: the tetrahedron (4 equilateral triangles), hexahedron (cube) (6 squares), octahedron (8 triangles), dodecahedron (12 pentagons), and icosahedron (20 equilateral triangles). There are 14 semiregular (or Archimedian) solids, or polyhedra with more than one kind of regular solid as faces. For instance, the truncated icosahedron (White et al., 1992) has 12 pentagons and 20 hexagons.

Any further subdivision of the regular or semiregular polyhedra produces subregions of either unequal shapes, unequal areas, or both. In the case of triangles, each can be recursively subdivided into four smaller triangles. At a given recursion level, triangles at the same latitude band have the same area, but these areas decrease monotonically with distance from the equator

according to the sinc of latitude (Goodchild and Shiren, 1992). In the recursion procedure, the four generated triangles do not have the same angles as one another.

An empirical tessellation is an adaptive or data-driven decomposition based upon the *spherical Voronoi diagram*. This tessellation is derived using a set of fixed *generating points* $\mathbf{x}_1,...\mathbf{x}_n$. Each point within the convex hull of this set is assigned to its nearest $\mathbf{x}_i$. The resulting assignment produces the regions:

$$V_i = \{\mathbf{x} | \, s(\mathbf{x},\mathbf{x}_i) < s(\mathbf{x},\mathbf{x}_j)\}, \quad j = 1,..., i-1, i+1,..., n \tag{2.5.1}$$

Generation of a spherical Voronoi diagram is very similar to the generation of a planar Voronoi diagram. The minor modifications required are discussed in Okabe et al. (1992), Lawson (1984), Renka (1984), and Traversoni (1990). Arc distance replaces Euclidean distance and the boundedness of the sphere produces changes in the planar equations relating the number of points, edges, and Voronoi vertices.

Dual to the Voronoi diagram is a triangulation, formed by connecting points having a common boundary in the Voronoi diagram. In a triangulation, there are 2n-4 triangles if the convex hull is the entire sphere (Lawson, 1984). If instead, the convex hull is a proper subset of the sphere (resulting if the points are hemispheric), the relation for the number of triangles is unchanged from the planar form, with 2n-b-2 triangles for a convex hull having b boundaries.

An arbitrary point within a triangle can be represented in terms of local, barycentric coordinates of either the spherical triangle or its projection onto a planar triangle. The resulting coordinates are not equivalent in these two representations. The use of spherical barycentric coordinates is discussed in Baumgardner and Frederickson (1985), while the planar form using a projection onto the polyhedra solids is discussed in Goodchild and Shiren (1982).

2.6 Calculus

The derivative and integral of a function on the sphere are defined similarly to their planar forms. Table 2.1 defines a few common calculus operations on the sphere. Calculus operations are generally simpler to express in spherical coordinates, and such coordinates will be used exclusively in this section.

The finite difference form of Laplace's equation on the sphere is (Swarztrauber, 1974):

$$\begin{aligned}\nabla^2 g_{i,j} \approx & \frac{1}{(r_s \Delta\varphi)^2 \cos\varphi}\left[\cos\varphi_{(i+1)/2}(g_{i+1,j} - g_{i,j}) - \cos\varphi_{(i-1)/2}(g_{i,j} - g_{i-1,j})\right] \\ & + \frac{1}{(r_s \cos\varphi \Delta\lambda)^2}(g_{i,j+1} - 2g_{i,j} + g_{i,j-1})\end{aligned} \tag{2.6.1}$$

where subscript i is the latitude index and j is the longitude index. Algorithms for solving Laplace's equation or the Poisson equation on the sphere are described in Swarztrauber and Sweet (1979). FORTRAN subroutines can be obtained from the National Center for Atmospheric Research in Boulder, Colorado. A geographic application is discussed in Tobler (1994).

Advection is defined as the transport of a substance by a fluid in motion. Advection of a scalar quantity g on a sphere is defined as (Smolarkiewicz and Rasch, 1991):

$$-\nabla(\rho g \cos\varphi)$$

where ρ is the density of the advecting medium (such as air or water), and is typically treated as a constant.

The length of an arbitrary arc can be computed by the line integration $\int g \, ds$, where the line element ds is:

Gradient	$\nabla g = \frac{1}{r_s}\begin{bmatrix} 1/\cos\varphi \ \partial g/\partial\lambda \\ \partial g/\partial\varphi \end{bmatrix}$
Laplacian	$\nabla^2 g = \frac{1}{r_s^2}\left[\frac{1}{\cos^2\varphi}\frac{\partial^2 g}{\partial\lambda^2} + \frac{1}{\cos\varphi}\frac{\partial}{\partial\varphi}\cos\varphi\frac{\partial g}{\partial\varphi}\right]$
Biharmonic	$\nabla^4 g = \nabla^2(\nabla^2 g)$
	$= r_s^{-4}\left[\frac{1}{\cos^4\varphi}\frac{\partial^4 g}{\partial\lambda^4} + \frac{1}{\cos^3\varphi}\left(\frac{\partial^3}{\partial\varphi\partial^2\lambda}\cos\varphi\frac{\partial g}{\partial\varphi} + \frac{\partial}{\partial\varphi}\cos\varphi\frac{\partial^3 g}{\partial\varphi\partial^2\lambda}\right) + \frac{1}{\cos^2\varphi}\frac{\partial}{\partial\varphi}\cos\varphi\frac{\partial^2}{\partial\varphi^2}\cos\varphi\frac{\partial g}{\partial\varphi}\right]$
Divergence	$\nabla\bullet\mathbf{V} = (r_S\cos\varphi)^{-1}(\partial\mathbf{V}_\lambda/\partial\lambda + \partial/\partial\varphi(\mathbf{V}_\varphi\cos\varphi))$
Curl (vertical component)	$(\nabla\times\mathbf{V})\bullet\hat{\mathbf{z}} = (r_s\cos\varphi)^{-1}[\partial\mathbf{V}_\varphi/\partial\lambda - \partial(\mathbf{V}_\lambda\cos\varphi)/\partial\varphi]\hat{\mathbf{z}}$
Integral	$\int_S g\,dS = r_s^2\int_0^{2\pi}\int_{-\pi/2}^{\pi/2} g\cos\varphi\,d\varphi d\lambda$
Line element	$ds = r_S[\cos^2\varphi\,(d\lambda)^2 + (d\varphi)^2]^{1/2}$
(meridional only)	$ds_\varphi = r_S\,d\varphi$
(zonal only)	$ds_\lambda = r_S\cos\varphi\,d\lambda$
Area element	$dS = r_S^2\cos\varphi\,d\varphi\,d\lambda$

Table 2.1 Calculus operators on the sphere. The notations $\mathbf{V}_\varphi$ and $\mathbf{V}_\lambda$ refer to the components of $\mathbf{V}$ in the meridional and zonal directions, respectively.

$$ds = r_S [\cos^2 \varphi \, (d\lambda)^2 + (d\varphi)^2]^{1/2} \quad . \tag{2.6.2}$$

The area of an arbitrary region can be obtained from $\int g \, dS$, where the area element dS is:

$$dS = r_S^2 \cos \varphi \, d\varphi \, d\lambda \quad . \tag{2.6.3}$$

2.7 Geometric aspects of spherical coordinates

Much of the difficulty associated with spherical analysis is based upon the singularities of the spherical coordinate system at the poles. Many of the complexities disappear when working with the direction cosine representation (Swarztrauber, 1981).

To represent a function $g(\varphi, \lambda)$ with continuity across the poles, it must be independent of longitude there:

$$\begin{aligned} g(\pi/2, 0) &= g(\pi/2, \lambda) \qquad 0 \le \lambda \le 2\pi \\ g(-\pi/2, 0) &= g(-\pi/2, \lambda) \qquad 0 \le \lambda \le 2\pi \quad . \end{aligned} \tag{2.7.1}$$

For the stronger condition of C^1 continuity, or continuous first derivatives, the additional requirements are:

$$\begin{aligned} \frac{\partial g(\pi/2, \lambda)}{\partial \theta} &= \cos\lambda \frac{\partial g(\pi/2, 0)}{\partial \theta} + \sin\lambda \frac{\partial g(\pi/2, \pi/2)}{\partial \theta} \\ \frac{\partial g(-\pi/2, \lambda)}{\partial \theta} &= \cos\lambda \frac{\partial g(-\pi/2, 0)}{\partial \theta} + \sin\lambda \frac{\partial g(-\pi/2, \pi/2)}{\partial \theta} \end{aligned} \qquad 0 \le \lambda \le 2\pi \tag{2.7.2}$$

A more general treatment of the "pole problem" can be found in Boyd (1989).

Vector-valued functions are particularly affected by the singularities at the poles. For example, a northward wind abruptly becomes a southward wind upon crossing a pole, and an eastward wind becomes westward if it slides past a pole. In general, the zonal and meridional components of a vector are discontinuous at the poles because of this change of sign. Analysis of vector-valued functions is discussed in Section 5.4.

2.8 Spherical harmonic transform

The *spherical harmonic transform* represents the natural generalization of the spatial Fourier transform to the domain of the sphere. When applied to a spherical data set, the transform produces a change of independent variable from geographic location (φ, λ) to an inverse measure of spatial scale (m,n) (where large m and n represent small spatial scales). Index m represents the zonal spatial scale (along latitude parallels alone) and index n represents the total spatial scale independent of any preferred direction. The value of n-m represents the meridional spatial scale, as the zonal and meridional spatial scales are additive, unlike the planar case where $k_{total} = (k_x^2 + k_y^2)^{1/2}$. Both m and n are non-negative in the form of the spherical harmonics presented here, and satisfy the condition $m \le n$. The indices m and n can also be interpreted as *wave numbers* in the zonal and total directions, respectively.

The *spherical harmonic functions* are defined as products of Fourier functions in longitude and normalized associated Legendre polynomials in sine of latitude:

$$\cos(m\lambda) \, P_n^m(\sin\varphi) \qquad \sin(m\lambda) \, P_n^m(\sin\varphi) \tag{2.8.1}$$

The transform of a function $g(\varphi, \lambda)$ is computed from its projection onto each of the spherical harmonic functions (2.8.1):

$$a_{mn} = \int_S g(\varphi,\lambda)\ \cos(m\lambda)\ P_n^m(\sin\varphi)\ dS$$

$$b_{mn} = \int_S g(\varphi,\lambda)\ \sin(m\lambda)\ P_n^m(\sin\varphi)\ dS \qquad (2.8.2)$$

The a_{mn} and b_{mn} are the spherical harmonic *coefficients* used to represent and recreate the function g as a linear combination of spherical harmonics:

$$g(\varphi,\lambda) = \sum_{n=0}^{N}\sum_{m=0}^{n}[a_{mn}\cos(m\lambda) + b_{mn}\sin(m\lambda)]\ P_n^m(\sin\varphi) \qquad (2.8.3)$$

The spherical harmonic representation provides equal area resolution over the entire sphere and produces fields that are continuously differentiable at the poles from all directions. The selected value of N in (2.8.3) is the *truncation* level and implies that the summation in (2.8.3) will have $(N+1)^2$ terms.

Parseval's relation relates the square of the coefficients to the square of the function:

$$\int f^2\ dS = \Sigma\Sigma\ \rho(m,n) \qquad (2.8.4)$$

where

$$\rho(m,n) = a_{mn}^2 + b_{mn}^2 \qquad (2.8.5)$$

is known as the *power spectrum*. Equation (2.8.4) represents a decomposition of the integrated value of f^2 by *spatial scale*. The power spectrum as a function of n alone is obtained by summation over all values of m for each n:

$$\rho(n) = \sum_m (a_{mn}^2 + b_{mn}^2) \qquad (2.8.6)$$

This value is independent of the placement of the poles, as expected from the definition of total wave number.

Once the spherical harmonic coefficients are known, the field can be easily created relative to a rotated pole. This transformation is carried out by replacing a_{mn} and b_{mn} with:

$$\tilde{a}_{mn} = \sum_{p=0}^{n} c_{pmn}a_{pn} \qquad \tilde{b}_{mn} = \sum_{p=0}^{n} c_{pmn}b_{pn} \qquad (2.8.7)$$

where c_{pmn} are a function only of the distance that the pole has been rotated.

The spherical harmonic functions are ordinarily computed on a regular grid, recursively from known values at small m and n. The grid is usually equally spaced in longitude; the latitude grid may be either equally spaced or with points that are the zeros of the Legendre polynomials. Spherical harmonic functions and transforms are available in many standard mathematical software packages (Mathematica, MATLAB, IDL). FORTRAN programs for performing spherical harmonic analysis and synthesis are also available from the National Center for Atmospheric

Research in Boulder, Colorado.

For some applications, it will be desirable to know expressions for the gradient and Laplacian of the spherical harmonic functions. Using standard identities for derivatives of associated Legendre polynomials (Swarztrauber, 1981), the gradient of the spherical harmonics is:

$$\nabla[\cos(m\lambda)P_n^m(\sin\varphi)] = \frac{\sqrt{n(n+1)}}{r_s}\begin{bmatrix} -\sin(m\lambda)W_n^m \\ \cos(m\lambda)V_n^m \end{bmatrix}$$
$$\nabla[\sin(m\lambda)P_n^m(\sin\varphi)] = \frac{\sqrt{n(n+1)}}{r_s}\begin{bmatrix} \cos(m\lambda)W_n^m \\ \sin(m\lambda)V_n^m \end{bmatrix} \qquad (2.8.8a)$$

The basis functions V and W are linear combinations of P using the relations:

$$W_n^m = c_1 P_{n-1}^{m-1} + c_2 P_{n-1}^{m+1}$$
$$V_n^m = c_3 P_n^{m+1} + c_4 P_n^{m-1} \qquad (2.8.8b)$$

where

$$c_1 = 1/2\ [(2n+1)(n+m)(n+m-1)/n(n+1)(2n-1)]^{1/2}$$
$$c_2 = 1/2\ [(n-m)(n-m-1)/n(n+1)]^{1/2}$$
$$c_3 = 1/2\ [(n-m)(n+m+1)/n(n+1)]^{1/2}$$
$$c_4 = -1/2\ [(n+m)(n-m+1)/n(n+1)]^{1/2} \quad .$$

The Laplacian of the spherical harmonic functions is of a simpler form, as $n(n+1)/r_n^2$ is the eigenvalue for that operator:

$$\nabla^2[\cos(m\lambda)P_n^m(\sin\varphi)] = \frac{n(n+1)}{r_s^2}\cos(m\lambda)P_n^m(\sin\varphi)$$
$$\nabla^2[\sin(m\lambda)P_n^m(\sin\varphi)] = \frac{n(n+1)}{r_s^2}\sin(m\lambda)P_n^m(\sin\varphi) \qquad (2.8.9)$$

Spherical harmonics are usually applied to functions defined over the entire sphere. However, for data defined only on a spherical cap, spherical cap harmonics can be used. In this case, the index n is real-valued, rather than integer-valued, although only a finite number of harmonics are required. The values of n are computed to satisfy simple boundary conditions at the perimeter of the cap (Haines, 1985). Software for computing the appropriate values of n and for carrying out the transform and its inverses are described in Haines (1988).

3. Spherical probability distributions

In this section, probability density functions (pdf) on the sphere are defined. The sections are divided into continuous and discrete distributions, although distributions for one can be applied to the other. Thus, point data can be well represented with continuous distributions where the number of data points is large.

3.1 Continuous distributions

Many continuous probability distributions on the unit sphere have been studied in great detail. Summaries can be found in Fisher et al. (1987) and Upton and Fingleton (1989; Chapter 10). Most of the interest in these distributions derives from directional statistics, in which the unit sphere is a conceptual model for three-dimensional direction. Very few applications consider the earth surface as the spherical domain. In this section, the relevant distributions for use in geographical applications are described.

A probability density function (pdf) can be expressed in terms of either spherical coordinates $f(\varphi, \lambda)$ or direction cosines $f(\mathbf{x})$, and there is an important distinction between the two forms. Consider a random variable on the sphere, such as the geographic distribution of birds, with pdf $f(\varphi, \lambda)$. The proportion of birds in the latitude-longitude range

$$(\varphi,\varphi+d\varphi) \text{ and } (\lambda,\lambda+d\lambda) \tag{3.1.1}$$

is by definition $f(\varphi, \lambda)\, d\varphi\, d\lambda$. Note that on a unit sphere, $d\varphi\, d\lambda \neq dS$; therefore, $f(\varphi, \lambda)$ cannot be interpreted as the bird density per unit area. In contrast, $f(\mathbf{x})$ does have the desirable interpretation of an areal density. The proportion of birds in the range:

$$(\mathbf{x}, \mathbf{x}+\mathbf{dx}) = (x, x+dx), (y, y+dy), \text{ and } (z, z+dz) \qquad \mathbf{x}+\mathbf{dx} \in S^2 \tag{3.1.2}$$

is $f(\mathbf{x})\, \mathbf{dx}$, where $\mathbf{dx}$ is in fact an areal unit. The difference between the two forms becomes apparent using the Jacobian of the transformation (2.7.2):

$$f(\varphi, \lambda) = \cos\varphi\, f(\mathbf{x}) \qquad ; \tag{3.1.3}$$

there is a $\cos\varphi$ factor is absorbed into the spherical coordinate form of the density $f(\varphi, \lambda)$. As a result of this difference, some authors define the probability density element (pde) dh:

$$\begin{aligned} dh &= f(\mathbf{x})\, d\mathbf{x} \\ &= f(\varphi, \lambda)\, d\varphi\, d\lambda \end{aligned} \tag{3.1.4}$$

The pde is a probability (rather than a density) and is independent of the coordinate system.

All of the major continuous distributions on the sphere have pdf expressible as:

$$f(\mathbf{x}) = c\, e^{A(\mathbf{x})} \qquad \mathbf{x} \in S^2 \qquad . \tag{3.1.5}$$

A summary of these distributions appears in Table 3.1. The constants c are not shown in the table, as they are often of a complicated form. A trivial special case is the uniform distribution, with pdf:

$$f(\mathbf{x}) = (4\pi)^{-1} \qquad . \tag{3.1.6}$$

This distribution is often used as a null hypothesis when testing for the presence of other distributions.

The rotationally symmetric class of distributions in Table 3.1 includes unimodal, bipolar, great circle, and small circle distributions. Distributions in this class are of the form $A(\mathbf{x}\bullet\mu)$ and are symmetric about the location $\mu \in S^2$. They are described in Sections 3.1.1 through 3.1.3. Distributions lacking rotational symmetry are discussed in Section 3.1.4.

Many statistical tests have been derived that assume an underlying unimodal distribution. This is analogous to the situation on the plane, where the t test and F test assume an underlying

normal distribution. Other tests are slightly less restrictive, and only require an underlying distribution that is rotationally symmetric (such as those listed as such in Table 3.1). A summary of these tests can be found in Fisher et al. (1987). Few non-parametric tests have been adapted to the spherical domain.

Other densities can be expressed non-parametrically either in terms of its moments or as an expansion in orthogonal functions. This subject has been discussed by Beran (1979) and Watson (1983) using a Fourier expansion, but has not been applied in practice. Sneyers and Isacker (1980) discuss a similar approach on the unit circle for meteorological data.

3.1.1 Fisher distribution

The most useful unimodal distribution is known alternatively as the Fisher, von Mises, or Langevin distribution. This distribution, denoted $F(\mu,\kappa)$, comes closest to representing the generalization of the normal distribution to the sphere. While not satisfying all of the normal distribution's properties, there are remarkable numerical agreements between it and competing distributions (Watson, 1983), and it is reasonable to use the Fisher distribution as an all-purpose unimodal distribution. The Fisher distribution thus represents a useful model for the random errors associated with a regression or other least squares fit.

The parameters of the distribution are the mean location μ and the dispersion or concentration parameter κ, with $\kappa>0$. The concentration parameter is an *inverse* measure of dispersion, with large κ representing highly concentrated data and small κ representing widely dispersed data. For relatively large κ,

$$\kappa^{-1} \approx \sigma^2 \tag{3.1.7}$$

so that κ can be interpreted directly as an inverse variance. The Fisher distribution has pdf:

$$\begin{aligned} f(\mathbf{x}) &= c\, e^{\kappa \boldsymbol{\mu} \cdot \mathbf{x}} \\ &= c\, e^{\kappa \cos\theta} \end{aligned} \tag{3.1.8a}$$

where θ is the angle between $\mathbf{x}$ and μ using (2.1.7). The constant c is:

$$\begin{aligned} c &= \kappa\, [4\pi \sinh \kappa]^{-1} \\ &= \kappa\, e^{-\kappa}\, [2\pi\, (1 - e^{2\kappa})]^{-1} \end{aligned} \tag{3.1.8b}$$

This pdf has its maximum density at $\mathbf{x}= \mu$ (or $\theta= 0$) where

$$f(\mu) = \kappa\, [2\pi\, (1 - e^{2\kappa})]^{-1} \tag{3.1.9}$$

and a minimum at the antipode $-\mu$ where

$$f(-\mu) = e^{-2\kappa}\, f(\mu) \tag{3.1.10}$$

This distribution approximates the multivariate normal distribution for either small angles θ or large κ ($\kappa>3$) (Clark and Morrison, 1983). The approximating normal density has mean location μ, variance κ^{-1}, and zero correlation. In the limit $\kappa \rightarrow \infty$, the Fisher distribution

approaches a "point" distribution, while for $\kappa \to 0$, it approaches the uniform distribution.

In the spherical coordinate form,

$$f(\varphi,\lambda)= c \cos\varphi \exp \{ \kappa \sin\varphi \sin\varphi_0 + \cos\varphi \cos\varphi_0 \cos(\lambda-\lambda o) \} \qquad (3.1.11a)$$

or with coordinates rotated so that μ is at the north pole:

$$f(\varphi,\lambda)= c \cos\varphi \, e^{\kappa \sin\varphi} \quad . \qquad (3.1.11b)$$

In (3.1.11b), the density is zero at the north pole $(\pi/2,0)$, where the areal density form (3.1.8a) has a global maximum! The spherical coordinate form (3.1.11b) conveys a geometric property of the coordinate system at the point (the vanishing of area at latitude band $\varphi= \pi/2$) producing the minimum there, rather than the property of the function itself. Diagrams comparing these two forms are given in Fisher et al. (1987) and Mardia (1972).

The Fisher distribution is separable in latitude and longitude. The marginal distribution of λ is uniform over $[0,2\pi]$. The marginal distribution of φ is:

$$\begin{aligned} f(\varphi) &= \kappa \cos\varphi \, [2 \sinh\kappa]^{-1} \, e^{\kappa \sin\varphi} \qquad (0 \le \varphi \le \pi) \\ &= \kappa \cos\varphi \, [1 - e^{-2\kappa}]^{-1} \, e^{-\kappa(1-\sin\varphi)} \quad , \qquad (0 \le \varphi \le \pi) \end{aligned} \qquad (3.1.12a)$$

a distribution discussed in Clark (1983). The quantity $(1 - \sin\varphi)$ in (3.1.12b) has a truncated exponential distribution on [0, 2] with rate κ. For large κ, the $[1 - e^{-2\kappa}]^{-1}$ factor is very close to 1 and can be neglected. For this case, the mean value of $f(\varphi)$ is approximately κ^{-1}.

For later reference, it will be useful to make a change of variables and express (3.1.12b) in terms of the distance of an arbitrary point from the north pole μ. Converting to colatitude and using (2.2.1a),

$$f(s)= \kappa/r_s \, (1-e^{-2\kappa})^{-1} \sin(s/r_s) \exp\{-\kappa[1-\cos(s/r_s)]\} \qquad (0 \le s \le \pi r_s) \qquad (3.1.13)$$

The inverse r_s factor arises from the Jacobian of the transformation into dimensional units.

3.1.2 Circle and bipolar distributions

Bipolar and great circle distributions are *axial*, or symmetric about an axis, and satisfy the condition $f(\mathbf{x})= f(-\mathbf{x})$. These two classes of distributions share the same mathematical forms and differ only in the sign of A in the exponent of (3.1.5); the sign reversal interchanges the roles of maximum and minimum locations. The most common distribution of this type is the Dimroth-Watson distribution (Watson, 1983):

$$f(\mathbf{x})= c \exp\{ \kappa(\mu \bullet \mathbf{x})^2 \} \, . \qquad (3.1.14)$$

At the surface of the earth, there are relatively few axial phenomena (magnetic poles and tides represent exceptions) and relatively few great circle phenomena.

Processes with a peak density along a parallel (or other small circle) can be fitted to a *small circle* distribution. Such a distribution is rotationally symmetric about an axis, with a peak density at a small circle and local *minima* at each pole. Two small circle densities have been studied in the statistical literature.

The Mardia-Gadsden distribution (Mardia and Gadsen, 1977) can be expressed in any of the forms:

$$f(\mathbf{x}) = c \exp \{\kappa_1 \mu \bullet \mathbf{x} + \kappa_2 |\mu \times \mathbf{x}| \} \quad (3.1.15a)$$

$$= c \exp \{ \kappa_1 \cos \theta + \kappa_2 \sin \theta \} \quad (3.1.15b)$$

$$= c \exp \{\kappa \cos (\theta - \theta^*)\} \quad . \quad (3.1.15c)$$

Contours of this distribution are shown in Figure 3.1. In (3.1.15a), μ is the axis of rotation, and κ_1 and κ_2 represent polar and equatorial concentration parameters, respectively. In the other two forms, θ is the angle between μ and $\mathbf{x}$. In (3.1.15c), θ^* represents the colatitude maximum and a single concentration parameter is used. For $\theta^* = 0$ (or $\kappa_2 = 0$), the Mardia-Gadsden distribution reduces to the Fisher distribution with the same parameters κ (or κ_1) and μ. For $\theta = \pi/2$ (or $\kappa_1 = 0$), this distribution reduces to a great circle distribution. The constant c is complicated; it is expressible using sums of Bessel or incomplete beta functions (Mardia and Gadsen, 1977).

The Bingham-Mardia distribution (Bingham and Mardia, 1978) has pdf:

$$f(\mathbf{x}) = c \exp \{\kappa_1 \mu \bullet \mathbf{x} + \kappa_2 (\mu \bullet \mathbf{x})^2\} \quad (3.1.16a)$$

$$= c \exp \{-\kappa (\cos \theta - \cos \theta^*)^2\} \quad . \quad (3.1.16b)$$

The parameters κ_1, κ_2, μ, and $\theta*$ have the same interpretation as above.

Comparisons of the two small circle distributions are described in Mardia (1981). The Mardia-Gadsden distribution appears to produce tighter confidence interval bounds, whereas the Bingham-Mardia distribution produces smoother contours at the poles and a more easily computable analysis of variance. Most of the theoretical sampling properties of these distributions are restricted to the special case of large κ.

3.1.3 Distributions lacking rotational symmetry

Three distributions lacking rotational symmetry are listed in Table 3.1. The Bingham distribution (Bingham, 1974) possesses *oval symmetry* and has density expressible in either of the forms:

$$f(\mathbf{x}) = c \exp \{ \kappa (\mu_1 \bullet \mathbf{x})(\mu_2 \bullet \mathbf{x}) \} \quad (3.1.17a)$$

$$= c \exp \{\mathbf{x}' \mathbf{K} \mathbf{x}) \quad . \quad (3.1.17b)$$

In (3.1.17a), μ_1 and μ_2 are locational parameters (the foci of an ellipse) and κ has its usual interpretation. In (3.1.17b), $\mathbf{K}$ is a symmetric 3x3 matrix. This distribution is axial and has five parameters.

The Fisher-Watson distribution is the product of Fisher and Dimroth-Watson distributions:

$$f(\mathbf{x}) = c \exp \{ \kappa_1 \mu_1 \bullet \mathbf{x} + \kappa_2 (\mu_2 \bullet \mathbf{x})^2 \} . \quad (3.1.18)$$

This six-parameter distribution possesses *plane symmetry.*

Most of the distributions in Table 3.1 are special cases of the 8-parameter Fisher-Bingham distribution:

$$f(\mathbf{x}) = c \exp \{ \kappa_1 \mu_1 \bullet \mathbf{x} + \kappa_2 (\mu_2 \bullet \mathbf{x})(\mu_3 \bullet \mathbf{x}) \} \quad (3.1.19)$$

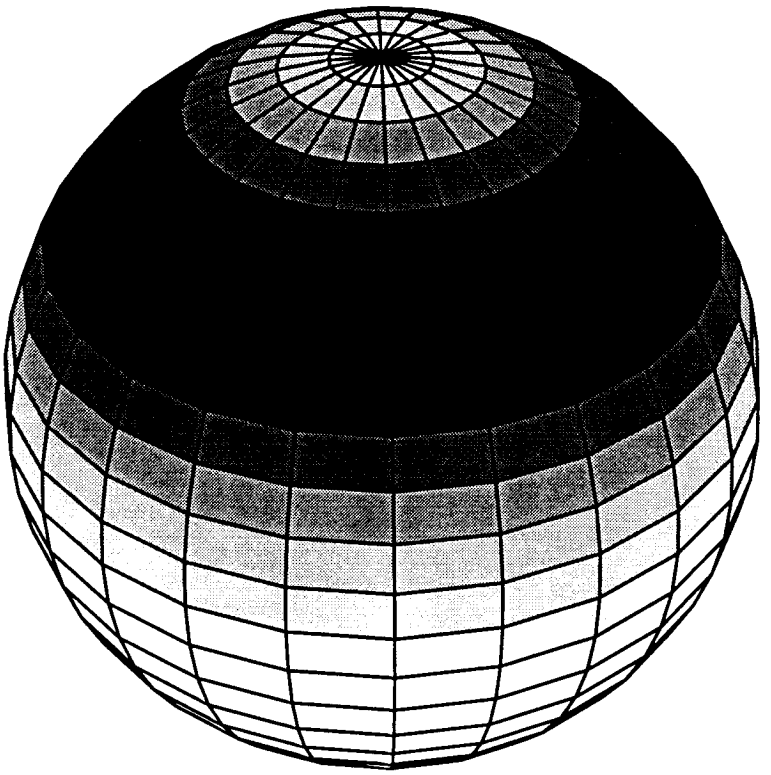

Figure 3.1 Contours of a Mardia-Gadsden (or small circle) distribution centered at $\theta = \pi/4$ (45° N).

Distribution type	A(**x**)	Name	Parameters
Uniform	0	Uniform	none
Distributions with rotational symmetry			
Unimodal	$\kappa\mu\bullet\mathbf{x}$	Fisher	$\kappa>0$, μ
Great circle/Bipolar	$\kappa(\mu\bullet\mathbf{x})^2$	Dimroth-Watson	κ^*, μ
	$\kappa\lvert\mu\times\mathbf{x}\rvert$	Selby	κ^*, μ
Small circle	$\kappa_1\mu\bullet\mathbf{x}+\kappa_2(\mu\bullet\mathbf{x})^2$	Bingham-Mardia	$\kappa_1>0$, $\kappa_2>0$, μ
	$\kappa_1\mu\bullet\mathbf{x}+\kappa_2\lvert\mu\times\mathbf{x}\rvert$	Mardia-Gadsden	$\kappa_1>0$, $\kappa_2>0$, μ
Distributions lacking rotational symmetry			
Oval symmetry	$\kappa(\mu_1\bullet\mathbf{x})(\mu_2\bullet\mathbf{x})$	Bingham	κ, μ_1, μ_2
Planar symmetry	$\kappa_1\mu_1\bullet\mathbf{x}+\kappa_2(\mu_2\bullet\mathbf{x})^2$	Fisher-Watson	κ_1, κ_2,μ_1,μ_2
General	$\kappa_1\mu_1\bullet\mathbf{x}+\kappa_2(\mu_2\bullet\mathbf{x})(\mu_3\bullet\mathbf{x})$	Fisher-Bingham	κ_1, κ_2,μ_1,μ_2, μ_3

Note that $\mu\bullet\mathbf{x}=\cos\theta$ and $\lvert\mu\times\mathbf{x}\rvert=\sin\theta$, where θ is the angle between μ and $\mathbf{x}$.

* $\kappa>0$ for great circle distributions, $\kappa<0$ for bipolar distributions

Table 3.1 Distributions on the sphere: $f(\mathbf{x})=c\,e^{A(\mathbf{x})}$ where c is a normalizing constant. Parameters μ are location parameters; parameters κ are spread parameters.

which is a product of Fisher and Bingham distributions. Comparisons of the various members of the family, including goodness-of-fit tests are found in Wood (1988). Methods of simulating data from these distributions are described in Wood (1987).

3.2 Discrete distributions

Very few results for discrete distributions appear in the literature. The discussion will be limited to the Poisson distribution, as there appear to be no references to other discrete spherical distributions.

3.2.1 Poisson distribution

The Poisson distribution on the sphere is very similar to its two-dimensional planar form. The number of points j in an area S is distributed as:

$$p(j)= e^{-\Omega S} \ (\Omega S)^j / j! \quad \text{for } j= 0, 1, 2, \ldots \tag{3.2.1}$$

where Ω is the expected number of points per unit area, or rate of the process. Typically, the area of interest is a spherical cap (cf. (2.4.8a)). Poisson distributed data are randomly distributed over the domain, a condition known as complete spatial randomness.

Common tests for complete spatial randomness are based upon nearest neighbor distances, but no nearest neighbor tests have been adapted to the sphere, to date. Tests of this type require formulas for the mean and standard deviation of the kth nearest neighbor distance from an arbitrary point. A complication associated with a bounded domain such as the sphere, is that a kth nearest neighbor may not exist at all. The probability of this occurring will be negligibly small in most applications, but from a theoretical perspective, a nonzero probability of a distance taking on the value infinity implies that the mean distance will be undefined. Thus, it is necessary to supply additional information to insure that the moments of the nearest neighbor distances exist.

The simplest additional assumption is to condition upon knowledge that the relevant nearest neighbor exists. It can be shown that given this assumption, the density function for the (first) nearest neighbor distance has a marginal Fisher distribution (3.1.13). As this result appears to be new, its derivation is given in Appendix A3.1.

A slightly stronger assumption is to condition upon knowledge that the number of realizations of the random process is equal to $n= 4\pi r_S^2\Omega$, where Ω is the mean rate of the process. This approach was followed by Scott and Tout (1989), who derived the pdf, mean, and variance for the kth nearest neighbor. (Scott and Tout derive the form for n-1 random points, but the formulas below are converted to their corresponding values for n random points.) The density function for the kth nearest neighbor distance is:

$$f_k(s) = \frac{n!}{2^n r_s (n-k)!(k-1)!} \sin(s / r_s)[1+\cos(s / r_s)]^{n-k}[1-\cos(s / r_s)]^{k-1} \qquad 0 \le s \le \pi r_S \ . \tag{3.2.2}$$

In particular, the nearest neighbor (k=1) has density:

$$f_1(s) = \frac{n}{2^n r_s} \sin(s / r_s)[1+\cos(s / r_s)]^{n-1} \qquad 0 \le s \le \pi r_S \tag{3.2.3}$$

with expected value

$$E(s) = \frac{(2n)!}{2^{2n} n!^2} \pi r_s \ . \tag{3.2.4}$$

For large n, this reduces to the planar form using Stirling's approximation:

$$E(s) \approx (\pi/n)^{1/2} r_S \tag{3.2.5}$$

$$= 1/2\ \Omega^{-1/2} \tag{3.2.6}$$

using $n = 4\pi r_S^2 \Omega$. Scott and Tout (1989) state that the exact form for the standard deviation is complicated and do not present it; the planar form for large n (Cressie, 1991) is:

$$\sigma \approx (1/\pi - 1/4)^{1/2}\ \Omega^{-1/2} \tag{3.2.7}$$

3.3 Bivariate distributions

The joint distribution of a pair of locations has been considered by several authors, although the mathematics are quite complicated even for simple cases. A general form

$$f(\mathbf{x},\mathbf{y}) = c \exp\{\kappa_1(\mu_1 \bullet \mathbf{x}) + \kappa_2(\mu_2 \bullet \mathbf{y}) + \mathbf{x}^{/}\mathbf{A}\mathbf{y}\} \tag{3.3.1}$$

where **A** is a matrix, is described in Mardia (1975) and summarized in Jupp and Mardia (1989). In (3.3.1), the conditional distributions of **x** and **y** are Fisher distributed, but the marginal distributions have a more complicated form. Rivest (1989) considers the special case of (3.3.1) when rotational symmetry can be assumed.

3.4 Estimation of parameters

Estimates of the parameters of the Fisher distribution are usually obtained from the maximum likelihood estimators:

$$\hat{\mu} = \Sigma \mathbf{x}_i / n \tag{3.4.1a}$$

$$\coth(\hat{\kappa}) - 1/\hat{\kappa} = R/n \quad . \tag{3.4.1b}$$

where the *resultant length* R is defined as:

$$R = [\,(\Sigma x_i)^2 + (\Sigma y_i)^2 + (\Sigma z_i)^2\,]^{1/2} \quad . \tag{3.4.1c}$$

For large κ, $\coth(\kappa) \approx 1$, and (3.4.1b) reduces to:

$$\hat{\kappa}^{-1} = 1 - \left(R/n\right) \quad . \tag{3.4.2}$$

This estimate is biased, and some modified forms have been suggested (Fisher et al., 1987).

Estimation of the parameters μ and θ^* in the small circle distribution (3.1.11c) is achieved by simultaneous solution of the set:

$$\Sigma\, \mu \bullet \mathbf{x_i} = \cot\theta^* \ \Sigma\, |\mu \times \mathbf{x_i}|$$

$$\Sigma \mathbf{x_i} = \tan\theta^* \ \Sigma\, \mathbf{x_i}\, \mu\, \mathbf{x_i} / |\mu \times \mathbf{x_i}| + \omega\mu \tag{3.4.3}$$

where ω is a Lagrange multiplier to constrain the solution to the sphere. Iterative solutions are described in Mardia and Gadsden (1977). Gray et al. (1980), Mancktelow (1981), and Schott et al. (1991) describe additional procedures for estimating small circle parameters.

Appendix 3.A Proof that (conditional) nearest neighbor distance has a marginal Fisher distribution

Consider a set of points generated by the Poisson process as in (3.2.1). A spherical cap with area S contains no points with probability

$$p(0)=e^{-\Omega S} \tag{3.A.1}$$

If we follow the analogous reasoning as on the plane, the distance from the origin of the spherical cap to its nearest neighbor has cumulative distribution function (cdf):

$$F_1(s)= 1 - e^{-\Omega S} \tag{3.A.2}$$

and pdf:

$$\begin{aligned} f_1(s)&= dF_1/ds \\ &= -\Omega\, dS/ds\, e^{-\Omega S} \end{aligned} \tag{3.A.3}$$

However, (3.A.2) and (3.A.3) are not correct as stated, because distance has not been restricted to one-half the sphere circumference. This complication will be ignored temporarily, and will be corrected below, by conditioning upon the distance lying in its proper domain. Using the spherical cap formulas (2.4.8a) and (2.4.10a), (3.A.3) becomes

$$f_1(s)=2\pi\Omega r_S \sin (s/r_S) \exp \{-2\pi\Omega r_S^2 (1- \cos (s/r_S))\} \qquad (s\geq 0). \tag{3.A.4}$$

Defining

$$\kappa= 2\pi\Omega r_S^2, \tag{3.A.5}$$

the nearest neighbor distance has density

$$f_1(s)= \kappa/r_S \sin (s/r_S) \exp \{-\kappa(1- \cos (s/r_S))\} \qquad (s\geq 0) \ . \tag{3.A.6}$$

This differs from (3.1.13) by a constant and has distance properly restricted to the spherical domain. If in (3.A.6), we condition upon knowledge that the nearest neighbor exists, this is equivalent to conditioning upon $0 \leq s \leq \pi r_S$, or multiplication by a constant. Thus, for it to integrate to 1, its density can be no other than (3.1.13).

From (3.A.5), the parameter 2κ in the Fisher density function can be interpreted as a nondimensional rate parameter for a Poisson process on a spherical domain.

4. Analysis of point data

In this section, the analysis of point data is considered. Thus the dependent variable is location on the sphere surface. The first section discusses sample statistics and exploratory data analysis methods. The next section discusses regression and correlation measures for spherical point data. Representation and search are the subjects of the following two sections, respectively, topics that are particularly relevant to GIS applications. The section closes with an example from location analysis.

4.1 Sample statistics

4.1.1 Measures of central tendency

Consider a random sample of point data $\mathbf{x}_1 = (x_1, y_1, z_1), \ldots, \mathbf{x}_n = (x_n, y_n, z_n)$ where $\mathbf{x}_i \in S^2$ for each sample i. The sample *spherical mean* is the location:

$$\bar{\mathbf{x}} = (\Sigma x_i / R, \Sigma y_i / R, \Sigma z_i / R)' \tag{4.1.1}$$

where the resultant length R is

$$R = [\, (\Sigma x_i)^2 + (\Sigma y_i)^2 + (\Sigma z_i)^2 \,]^{1/2} . \tag{4.1.2}$$

The spherical mean is a point on the sphere surface; in contrast, the *center of mass*:

$$\begin{aligned} \mathbf{x_c} &= (\Sigma x_i/n, \Sigma y_i/n, \Sigma z_i/n)' \\ &= R/n \ \bar{\mathbf{x}} \end{aligned} \tag{4.1.3}$$

is interior to the sphere unless all the mass is concentrated at a single point.

Fisher and Lee (1983) discuss pooled estimation of the mean for the case where several distributions are present. Tests for comparing mean locations are discussed in Fisher et al. (1987) and Fisher and Hall (1990).

The *spherical median* is defined as the location $\mathbf{x}$ that minimizes the sum of distances to the data points (Fisher 1985):

$$r_s \min \sum_{i=1}^{n} \cos^{-1}(\mathbf{x} \bullet \mathbf{x}_i) \quad . \tag{4.1.4}$$

Equation (4.1.4) is solvable numerically using most standard optimization package. The spherical median and mean differ in that the former minimizes the average of $\cos^{-1}(\mathbf{x} \bullet \mathbf{x_i})$ and the latter minimizes the average of $\mathbf{x} \bullet \mathbf{x_i}$.

4.1.2 Measures of dispersion

Several measures of dispersion have been defined on the sphere. The simplest dispersion measure is the distance of the center of mass (4.1.3) from the sphere surface:

$$r_S(1 - R/n) \quad .$$

This distance is the *spherical standard deviation* and ranges from 0 for data entirely concentrated at a single point to r_S for spherically symmetric data. In the literature, this measure is incorrectly referred to as a spherical variance.

Another measure of dispersion (Woodcock, 1977; Woodcock and Naylor, 1984) is provided by the eigenvalues and eigenvectors of the matrix of second moments (or *orientation* matrix) $\mathbf{T}$ where:

$$\mathbf{T} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' = \frac{1}{n}\begin{bmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i z_i \\ \sum y_i x_i & \sum y_i^2 & \sum y_i z_i \\ \sum z_i x_i & \sum z_i y_i & \sum z_i^2 \end{bmatrix} \tag{4.1.5}$$

$$= E(\mathbf{x}\,\mathbf{x}^{/}).$$

This method should be used with caution, as it measures dispersion based upon perpendicular distances from axes, rather than geodesic distances from points. Nevertheless, the eigenvalue method is in common use and is described below.

Let (τ_1, τ_2, τ_3) represent the eigenvalues of **T**, with $\tau_1 \leq \tau_2 \leq \tau_3$. The eigenvalues sum to 1 because this sum is equal to that of the diagonal elements of **T**:

$$\tau_1 + \tau_2 + \tau_3 = \underbrace{\Sigma x_i^2 + \Sigma y_i^2 + \Sigma z_i^2}_{=1} . \tag{4.1.6}$$

The eigenvalues are also nonnegative because the matrix is nonnegative definite, which is a consequence of:

$$\mathbf{x}'\mathbf{T}\mathbf{x} = \frac{1}{n}\sum_{i=1}^{3}(\mathbf{x}\mathbf{x}_i^{/})^2 \tag{4.1.7}$$

The magnitude of the largest eigenvalue τ_3 is a measure of the degree of clustering of the data, and its associated eigenvector $\mathbf{e_3}$ identifies the location or *principal axis* about which the clustering occurs. τ_3 satisfies $1/3 \leq \tau_3 \leq 1$, with the lower limit corresponding to uniformly distributed data and no tendency to cluster. Larger values of τ_3 correspond to increasingly clustered data about the axis $\mathbf{e_3}$; the upper limit $\tau_3 = 1$ corresponding to data confined to this single axis. The measure τ_3 does not discriminate between clustering about a point and clustering about a point and its antipode. However, these two cases can be distinguished using the auxiliary measure R/n, as $R/n \approx 1$ implies a single cluster and $R/n \approx 0$ represents an antipodal cluster pair. No such ambiguity exists about the eigenvalue $\mathbf{e_3}$, as it is *directed* and identifies the *single* cluster center except where this is not possible (as in the limiting cases of uniformly distributed or bipolar data).

The second largest eigenvalue τ_2 ranges in value from 0 to τ_3 and measures the degree to which the data fall along a *great circle*. This great circle lies in the plane spanned by the intersection of eigenvectors $\mathbf{e_2}$ and $\mathbf{e_3}$. The minimum value $\tau_2 = 0$ corresponds to the $\tau_3 = 1$ case described above, with data concentrated along an axis. The maximum value $\tau_2 = \tau_3$ occurs for data entirely confined to a great circle (except in the special case of uniformly distributed data).

Eigenvalue τ_1 should be interpreted as a residual value. It takes on its lower value of zero for perfect fits to an axis or great circle, whereas, uniformly distributed data produce the maximum value $\tau_1 = 1/3$.

For large sample sizes, the standard error of the estimate of the spherical mean can be estimated using the central limit theorem (Fisher and Lewis, 1983):

$$\sigma = r_S\,(n - \Sigma\,(\mu \bullet \mathbf{x}_i)^2)^{1/2} / R \tag{4.1.8}$$

Unfortunately, this quantity is often referred to in the literature as a spherical variance.

Several tests for complete spatial randomness are compared in Fisher et al. (1985). One common method uses the values of the eigenvalues of the second moment matrix **T** (4.1.5). These eigenvalues are equal for uniformly distributed data. Support for the hypothesis of uniformity occurs if the observed χ^2 value, defined by

$$\chi^2 = 15n/2 \sum (\tau_i - 1/3)^2 \qquad (4.1.9)$$

exceeds the critical χ^2_c value with 5 degrees of freedom.

4.1.3 Confidence intervals

Using (4.1.8), confidence regions can be placed around estimates of the spherical mean or median. Two cases are considered, depending upon assumptions made about the underlying distribution. If this distribution can be assumed to be rotationally symmetric, the confidence region about the spherical mean is a spherical cap. For large sample sizes, $\bar{\mathbf{x}}$ is approximately distributed as a normal random variable. The cap has spherical radius:

$$s = r_S \sin^{-1} [(-\ln \alpha)^{1/2} \sigma / r_S] \qquad (4.1.10)$$

where $100(1-\alpha)\%$ is the confidence level (typically $\alpha = .05$). If rotational symmetry cannot be assumed, an ellipsoidal cap about an estimate of the *median* can be computed (Fisher et al., 1987).

Where only small samples are available, confidence intervals can be generated using bootstrap methods (Fisher and Hall, 1989). Confidence intervals for the best fitting small circle latitude are discussed in Kelker and Langenberg (1982).

4.2 Regression and correlation

Regression on the sphere is the subject of the review article of Chang (1993). The general spherical regression problem is to find the best fitting orthogonal matrix **A** such that

$$\mathbf{y} = \mathbf{A}\mathbf{x} \qquad (4.2.1)$$

where **x** and **y** are points on the sphere (Jupp and Mardia, 1980). There is no constant term in (4.2.1) because all spherical translations are rotations.

A method for estimating **A** can be found in Fisher and Lee (1986). The residual of the fit is usually assumed to be Fisher distributed with mean location **Ax** and unknown concentration parameter κ. Decision analysis problems relating to spherical regression are discussed in Kim (1991). Matched regression on pairs of points is the subject of Prentice (1989). Jupp and Kent (1987) find the best fitting time trajectory of spherical data; a continuous function **R**(t) is found on the plane using the equidistant azimuthal projection of the sphere.

Several correlation coefficient measures have been defined for point data on a spherical domain. The measure of Fisher and Lee (1986) will be described here, as it most closely resembles correlation measures on the plane; others are described in Jupp and Mardia (1989). Two variables **x** and **y** have correlation $\rho = +1$ if there exists an orthogonal transformation **A** such that $\mathbf{y} = \mathbf{A}\mathbf{x}$ with $\det \mathbf{A} = +1$ (Fisher and Lee, 1986; Stephens, 1979). If $\mathbf{y} = \mathbf{A}\mathbf{x}$ with $\det \mathbf{A} = -1$, then $\rho = -1$ and **A** represents a rotation plus a reflection. More generally, $-1 \leq \rho \leq 1$, with larger magnitudes representing better fits to the respective models. The correlation coefficient ρ is calculated as:

$$\rho = S_{xy} / (S_{xx} S_{yy})^{1/2} \qquad (4.2.2)$$

where

$$S_{xy} = \det \sum \mathbf{x_i} \mathbf{y_i}'$$

$$S_{XX} = \det \sum \mathbf{x_i}\, \mathbf{x_i}^{/}$$
$$S_{yy} = \det \sum \mathbf{y_i}\, \mathbf{y_i}^{/} \quad . \qquad (4.2.3)$$

Critical values of ρ are difficult to obtain, in general (Fisher et al., 1987).

One of the most successful applications of spherical analysis has been in the field of plate tectonics, or continental drift. According to plate tectonics theory, continents in the distant past have drifted as a rigid body (with occasional loss and gain of land mass). The problem is to find the best fitting rotation matrix relating current continental positions to those suspected in the past (Chang, 1986; 1993; Cox et al., 1990).

No autocorrelation measures for spherical data have been developed, and this area represents a wide open field for research pursuits.

4.3 Representation

Voronoi polygons on the sphere were briefly described in Section 2.9. They can be computed in an analogous manner to those on the plane (Lawson, 1984; Turner, 1986; Okabe et al., 1992; Miles 1971). Lawson found that the time required to generate a triangulation on the sphere is proportional to $n^{5/4}$, where n is the number of points. This relation was derived empirically, using values of n ranging from 25 to 500. Equations (2.1.8) and (2.3.5) are used to find the midpoint and perpendicular bisector, respectively, of a pair of points.

If the generated polygons are triangles, the triangulation is not unique. For each pair of triangles sharing a common side, a different triangulation results from switching the diagonal element in the quadrilateral formed by the associated 4 vertices. Usually, sides of nearly equal length are preferred to long, skinny triangles. Criteria for choosing between the two choices include: maximizing the minimum triangle angle (Nielson and Ramaraj, 1987) and minimizing the radius of the inscribed circle (Renka, 1984; Lawson, 1984).

Edges of a spherical polygon are usually stored by specifying the coordinates of the endpoints. To represent an edge of angular length π, an additional vertex (or *pseudo-vertex*) is usually introduced along the edge.

White et al. (1992) and Brown (1993, 1994a, 1994b) discuss issues associated with sampling on the sphere. Goodchild and Shiren (1992) discuss addressing schemes for regions decomposed by the regular polyhedral solids (cf. Section 2.5).

Several polyhedral-based GIS have been proposed. The quaternary triangular mesh (QTM) (Dutton, 1984; White et al., 1992) subdivides the globe based upon a projection of the sphere onto the octahedron. Projection onto the dodecahedron is described in Wickman, Elvers, and Edvarson (1984). Projection onto the dodecahedron has been proposed by Fekete (1990), Fekete and Davis (1984), and Fekete and Treinish (1990), and onto the truncated icosahedron (producing 12 pentagons and 20 hexagons) by White et al. (1992).

Very few working examples of GIS on a spherical domain are known. The large geographic extent of a global GIS suggests the use of hierarchical data structures (Tobler and Chen, 1986; Mark and Lauzon, 1985), to provide ready access to information at various resolutions. The Hipparchus GIS (Lukatela, 1987) for geopositioning uses direction cosines as the basic storage units, and three values are required to reference a location. This represents a favorable tradeoff of storage vs. computational efficiency relative to the spherical coordinate form (requiring only two storage values per location), as the number of trigonometric calculations is greatly reduced.

4.4 Search and query algorithms

Hodgson (1992) described and tested an algorithm for finding the kth nearest neighbor on the sphere using spherical coordinates. The essence of this algorithm is two criteria for reducing the distance computation. The points are sorted according to latitude distance alone. If this distance exceeds the kth closest distance identified thus far in the search, no further evaluations are necessary. This result is justified using (2.2.6a) (distance on the sphere is at least that considering

latitude distance alone). Besides this stopping criterion, there is a partial distance evaluation method. If the one-dimensional longitudinal difference is sufficiently large according to (2.2.6b), it is not necessary to compute the full spherical distance. Hodgson noted that this algorithm is only one of the commonly used search algorithms that could be adapted to spherical geometry.

A basic operation of a GIS is to determine whether or not a point is located within a polygonal region. Bevis and Chatelain (1989) describe an algorithm and computer code for carrying out this query for an arbitrary spherical polygon. On the sphere, it is necessary to specify both the vertices of the polygon and another flag to indicate whether the "inside" or "outside" of the polygon is to be represented. This redundancy can be resolved using the direction of vertex numbering: clockwise numbering of vertices implies the outside polygon and counterclockwise numbering represents the inside polygon (Bevis and Cambareri, 1987).

Bevis and Chatelain's (1989) algorithm resolves the ambiguity differently, by having the user specify an interior point X. This point is used as the north pole in subsequent calculations. The algorithm determines the number of polygon side crossings between the point of interest and X; an even number of crossings implies that the point is inside and an odd number of crossings indicates outside. A number of restrictions apply to the specification of this orienting point, as it cannot be the antipode of a vertex and cannot lie on a great circle connecting neighboring vertices of the polygon boundary. If either of these conditions are not met, another interior point must be specified.

Renka (1984) and Lawson (1984) describe a query to determine which "side" of a directed arc contains a specified point, using (2.3.9). They also discuss a query for determining whether a point is interior or exterior to a small circle, using (2.3.18). General concepts of data depth and distance using various metrics are discussed in Liu and Singh (1992).

4.5 Location analysis

The most developed application of spherical analysis in human geography is the spherical location analysis problem. Its objective is to find the location $\mathbf{x}^*$ that minimizes the sum of weighted arc distances from a set of points $\mathbf{x}_i$:

$$\min \sum w_i \, r_S \cos^{-1} (\mathbf{x}_i \bullet \mathbf{x}^*) \qquad (4.5.1)$$

The planar version of (4.5.1) is known as the Weber problem. The spherical version is the subject of the reviews of Litwhiler (1977) and Wesolowsky (1985).

In the spherical version, the objective function is not convex, in general. A simplification results if the points lie along a great circle, in which case the optimum will be located at one of the points $\mathbf{x}_i$ (Drezner, 1981). Another simplification results if the points are hemispheric. In this case, the search can be limited to the spherical convex hull of the points (Aly et al., 1979). If all the points lie within a cap of spherical radius $\pi/4 \, r_S$, the distance function is convex, and a local maximizer is guaranteed to be a global one. When the weights are sufficiently large, it is known that the minimizing point must coincide with one of the data points. Conditions for this occurrence are presented in (Katz and Cooper, 1980). Extrema solutions to (4.5.1) occur in antipodal pairs, as a local minimum has the dual solution of a local maximum at its antipode.

A gradient method for solving (4.5.1) is described in Katz and Cooper (1980). Litwhiler and Aly (1979) describe two additional algorithms: the point projection method and the alternating direction search method. In the projection algorithm, the points are projected onto a plane with the current best location as the pole. Standard planar search methods are applied at this point. The cyclic meridian search alternates the search along a meridian and an orthogonal great circle. Xue (1994) presents an algorithm for the spherical facility location problem that is globally convergent. However, his study is restricted to points contained within a cap of radius of $\pi/4$, where convexity is assured. Drezner (1985) developed a finite algorithm without these restrictions. Extreme point solutions to (4.5.1) are found by differentiating with respect to φ and λ and setting each to zero.

5. Analysis of functions on the sphere

This section describes the analysis of functions defined on a spherical domain. Natural applications of this material are to global datasets obtained by satellite or computer simulation. The primary emphasis is upon fields, or continuous-valued functions that are everywhere defined. The first three sections discuss scalar-valued fields, such as surface temperature, elevation, and population density. These sections cover interpolation, gradient estimation, and correlations. The fourth section describes vector-valued fields, such as wind and monetary flow. The final section discusses representation of fields.

A field is characterized mathematically by a finite-valued function of location $g(\varphi,\lambda)$ or $g(\mathbf{x})$. The mean value of a field is defined as:

$$\bar{g} = \frac{1}{4\pi} \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} g(\varphi,\lambda) \cos\varphi \, d\varphi d\lambda \quad . \tag{5.1}$$

The center of mass $\mathbf{x}_c$ may be located either interior to, exterior to, or on the sphere and is defined as:

$$\mathbf{x}_c = \frac{\int \mathbf{x} g(\mathbf{x}) dS}{\int g(\mathbf{x}) dS} \quad . \tag{5.2}$$

The *spherical center of mass* is the extension of this point to the surface:

$$\bar{\mathbf{x}} = \frac{\mathbf{x}_c}{|\mathbf{x}_c|} \quad . \tag{5.3}$$

This formula can be used to find the center of population of a large region.

5.1 Interpolation

The general interpolation problem is to estimate the value of $g(\mathbf{x})$ using known values of $g(\mathbf{x}_i)$ for i= 1,...n. Several interpolation methods are in use on spherical domains. The various methods differ in the number of points used in the estimation procedure and in the way that the $g(\mathbf{x}_i)$ values are combined. The methods are discussed below in order of locality, beginning with local methods and ending with global methods. Source code in most cases is available from the authors cited below.

The tessellation method (Renka, 1984; Lawson, 1984) is based upon an empirical tessellation of the sphere using the $\mathbf{x}_i$ as generating points. This method uses the value of $g(\mathbf{x})$ at the nearest generating point $\mathbf{x}_i$ and this point's three nearest neighbors. The convex hull of the points is decomposed into spherical triangles using a Delauny triangulation (Okabe et al., 1992). The gradient of the function is then estimated at each vertex $\mathbf{x}_i$, using a technique described in the next section. The interpolated value at an arbitrary point $\mathbf{x}$ is obtained by identifying the triangle that contains $\mathbf{x}$ and using a cubic Hermite interpolation based upon the values of g and ∇g at the neighboring vertices.

In the more general local method, $g(\mathbf{x})$ is estimated using

$$g(\mathbf{x}) = \sum_{i=1}^{n} w(\mathbf{x}_i) g(\mathbf{x}_i) \tag{5.1.1}$$

with w an inverse function of the distance $s(\mathbf{x},\mathbf{x_i})$. In (5.1.1), the number of points n' with nonzero weights is dynamically determined based upon the amount of data in that region. Often, only points within a designated radius (the *radius of influence*) are used (Theibaux, 1987). A common choice for the weighting functions w is Fisher distribution functions with mean μ and a modified concentration parameter k, centered at each of the n' estimation points. These functions are normalized through a division by n':

$$\begin{aligned} w(\mathbf{x_i}) &= F(\mu,\kappa)/n \\ &= k(4\pi n \sinh k)^{-1} e^{k\mu\bullet\mathbf{x_i}} \quad . \end{aligned} \tag{5.1.2}$$

The amount of smoothing is controlled by the concentration parameter k, with smaller k producing more smoothing (i.e. smoothing over a broader range). More smoothing (smaller k) is appropriate where there are fewer data points and where the underlying data have greater variance.

For data that are relatively unimodal and isotropic, Diggle and Fisher (1985) suggest the formula

$$k = \kappa\, n^{1/3} \tag{5.1.3}$$

where the actual concentration parameter κ is estimated using (3.4.1b). Willmott et al. (1985) and Robeson and Willmott (1993) discuss applications of this method in climatology.

For more general data sets, the *cross-validation* method can be used to identify a suitable k. The value of k is selected to maximize the log-likelihood function:

$$\max \sum_i \log\, f(\mathbf{x}_i) \tag{5.1.4}$$

where f is computed by deleting point i in the calculation:

$$f(\mathbf{x}_i) = \sum_{i\neq j} w(\mathbf{x}_j,\mathbf{x}_i) \tag{5.1.5}$$

An iterative method must be used to find the κ that maximizes (5.1.4). Diggle and Fisher (1985) have produced FORTRAN source code to carry out this calculation.

Dierckx (1984; 1986) describes the fitting of bicubic splines to spherical data. These splines are semi-global, in that values of g at greater distances are included in the estimation procedure. Global spline fits are described by Wahba (1981) and Freeden (1981) who use thin plate splines to fit trigonometric functions in latitude and longitude. These splines are linear combinations of spherical harmonics and have as many coefficients as the original field.

If the spherical harmonic representation of a field is known, interpolation is carried out by evaluating (2.8.3) at the point of interest. This interpolation is "exact" if the coefficients are exactly known. However, the usual interpolation problem is to *obtain* the coefficients using limited and scattered data, a task requiring a method such as Wahba (1981) or Freeden (1981).

5.2 Gradient estimation

The gradient of a field is important both in its own right and as an intermediate function in other calculations, such as interpolation. The gradient on the sphere is defined as:

$$\nabla g = \frac{1}{r_s}\begin{bmatrix} 1/\cos\varphi\ \ \partial g/\partial\lambda \\ \partial g/\partial\varphi \end{bmatrix} \tag{5.2.1}$$

As with interpolation, the gradient may be estimated using either local or global information, or may be computed exactly using spherical harmonic coefficients.

Estimation using local information is described in Renka (1984) and Lawson (1984). They estimate the gradient at **x**, by fitting a bivariate quadratic function to data projected onto a plane. An azimuthal planar projection is used with **x** as the point of zero distortion. The estimated gradient is that of the fitted quadratic. This method is most suitable for interpolation over a relatively small region to minimize projection errors.

A global fit provides potentially more accurate estimation of the gradient, at a greater computational cost. Renka (1984) obtains the gradient of g as that which minimizes the curvature of g, that is, the g that minimizes

$$Q = \int_S g'' dS \tag{5.2.2}$$

This minimization problem is solved by differentiating Q with respect to the gradient and setting it to zero, producing a set of simultaneous equations that can be solved numerically.

If the spherical harmonic coefficients of a field are known, it is simple to compute the gradient of the field. The calculation is exact if the spherical harmonic coefficients are exactly known. This is accomplished by applying the gradient operator directly to the expansion:

$$g(\varphi, \lambda) = \Sigma a_{mn} \cos(m\lambda)\ P_n^m(\sin\varphi) + \Sigma b_{mn} \sin(m\lambda)\ P_n^m(\sin\varphi)\ , \tag{5.2.3}$$

and noting that the coefficients are not a function of space:

$$\nabla g(\varphi, \lambda) = \Sigma a_{mn} \nabla[\cos(m\lambda)\ P_n^m(\sin\varphi)] + \Sigma b_{mn} \nabla[\sin(m\lambda)\ P_n^m(\sin\varphi)]\ . \tag{5.2.4}$$

Equation (5.2.4) can be rewritten using (2.8.8) as:

$$\nabla g(\varphi,\lambda) = \frac{1}{r_s} \sum_{n=0}^{N} \sqrt{n(n+1)} \sum_{m=0}^{n} \{ a_{mn} \begin{bmatrix} -\sin(m\lambda) W_m^n \\ \cos(m\lambda) V_m^n \end{bmatrix} + b_{mn} \begin{bmatrix} \cos(m\lambda) W_m^n \\ \sin(m\lambda) V_m^n \end{bmatrix} \} \tag{5.2.5}$$

where V and W were defined in (2.8.8b). Thus the gradient of a field can be generated using the same spherical harmonic coefficients as the field itself, but with the modified basis set V and W.

5.3 Correlations

The autocorrelation of a field and the cross-correlation of two fields can be computed using the spherical harmonics expansions of the fields. For isotropic processes, the spatial autocorrelation is (Kaula, 1967):

$$C(\theta) = 1/8\pi^2 \ \Sigma\Sigma\ (\ a_{mn}^2 + b_{mn}^2\ )\ P_n(\cos\theta)\ . \tag{5.3.1}$$

where θ is the angular separation of a pair of points. More generally, for two variables g_1 and g_2, the cross-correlation is:

$$C(\theta) = 1/8\pi^2 \ \Sigma\Sigma\ (\ a_{1mn}\, a_{2mn} + b_{1mn}\, b_{2mn})\ P_n(\cos\theta). \tag{5.3.2}$$

Individual terms in the summation of (5.3.1) or (5.3.2) can be interpreted as the contribution to the total autocorrelation or cross-correlation at each spatial scale. Kaula (1967) discusses other

statistical analyses using spherical harmonic coefficients.

For specific probability distributions on the sphere, it is possible to compute the corresponding autocorrelation functions in closed form. For example, for uniformly distributed data with mean density Ω, the correlation function takes on the following form (Orsingher, 1984):

$$C(\theta)= 4\pi\Omega \sum\sum (a_{mn}^2 + b_{mn}^2) P_n(\cos \theta) . \qquad (5.3.3)$$

5.4 Vectors

In some applications, the dependent variable is a vector-valued function of location. Examples of vector functions include horizontal wind, heat or water flux, population migration, topographic gradient, and ocean currents. Note that this use of the word "vector" differs from that in Fisher et al. (1987), where geographic location is the dependent variable interpreted as a unit vector relative to the center of the earth.

It was noted in Section 2.7, that a continuous vector field will have a discontinuous representation in spherical coordinates at the poles, because the vector components change signs. For applications where the function is zero at and near the poles, the discontinuity is not a discontinuity at all (negative zero is still zero). In the more general case, it is not desirable to compute the spherical harmonic decomposition of a vector by transforming the zonal and meridional components as scalars, because the discontinuity of the field at the poles will "corrupt" the spectrum. Instead, an alternative decomposition can be used to retain continuity of the components across the poles.

This alternative decomposition is based upon Helmholz's theorem from fluid dynamics, which states that a vector can be expressed as the sum of *divergent* and *rotational* components. The associated spherical harmonic components are known as *vector spherical harmonics.* Equation (5.2.5) represents essentially one half of the necessary expansion, because a gradient is a purely divergent function, without a rotational component. It is trivial to generate the basis set for the rotational contribution, because for any (m,n) pair, the divergent basis is orthogonal to the rotational basis. Therefore, an arbitrary vector can be decomposed into divergent and rotational components (represented with superscripts D and R, respectively) as (Swarztrauber; 1981, 1993):

$$\mathbf{V}(\varphi,\lambda) = \sum_{n=0}^{N}\sum_{m=0}^{n}\{a_{mn}^{D}\begin{bmatrix} -\sin(m\lambda)W_m^n \\ \cos(m\lambda)V_m^n \end{bmatrix} + b_{mn}^{D}\begin{bmatrix} \cos(m\lambda)W_m^n \\ \sin(m\lambda)V_m^n \end{bmatrix}$$

$$+ a_{mn}^{R}\begin{bmatrix} \cos(m\lambda)V_m^n \\ -\sin(m\lambda)W_m^n \end{bmatrix} + b_{mn}^{R}\begin{bmatrix} \sin(m\lambda)V_m^n \\ -\cos(m\lambda)W_m^n \end{bmatrix}\} \qquad (5.4.1)$$

In general, a vector field requires two sets of coefficients, to represent the divergent and rotational components, respectively. However, if the field is purely divergent or purely rotational, it can be decomposed with a single a_{mn}, b_{mn} pair, although it has both zonal and meridional components.

Most of the general properties of scalar spherical harmonic coefficients described in Section 2.8 have analogs for vector harmonic coefficients. Thus, the power spectrum of a vector is the sum of the power in the divergent and rotational components:

$$\rho(m,n)= \rho^D(m,n) + \rho^R(m,n) \qquad (5.4.2)$$

where

$$\rho^D(m,n) = (a_{mn}^D)^2 + (b_{mn}^D)^2 \qquad (5.4.3a)$$

$$\rho^R(m,n) = (a_{mn}^R)^2 + (b_{mn}^R)^2 \qquad (5.4.3b)$$

Like the scalar harmonics, the vector harmonics represent an orthonormal set of functions.

Just as the gradient of a scalar field can be created using the same scalar coefficients as the field itself, several derived quantities can be created using the same vector coefficients as the vector field itself. The divergence, vorticity, streamfunction, and potential function of a vector field are respectively,

$$-(\nabla \bullet V)(\varphi,\lambda) = \frac{1}{r_s}\sum_{n=0}^{N}\sqrt{n(n+1)}\sum_{m=0}^{n}\{a^D_{mn}\cos(m\lambda)P^n_m + b^D_{mn}\sin(m\lambda)P^n_m\} \qquad (5.4.4a)$$

$$\xi(\varphi,\lambda) = \frac{1}{r_s}\sum_{n=0}^{N}\sqrt{n(n+1)}\sum_{m=0}^{n}\{a^R_{mn}\cos(m\lambda)P^n_m + b^R_{mn}\sin(m\lambda)P^n_m\} \qquad (5.4.4c)$$

$$\Psi(\varphi,\lambda) = r_s\sum_{n=0}^{N}\frac{1}{\sqrt{n(n+1)}}\sum_{m=0}^{n}\{a^R_{mn}\cos(m\lambda)P^n_m + b^R_{mn}\sin(m\lambda)P^n_m\} \qquad (5.4.4d)$$

$$\Phi(\varphi,\lambda) = r_s\sum_{n=0}^{N}\frac{1}{\sqrt{n(n+1)}}\sum_{m=0}^{n}\{a^D_{mn}\cos(m\lambda)P^n_m + b^D_{mn}\sin(m\lambda)P^n_m\} \qquad (5.4.4b)$$

As with scalar harmonics, the synthesis of a field using vector spherical harmonics provides a smoothing over the domain. Wahba (1982) and Freeden and Gervens (1993) discuss interpolation of vector-valued data on the sphere; Freeden and Gervens use vector spherical harmonics to carry out the interpolation.

5.5 Representation

Spherical harmonics can be used as a means of storage of a field. If the spatial variations are generally caused by variations at only a few spatial scales, and these scales are relatively consistent across the domain, the number of coefficients of non-negligible magnitude will be relatively few. Meteorological phenomena tend to be driven by atmospheric waves, and represent a natural application for spherical harmonic storage.

Bess et al. (1989) describes how a ten-year monthly data set of Nimbus-6 and Nimbus-7 satellite data were stored as spherical harmonic coefficients. A total of 169 coefficients were retained in storage, corresponding to N= 12 in (2.8.3). Killeen et al. (1989) used a spherical harmonic representation to store upper atmospheric climatology from many runs of a general circulation model (GCM). Vector spherical harmonic coefficients were used to represent the winds and scalar harmonics were used to represent temperature and constituent densities.

A spherical harmonic representation of the earth's topography is given in Balmino (1993). A spherical harmonic representation of the 0-1 global land-ocean function is given in Koppelt and Biegel (1989). This function takes on the value 0 over land and 1 over the oceans and demonstrates that a function need not be continuous to be represented with spherical harmonics. Another geographical example is Tobler (1992), who represents the world population with spherical harmonics.

6. Future directions

Surprisingly few applications on the spherical domain appear in the geographical literature. This situation is apparently due to the preference for analysis over relatively small spatial scales relative to the sphere. The success of the projection approach over these domains is undoubtedly a contributing factor. The Hipparchus geopositioning system represents a lone example of a full-scale GIS that accounts for spherical geometry. It is hoped that this review article will remedy this situation. The Hipparchus GIS uses direction cosines rather than latitude-longitude as the basic locational units, and the formulations in Section 2 suggest that this is a good idea for reducing

computational time requirements.

Goodchild (1988) and Tobler (1993) discuss several potential applications of spherical analysis. Perhaps the most promising application of spherical analysis will be to global change research. Nevertheless, few such applications using spherical geometry have been developed to date.

ACKNOWLEDGEMENTS

This report was funded by a Visiting Fellowship provided by the National Center for Geographical Information and Analysis at the University of California, Santa Barbara. Waldo Tobler provided the inspiration for this work, and he kindly provided editorial assistance and other very helpful suggestions.

APPENDIX 1 Definition of symbols

A	argument of exp in spherical statistical distributions
$\mathbf{c}$	axis of great circle
$C(\theta)$	autocorrelation function
$\mathbf{e_i}$	eigenvector
f	probability density function (pdf) on the sphere
$F(\mu,\kappa)$	Fisher distribution with mean μ and concentration parameter κ
g	arbitrary function on the sphere
I	moment of inertia
$\mathbf{I}$	identity matrix
m	spherical harmonic zonal wave number (along parallel of latitude)
n	spherical harmonic total wave number
n	number of points
N	truncation or upper limit for spherical harmonic total wave number
P_n^m	Legendre polynomial
$\rho(m,n)$	power at wave numbers m,n
$P_n(\theta)$	zonal-average Legendre polynomial (m= 0)
R	resultant length of a data set
$\mathcal{R}^2$	two-dimensional Euclidean space
$\mathcal{R}^3$	three-dimensional Euclidean space
$\mathbf{R}$	rotation matrix
r_S	radius of sphere (for earth, r_S = 6371 km)
s	distance
$\tilde{s}$	arc length
S	surface area
S^2	the unit sphere
$\mathbf{T}$	matrix of second moments of a data set
$\mathbf{x}$	arbitrary point on sphere
$\mathbf{V}$	arbitrary vector-valued function on the sphere
V_n^m	zonal derivative of P_{mn}
W_n^m	meridional derivative of P_{mn}
$\mathbf{x_c}$	center of mass
x	direction cosine along x-axis

y	direction cosine along y-axis
z	direction cosine along z-axis
κ	concentration parameter of probability distribution
θ	angle between two points
α_i	polygon angle
λ	longitude
τ_i	eigenvalue
φ	latitude
μ	mean location of probability distribution
Ω	rate of Poisson process

BIBLIOGRAPHY

Abbot, E. A., Flatland; a Romance of Many Dimensions, Seeley, London, 1884.

Aly, A. A., D. C. Kay, and D. W. Litwhiler, Jr., Location dominance on spherical surfaces, Operations Research, 27, 972-81, 1979.

Balmino, G., The spectra of the topography of the Earth, Venus and Mars, Geophys. Res. Lett., 20, 1063-6, 1993.

Barone, P., A discrete orthogonal transform based upon spherical harmonics, J. Comput. Appl. Math., 33, 29-34, 1990.

Barton, B. A., A note on the transformation of spherical coordinates, American Cartographer, 3, 161-68, 1976.

Baumgardner J. R. and P. O. Frederickson, Icosahedral discretization of the two-sphere, SIAM J. Numer. Anal., 22, 1107-1115, 1985.

Beran, R., Exponential models for directional data, Annals of Statistics, 7, 1162-1178, 1979.

Bess, T. D., G. L. Smith, and T. P. Charlock, A ten-year monthly dataset of outgoing longwave radiation from Nimbus-6 and Nimbus-7 satellites, Bull. Amer. Met. Soc., 70, 480-9, 1989.

Bevis, M. and G. Cambareri, Computing the area of a spherical polygon of arbitrary shape, Mathematical Geology, 19, 335-346, 1987.

Bevis, M. and J. Chatelain, Locating a point on a spherical surface relative to a spherical polygon of arbitrary shape, Mathematical Geology, 811-829, 1989.

Bingham, C., An antipodally symmetric distribution on the sphere, Ann. Statist., 2, 1201-1225, 1974.

Bingham C. and K. V. Mardia, A small circle distribution on the sphere, Biometrika, 65, 379-389, 1978.

Boyd, J. P., Chebyshev and Fourier Spectral Methods, Springer-Verlag, Berlin, 798 pp., 1989.

Brown, J. J., Sampling plans for an isotropic random sphere, *submitted*, 1993.

Brown, J. J., A central limit theorem for an isotropic random sphere, Stochastic Geometry and its Applications, *to appear*, 1994a.

Brown, J. J., Bootstrap for the random sphere, *submitted*, 1994b.

Burger, D., Sphereland; a Fantasy About Curved Spaces and an Expanding Universe, Crowell, New York, 1965.

Cabrera, J. and G. S. Watson, On a spherical median related distribution, Commun. Statist.-Theory Meth., 19, 1973-86, 1990.

Chang, T., Spherical regression, Annal. Stat., 14, 907-24, 1986.

Chang, T., Spherical regression and the statistics of tectonics plate reconstructions, Int'l. Stat.

Rev., 61, 299-316, 1993.

Clark, R. M., Estimation of parameters in the marginal Fisher distribution, Austral. J. Statist., 25, 227-237, 1983.

Clark, R. M. and B. J. Morrison, A normal approximation to the Fisher distribution, Austal. J. Statist., 25, 96-104, 1983.

Cox, D. R. and V. Isham, Point Processes, Chapman and Hall, London, 188 pp., 1980.

Cressie, N. A. C., Statistics for Spatial Data, John Wiley & Sons, New York, 900 pp., 1991.

Daley, R., Atmospheric Data Analysis, Cambridge University Press, Cambridge, 457 pp., 1991.

Dierckx, P., Algorithms for smoothing data on the sphere with tensor product splines, Computing, 32, 319-42, 1984.

Dierckx, P., The spectral approximation of bicubic splines on the sphere, SIAM J. Sci. Stat. Comput., 7, 611-623, 1986.

Diggle, P. and N. Fisher, Sphere - A contouring program for spherical data, Comp. & Geosci., 11, 6, 725-66, 1985.

Diggle, P., N. Fisher, and A. J. Lee, A comparison of tests of uniformity of spherical data. Austral. J. Statist., 27, 53-59, 1985.

Drezner, H. Z., On location dominance on spherical surfaces, Operations Research, 29, 1218-9, 1981.

Drezner, H. Z., Constrained location problems in the plane and on a sphere, IIE Transactions, 15, 300-4, 1983.

Drezner, H. Z., A solution to the Weber location problem on a sphere, J. Operat. Res. Soc., 36, 333-4, 1985.

Drezner, Z. and G. O. Wesolowsky, Facility location on a sphere, J. Oper. Res. Soc., 29, 997-1004, 1978.

Drezner, Z. and G. O. Wesolowsky, Minimax and maximin facility location problems on a sphere, Nav. Res. Logist. Quart., 30, 305-12, 1983.

Dutton, G., Geodesic modeling of planetary relief, Cartographica, 21, 188-207, 1984.

Fekete, G., Rendering and managing spherical data with sphere quadtrees, in Proceedings, Visualization '90, San Francisco, CA, 1990.

Fekete, G. and L. S. Davis, Property spheres: A new representation for 3-D object recognition, 1984.

Fekete, G. and L. Treinish, Sphere quadtrees: a new data structure to support the visualization of spherically distributed data, in SPIE Vol. 1259 Extracting meaning from complex data: Processing, display, interaction, 242-253, 1990.

Fisher, N. I., Spherical medians, J. Roy. Stat. Soc. B, 47, 342-8, 1985.

Fisher N. I. and P. Hall, New statistical methods for directional data--I. Bootstrap comparison of mean directions and the fold test in palaeomagnetism, Geophys. J. Int., 101, 305-313.

Fisher, N. I. and A. J. Lee, Correlation coefficients for random variables on a unit sphere or hypersphere, Biometrika, 73, 159-164, 1986.

Fisher, N. I. and T. Lewis, Estimating the common mean direction of several circular or spherical distributions with differing dispersions, Biometrika, 70, 333-341, 1983.

Fisher, N. I., T. Lewis, and B. Embleton, Statistical Analysis of Spherical Data, Cambridge University Press, Cambridge, 329 pp., 1987.

Fisher, S. R., Dispersion on a sphere, Proc. Roy. Stat. Soc. A, 217, 295-305, 1953.

Freeden, W., On spherical spline interpolation and approximation, Math. Meth. Appl. Sci., 3, 551-75, 1981.

Freeden, W. and T. Gervens, Vector spherical spline interpolation - Basic theory and computational aspects, Math. Meth. Appl. Sci., 16, 151-183, 1993.

Getis A. and B. Boots, Models of Spatial Processes, Cambridge University Press, Cambridge, 198 pp., 1978.

Giacaglia, G. E. O. and C. A. Lundquist, Sampling functions for geophysics, SAO Special Report No. 344, Smithsonian Institution, Astrophysical Observatory, Cambridge, Mass., 1972.

Goodchild, M., The issue of accuracy in global databases in Building Databases for Global Science, H. Mounsey, ed., Taylor & Francis, 1988.

Goodchild, M. F. and Y. Shiren, A hierarchical spatial data structure for global geographic information systems. CVGIP: Graphical Models and Image Processing 54, 31-44, 1992.

Gray, N. H., P. A. Geiser, and J. R. Geiser, On the least-squares fit of small and great circles to spherically projected orientation data, Math. Geol., 12, 173-185, 1980.

Haines, G. V., Spherical cap harmonic analysis, J. Geophys. Res., 90, B3, 2583-2591, 1985.

Haines, G. V., Computer programs for spherical cap harmonic analysis of potential and general fields, Computers & Geosciences, 14, 413-447, 1988.

Hall, P., G. S. Watson, and J. Cabrera, Kernel density estimation with spherical data, Biometrika, 74, 751-62, 1987.

Hanson, B., K. Klink, K. Matsuura, S. M. Robeson, and C. J. Willmott, Vector correlation: Review, exposition, and geographic application, Annals Assoc. Amer. Geog., 82, 103-116, 1992.

Hodgson, M. E., Rapid k-neighbor identification for the sphere, Cart. Geog. Info. Sys., 19, 165-74, 1992.

Jupp P. E. and J. T. Kent, Fitting smooth paths to spherical data, Appl. Statist., 36:34-46, 1987.

Jupp P. E. and K. V. Mardia, A general correlation coefficient for directional data and related

regression problems, Biometrika, 67, 163-73, 1980.

Jupp P. E. and K. V. Mardia, A unified view of the theory of directional statistics, 1975-1988, International Statistical Review, 261-294, 1989.

Katz, I. N. and L. Cooper, Optimal location on a sphere, Computers Math. Applic., 6, 175-196, 1980.

Kaula, W., Theory of statistical analysis of data distributed over a sphere, Rev. Geophys., 1, 83-107, 1967.

Kelker, D. and C. W. Langenberg, A mathematical model for orientation data from macroscopic conical folds, Math. Geol., 14, 289-307, 1982.

Kent, J. T., The Fisher-Bingham distribution on the sphere, J. R. Statist. Soc. B, 44, 71-80, 1982.

Kim, P. T., Decision theoretic analysis of spherical regression, J. Multivariate Analysis, 38, 233-240, 1991.

Kimerling, A. J., Area computation from geodetic coordinates on the spheroid, Surveying and Mapping, 4, 343-351, 1984.

Koppelt, U. and M. Biegel, Spherical harmonic expansion of the continents and oceans distribution function, Studia Geoph. et Geod., 33, 315-21, 1989.

Lawson, L. C^1 surface interpolation for scattered data on a sphere, Rocky Mtn. J. Math., 14, 177-202, 1984.

Litwhiler, D. W., Large region location problems, Ph.D. thesis (University of Oklahoma, Norman, OK), 1977.

Litwhiler, D.W. and A.A. Aly, Large region location problems, Comput. and Op. Res., 6, 1-12, 1979.

Liu, R. Y. and K. Singh, Ordering directional data: Concepts of data depth on circles and spheres, Annals of Statistics, 20, 1468-1484, 1992.

Lo, J. T. and L. R. Eshleman, Exponential Fourier densities on S2 and optimal estimation and detection for directional processes, IEEE Trans. Info. Theor., IT-23, 321-339, 1977.

Lukatela, H., Hipparchus geopositioning model: An overview, Proceedings, AutoCarto 87-96, ASPRS & ACSM, Baltimore, 1987.

Maling, D. H., Coordinate Systems and Map Projections, Pergamon Press, Oxford, 476 pp., 1992.

Mancktelow, N. S., A least squares method for determining the best-fit point maximum, great circle, and small circle to nondirectional orientation data, Math. Geol., 13, 507-521, 1981.

Mardia, K. V., Statistics of Directional Data, Academic Press, New York, 357 pp, 1972.

Mardia, K. V., Statistics of directional data (with discussion), J. R. Statist. Soc. B, 37, 349-393,

1975.

Mardia, K. V., Directional statistics in geosciences, Commun. Statist.-Theor. Meth., A10, 1523-1543, 1981.

Mardia, K. V. and R. J. Gadsden, A small circle of best fit for spherical data and areas of vulcanism, J. Roy. Statist. Soc., C, 26, 238-245, 1977.

Mark, D. M. and J. P. Lauzon, Approaches for quadtree-based geographic information systems at continental or global-scales, Proceedings, Auto-Cart 7, 355-364, 1985.

Miles, R. E., Random points, sets and tessellations on the surface of a sphere, Sankhya, 33A,145-174, 1971.

Moritz, H., Advanced Physical Geodesy, Abacus Press, Tunbridge Wells, Kent, 500 pp., 1980.

Mounsey H. and R. Tomlinson eds., Building Databases for Global Science, Taylor and Francis, London, 419 pp., 1989.

Nielson, G. M. and R. Ramaraj, Interpolation over a sphere based upon a minimum norm network, Computer Aided Design, 4, 41-57, 1987.

Okabe, A., B. Boots, and K. Sugihara, Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, John Wiley & Sons, Chichester, 532 pp., 1992.

Orsingher, E., Shot noise fields on the sphere, Bollettino U. M. I., (6) 3-B, 477-96, 1984.

Prentice, M. J., Spherical regression on matched pairs of orientation statistics, J. Roy. Statist. Soc., B, 51, 241-248, 1989.

Renka, R., Interpolation of data on the surface of a sphere, ACM Trans. Math Softw., 10:417-436, 1984.

Rivest, L.-P., The bivariate Fisher-von Mises distributions, Ann. Statist., 17, 307-317, 1989.

Robeson, S. M. and C. J. Willmott, Spherical spatial interpolation and terrestrial air temperature variability, in Second International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modeling, Breckenridge, Colorado, September, 1993.

Robinson, A. H., Circle of best fit, Geographical Review, 53, 115-116, 1963.

Schott, J. J., J. C. Turlot, and J. Thomann, The fit of polar wander curves by small circles using Marquardt's algorithm: statistical properties, Geophys. J. Int., 105, 731-745, 1991.

Scott, D. and C. A. Tout, Nearest neighbour analysis of random distributions on a sphere, Mon. Notes Roy. Astr. Soc., 241, 109-117, 1989.

Sneyers, R. and J. V. Isacker, A generalized circular distribution, in Statistical Climatology. Developments in Atmospheric Science, 13, S. Ikeda et al., ed., Elsevier, pp. 21-24, 1980.

Stephens, M. A., Vector correlation, Biometrika, 66:41-48, 1979.

Swarztrauber, P., The direct solution of the discrete Poisson equation on the surface of a sphere, J. Comput. Phys., 15, 46-54, 1974.

Swarztrauber, P., On the spectral approximation of discrete scalar and vector functions on the sphere, SIAM J. Num. Anal., 16,934-949, 1979.

Swarztrauber, P., The approximation of vector functions and their derivatives on the sphere, SIAM J. Numer. Anal., 18, 934-949, 1981.

Swarztrauber, P. The vector harmonic transform for solving partial differential equations in spherical geometry, Month. Weath. Rev., 121, 3415-3437, 1993.

Tobler, W., Preliminary representation of world population by spherical harmonics, Proceedings of the Naitonal Academy of Sciences of the USA, 89:6262-64, 1992.

Tobler, W., Three presentations on geographical analysis and modeling, National Center for Geographic Information and Analysis Technical Report 93-1, 1993.

Tobler, W., Smooth pycnophylactic interpolation on the sphere, unpublished manuscript, 1994.

Tobler, W. R. and Z. Chen, A quadtree for global information storage, Geographical Analysis 18, 360-71, 1986.

Traversoni, L., Delaunay's tetrahedronalization; an efficient algorithm for 3D triangulation, SPIE Vo. 1251 Curves and Surfaces in Computer Vision and Graphics, 1990.

Turner, S. M., Weighting irregularly-spaced data points on the sphere for use in spatial analysis, M.S. Thesis, University of Delaware, 1986.

Upton G. J. G. and B. Fingleton, Spatial Data Analysis by Example, Volume 2: Categorical and Directional Data, Wiley & Sons, Chichester, 1989.

Wahba, G., Spline interpolation and smoothing on the sphere, SIAM J. Sci. Stat. Comp., 2:5-16 1981.

Wahba, G., Vector splines on the sphere, with applications to the estimation of vorticity and divergence from discrete, noisy data, in Multivariate Approximation Theory II, W. Schempp and K. Zeller, eds., Birkhauser Verlag, Basel, pp. 407-429, 1982.

Watson, G., Statistics on Spheres, J. Wiley, New York, 238 pp., 1983.

Wendell, R. E., Some aspects in the theory of location, Ph.D. thesis (Northwestern University, Evanston, IL), 1971.

Wesolowsky, G. O., Location problems on a sphere, Reg. Sci. & Urban Econ., 12, 495-508, 1983.

White, D., J. Kimberling, and W. Overton, Cartographic and geometric components of a global sampling design for environmental monitoring, Cartog. & Geog. Inf. Sys., 19, 5-22, 1992.

Wickman, F. E., E. Elvers, and K. Edvarson, A system of domains for global sampling problems, Geografiska Annaler, 56A, 201-212, 1974.

Willmott, C., C. Rowe, W. Philpot, Small-scale climate maps - A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring, Amer. Cartog., 12, 5-16, 1985.

Wood, A. T. A., The simulation of spherical distributions in the Fisher-Bingham family, Commun. Statist. -Simula., 16, 885-898, 1987.

Wood, A. T. A., Some notes on the Fisher-Bingham family on the sphere, Commun. Statist. -Theory Meth., 17, 3881-3897, 1988.

Woodcock N. H., Specification of fabric shapes using an eigenvalue method, Geol. Soc. Amer. Bull., 88, 1231-6, 1977.

Woodcock N. H. and M. A. Naylor, Randomness testing in three-dimensional orientation data, J. Struct. Geol., 5, 539-48, 1983.

Xue, G.-L., A globally convergent algorithm for facility location on a sphere, Comp. Math. Appl., 27, 37-50, 1994.

Zhen-Chang-An, Shi-Zhuang-Ma, Dong-Hai-Tan, Barraclough, -D. -R, and Kerridge, -D. -J., A spherical cap harmonic model of the satellite magnetic anomaly field over China and adjacent areas, J. Geomag. & Geoelec., 44, 243-52, 1992.