THE ASTROPHYSICAL JOURNAL, 459:779–791, 1996 March 10

INFERRING SPATIAL VARIATION OF SOLAR PROPERTIES FROM HELIOSEISMIC DATA

D. O. GOUGH,[1,2,3] T. SEKII,[2] AND P. B. STARK[4]
Received 1995 July 21; accepted 1995 August 22

ABSTRACT

A common method to infer that solar properties vary with position is to compare linear estimates of averages of those properties centered at different locations. If some of the confidence intervals for the averages do not overlap, one concludes that the property varies. In order for this conclusion to be statistically valid, the lengths of the intervals must be adjusted to obtain the correct "simultaneous coverage probability." We illustrate the notion of simultaneous coverage probability using coin tossing as an example. We present four methods for adjusting the lengths of confidence intervals for linear estimates, and a complementary approach to infer changes based on constructing a linear estimator that is directly sensitive to changes. The first method for constructing simultaneous confidence intervals is based on Bonferroni's inequality, and applies generally to confidence intervals for any set of parameters, from dependent or independent observations. The second method is based on a χ^2 measure of fit to the data, which allows one to compute simultaneous confidence intervals for any number of linear functionals of the model. The third method uses a χ^2 distribution in the space of estimates, which yields "Scheffé" confidence intervals for the functionals. The fourth method, which produces the shortest confidence intervals, uses the infinity-norm in the space of estimates to construct "maximum-modulus" confidence intervals. We apply the four methods to search for radial changes in averages of solar angular velocity, using data from Big Bear Solar Observatory (BBSO) averaged for the 4 yr 1986, 1988–1990. Finally, we apply the new differencing estimator to the BBSO data, finding strong evidence that the average solar angular velocity is lower near the poles than near the equator over a range of depths, as is observed at the surface as well.

Subject headings: methods: statistical — Sun: oscillations — Sun: rotation

1. INTRODUCTION

Backus-Gilbert theory (Backus & Gilbert 1968, 1970; Backus 1970a, b, c) and the SOLA method described by Pijpers & Thompson (1992, 1994) estimate certain weighted spatial averages of some physical property from indirect linear measurements. Often, averages sensitive to different regions in the object of study are compared to try to identify differences in properties between the regions. For example, in both geophysics (e.g., Backus & Gilbert 1970; Parker 1970; Johnson & Gilbert 1972; Masters 1979; Oldenburg 1979, 1981) and astrophysics (e.g., Christensen-Dalsgaard, Schou, & Thompson 1990; Däppen et al. 1991; Gough 1985; Gough & Toomre 1991), estimates of averages centered at different depths are frequently plotted in the same figure, perhaps inviting the reader to compare the estimated values at different depths. A common plotting method is to use upright crosses centered at the depths and values of the estimates, with the lengths of the horizontal bars corresponding to the nominal resolutions, and the lengths of the vertical bars corresponding to the nominal uncertainties. It does not appear to be appreciated generally that the nominal uncertainties must be magnified somewhat for comparisons of different confidence intervals to be valid. In this paper, we illustrate the issue of simultaneous versus individual confidence intervals, and give four methods to compute simultaneous confidence intervals for linear estimates of linear functionals, including "Backus-Gilbert" and SOLA estimates, that can be used to compare estimates of averages centered at different points. We apply the methods to Big Bear Solar Observatory (BBSO) normal-mode frequency estimates averaged over the 4 yr 1986 and 1988–1990 (Woodard & Libbrecht 1993) to test for radial changes in a quantity related to the solar angular-momentum density. Finally, we present a complementary method to infer spatial changes of physical properties, based on tailoring a linear estimate to be sensitive to the change in question, and apply the method to BBSO data to infer that over a range of depths, the Sun rotates more slowly near the poles than near the equator.

2. LINEAR ESTIMATES OF SOLAR ROTATION

The algebra in this section applies generally to linear inverse problems, but we shall use notation particular to the inverse problem of estimating solar angular velocity from eigenfrequency splittings. The physical problem we address is to learn about the distribution of angular velocity in the solar interior, $\Omega(r, \mu)$, where r is radius and μ is the cosine of colatitude. We do so from the deviations of the frequencies of normal modes of the Sun from the degenerate multiplet frequencies $\{\omega_{nl}\}$ at which the Sun would oscillate were it not rotating, neglecting other sources of splitting such as aspherical structure and magnetic fields (n is the principal order, l is the degree, and m is the azimuthal order). Observers usually report these deviations, $\{\Delta\omega_{nlm}\}$, as "a-

[1] Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge, CB3 0EZ, England, UK.
[2] Institute for Astronomy, Madingley Road, Cambridge, CB3 0HA, England, UK.
[3] JILA, Campus Box 440, University of Colorado, CO 80309-0440.
[4] Department of Statistics, and Space Sciences Laboratory, University of California, Berkeley, CA 94720-3860.

coefficients" $\{a_{jnl}\}$, which are estimates of the coefficients in the expansion

$$\frac{\Delta\omega_{nlm}}{2\pi} = L \sum_{j=1}^{2l+1} a_{jnl} P_j\left(\frac{m}{L}\right), \quad (1)$$

where P_j is the Legendre polynomial of degree j. The value of L has been variously chosen to be l, $[l(l+1)]^{1/2}$, and $l+\frac{1}{2}$. In practice, observers estimate a-coefficients only up to some maximum index $j_{\max}$, which is eventually less than $2l+1$. Observers also report standard deviations $\{\sigma_{jnl}\}$ of the estimates of the a-coefficients. Below we use estimates of a-coefficients with $j_{\max} = 12$; to first order in Ω, only the odd-numbered coefficients are related to the angular velocity Ω. It is often asserted that the errors $\{\epsilon_{jnl}\}$ in the estimates of $\{a_{jnl}\}$ are essentially independent and Gaussian-distributed; we shall assume that that is true in our data analysis, but in our derivations we allow arbitrary error correlation.

To first order in the angular velocity Ω, the relation between the angular velocity and the part of the splitting that is odd in m, $\Delta_o \omega_{nlm}$, is the linear integral

$$\frac{\Delta_o \omega_{nlm}}{m} = \int_0^R dr \int_{-1}^{1} d\mu K_{nlm}(r, \mu)\Omega(r, \mu), \quad (2)$$

where the "splitting kernels" are of the form

$$K_{nlm}(r, \mu) \equiv K_{nl}(r)W_{lm}(\mu) + L_{nl}(r)X_{lm}(\mu). \quad (3)$$

Expressions for the components of the right-hand side in terms of the eigenfunctions and density of the reference model are given by Sekii (1993), and in the Appendix. A property of the expansion (3) we shall exploit below is that L_{nl} is small compared with $K_{nl}(r)$ except near the inner turning point of the (n, l) multiplet, and thus

$$K_{nlm}(r, \mu) \approx K_{nl}(r)W_{lm}(\mu). \quad (4)$$

This property was employed by Sekii (1993, 1995) to develop an efficient method to estimate the two-dimensional distribution of angular velocity in the Sun from eigenfrequency splittings.

Let J denote the multi-index (j, n, l) and $\mathcal{J}$ the set of values J can take. We denote the number of elements in the set $\mathcal{J}$ by N. The set $\mathcal{J}$ contains ordered triples with all combinations of values of n and l corresponding to the observed multiplets, but only odd values of j between 1 and $j_{\max}$. We write the relation between the observations and the angular velocity abstractly as follows:

$$\delta_J = K_J\Omega + \epsilon_J, \quad J \in \mathcal{J}. \quad (5)$$

Here $K_J\Omega$ is a linear functional of the angular velocity Ω and may be written as an integral $K_J\Omega = \int K_J(r, \mu)\Omega(r, \mu)$, with data kernel $K_J(r, \mu)$ related to the kernels $K_{nlm}(r, \mu)$ of equation (3) through equation (1)—see Brown et al. (1989) for precise expressions. We deliberately use the same notation for the data kernel and the linear functional derived by integrating against the data kernel; context should obviate possible confusion. The noise terms $\{\epsilon_J\}$ are assumed to be realizations of zero-mean Gaussian random variables with covariance matrix Σ; this notation requires that we impose an arbitrary but fixed ordering on the set $\mathcal{J}$. We abbreviate the N equations (5) using vector notation:

$$\delta = K\Omega + \epsilon. \quad (6)$$

Provided there is no repeated measurement of any a_J, the functionals $\{K_J\}_{J\in\mathcal{J}}$ are linearly independent, a property we shall use below.

Let Ψ denote an arbitrary angular-velocity distribution. For any $\lambda = (\lambda_J)_{J\in\mathcal{J}}$ and $\gamma = (\gamma_J)_{J\in\mathcal{J}}$, define

$$\lambda \cdot \gamma \equiv \sum_{J\in\mathcal{J}} \lambda_J \gamma_J, \quad (7)$$

$$\lambda \cdot K(r, \mu) = \sum_{J\in\mathcal{J}} \lambda_J K_J(r, \mu), \quad (8)$$

and

$$\lambda \cdot K\Psi \equiv (\lambda \cdot K)\Psi \equiv \lambda \cdot (K\Psi) \equiv \sum_{J\in\mathcal{J}} \lambda_J K_J \Psi. \quad (9)$$

Define the two-norm in the usual way, $\|\lambda\| = (\lambda \cdot \lambda)^{1/2}$, and (for a positive-definite N by N matrix Σ) the weighted two-norm

$$\|\lambda\|_\Sigma \equiv \sqrt{\lambda \cdot \Sigma \cdot \lambda}. \quad (10)$$

From this point forward, except in the Appendix, L denotes a linear functional (such as a spatially weighted average) defined for rotation models. Suppose we wish to estimate the linear functional $L\Omega$ from data δ. A fundamental result of linear inverse theory is that this is impossible without additional information unless one can write L as a linear combination of the data functionals $\{K_J\}$; i.e., one can estimate $L\Omega$ with finite uncertainty only if $L\Psi = \lambda \cdot K\Psi$ for some λ and all Ψ (Backus & Gilbert 1970). The corresponding linear combination of the data, $\lambda \cdot \delta$, is an unbiased estimate of $L\Omega$. Much of Backus-Gilbert theory consists of methods for selecting λ so that the functional L is a localized average of Ω and can be estimated with reasonable uncertainty. Pijpers & Thompson (1992, 1994) discuss an alternative, computationally efficient strategy for choosing λ, called SOLA.

This paper concerns estimating M such linear functionals $\{L_k\}_{k=1}^M$ with simultaneous confidence. Let $\{\Lambda_k\}_{k=1}^M$ be the collection of N-vectors such that $L_k = \Lambda_k \cdot K$; since $\{K_J\}$ is a linearly independent set, only one Λ_k can solve this equation. We assume for convenience that $\{\Lambda_k\}$ are linearly independent, but their choice is otherwise arbitary, so Backus-Gilbert and SOLA estimates are included as special cases, as are the tailored differencing kernels we derive below. The assumption of linear independence is for convenience of notation, not necessity. A consequence of the linear independence is that $M \le N$; i.e., we restrict ourselves to estimating no more linear functions than we have data.

Each estimate $\Lambda_k \cdot \delta$ is an unbiased estimate of the corresponding functional $L_k\Omega$; under our assumption of Gaussian errors, the estimates are jointly Gaussian since they are linear combinations of Gaussian random variables. Let $L\Omega$ be the M-vector $(L_k\Omega)_{k=1}^M$, and let Λ (with no subscript) be the M by N matrix whose $(k, J)^{\text{th}}$ entry is Λ_{kJ}. It is well known that $\Lambda \cdot \delta$ has the M-dimensional Gaussian distribution with mean vector $L\Omega$ and covariance matrix

$$\Xi = \Lambda \cdot \Sigma \cdot \Lambda'. \quad (11)$$

The variance of the estimate $\Lambda_k \cdot \delta$ is

$$\tau_k^2 = \Xi_{kk}. \quad (12)$$

It will be useful as well to have the correlation matrix Γ of the estimates, a normalized version of the covariance

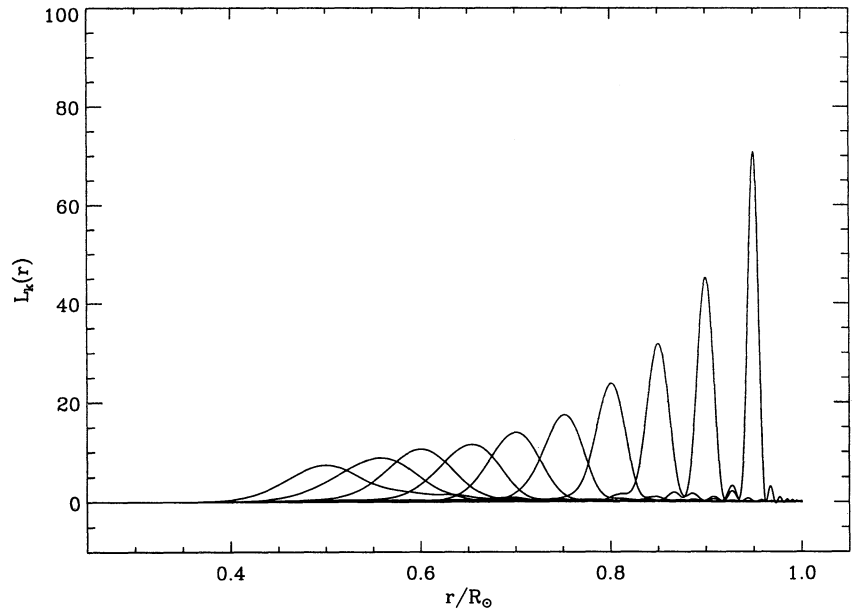

FIG. 1.—The 10 averaging kernels for Ω_1 as a function of radius. The kernels were derived using the Backus-Gilbert method.

matrix:

$$\Gamma = \tilde{\Lambda} \cdot \Sigma \cdot \tilde{\Lambda}' , \tag{13}$$

where

$$\tilde{\Lambda}_{kJ} = \frac{\Lambda_{kJ}}{\tau_k} . \tag{14}$$

From BBSO splitting data (Woodard & Libbrecht 1993), we shall try to estimate averages of the quantity

$$\Omega_1(r) = \frac{3}{4} \int_{-1}^{1} (1 - \mu^2)\Omega(r, \mu) d\mu , \tag{15}$$

which is related to the angular momentum

$$dh = \Omega_1(r)\rho(r)r^4 \, dr . \tag{16}$$

This quantity is interesting because the existence of a region with rapid variation would be evidence for a torque, if stress were proportional to shear. Since Ω_1 involves the angular projection of $\Omega(r, \mu)$ onto the function $1 - \mu^2$, asymptotically as $l \to \infty$, it affects only the first a-coefficients $\{a_{1nl}\}$. Conversely, asymptotically $\{a_{1nl}\}$ contain information about only the Ω_1 component of Ω. The data kernel for a_{1nl} is $K_{nl}(r)$, which is defined in the Appendix. Of the a-coefficients, $\{a_{1nl}\}$ are the most accurately measured. For these reasons, we used only the first a-coefficients in our study of Ω_1.

We averaged the reported BBSO $\{a_{1nl}\}$ splitting coefficients using equal weights for the 4 yr 1986, 1988–1990, for 1336 multiplets in the range $20 \leq l \leq 140$. We assumed that the standard deviations of the errors in the coefficients are as reported, unless they were reported to be less than 0.1 nHz, in which case we set them to 0.1 nHz. We assumed that the errors are independent. We used the Backus-Gilbert method to construct ten averaging kernels for Ω_1 centered at different depths. Figure 1 plots the 10 kernels; Table 1 lists the centers of the averaging kernels, their widths as defined by Backus & Gilbert (1968), the corresponding linear combination of the data ($\Lambda_k \cdot \delta$), and the

TABLE 1

SUMMARY OF THE AVERAGING KERNELS FOR THE ESTIMATED RADIAL AVERAGE OF Ω_1

Kernel	Center/$R_\odot$	Width/$R_\odot$	Estimate/nHz	Formal Error/nHz
1.....	0.510	0.111	429.140	2.873
2.....	0.555	0.072	432.812	1.343
3.....	0.599	0.057	434.561	1.002
4.....	0.651	0.052	436.369	0.950
5.....	0.699	0.043	437.981	0.767
6.....	0.750	0.034	444.395	0.804
7.....	0.800	0.024	441.784	0.768
8.....	0.850	0.018	438.367	0.850
9.....	0.900	0.013	445.104	0.680
10.....	0.950	0.009	450.712	0.908

NOTES.—Col. (2) radii of the "centers" of the averaging kernels ($\int rL_k^2 / \int L_k^2$), in solar radii. Col. (3): widths of the kernels using the definition of Backus & Gilbert 1968. Col. (4): estimated averages of $\Omega_1/2\pi$. Col. (5): standard deviations of the estimates, computed using eq. (12).

TABLE 2

CORRELATION MATRIX Γ CORRESPONDING TO THE 10 AVERAGING KERNELS

1.0000	0.3061	0.1282	0.0901	0.0463	0.0516	0.1325	−0.0751	−0.0517	0.1208
0.3061	1.0000	0.1230	−0.0724	−0.0290	−0.0616	−0.0092	−0.0570	0.0092	0.0934
0.1282	0.1230	1.0000	−0.0660	0.0794	0.0556	0.0800	−0.0525	0.0175	0.1166
0.0901	−0.0724	−0.0660	1.0000	−0.4593	0.1386	0.1234	0.0272	0.1193	0.0680
0.0463	−0.0290	0.0794	−0.4593	1.0000	−0.3951	−0.1508	0.0732	0.0131	0.2108
0.0516	−0.0616	0.0556	0.1386	−0.3951	1.0000	0.0366	−0.0608	0.1048	0.1650
0.1325	−0.0092	0.0800	0.1234	−0.1508	0.0366	1.0000	0.2446	0.2791	0.1886
−0.0751	−0.0570	−0.0525	0.0272	0.0732	−0.0608	0.2446	1.0000	0.3904	0.1910
−0.0517	0.0092	0.0175	0.1193	0.0131	0.1048	0.2791	0.3904	1.0000	0.3419
0.1208	0.0934	0.1166	0.0680	0.2108	0.1650	0.1886	0.1910	0.3419	1.0000

standard deviations of the estimates. Table 2 contains the correlation matrix Γ of the 10 estimates.

3. SIMULTANEOUS CONFIDENCE INTERVALS FOR LINEAR FUNCTIONALS

3.1. *Simultaneous Coverage Probability*

Consider tossing a (possibly loaded) coin independently 100 times to find a confidence interval for the probability p with which the coin lands heads. The distribution of the number of heads in 100 tosses is, under these assumptions, binomial with 100 trials and probability p of success in each trial. Provided p is not too close to zero or one, this distribution is approximated well by a Gaussian distribution with mean $100p$ and standard deviation $10[p(1-p)]^{1/2}$. With high probablity, $10[\hat{p}(1-\hat{p})]^{1/2}$ will be close to the standard deviation, where $\hat{p}$ = (number of heads observed)/100 is the sample proportion. Provided p is not too close to zero or one, the distribution of the sample proportion is approximately Gaussian with mean p and standard deviation $\hat{\sigma} = [\hat{p}(1-\hat{p})]^{1/2}/10$. Since the chance that a standard Gaussian random variable is between -1.96 and 1.96 is 0.95, $\Pr\{|p-\hat{p}| \le 1.96\hat{\sigma}\} \approx 0.95$. Consequently, an approximate 95% confidence interval for p is given by

$$I = [\hat{p} - 1.96\hat{\sigma}, \hat{p} + 1.96\hat{\sigma}] . \quad (17)$$

The coverage probability of the confidence interval I is the chance that I contains p. Note that this probability makes sense only before the data are observed: the coverage probability refers to properties of the procedure applied to random data, not to the ultimate numerical values one computes after observing the data. After observing the data and calculating a confidence interval, the interval either does or does not contain p, and there is no more randomness in the problem. At that point, the "confidence level," which is equal to the coverage probability of the random interval before observing the data, quantifies our uncertainty about whether in fact the confidence interval contains p.

Now suppose that we wish to know whether there is a difference between the probablities p_1 and p_2 of heads of two possibly biased coins. One way to do this is to compare $\hat{p}_1$ and $\hat{p}_2$, the sample proportions of heads in 100 independent tosses of each coin separately. Intuition and common practice suggest that we can use $\hat{p}_1$ and $\hat{p}_2$ to construct 95% confidence intervals for p_1 and p_2, then conclude that $p_1 \neq p_2$ if the confidence intervals do not overlap. In fact, testing the hypothesis that $p_1 = p_2$ in this way has a higher significance level (lower confidence level, loosely speaking) than is commonly recognized: if $\hat{p}_1$ and $\hat{p}_2$ are independent, then the chance that both $\hat{p}_1 - 1.96\hat{\sigma}_1 \le p_1 \le \hat{p}_1 + 1.96\hat{\sigma}_1$ and $\hat{p}_2 - 1.96\hat{\sigma}_2 \le p_2 \le \hat{p}_2 + 1.96\hat{\sigma}_2$ is only about $0.95 \times 0.95 = 0.9025$. Thus if the two 95% confidence intervals do not overlap, we have only about 90% confidence that $p_1 \neq p_2$. (More precisely, we could reject the hypothesis that $p_1 = p_2$ at significance level approximately 0.1, not 0.05.) The essential point is that even though each interval I_j, $j = 1, 2$, contains its corresponding p_j with probability $1-\alpha$, part of the time that I_1 contains p_1, I_2 will not contain p_2, and vice versa. As a result, the chance that both intervals contain their parameters, the simultaneous coverage probability, is less than $1-\alpha$. Figure 2 illustrates this point. If we try to compare the probabilities of heads for N coins in this way, we end up with a simultaneous confidence level of about $0.95^N \times 100\%$. Conversely, if we want to end up with simultaneous 95% confidence, we need to begin with $0.95^{1/N} \times 100\%$ confidence intervals for each p_j. The situation is more complicated when the sample proportions $\hat{p}_j$ are not independent, which is more directly analogous to comparing several linear estimates in an inverse problem, since estimates centered at different points typically involve correlated linear combinations of the same data, as in our helioseismological example of § 2.

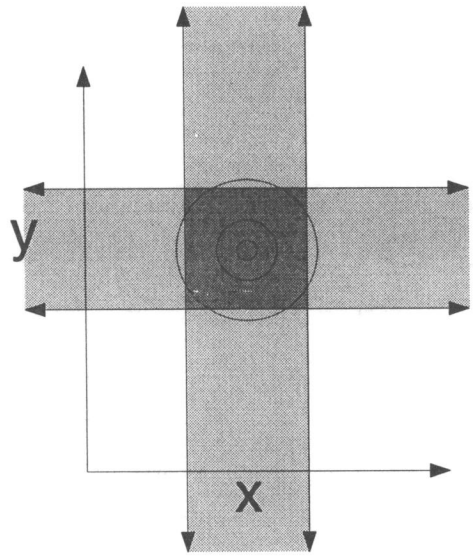

FIG. 2.—Illustration of the issue of simultaneity. The circles represent contours of the joint probability density function of the random variables X and Y. The chance that X is in the (infinite) shaded vertical slice is $1-\alpha$, as is the chance that Y is in the horizontal slice. However, the chance that X is in the shaded region and Y is in the shaded region is the probability that (X, Y) is in the small square, which is less than $1-\alpha$. In this sketch, X and Y are independent, so the probability that (X, Y) is in the square is $(1-\alpha)^2$. In general, the probability that both X and Y are in a pair of ranges depends on their joint distribution. The text gives four ways to adjust the lengths of a set of M confidence intervals so that the chance the intervals simultaneously contain their M associated parameters is at least $1-\alpha$.

3.2. *Bonferroni Simultaneous Confidence Intervals*

Bonferroni's inequality (e.g., Bickel & Doksum 1977) says that the probability that at least one of the events $\{A_k\}_{k=1}^M$ occurs is no larger than the sum of the chances that each occurs; i.e., Pr $\{A_1$ occurs or A_2 occurs or ... or A_M occurs$\} \leq \sum_{k=1}^M$ Pr $\{A_k$ occurs$\}$.

Suppose we have a procedure for producing a set of (not necessarily simultaneous) $1 - \alpha'$ confidence intervals for the parameters $L_k\Omega$, $k = 1, \ldots, M$. The chance that at least one of these intervals fails to contain its corresponding parameter is the chance that at least one of the events $\{A_k\}_{k=1}^M$ occurs, where A_k is the event that the kth confidence interval fails to contain its parameter. Each of these events has (by assumption) probability α'. Bonferroni's inequality implies that the chance that one or more of the M intervals fails to contain its corresponding parameter is no larger than $M\alpha'$, regardless of the correlation among the estimates. Consequently, the chance that all the confidence intervals simultaneously include their corresponding parameters is at least $1 - M\alpha'$. It follows that if $\alpha' = \alpha/M$, so that the original intervals have coverage probability $1 - \alpha/M$, the set of M confidence intervals has simultaneous coverage probablity at least $1 - \alpha$, as desired. Note that Bonferroni's inequality does not constrain us to use the same confidence level for each interval; we need only insist that the sum of the noncoverage probabilities of each of the intervals equal at most α. This gives one the freedom to construct shorter confidence intervals for some of the functionals of interest, to minimize the maximum length of the intervals, or to minimize some weighted sum of the lengths of the intervals, however the practitioner may choose.

The Bonferroni approach to simultaneity can be quite useful, especially when the number of simultaneous confidence intervals one desires is not too large. It is particularly easy to use, since we need no assumption about the dependence of the estimates. In our helioseismological example, the covariance among the estimates is such that Bonferroni's inequality is nearly sharp, and little is gained by more complicated analyses.

Suppose we desire to have a set of simultaneous 95% confidence intervals for the 10 averages of angular velocity described above. Bonferroni's inequality implies that if we start with confidence intervals with individual coverage probabilities $1 - 0.05/10 = 0.995$, the intervals will have simultaneous coverage probability at least 95%. Since the chance that a standard Gaussian random variable is between -2.81 and 2.81 is 0.995, the intervals

$$\{[\Lambda_k \cdot \delta - 2.81\tau_k, \Lambda_k \cdot \delta + 2.81\tau_k]\}_{k=1}^{10} \quad (18)$$

have simultaneous 95% coverage probablity. These intervals are plotted in Figure 3. A 95% confidence interval for a single estimate would have length $2 \times 1.96\tau_k$, so the increase in length required for simultaneity is $(2 \times 2.81\tau_k)/(2 \times 1.96\tau_k) \approx 1.43$. This ratio increases with the number of intervals, and for Gaussian errors, decreases with increasing confidence level: for example, for 99% simultaneous confidence in this same problem, the ratio would be about $3.27/2.56 \approx 1.28$.

3.3. *Simultaneous Confidence Intervals Based on χ^2 Misfit to the Data*

Since the errors $\{\epsilon_J\}$ are zero-mean Gaussian random variables with covariance matrix Σ, the distribution of $\|\epsilon\|_{\Sigma^{-1}}^2$ is χ^2 with N degrees of freedom (recall that N is the cardinality of J). If $\chi = (\chi_{N,\alpha}^2)^{1/2}$ is the square root of the $(1 - \alpha) \times 100$ percentage point of the χ^2 distribution with N degrees of freedom, by definition

$$\Pr\{\|\epsilon\|_{\Sigma^{-1}} \leq \chi\} = 1 - \alpha\,. \quad (19)$$

Thus

$$\Pr\{\|\delta - K\Omega\|_{\Sigma^{-1}} \leq \chi\} = 1 - \alpha\,. \quad (20)$$

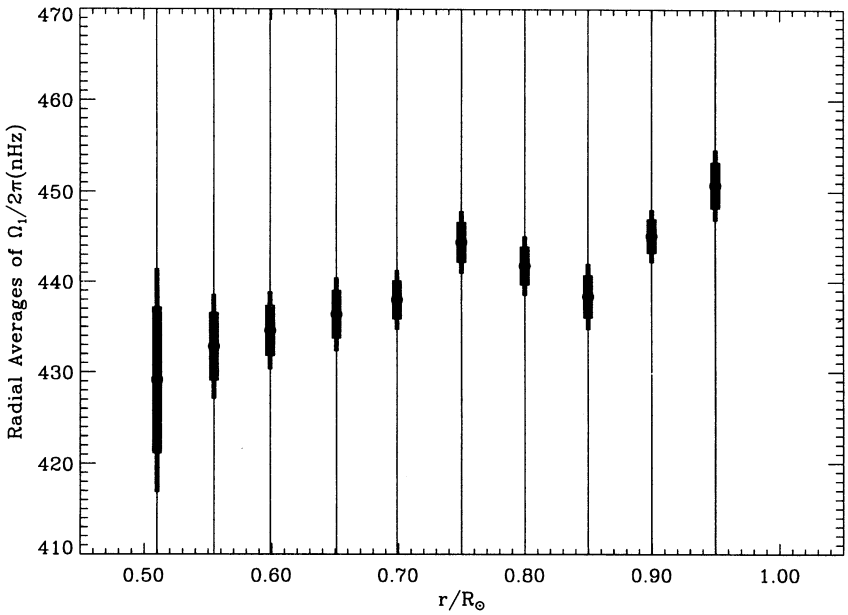

FIG. 3.—Four sets of simultaneous confidence intervals for the 10 averages of Ω_1. The lengths of the intervals are proportional to the standard deviations of the estimates; other choices are possible. The shortest intervals (*heaviest lines*) are the maximum-modulus intervals; on this scale, they are indistinguishable from the Bonferroni intervals, which are the next shortest. The third-shortest intervals (medium-weight lines) are the Scheffé intervals, and the widest (thin lines extending beyond the boundary of the plot) are those derived using the χ^2 distribution in the 1336-dimensional data space. One cannot pass a straight line through any of the confidence intervals except the widest, so we may reject the hypothesis that Ω_1 is constant at significance level 0.05.

For any $L_k = \Lambda_k \cdot \boldsymbol{K}$, define

$$l_k = \inf \{L_k \Psi : \|\delta - \boldsymbol{K}\Psi\|_{\Sigma^{-1}} \leq \psi\} \tag{21}$$

and

$$u_k = \sup \{L_k \Psi : \|\delta - \boldsymbol{K}\Psi\|_{\Sigma^{-1}} \leq \chi\} . \tag{22}$$

Then

$$\Pr \{l_k \leq L_k \Omega \leq u_k\} \geq 1 - \alpha ; \tag{23}$$

i.e., $[l_k, u_k]$ is a $1 - \alpha$ confidence interval for $L_k\Omega$. Note that the constraint set for this optimization problem, $\{\Psi : \|\delta - \boldsymbol{K}\Psi\|_{\Sigma^{-1}} \leq \chi\}$, does not depend on L_k. The optimization problems measure different properties of a single set that has probablity $1 - \alpha$ of containing the noise-free observations. As a result, the simultaneous coverage probability of the M intervals is $1 - \alpha$:

$$\Pr \{l_1 \leq L_1\Omega \leq u_1 \text{ and } l_2 \leq L_2\Omega \leq u_2 \text{ and } \ldots \text{ and } l_M \leq L_M\Omega \leq u_M\} \geq 1 - \alpha . \tag{24}$$

It is possible to find l_k and u_k explicitly: if $\|\boldsymbol{K}\Omega - \delta\|_{\Sigma^{-1}} \leq \chi$ (which occurs with probability $1 - \alpha$),

$$\begin{aligned} |\Lambda_k \cdot \delta - \Lambda_k \cdot \boldsymbol{K}\Omega| &= |\Lambda_k \cdot (\boldsymbol{K}\Omega + \epsilon) - \Lambda_k \cdot \boldsymbol{K}\Omega| \\ &= |\Lambda_k \cdot \epsilon| \\ &\leq \|\Lambda_k\|_\Sigma \|\epsilon\|_{\Sigma^{-1}} \\ &\leq \chi \|\Lambda_k\|_\Sigma , \qquad (25) \\ &= \chi\tau_k . \qquad (26) \end{aligned}$$

This bound is attained provided $\{K_J\}$ are linearly independent (which we have assumed), so $l_k = \Lambda \cdot \delta - \chi\|\Lambda_k\|_\Sigma$ and $u_k = \Lambda \cdot \delta + \chi\|\Lambda_k\|_\Sigma$. See also Stark (1992). This gives an explicit solution to the problem of constructing simultaneous confidence intervals for any number of linear functionals of Ω. The lengths of the intervals are proportional to the standard deviations of the corresponding estimates, with constant of proportionality $(\chi^2_{n,\alpha})^{1/2}$. The intervals thus have the same relative lengths as we found using Bonferroni's inequality (for equal coverage probability $1 - \alpha'$ for each interval).

For $N = 1336$ data, the value of χ needed to get 95% simultaneous coverage probability is about 37.71. The resulting confidence intervals are plotted in Figure 3. The simultaneous confidence intervals for the 10 averages of solar angular velocity derived using the χ^2 distribution with N degrees of freedom are longer than the Bonferroni intervals by a factor of $37.71/2.81 \approx 13.42$. The confidence intervals obtained this way are unnecessarily wide, especially when one is interested in few linear combinations compared with the number of data.

3.4. *Confidence Intervals Based on the M-Dimensional Gaussian Distribution*

There are innumerable choices of simultaneous confidence intervals for the collection of functionals $\{L_k\Omega\}$, and one might consider using a procedure that gives intervals that are in some sense optimal for the problem at hand. For example, one might seek to minimize the length of the longest of the joint confidence intervals, to minimize a weighted sum of the lengths of the intervals, or to construct intervals that support certain inferences by being likely to exclude certain values (Banjamini & Stark 1996 construct simultaneous confidence intervals that are less likely to contain zero). The two procedures presented above have the property that the lengths of the intervals are proportional to the standard deviations of the estimates, which is a reasonable way to put the estimates on an equal footing. In other words, both give values of a constant c so that the set of intervals $\{[\Lambda_k \cdot \delta - c\tau_k, \Lambda_k \cdot \delta + c\tau_k]\}_{k=1}^m$ are simultaneous $1 - \alpha$ confidence intervals for $\{L_k\Omega\}$.

This section presents two standard choices for joint confidence intervals based on the joint distribution of the estimates: "Scheffé" intervals and "maximum-modulus" intervals (e.g., Lehmann 1986). Scheffé intervals derive from the event that the vector of estimates lies in an ellipsoid centered at the vector $\boldsymbol{L}\Omega$ of true values of the functionals, and their lengths are related to the χ^2 distribution with M degrees of freedom. Maximum-modulus intervals derive from the event that the vector of estimates lies in a hyperrectangle centered at $\boldsymbol{L}\Omega$, and require integrating a correlated multidimensional Gaussian density. They give the smallest constant c described above.

3.4.1. *Scheffé Intervals*

Since $\Lambda \cdot \delta$ has a Gaussian distribution with mean $\boldsymbol{L}\Omega$ and covariance matrix Ξ, $\|\Lambda \cdot \delta - \boldsymbol{L}\Omega\|^2_{\Xi^{-1}}$ has the χ^2 distribution with M degrees of freedom. Thus

$$\Pr \{\|\Lambda \cdot \delta - \boldsymbol{L}\Omega\|^2_{\Xi^{-1}} \leq \chi^2_{M,\alpha}\} = 1 - \alpha . \tag{27}$$

As in § 3.3, we can find a $1 - \alpha$ confidence interval for the linear functional $\gamma \cdot \boldsymbol{L}\Omega$ by finding the smallest and largest values $\gamma \cdot \beta$ can take for β satisfying $\|\Lambda \cdot \delta - \beta\|^2_{\Xi^{-1}} \leq \chi^2_{M,\alpha}$. We are particularly interested in the components of $\boldsymbol{L}_\Omega$, which corresponds to taking $\gamma = e_k$, $k = 1, \ldots, M$, where e_k is the M-vector all of whose components are zero except the kth, which is unity.

Essentially the same derivation as in § 3.3 shows that

$$\{[\Lambda \cdot \delta - \chi\|e_k\|_\Xi, \Lambda \cdot \delta + \chi\|e_k\|_\Xi]\}_{k=1}^M , \tag{28}$$

where $\chi = (\chi^2_{M,\alpha})^{1/2}$ are simultaneous $1 - \alpha$ confidence intervals for $\{L_k\Omega\}$. Since $\|e_k\|_\Xi = \tau_k$, we again end up with confidence intervals whose lengths are proportional to the standard deviations of the estimates, but this time the constant of proportionality is $(\chi^2_{M,\alpha})^{1/2}$, which is smaller than $(\chi^2_{N,\alpha})^{1/2}$ when the number of data N exceeds the number of estimated functionals M, approximately by the ratio $(M/N)^{1/2}$. For 95% confidence intervals for the 10 functionals in our helioseismological example, we compute $c = (\chi^2_{0.05,10})^{1/2} = 4.279$, so the resulting intervals are long than the Bonferroni intervals by a factor of 1.52, but are 8.81 times shorter than those based on the χ^2 distribution with N degrees of freedom. These intervals are plotted in Figure 3.

3.4.2. *Maximum-Modulus Intervals*

We have now seen a variety of methods that give simultaneous confidence intervals whose half-lengths are a common multiple c of the standard deviations $\{\tau_k\}$ of the estimates. Maximum-modulus intervals are of this form, but have the smallest value of c that gives the correct simultaneous coverage probability:

$$\Pr \left\{ \max_{k=1}^m \left| \frac{\Lambda_k \cdot \delta - L_k\Omega}{\tau_k} \right| \leq c \right\} = 1 - \alpha . \tag{29}$$

If $\alpha = 0.05$, for example, we know that for $M = 1$, $c \approx 1.96$ suffices. The first section of this paper shows that if $M = 2$ and the estimates $\Lambda_1 \cdot \delta$ and $\Lambda_2 \cdot \delta$ are statistically

independent, it suffices to take c to be the $100(1-\alpha/2)^{1/2} = 100(0.975)^{1/2} \approx 98.7$ percentage point of the Gaussian distribution, which is about 2.23. For Gaussian errors, the estimates are independent when the coefficients in the linear combinations are orthogonal vectors, for example, when they depend on mutually exclusive subsets of the data. Typically, the estimates are dependent. Note that if we wish to compare only two linear estimates, we get more precise inferences by finding a single confidence interval for the difference functional $(\mathbf{\Lambda}_1 - \mathbf{\Lambda}_2) \cdot \boldsymbol{K}\Omega$; we take this approach in § 4 below. Only when we want to compare more than two averages does it make sense to use simultaneous confidence intervals, and even then, if our only goal is to make comparisons, better procedures exist than those based on confidence intervals.

We can find c by integrating the joint density of $v_k = (\mathbf{\Lambda}_k \cdot \delta - L_k\Omega)/\tau_k$, $k = 1, \ldots, M$, over hypercubes with various side lengths, to find the side half-length that gives probability $1 - \alpha$. By construction, $\{v_k\}$ are jointly normal with zero mean and unit variance; their covariance matrix is the correlation matrix $\mathbf{\Gamma}$ defined in equation (13). The constant c thus solves

$$(2\pi)^{-M/2} |\mathbf{\Gamma}|^{-1/2} \int_{-c}^{c} dv_1 \int_{-c}^{c} dv_2 \cdots \int_{-c}^{c} dv_M \times \exp\left(-\frac{1}{2|\mathbf{\Gamma}|} v^T \cdot \mathbf{\Gamma} \cdot v\right) = 1 - \alpha\,, \quad (30)$$

where $|\mathbf{\Gamma}|$ is the determinant of $\mathbf{\Gamma}$. For a general covariance matrix $\mathbf{\Gamma}$, c cannot be found in closed form. However, it can be found iteratively by performing the integral (30) numerically for trial values of c. Since the integral is monotone in c, a bisection or other search can find c to any desired precision.

Alternatively (and much less expensively), one may approximate the probability by counting the fraction of the time a pseudorandom multivariate Gaussian variable lands in a hypercube with side length $2c$. This is easily accomplished in the computer data analysis language MATLAB (The Math Works 1985–1990). Let `g` = $\mathbf{\Gamma}$, let `nsim` be the number of pseudorandom variates to compute, and let `M` be the number of averages (so that `g` is an `M` by `M` matrix). Then the command

```
p=sum(all(abs(chol(g)'*randn(M,nsim))
<=c))/nsim
```

estimates the coverage probability of a hypercube with side $2c$.

We performed a sequence of such simulations using the matrix $\mathbf{\Gamma}$ given in Table 2 to estimate the value of c corresponding to 95% simultaneous coverage probability for the 10 averages of angular velocity. We found that $c = 2.789$ gave an estimated coverage probability of 0.9502 in 10^6 trials. If one believes that pseudorandom numbers behave like random numbers in this problem, the results of § 3.1 yield a confidence interval for the actual coverage probability corresponding to c. The estimated standard deviation of the estimated probability is $[10^{-6}0.9502(1-0.9502)]^{1/2} \approx 2.17 \times 10^{-4}$, so a 95% confidence interval for the coverage probability would be [0.9498, 0.9506].

This choice of c gives confidence intervals that, in this application, are shorter than the Bonferroni intervals by the factor $2.79/2.81 \approx 0.99$, which is not much savings for the extra effort. For other covariance matrices, the savings can be greater, and the savings typically increase as the number of estimates M grows. Table 3 displays the normalization constants c for the four methods discussed, for the helioseismological example; Figure 3 plots the four sets of confidence intervals.

TABLE 3

Method	c
Maximum modulus.....	2.789
Bonferroni	2.807
Scheffé	4.279
Data χ^2	37.711

NOTES.—Constants c derived by various means described in the text, such that the intervals $\{[\mathbf{\Lambda} \cdot \delta - c\tau_k, \mathbf{\Lambda} \cdot \delta + c\tau_k]\}$ are simultaneous 95% confidence intervals for the 10 radial averages of Ω_1 corresponding to the averaging kernels in Fig. 1.

4. AVERAGING KERNELS FOR SPATIAL CHANGES

In this section, we construct averaging kernels that are sensitive to the difference between averages of the angular velocity, rather than localized averaging kernels.

4.1. *Differences in Radial Averages of* Ω_1

Consider testing the hypothesis that the average of Ω_1 near the radius $0.55R$ ($L_2\Omega$) differs from that near $0.70R$ ($L_5\Omega$). If we base the inference on the shortest simultaneous confidence intervals we have constructed for the 10 averages (with $c = 2.789$), we would not reject the hypothesis that the averages of Ω_1 are equal at significance level 0.05, because the simultaneous confidence intervals, [429.07 nHz, 436.56 nHz] and [435.84 nHz, 440.12 nHz], overlap. However, we would expect to have done better had we considered only the two averages. Had we constructed simultaneous 95% confidence intervals for just $L_2\Omega$ and $L_5\Omega$, c would be smaller, but it could not be less than 1.96, the value for a single estimate and for any number of perfectly correlated estimates.

Suppose that we had begun this study with the hypothesis that these two averages of Ω_1 were equal, and that we wanted to concentrate all our resources to determine whether there was in fact a difference between them. We could then have proceeded by estimating the difference $(L_2 - L_5)\Omega$ directly. The estimate $(\mathbf{\Lambda}_2 - \mathbf{\Lambda}_5) \cdot \delta = -5.169$ nHz has standard deviation 1.566 nHz, so

$$[-5.169 - 1.96 \times 1.566 \text{ nHz}, -5.169 + 1.96 \times 1.566 \text{ nHz}] = [-8.238 \text{ nHz}, -2.099 \text{ nHz}] \quad (31)$$

is a 95% confidence interval for the difference of the averages. This interval does not contain zero, so we would have rejected the hypothesis that the averages of Ω_1 near the two depths are the same at significance level 0.05.

We caution the reader that it would not be statistically sound to reject the hypothesis that $L_2\Omega = L_5\Omega$ based on our analysis here: the quoted 0.05 significance level is incorrect, since we first looked at the estimates before deciding to test that hypothesis. In fact, we chose to compare $L_2\Omega$ and $L_5\Omega$ precisely because the test based on simultaneous confidence intervals would fail to reject, while the test based directly on the single estimate $(\mathbf{\Lambda}_2 - \mathbf{\Lambda}_5) \cdot \delta$ would reject.

As intimated above, simultaneous 95% confidence intervals for $L_2\Omega$ and $L_5\Omega$ with lengths proportional to the standard deviations of the estimates can be no shorter than the intervals [430.180 nHz, 435.444 nHz] and [436.478 nHz, 439.484 nHz], which use $c = 1.96$. These intervals fail to intersect by 1.034 nHz, which is an upper bound on what we could hope to achieve by comparing confidence intervals; in comparison, the confidence interval for the difference does better: it fails to include zero by 2.099 nHz. Thus a confidence interval derived directly for the difference is more able to detect a difference than a procedure that compares simultaneous confidence intervals for the separate averages. This is true quite generally, and results from the facts that the standard deviation of the estimator of the difference is no larger than the sum of standard deviations of the two estimates (which follows from the triangle inequality), and that c is a nondecreasing function of the number of functionals one estimates (in the class of simultaneous confidence intervals whose lengths are proportional to the standard deviations of the estimates, simultaneous intervals must be at last as long as nonsimultaneous intervals).

4.2. *Differences in Angular Averages of* Ω

Figure 4 shows the angular dependence of three localized averaging kernels for angular averages of angular velocity. The kernels are not very localized, but trying to construct kernels more localized than these results in even larger negative sidelobes. Among the advantages of estimating differences are (i) it may avoid the question of simultaneity, since the value of one functional is enough to tell whether rotation varies with latitude, and (ii) we can construct a satisfactory "differencing kernel" in some cases (as here) where it is impossible to find a localized linear combination of the data kernels.

To estimate an average difference of angular velocity at low and high latitudes, we need to find a linear combination of the data kernels that is nonnegative over some range of latitudes, nonpositive on another set of latitudes, zero elsewhere, and such that its integral is zero. The approximation (4) shows that the radial part and angular part of the data kernel K_{nlm} can be separated approximately, and that the radial part does not depend on m. Since the a-coefficients $\{a_{jnl}\}_{j=1}^{j_{max}}$ are linear combinations of splittings for different values of m but a fixed (n, l) multiplet, the radial dependence of the data kernels for those a-coefficients is the same. Thus we obtain a simpler problem if we use only one (n, l) multiplet at a time; we need worry about only the angular dependence of the linear combination of data kernels, and the radial dependence will take care of itself, to a first approximation. We can later take linear combinations of the resulting averaging kernels for different (n, l) multiplets to bring more data to bear on the estimate.

We found a linear combination of the angularly dependent functions $W_{lm}(\mu)$ of equation (3) such that the combination is positive at low colatitudes and negative at high colatitudes, using the SOLA method described by Pijpers & Thompson (1992, 1994), with target function satisfying

$$T(\mu)\begin{cases} \le 0\,, & \mu \ge \mu_0, \\ \ge 0\,, & \mu \le \mu_0, \end{cases} \tag{32}$$

$$\int_0^{\mu_0} T(\mu)d\mu = -1\,, \tag{33}$$

and

$$\int_{\mu_0}^{1} T(\mu)d\mu = 1\,. \tag{34}$$

The constant μ_0 is the colatitude at which the target function changes sign; the precise form of T is given in the Appendix. The corresponding linear combination of the full data kernels, $\boldsymbol{\lambda} \cdot \boldsymbol{K}(r, \mu)$, changes sign at a radially dependent colatitude, $\mu_0(r)$, which is in fact quite close to μ_0. Once the radial variation of the complete data kernels is included, the integral of $\boldsymbol{\lambda} \cdot \boldsymbol{K}(r, \mu)$ on each side of $\mu_0(r)$ differs slightly from ± 1 by the multiplicative factor C_{nl}, which is derived in

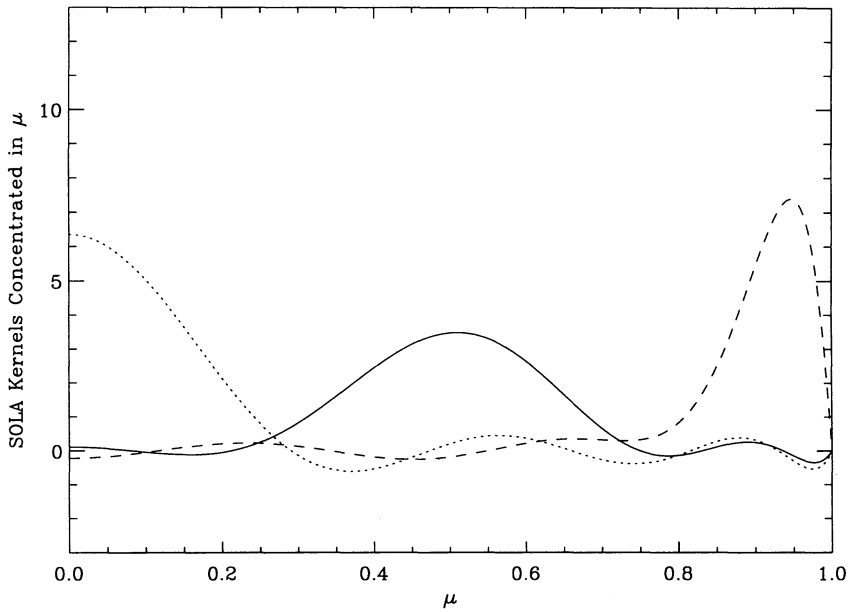

FIG. 4.—Plot of the angular part of three localized averaging kernels for an angular average of rotation. They are poorly localized, but kernels more localized than these have even larger negative sidelobes.

the Appendix. As a result, the linear functional

$$L = \frac{1}{2\pi C_{nl}} \boldsymbol{\lambda} \cdot \boldsymbol{K} \tag{35}$$

differences averages of the rotation rate (in cyclic frequency) at low and high latitudes, and an unbiased estimate of $L\Omega$ is given by

$$\frac{1}{2\pi C_{nl}} \boldsymbol{\lambda} \cdot \boldsymbol{\delta} \, . \tag{36}$$

The BBSO splitting data for the years 1986, 1988–1990 have 1336 multiplets in common. We averaged corresponding a-coefficients for those multiplets for the 4 yr, with equal weights. The variation of the latitude of the sign change is small for all the kernels we constructed for individual multiplets: $|\mu_0(r) - \mu_0|$ never exceeded 0.004. There is no other sign change in the averaging kernels, as is desired.

We averaged groups of 100 differencing kernels for individual multiplets with similar turning depths to obtain 14 kernels for differences of angular averages of Ω sensitive to different ranges of radius (the most deeply penetrating group had only 36 multiplets). Figure 5 shows an example one of these average kernels. Since all the averaging kernels change sign extremely close to the cone $\mu = \mu_0$, so do averages of them. Since the averaging kernels all integrate to zero, so do averages of them. Since, by assumption, measurement errors for different a-coefficients are independent, the estimated averages are independent as well. As a result, it is simple to compute maximum-modulus confidence intervals for the averages; as in § 3.1, the individual intervals should have coverage probability $0.95^{1/14} \approx 0.9963$, which for Gaussian random variables gives $c = 2.678$. Figure 6 plots maximum-modulus simultaneous 95% confidence intervals for the difference between angular averages of Ω against an average of the turning depths of the multiplets that contribute to the 14 average kernels. The estimated difference is roughly constant among all but the deepest of the averages, and none of the confidence intervals but that for the deepest average includes zero. This indicates that over a range of depths, the low-latitude region rotates faster on average than the high-latitude region, as is observed at the surface of the Sun.

5. CONCLUSIONS

The lengths of confidence intervals for estimated solar properties must be adjusted if one wishes to compare estimates, for example, to infer that those properties vary spatially. The adjustment is straightforward and can be accomplished in a number of ways. We examined four ways to construct simultaneous confidence intervals, within the class of procedures that give intervals whose lengths are proportional to the standard deviations of the estimates. The method that yields the shortest intervals is based on finding a hypercube that has probability $1 - \alpha$ of containing the correlated errors in the estimates, which yields "maximum-modulus" confidence intervals. The next shortest intervals in our helioseismological example derive from the Bonferroni inequality, followed by "Scheffé" confidence intervals, which derive from an ellipsoid with probability $1 - \alpha$ of containing the correlated errors in the estimates, and finally by confidence intervals that derive from an ellipsoid with probability $1 - \alpha$ of containing the original, possibly correlated, data errors. The ordering of the lengths of the latter three sets of intervals depends on the correlation of the data errors, the correlation of the estimates, the number estimated functionals, and the number of data. It is always true that the maximum-modulus intervals will be shortest in the class of procedures that yield interval lengths proportional to the standard deviations of the estimates, and that whenever the number of estimates is smaller than the number of data, the Scheffé intervals will be shorter than those based on the ellipsoid in data space. It is difficult to say a priori where the lengths of the Bonferroni-derived intervals will lie in the hierarchy.

We have assumed throughout this paper that the data errors are realizations of Gaussian random variables with

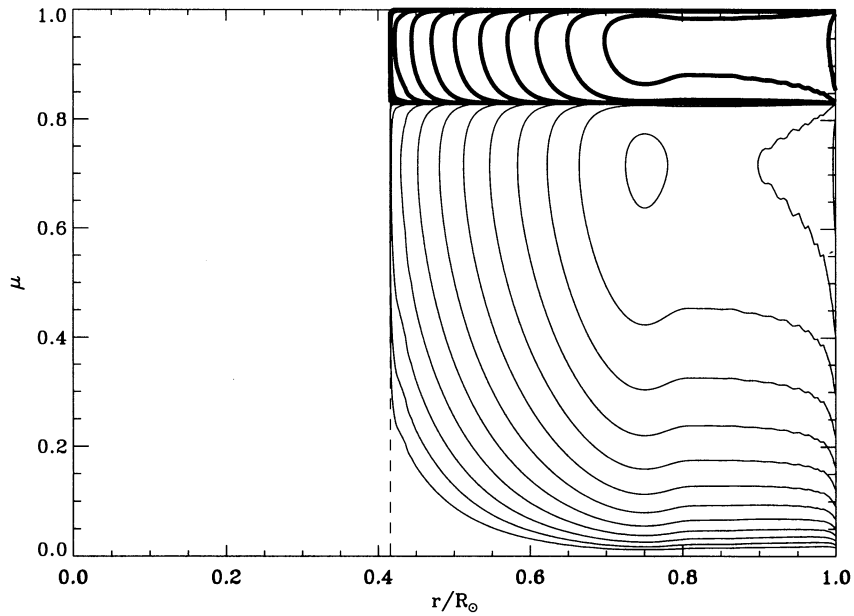

FIG. 5.—A kernel for the difference between averages of the angular velocity at low and high latitudes constructed by averaging 100 such kernels for individual multiplets with similar turning depths. The kernel changes sign on a single surface that lies almost exactly at latitude 56°.3, and its integral is zero.

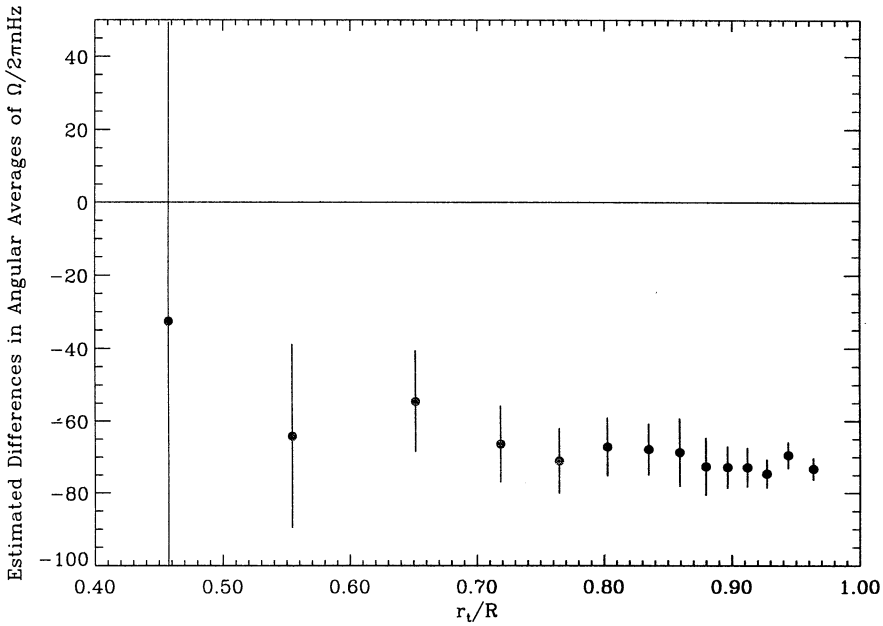

FIG. 6.—Maximum-modulus simultaneous 95% confidence intervals for differences in averages of angular velocity above and below colatitude $\mu_0 = 3 \times 13^{-1/2}$. The radial and angular weights in the averages differ, but they change sign almost exactly at μ_0, and integrate to ± 1 above and below μ_0. The abscissa of each interval is an average of the turning depths of the multiplets contributing to the estimate.

mean zero and known covariance matrix. This assumption might be unrealistic but is often the best one can adopt: the details of the data reduction, which would be necessary for characterizing the errors more accurately, are typically unavailable. The application of the methods we have presented for constructing simultaneous confidence intervals depends on the joint distribution of the data errors. In particular, the confidence interval lengths the Bonferroni method gives depend critically on the fact that the tails of a normal distribution are very thin. If the true error distribution has thicker tails (for example, because of outliers), even if the covariance matrix were known exactly, the Bonferroni bound could change substantially. The maximum-modulus intervals are the next most sensitive to the tail behavior of the error distribution, followed by the procedures derived from the χ^2 distribution.

We used all four methods to compute simultaneous confidence intervals for 10 radial averages of Ω_1, a quantity related to the angular-momentum density, from BBSO splitting data averaged for the 4 yr 1986, 1988–1990. All sets of intervals but that based on the χ^2 distribution in data space support the inference that Ω_1 varies with radius. (We can reject the hypothesis that Ω_1 is constant at significance level 0.05.) In this example, the maximum-modulus intervals are shortest, followed by Bonferroni's intervals, the Scheffé intervals, and the intervals based on the χ^2 distribution in data space.

Directly estimating the change in the property of interest generally leads to more precise inferences than does comparing simultaneous estimates. It is sometimes possible to estimate changes in circumstances where it is not possible to construct localized averages. We constructed linear combinations of a-coefficient kernels of individual (n, l) multiplets that change sign only once, near a prescribed boundary between a high- and a low-latitude region, by exploiting the degeneracy of the data kernel to decompose the problem approximately. The linear combinations have zero integrals, and therefore serve as kernels for the difference between averages of angular velocity in the two regions. By averaging these differencing functionals for modes with similar turning depths, we obtained functionals sensitive to the difference of the average angular velocity between low and high latitudes over different ranges of radius. We estimated these functionals of the Sun's angular velocity from BBSO splitting data and found strong evidence that over a range of depths, the low-latitude region rotates faster on average than the high-latitude region.

An obvious next step is to construct linear combinations of a-coefficient splitting kernels with isolated sign changes in both radius and colatitude, to estimate two-dimensional variations of solar angular velocity. Constraining the sign changes of the linear combination will require solving high-dimensional quadratic programs, which is far costlier than the method we employed here to reduce the dimension of the problem to unity.

We are grateful to K. G. Libbrecht and M. F. Woodard for making BBSO splitting data available to us. P. B. Stark thanks Andrew Gelman and Peter Wilson for helpful discussions. D. O. Gough and T. Sekü acknowledge support by PPARC (UK) and NASA grant NAS 5-30386. P. B. Stark was supported by NSF Presidential Young Investigator award DMS 89-57573 and grants DMS 94-04276 and AST 95-04410, and NASA grant NAGW-2515. Much of this work was performed while the authors were at JILA, University of Colorado, Boulder, using the computational facilities of the Laboratory for Computational Dynamics.

APPENDIX

DERIVATION OF THE CONSTANTS C_{nl}

In this Appendix, we return to the notation at the beginning of § 2; L will no longer denote a linear functional, but instead a measure of the total horizontal wavenumber, and K will no longer be associated with a-coefficients.

We consider adiabatic eigenfunctions satisfying a perfectly reflecting boundary condition; to a first approximation, the displacement eigenfunctions $\boldsymbol{\xi}$ can be written in polar coordinates (r, θ, ϕ) in the form

$$\boldsymbol{\xi}(\mathbf{r}, t) = R[(\xi_r P_l^m, \xi_h L^{-1} dP_l^m/d\theta, \xi_h imL^{-1} \csc \theta P_l^m) e^{im\phi - i\omega_{nlm}t}] , \tag{A1}$$

where (here) $L = [l(l+1)]^{1/2}$ and $P_l^m(\cos\theta)$ is the associated Legendre function of the first kind, of degree l and order m, normalized so that the integral of its square from 0 to 1 is unity. The frequency ω_{nlm} and the amplitude functions $\xi_r(r)$ and $\xi_h(r)$ of the radial and horizontal components of the displacement eigenfunctions are real. The amplitudes ξ_r and ξ_h depend on n and l, but we do not include the dependence in the notation.

Equation (3) exhibits the splitting kernel for the singlet (n, l, m) in its degenerate form. Note that the symbols L_{nl} and K_{nl} again pertain to the (n, l) multiplet. The factors in equation (3) that depend on the radial coordinate r are given by

$$K_{nl}(r) = (\xi_r^2 + \xi_h^2 - 2L^{-1}\xi_r \xi_h)\rho r^2 / I_{nl} , \tag{A2}$$

and

$$L_{nl}(r) = \xi_h^2 \rho r^2 / I_{nl} , \tag{A3}$$

where

$$I_{nl} = \int_0^R (\xi_r^2 + \xi_h^2)\rho r^2 \, dr , \tag{A4}$$

and ρ is the density of the background state. The components of the kernel that depend on the angular variable $\mu = \cos\theta$ are

$$W_{lm}(\mu) = P_l^m(\mu)^2 , \tag{A5}$$

and

$$X_{lm}(\mu) = L^{-2}\left[\frac{d}{d\mu} H_{lm}(\mu) - W_{lm}(\mu)\right], \tag{A6}$$

where

$$H_{lm}(\mu) - \frac{1}{2}(1-\mu^2)\frac{dW_{lm}(\mu)}{d\mu} + \mu W_{lm}(\mu) . \tag{A7}$$

In the case of high-order solar p-modes, the term $L_{nl}(r)$, which is proportional to ξ_h^2, is small compared with $K_{nl}(r)$ except near the inner turning point, and therefore

$$K_{nlm}(r, \mu) \simeq K_{nl}(r) W_{lm}(\mu) . \tag{A8}$$

Since the a-coefficient a_J is the linear combination

$$a_J = \sum_{m=-l}^{l} \gamma_{lm} \Delta_o \omega_{nlm} , \tag{A9}$$

the functions

$$\mathcal{W}_{jl} \equiv \sum_{m=-l}^{l} \gamma_{lm} W_{lm} \tag{A10}$$

and

$$\mathcal{X}_{jl} \equiv \sum_{m=-l}^{l} \gamma_{lm} X_{lm} \tag{A11}$$

satisfy

$$K_J(r, \mu) = K_{nl}(r)\mathcal{W}_{jl}(\mu) + L_{nl}(r)\mathcal{X}_{jl}(\mu) , \tag{A12}$$

where one can obtain γ by inverting the linear relations (1).

We seek a vector $\boldsymbol{\lambda}$ that is zero in all components except those corresponding to the (n, l) multiplet (in the example we consider explicitly, there are $j_{\max}/2 = 6$ values of j corresponding to this multiplet, and hence 6 indices in J for which $\lambda_J \neq 0$), such that the linear combination of the a-coefficient kernels have the properties discussed in § 4, namely,

$$\boldsymbol{\lambda} \cdot \boldsymbol{K}(r, \mu) \begin{cases} \leq 0 , & \mu \geq \mu_0(r) \\ \leq 0 , & \mu \leq \mu_0(r) \end{cases} , \tag{A13}$$

$$\int \boldsymbol{\lambda} \cdot \boldsymbol{K}(r, \mu) d\mu = 0 , \tag{A14}$$

$$\int dr \int_0^{\mu_0(r)} d\mu \boldsymbol{\lambda} \cdot \boldsymbol{K}(r, \mu) = -1 , \qquad \int dr \int_{\mu_0(r)}^1 d\mu \boldsymbol{\lambda} \cdot \boldsymbol{K}(r, \mu) = 1 . \tag{A15}$$

We use the degeneracy of the a-coefficient kernels to treat the angular dependence separately as a first approximation. The functions W_{jl} dominate the angular dependence of the kernels. We find the coefficients $\boldsymbol{\lambda}$ so that the linear combination $\sum_j \lambda_{jnl} W_{jl}$ satisfies

$$\int_0^1 \sum_j \lambda_{jnl} \mathscr{W}_{jl}(\mu) d\mu = 0 , \tag{A16}$$

using the SOLA method described by Pijpers & Thompson (1992, 1994) with target function

$$T(\mu) = -A_\beta \mu^{2\beta} (\mu^2 - 1)(\mu^2 - \mu_0^2) , \tag{A17}$$

where

$$\beta = \frac{5\mu_0^2 - 1}{2(1 - \mu_0^2)} , \tag{A18}$$

$$A_\beta = \frac{1}{16} \frac{(2\beta + 5)^{\beta + 7/2}}{(2\beta + 1)^{\beta + 1/2}} , \tag{A19}$$

and μ_0 is the specified colatitude at which T changes sign.

In our computational example, we used $\beta = 4$. For $j_{\max} = 12$, the asymptotic form of the a-coefficient kernels shows that it is impossible to fit the target function adequately for $\beta \geq 5$, while one can fit it almost perfectly when $\beta \leq 4$. We chose $\beta = 4$ because it gives the highest value of μ_0, namely $3/(13)^{1/2}$, which corresponds to a latitude of 56°.3.

The functions $H_{lm}(\mu)$ vanish at $\mu = 0$ and $\mu = 1$. This follows from their definition (A7) together with the facts that $\{W_{lm}\}$ are even functions of μ and that $W_{lm}(1) = 0$. Thus the constraint (A16), together with the definition of $X_{jl}(\mu)$ (determined by eqns. [A6], [A7], [A10], and [A11]), imply that the integral from $\mu = 0$ to $\mu = 1$ of $\sum_j \lambda_{jnl} X_{jl}(\mu)$ vanishes also. Consequently, condition (A16) ensures that the corresponding linear combination of the full data kernels satisfies condition (A14).

We find that for the value of $\boldsymbol{\lambda}$ determined using SOLA, the linear combination of the full kernels, $\boldsymbol{\lambda} \cdot \boldsymbol{K}(r, \mu)$, changes sign only once, at $\mu_0(r)$. Therefore, the conditions (A15) can be satisfied simply by scaling $\boldsymbol{\lambda}$. It is straightforward to show that

$$C_{nl} \equiv \int_0^R dr \int_{\mu_0(r)}^1 d\mu \boldsymbol{\lambda} \cdot \boldsymbol{K}(r, \mu) \tag{A20}$$

$$= \int_0^R [1 - \mathscr{E}(r)][K_{nl}(r) - L_{nl}(r)] - L_{nl}(r) H_{nl}(\mu_0)(r) dr \tag{A21}$$

$$\simeq 1 , \tag{A22}$$

where the small parameter $\mathscr{E}(r)$ is given by

$$\mathscr{E}(r) = 1 - \int_{\mu_0(r)}^1 \sum_j \lambda_{jnl} \mathscr{W}_{jl}(\mu) d\mu . \tag{A23}$$

Thus a kernel for the difference between the average angular velocity at low and at high latitides is given by

$$\frac{1}{C_{nl}} \boldsymbol{\lambda} \cdot \boldsymbol{K}(r, \mu) , \tag{A24}$$

and an unbiased estimator of the difference is

$$\frac{1}{C_{nl}} \boldsymbol{\lambda} \cdot \boldsymbol{\delta} . \tag{A25}$$

Since the kernels $C_{nl}^{-1}\boldsymbol{\lambda} \cdot \boldsymbol{K}(r, \mu)$ all integrate to zero and have a single sign change quite close to the cone $\mu = \mu_0$, so do linear combinations of them. Therefore, by averaging collections of these kernels, we can obtain new kernels that satisfy the constraints but involve more of the data with smaller coefficients, decreasing the variance of the corresponding estimate. We followed this strategy, averaging sets of 100 kernels for different multiplets with similar turning depths, to estimate differences in angular velocity across the surface $\mu = \mu_0$ averaged over different ranges of radius. The resulting estimates have uncertainties smaller than the individual estimates $C_{nl}^{-1}\boldsymbol{\lambda} \cdot \boldsymbol{\delta}$ possess.

REFERENCES

Backus, G. E. 1970a, Proc. Natl. Acad. Sci., 65, 1
———. 1970b, Proc. Natl. Acad. Sci., 65, 281
———. 1970c, Proc. Natl. Acad. Sci., 67, 282
Backus, G. E., & Gilbert, F. 1968, Geophys. J.R.A.S., 16, 169
———. 1970, Phil. Trans. Roy. Soc. Lond. A., 266, 123
Benjamini, Y., & Stark, P. B. 1996, J. Am. Stat. Assoc., in press
Bickel, P. J., & Doksum, K. A. 1977, Mathematical Statistics: Basic Ideas and Selected Topics (San Francisco: Holden-Day)
Brown, T. M., Christensen-Dalsgaard, J., Dziembowski, W. A., Goode, P., Gough, D. O., & Morrow, C. A. 1989, ApJ, 343, 526
Christensen-Dalsgaard, J., Schou, J., & Thompson, M. J. 1990, MNRAS, 242, 353
Däppen, W., Gough, D. O., Kosovichev, A. G., & Thompson, M. J. 1991, in Challenges to Theories of the Structure of Moderate-Mass Stars, ed. D. O. Gough & J. Toomre (New York: Springer-Verlag), 111
Gough, D. O. 1985, Sol. Phys., 100, 65
Gough, D. O., & Toomre, J. 1991, ARA&A, 29, 627
Johnson, L. E., & Gilbert, F. 1972, Methods Comput. Phys., 12, 231
Lehmann, E. L. 1986, Testing Statistical Hypotheses (2d ed.; New York: Wiley)
Masters, T. G. 1979, Geophys. J.R.A.S., 57, 507
The Math Works, Inc. 1985–1990, MATLAB for Unix Computers: User's Guide (Natick, MA: The Math Works, Inc.)
Oldenburg, D. W. 1979, Geophys., 44, 1218
———. 1981, Geophys. J.R.A.S., 65, 359
Parker, R. L. 1970, Geophys. J.R.A.S., 22, 121
Pijpers, F. P., & Thompson, M. J. 1992, A&A, 262, L33
———. 1994, A&A, 281, 231
Sekii, T. 1993, MNRAS, 264, 1018
———. 1995, MNRAS, submitted
Stark, P. B. 1992, J. Geophys. Res., 97, 14055
Woodard, M. F., & Libbrecht, K. G. 1993, ApJ, 402, L77