Some linear model techniques for analyzing small circle spherical data *

Louis-Paul Rivest
Département de mathématiques et de statistique
Université Laval, Ste-Foy
Québec G1K 7P4
email: lpr@mat.ulaval.ca
FAX: 418-656-2817
tel.: 418-656-7353

Running title: Small circle spherical data
Accepted for publication in January 99

*The support of the Natural Sciences and Engineering Research Council of Canada and of the Fonds pour la formation de chercheurs et l'aide à la recherche du Gouvernement du Québec is gratefully acknowledged.

Some linear model techniques for analyzing small circle spherical data

Key words and phrases: Conical folds; Directional data; Fisher distribution; Girdle distribution; Rotationally symmetric distribution; Spherical ellipse; Watson distribution.
AMS 1991 subject classifications: Primary 62H11, secondary 62J05 62M10.

Abstract

The author investigates least squares as a method for fitting small circle models to a sample of unit vectors in $\mathbb{R}^3$. He highlights a local linear model underlying the estimation of the circle's parameters. This model is used to construct an estimation algorithm and regression type inference procedures for the circle's parameters. It makes it possible to compare the fit of a small circle with that of a spherical ellipse. The limitations of the least squares approach are emphasized: when the errors are bounded away from 0, the least squares estimators are not consistent as the sample size goes to infinity. Two examples, concerned with the migration of elephant seals and with the classification of geological folds, are analyzed using the linear model techniques proposed in this work.

Résumé

L'auteur étudie l'ajustement d'un petit cercle à la surface de la sphère unité à un échantillon de vecteurs unitaires de $\mathbb{R}^3$ à l'aide de la méthode des moindres carrés. Il met en lumière un modèle linéaire local sous-jacent à l'estimation des paramètres du cercle. Ce modèle est utilisé pour construire un algorithme pour le calcul des estimations de même que des tests et des intervalles de confiance concernant les paramètres du cercle. Il permet également de déterminer si une ellipse sphérique donne un meilleur ajustement qu'un petit cercle. L'auteur montre également que la méthode des moindres carrés ne s'applique qu'à des échantillons où les vecteurs sont très proches du cercle échantillonnal. Si les erreurs expérimentales sont grandes, les estimateurs des moindres carrés ne sont pas convergents lorsque la taille d'échantillon tend vers l'infini. Deux exemples, portant sur la migration des phoques et sur la classification des plis géologiques, sont traités à l'aide des méthodes des moindres carrés mises de l'avant dans cet article.

1 Introduction

Let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ be a sample of unit vectors collected in S^2, the unit sphere in $I\!R^3$. The distributions of Fisher (Watson, 1983) and of Kent (1982) provide convenient statistical models when the data points are clustered around a modal direction (Jupp and Mardia, 1989). This paper focuses on situations where the sample points are scattered around a circle on the surface of the unit sphere. When the circle has radius 1, one gets a so-called girdle, or equatorial, sample while radiuses less than 1 yield small circle samples.

Let $\mathbf{e}_1 = (1, 0, 0)'$ represent the north pole of the unit sphere. As the longitude ω goes from 0 to 2π, the unit vector $(\cos\theta, \sin\theta\cos\omega, \sin\theta\sin\omega)'$ goes around a circle at colatitude θ, $\pi \geq \theta \geq 0$, on the surface of the unit sphere. Since this circle lies in a plane orthogonal to the north pole, the direction of this circle is said to be $(1, 0, 0)'$. Small circles with an arbitrary direction μ are obtained by premultiplying the above unit vector by $\mathbf{M}(\mu) = [\mu; \mu_{(\cdot)}]$, a 3×3 rotation matrix whose first column is equal to unit vector μ, and where $\mu_{(\cdot)} = [\mu_{(1)}; \mu_{(2)}]$ is a 3×2 matrix containing an orthogonal complement to μ. This yields the following expression for a circle at colatitude θ with direction μ,

$$\mathcal{C}(\theta, \mu) = \left\{ \mathbf{u}(\omega) = \cos(\theta)\mu + \sin(\theta)\mu_{(\cdot)} \begin{pmatrix} \cos\omega \\ \sin\omega \end{pmatrix} : \omega \text{ in } (0, 2\pi) \right\}. \tag{1}$$

A great circle obtains by setting $\theta = \pi/2$.

In small circle data sets, $\mathbf{v}_i$ is, for $i = 1, \ldots, n$, equal to $\mathbf{u}(\omega_i)$, a unit vector in $\mathcal{C}(\theta, \mu)$, perturbed by experimental errors. The ω_i's are, for $i = 1, \ldots, n$, unknown angles representing the coordinates, on $\mathcal{C}(\theta, \mu)$, of the modal values for the $\mathbf{v}_i$'s. The aim of the analysis is to estimate the circle's parameters, θ and μ. The models for the analysis of such data assume that the ω_i's are uniformly distributed on $(0, 2\pi]$. Examples are provided by the models of Watson (1965), Mardia and Gadsden (1977), and Bingham and Mardia (1978). As exemplified in Sections 6 and 7, the uniformity assumption is seldom tenable. This work develops inference procedures that do not rely on this hypothesis. The ω_i's are instead regarded as unknown nuisance parameters.

The estimates are defined as the values of $(\theta, \mu, \omega_i, i = 1, \ldots, n)$ minimizing the least

squares criterion $\sum \|\mathbf{v}_i - \mathbf{u}(\omega_i)\|^2$, where $\|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{x}$ denotes the squared Euclidean length. Since $\|\mathbf{v}_i\| = \|\mathbf{u}(\omega_i)\| = 1$, the least squares estimates maximize the sum of the cosines between observed and predicted values, $\sum \mathbf{v}_i'\mathbf{u}(\omega_i)$. They are maximum likelihood estimates when the scatter of $\mathbf{v}_i$ around $\mathbf{u}(\omega_i)$ is assumed to follow a Fisher distribution for $i = 1, \ldots, n$. Least squares estimating equations and several variants have been studied in the geological literature; see Cruden and Charlesworth (1972), Gray, Geisser, and Geisser (1980), Stockmal and Spang (1982) and Kelker and Langenberg (1982, 1987).

Sections 2, 3 and 4 highlight a local linear model underlying the estimation of (θ, μ). Section 2 gives a simple least squares algorithm for estimating the circle's parameters. Section 3 shows that when the errors in the model are small (this corresponds to the large concentration asymptotics of directional data), the distribution of the estimators of the circle's parameters is determined by the local linear model. Section 4 proposes a simple method for fitting a spherical ellipse, a five-parameter generalization to $\mathcal{C}(\theta, \mu)$. Section 5 investigates the large-sample behavior of the least squares estimators when the errors are bounded away from 0. The large number of nuisance parameters ω_i makes least squares estimators inconsistent. Sections 6 and 7 present data analyses. All the proofs are given in an appendix.

2 An algorithm for fitting the small circle model

This section constructs a least squares algorithm to calculate the estimates of the circle's parameters (θ, μ). Several algorithms have already been proposed for this problem; see, for instance Fisher, Lewis and Embleton (1987) p. 141, Gray et al. (1982), and Stockman and Spang (1982). The one proposed here is relatively simple; it is based on the design matrix $\mathbf{X}$ and the residual vector $\mathbf{r}$ of a linear model underlying the estimation of the circle's parameters.

The following notation is used throughout the paper. To any 3×1 unit vector μ, $\mu \neq -\mathbf{e}_1$, one associates the following rotation matrix (this is a rotation of angle π around axis $\mathbf{e}_1 + \mu$

where $\mathbf{e}_1$ is the north pole),

$$\mathbf{M}(\mu) = \frac{(\mathbf{e}_1 + \mu)(\mathbf{e}_1 + \mu)'}{1 + \mathbf{e}_1'\mu} - \mathbf{I}_3,$$

where $\mathbf{I}_3$ denotes the 3×3 identity matrix. One has $\mathbf{M}(\mu) = [\mu; \mu_{(\cdot)}]$ where $\mu_{(\cdot)}$ is a 3×2 matrix containing an orthogonal complement to μ. Unit vectors μ^* in a neighborhood of μ can be expressed as

$$\mu^* = \mathbf{M}(\mu) \begin{pmatrix} (1 - \gamma'\gamma)^{1/2} \\ \gamma \end{pmatrix} = \left(1 - \frac{\gamma'\gamma}{2}\right) \mu + \mu_{(\cdot)}\gamma + o(\gamma'\gamma), \tag{2}$$

where γ is a 2×1 vector whose components are small.

In the notation of (1), the least squares estimates are obtained by maximizing

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{v}_i'\mathbf{u}(\omega_i) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \cos(\theta)\mathbf{v}_i'\mu + \sin(\theta)\mathbf{v}_i'\mu_{(\cdot)} \begin{pmatrix} \cos\omega_i \\ \sin\omega_i \end{pmatrix} \right\}.$$

For fixed values of θ and μ, the least squares estimate of ω_i is easily derived. One has

$$\mathbf{v}_i'\mu_{(\cdot)} \begin{pmatrix} \cos\omega_i \\ \sin\omega_i \end{pmatrix} \leq \{1 - (\mathbf{v}_i'\mu)^2\}^{1/2},$$

and the angle that yields this maximum value is

$$\omega_i = \operatorname{atan}(\mu_{(2)}'\mathbf{v}_i, \mu_{(1)}'\mathbf{v}_i), \tag{3}$$

where $\operatorname{atan}(a, b)$ is the angle ω in $[0, 2\pi)$ such that $\cos\omega = b/(a^2 + b^2)^{1/2}$ and $\sin\omega = a/(a^2 + b^2)^{1/2}$. Estimates for θ and μ are obtained by maximizing

$$\bar{R}(\theta, \mu) = \frac{1}{n} \sum_{i=1}^{n} \left[\cos(\theta)\mathbf{v}_i'\mu + \sin(\theta)\{1 - (\mathbf{v}_i'\mu)^2\}^{1/2} \right].$$

This expression can be rewritten as $\sum \cos\{\theta - \arccos(\mathbf{v}_i'\mu)\}/n$. This is the loss function of Cruden and Charlesworth (1972) and Mardia and Gadsden (1977). Other authors, for instance, Gray, Geisser, and Geisser (1980), Stockmal and Spang (1982) and Kelker and Langenberg (1982, 1987), suggested minimizing $\sum\{\theta - \arccos(\mathbf{v}_i'\mu)\}^2/n$. Since for small values of x, $\cos(x) \approx 1 - x^2/2$, these two proposals lead to similar estimates, especially when the data vectors are concentrated around the modal circle.

A simple algorithm uses the second order Taylor series expansion of $\bar{R}(\theta^*, \mu^*)$ with respect to θ and μ. The expansion is obtained by differentiating $\bar{R}(\theta^*, \mu^*)$ with respect to θ and γ twice, where (2) gives the relationship between γ and μ. As shown in the appendix, if $\beta = (\theta^* - \theta, \gamma')'$, one has

$$\bar{R}(\theta^*, \mu^*) \approx \bar{R}(\theta, \mu) + \bar{R}_1(\theta, \mu)\beta + \beta' \bar{R}_{11}(\theta, \mu)\beta/2.$$

The 3×1 vector of the first order terms is given by, $\bar{R}_1(\theta, \mu) = \sum \mathbf{X}_i r_i/n$ where $\mathbf{X}_i$ is a 3×1 vector equal to

$$\mathbf{X}_i = \begin{pmatrix} 1 \\ \dfrac{\mathbf{v}_i' \mu_{(\cdot)}}{\{1 - (\mathbf{v}_i'\mu)^2\}^{1/2}} \end{pmatrix}. \tag{4}$$

and r_i is the residual for the i^{th} data point,

$$r_i = \{1 - (\mu' \mathbf{v}_i)^2\}^{1/2} \cos(\theta) - \mu' \mathbf{v}_i \sin(\theta). \tag{5}$$

Note that from (3), one has $\mathbf{X}_i = (1, \cos\omega_i, \sin\omega_i)'$. The residual r_i has a geometric interpretation; it is the sine of the difference of the angle between $\mathbf{v}_i$ and μ minus θ. It is shown in the appendix that the 3×3 matrix of the second order terms, $\bar{R}_{11}(\theta, \mu)$, is approximately equal to $-\mathbf{X}'\mathbf{X}/n$ where $\mathbf{X}$ is the $n \times 3$ matrix whose ith row is given by $\mathbf{X}_i'$. This leads to the following approximation,

$$\bar{R}(\theta^*, \mu^*) \approx \bar{R}(\theta, \mu) + \mathbf{r}'\mathbf{r}/2 - (\mathbf{r} - \mathbf{X}\beta)'(\mathbf{r} - \mathbf{X}\beta)/2,$$

where $\mathbf{r}$ is vector of the r_i's. The regression of $\mathbf{r}$ on $\mathbf{X}$ gives a suitable β-value for updating the estimates.

For a fixed μ, the least squares estimate of θ can be calculated explicitly. One has $\bar{R}(\theta, \mu) \leq \bar{R}(\theta_\mu, \mu)$ where

$$\theta_\mu = \operatorname{atan}\left(\sum_{i=1}^{n} \{1 - (\mathbf{v}_i'\mu)^2\}^{1/2}, \sum_{i=1}^{n} \mathbf{v}_i'\mu \right).$$

This suggests the following algorithm for estimating the circle's parameters:

- Take as starting value for μ the eigenvector corresponding to the smallest eigenvalue of $\sum(\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})'$; this preliminary estimator for μ is discussed in Section 3.3;
- Let $\theta_k = \text{atan}(\sum_{i=1}^n \{1 - (\mathbf{v}_i'\mu_{k-1})^2\}^{1/2}, \sum_{i=1}^n \mathbf{v}_i'\mu_{k-1})$;
- To update μ at iteration k, let $\mathbf{X}_k$ and r^k be respectively the current $\mathbf{X}$-matrix and residual vector, calculated with (θ_k, μ_{k-1}). Let γ_k be the 2×1 vector of the second and third components of $(\mathbf{X}_k'\mathbf{X}_k)^{-1}\mathbf{X}_k'r^k$, then take $\mu_k = (1 - \gamma_k'\gamma_k)^{1/2}\mu_{k-1} + \mu_{k-1(\cdot)}\gamma_k$;
- Test for convergence and go to iteration $k + 1$ if necessary.

When the data are concentrated around the modal circle, the proposed algorithm converges very rapidly since it is almost a Newton-Raphson algorithm. As a by-product, it provides the matrix $\mathbf{X}$ and the residual vector $\mathbf{r}$ that are used in the next section to derive statistical procedures for the circle's parameters.

3 Large concentration asymptotics in the small circle model

The exact sampling distribution of the least squares estimators obtained in Section 2 is hard to evaluate. This section derives approximate distributions that are valid when the sampling errors are small and when the sample size n is fixed. If there were no sampling errors, one would have $\mathbf{v}_i = \mathbf{u}(\omega_i)$ for $i = 1, \ldots, n$ and the least squares estimators would be exactly equal to the parameter values: $\hat{\theta} = \theta$ and $\hat{\mu} = \mu$. When the sampling errors are small, or $O(\kappa^{-1/2})$, $\hat{\theta} - \theta$ is $O(\kappa^{-1/2})$, $\hat{\mu} = (1 - \hat{\gamma}'\hat{\gamma})^{1/2}\mu + \mu_{(\cdot)}\hat{\gamma}$ where $\hat{\gamma}$ is $O(\kappa^{-1/2})$, and κ indexes the magnitude of the errors. Simple approximations to $\hat{\theta} - \theta$ and $\hat{\gamma}$ in terms of linear functions of the sampling errors are derived in this section. These approximations involve the local linear model introduced in Section 2.

To introduce errors in the model, it is convenient to work in the tangent plane to the unit sphere at $\mathbf{u}(\omega_i)$. A convenient basis for this 2-dimensional vector space consists of

$\mathbf{M}(\mu)(-\sin\theta, \cos\theta\cos\omega_i, \cos\theta\sin\omega_i)'$ and $\mathbf{M}(\mu)(0, -\sin\omega_i, \cos\omega_i)'$. These are two orthogonal unit vectors orthogonal to $\mathbf{u}(\omega_i)$ such that $\mathbf{M}(\mu)(-\sin\theta, \cos\theta\cos\omega_i, \cos\theta\sin\omega_i)'$ is orthogonal to the small circle at $\mathbf{u}(\omega_i)$ while $\mathbf{M}(\mu)(0, -\sin\omega_i, \cos\omega_i)'$ is parallel to it. The errors for the ith data point, $(\epsilon_{i1}, \epsilon_{i2})$, are the coordinates of $\mathbf{v}_i$ in this tangent plane. Thus, one can write

$$\mathbf{v}_i = (1 - \epsilon_{i1}^2 - \epsilon_{i2}^2)^{1/2}\mathbf{u}(\omega_i) + \epsilon_{i1}\mathbf{M}(\mu)\begin{pmatrix} -\sin\theta \\ \cos\theta\cos\omega_i \\ \cos\theta\sin\omega_i \end{pmatrix} + \epsilon_{i2}\mathbf{M}(\mu)\begin{pmatrix} 0 \\ -\sin\omega_i \\ \cos\omega_i \end{pmatrix}. \quad (6)$$

The error ϵ_{i1} represents variability about the circle $\mathcal{C}(\mu, \theta)$ while ϵ_{i2} is related to variability within this circle. Indeed, if $\epsilon_{i1} = 0$ in (6), then $\mathbf{v}_i$ belongs to $\mathcal{C}(\mu, \theta)$. Approximations for $\hat{\theta} - \theta$ and $\hat{\gamma}$ are now derived in terms of terms of linear functions of the $\epsilon'_{ij}s$ which are $O(\kappa^{-1/2})$ unknown quantities.

Let θ^* and μ^* be parameters values in $O(\kappa^{-1/2})$ neighborhoods of θ and μ. Writing $\mu^* = (1 - \gamma'\gamma)^{1/2}\mu + \mu_{(\cdot)}\gamma$, from(6),

$$\mathbf{v}'_i\mu^* = \cos(\theta) + \sin(\theta)\gamma'\begin{pmatrix} \cos\omega_i \\ \sin\omega_i \end{pmatrix} - \sin(\theta)\epsilon_{i1} + O(\kappa^{-1}).$$

Thus,

$$\theta^* - \arccos(\mathbf{v}'_i\mu^*) = (\theta^* - \theta) + \gamma_1\cos(\omega_i) + \gamma_2\sin(\omega_i) - \epsilon_{1i} + O(\kappa^{-1}),$$

where γ_1 and γ_2 are the two components of γ. Therefore one has

$$\begin{aligned} \bar{R}(\theta^*, \mu^*) &= \frac{1}{n}\sum_{i=1}^{n}\cos\{\theta^* - \arccos(\mathbf{v}'_i\mu^*)\} \\ &= 1 - \frac{1}{2n}\sum_{i=1}^{n}\{\epsilon_{1i} - (\theta^* - \theta) - \gamma_1\cos(\omega_i) - \gamma_2\sin(\omega_i)\}^2 + O(\kappa^{-3/2}). \end{aligned}$$

Since $\hat{\theta}$ and $\hat{\mu}$ are the parameter values maximizing $\bar{R}(\theta^*, \mu^*)$, approximate estimators are obtained by maximizing the leading $O(\kappa^{-1})$ term in the expansion. These estimators can be expressed in terms of the regression of the ϵ_{i1}'s on $(\cos\omega_i, \sin\omega_i)$. Let $\mathbf{X}$ be the $n \times 3$ design matrix for this regression, with row i equal to $(1, \cos\omega_i, \sin\omega_i)$ and $\beta = (\theta^* - \theta, \gamma')'$. One has $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon + O(\kappa^{-1})$ and $\bar{R}(\hat{\mu}, \hat{\theta}) = 1 - \epsilon'\{\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}\epsilon/(2n) + O(\kappa^{-3/2})$, where

ϵ is the $n \times 1$ vector of the ϵ_{i1}'s. Thus the large concentration distribution of $\hat{\beta}$ is identical to that of the coefficients of the regression of ϵ on $\mathbf{X}$ while $2n\{1 - \bar{R}(\hat{\mu}, \hat{\theta})\}$ is distributed as the sum of the squared residuals of that regression. An explicit formula for the $O(\kappa^{-1/2})$ approximation to $\hat{\gamma}$ is

$$\hat{\gamma} \approx \left[\sum \begin{pmatrix} \cos\omega_i - m_{\cos} \\ \sin\omega_i - m_{\sin} \end{pmatrix} \begin{pmatrix} \cos\omega_i - m_{\cos} \\ \sin\omega_i - m_{\sin} \end{pmatrix}' \right]^{-1} \sum \begin{pmatrix} \cos\omega_i - m_{\cos} \\ \sin\omega_i - m_{\sin} \end{pmatrix} \epsilon_{i1}, \tag{7}$$

where $m_{\cos}$ and $m_{\sin}$ represent the averages of the $\cos\omega_i$'s and the $\sin\omega_i$'s respectively.

The design matrix of the linear model underlying the fitting of a small circle is not known. This is typical of models for directional data; see Rivest (1998). The last iteration of the estimation algorithm of Section 2 gives a convenient estimate $\hat{\mathbf{X}}$ for $\mathbf{X}$; it also gives a residual vector $\mathbf{r}$ which is approximately distributed as the residual for the local linear model. One has $\hat{\mathbf{X}} = \mathbf{X} + O(\kappa^{-1/2})$ and $\mathbf{r} = \{\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}\epsilon + O(\kappa^{-1})$. These expansions are distribution-free since they do not use any assumption concerning the stochastic behavior of the errors ϵ.

3.1 Independent, Identically Distributed Normal Errors

Assuming that the errors are normal, the standard distributional results of the theory of linear models characterize the large concentration distribution of the least squares estimators of the small circle.

PROPOSITION 3.1. *If, as κ goes to infinity, $\kappa^{1/2}\epsilon$ converges in distribution to a n-variate standardized normal distribution then,*

1. *$2n\kappa\{1 - \bar{R}(\hat{\mu}, \hat{\theta})\}$ converges to a chi-squared distribution with $n - 3$ degrees of freedom;*
2. *$\kappa^{1/2}\hat{\beta}$ converges to $N_3(0, (\mathbf{X}'\mathbf{X})^{-1})$, a three-dimensional normal distribution with mean 0 and variance-covariance matrix $(\mathbf{X}'\mathbf{X})^{-1}$,*
3. *$\bar{R}(\hat{\mu}, \hat{\theta})$ and $(\hat{\mu}, \hat{\theta})$ are asymptotically independent.*

Proposition 3.1 emphasizes that all the statistical techniques developed for linear models can be applied to the small circle model. The assumption of normality can be ascertained with a residual q-q plot. The residual variance can be estimated by $2n\{1-\bar{R}(\hat{\theta},\hat{\mu})\}/(n-3)$ and an estimate of the variance-covariance matrix of $\hat{\beta}$ is

$$(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\frac{2n\{1-\bar{R}(\hat{\theta},\hat{\mu})\}}{n-3}. \tag{8}$$

Consider the hypothesis $H_0:\ \theta=\theta_0$. Using standard linear model techniques, an F-test for this hypothesis is easily constructed. It rejects H_0 if

$$F_{\text{obs}}=\frac{\bar{R}(\hat{\theta},\hat{\mu})-\bar{R}(\hat{\theta}_0,\hat{\mu}_0)}{\{1-\bar{R}(\hat{\theta},\hat{\mu})\}/(n-3)}$$

is large, where $\hat{\mu}_0$ is the least squares estimator of μ when $\theta=\theta_0$, as discussed in Section 3.4. The numerator of F_{obs} can be regarded as the difference between the sums of squared residuals of two nested linear models. Thus, the large concentration distribution of F_{obs} under the null hypothesis is an $F_{1,n-3}$. This F test-statistic has been proposed by Gray et al. (1980) who also suggested that its distribution be approximated by an $F_{1,n-3}$. The derivations presented here provide a theoretical foundation to their proposal.

The large concentration variance-covariance matrix, $v(\hat{\gamma})$, of $\hat{\gamma}$ that governs the distribution of $\hat{\mu}$ is proportional to the inverse of the variance-covariance matrix of $\{(\cos\hat{\omega}_i,\sin\hat{\omega}_i)':\ i=1,...,n\}$, where $\hat{\omega}_i$ is given by (3) with $\mu_{(.)}$ replaced by $\hat{\mu}_{(.)}$. An estimate is easily obtained from (8). Thus, to get a precise estimator of μ, one needs sample points approximately uniformly distributed on the small circle. The more the points are clustered together, the larger the uncertainty concerning μ. A 100(1-α)% confidence region for μ is given by

$$\left\{\mu=(1-\gamma'\gamma)^{1/2}\hat{\mu}+\hat{\mu}_{(.)}\gamma:\ \gamma' v(\hat{\gamma})^{-1}\gamma\leq \mathrm{F}_{2,n-3,\alpha}\right\}.$$

3.2 Correlated errors

It may happen that the data collected at neighboring sites on the circle are correlated. This is exemplified in Section 7. Assuming $\omega_i\leq\omega_{i+1}$, for $i=1,\ldots,n-1$ one expects that the

correlation between ϵ_{i1} and $\epsilon_{(i+j)1}$ decreases with j. Methods for estimating the correlation structure of the errors and for correcting the tests for this correlation are proposed in this section.

The residuals r_i defined in (5) are, up to $O(\kappa^{-1})$ terms, distributed as the residuals from a linear model with design matrix given by $\mathbf{X}$. If $\Sigma(\psi)$ is the $n \times n$ assumed variance-covariance matrix for ϵ, where ψ is a vector-valued unknown parameter, one can estimate ψ using restricted maximum likelihood techniques (see Searle, Casella, and McCulloch, 1992, ch. 6) applied to the r_i's. A sandwich variance estimator for $(\hat{\theta}, \hat{\gamma}')'$ is given by $(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\Sigma(\hat{\psi})\hat{\mathbf{X}}(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}$. It can be used to construct tests and confidence regions adjusted for spatial correlation.

In many applications, a simple autoregressive structure of order 1 is adequate to model the correlation. If the data points are approximately equally spaced on the circle, that is $\omega_{i+1} - \omega_i$ is independent of i, then ρ, the first order autocorrelation of ϵ, characterizes the dependence structure. A suitable estimate is $\hat{\rho}$, the empirical first order autocorrelation calculated on the residuals r_i. An approximate sandwich variance estimator is then given by

$$(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\frac{2n\{1-\bar{R}(\hat{\mu},\hat{\theta})\}(1+\hat{\rho})}{(n-3)(1-\hat{\rho})},$$

where $(1+\hat{\rho})/(1-\hat{\rho})$ is the correction factor for the variance of the sample mean for an autoregressive process of order 1 (Box and Jenkins, 1976, p. 194). The F-test of Section 3.1 can easily be corrected for autocorrelation. The null hypothesis $H_0 : \theta = \theta_0$ is rejected at level α when

$$F_{\text{obs}} = \frac{\{\bar{R}(\hat{\mu},\hat{\theta}) - \bar{R}(\hat{\mu}_0,\theta_0)\}(1-\hat{\rho})}{(1+\hat{\rho})\{1-\bar{R}(\hat{\mu},\hat{\theta})\}/(n-3)}$$

is larger than $F_{\alpha,1,n-3}$, the α-level critical value from an $F_{1,n-3}$ distribution.

3.3 An alternative estimator for μ

Considered as observations in $\mathbb{R}^3$, the data points $\mathbf{v}_i$ are approximately coplanar. They are located about a plane orthogonal to μ at a distance $\cos\theta$ from the origin. If the experimental errors about the circle are all null, then the $\mathbf{v}_i$'s are exactly coplanar. In this case the sample

sum of squares and product matrix, $\sum(\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})'$, has rank 2 and μ is equal to the eigenvector corresponding to its null eigenvalue. When the errors ϵ_{i1} are small, or $O(\kappa^{-1/2})$, the eigenvector, say $\hat{\mu}_s$, corresponding to the smallest eigenvalue of $\sum(\mathbf{v}_i - \bar{\mathbf{v}})(\mathbf{v}_i - \bar{\mathbf{v}})'$ is a simple estimator for μ; see Sheringer (1971) and Shomaker, Waser, Marsh, and Bergman (1959). Its properties are studied in Prentice (1988). This section shows that, in concentrated samples, $\hat{\mu}_s$ and $\hat{\mu}$ are equivalent in the sense that $\hat{\mu}_s - \hat{\mu}$ is $o(\kappa^{-1/2})$.

To prove this result, one can use the large concentration expansion for $\hat{\mu}_s$ established in Rivest (1995). One has

$$\hat{\mu}_s = \mu + \mu_{(\cdot)}\hat{\gamma}_s + O(\kappa^{-1}),$$

where $\hat{\gamma}_s$ can be expressed in terms of a regression similar to (7). It suffices to replace, in (7), the errors ϵ_{i1} and the circle coordinates $(\cos\omega_i, \sin\omega_i)$ by the errors orthogonal to the plane containing the circle, $\sin(\theta)\epsilon_{i1}$, and by the planar coordinates of the sampled points respectively. Taking $\mu\cos\theta$ as the plane origin, these coordinates are given by $\sin\theta(\cos\omega_i, \sin\omega_i)$. The $\sin\theta$'s cancel out and we get (7) as an expansion for $\hat{\gamma}_s$. Thus $\hat{\mu}_s$ and $\hat{\mu}$ have the same first order large concentration distribution. This shows that $\hat{\mu}_s$ is a good starting value for the algorithm of Section 2. Indeed, all the large concentration results of this section apply to the estimates obtained after only one iteration of the algorithm.

3.4 The case of a known θ

Supposing that θ is known and equal to θ_0; the case $\theta_0 = \pi/2$ is of special interest since it gives a great circle, also called a girdle, on the surface on the sphere. All the techniques seen so far apply to the estimation of μ.

The design matrix of the underlying local linear model is $\mathbf{X}_0$, the $n \times 2$ matrix whose ith row is given by $(\cos\omega_i, \sin\omega_i)$. The expression $2n\{1 - \bar{R}(\hat{\mu}_0, \theta_0)\}$ is, to a first order of approximation, equal to the sum of the squared residuals for this linear model. The algorithm of Section 2 is easily modified to calculate $\hat{\mu}_0$, the least squares estimate of μ. It suffices to remove the steps concerned with the estimation of θ and to use $\mathbf{X}_0$ in the calculations. The residuals are given by (5) with θ replaced by the known θ_0. A first order expansion for the $\hat{\gamma}$

characterizing the discrepancy between $\hat{\mu}_0$ and the true value μ is given by (7) with both $m_{\cos}$ and $m_{\sin}$ equal to 0. Thus the variance-covariance matrix of $\hat{\gamma}$ for a great circle fit involves the matrix of second moments of $(\cos\omega_i, \sin\omega_i)$ rather than their variance-covariance that obtains in a small circle fit. This difference is most important when all the data points are clustered together on the modal circle. The great circle estimator of μ is then much more precise than its small circle alternative.

The techniques of Section 3.2 can be used to handle correlated errors in great circle fits. As in Section 3.3, one can construct a simple alternative estimator to $\hat{\mu}_0$ that has the same large concentration distribution. One can show that the eigenvector corresponding to the smallest eigenvalue of $\sum \mathbf{v}_i\mathbf{v}_i'$ is, for concentrated errors, equivalent to $\hat{\mu}_0$.

4 Approximate fitting of a spherical ellipse

The spherical ellipse has been introduced by Kelker and Langenberg (1987) as the intersection of an elliptical cone and the unit sphere. This subset of the sphere can be expressed as

$$\mathcal{C}_e(\theta,\mu) = \left\{ \mathbf{u}_e(\omega) = \cos(\theta_\omega)\mu + \sin(\theta_\omega)\mu_{(\cdot)} \begin{pmatrix} \cos\omega \\ \sin\omega \end{pmatrix} : \omega \text{ in } (0, 2\pi) \right\},$$

where

$$\theta_\omega = \operatorname{arccot}[\{(a^2+c^2)/2 + (a^2-c^2)\cos(2\omega)/2 + b\sin(2\omega)\}^{1/2}].$$

Here, arccot is the inverse of the cotangent function and a, b, and c are unknown parameters satisfying $a^2c^2 > b^2$ (remember that the cotangent of an angle is equal to 1 over its tangent). As ω goes from 0 to 2π, unit vector $\mathbf{u}_e(\omega)$ defined in $\mathcal{C}_e(\theta,\mu)$ goes around the spherical ellipse once. Fitting a spherical ellipse by least squares is no simple matter since, for fixed $\mu, a, b,$ and c, there is no explicit expression, such as (3), for the value of ω_i minimizing $\|\mathbf{v}_i - \mathbf{u}_e(\omega_i)\|^2$. Nevertheless, Kelker and Langenberg (1987) use (3) to project a point on a spherical ellipse; thus their algorithm is not really of the least squares type.

The linear model framework laid out in this paper provides simple techniques to investigate whether a spherical ellipse would fit better than a small circle. When $a^2 - c^2 = b = 0$,

the spherical ellipse reduces to a small circle with angle $\theta = \text{arccot}[\{(a^2+c^2)/2\}^{1/2}]$. Suppose now that $a^2 - c^2$ and b are small or $O(\kappa^{-1/2})$. Expanding in a Taylor series the above expression for θ_ω yields

$$\theta_\omega = \text{arccot}\left\{\left(\frac{a^2+c^2}{2}\right)^{1/2}\right\} - \frac{(a^2-c^2)\cos(2\omega) + 2b\sin(2\omega)}{\{2(a^2+c^2)\}^{1/2}(2+a^2-c^2)} + O(\kappa^{-1}). \tag{9}$$

Let $\delta(\omega)$ denote the $O(\kappa^{-1/2})$ term of this expansion. One has the following approximate expression for $\mathbf{u}_e(\omega)$, a point on the spherical ellipse,

$$\mathbf{u}_e(\omega) = \mathbf{M}(\mu)\begin{pmatrix}\cos\theta\\ \sin\theta\cos\omega\\ \sin\theta\sin\omega\end{pmatrix} + \delta(\omega)\mathbf{M}(\mu)\begin{pmatrix}-\sin\theta\\ \cos\theta\cos\omega\\ \cos\theta\sin\omega\end{pmatrix} + O(\kappa^{-1}).$$

Substituting this approximation for $\mathbf{u}_e(\omega_i)$ in (6) gives the following model for data gathered around a spherical ellipse,

$$\begin{aligned}\mathbf{v}_i = \mathbf{M}(\mu)\begin{pmatrix}\cos\theta\\ \sin\theta\cos\omega_i\\ \sin\theta\sin\omega_i\end{pmatrix} &+ \{\epsilon_{i1} + \delta(\omega_i)\}\mathbf{M}(\mu)\begin{pmatrix}-\sin\theta\\ \cos\theta\cos\omega_i\\ \cos\theta\sin\omega_i\end{pmatrix}\\ &+ \epsilon_{i2}\mathbf{M}(\mu)\begin{pmatrix}0\\ -\sin\omega_i\\ \cos\omega_i\end{pmatrix} + O(\kappa^{-1}).\end{aligned}$$

Thus, a spherical ellipse can be described as a small circle model whose errors about the small circle have a systematic component given by $\delta(\omega)$. In other words, a spherical ellipse improves the fit if the small circle residuals r_i are correlated to $(\cos(2\hat{\omega}_i), \sin(2\hat{\omega}_i))$.

A score test for a spherical ellipse is easily constructed. It suffices to regress r_i on $\hat{\mathbf{X}}_e$, the $n \times 5$ design matrix of $(1, \cos(\hat{\omega}_i), \sin(\hat{\omega}_i), \cos(2\hat{\omega}_i), \sin(2\hat{\omega}_i))$. Let SSE denote the sum of the squared residuals for this regression. The small circle fit would be deficient at level α if

$$\frac{[2n\{1-\bar{R}(\hat{\mu},\hat{\theta})\} - SSE]/2}{SSE/(n-5)} > F_{2,n-5,\alpha}.$$

Estimates for the parameter $(\mu_e, a, b, c,)$ of the spherical ellipse can be calculated using the least squares estimates $(\hat{\beta}_j, j = 1, \ldots, 5)$ from the above fit. One has

$$\hat{\mu}_e = (1 - \hat{\beta}_2^2 - \hat{\beta}_3^2)^{1/2}\hat{\mu} + \hat{\beta}_2\hat{\mu}_{(1)} + \hat{\beta}_3\hat{\mu}_{(2)}$$

and $(\hat{a}, \hat{b}, \hat{c})$ are obtained by solving the equations setting $\hat{\theta}+\hat{\beta}_1 = \text{arccot}[\{(a^2+c^2)/2\}^{1/2}]$ and $(\hat{\beta}_4,\ \hat{\beta}_5)$ equal to the coefficients for $\cos(2\omega)$ and $\sin(2\omega)$ appearing in (9). These solutions may fail to satisfy the constraints $\hat{a}^2 > 0$, $\hat{c}^2 > 0$, and $\hat{c}^2\hat{a}^2 > \hat{b}^2$. To overcome this difficulty, one may replace $\hat{\beta}$ by $\lambda\hat{\beta}$ in the calculations, where λ is the largest value less than or equal to 1 for which the shape parameters calculated with $\lambda\hat{\beta}$ meet the constraints. This method always gives a unique solution since $\lambda = 0$ corresponds to the least squares small circle with $\hat{a} = \hat{c}$ and $\hat{b} = 0$. It will hopefully provide a good approximation to the least squares spherical ellipse whose parameters minimize $\sum \|\mathbf{v}_i - \mathbf{u}_e(\omega_i)\|^2$.

5 Large sample asymptotics

This section studies the behavior of the least squares estimators of the small circle parameters as the sample size goes to infinity, when the errors are bounded away from 0. The least squares estimators are, in this situation, inconsistent. Neymann and Scott (1948) showed that too many nuisance parameters can jeopardize the convergence of the maximum likelihood estimator. For the small circle model, the problem comes from estimating the ω_i's.

Assume that $\mathbf{v}_i$ has a rotationally symmetric distribution with mean direction $\mathbf{u}(\omega_i)$ whose density with respect to the Lebesgue measure on the unit sphere is $f\{\mathbf{v}'\mathbf{u}(\omega_i)\}$. Watson (1983), p. 92, shows that such a distribution is determined by a random variable T defined in $(-1, 1)$, with density $2\pi f(t)$. Indeed, if ϕ is distributed uniformly on $(0, 2\pi]$ independently of T, the general form of a rotationally symmetric distribution around $\mathbf{u}(\omega_i)$ is:

$$\mathbf{v}_i = T\mathbf{u}(\omega_i)+(1-T^2)^{1/2}\cos(\phi)\mathbf{M}(\mu)\begin{pmatrix} -\sin\theta \\ \cos\theta\cos\omega_i \\ \cos\theta\sin\omega_i \end{pmatrix}+(1-T^2)^{1/2}\sin(\phi)\mathbf{M}(\mu)\begin{pmatrix} 0 \\ -\sin\omega_i \\ \cos\omega_i \end{pmatrix}.$$

The least squares estimators for θ and μ converge to the values maximizing

$$E\{\bar{R}(\mu^*, \theta^*)\} = \frac{1}{n}\sum_{i=1}^{n}\left[\cos(\theta^*)E(\mathbf{v}_i'\mu^*) + \sin(\theta^*)E[\{(1-(\mathbf{v}_i'\mu^*)^2)\}^{1/2}]\right].$$

This expectation depends on the distribution of the random variable $\mathbf{v}_i'\mu^*$. Its evaluation is not straightforward. To prove that the least squares estimators are *not* convergent, we focus

on θ and approximate the value θ^* maximizing $E\{\bar{R}(\mu, \theta^*)\}$. In other words we investigate the simpler problem of estimating θ when μ is known.

The above expression for $\mathbf{v}_i$ yields

$$\mathbf{v}_i'\mu = T\cos\theta + (1-T^2)^{1/2}\sin\theta\cos(\phi - \omega_i). \tag{10}$$

The distribution of this quantity does not depend on ω_i since $\phi - \omega_i$ is uniformly distributed on $[0, 2\pi)$ for all values of ω_i. Thus,

$$E\{\bar{R}(\mu, \theta^*)\} = \cos(\theta^*)E(\mathbf{v}_1'\mu) + \sin(\theta^*)E[\{(1-(\mathbf{v}_1'\mu)^2)\}^{1/2}]$$

and $\theta^* = \text{atan}\left[E[\{(1-(\mathbf{v}_1'\mu)^2\}^{1/2}], E(\mathbf{v}_1'\mu)\right]$ maximizes this expression. A Taylor series expansion derived in the appendix shows that:

$$\theta^* = \theta + E(1-T)\,\frac{\cos\theta}{2\sin\theta} + o\{E(1-T)\} \tag{11}$$

for random variables T whose expectation is close to 1. In (11), θ^* is closer to $\pi/2$ than θ. Thus the least squares estimation procedure is biased towards great circles. Note also that the maximum value of $E\{\bar{R}(\mu, \theta^*)\}$ derived in the appendix is equal to $\{E(T)\}^{1/2} + o\{E(1-T)\}$.

To further investigate the bias of the least squares estimators, a small Monte Carlo experiment was carried out. The sample size was fixed at $n = 50$, θ was set equal to $\pi/4$ (0.786) radians and $\mu = (1, 0, 0)'$. Two sets of ω_i's, simulated from wrapped normal distributions with mean resultant length σ_ω equal to .5 and to .75, were used in the study. The mean resultant length σ_ω measures the clustering of the ω_i's; it varies between 0 and 1 with larger values corresponding to clustered points. Taking the expectation with respect to the wrapped normal distribution used for the study, $E(\cos\omega) = \sigma_\omega$ and $E(\sin\omega) = 0$. The sampling errors of $\mathbf{v}$ around $\mathbf{u}(\omega)$ were simulated using the Fisher distribution. The density of $\mathbf{v}$ is then (see Watson, 1983)

$$f(\mathbf{v}|\omega) = \frac{\kappa\exp\{\kappa\mathbf{v}'\mathbf{u}(\omega)\}}{2\pi\{\exp(\kappa) - \exp(-\kappa)\}},$$

where κ is related to the clustering of the data points around the modal circle. Large values of κ give small errors. The least squares estimators are maximum likelihood estimators

	$\sigma_\omega = .5$				$\sigma_\omega = .75$			
$E(T)$	$E(\hat{\theta})$	$E(\hat{\mu}')$			$E(\hat{\theta})$	$E(\hat{\mu}')$		
0.5	1.231	0.985	0.163	-0.05	1.214	0.96	0.278	-0.033
0.75	0.955	0.987	0.147	-0.062	0.912	0.931	0.362	-0.053
0.9	0.836	0.999	0.04	-0.02	0.811	0.987	0.156	-0.025
0.95	0.811	1	0.008	-0.003	0.808	1	0.027	-0.004

Table 1: Simulation results for the least squares estimators ($n = 50$)

for the Fisher distribution. Four values of κ, 1.4, 4, 10 and 20 corresponding to random variables T with expectations given by 0.5, 0.75, 0.9, and 0.95 respectively were used in the simulations. The results appear in Table 1; they are based on Monte Carlo samples of size 10,000. The estimates of $E(\hat{\theta})$ and the first two components of $E(\hat{\mu}')$ have respective Monte Carlo standard errors of 0.05, 0.003, and 0.02. The third components of $E(\hat{\mu}')$ are not significantly different from 0.

The simulations reveal that $\hat{\theta}$ has a huge positive bias that decreases with the clustering of the points around the modal circle. An approximate formula for the bias of $\hat{\theta}$ derived from (11) is $\{1 - E(T)\}/2$. The simulated biases given in Table 1 agree with this approximation when $E(T)$ is bigger than or equal to 0.9. There is also some bias in $\hat{\mu}$: the least squares estimator of μ appears to have been pulled towards $(0, 1, 0)'$. From Watson (1983), p. 92, random unit vectors from the simulated model satisfy $E(\mathbf{v}|\omega) = E(T)\mathbf{u}(\omega)$, where T is as in (10). The unconditional expectation is given by $E(\mathbf{v}) = E(T)(\cos\theta, \sin\theta\sigma_\omega, 0)'$. Thus the least squares estimator of μ appears to be biased towards the mean direction of $\mathbf{v}$, with the bias increasing with σ_ω. As expected, the bias decreases as $E(T)$ increases. In concentrated samples, $\bar{R}(\hat{\mu}, \hat{\theta})$ estimates $\{E(T)\}^{1/2}$; its value provides a rough guideline to investigate whether large concentration asymptotics apply. The simulations suggest that the large concentration inference procedures of Sections 3 and 4 should not be used unless $\bar{R}^2(\hat{\mu}, \hat{\theta})$ is greater than 0.95.

6 Example: The detection of elliptical conical folds

When layered rocks are deformed in geological times, they may have curved surfaces that are also called folds. Folds exhibit different forms; the determination of the form of a fold is based on statistical techniques (Kelker and Langenberg, 1988). The data consist of unit vectors orthogonal to the surface of the fold. The fold is cylindrical if these unit vectors belong, up to experimental errors, to a great circle of the unit sphere. It is circular conical if the sample unit vectors are on a small circle. Kelker and Langengerg (1987) introduce elliptical conical folds as folds whose normal unit vectors belong to a spherical ellipse.

Consider the $n = 78$ data points collected on the Wynd Syncline, Alberta (Charlesworth, Langenberg and Ramsden, 1976). The data appear in Appendix A of Kelker and Langenberg (1987), and are available at http://www.mat.ulaval.ca/pages/lpr/. The dip (d) and dip direction (dd), given in degrees, measure the orientation of the tangent plane to the fold with respect to both the horizontal plane and the north direction, at a sampled location (Fisher et al., 1987, p. 19). They are transformed into unit vectors by $v = (-\cos(d+90)\cos(180+dd), \cos(d+90)\sin(180+dd), -\sin(d+90))'$. For the small circle fit $\bar{R}(\hat{\mu}, \hat{\theta}) = 0.99618$; thus large concentration asymptotics apply to these data. The sum of the squared residuals for the underlying linear model is $2 \times 78 \times \{1 - \bar{R}(\hat{\mu}, \hat{\theta})\} = .596$.

insert Figure 1

The F-test for investigating the fit of the small circle is obtained by regressing the small circle residuals r_i on $\hat{\mathbf{X}}_e$, defined in Section 4. This yields a $F_{2,73}$ statistic of 4.89 (p-value = .01), as compared to 4.38 for Kelker and Langenberg's F-statistics obtained after fitting a spherical ellipse. The residuals are not autocorrelated; these F-tests appear to be valid. To study the fit of the spherical ellipse, one can test whether adding $(\cos(3\hat{\omega}_i), \sin(3\hat{\omega}_i))$ as independent variables improves the model. The $F_{2,71}$ statistic for these two variables is 0.92 (p-value = 0.60). Thus a spherical ellipse is appropriate for this data set.

The unit vector $\hat{\mu} = (0.140, -0.404, -0.904)'$ determines the orientation of the best fitting small circle. The shape parameters for the spherical ellipse estimated with $\hat{\beta}$, the vector of

coefficients of the least squares regression of the small circle residual r_i on $\hat{\mathbf{X}}_e$, are not admissible; the largest λ-value that yields admissible estimates is 0.44. The estimate of the spherical ellipse axis calculated from $0.44\hat{\beta}$ is $\hat{\mu}_e = (.260, -0.436, -0.862)'$ as compared to $(0.278, -0.478, -0.833)'$ obtained by Kelker and Langenberg (1987). The estimates of the shape parameters are $\hat{a} = 0.0437$, $\hat{b} = 0.0075$, and $\hat{c} = 0.2246$

Figure 1 compares the fit of the small circle model to that of the spherical ellipse. The $\mathbf{v}_i$'s are represented by their equal area projection on a plane orthogonal to $\hat{\mu}_e$ (see Fisher, Lewis and Embleton, 1987 p. 36). Each $\mathbf{v}$-vector is first premultiplied by the rotation $\mathbf{M}(\hat{\mu}_e)$, that maps $\hat{\mu}_e$ to the north pole; this yields $\mathbf{w} = \mathbf{M}(\hat{\mu}_e)\mathbf{v}$. The equal area projection of $\mathbf{v}$ is then $(w_2, w_3)'/(1 + w_1)^{1/2}$. If all the directions were on a small circle of colatitude θ orthogonal to $\hat{\mu}_e$, the projected points would lie on a circle of radius $2^{1/2}\sin(\theta/2)$. In Figure 1, the best small circle and spherical ellipse fit are also represented by their equal area projections. The spherical ellipse fits better!

7 Example: The characterization of the migration path of an elephant seal

Consider the $n = 73$ daily measurements of the position of a migrating elephant seal presented in Brillinger and Stewart (1998). This migration is a return trip, from the South Californian coast to locations in the mid-north Pacific. These 73 latitude and longitude values are easily transformed into unit vectors on the unit sphere. Brillinger and Stewart (1998) fit a model suggesting that the elephant seal migrated along a great circle path. The techniques of this paper provide an alternative analysis for these data. The small circle fit gives $\bar{R}(\hat{\theta}, \hat{\mu}) = 1.72 \times 10^{-5}$ and the large concentration methods of this paper apply.

Gray's et al. $F_{1,70}$ statistic for a great circle model is 2.90 (p-value = 0.093); this raises doubts about the validity of the great circle assumption. Since these data are daily seal positions, one may expect to get correlation between the residuals r_i observed at successive days. These residuals are signed distances between the seal's positions and the fitted circle; a

positive correlation between two successive distances is likely. The technique of Section 3.3, with the residuals ordered in time comes in handy to investigate a possible autocorrelation. The first order autocorrelation of r_i is $\hat{\rho} = 0.54$. Correcting the F-test for autocorrelation, as proposed in Section 3.2, yields an $F_{1,70}$ statistic of 0.87 (p-value = 0.65); the great circle assumption is not rejected. The estimated residual standard deviation is $[2 \times 73 \times \{1 - \bar{R}(\pi/2, \hat{\mu})\}/71]^{1/2} = 0.0089$, this corresponds to 57 km on the earth's surface . The range of the estimated latitude $\hat{\omega}_i$ is 0.422; thus the distance between the South Californian coast and the migration's destination is about 2700 km.

To test whether the forward journey (data points 1 to 39) and the return trip (data points 40 to 73) took place on possibly different small circles, one can calculate the following $F_{2,69}$ statistic,

$$\frac{(1-\hat{\rho})[39\{\bar{R}_1(\hat{\theta}_1, \hat{\mu}_1) - \bar{R}_1(\pi/2, \hat{\mu}_{01})\} + 34\{\bar{R}_2(\hat{\theta}_2, \hat{\mu}_2) - \bar{R}_2(\pi/2, \hat{\mu}_{02})\}]}{(1+\hat{\rho})[39\{1 - \bar{R}_1(\hat{\theta}_1, \hat{\mu}_1)\} + 34\{1 - \bar{R}_2(\hat{\theta}_2, \hat{\mu}_2)\}]/69},$$

where $\bar{R}_1$ and $\bar{R}_2$ are obtained by fitting distinct great circles to the outbound and the inbound trips. This F-statistic is 0.63 (p-value = .54). Regressing the great circle residuals on $(\cos\hat{\omega}_i, \sin\hat{\omega}_i, \cos(2\hat{\omega}_i), \sin(2\hat{\omega}_i))$ permits to investigate the fit of elliptical alternatives. The $F_{2,69}$ goodness of fit statistic is 1.55 (p-value = 0.22). These tests reveal that the great circle model cannot be improved upon.

The unit vector orthogonal to the fitted great circle is $\hat{\mu}_0 = (-0.7797, 0.0257, -0.6256)'$, or 38.72 degree of latitude south and 1.47 degree of longitude west. Using the results of Sections 3.1, 3.2, and 3.4, a 95% confidence region for μ can be shown to contain the unit vectors

$$(1-\gamma_1^2-\gamma_2^2)^{1/2}\begin{pmatrix} -0.7797 \\ 0.0257 \\ -0.6256 \end{pmatrix} + \gamma_1\begin{pmatrix} -0.5236 \\ 0.5212 \\ 0.6740 \end{pmatrix} + \gamma_2\begin{pmatrix} -0.3433 \\ 0.8531 \\ 0.3921 \end{pmatrix}$$

such that $85137\,\gamma_1^2 + 1294\,\gamma_2^2 \leq 1$. To validate the approximations underlying the calculations of this confidence region, its real coverage was calculated using 1000 Monte Carlo samples and found to be equal to 88.4%. In the simulations, all the parameters were set equal to their estimates. Thus, the experimental errors were generated form (6), with (ϵ_{i1}) as an

$AR(1)$ sequence with $\rho = 0.54$ and standard error 0.0089 and $\epsilon_{i2} = 0$ for each i. As proposed in Section 3.4, μ was estimated by the eigenvector corresponding to the smallest eigenvalue of $\sum \mathbf{v}_i \mathbf{v}_i'$. The relatively poor coverage of the confidence region might be caused by the estimation of ρ; the first order residual autocorrelation underestimates ρ by 13%.

It is interesting to contrast this analysis to that of Brillinger and Stewart. They do an eye fit of the great circle; this could be improved upon by using the algorithm of Section 2. The $\hat{\omega}_i$'s, properly centered, corresponds to their $\tilde{\theta}_t$'s while the residuals r_i are approximately equal to their $\tilde{\phi}_t$'s. They use the pairs $(\tilde{\theta}_t, \tilde{\phi}_t)$ to estimate the parameters of a diffusion model for elephant seal migration along a great circle and they use residual plots to investigate the great circle assumption. The test statistics for small circle and elliptical alternatives yield a more direct assessment of this assumption.

Acknowledgments

I am grateful to Jérôme Asselin for carrying out the simulations presented in Section 5, to Ted Chang for providing the codes for the simulations of Section 7, and to the Associate Editor and to the Editor for their editorial suggestions.

Appendix

Derivation of the partial derivatives of $\bar{R}(\theta, \mu)$

Partial derivatives with respect to γ can be obtained from the Taylor series expansions for $\mathbf{v}_i'\mu^*$ and $\{1 - (\mathbf{v}_i'\mu^*\}^2)^{1/2}$, the two expressions appearing in $\bar{R}(\theta^*, \mu^*)$ that involve μ^*. The first one is easily derived from (2),

$$\mathbf{v}_i'\mu^* = \mathbf{v}_i'\mu(1 - \gamma'\gamma/2) + \mathbf{v}_i'\mu_{(\cdot)}\gamma + o(\gamma'\gamma)\,.$$

To derive an expansion for $\{1 - (\mathbf{v}_i'\mu^*)^2\}^{1/2}$, let $c_i = \mathbf{v}_i'\mu$ and $a_i = \mathbf{v}_i'\mu_{(\cdot)}$. One has $(\mathbf{v}_i'\mu^*)^2 = c_i^2(1 - \gamma'\gamma) + 2c_i a_i'\gamma + (a_i'\gamma)^2 + o(\gamma'\gamma)$. Thus, using $(1 + x)^{1/2} = 1 + x/2 - x^2/8 + o(x^2)$, one

gets, neglecting $o(\gamma'\gamma)$ terms,

$$\begin{aligned}\{1-(\mathbf{v}_i'\mu^*)^2\}^{1/2} &= \left[(1-c_i^2)\left\{1+\frac{\gamma'\gamma c_i^2}{1-c_i^2}-2\frac{c_i a_i'\gamma}{1-c_i^2}-\frac{(a_i'\gamma)^2}{1-c_i^2}\right\}\right]^{1/2} \\ &= (1-c_i^2)^{1/2}\left\{1-\frac{c_i a_i'\gamma}{1-c_i^2}-\frac{c_i^2(a_i'\gamma)^2}{2(1-c_i^2)^2}-\frac{(a_i'\gamma)^2}{2(1-c_i^2)}+\frac{\gamma'\gamma c_i^2}{2(1-c_i^2)}\right\} \\ &= (1-c_i^2)^{1/2}-\frac{c_i a_i'\gamma}{(1-c_i^2)^{1/2}}-\frac{(a_i'\gamma)^2}{2(1-c_i^2)^{3/2}}+\frac{\gamma'\gamma c_i^2}{2(1-c_i^2)^{1/2}}\,.\end{aligned}$$

This is equivalent to

$$\{1-(\mathbf{v}_i'\mu^*)^2\}^{1/2}=\{1-(\mathbf{v}_i'\mu)^2\}^{1/2}-\frac{\mathbf{v}_i'\mu\;\mathbf{v}_i'\mu_{(\cdot)}\gamma}{\{1-(\mathbf{v}_i'\mu)^2\}^{1/2}}$$

$$-\frac{(\mathbf{v}_i'\mu_{(\cdot)}\gamma)^2}{2\{1-(\mathbf{v}_i'\mu)^2\}^{3/2}}+\frac{\gamma'\gamma(\mathbf{v}_i'\mu)^2}{2\{1-(\mathbf{v}_i'\mu)^2\}^{1/2}}+o(\gamma'\gamma).$$

The partial derivative of $\bar{R}(\theta,\mu)$ with respect to θ is easily evaluated,

$$\frac{\partial\bar{R}(\theta,\mu)}{\partial\theta}=\frac{1}{n}\sum_{i=1}^{n}r_i$$

where the residuals r_i are defined in (5). Combining this with the first order terms in the above expansion yields the vector of the first partial derivatives, $\bar{R}_1(\theta,\mu)=\sum r_i\mathbf{X}_i/n$. Furthermore, the 2×2 matrix of the second order terms involving unit vector μ is given by

$$\begin{aligned}\bar{R}_{11}(\theta,\mu)_{\mu\,\mu}=-\frac{1}{n}\Bigg[&\sin(\theta)\sum_{i=1}^{n}\frac{\mu_{(\cdot)}'\mathbf{v}_i\mathbf{v}_i'\mu_{(\cdot)}}{\{1-(\mathbf{v}_i'\mu)^2\}^{3/2}} \\ &+\sum_{i=1}^{n}\left[\cos(\theta)-\sin(\theta)\frac{\mathbf{v}_i'\mu}{\{1-(\mathbf{v}_i'\mu)^2\}^{1/2}}\right]\mathbf{v}_i'\mu\mathbf{I}_2\Bigg]\,.\end{aligned}$$

The (θ,θ) and (θ,μ) terms of $\bar{R}_{11}(\theta,\mu)$ are obtained by differentiating $\bar{R}_1(\theta,\mu)$ with respect to θ,

$$\begin{pmatrix}\bar{R}_{11}(\theta,\mu)_{\theta\,\theta}\\ \bar{R}_{11}(\theta,\mu)_{\theta\,\mu}\end{pmatrix}=-\frac{1}{n}\sum_{i=1}^{n}\left[\sin(\theta)\{1-(\mathbf{v}_i'\mu)^2\}^{1/2}+\cos(\theta)\mathbf{v}_i'\mu\right]\mathbf{X}_i.$$

When the data are concentrated about the modal circle, one has $\{1-(\mathbf{v}_i'\mu)^2\}^{1/2}\approx\sin\theta$ and $\mathbf{v}_i'\mu\approx\cos\theta$; thus, $\bar{R}_{11}(\theta,\mu)\approx-\sum\mathbf{X}_i\mathbf{X}_i'/n$.□

Proof of the expansion for θ^*

In (10), let $T = 1 - \delta$. One has $E(\mathbf{v}_i'\mu) = \cos(\theta)\{1 - E(\delta)\}$. To evaluate $E[\{1 - (\mathbf{v}_i'\mu)^2\}^{1/2}]$, a useful expansion derived using (10) is

$$\{1 - (\mathbf{v}_i'\mu)^2\}^{1/2} = \sin\theta \left(1 + \delta\frac{\cos^2\theta - \cos^2\phi}{\sin^2\theta} - (2\delta)^{1/2}\frac{\cos\theta\cos\phi}{\sin\theta}\right) + o_p(\delta).$$

Since ϕ is uniformly distributed on $[0, 2\pi)$, $E(\cos\phi) = 0$ and $E(\cos^2\phi) = 1/2$,

$$E[\{1 - (\mathbf{v}_i'\mu)^2\}^{1/2}] \approx \sin\theta \left\{1 + \frac{E(\delta)}{2}\frac{\cos^2\theta}{\sin^2\theta} - \frac{E(\delta)}{2}\right\}.$$

Putting this back into the expression for θ^* yields $\theta^* = \text{atan}[\tan(\theta) + E(\delta)/\{2\sin(\theta)\cos(\theta)\}]$. Expanding the atan in a Taylor series yields the approximation for θ^* presented in Section 5. The maximum value for $E\{\bar{R}(\theta, \mu^*)\}$ is equal to $\{E(T)\}^{1/2} + O\{E(1 - T)\}$. □

REFERENCES

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

Box, G. E. P. and Jenkins, M. J. (1976). *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco.

Brillinger, D. R. and Stewart, B. S. (1998). Elephant seal movements: Modeling migration. *Canadian Journal of Statistics*, **26**, 431-443.

Charlesworth, H.A.K., Langenberg,C. W., and Ramsden, J. (1976). Determining axes, axial planes, and sections of macroscopic folds. *Canadian Journal of Earth Sciences*, **13**, 54-65.

Cruden, D.M. and Charlesworth, H.A.K. (1972). Observations on the numerical determination of axes of cylindrical and conical folds. *Geol. Soc. Amer. Bull.*, **83**, 2019-2024.

Fisher, N. I. Lewis, T. and Embleton, B.J.J. (1987). *Statistical Analysis of Spherical Data.* Cambridge University Press, London.

Gray, N.H., Geiser, P.A. and Geiser, J.R. (1980). On the least squares fit of small and great circles to spherically projected orientation data. *Mathematical Geology*, **12**, 173-184.

Jupp, P.E. and Mardia, K.V. (1989). A unified view of the theory of directional statistics, 1975-1988. *International Statistical Review*, **53**, 261-294.

Kelker, D. and Langenberg, C. W. (1982). A mathematical model for orientation data from macroscopic conical folds. *Mathematical Geology*, **14**, 289-307.

Kelker, D. and Langenberg, C. W. (1987). A mathematical model for orientation data from macroscopic elliptical conical folds. *Mathematical Geology*, **19**, 729-743.

Kelker, D. and Langenberg, C. W. (1988). Statistical classification of macroscopic folds as cylindrical, circular conical, or elliptical conical. *Mathematical Geology*, **20**, 717-730.

Kent, J. T.(1982). The Fisher-Bingham distribution on the sphere. *Journal of the Royal Statistical Society, Series B*, **44**, 71-80.

Mardia, K.V. and Gadsden, R.J. (1977). A circle of best fit for spherical data and areas of volcanism. *Applied Statistics*, **26**, 238-245.

Neymann, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1-32.

Prentice, M. J. (1988). Using principal axes to estimate small circles and ellipses on the sphere. *Journal of Applied Statistics*, **15**, 171-177.

Rivest, L.-P. (1995). Some local linear models for the assessment of geometric integrity. *Statistica Sinica*, **5**, 204-210.

Rivest, L.-P. (1998). Some linear models for the estimation of the motion of rigid bodies with applications to geometric quality assurance. *Journal of the American Statistical Association*, **93**, 632-642.

Searle, S.R., Casella, G., and McCulloch, C. E. (1992). *Variance Components.* Wiley, New York.

Sheringer, C. (1971). A method for fitting a plane to a set of points by least squares. *Acta Cryst.*, B**27**, 1470-1472.

Shomaker, V., Waser, J., Marsh, R. E. and Bergman, G. (1959). To fit a plane or a line to a set of points by least squares. *Acta Cryst.*, **12**, 600-604.

Stockmal, G.S. and Spang, J.H. (1982). A method for the distinction of circular conical from cylindrical folds. *Canadian Journal of Earth Sciences*, **19**, 1101-1105.

Watson, G.S. (1965). Equatorial distributions on a sphere. *Biometrika*, **52**, 193-201.

Watson, G.S. (1983). *Statistics on Spheres. University of Arkansas Lecture Notes in the Mathematical Sciences* **6**. Wiley, New York.

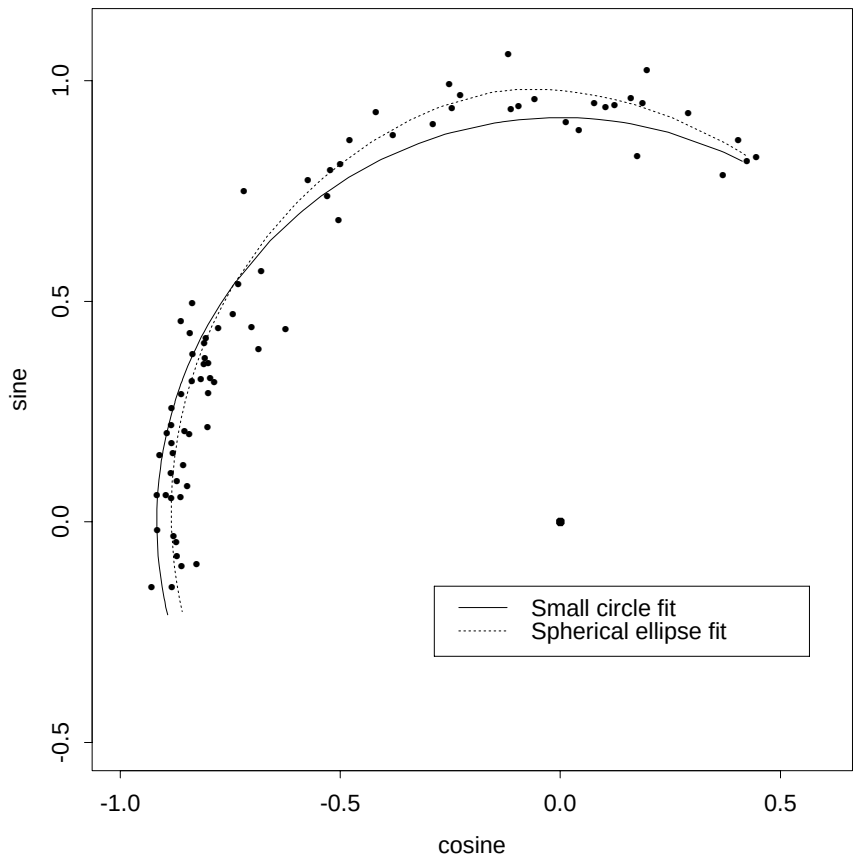

Figure 1: Equal area projection of the Wynd Syncline directions with the origin equal to the projection of $\hat{\mu}_e$, the spherical ellipse axis.