

Spectral estimation on a sphere in geophysics and cosmology

F. A. Dahlen* and Frederik J. Simons†

Department of Geosciences, Princeton University, Princeton, NJ 08544, USA. E-mail: fjsimons@alum.mit.edu

Accepted 2008 May 15. Received 2008 May 15; in original form 2007 May 21

SUMMARY

We address the problem of estimating the spherical-harmonic power spectrum of a statistically isotropic scalar signal from noise-contaminated data on a region of the unit sphere. Three different methods of spectral estimation are considered: (i) the spherical analogue of the one-dimensional (1-D) periodogram, (ii) the maximum-likelihood method and (iii) a spherical analogue of the 1-D multitaper method. The periodogram exhibits strong spectral leakage, especially for small regions of area $A \ll 4\pi$, and is generally unsuitable for spherical spectral analysis applications, just as it is in 1-D. The maximum-likelihood method is particularly useful in the case of nearly-whole-sphere coverage, $A \approx 4\pi$, and has been widely used in cosmology to estimate the spectrum of the cosmic microwave background radiation from spacecraft observations. The spherical multitaper method affords easy control over the fundamental trade-off between spectral resolution and variance, and is easily implemented regardless of the region size, requiring neither non-linear iteration nor large-scale matrix inversion. As a result, the method is ideally suited for most applications in geophysics, geodesy or planetary science, where the objective is to obtain a spatially localized estimate of the spectrum of a signal from noisy data within a pre-selected and typically small region.

Key words: Time series analysis; Fourier analysis; Inverse theory; Spatial analysis.

1 INTRODUCTION

Problems involving the spectral analysis of data on the surface of a sphere arise in a variety of geodetic, geophysical, planetary, cosmological and other applications. In the vast majority of such applications the data are either inherently unavailable over the whole sphere, or the desired result is an estimate that is localized to a geographically limited portion thereof. In geodesy, statistical properties of gravity fields often need to be determined using data from an incompletely sampled sphere (e.g. Hwang 1993; Albertella *et al.* 1999; Pail *et al.* 2001; Swenson & Wahr 2002; Simons & Dahlen 2006). Similar problems arise in the study of (electro)magnetic anomalies in earth, planetary (e.g. Lesur 2006; Thébault *et al.* 2006) and even medical (e.g. Maniar & Mitra 2004; Chung *et al.* 2007) contexts. More specifically, in geophysics and planetary science, the local mechanical strength of the terrestrial or a planetary lithosphere can be inferred from the cross-spectrum of the surface topography and gravitational anomalies (e.g. McKenzie & Bowin 1976; Turcotte *et al.* 1981; Simons *et al.* 1997; Wicczorek & Simons 2005; Wicczorek 2007). Workers in astronomy and cosmology seek to estimate the spectrum of the pointwise function that characterizes the angular distribution of distant galaxies catalogued in sky surveys (e.g. Hauser & Peebles 1973; Peebles 1973; Tegmark 1995). An even more important problem in cosmology is to estimate the spectrum of the cosmic microwave background or CMB radiation, either from ground-based temperature data collected in a limited region of the sky or from spacecraft data that are contaminated by emission from our own galaxy and other bright non-cosmological radio sources (e.g. Górski 1994; Bennett *et al.* 1996; Tegmark 1996, 1997; Tegmark *et al.* 1997; Bond *et al.* 1998; Oh *et al.* 1999; Wandelt *et al.* 2001; Hivon *et al.* 2002; Mortlock *et al.* 2002; Hinshaw *et al.* 2003; Efstathiou 2004). In this paper, consider the statistical problem of estimating the spherical-harmonic power spectrum of a noise-contaminated signal within a spatially localized region of a sphere. All of the methods that we discuss can easily be generalized to the multivariate case.

The material we discuss is mathematical, and some of the notation and the general conceptual framework will be more familiar to the cosmologist than to the geophysicist. To guide the novice reader, we offer the following. To represent scalar functions on a spherical surface (see Section 2.1), spherical harmonics (see Section 2.2) are ideal, since, properly normalized, they constitute an orthonormal basis on the

*Deceased.

†Formerly at: University College London, Department of Earth Science, Gower Street, London WC1E 6BT, UK.

sphere. In global geophysics, their use is widespread in geodesy (e.g. Lambeck 1988), geomagnetism (e.g. Blakely 1995), and seismology (e.g. Dahlen & Tromp 1998). The orthonormality is such that when the product of any two spherical harmonics is integrated over the surface of the entire sphere, the result is either one (if both harmonics are of identical degree and order) or zero (if they differ)—see eq. (8). Furthermore, the orthonormality helps produce convenient expressions for the integrals of three spherical harmonics (see Section 2.3), which are frequently needed and that, like the spherical harmonics themselves, can be computed via recursion.

As often arises in the physical sciences, we do not have access to, or may simply not be interested in, the values or the properties of the function outside some particular subregion of the sphere. In such cases it is convenient to think of the spatially restricted function as the projection of a function that is, itself, globally defined (see Section 2.4). When projected onto a subregion of the sphere, the orthonormality of the spherical harmonics is, unfortunately, lost. Instead of the mathematically very convenient delta function, the integrated product of two spherical harmonics now yields a non-diagonal quantity—see eq. (53). The properties of this ‘spherical localization kernel’ were studied in detail by Simons *et al.* (2006) and further developed by Simons & Dahlen (2007), who termed its eigenfunctions ‘spherical Slepian functions’, after David Slepian (1923–2007), whose seminal work on the eigenfunctions of the analogous 1-D Dirichlet kernel (Slepian 1983) led to the 1-D multitaper method of spectral analysis (Thomson 1982; Haykin 1991; Percival & Walden 1993).

Practical applications most commonly involve real-valued measurements that are contaminated by additive noise, for which cosmologists, though admittedly, rarely, geophysicists, are comfortable adopting idealized models (see Section 2.5). Two different statistical problems in the treatment of incomplete and noisy spherical data naturally arise in this context, namely, (i) how to find the ‘best’ estimate of the signal (e.g. by finding its spherical harmonic expansion coefficients) given such data, and (ii) how to construct from such data the ‘best’ estimate of the power spectral density of the signal (in other words, a particular quadratic combination of its spherical harmonic coefficients).

Various solutions to problem (i) were presented by Simons & Dahlen (2006). Solving this problem in the spherical harmonic domain requires the damped inversion of the ill-conditioned localization kernel and produces estimates of the coefficients whose errors are strongly correlated. As Simons & Dahlen (2006) showed, the more intuitive solution expands the unknown signal as a truncated series of Slepian functions and solves directly for their expansion coefficients. This is conceptually as well as computationally easier, and produces estimated coefficients with much less statistical correlation between them. There is no free lunch, however: no information is gained by merely changing bases, and the overall conclusive metric—how well one can recover the signal from incomplete and noisy data in a root-mean-square sense—can be optimized to almost equal levels by either the damped spherical harmonic or the truncated Slepian function approach. Slepian functions are orthogonal on both the entire as well as the cut sphere (Gilbert & Slepian 1977; Grünbaum *et al.* 1982; Simons *et al.* 2006). In the sense that the underlying difficulty of problem (i) is the non-orthogonality of the spherical harmonics over partially or irregularly sampled observation domains (see, e.g. Sneeuw 1994), the Slepian approach is a relative of the Gram-Schmidt solution strategies proposed by Kaula (1967) and Hwang (1993) in geodesy, and by Górski (1994) in cosmology, or of the singular-value-decomposition approaches popular in contemporary geodetic inversions (Xu 1992a,b, 1998) and elsewhere (Wingham 1992). All of these attempt to find a new, orthogonal basis for the estimation problem, and while they differ in computational details and statistical performance, the Slepian philosophy is essentially their limiting case for regular samplings.

Problem (ii), estimating the spherical-harmonic power spectral density of an incompletely and noisily observed field on the surface of the sphere, is restated in Section 3. We hold it as scientifically self-evident that knowledge of the power spectrum of some process is often all we want to recover from the measured data. Indeed, physical properties of interest often end up as the unknown parameters directly influencing the spectrum that, itself, needs to be estimated from spatially distributed observations. This is the case, for instance, in modelling the source depth of planetary gravitational or magnetic fields, in determining the strength of planetary lithospheres from the cross-spectral density of surface topography and gravitational anomalies, or in deriving the parameters of cosmological models via their effect on the power spectrum of the cosmic microwave background radiation. We also take it for granted that we want to retain the flexibility and physical appeal of the spherical harmonics as long as practical before switching to 2-D Cartesian formulations (to which they are asymptotically equivalent, see Simons *et al.* 2006; Simons & Dahlen 2007). We finally stress that solving problem (ii) is not the same as first solving problem (i) and then constructing the power spectrum. The ‘localized’ spherical harmonic expansion coefficients of an incompletely observed field relate to the ‘global’ ones of the underlying whole-sphere process via the spatio-spectral localization kernel mentioned above (and which is the unifying force behind all of our considerations)—see eq. (23). However the power spectral density of a locally observed process is coupled to the ‘true’ value that can be recovered by observing the whole sphere via a particular combination of the squares of the elements of the same kernel—see eq. (57). Thus, designing the mathematical machinery to recover the actual spectrum from incomplete and noisy data requires a separate treatment, the subject of this paper, in which, however, we once again give a starring role to the spherical Slepian functions. For mathematical convenience, we deal with regularly (if incompletely) sampled fields such as are available to cosmologists, and sometimes geophysicists, via satellite surveys. Extensions of the method to non-uniform sampling such as may be more characteristic of data acquisition for terrestrial geophysics are likely to lead to generalizations of our method (see, e.g. Bronez 1988, for a discussion in the Cartesian plane), that should be better understood with the Slepian methodology presented here as their limiting case.

Not having to worry about partial coverage is discussed in Section 4; the disastrous effects of not worrying about it are discussed in Section 5. Statistical approaches to solving problem (ii) that have been popular in cosmology appear in Section 6, and then, finally, our new Slepian multitaper approach is discussed in detail in Section 7. Practical formulas to characterize bias and variance of the various spectral estimates are derived in Section 8. A comparison of their performance is given in Section 9, and an example taken from cosmology in Section 10.

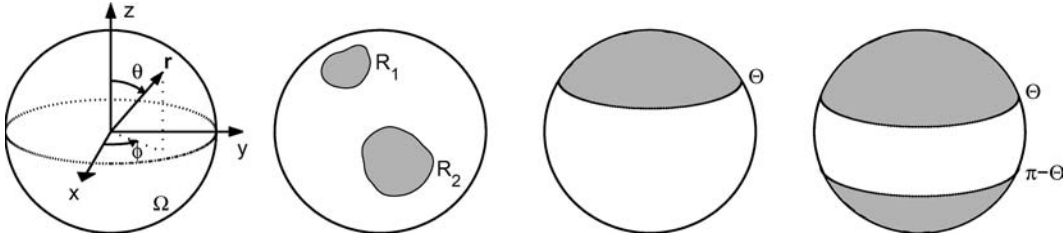


Figure 1. Geometry of the unit sphere $\Omega = \{\mathbf{r} : \|\mathbf{r}\| = 1\}$, showing, from left- to right-hand side, colatitude $0 \leq \theta \leq \pi$ and longitude $0 \leq \phi < 2\pi$, an arbitrary spacelimited region $R = R_1 \cup R_2 \cup \dots$; an axisymmetric polar cap $\theta \leq \Theta$; and a double polar cap $\theta \leq \Theta$ and $\pi - \Theta \leq \theta \leq \pi$.

2 PRELIMINARIES

We denote points on the unit sphere Ω by \mathbf{r} rather than the more commonly used $\hat{\mathbf{r}}$, reserving the circumflex to identify an estimate of a statistical variable. We use R to denote a region of Ω within which we have data from which we wish to extract a spatially localized spectral estimate; the region may consist of a number of unconnected subregions, $R = R_1 \cup R_2 \cup \dots$, and it may have an irregularly shaped boundary, as shown in Fig. 1. We shall illustrate our results using two more regularly shaped regions, namely a polar cap of angular radius Θ and a pair of antipodal caps of common radius Θ , separated by an equatorial cut of width $\pi - 2\Theta$, as shown in the rightmost two panels of Fig. 1. An axisymmetric cap, which may be rotated to any desired location on the sphere, is an obvious initial choice for conducting localized spatio-spectral analyses of planetary or geodetic data whereas an equatorial cut arises in the spectral analysis of spacecraft CMB temperature data, because of the need to mask foreground contamination from our own galactic plane. The surface area of the region R is A .

2.1 Spatial, pixel and spectral bases

We switch back and forth among three different representations or bases which may be used to specify a given function on Ω .

- (i) The familiar ‘spatial basis’ in which a piecewise continuous function f is represented by its values $f(\mathbf{r})$ at points \mathbf{r} on Ω .
- (ii) The ‘pixel basis’ in which the region R we wish to analyse is subdivided into equal-area pixels of solid angle $\Delta\Omega = 4\pi J^{-1}$. A function f is represented in the pixel basis by a J -dimensional column vector $\mathbf{f} = (f_1 \ f_2 \ \dots \ f_J)^T$, where $f_j = f(\mathbf{r}_j)$ is the value of f at pixel j , and J is the total number of pixels. Equal-area pixelization of a 2-D function $f(\mathbf{r})$ on a portion R of Ω is analogous to the equispaced digitization of a finite 1-D time-series $f(t)$, $0 \leq t \leq T$. Integrals over the region R will be assumed to be approximated with sufficient accuracy by a Riemann sum over pixels:

$$\int_R f(\mathbf{r}) d\Omega \approx \Delta\Omega \sum_{j=1}^J f_j. \quad (1)$$

Henceforth, in transforming between the spatial and pixel bases, we shall ignore the approximate nature of the equality in eq. (1). In cosmology, such an equal-area pixelization scheme is commonly used in the collection and analysis of CMB temperature data (e.g. Górski *et al.* 2005); in the present paper we shall make extensive use of the pixel basis, even in the case that R is the whole sphere Ω , primarily because it enables an extremely succinct representation of expressions that would be much more unwieldy if expressed in the spatial basis. As simple examples we note that we can write integrals over R as

$$\int_R f(\mathbf{r}) \tilde{f}(\mathbf{r}) d\Omega = \Delta\Omega \mathbf{f}^T \tilde{\mathbf{f}} \quad \text{and} \quad \int_R F(\mathbf{r}, \mathbf{r}') f(\mathbf{r}') d\Omega' = \Delta\Omega \mathbf{F} \mathbf{f}, \quad (2)$$

for any function $f(\mathbf{r})$ and $F(\mathbf{r}, \mathbf{r}')$, and the double integral of the product of two symmetric functions as

$$\iint_R F(\mathbf{r}, \mathbf{r}') \tilde{F}(\mathbf{r}', \mathbf{r}) d\Omega d\Omega' = (\Delta\Omega)^2 \text{tr}(\tilde{\mathbf{F}} \mathbf{F}) = (\Delta\Omega)^2 \text{tr}(\mathbf{F} \tilde{\mathbf{F}}), \quad (3)$$

where \mathbf{F} and $\tilde{\mathbf{F}}$ are symmetric matrices of dimension $J \times J$ with elements $F_{jj'} = F(\mathbf{r}_j, \mathbf{r}_{j'})$ and $\tilde{F}_{jj'} = \tilde{F}(\mathbf{r}_j, \mathbf{r}_{j'})$, and we have blithely replaced the symbol \approx by $=$ as advertized. We shall consistently write pixel-basis column vectors and matrices using a bold, lower-case and upper-case, sans serif font, respectively, as above. It further follows that $\mathbf{f}^T \mathbf{F} \mathbf{f} = \text{tr}(\mathbf{f} \mathbf{f}^T \mathbf{F})$.

- (iii) The ‘spectral basis’ in which a function f is represented in terms of its spherical harmonic expansion coefficients:

$$f(\mathbf{r}) = \sum_{lm} f_{lm} Y_{lm}(\mathbf{r}), \quad \text{where} \quad f_{lm} = \int_{\Omega} f(\mathbf{r}) Y_{lm}^*(\mathbf{r}) d\Omega. \quad (4)$$

The harmonics $Y_{lm}(\mathbf{r})$ used in this paper are the complex surface spherical harmonics defined by Edmonds (1996), with properties that we review briefly in the next section. An asterisk in eq. (4) and elsewhere in this paper denotes the complex conjugate.

2.2 Spherical harmonics

The functions $Y_{lm}(\mathbf{r}) = Y_{lm}(\theta, \phi)$ are defined by the relations (e.g. Jackson 1962; Edmonds 1996; Dahlen & Tromp 1998)

$$Y_{lm}(\theta, \phi) = X_{lm}(\theta) \exp(im\phi), \quad (5)$$

$$X_{lm}(\theta) = (-1)^m \left(\frac{2l+1}{4\pi} \right)^{1/2} \left[\frac{(l-m)!}{(l+m)!} \right]^{1/2} P_m(\cos \theta), \quad (6)$$

$$P_{lm}(\mu) = \frac{1}{2^l l!} (1 - \mu^2)^{m/2} \left(\frac{d}{d\mu} \right)^{l+m} (\mu^2 - 1)^l, \quad (7)$$

where $0 \leq \theta \leq \pi$ is the colatitude and $0 \leq \phi < 2\pi$ is the longitude. The integer $0 \leq l \leq \infty$ is the angular degree of the spherical harmonic and $-l \leq m \leq l$ is its angular order. The function $P_{lm}(\mu)$ defined in eq. (7) is the associated Legendre function of degree l and order m . The choice of the multiplicative constants in eqs (5)–(7) orthonormalizes the spherical harmonics on the unit sphere so that there are no $\sqrt{4\pi}$ factors in the spatial-to-spectral basis transformation (4):

$$\int_{\Omega} Y_{lm}^*(\mathbf{r}) Y_{l'm'}(\mathbf{r}) d\Omega = \delta_{ll'} \delta_{mm'}. \quad (8)$$

The spherical harmonics $Y_{lm}(\mathbf{r})$ are eigenfunctions of the Laplace–Beltrami operator, $\nabla^2 = \partial_\theta^2 + \cot \theta \partial_\theta + (\sin \theta)^{-2} \partial_\phi^2$, with associated eigenvalues $-l(l+1)$. Harmonics of negative and positive order are related by $Y_{l-m}(\mathbf{r}) = (-1)^m Y_{lm}^*(\mathbf{r})$. The $l \rightarrow \infty$ asymptotic wavenumber of a spherical harmonic of degree l is $[l(l+1)]^{1/2} \approx l + 1/2$ (Jeans 1923). A 2-D Dirac delta function on the sphere Ω , with the replication property

$$\int_{\Omega} \delta(\mathbf{r}, \mathbf{r}') f(\mathbf{r}') d\Omega' = f(\mathbf{r}), \quad (9)$$

can be expressed as a spherical harmonic expansion in the form

$$\delta(\mathbf{r}, \mathbf{r}') = \sum_{lm} Y_{lm}(\mathbf{r}) Y_{lm}^*(\mathbf{r}') = \frac{1}{4\pi} \sum_l (2l+1) P_l(\mathbf{r} \cdot \mathbf{r}'), \quad (10)$$

where $P_l(\mu) = P_{l0}(\mu)$ is the Legendre polynomial of degree l and the second equality is a consequence of the spherical harmonic addition theorem. A 1-D Dirac delta function can be expanded in terms of Legendre polynomials as

$$\delta(\mu - \mu') = \frac{1}{2} \sum_l (2l+1) P_l(\mu) P_l(\mu'). \quad (11)$$

In eqs (4), (10), (11) and throughout this paper we refrain from writing the limits of sums over spherical harmonic indices except in instances where we wish to be emphatic or it is essential. All spherical-harmonic or spectral-basis sums without specifically designated limits will either be infinite, as in the case of the sums over degrees $0 \leq l \leq \infty$ above, or they will be limited naturally, for example, by the restriction upon the orders $-l \leq m \leq l$ or by the selection rules governing the Wigner 3- j symbols which we discuss next.

2.3 Wigner 3- j and 6- j symbols

We shall make frequent use of the well-known formula for the surface integral of a product of three spherical harmonics:

$$\int_{\Omega} Y_{lm}(\mathbf{r}) Y_{pq}(\mathbf{r}) Y_{l'm'}(\mathbf{r}) d\Omega = \left[\frac{(2l+1)(2p+1)(2l'+1)}{4\pi} \right]^{1/2} \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & l' \\ m & q & m' \end{pmatrix}, \quad (12)$$

where the arrays of integers are Wigner 3- j symbols (Edmonds 1996; Messiah 2000). Both of the 3- j symbols in eq. (12) are zero except when (i) the bottom-row indices sum to zero, $m + q + m' = 0$, and (ii) the top-row indices satisfy the triangle condition $|l - l'| \leq p \leq l + l'$. The first symbol, with all zeroes in the bottom row, is non-zero only if $l + p + l'$ is even. A product of two spherical harmonics can be written as a sum of harmonics in the form

$$Y_{lm}(\mathbf{r}) Y_{l'm'}(\mathbf{r}) = \sum_{pq} \left[\frac{(2l+1)(2p+1)(2l'+1)}{4\pi} \right]^{1/2} \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & l' \\ m & q & m' \end{pmatrix} Y_{pq}^*(\mathbf{r}). \quad (13)$$

The analogous formulas governing the Legendre polynomials $P_l(\mu)$ are

$$\int_{-1}^1 P_l(\mu) P_p(\mu) P_{l'}(\mu) d\mu = 2 \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2 \quad \text{and} \quad P_l(\mu) P_{l'}(\mu) = \sum_p (2p+1) \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2 P_p(\mu). \quad (14)$$

Two orthonormality relations governing the 3- j symbols are useful in what follows:

$$\sum_{st} (2s+1) \begin{pmatrix} l & p & s \\ m & q & t \end{pmatrix} \begin{pmatrix} l & p & s \\ m' & q' & t \end{pmatrix} = \delta_{mm'} \delta_{qq'}, \quad (15)$$

$$\sum_{mm'} \begin{pmatrix} l & p & l' \\ m & q & m' \end{pmatrix} \begin{pmatrix} l & p' & l' \\ m & q' & m' \end{pmatrix} = \frac{1}{2p+1} \delta_{pp'} \delta_{qq'}, \quad (16)$$

provided the enclosed indices satisfy the triangle condition. The Wigner 6- j symbol is a particular symmetric combination of six degree indices which arises in the quantum mechanical analysis of the coupling of three angular momenta; among a welter of formulas relating the 3- j and 6- j symbols, the most useful for our purposes are (Varshalovich *et al.* 1988; Messiah 2000)

$$\sum_{t't'v'v'q} (-1)^{u+u'+p+v+v'+q} \begin{pmatrix} s & e & s' \\ t & f & t' \end{pmatrix} \begin{pmatrix} u & e' & u' \\ -v & f' & v' \end{pmatrix} \begin{pmatrix} s & p & u' \\ t & q & -v' \end{pmatrix} \begin{pmatrix} u & p & s' \\ v & -q & t' \end{pmatrix} = \frac{\delta_{ee'} \delta_{ff'}}{2e+1} \begin{Bmatrix} s & e & s' \\ u & p & u' \end{Bmatrix}, \quad (17)$$

$$\sum_e (-1)^{p+e} (2e+1) \begin{Bmatrix} s & e & s' \\ u & p & u' \end{Bmatrix} \begin{pmatrix} s & e & s' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u & e & u' \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} s & p & u' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u & p & s' \\ 0 & 0 & 0 \end{pmatrix}, \quad (18)$$

where the common array in curly braces is the 6- j symbol. Two simple special cases of the 3- j and 6- j symbols will be needed:

$$\begin{pmatrix} l & 0 & l' \\ 0 & 0 & 0 \end{pmatrix} = \frac{(-1)^l}{\sqrt{2l+1}} \delta_{ll'} \quad \text{and} \quad \begin{Bmatrix} s & 0 & s' \\ u & p & u' \end{Bmatrix} = \frac{(-1)^{s+p+u}}{\sqrt{(2s+1)(2u+1)}} \delta_{ss'} \delta_{uu'}. \quad (19)$$

Finally, we shall have occasion to use an asymptotic relation for the 3- j symbols, namely

$$(2p+1) \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2 \approx \frac{4\pi}{2l+1} [X_{p|l-l'|}(\pi/2)]^2 \approx \frac{4\pi}{2l'+1} [X_{p|l-l'|}(\pi/2)]^2 \quad (20)$$

which is valid for $l \approx l' \gg p$ (Brussaard & Tolhoek 1957; Edmonds 1996). All of the degree and order indices in eqs (12)–(20) and throughout this paper are integers.

Well-known recursion relations allow for the numerically stable computation of spherical harmonics (Libbrecht 1985; Dahlen & Tromp 1998; Masters & Richards-Dinger 1998) and Wigner 3- j and 6- j symbols (Schulten & Gordon 1975; Luscombe & Luban 1998) to high degree and order. The numerous symmetry relations of the Wigner symbols can be exploited for efficient data base storage (Rasch & Yu 2003).

2.4 Projection operator

We use $f^R(\mathbf{r})$ to denote the restriction of a function $f(\mathbf{r})$ defined everywhere on the sphere Ω to the region R , that is,

$$f^R(\mathbf{r}) = \begin{cases} f(\mathbf{r}) & \text{if } \mathbf{r} \in R, \\ 0 & \text{otherwise.} \end{cases} \quad (21)$$

In the pixel basis restriction to the region R is accomplished with the aid of a projection operator:

$$\mathbf{f}^R = \mathbf{D}\mathbf{f} \quad \text{where} \quad \mathbf{D} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \quad (22)$$

In writing eqs (22) we have assumed that the entire sphere has been pixelized with those pixels located within R grouped together in the upper left-hand corner, so that \mathbf{I} is the identity operator within R . It is evident that $\mathbf{D}^2 = \mathbf{D}$ and $\mathbf{D} = \mathbf{D}^T$, as must be true for any (real) projection operator. In the spectral basis it is easily shown that the spherical harmonic expansion coefficients of $f^R(\mathbf{r})$ are given by

$$f_{lm}^R = \sum_{l'm'} D_{lm,l'm'} f_{l'm'}, \quad \text{where} \quad D_{lm,l'm'} = \int_R Y_{lm}^*(\mathbf{r}) Y_{l'm'}(\mathbf{r}) d\Omega. \quad (23)$$

The quantities $D_{lm,l'm'}$ are the elements of a spectral-basis projection operator, a localization kernel, with properties analogous to those of the pixel-basis projector \mathbf{D} , namely

$$\sum_{pq} D_{lm,pq} D_{pq,l'm'} = D_{lm,l'm'} \quad \text{and} \quad D_{lm,l'm'} = D_{l'm',lm}^*. \quad (24)$$

The first of eqs (24) can be verified by using the definition (23) of $D_{lm,l'm'}$ together with the representation (9)–(10) of the Dirac delta function. Neither the pixel-basis projection operator \mathbf{D} nor the infinite-dimensional spectral-basis projection operator $D_{lm,l'm'}$ is invertible, except in the trivial case of projection onto the whole sphere, $R = \Omega$.

2.5 Signal, noise and data

We assume that the real-valued spatial-basis ‘signal’ of interest, which we denote by

$$s(\mathbf{r}) = \sum_{lm} s_{lm} Y_{lm}(\mathbf{r}), \quad (25)$$

is a realization of a zero-mean, Gaussian, isotropic, random process, with spherical harmonic coefficients s_{lm} satisfying

$$\langle s_{lm} \rangle = 0 \quad \text{and} \quad \langle s_{lm} s_{l'm'}^* \rangle = S_l \delta_{ll'} \delta_{mm'}, \quad (26)$$

where the angle brackets denote an average over realizations. Such a stochastic signal is completely characterized by its angular power spectrum S_l , $0 \leq l \leq \infty$, for which we further stipulate that $0 < S_l < \infty$. The second of eqs (26) stipulates that the covariance of the signal

is diagonal in the spectral representation. We denote the signal covariance matrix in the pixel basis by $\mathbf{S} = \langle \mathbf{s} \mathbf{s}^T \rangle$, where $\mathbf{s} = (s_1 \ s_2 \ \dots \ s_J)^T$ and $s_j = s(\mathbf{r}_j)$. To evaluate \mathbf{S} we note that

$$\begin{aligned} \langle s(\mathbf{r}_j) s(\mathbf{r}_{j'}) \rangle &= \sum_{lm} \sum_{l'm'} \langle s_{lm} s_{l'm'}^* \rangle Y_{lm}(\mathbf{r}_j) Y_{l'm'}^*(\mathbf{r}_{j'}) \\ &= \sum_{lm} S_l Y_{lm}(\mathbf{r}_j) Y_{lm}^*(\mathbf{r}_{j'}) \\ &= \frac{1}{4\pi} \sum_l (2l+1) S_l P_l(\mathbf{r}_j \cdot \mathbf{r}_{j'}). \end{aligned} \quad (27)$$

It is convenient in what follows to introduce the $J \times J$ symmetric matrix \mathbf{P}_l with elements

$$(\mathbf{P}_l)_{jj'} = \sum_m Y_{lm}(\mathbf{r}_j) Y_{lm}^*(\mathbf{r}_{j'}) = \left(\frac{2l+1}{4\pi} \right) P_l(\mathbf{r}_j \cdot \mathbf{r}_{j'}). \quad (28)$$

In particular, the pixel-basis covariance matrix may be written using this notation in the succinct form

$$\mathbf{S} = \sum_l S_l \mathbf{P}_l. \quad (29)$$

Eq. (29) shows that the signal covariance is not diagonal in the pixel representation. The total power of the signal integrated over the whole sphere is

$$S_{\text{tot}} = \int_{\Omega} \langle s^2(\mathbf{r}) \rangle d\Omega = \sum_l (2l+1) S_l, \quad (30)$$

and the power contained within the region R of area $A \leq 4\pi$ is

$$S_{\text{tot}}^R = \int_R \langle s^2(\mathbf{r}) \rangle d\Omega = \Delta\Omega \text{tr} \mathbf{S} = \frac{A}{4\pi} S_{\text{tot}}. \quad (31)$$

In general the signal $s(\mathbf{r})$ in eq. (25) is contaminated by random measurement noise,

$$\mathbf{n}(\mathbf{r}) = \sum_{lm} n_{lm} Y_{lm}(\mathbf{r}), \quad (32)$$

which we will also assume to be zero-mean, Gaussian and isotropic,

$$\langle n_{lm} \rangle = 0 \quad \text{and} \quad \langle n_{lm} n_{l'm'}^* \rangle = N_l \delta_{ll'} \delta_{mm'}, \quad (33)$$

with a known angular power spectrum N_l , $0 \leq l \leq \infty$. The covariance of the noise in the pixel basis is given by the analogue of eq. (29), namely $\mathbf{N} = \langle \mathbf{n} \mathbf{n}^T \rangle = \sum_l N_l \mathbf{P}_l$. The simplest possible case is that of white noise, $N_l = N = \Delta\Omega \sigma^2$; the pixel-basis noise covariance then reduces to $\mathbf{N} = \sigma^2 \mathbf{I}$, where σ is the root-mean-square measurement noise per pixel and \mathbf{I} is the $J \times J$ identity, by virtue of the pointwise relation

$$\sum_l \mathbf{P}_l = (\Delta\Omega)^{-1} \mathbf{I}. \quad (34)$$

Eq. (34) is the pixel-basis analogue of the spatial-basis representation (9)–(10) of the Dirac delta function. The covariance of white noise is diagonal in both the spectral and pixel bases.

The measured data, which we denote by $d(\mathbf{r})$ or $\mathbf{d} = (d_1 \ d_2 \ \dots \ d_J)^T$, consist of the signal plus the noise:

$$d(\mathbf{r}) = s(\mathbf{r}) + n(\mathbf{r}) \quad \text{or} \quad \mathbf{d} = \mathbf{s} + \mathbf{n}. \quad (35)$$

We assume that the signal and noise are uncorrelated; that is, $\langle \mathbf{n} \mathbf{s}^T \rangle = \langle \mathbf{s} \mathbf{n}^T \rangle = \mathbf{0}$. The pixel-basis covariance matrix of the data under these assumptions is

$$\mathbf{C} = \langle \mathbf{d} \mathbf{d}^T \rangle = \langle \mathbf{s} \mathbf{s}^T \rangle + \langle \mathbf{n} \mathbf{n}^T \rangle = \mathbf{S} + \mathbf{N} = \sum_l (S_l + N_l) \mathbf{P}_l. \quad (36)$$

It is noteworthy that there are two different types of stochastic averaging going on in the above discussion: $\langle s_{lm} s_{l'm'}^* \rangle$ or $\langle \mathbf{s} \mathbf{s}^T \rangle$ is planetary or cosmic averaging over all realizations of the signal $s(\mathbf{r})$ or \mathbf{s} , whereas $\langle n_{lm} n_{l'm'}^* \rangle$ or $\langle \mathbf{n} \mathbf{n}^T \rangle$ is averaging over all realizations of the measurement noise $n(\mathbf{r})$ or \mathbf{n} . In what follows we will use a single pair of angle brackets to represent both averages: $\langle \cdot \rangle = \langle \langle \cdot \rangle_{\text{signal}} \rangle_{\text{noise}} = \langle \langle \cdot \rangle_{\text{noise}} \rangle_{\text{signal}}$.

In practice the CMB temperature data $\mathbf{d} = \mathbf{s} + \mathbf{n}$ in a cosmological experiment are convolved with the beam response of the measurement antenna or antennae, which must be determined independently. Harmonic degrees l whose angular scale is less than the finite aperture of the beam cannot be resolved; for illustrative purposes in Section 10 we adopt a highly idealized noise model that accounts for this effect, namely

$$N_l = \Delta\Omega \sigma^2 \exp\left(\frac{l^2 \theta_{\text{fwhm}}^2}{8 \ln 2}\right), \quad (37)$$

where θ_{fwhm} is the full width at half-maximum of the beam, which is assumed to be Gaussian (Knox 1995). For moderate angular degrees the noise (37) is white but for the unresolvable degrees, $l \gg \sqrt{8 \ln 2} / \theta_{\text{fwhm}}$, it increases exponentially. Two other complications that arise in real-world cosmological applications will be ignored: (i) In general some pixels are sampled more frequently than others; in that case, the constant noise per pixel σ must be replaced by $\sigma_0 \nu_j^{-1/2}$, where ν_j is the number of observations of sample j . The resulting noise covariance

is then non-diagonal in both the spectral and pixel bases. (ii) CMB temperature data are generally collected in a variety of microwave bands, requiring consideration of the cross-covariance $\mathbf{C}_{\lambda\lambda'}$ between different wavelengths λ and λ' .

3 STATEMENT OF THE PROBLEM

We are now in a position to give a formal statement of the problem that will be addressed in this paper: given data $\mathbf{d} = \mathbf{s} + \mathbf{n}$ over a region R of the sphere Ω and given the noise covariance \mathbf{N} , estimate the spectrum S_l , $0 \leq l \leq \infty$, of the signal. This is the 2-D spherical analogue of the more familiar problem of estimating the power spectrum $S(\omega)$ of a 1-D time-series, given noise-contaminated data $d(t) = s(t) + n(t)$ over a finite time interval $0 \leq t \leq T$. The 1-D spectral estimation problem has been extremely well studied and has spawned a substantial literature (e.g. Thomson 1982, 1990; Haykin 1991; Mullis & Scharf 1991; Percival & Walden 1993). We shall compare three different spectral estimation methods: (i) the spherical analogue of the classical periodogram, which is unsatisfactory for the same strong spectral leakage reasons as in 1-D; (ii) the maximum-likelihood method, which has been developed and widely applied in CMB cosmology (e.g. Bond *et al.* 1998; Oh *et al.* 1999; Hinshaw *et al.* 2003) and (iii) a spherical analogue of the 1-D multitaper method (Wieczorek & Simons 2005; Simons & Dahlen 2006; Simons *et al.* 2006; Wieczorek & Simons 2007).

4 WHOLE-SPHERE DATA

It is instructive to first consider the case in which usable data $\mathbf{d} = \mathbf{s} + \mathbf{n}$ are available over the whole sphere, that is, $R = \Omega$. An obvious choice for the spectral estimator in that case is (e.g. Jones 1963; Kaula 1967; Grishchuk & Martin 1997)

$$\hat{S}_l^{\text{WS}} = \frac{1}{2l+1} \sum_m \left| \int_{\Omega} d(\mathbf{r}) Y_{lm}^*(\mathbf{r}) d\Omega \right|^2 - N_l, \quad (38)$$

where the first term is the conventional definition of the degree- l power of the data $d(\mathbf{r})$ and—as we shall show momentarily—the subtracted constant N_l corrects the estimate for the bias due to noise. In the pixel basis eq. (38) is rewritten in the form

$$\hat{S}_l^{\text{WS}} = \frac{(\Delta\Omega)^2}{2l+1} [\mathbf{d}^T \mathbf{P}_l \mathbf{d} - \text{tr}(\mathbf{N} \mathbf{P}_l)]. \quad (39)$$

The equivalence of eqs (38) and (39) can be confirmed with the aid of the whole-sphere double-integral identity

$$\text{tr}(\mathbf{P}_l \mathbf{P}_{l'}) = (\Delta\Omega)^{-2} (2l+1) \delta_{ll'}. \quad (40)$$

To verify the relation (40) it suffices to substitute the definition (28), transform from the pixel to the spatial basis, and utilize the spherical harmonic orthonormality relation (8). The superscript WS identifies the equivalent expressions (38)–(39) as the ‘whole-sphere estimator’; \hat{S}_l^{WS} is said to be a ‘quadratic estimator’ because it is quadratic in the data \mathbf{d} . Every spectral estimator that we shall consider subsequently, in the more general case $R \neq \Omega$, has the same general form as eqs (38)–(39): a first term that is quadratic in \mathbf{d} and a second, subtracted constant term that corrects for the bias due to noise.

The expected value of the whole-sphere estimator \hat{S}_l^{WS} is

$$\begin{aligned} \langle \hat{S}_l^{\text{WS}} \rangle &= \frac{(\Delta\Omega)^2}{2l+1} [\text{tr}(\mathbf{C} \mathbf{P}_l) - \text{tr}(\mathbf{N} \mathbf{P}_l)] \\ &= \frac{(\Delta\Omega)^2}{2l+1} \text{tr}(\mathbf{S} \mathbf{P}_l) \quad \text{noise bias cancels} \\ &= \frac{(\Delta\Omega)^2}{2l+1} \sum_{l'} S_{l'} \text{tr}(\mathbf{P}_l \mathbf{P}_{l'}) \\ &= S_l, \end{aligned} \quad (41)$$

where the first equation follows from $\langle \mathbf{d}^T \mathbf{P}_l \mathbf{d} \rangle = \text{tr}(\mathbf{C} \mathbf{P}_l)$ through eq. (36). The result (41) shows that, when averaged over infinitely many realizations, the whole-sphere expressions (38)–(39) will return an estimate that will coincide exactly with the true spectrum: $\langle \hat{S}_l^{\text{WS}} \rangle = S_l$. Such an estimator is said to be unbiased.

We denote the covariance of two whole-sphere estimates \hat{S}_l^{WS} and $\hat{S}_{l'}^{\text{WS}}$ at different angular degrees l and l' by

$$\Sigma_{ll'}^{\text{WS}} = \text{cov}(\hat{S}_l^{\text{WS}}, \hat{S}_{l'}^{\text{WS}}), \quad (42)$$

where as usual by $\text{cov}(d, d')$ we mean

$$\text{cov}(d, d') = \langle (d - \langle d \rangle)(d' - \langle d' \rangle) \rangle = \langle dd' \rangle - \langle d \rangle \langle d' \rangle. \quad (43)$$

To compute the covariance of a quadratic estimator such as (38)–(39) we make use of an identity due to Isserlis (1916),

$$\text{cov}(d_1 d_2, d_3 d_4) = \text{cov}(d_1, d_3) \text{cov}(d_2, d_4) + \text{cov}(d_1, d_4) \text{cov}(d_2, d_3), \quad (44)$$

which is valid for any four real scalar Gaussian random variables d_1, d_2, d_3 and d_4 . Using eq. (44) and the symmetry of the matrices $\mathbf{P}_l, \mathbf{P}_{l'}$ and \mathbf{C} to reduce the expression $\text{cov}(\mathbf{d}^T \mathbf{P}_l \mathbf{d}, \mathbf{d}^T \mathbf{P}_{l'} \mathbf{d})$, it is straightforward to show that

$$\Sigma_{ll'}^{\text{WS}} = \frac{2(\Delta\Omega)^4}{(2l+1)(2l'+1)} \text{tr}(\mathbf{C} \mathbf{P}_l \mathbf{C} \mathbf{P}_{l'}), \quad (45)$$

where the factor of two arises because the two terms on the right-hand side of the Isserlis identity are in this case identical. To evaluate the scalar quantity $\text{tr}(\mathbf{C} \mathbf{P}_l \mathbf{C} \mathbf{P}_{l'})$ we substitute the representation (36) of the data covariance matrix \mathbf{C} , and transform the result into a fourfold integral over the sphere Ω in the spatial basis. Spherical harmonic orthonormality (8) obligingly eliminates almost everything in sight, leaving the simple result

$$\Sigma_{ll'}^{\text{WS}} = \frac{2}{2l+1} (S_l + N_l)^2 \delta_{ll'}. \quad (46)$$

The Kronecker delta $\delta_{ll'}$ in eq. (46) is an indication that whole-sphere estimates $\hat{S}_l^{\text{WS}}, \hat{S}_{l'}^{\text{WS}}$ of the spectrum $S_l, S_{l'}$ are uncorrelated as well as unbiased.

The formula for the variance of an estimate,

$$\text{var}(\hat{S}_l^{\text{WS}}) = \Sigma_{ll}^{\text{WS}} = \frac{2}{2l+1} (S_l + N_l)^2, \quad (47)$$

can be understood on the basis of elementary statistical considerations (Jones 1963; Knox 1995; Grishchuk & Martin 1997). The estimate \hat{S}_l^{WS} in eq. (38) can be regarded as a linear combination of $2l+1$ samples of the power $|d_{lm}|^2$, $-l \leq m \leq l$, where d_{lm} is drawn from a Gaussian distribution with variance $S_l + N_l$. Every term d_{lm} , except where $m=0$, is complex, and thus responsible for two degrees of freedom, but for real signals, $d_{l-m} = (-1)^m d_{lm}^*$, and thus the total number of degrees of freedom in the expression (38) is $2l+1$. The resulting statistic has a chi-squared distribution with a variance equal to twice the squared variance of the underlying Gaussian distribution divided by the number of samples (e.g. Bendat & Piersol 2000); this accounts for the factors of $2/(2l+1)$ and $(S_l + N_l)^2$ in eq. (47). It may seem surprising that $\text{var}(\hat{S}_l^{\text{WS}}) > 0$ even in the absence of measurement noise, $N_l = 0$; however, there is always a sampling variance when drawing from a random distribution no matter how precisely each sample is measured. This noise-free ‘planetary’ or ‘cosmic variance’ sets a fundamental limit on the uncertainty of a spectral estimate that cannot be reduced by experimental improvements.

In applications where we do not have any *a priori* knowledge about the statistics of the noise \mathbf{n} , we have no choice but to omit the terms N_l and $\text{tr}(\mathbf{N} \mathbf{P}_l)$ in eqs (38)–(39). The estimate \hat{S}_l^{WS} is then biased by the noise, $\langle \hat{S}_l^{\text{WS}} \rangle = S_l + N_l$; nevertheless, the formula (46) for the covariance remains valid. Similar remarks apply to the other estimators that we shall consider in the more general case $R \neq \Omega$. We shall employ the whole-sphere variance $\text{var}(\hat{S}_l^{\text{WS}})$ of eq. (47) as a ‘gold standard’ of comparison for these other estimators.

5 CUT-SPHERE DATA: THE PERIODOGRAM

Suppose now that we only have (or more commonly in geophysics we only wish to consider) data $d(\mathbf{r})$ or $\mathbf{d} = (d_1 \ d_2 \ \dots \ d_J)^T$ over a portion R of the sphere Ω , with surface area $A < 4\pi$.

5.1 Boxcar window function

It is convenient in this case to regard the data $d(\mathbf{r})$ as having been multiplied by a unit-valued boxcar window function,

$$b(\mathbf{r}) = \sum_{pq} b_{pq} Y_{pq}(\mathbf{r}) = \begin{cases} 1 & \text{if } \mathbf{r} \in R, \\ 0 & \text{otherwise,} \end{cases} \quad (48)$$

confined to the region R . The power spectrum of the boxcar window (48) is

$$B_p = \frac{1}{2p+1} \sum_q |b_{pq}|^2. \quad (49)$$

Using a classical Legendre integral formula due to Byerly (1893) it can be shown that eq. (49) reduces, in the case of a single axisymmetric polar cap of angular radius Θ and a double polar cap complementary to an equatorial cut of width $\pi - 2\Theta$, to

$$B_p^{\text{cap}} = \pi(2p+1)^{-2} [P_{p-1}(\cos \Theta) - P_{p+1}(\cos \Theta)]^2, \quad (50)$$

$$B_p^{\text{cut}} = \begin{cases} 4B_p^{\text{cap}} & \text{if } p \text{ is even,} \\ 0 & \text{if } p \text{ is odd,} \end{cases} \quad (51)$$

where $P_{-1}(\mu) = 1$. As a special case of eqs (50)–(51), the power of the $p=0$ or dc component in these two instances is $B_0^{\text{cap}} = A^2/(4\pi) = \pi(1 - \cos \Theta)^2$, $B_0^{\text{cut}} = 4B_0^{\text{cap}} = A^2/(4\pi)$. In fact, the dc power of any boxcar $b(\mathbf{r})$, no matter how irregularly shaped, is $B_0 = A^2/(4\pi)$.

The whole-sphere identity (40) is generalized in the case $R \neq \Omega$ to

$$\text{tr}(\mathbf{P}_l \mathbf{P}_{l'}) = (\Delta\Omega)^{-2} \sum_{mm'} |D_{lm,l'm'}|^2, \quad (52)$$

where the quantities

$$D_{lm,l'm'} = \int_R Y_{lm}^*(\mathbf{r}) Y_{l'm'}(\mathbf{r}) d\Omega \quad (53)$$

are the matrix elements of the spectral-basis projection operator defined in eq. (23). We can express this in terms of the power spectral coefficients B_p by first using the boxcar (48) to rewrite eq. (53) as an integral over the whole sphere Ω , and then making use of the formula for integrating a product of three spherical harmonics, eq. (12):

$$\begin{aligned} \text{tr}(\mathbf{P}_l \mathbf{P}_{l'}) &= (\Delta\Omega)^{-2} \sum_{mm'} \left| \sum_{pq} b_{pq} \int_{\Omega} Y_{lm}^*(\mathbf{r}) Y_{pq}(\mathbf{r}) Y_{l'm'}(\mathbf{r}) d\Omega \right|^2 \\ &= \frac{(2l+1)(2l'+1)}{4\pi(\Delta\Omega)^2} \sum_{pq} \sum_{p'q'} \sqrt{(2p+1)(2p'+1)} b_{pq} b_{p'q'}^* \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p' & l' \\ 0 & 0 & 0 \end{pmatrix} \sum_{mm'} \begin{pmatrix} l & p & l' \\ m & q & m' \end{pmatrix} \begin{pmatrix} l & p' & l' \\ m & q' & m' \end{pmatrix}. \end{aligned} \quad (54)$$

The 3- j orthonormality relation (16) can be used to reduce the final double sum in eq. (54), leading to the simple result

$$\text{tr}(\mathbf{P}_l \mathbf{P}_{l'}) = \frac{(2l+1)(2l'+1)}{4\pi(\Delta\Omega)^2} \sum_p (2p+1) B_p \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2. \quad (55)$$

In the limit $A \rightarrow 4\pi$ of whole-sphere coverage, $B_p \rightarrow 4\pi\delta_{p0}$ and the 3- j symbol with $p = 0$ is given by the first of eqs (19), so that eq. (55) reduces to the result (40) as expected.

Fig. 2 shows the normalized boxcar power spectra B_p/B_0 associated with axisymmetric single and double polar caps of various angular radii. For a given radius Θ , eqs (50)–(51) show that $(B_p/B_0)^{\text{cut}}$ has a shape identical to $(B_p/B_0)^{\text{cap}}$, but with the odd degrees removed; to avoid duplication, we illustrate the spectra for single caps of radii $\Theta = 10^\circ, 20^\circ, 30^\circ$ and double caps of common radii $\Theta = 60^\circ, 70^\circ, 80^\circ$. The scales along the top of each plot show the number of asymptotic wavelengths that just fit within either the single cap or one of the two double caps; one perfectly fitting wavelength corresponds to a spherical harmonic of degree p_Θ given by $[p_\Theta(p_\Theta + 1)]^{1/2} = 180^\circ/\Theta$, two wavelengths

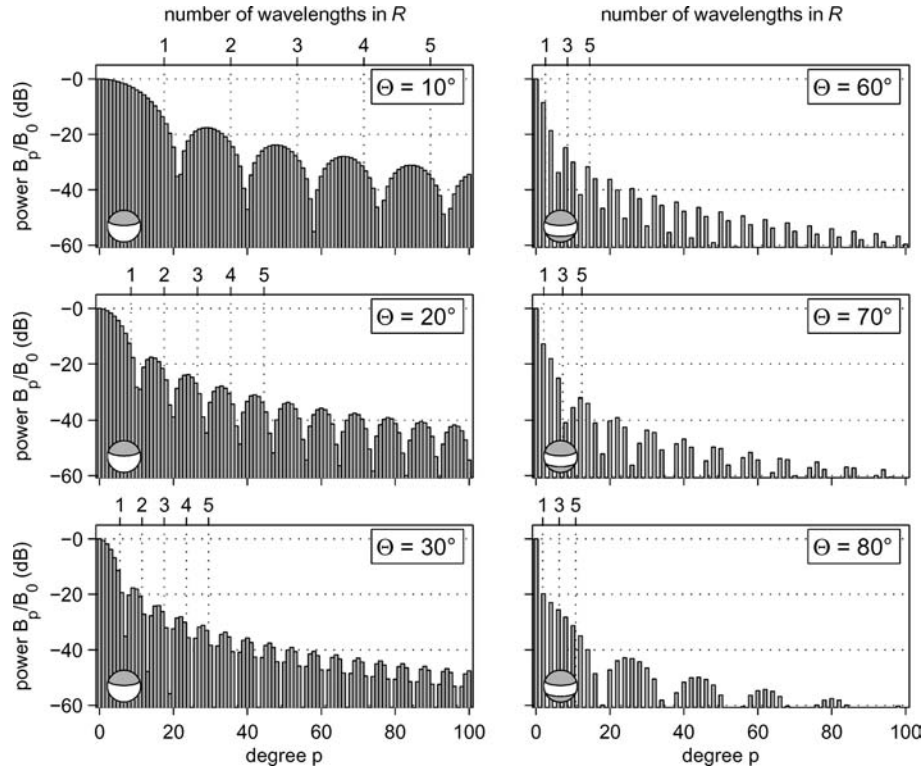


Figure 2. Bar plots of the normalized power B_p/B_0 versus angular degree p for various boxcar windows $b(\mathbf{r})$ as defined by eq. (48). Inset schematic thumbnails show the shapes of the regions considered: axisymmetric polar caps of angular radii $\Theta = 10^\circ, 20^\circ, 30^\circ$ (left-hand panels) and double polar caps of common radii $\Theta = 60^\circ, 70^\circ, 80^\circ$ (right-hand panels). Abscissa in all cases is logarithmic, measured in $\text{dB} = 10 \log_{10}(B_p/B_0)$. Topmost scales show the number of asymptotic wavelengths that just fit within either a single cap (left-hand panels) or one of the two double polar caps (right-hand panels). The odd-degree values of the double-cap power B_p are all identically zero for reasons of symmetry; see eq. (51).

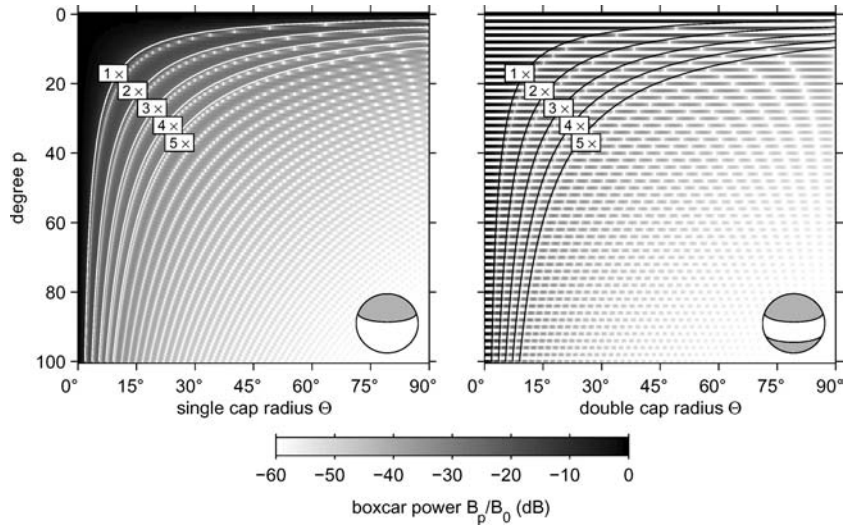


Figure 3. Grey-scale contour plots of the normalized boxcar power B_p/B_0 , measured in dB, versus angular degree $0 \leq p \leq 100$, measured downward on the vertical axis, and single or double polar cap radius $0^\circ \leq \Theta \leq 90^\circ$, on the horizontal axis. Isolines $[p(p+1)]^{1/2} = \{1-5\} \times (180^\circ/\Theta)$ designate the number $\{1-5\}$ of asymptotic wavelengths that just fit within a single polar cap. Thumbnail insets again show the shapes of the regions considered. The double-cap power is ‘striped’ because $B_p^{\text{cut}} = 0$ for odd p .

to a degree $p_{\Theta/2} \approx 2p_\Theta$, and so on. Extending the concept of ‘Rayleigh resolution’ (Thomson & Chave 1991) to the spherical case, a rough rule-of-thumb is that $B_p \ll B_0$ (say 10–20 dB down from the maximum) for all harmonics that are large enough to easily accommodate at least one or two wavelengths within a cap, that is, for all $p \geq \{1-2\} \times p_\Theta$.

Fig. 3 shows a contour plot of the normalized power B_p/B_0 for spherical harmonic degrees $0 \leq p \leq 100$ and single caps (left) and double caps (right) of radii $0^\circ \leq \Theta \leq 90^\circ$. A double cap of common radius $\Theta = 90^\circ$ covers the whole sphere and has power $B_p = 4\pi\delta_{p0}$. The curves labeled $\{1-5\} \times$ are isolines of the functions $[p(p+1)]^{1/2} = \{1-5\} \times (180^\circ/\Theta)$, which correspond to the specified number of asymptotic wavelengths just fitting within a single polar cap. These isolines roughly coincide with the $\{1-5\} \times (-10 \text{ dB})$ contours of the power B_p/B_0 , respectively, confirming the conclusion inferred from Fig. 2 that $B_p \ll B_0$ for all spherical harmonic degrees p that are able to comfortably fit one or two wavelengths within either a single or double cap of arbitrary radius $0^\circ \leq \Theta \leq 90^\circ$. Sums involving B_p such as eq. (55) converge relatively rapidly as a result of this strong decay of the high-degree boxcar power.

5.2 Periodogram estimator

A ‘naive’ (Percival & Walden 1993) estimator of the signal power S_l in the case $R \neq \Omega$ is the spherical analogue of what Schuster (1898) named the periodogram in the context of 1-D time-series analysis:

$$\hat{S}_l^{\text{SP}} = \left(\frac{4\pi}{A} \right) \frac{1}{2l+1} \sum_m \left| \int_R d(\mathbf{r}) Y_{lm}^*(\mathbf{r}) d\Omega \right|^2 - \sum_{l'} K_{ll'} N_{l'}, \quad (56)$$

where we have introduced the matrix

$$K_{ll'} = \left(\frac{4\pi}{A} \right) \frac{1}{2l+1} \sum_{mm'} |D_{lm,l'm'}|^2 = \left(\frac{2l'+1}{A} \right) \sum_p (2p+1) B_p \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2 = \left(\frac{4\pi}{A} \right) \frac{(\Delta\Omega)^2}{2l+1} \text{tr}(\mathbf{P}_l \mathbf{P}_{l'}). \quad (57)$$

The subtracted term in eq. (56) is simply a known constant which—as we will show—corrects the estimate for the bias due to noise. In the pixel basis eqs (56)–(57) become

$$\hat{S}_l^{\text{SP}} = \left(\frac{4\pi}{A} \right) \frac{(\Delta\Omega)^2}{2l+1} [\mathbf{d}^T \mathbf{P}_l \mathbf{d} - \text{tr}(\mathbf{N} \mathbf{P}_l)], \quad (58)$$

the only difference with the whole-sphere estimator (39) being the leading factor of $4\pi/A$ and the fact that the vector and matrix multiplications represent spatial-basis integrations over the region R rather than over the whole sphere Ω . The superscript SP identifies eqs (56) and (58) as the ‘spherical periodogram’ estimator. When $A = 4\pi$, $K_{ll'} = \delta_{ll'}$.

5.3 Leakage bias

To find the expected value of \hat{S}_l^{SP} we proceed just as in reducing eq. (41):

$$\begin{aligned}
 \langle \hat{S}_l^{\text{SP}} \rangle &= \left(\frac{4\pi}{A} \right) \frac{(\Delta\Omega)^2}{2l+1} [\text{tr}(\mathbf{C}\mathbf{P}_l) - \text{tr}(\mathbf{N}\mathbf{P}_l)] \\
 &= \left(\frac{4\pi}{A} \right) \frac{(\Delta\Omega)^2}{2l+1} \text{tr}(\mathbf{S}\mathbf{P}_l) \quad \text{noise bias cancels} \\
 &= \left(\frac{4\pi}{A} \right) \frac{(\Delta\Omega)^2}{2l+1} \sum_{l'} S_{l'} \text{tr}(\mathbf{P}_l \mathbf{P}_{l'}) \\
 &= \sum_{l'} K_{ll'} S_{l'},
 \end{aligned} \tag{59}$$

where we used the definition (57) of $K_{ll'}$ to obtain the final equality. The calculation in eq. (59) confirms the equivalence of eqs (56) and (58), and shows that, unlike the whole-sphere estimator \hat{S}_l^{WS} , the periodogram \hat{S}_l^{SP} is biased, inasmuch as $\langle \hat{S}_l^{\text{SP}} \rangle \neq S_l$. The source of this bias is leakage from the power in neighbouring spherical harmonic degrees $l' = l \pm 1, l \pm 2, \dots$. The matrix $K_{ll'}$ was introduced in an astrophysical context by Peebles (1973) and is known as the (periodogram) ‘leakage’ matrix in helioseismology (Schou & Brown 1994; Appourchaux *et al.* 1998). We shall refer to it as the periodogram ‘coupling’ matrix, which is more in line with cosmological usage (Wandelt *et al.* 2001; Hivon *et al.* 2002). The 3-*j* identity

$$\sum_{l'} (2l' + 1) \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2 = 1, \tag{60}$$

which is a special case of the orthonormality relation (15), guarantees that every row of $K_{ll'}$ sums to unity,

$$\sum_{l'} K_{ll'} = \frac{1}{A} \sum_p (2p + 1) B_p = \frac{1}{A} \int_{\Omega} b^2(\mathbf{r}) d\Omega = 1, \tag{61}$$

so that there is no leakage bias only in the case of a perfectly white spectrum:

$$\langle \hat{S}_l^{\text{SP}} \rangle = S \quad \text{if} \quad S_l = S. \tag{62}$$

This is in fact why we introduced the factor of $4\pi/A$ in eqs (56) and (58): to ensure the desirable result (62). For pixelized measurements with a white noise spectrum, $N_l = N = \Delta\Omega \sigma^2$, the subtracted noise-bias correction term in eq. (56) reduces to $N = \Delta\Omega \sigma^2$, as in eq. (38). In the whole-sphere limit, $B_p \rightarrow 4\pi \delta_{p0}$ so that $K_{ll'} \rightarrow \delta_{ll'}$ and $\langle \hat{S}_l^{\text{SP}} \rangle \rightarrow S_l$, as expected.

In the opposite limit of a connected, infinitesimally small region,

$$A \rightarrow 0 \quad \text{and} \quad \sum_l (2l + 1) \rightarrow \infty \quad \text{with} \quad \left(\frac{A}{4\pi} \right) \sum_l (2l + 1) = 1 \quad \text{held fixed}, \tag{63}$$

the inverse-area-scaled boxcar $A^{-1} b(\mathbf{r})$ tends to a Dirac delta function $\delta(\mathbf{r}, \mathbf{R})$, where \mathbf{R} is the pointwise location of the region R , so that the boxcar power is white: $B_p \rightarrow A^2/(4\pi)$. The spectral-basis projector (53) tends in the same limit to $D_{lm,l'm'} \rightarrow A Y_{lm}^*(\mathbf{R}) Y_{l'm'}(\mathbf{R})$, so that the coupling matrix (57) reduces to

$$K_{ll'} \rightarrow \frac{A}{4\pi} (2l' + 1) \quad \text{for all} \quad 0 \leq l \leq \infty. \tag{64}$$

Eq. (64) highlights the fact that there is strong coupling among all spherical harmonic degrees l, l' in the limit (63); in fact, the expected value of the periodogram estimate is then simply the total signal power contained within the infinitesimal measurement region: $\langle \hat{S}_l^{\text{SP}} \rangle \rightarrow S_{\text{tot}}^R$. The fixity constraint upon the limit (63) guarantees that the rows of the coupling matrix (64) sum to unity, in accordance with eq. (61).

In Fig. 4, we illustrate the periodogram coupling matrix $K_{ll'}$ for the same single polar caps of radii $\Theta = 10^\circ, 20^\circ, 30^\circ$ and double polar caps of common radii $\Theta = 60^\circ, 70^\circ, 80^\circ$ as in Figs 2 and 3. In particular, for various values of the target angular degree $l = 0, 20, 40, 60$, we exhibit the variation of $K_{ll'}$ as a function of the column index l' ; this format highlights the spectral leakage that is the source of the bias described by eq. (59). The quantity we actually plot is $100 \times K_{ll'}$, so that the height of each bar reflects the per cent leakage of the power at degree l' into the periodogram estimate \hat{S}_l^{SP} , in accordance with the constraint that all of the bars must sum to 100 per cent, by virtue of eq. (61). At small target degrees $l \approx 0$ the variation of $K_{ll'}$ with l' is influenced by the triangle condition that applies to the 3-*j* symbols in eq. (57), but in the limit $l \rightarrow \infty$ the coupling matrix takes on a universal shape that is approximately described by

$$K_{ll'} \approx \left(\frac{4\pi}{A} \right) \sum_p B_p [X_{p|l-l'|}(\pi/2)]^2, \tag{65}$$

as a consequence of the 3-*j* asymptotic relation (20); this satisfies the constraint eq. (61). This tendency for $K_{ll'}$ to maintain its shape and just translate to the next large target degree is apparent in all of the plots.

It is evident from both eq. (57) and the plots of $K_{ll'}$ in Fig. 4 that a small measurement region, with $A \ll 4\pi$, gives rise to much more extensive coupling and broad-band spectral leakage than a large region, with $A \approx 4\pi$. We quantify this relation between the extent of the coupling and the size of the region R in Fig. 5, in which we plot the large- l limits of the matrix $K_{ll'}$ in eq. (57) as a function of the offset from the target degree for the same single-cap and double-cap regions as in Fig. 4. The common abscissa in all plots is measured in asymptotic

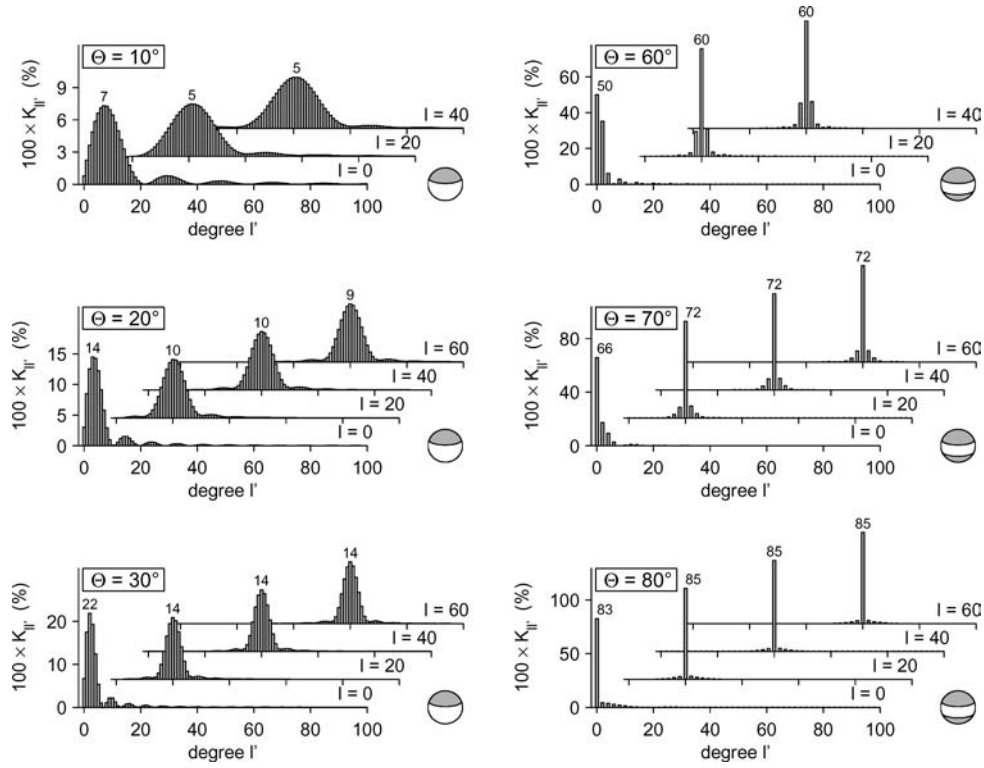


Figure 4. Bar plots of the periodogram coupling matrix $100 \times K_{ll'}$ for single polar caps of radii $\Theta = 10^\circ, 20^\circ, 30^\circ$ (left-hand panels) and double caps of common radii $\Theta = 60^\circ, 70^\circ, 80^\circ$ (right-hand panels). The tick marks are at $l' = 0, 20, 40, 60, 80, 100$ on every offset abscissa; the target degrees $l = 0, 20, 40, 60$ are indicated on the right. Numbers on top are the maximum diagonal value $100 \times K_{ll}$ for every target degree l . The double-cap matrix is alternating, $K_{ll'} = 0$ if $|l - l'|$ odd, since the 3- j symbols are zero whenever $l + p + l'$ is odd and $B_p^{\text{cut}} = 0$ if p odd.

wavelengths, $-3 \leq \nu \leq 3$, defined by $|l' - l| = p_\Theta/|\nu|$, or indeed $l' - l \approx \nu p_\Theta$ where $[p_\Theta(p_\Theta + 1)]^{1/2} = 180^\circ/\Theta$, and delineated along the top; the $l' - l$ scales along the bottom vary depending upon the cap size Θ . It is clear from this format that $K_{ll'}$ is always substantially less than its peak diagonal value K_{ll} , so that the coupling and spectral leakage are weak, whenever $|l' - l| \geq \{1-2\} \times p_\Theta$. The extent of the periodogram coupling thus scales directly with the radius Θ of a single or double polar cap. The resulting broad-band character of the spectral leakage for small regions ($A \ll 4\pi$) is highly undesirable and makes the periodogram ‘hopelessly obsolete’ (Thomson & Chave 1991).

5.4 Periodogram covariance

Making use of the Isserlis identity (44) we find that the covariance of two periodogram estimates \hat{S}_l^{SP} and $\hat{S}_{l'}^{\text{SP}}$ at different degrees l and l' is given by a pixel-basis formula very similar to eq. (45),

$$\Sigma_{ll'}^{\text{SP}} = \text{cov}(\hat{S}_l^{\text{SP}}, \hat{S}_{l'}^{\text{SP}}) = \frac{2(4\pi/A)^2(\Delta\Omega)^4}{(2l+1)(2l'+1)} \text{tr}(\mathbf{C}\mathbf{P}_l\mathbf{C}\mathbf{P}_{l'}), \quad (66)$$

with the important difference that $\text{tr}(\mathbf{C}\mathbf{P}_l\mathbf{C}\mathbf{P}_{l'})$ now represents a fourfold integral over the region R rather than over the whole sphere Ω . Inserting the representation (36) of the data covariance matrix \mathbf{C} and transforming to the spatial basis, we obtain the result

$$\Sigma_{ll'}^{\text{SP}} = \frac{2(4\pi/A)^2}{(2l+1)(2l'+1)} \sum_{mm'} \left| \sum_{pq} (S_p + N_p) D_{lm,pq} D_{pq,l'm'} \right|^2, \quad (67)$$

which reduces to eq. (46) in the limit of whole-sphere data coverage, when $D_{lm,l'm'} = \delta_{ll'}\delta_{mm'}$. Using the boxcar function $b(\mathbf{r})$ to rewrite $D_{lm,l'm'}$ as an integral over the whole sphere Ω as in our reduction of eq. (52) we can express the covariance of a periodogram spectral estimate in terms of Wigner 3- j symbols:

$$\Sigma_{ll'}^{\text{SP}} = \frac{2}{A^2} \sum_{mm'} \left| \sum_{pq} (2p+1)(S_p + N_p) \sum_{st} \sum_{s't'} \sqrt{(2s+1)(2s'+1)} b_{st} b_{s't'}^* \begin{pmatrix} l & p & s \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l' & p & s' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & s \\ m & q & t \end{pmatrix} \begin{pmatrix} l' & p & s' \\ m' & q & t' \end{pmatrix} \right|^2. \quad (68)$$

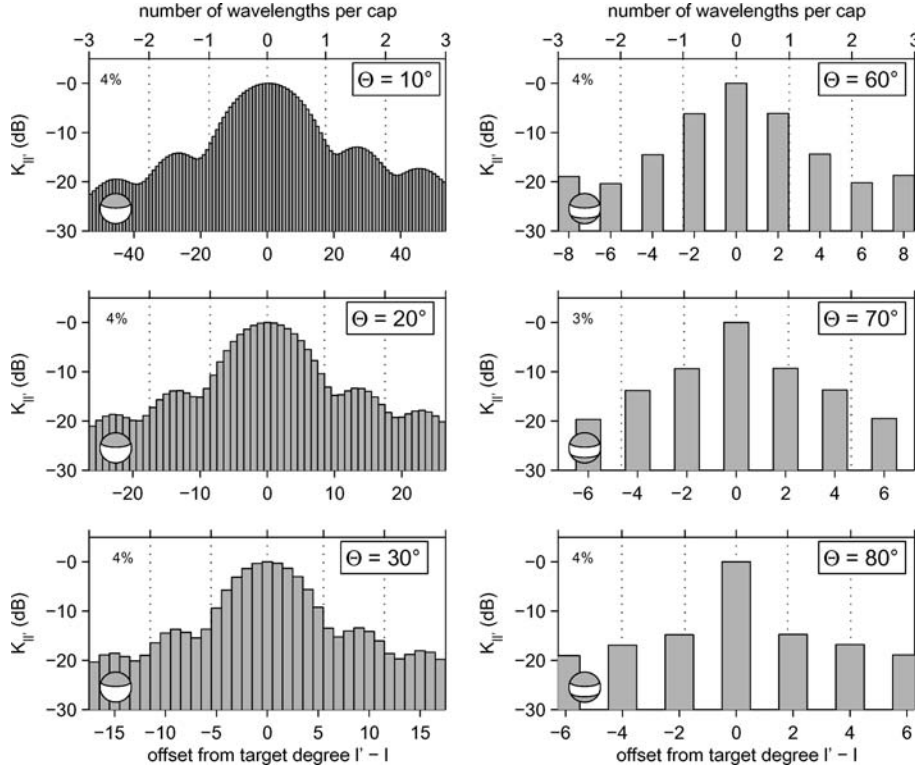


Figure 5. Large- l limits of the periodogram coupling matrix $K_{ll'}$ for single polar caps of radii $\Theta = 10^\circ, 20^\circ, 30^\circ$ (left-hand panels) and double caps of radii $\Theta = 60^\circ, 70^\circ, 80^\circ$ (right-hand panels). The common abscissa is the offset from the target angular degree, measured in asymptotic wavelengths, $l' - l \approx \nu p_\Theta$. The limiting shapes were found empirically by increasing l until the plots no longer changed visibly. The exact coupling matrix (57) is asymmetric because of the leading factor of $2l' + 1$; the slight left–right asymmetry visible here is not retained in the asymptotic result (65). Small numbers in upper left corner give the per cent coupling outside the boundaries $-3 \leq \nu \leq 3$ of each plot.

Eqs (67) and (68) are exact and show that every element of the periodogram covariance is non-negative: $\Sigma_{ll'}^{\text{SP}} \geq 0$, with equality prevailing only for $l \neq l'$ in the limit of whole-sphere coverage, $A = 4\pi$. We shall obtain a more palatable approximate expression for $\Sigma_{ll'}^{\text{SP}}$, valid for a moderately coloured spectrum, in Section 8.1.

5.5 Deconvolved periodogram

In principle it is possible to eliminate the leakage bias in the periodogram estimate \hat{S}_l^{SP} by numerical inversion of the coupling matrix $K_{ll'}$. The expected value of the ‘deconvolved periodogram’ estimator, defined by

$$\hat{S}_l^{\text{DP}} = \sum_{l'} K_{ll'}^{-1} \hat{S}_{l'}^{\text{SP}}, \quad (69)$$

is clearly $\langle \hat{S}_l^{\text{DP}} \rangle = S_l$. The corresponding covariance is given by the usual formula for the covariance of a linear combination of estimates (Menke 1989):

$$\Sigma_{ll'}^{\text{DP}} = \text{cov}(\hat{S}_l^{\text{DP}}, \hat{S}_{l'}^{\text{DP}}) = \sum_{pp'} K_{lp}^{-1} \Sigma_{pp'}^{\text{SP}} K_{p'l'}^{-T} \quad (70)$$

where $K_{p'l'}^{-T} = K_{l'p'}^{-1}$. In practice the deconvolution (69) is only feasible when the region R covers most of the sphere, $A \approx 4\pi$; for any region whose area A is significantly smaller than 4π , the periodogram coupling matrix (57) will be too ill-conditioned to be invertible.

6 MAXIMUM-LIKELIHOOD ESTIMATION

In this section we review the maximum-likelihood method of spectral estimation, which has been developed and applied by a large number of cosmological investigators to CMB temperature data from ground-based surveys as well as two space missions: the ‘Cosmic Background Explorer’ (COBE) satellite and the ‘Wilkinson Microwave Anisotropy Project’ (WMAP). Our discussion draws heavily upon the analyses by Tegmark (1997), Tegmark *et al.* (1997), Bond *et al.* (1998), Oh *et al.* (1999) and Hinshaw *et al.* (2003). For more rigorous theoretical considerations, we refer to Cox & Hinkley (1974). In particular, we caution that if the number of parameters to be estimated is large compared to the number of data, maximum-likelihood estimators often behave poorly.

6.1 Likelihood function

The starting point of the analysis is the likelihood $\mathcal{L}(S_l, \mathbf{d})$ that one will observe the pixel-basis data $\mathbf{d} = (d_1 \ d_2 \ \dots \ d_J)^T$ given the spectrum S_l . We model this likelihood as Gaussian:

$$\mathcal{L}(S_l, \mathbf{d}) = \frac{\exp\left(-\frac{1}{2} \mathbf{d}^T \mathbf{C}^{-1} \mathbf{d}\right)}{(2\pi)^{J/2} \sqrt{\det \mathbf{C}}}, \quad (71)$$

where \mathbf{C}^{-1} is the inverse of the data covariance matrix defined in eq. (36), $\mathbf{C}^{-1} \mathbf{C} = \mathbf{C} \mathbf{C}^{-1} = \mathbf{I}$, and J is the total number of observational pixels as before. The notation is intended to imply that $\mathcal{L}(S_l, \mathbf{d})$ depends upon *all* of the spectral values S_l , $0 \leq l \leq \infty$; the ‘maximum-likelihood’ estimator is the spectrum S_l that maximizes the multivariate Gaussian likelihood function (71) for measured data \mathbf{d} .

Maximization of $\mathcal{L}(S_l, \mathbf{d})$ is equivalent to minimization of the logarithmic likelihood

$$L(S_l, \mathbf{d}) = -2 \ln \mathcal{L}(S_l, \mathbf{d}) = \ln(\det \mathbf{C}) + \mathbf{d}^T \mathbf{C}^{-1} \mathbf{d} + J \ln(2\pi). \quad (72)$$

To minimize $L(S_l, \mathbf{d})$ we differentiate with respect to the unknowns S_l using the identity $\ln(\det \mathbf{C}) = \text{tr}(\ln \mathbf{C})$ and

$$\frac{\partial \mathbf{C}}{\partial S_l} = \mathbf{P}_l, \quad \frac{\partial \mathbf{C}^{-1}}{\partial S_l} = -\mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1}, \quad \frac{\partial(\ln \mathbf{C})}{\partial S_l} = \mathbf{C}^{-1} \mathbf{P}_l. \quad (73)$$

The first equality in eq. (73) follows from eq. (36), the others are the result of matrix identities. The resulting minimization condition is

$$\frac{\partial L}{\partial S_l} = -\mathbf{d}^T \mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1} \mathbf{d} + \text{tr}(\mathbf{C}^{-1} \mathbf{P}_l) = 0. \quad (74)$$

The ensemble average of eq. (74) is

$$\left\langle \frac{\partial L}{\partial S_l} \right\rangle = -\text{tr}(\mathbf{C}^{-1} \mathbf{P}_l) + \text{tr}(\mathbf{C}^{-1} \mathbf{P}_l) = 0, \quad (75)$$

verifying that the maximum-likelihood estimate is correct on average in the sense that the average slope $\langle \partial L / \partial S_l \rangle$ is zero at the point corresponding to the true spectrum S_l . The curvature of the logarithmic likelihood function $L(S_l, \mathbf{d})$ is

$$\frac{\partial^2 L}{\partial S_l \partial S_{l'}} = \mathbf{d}^T \mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1} \mathbf{d} + \mathbf{d}^T \mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1} \mathbf{d} - \text{tr}(\mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1} \mathbf{P}_{l'}). \quad (76)$$

In the vicinity of the minimum we can expand $L(S_l, \mathbf{d})$ in a Taylor series:

$$L(S_l + \delta S_l, \mathbf{d}) = L(S_l, \mathbf{d}) + \sum_l \left(\frac{\partial L}{\partial S_l} \right) \delta S_l + \frac{1}{2} \sum_{ll'} \delta S_l \left(\frac{\partial^2 L}{\partial S_l \partial S_{l'}} \right) \delta S_{l'} + \dots \quad (77)$$

The quantities $\partial^2 L / \partial S_l \partial S_{l'}$ are the elements of the Hessian of the logarithmic likelihood function; likewise, we shall write $(\partial^2 L / \partial S_l \partial S_{l'})^{-1}$ to denote the elements of its inverse. Ignoring the higher-order terms ... in eq. (77) we can write the minimization condition (74) in the form

$$\delta S_l = \sum_{l'} \left(\frac{\partial^2 L}{\partial S_l \partial S_{l'}} \right)^{-1} \left(-\frac{\partial L}{\partial S_{l'}} \right) = \sum_{l'} \left(\frac{\partial^2 L}{\partial S_l \partial S_{l'}} \right)^{-1} [\mathbf{d}^T \mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1} \mathbf{d} - \text{tr}(\mathbf{C}^{-1} \mathbf{P}_{l'})]. \quad (78)$$

Eq. (78) is the classical Newton–Raphson iterative algorithm for the minimization of $L(S_l, \mathbf{d})$. Starting with an initial guess for the spectrum S_l the method uses eq. (78) to find δS_l , updates the spectrum $S_l \rightarrow S_l + \delta S_l$, re-evaluates the right-hand side, and so on until convergence, $\delta S_l \rightarrow 0$, is attained (see, e.g. Strang 1986; Press *et al.* 1992).

6.2 Quadratic estimator

For large data vectors \mathbf{d} computation of the logarithmic likelihood curvature (76) is generally prohibitive and it is customary to replace $\frac{1}{2}(\partial^2 L / \partial S_l \partial S_{l'})$ by its ensemble average, which is known as the ‘Fisher matrix’:

$$F_{ll'} = \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial S_l \partial S_{l'}} \right\rangle = \frac{1}{2} \text{tr}(\mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1} \mathbf{P}_{l'}). \quad (79)$$

Note that like the curvature (76) itself the Fisher matrix (79) is symmetric, $F_{ll'} = F_{l'l}$, and positive definite. Upon substituting $\frac{1}{2} F_{ll'}^{-1}$ for the inverse Hessian $(\partial^2 L / \partial S_l \partial S_{l'})^{-1}$ in eq. (78), we obtain a Newton–Raphson algorithm that is computationally more tractable, and guaranteed to converge (albeit by a different iteration path) to the same local minimum:

$$\delta S_l = \frac{1}{2} \sum_{l'} F_{ll'}^{-1} [\mathbf{d}^T \mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1} \mathbf{d} - \text{tr}(\mathbf{C}^{-1} \mathbf{P}_{l'})]. \quad (80)$$

The second term in brackets in eq. (80) can be manipulated as follows:

$$\text{tr}(\mathbf{C}^{-1} \mathbf{P}_{l'}) = \text{tr}(\mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1} \mathbf{C}) = \sum_n \text{tr}(\mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1} \mathbf{P}_n) (S_n + N_n) = 2 \sum_n F_{l'n} (S_n + N_n). \quad (81)$$

This enables us to rewrite the iteration (80) in the form

$$S_l + \delta S_l = \frac{1}{2} \sum_{l'} F_{ll'}^{-1} [\mathbf{d}^T \mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1} \mathbf{d} - \text{tr}(\mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1} \mathbf{N})]. \quad (82)$$

In particular, at the minimum, where $\delta S_l = 0$, the minimum conditions (74) are satisfied and eq. (82) reduces to

$$\hat{S}_l^{\text{ML}} = \mathbf{d}^T \mathbf{Z}_l \mathbf{d} - \text{tr}(\mathbf{N} \mathbf{Z}_l), \quad (83)$$

where we have defined a new symmetric matrix,

$$\mathbf{Z}_l = \frac{1}{2} \sum_{l'} F_{ll'}^{-1} (\mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1}). \quad (84)$$

The superscript ML designates \hat{S}_l^{ML} as the maximum-likelihood estimator. Eq. (83) is quadratic in the data \mathbf{d} and has the same form as the whole-sphere and periodogram estimators \hat{S}_l^{WS} and \hat{S}_l^{SP} , but with an important difference: the right-hand sides of eqs (39) and (58) are independent of the spectrum S_l whereas the matrix \mathbf{Z}_l in eq. (84) depends upon S_l . In fact, eq. (83) can be regarded as a fixed-point equation of the form $\hat{S}_l^{\text{ML}} = f(\mathbf{d}, \hat{S}_l^{\text{ML}})$, where the right-hand side exhibits a quadratic dependence upon \mathbf{d} but a more general dependence upon the unknown spectral estimates \hat{S}_l^{ML} , $0 \leq l \leq \infty$. Maximum-likelihood estimation is inherently non-linear, requiring iteration to converge to the local minimum \hat{S}_l^{ML} .

6.3 Mean and covariance

The maximum-likelihood method yields an unbiased estimate of the spectrum inasmuch as

$$\begin{aligned} \langle \hat{S}_l^{\text{ML}} \rangle &= \text{tr}(\mathbf{C} \mathbf{Z}_l) - \text{tr}(\mathbf{N} \mathbf{Z}_l) \\ &= \text{tr}(\mathbf{S} \mathbf{Z}_l) \quad \text{noise bias cancels} \\ &= \frac{1}{2} \sum_{l'} F_{ll'}^{-1} \sum_p S_p \text{tr}(\mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1} \mathbf{P}_p) \\ &= \sum_{l'} F_{ll'}^{-1} \sum_p F_{l'p} S_p \\ &= S_l. \end{aligned} \quad (85)$$

Using the Isserlis identity (44) to compute the covariance of two estimates \hat{S}_l^{ML} and $\hat{S}_{l'}^{\text{ML}}$, we find that

$$\begin{aligned} \Sigma_{ll'}^{\text{ML}} &= \text{cov}(\hat{S}_l^{\text{ML}}, \hat{S}_{l'}^{\text{ML}}) \\ &= 2 \text{tr}(\mathbf{C} \mathbf{Z}_l \mathbf{C} \mathbf{Z}_{l'}) \\ &= \frac{1}{2} \text{tr} \left(\mathbf{C} \sum_p F_{lp}^{-1} \mathbf{C}^{-1} \mathbf{P}_p \mathbf{C}^{-1} \mathbf{C} \sum_{p'} F_{l'p'}^{-1} \mathbf{C}^{-1} \mathbf{P}_{p'} \mathbf{C}^{-1} \right) \\ &= \frac{1}{2} \sum_p F_{lp}^{-1} \sum_{p'} F_{l'p'}^{-1} \text{tr}(\mathbf{C}^{-1} \mathbf{P}_p \mathbf{C}^{-1} \mathbf{P}_{p'}) \\ &= \sum_p F_{lp}^{-1} \sum_{p'} F_{l'p'}^{-1} F_{p'p} \\ &= F_{ll'}^{-1}. \end{aligned} \quad (86)$$

The calculation in eq. (86) shows that the maximum likelihood covariance $\Sigma_{ll'}^{\text{ML}}$ is the inverse $F_{ll'}^{-1}$ of the ubiquitous Fisher matrix (79). The method depends upon our ability to invert $F_{ll'}$ and, as we shall elaborate in Section 6.6, this is only numerically feasible in the case of nearly-whole-sphere coverage, $A \approx 4\pi$. In other words, in many practical (geophysical) applications, a maximum-likelihood estimate may not exist.

6.4 The Fisher matrix

Pixel-basis computation of the Fisher matrix $F_{ll'} = \frac{1}{2} \text{tr}(\mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1} \mathbf{P}_{l'})$ requires numerical inversion of the $J \times J$ covariance matrix \mathbf{C} . Transforming to the spatial basis, we can instead write the definition (79) in terms of the inverse data covariance function $C^{-1}(\mathbf{r}, \mathbf{r}')$ equivalent to the pixel-basis inverse $(\Delta\Omega)^{-2} \mathbf{C}^{-1}$ in the form

$$F_{ll'} = \frac{1}{2} \sum_{mm'} |V_{lm, l'm'}|^2, \quad (87)$$

where

$$V_{lm, l'm'} = \iint_R Y_{lm}^*(\mathbf{r}) C^{-1}(\mathbf{r}, \mathbf{r}') Y_{l'm'}(\mathbf{r}') d\Omega d\Omega'. \quad (88)$$

Among other things, eq. (87) shows that every element of the Fisher matrix is non-negative: $F_{ll'} \geq 0$. To compute the matrix elements (88) in the absence of an explicit expression for $C^{-1}(\mathbf{r}, \mathbf{r}')$ in the case $R \neq \Omega$ we can find the auxiliary spacelimited function

$$V_{l'm'}(\mathbf{r}) = \int_R C^{-1}(\mathbf{r}, \mathbf{r}') Y_{l'm'}(\mathbf{r}') d\Omega' = \sum_{lm} V_{lm,l'm'} Y_{lm}(\mathbf{r}) \quad (89)$$

by solving the spatial-basis integral equation

$$\int_R C(\mathbf{r}, \mathbf{r}') V_{l'm'}(\mathbf{r}') d\Omega' = Y_{l'm'}(\mathbf{r}), \quad \mathbf{r} \in R, \quad (90)$$

where

$$C(\mathbf{r}, \mathbf{r}') = \sum_{pq} (S_p + N_p) Y_{pq}(\mathbf{r}) Y_{pq}^*(\mathbf{r}') = \frac{1}{4\pi} \sum_p (2p+1)(S_p + N_p) P_p(\mathbf{r} \cdot \mathbf{r}'). \quad (91)$$

Alternatively, we can transform eq. (90) to the spectral basis and solve

$$\sum_{st} \sum_{pq} D_{lm,pq} (S_p + N_p) D_{pq,st} V_{st,l'm'} = D_{lm,l'm'}. \quad (92)$$

In the case of an axisymmetric region such as a polar cap or equatorial cut, the spatial-basis and spectral-basis inverse problems (90) and (92) can be decomposed into a series of simpler problems, one for each fixed, non-negative order m ; this axisymmetric reduction is straightforward and will not be detailed here.

In the limiting case of whole-sphere coverage, $R = \Omega$, the pixel-basis covariance matrix (36) can be inverted analytically, $\mathbf{C}^{-1} = (\Delta\Omega)^2 \sum_l (S_l + N_l)^{-1} \mathbf{P}_l$, and the Fisher matrix (79) reduces to

$$F_{ll'} = \frac{1}{2} (2l+1)(S_l + N_l)^{-2} \delta_{ll'}, \quad (93)$$

where we have used the whole-sphere identity (40). The result (93) can also be obtained from eqs (87) and (92) by recalling that $D_{lm,l'm'} = \delta_{ll'} \delta_{mm'}$ if $R = \Omega$. In fact, the maximum-likelihood estimate (83) coincides in this limiting case with the whole-sphere estimate (38), $\hat{S}_l^{\text{ML}} = \hat{S}_l^{\text{WS}}$, and the covariance (86) reduces to $\Sigma_{ll'}^{\text{ML}} = F_{ll'}^{-1} = 2(2l+1)^{-1} (S_l + N_l)^2 \delta_{ll'}$, in agreement with eq. (46), as expected. We give an explicit approximate formula that generalizes eq. (93) to the case of a region $R \neq \Omega$ in Section 8.2.

6.5 Cramér–Rao lite

Maximum-likelihood estimation is the method of choice in a wide variety of statistical applications, including CMB cosmology. In large part this popularity is due to a powerful theorem due to Fisher, Cramér and Rao, which guarantees that the maximum-likelihood method yields the ‘best unbiased estimator’ in the sense that it has lower variance than any other unbiased estimate; that is, in this spherical spectral estimation problem,

$$\text{var}(\hat{S}_l^{\text{ML}}) = F_{ll}^{-1} \leq \text{var}(\hat{S}_l) \quad \text{for any } \hat{S}_l \text{ satisfying } \langle \hat{S}_l \rangle = S_l. \quad (94)$$

A general statement and proof of this so-called ‘Cramér–Rao inequality’ is daunting (see, e.g. Kendall & Stuart 1969); however, it is straightforward to prove the limited result (94) if we confine ourselves to the class of quadratic estimators, of the form

$$\hat{S}_l = \mathbf{d}^T \mathbf{Z}_l \mathbf{d} - \text{tr}(\mathbf{N} \mathbf{Z}_l), \quad (95)$$

where the second term corrects for the bias due to noise as usual, and where the symmetric matrix \mathbf{Z}_l remains to be determined. The ensemble average of eq. (95) is

$$\langle \hat{S}_l \rangle = \sum_{l'} Z_{ll'} S_{l'} \quad \text{where} \quad Z_{ll'} = \text{tr}(\mathbf{Z}_l \mathbf{P}_{l'}), \quad (96)$$

so that the condition that there be no leakage bias, that is, $\langle \hat{S}_l \rangle = S_l$, is that $Z_{ll'} = \delta_{ll'}$; and the covariance between two estimates of the form (95), by another application of the Isserlis identity (44), is

$$\Sigma_{ll'} = \text{cov}(\hat{S}_l, \hat{S}_{l'}) = 2 \text{tr}(\mathbf{C} \mathbf{Z}_l \mathbf{C} \mathbf{Z}_{l'}). \quad (97)$$

To find the minimum-variance, unbiased quadratic estimator, we therefore, seek to minimize $\text{var}(\hat{S}_l) = 2 \text{tr}(\mathbf{C} \mathbf{Z}_l \mathbf{C} \mathbf{Z}_l)$ subject to the constraints that $Z_{ll'} = \text{tr}(\mathbf{Z}_l \mathbf{P}_{l'}) = \delta_{ll'}$. Introducing Lagrange multipliers $\eta_{l'}$ we are led to the variational problem

$$\Phi_l = \text{tr}(\mathbf{C} \mathbf{Z}_l \mathbf{C} \mathbf{Z}_l) - \sum_{l'} \eta_{l'} [\text{tr}(\mathbf{Z}_l \mathbf{P}_{l'}) - \delta_{ll'}] = \text{minimum}. \quad (98)$$

Demanding that $\delta\Phi_l = 0$ for arbitrary variations $\delta\mathbf{Z}_l$ of the unknowns \mathbf{Z}_l gives the relation

$$2(\mathbf{C} \mathbf{Z}_l \mathbf{C}) = \sum_{l'} \eta_{l'} \mathbf{P}_{l'} \quad \text{or} \quad \mathbf{Z}_l = \frac{1}{2} \sum_{l'} \eta_{l'} (\mathbf{C}^{-1} \mathbf{P}_{l'} \mathbf{C}^{-1}). \quad (99)$$

To find the multipliers $\eta_{l'}$ that render $\text{tr}(\mathbf{Z}_l \mathbf{P}_{l'}) = \delta_{ll'}$ we multiply eq. (99) by $\mathbf{P}_{l''}$ and take the trace:

$$\sum_{l'} \eta_{l'} F_{l'l''} = \text{tr}(\mathbf{Z}_l \mathbf{P}_{l''}) = \delta_{ll''} \quad \text{or} \quad \eta_{l'} = F_{ll'}^{-1}. \quad (100)$$

Upon substituting eq. (100) into eq. (99) we obtain the final result

$$\mathbf{Z}_l = \frac{1}{2} \sum_{l'} F_{ll'}^{-1} (\mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1}), \quad (101)$$

which is identical to eq. (84). This argument, due to Tegmark (1997), shows that the maximum-likelihood estimator (83) is the best unbiased quadratic estimator of the spectrum, in the sense (94).

6.6 To bin or not to bin

The maximum-likelihood method as described above is applicable only to measurements \mathbf{d} that cover most of the sphere, for example, to spacecraft surveys of the whole-sky CMB temperature field with a relatively narrow galactic cut. For smaller regions the method fails because the degree-by-degree Fisher matrix $F_{ll'}$ is too ill-conditioned to be numerically invertible. Fundamentally, this is due to the strong correlation among adjacent spectral estimates $\hat{S}_l^{\text{ML}}, \hat{S}_{l'}^{\text{ML}}$ within a band of width $|l' - l| \approx \{1-2\} \times p_\Theta$, where as before p_Θ is the degree of the spherical harmonic that just fits a single asymptotic wavelength into the region of dimension $\Theta \approx (2A/\pi)^{1/2}$. In view of this strong correlation it is both appropriate and necessary to sacrifice spectral resolution, and seek instead the best unbiased estimates \hat{S}_B^{ML} of a sequence of binned linear combinations of the individual spectral values S_l , of the form

$$S_B = \sum_l W_{Bl} S_l. \quad (102)$$

We shall assume that the bins B are sufficiently non-overlapping for the non-square weight matrix W_{Bl} to be of full row rank, and we shall stipulate that every row sums to unity, that is, $\sum_l W_{Bl} = 1$, to ensure that $\langle \hat{S}_B^{\text{ML}} \rangle = S$ in the case of a white spectrum, $S_l = S$. Apart from these constraints, the weights can be anything we wish; for example, a boxcar or uniformly weighted average $W_{Bl} = \delta_{l \in B} / \sum_{l' \in B}$, where $\delta_{l \in B}$ is one if degree l is in bin B and zero otherwise, and the denominator is the width of the bin.

Because we must resort to estimating band averages S_B we are obliged to adopt a different statistical viewpoint in the maximum-likelihood estimation procedure; specifically, we shall suppose that S_l can be adequately approximated by a coarser-grained spectrum,

$$S_l^\dagger = \sum_B W_{lB}^\dagger S_B, \quad (103)$$

where W_{lB}^\dagger is the Moore–Penrose generalized inverse or pseudo-inverse of the weight matrix W_{Bl} (Strang 1998). Because W_{Bl} is of full row rank, W_{lB}^\dagger is the purely underdetermined pseudo-inverse, given by

$$W_{lB}^\dagger = \sum_{B'} W_{lB'}^T \left(\sum_{l'} W_{B'l'} W_{l'B}^T \right)^{-1}, \quad (104)$$

where $W_{lB}^T = W_{Bl}$ and the second term is the inverse of the enclosed symmetric matrix (Menke 1989; Gubbins 2004). The coarse-grained spectrum (103) is the minimum-norm solution of eq. (102) with no component in the null-space of W_{Bl} ; in other words, S_l^\dagger is the part of S_l that can be faithfully recovered from the binned values S_B . Since W_{lB}^\dagger in eq. (104) is a right inverse of W_{Bl} , that is, $\sum_l W_{Bl} W_{lB'}^\dagger = \delta_{BB'}$, the spectra S_l^\dagger and S_l have identical binned averages, $S_B^\dagger = \sum_l W_{Bl} S_l^\dagger = S_B$. For the simplest case of contiguous, boxcar-weighted bins, $W_{lB}^\dagger = (\delta_{l \in B})^T$ so that S_l^\dagger is a staircase spectrum, constant and equal to S_B in every bin B . Stoica & Sundin (1999) imposed this before comparing maximum-likelihood with other estimators.

The coarse-grained spectrum S_l^\dagger gives rise to an associated, coarse-grained representation \mathbf{C}^\dagger of the data covariance matrix \mathbf{C} in eq. (36), namely

$$\mathbf{C}^\dagger = \mathbf{S}^\dagger + \mathbf{N}^\dagger = \sum_l (S_l^\dagger + N_l^\dagger) \mathbf{P}_l = \sum_B (S_B + N_B) \mathbf{P}_B, \quad (105)$$

where N_B and N_l^\dagger are defined in terms of N_l by the analogues of eqs (102)–(103), and where the vector $\mathbf{P}_B = \partial \mathbf{C}^\dagger / \partial S_B$ is

$$\mathbf{P}_B = \sum_l \left(\frac{\partial \mathbf{C}^\dagger}{\partial S_l^\dagger} \right) \left(\frac{\partial S_l^\dagger}{\partial S_B} \right) = \sum_l \mathbf{P}_l W_{lB}^\dagger. \quad (106)$$

To estimate the binned spectrum (102) we consider a new likelihood function $\mathcal{L}(S_B, \mathbf{d})$ of the form (71) but with \mathbf{C}^{-1} replaced by the coarse-grained inverse matrix $\mathbf{C}^{-\dagger}$, and minimize by differentiating the log likelihood $L(S_B, \mathbf{d}) = -2 \ln \mathcal{L}(S_B, \mathbf{d})$ with respect to the unknowns S_B . Every step in the derivation leading to eq. (83) can be duplicated with the degree indices l and l' replaced by bin indices B and B' ; the resulting maximum-likelihood estimate of S_B is

$$\hat{S}_B^{\text{ML}} = \mathbf{d}^T \mathbf{Z}_B \mathbf{d} - \text{tr}(\mathbf{N}^\dagger \mathbf{Z}_B), \quad (107)$$

where

$$\mathbf{Z}_B = \frac{1}{2} \sum_{B'} F_{BB'}^{-1} (\mathbf{C}^{-\dagger} \mathbf{P}_B \mathbf{C}^{-\dagger}) \quad (108)$$

and

$$F_{BB'} = \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial S_B \partial S_{B'}} \right\rangle = \frac{1}{2} \text{tr}(\mathbf{C}^{-\dagger} \mathbf{P}_B \mathbf{C}^{-\dagger} \mathbf{P}_{B'}). \quad (109)$$

Upon utilizing eq. (106) we can express the band-averaged Fisher matrix (109) in terms of the generalized inverse (104) and the original unbinned Fisher matrix (79) in the form

$$F_{BB'} = \sum_{ll'} W_{Bl}^{\dagger T} F_{ll'} W_{l'B'}^{\dagger}, \quad (110)$$

where $W_{Bl}^{\dagger T} = W_{lB}^{\dagger}$. Eq. (107) is an unbiased estimator of the averaged quantity (102), that is, $\langle \hat{S}_B^{\text{ML}} \rangle = S_B$, by an argument analogous to that in eq. (85), and the covariance of two binned estimates is the inverse of the matrix (109)–(110),

$$\Sigma_{BB'}^{\text{ML}} = \text{cov}(\hat{S}_B^{\text{ML}}, \hat{S}_{B'}^{\text{ML}}) = F_{BB'}^{-1}, \quad (111)$$

by an argument analogous to that in eq. (86). The spacing of the bins B renders the matrix $F_{BB'}$ in eqs (109)–(110) invertible, enabling the quadratic estimator (107) to be numerically implemented and the associated covariance (111) to be determined. An argument analogous to that in Section 6.5 shows that the resulting estimate is minimum-variance, that is, $\text{var}(\hat{S}_B^{\text{ML}}) = F_{BB}^{-1} \leq \text{var}(\hat{S}_B)$ for any \hat{S}_B satisfying $\langle \hat{S}_B \rangle = S_B$. In the case of contiguous, boxcar-weighted bins the band-averaged Fisher matrix (110) is simply $F_{BB'} = \sum_{l \in B} \sum_{l' \in B'} F_{ll'}$.

6.7 The white album

The original unbinned maximum-likelihood estimate (83) can be computed without iteration in the special case that the signal and noise are both white: $S_l = S$ and $N_l = N$. Even for a region $R \neq \Omega$, the pixel-basis data covariance matrix can then be inverted:

$$\mathbf{C} = (S + N) \sum_l \mathbf{P}_l = (\Delta\Omega)^{-1} (S + N) \mathbf{I} \quad \text{so that} \quad \mathbf{C}^{-1} = \Delta\Omega (S + N)^{-1} \mathbf{I}. \quad (112)$$

The Fisher matrix obtained by substituting eq. (112) into (79) is related to the periodogram coupling matrix of (57) by

$$F_{ll'} = \frac{1}{2} \left(\frac{A}{4\pi} \right) \frac{2l + 1}{(S + N)^2} K_{ll'}, \quad (113)$$

so that the matrix defined in eq. (84) is given by $\mathbf{Z}_l = (4\pi/A)(\Delta\Omega)^2 \sum_{l'} K_{ll'}^{-1} (2l' + 1)^{-1} \mathbf{P}_{l'}$. Inserting this into eq. (83) and comparing with eq. (58) we find that the maximum-likelihood estimator coincides with the deconvolved periodogram estimator (69): $\hat{S}_l^{\text{ML}} = \hat{S}_l^{\text{DP}}$ if $S_l = S$ and $N_l = N$. The covariance computed using eq. (70) likewise coincides with the maximum-likelihood covariance (86):

$$\Sigma_{ll'}^{\text{DP}} = 2 \left(\frac{4\pi}{A} \right) \frac{(S + N)^2}{2l' + 1} K_{ll'}^{-1} = \Sigma_{ll'}^{\text{ML}}. \quad (114)$$

The deconvolved periodogram \hat{S}_l^{DP} is thus the best unbiased estimate of a white spectrum $S_l = S$ contaminated by white noise $N_l = N$.

6.8 Pros and cons

The maximum-likelihood method returns, if it exists, the unbiased spectral estimate with the lowest variance. This is obviously desirable, although we shall see that, by deliberately introducing bias, we may dramatically reduce the estimation variance of the resulting biased estimate. Furthermore, the maximum-likelihood method has a number of significant disadvantages:

(i) It is intrinsically non-linear, $\hat{S}_l^{\text{ML}} = f(\mathbf{d}, \hat{S}_l^{\text{ML}})$, requiring a good approximation to the spectrum S_l to begin the iteration, and such a good initial guess may not always be available. It is critical to start in the global minimum basin since the Newton–Raphson iteration (80) will only converge to the nearest local minimum.

(ii) Particularly for large data vectors $\mathbf{d} = (d_1 \ d_2 \ \dots \ d_J)^T$, computation of the inverse data covariance matrix \mathbf{C}^{-1} and the matrix products in eq. (80) can be a highly numerically intensive operation. The number of pixels in the WMAP cosmology experiment is $J \approx 3 \times 10^6$ at five wavelengths (Górski *et al.* 2005), and \mathbf{P}_l , $\mathbf{P}_{l'}$, \mathbf{C} and \mathbf{C}^{-1} are all non-sparse matrices. The nearly complete (80–85 per cent) sky coverage enabled the WMAP team to develop and implement a pre-conditioned conjugate gradient technique to compute the three ingredients needed to determine the estimate \hat{S}_l^{ML} and its covariance $\Sigma_{ll'}^{\text{ML}}$, namely $\mathbf{d}^T (\mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1}) \mathbf{d}$, $\text{tr}(\mathbf{C}^{-1} \mathbf{P}_l)$ and $\text{tr}(\mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1} \mathbf{P}_{l'})$ (Oh *et al.* 1999; Hinshaw *et al.* 2003). Computational demands continue to increase: the upcoming ‘Planck’ mission will detect $J \approx 50 \times 10^6$ pixels at nine wavelengths (Efstathiou *et al.* 2005).

(iii) Maximum-likelihood estimation of individual spectral values S_l is only numerically feasible for surveys such as WMAP that cover a substantial portion of the sphere; for smaller regions the method is limited to the estimation of binned values of the spectrum S_B , and it is necessary to assume that the true spectrum S_l can be adequately approximated by a coarse-grained spectrum S_l^{\dagger} that can be fully recovered from S_B . Even when $A \approx 4\pi$ it may be advantageous to plot binned or band-averaged values of the individual estimates, because $\text{var}(\hat{S}_l^{\text{ML}})$ may be very large, obscuring salient features of the spectrum.

The multitaper method—which we discuss next—is applicable to regions of arbitrary area $0 \leq A \leq 4\pi$, does not require iteration or large-scale matrix inversion, and gives the analyst easy control over the resolution-variance trade-off that is at the heart of spectral estimation. Under certain restrictive assumptions on the smoothness of the underlying spectrum of 1-D time-series, the maximum-likelihood and multitaper methods are approximately indistinguishable (Mullis & Scharf 1991; Stoica & Sundin 1999); however, even for the staircase spectra discussed in Section 6.6, proving a similar equivalence for anything but whole-sphere coverage has thus far eluded our attempts.

7 MULTITAPER SPECTRAL ESTIMATION

The multitaper method was first introduced into 1-D time-series analysis in a seminal paper by Thomson (1982), and has recently been generalized to spectral estimation on a sphere by Wiecezorek & Simons (2005, 2007). In essence, the method consists of multiplying the data by a series of specially designed orthogonal data tapers, and then combining the resulting spectra to obtain a single averaged estimate with reduced variance. In 1-D the tapers are the prolate spheroidal wavefunctions that are optimally concentrated in both the time and frequency domains (Slepian 1983; Percival & Walden 1993). We present a whirlwind review of the analogous spatio-spectral concentration problem on a sphere in the next section; for a more thorough discussion see Simons *et al.* (2006), which, we caution, however, used a real spherical harmonic basis rather than the complex eqs (5)–(7) used here for mathematical convenience.

7.1 Spherical Slepian functions

A ‘bandlimited’ spherical Slepian function is one that has no power outside of the spectral interval $0 \leq l \leq L$, that is,

$$g(\mathbf{r}) = \sum_{lm}^L g_{lm} Y_{lm}(\mathbf{r}), \quad (115)$$

but that has as much of its power as possible concentrated within a region R , that is,

$$\lambda = \frac{\int_R g^2(\mathbf{r}) d\Omega}{\int_{\Omega} g^2(\mathbf{r}) d\Omega} = \text{maximum}. \quad (116)$$

Functions (115) that render the spatial-basis Rayleigh quotient in eq. (116) stationary are solutions to the $(L+1)^2 \times (L+1)^2$ algebraic eigenvalue problem

$$\sum_{l'm'}^L D_{lm,l'm'} g_{l'm'} = \lambda g_{lm}, \quad (117)$$

where $D_{lm,l'm'} = D_{l'm',lm}^*$ are the spectral-basis matrix elements that we have encountered before, in eqs (23) and (53). The eigenvalues, which are a measure of the spatial concentration, are all real and positive, $\lambda = \lambda^*$ and $\lambda > 0$; in addition, the eigencolumns satisfy $g_{l-m} = (-1)^m g_{lm}^*$, so that the associated spatial eigenfunctions are all real, $g(\mathbf{r}) = g^*(\mathbf{r})$.

Instead of concentrating a bandlimited function $g(\mathbf{r})$ of the form (115) into a spatial region R , we could seek to concentrate a ‘spacelimited’ function,

$$h(\mathbf{r}) = \sum_{lm}^{\infty} h_{lm} Y_{lm}(\mathbf{r}), \quad \text{where} \quad h_{lm} = \int_R Y_{lm}^*(\mathbf{r}) h(\mathbf{r}) d\Omega, \quad (118)$$

that vanishes outside R , within a spectral interval $0 \leq l \leq L$. The concentration measure analogous to (116) in that case is

$$\lambda = \frac{\sum_{lm}^L |h_{lm}|^2}{\sum_{lm}^{\infty} |h_{lm}|^2} = \text{maximum}. \quad (119)$$

Functions (118) that render the spectral-basis Rayleigh quotient (119) stationary are solutions to the Fredholm integral eigenvalue equation

$$\int_R D(\mathbf{r}, \mathbf{r}') h(\mathbf{r}') d\Omega' = \lambda h(\mathbf{r}), \quad \mathbf{r} \in R, \quad (120)$$

where

$$D(\mathbf{r}, \mathbf{r}') = \sum_{lm}^L Y_{lm}(\mathbf{r}) Y_{lm}^*(\mathbf{r}') = \frac{1}{4\pi} \sum_l^L (2l+1) P_l(\mathbf{r} \cdot \mathbf{r}'). \quad (121)$$

In fact, the bandlimited and spacelimited eigenvalue problems (117) and (120) have the same eigenvalues λ and are each other’s duals. We are free to require that $h(\mathbf{r})$ and $g(\mathbf{r})$ coincide on the region of spatial concentration, that is, $h(\mathbf{r}) = g^R(\mathbf{r})$ or, equivalently,

$$h_{lm} = \sum_{l'm'}^L D_{lm,l'm'} g_{l'm'}, \quad 0 \leq l \leq \infty, \quad -l \leq m \leq l. \quad (122)$$

We shall focus primarily upon the bandlimited spherical Slepian functions $g(\mathbf{r})$ throughout the remainder of this paper.

We distinguish the $(L+1)^2$ eigensolutions by a Greek subscript, $\alpha = 1, 2, \dots, (L+1)^2$, and rank them in order of their concentration, that is, $1 > \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{(L+1)^2} > 0$. The largest eigenvalue λ_1 is strictly less than one because no function can be strictly contained within the spectral band $0 \leq l \leq L$ and the spatial region R simultaneously. The Hermitian symmetry $D_{lm,l'm'} = D_{l'm',lm}^*$ also guarantees that the eigencolumns $g_{\alpha,lm}$ in eq. (117) are mutually orthogonal; it is convenient in this application to adopt a normalization that is slightly different from that used by Simons *et al.* (2006), namely

$$\sum_{lm}^L g_{\alpha,lm}^* g_{\beta,lm} = 4\pi \delta_{\alpha\beta} \quad \text{and} \quad \sum_{lm}^L \sum_{l'm'}^L g_{\alpha,lm}^* D_{lm,l'm'} g_{\beta,l'm'} = 4\pi \lambda_{\alpha} \delta_{\alpha\beta} \quad (123)$$

or, equivalently,

$$\int_{\Omega} g_{\alpha}(\mathbf{r}) g_{\beta}(\mathbf{r}) d\Omega = 4\pi \delta_{\alpha\beta} \quad \text{and} \quad \int_R g_{\alpha}(\mathbf{r}) g_{\beta}(\mathbf{r}) d\Omega = 4\pi \lambda_{\alpha} \delta_{\alpha\beta}. \quad (124)$$

The eigenfunction $g_1(\mathbf{r})$ associated with the largest eigenvalue λ_1 is the bandlimited function that is most spatially concentrated within R , the eigenfunction $g_2(\mathbf{r})$ is the next best concentrated function of the form (115) orthogonal to $g_1(\mathbf{r})$, and so on.

The sum of the $(L + 1)^2$ eigenvalues is a diagnostic area-bandwidth product known as the ‘Shannon number’ which we denote by

$$K = \sum_{\alpha}^{(L+1)^2} \lambda_{\alpha} = \sum_{lm} D_{lm,lm} = \frac{A}{4\pi} (L + 1)^2. \quad (125)$$

A plot of λ_{α} versus the rank α resembles a step function, with the first K eigenfunctions $g_{\alpha}(\mathbf{r})$ having associated eigenvalues $\lambda_{\alpha} \approx 1$ and being well concentrated within the region R , and the remainder having associated eigenvalues $\lambda_{\alpha} \approx 0$ and being well concentrated within the complementary region $\Omega - R$. The eigenvalue-weighted sums of the product of two eigencolumns or eigenfunctions are given exactly by

$$\sum_{\alpha}^{(L+1)^2} \lambda_{\alpha} g_{\alpha,lm} g_{\alpha,l'm'}^* = 4\pi D_{lm,l'm'}, \quad (126)$$

$$\sum_{\alpha}^{(L+1)^2} \lambda_{\alpha} g_{\alpha}(\mathbf{r}) g_{\alpha}(\mathbf{r}') = 4\pi \sum_{lm}^L \sum_{l'm'}^L Y_{lm}(\mathbf{r}) D_{lm,l'm'} Y_{l'm'}^*(\mathbf{r}'). \quad (127)$$

Because of the steplike character of the λ_{α} versus α eigenvalue spectrum, we can approximate eqs (126)–(127) by unweighted sums over just the first K eigenfunctions:

$$\sum_{\alpha}^K g_{\alpha,lm} g_{\alpha,l'm'}^* \approx 4\pi D_{lm,l'm'}, \quad (128)$$

$$\sum_{\alpha}^K g_{\alpha}(\mathbf{r}) g_{\alpha}(\mathbf{r}') \approx 4\pi \sum_{lm}^L \sum_{l'm'}^L Y_{lm}(\mathbf{r}) D_{lm,l'm'} Y_{l'm'}^*(\mathbf{r}'). \quad (129)$$

Whenever the area of the region R is a small fraction of the area of the sphere, $A \ll 4\pi$, there will be many more well-excluded eigenfunctions $g_{\alpha}(\mathbf{r})$ with insignificant ($\lambda_{\alpha} \approx 0$) eigenvalues than well-concentrated ones with significant ($\lambda_{\alpha} \approx 1$) eigenvalues, that is, $K \ll (L + 1)^2$. In the opposite extreme of nearly whole-sphere coverage, $A \approx 4\pi$, there will be many more well-concentrated eigenfunctions $g_{\alpha}(\mathbf{r})$ than well-excluded ones, that is, $K \approx (L + 1)^2$.

The axisymmetry of a single or double polar cap enables the $(L + 1)^2 \times (L + 1)^2$ eigenvalue problem in eq. (117) to be decomposed into a series of $(L - m + 1) \times (L - m + 1)$ problems, one for each non-negative order $0 \leq m \leq L$. More importantly, the matrix governing each of these smaller fixed-order eigenvalue problems commutes with a tridiagonal matrix with analytically specified elements and a well-behaved spectrum, that can be diagonalized to find the bandlimited eigencolumns $g_{\alpha,lm}$ instead. We refrain from discussing this decomposition and the associated commuting matrix here, except to note that it makes the accurate computation of the well-concentrated eigenfunctions $g_{\alpha}(\mathbf{r})$ of even a large axisymmetric region R not only possible but essentially trivial (Grünbaum *et al.* 1982; Simons & Dahlen 2006; Simons *et al.* 2006).

7.2 Data availability

Thus far, in our discussion of the periodogram and maximum-likelihood estimators, we have taken the point of view that the available data $d(\mathbf{r})$ are strictly restricted to points \mathbf{r} within the region R . We shall henceforth adopt a slightly different viewpoint, namely that we are willing to allow data $d(\mathbf{r})$ from a narrow region on the periphery of R . This flexibility allows us to use the spatially concentrated, bandlimited tapers $g_{\alpha}(\mathbf{r})$ rather than the corresponding spectrally concentrated, spacelimited tapers $h_{\alpha}(\mathbf{r}) = g_{\alpha}^R(\mathbf{r})$ with spherical harmonic coefficients $h_{\alpha,lm}$ given by eq. (122). The small amount of spatial leakage from points \mathbf{r} outside of R that we accept is offset by the advantage that there is no broad-band bias in the resulting multitaper spectral estimates, as we shall see. The use of bandlimited rather than spacelimited tapers is natural in many geophysical applications, where we seek a spatially localized estimate of the spectrum S_l of a signal $s(\mathbf{r})$. In other applications the most natural viewpoint may be that the only available or usable data $d(\mathbf{r})$ truly are within a specified region R ; in that case, it is necessary to replace $g_{\alpha}(\mathbf{r})$ by $h_{\alpha}(\mathbf{r})$ in many of the formulas that follow, and the associated sums over $0 \leq l \leq L$ become sums over $0 \leq l \leq \infty$.

7.3 Single-taper spectral estimate

The first step in making a multitaper spectral estimate is to select the bandwidth L or the Shannon number $K = (A/4\pi)(L + 1)^2$ and compute the associated bandlimited tapers $g_{\alpha}(\mathbf{r})$, $\alpha = 1, 2, \dots, (L + 1)^2$ that are well concentrated in the region of interest R . To obtain the α th

single-taper estimate \hat{S}_l^α , we multiply the data $d(\mathbf{r})$ by $g_\alpha(\mathbf{r})$ prior to computing the noise-corrected power:

$$\hat{S}_l^\alpha = \frac{1}{2l+1} \sum_m \left| \int_\Omega g_\alpha(\mathbf{r}) d(\mathbf{r}) Y_{lm}^*(\mathbf{r}) d\Omega \right|^2 - \sum_{l'} M_{ll'}^\alpha N_{l'}. \quad (130)$$

The banded single-taper coupling matrix analogous to $K_{ll'}$ in eqs (56) and (57) is

$$M_{ll'}^\alpha = \left(\frac{2l'+1}{4\pi} \right) \sum_p (2p+1) G_{\alpha,p} \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2, \quad (131)$$

where

$$G_{\alpha,p} = \frac{1}{2p+1} \sum_q |g_{\alpha,pq}|^2, \quad 0 \leq p \leq L, \quad (132)$$

is the power spectrum of the bandlimited taper $g_\alpha(\mathbf{r})$. In the pixel basis eqs (130)–(131) become

$$\hat{S}_l^\alpha = \frac{(\Delta\Omega)^2}{2l+1} [\mathbf{d}^T \mathbf{G}_l^\alpha \mathbf{d} - \text{tr}(\mathbf{N} \mathbf{G}_l^\alpha)], \quad (133)$$

where \mathbf{G}_l^α is the $J \times J$ symmetric matrix with elements given by

$$(\mathbf{G}_l^\alpha)_{jj'} = g_\alpha(\mathbf{r}_j) \left[\sum_m Y_{lm}(\mathbf{r}_j) Y_{lm}^*(\mathbf{r}_{j'}) \right] g_\alpha(\mathbf{r}_{j'}) = \left(\frac{2l+1}{4\pi} \right) g_\alpha(\mathbf{r}_j) P_l(\mathbf{r}_j \cdot \mathbf{r}_{j'}) g_\alpha(\mathbf{r}_{j'}). \quad (134)$$

The expected value of the α th estimate (130) is

$$\begin{aligned} \langle \hat{S}_l^\alpha \rangle &= \frac{(\Delta\Omega)^2}{2l+1} [\text{tr}(\mathbf{C} \mathbf{G}_l^\alpha) - \text{tr}(\mathbf{N} \mathbf{G}_l^\alpha)] \\ &= \frac{(\Delta\Omega)^2}{2l+1} \text{tr}(\mathbf{S} \mathbf{G}_l^\alpha) \quad \text{noise bias cancels} \\ &= \frac{(\Delta\Omega)^2}{2l+1} \sum_{l'} S_{l'} \text{tr}(\mathbf{G}_l^\alpha \mathbf{P}_{l'}) \\ &= \sum_{l'} M_{ll'}^\alpha S_{l'}. \end{aligned} \quad (135)$$

To verify the final step in the reduction (135) and thereby confirm that the pixel-basis product

$$M_{ll'}^\alpha = \frac{(\Delta\Omega)^2}{2l+1} \text{tr}(\mathbf{G}_l^\alpha \mathbf{P}_{l'}) \quad (136)$$

is identical to the single-taper coupling matrix in eqs (131)–(132), we transform to the spatial basis and replace $b_{pq} \rightarrow g_{\alpha,pq}$ in the argument leading to eq. (55), to obtain the result

$$\text{tr}(\mathbf{G}_l^\alpha \mathbf{P}_{l'}) = \frac{(2l+1)(2l'+1)}{4\pi(\Delta\Omega)^2} \sum_p (2p+1) G_{\alpha,p} \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2. \quad (137)$$

Every row of the matrix $M_{ll'}^\alpha$ sums to unity,

$$\sum_{l'} M_{ll'}^\alpha = \frac{1}{4\pi} \sum_p (2p+1) G_{\alpha,p} = \frac{1}{4\pi} \int_\Omega g_\alpha^2(\mathbf{r}) d\Omega = 1, \quad (138)$$

by virtue of the 3- j identity (60). This is why we introduced the 4π normalization in eqs (123) and (124): to ensure that a single-taper spectral estimate \hat{S}_l^α has no leakage bias in the case of a perfectly white spectrum: $\langle \hat{S}_l^\alpha \rangle = S$ if $S_l = S$.

7.4 Multitaper estimate

A ‘multitaper’ spectral estimate is simply a weighted linear combination of single-taper estimates, of the form

$$\hat{S}_l^{\text{MT}} = \sum_\alpha c_\alpha \hat{S}_l^\alpha, \quad \text{where} \quad \sum_\alpha c_\alpha = 1. \quad (139)$$

The expected value of the estimate (139) is

$$\langle \hat{S}_l^{\text{MT}} \rangle = \sum_{l'} M_{ll'} \mathbf{S}_{l'}, \quad \text{where} \quad M_{ll'} = \sum_\alpha c_\alpha M_{ll'}^\alpha \quad (140)$$

is the multitaper coupling matrix. The constraint that the weights c_α in eq. (139) sum to unity guarantees that

$$\sum_{l'} M_{ll'} = 1 \quad \text{so that} \quad \langle \hat{S}_l^{\text{MT}} \rangle = S \quad \text{if} \quad S_l = S. \quad (141)$$

Apart from this constraint, the weights are at our disposal. Two simple choices are eigenvalue weighting of all $(L+1)^2$ tapers,

$$c_\alpha = K^{-1} \lambda_\alpha, \quad \alpha = 1, 2, \dots, (L+1)^2, \quad (142)$$

or equal weighting of only the first K tapers,

$$c_\alpha = \begin{cases} 1/K & \text{if } \alpha = 1, 2, \dots, K \\ 0 & \text{otherwise,} \end{cases} \quad (143)$$

where K is the Shannon number (125). We expect the two choices (142) and (143) to lead to nearly identical spectral estimates \hat{S}_l^{MT} for the same reason that eqs (128)–(129) are a good approximation to eqs (126)–(127). Eigenvalue weighting has theoretical advantages, enabling us to obtain a more succinct expression for the multitaper coupling matrix and covariance; however, uniform weighting of only the first K tapers is, in practice, the best way to compute an actual spectral estimate \hat{S}_l^{MT} , for reasons of efficiency. Truncation at the Shannon number K retains only the bandlimited tapers $g_\alpha(\mathbf{r})$ that are well concentrated within the region R , so that \hat{S}_l^{MT} can be viewed as a ‘spatially localized’ estimate of the spectrum S_l .

7.5 Leakage bias

The eigenvalue-weighted power spectrum of all $(L+1)^2$ tapers $g_\alpha(\mathbf{r})$ is simply

$$\sum_{\alpha}^{(L+1)^2} \lambda_\alpha G_{\alpha,p} = \frac{4\pi}{2p+1} \sum_q D_{pq,pq} = \int_R P_p(1) d\Omega = A \quad \text{for all } 0 \leq p \leq L, \quad (144)$$

by virtue of the identity (126). Because of this, the multitaper coupling matrix in eq. (140) reduces to

$$M_{ll'} = \frac{2l'+1}{(L+1)^2} \sum_p^L (2p+1) \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2. \quad (145)$$

It is remarkable that this result depends only upon the chosen bandwidth L and is completely independent of the size, shape or connectivity of the region R , even as $R = \Omega$. Eq. (145) is strictly valid only for eigenvalue weighting (142) but, as just noted, we expect it to be a very good approximation for uniform weighting of the first K tapers (143) as well. For $l, l' \gg L$ we can use the 3- j asymptotic relation (20) to approximate (145) further by

$$M_{ll'} \approx \frac{4\pi}{(L+1)^2} \sum_p^L [X_{p|l-l'|}(\pi/2)]^2. \quad (146)$$

This shows that for large l we expect $M_{ll'}$ to take on a universal shape that depends only upon L and the offset from the target degree $|l' - l|$. Both the exact asymmetric relation (145), as we have seen before, and the symmetric large- l approximation (146), by the spherical harmonic addition theorem, satisfy the constraint (141).

In Fig. 6, we illustrate the variation of the coupling matrix $M_{ll'}$ versus the column index $0 \leq l' \leq 100$ for various target degrees $l = 0, 10, 20, 30, 40, 50$ and two different bandwidths, $L = 20$ and $L = 10$. A major advantage of the multitaper method is the easy control that it affords over the spectral leakage and resolution; the coupling is strictly confined to the interval $|l' - l| \leq L$, of width $L + \min(l, L) + 1$, regardless of the size, shape or connectivity of the region R . The ‘triangular’ coupling to the monopole degree $l = 0$ is, by virtue of (19), exactly described by the relation $M_{0l'} = (2l' + 1)/(L + 1)^2$, $0 \leq l' \leq L$; that is, the degree-zero estimate \hat{S}_0^{MT} is really an estimate of the total power within the band $0 \leq l' \leq L$. As the target degree l increases the coupling matrix $M_{ll'}$ increasingly takes on a domelike universal shape that is approximately described by eq. (146). Fig. 7 shows a plot of this large- l limit for four different bandwidths, $L = 5, 10, 20, 30$; the abscissa is the offset from the target degree, $l' - l$, which is confined to the closed interval $[-L, L]$. Roughly speaking the shapes are all scaled versions of each other; recall that the height of the $2L + 1$ bars in every graph must sum to 100 per cent.

7.6 Multitaper covariance

The covariance of two multitaper estimates (139) is a doubly weighted sum over all of the single-taper cross-covariances:

$$\Sigma_{ll'}^{\text{MT}} = \text{cov}(\hat{S}_l^{\text{MT}}, \hat{S}_{l'}^{\text{MT}}) = \sum_{\alpha\beta} c_\alpha \Sigma_{ll'}^{\alpha\beta} c_\beta, \quad (147)$$

where, as usual via the Isserlis identity (44), we have

$$\Sigma_{ll'}^{\alpha\beta} = \text{cov}(\hat{S}_l^\alpha, \hat{S}_{l'}^\beta) = \frac{2(\Delta\Omega)^4}{(2l+1)(2l'+1)} \text{tr}(\mathbf{C}\mathbf{G}_l^\alpha \mathbf{C}\mathbf{G}_{l'}^\beta). \quad (148)$$

Transforming to the spatial basis as in the derivation of eq. (68) we obtain

$$\Sigma_{ll'}^{\alpha\beta} = \frac{2}{(2l+1)(2l'+1)} \sum_{mm'} \left| \sum_{pq} (S_p + N_p) \int_\Omega g_\alpha(\mathbf{r}) Y_{pq}^*(\mathbf{r}) Y_{lm}(\mathbf{r}) d\Omega \int_\Omega g_\beta(\mathbf{r}') Y_{pq}(\mathbf{r}') Y_{l'm'}^*(\mathbf{r}') d\Omega' \right|^2 \quad (149)$$

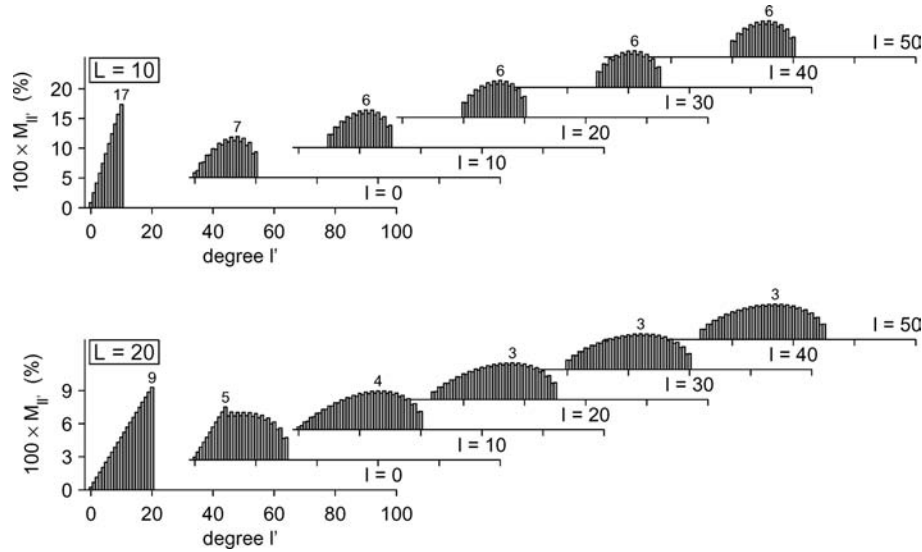


Figure 6. Bar plots of the multitaper coupling matrix $100 \times M_{l'l'}$ for bandwidths $L = 10$ (top panel) and $L = 20$ (bottom panel). The (occasionally obscured) tick marks are at $l' = 0, 20, 40, 60, 80, 100$ on every offset abscissa; the target degrees $l = 0, 10, 20, 30, 40, 50$ are indicated on the right-hand side. The height of each bar reflects the per cent leakage of the power at degree l' into the multitaper estimate \hat{S}_l^{MT} , in accordance with the constraint (141). Small numbers on top are the maximum value of $100 \times M_{l'l'}$ for every target degree l .

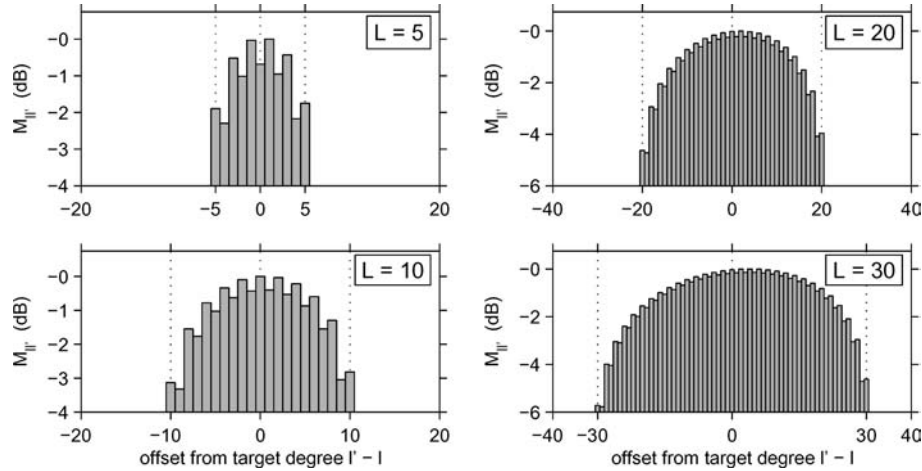


Figure 7. Large- l limits of the multitaper coupling matrix $M_{l'l'}$, plotted versus the offset $l' - l$ from the target angular degree, for bandwidths for $L = 5$ (top left-hand panel), $L = 10$ (bottom left-hand panel), $L = 20$ (top right-hand panel) and $L = 30$ (bottom right-hand panel). The limiting shapes were found empirically by increasing l until the plots no longer changed visibly. The slight asymmetry reflects the inaccuracy of the approximation (146); the exact coupling matrix (145) is asymmetric because of the leading factor of $2l' + 1$.

or, equivalently,

$$\Sigma_{l'l'}^{\alpha\beta} = \frac{1}{8\pi^2} \sum_{mm'} \left| \sum_{pq} (2p+1)(S_p + N_p) \sum_{st} \sum_{s't'}^L \sqrt{(2s+1)(2s'+1)} g_{\alpha,st} g_{\beta,s't'}^* \begin{pmatrix} l & p & s \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l' & p & s' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & p & s \\ m & q & t \end{pmatrix} \begin{pmatrix} l' & p & s' \\ m' & q & t' \end{pmatrix} \right|^2. \quad (150)$$

It is noteworthy that $\Sigma_{l'l'}^{\alpha\alpha} = \Sigma_{l'l}^{\alpha\alpha}$ and $\Sigma_{ll}^{\alpha\beta} = \Sigma_{ll}^{\beta\alpha}$; however, it is not in general true that $\Sigma_{l'l'}^{\alpha\beta} = \Sigma_{l'l}^{\alpha\beta} = \Sigma_{l'l'}^{\beta\alpha}$. Wiecezorek & Simons (2007) plot examples of eq. (150) for azimuthally symmetric tapers and parametrized input spectra S_p . We caution, however, that their notation and normalization conventions differ from the ones used in this paper. Eqs (149) and (150) show that every element of the multitaper-covariance matrix is positive, $\Sigma_{l'l'}^{\text{MT}} > 0$, as long as the weights are positive, $c_\alpha > 0$. We shall henceforth limit attention to the case of eigenvalue weighting, $c_\alpha = K^{-1} \lambda_\alpha$, $\alpha = 1, 2, \dots, (L+1)^2$. The eigenvalue-weighted multitaper covariance $\Sigma_{l'l'}^{\text{MT}}$ can be written in a relatively simple approximate form in the case of a moderately coloured spectrum, as we show in Section 8.3.

7.7 Bias and mean-squared error

The bias of an eigenvalue-weighted multitaper estimate $\hat{S}_l^{\text{MT}} = K^{-1} \sum_{\alpha} \lambda_{\alpha} \hat{S}_l^{\alpha}$ is the discrepancy between its expected value and the true spectrum:

$$\text{bias}(\hat{S}_l^{\text{MT}}) = \langle \hat{S}_l^{\text{MT}} \rangle - S_l = \sum_{|l' - l| \leq L} (M_{ll'} - \delta_{ll'}) S_{l'}. \quad (151)$$

The bandlimited character of the tapers $g_{\alpha}(\mathbf{r})$, $\alpha = 1, 2, \dots, (L + 1)^2$ ensures that the bias is purely local; there is no broad-band bias from harmonic degrees l' outside of the coupling interval $|l' - l| \leq L$. If the spectrum is not highly coloured within this band, in the sense $S_{l'} \approx S_l$, the bias will be small: $\sum_{|l' - l| \leq L} (M_{ll'} - \delta_{ll'}) S_{l'} \approx S_l \sum_{|l' - l| \leq L} (M_{ll'} - \delta_{ll'}) = 0$, by virtue of (141). The total estimation error is given by $\hat{S}_l^{\text{MT}} - S_l$ and the mean-squared error is the expectation of the square of this:

$$\text{mse}(\hat{S}_l^{\text{MT}}) = \langle (\hat{S}_l^{\text{MT}} - S_l)^2 \rangle. \quad (152)$$

As is true for any estimate (e.g. Cox & Hinkley 1974; Bendat & Piersol 2000), the mean-squared error is the sum of the variance and the square of the bias:

$$\text{mse}(\hat{S}_l^{\text{MT}}) = \text{var}(\hat{S}_l^{\text{MT}}) + \text{bias}^2(\hat{S}_l^{\text{MT}}). \quad (153)$$

In CMB analyses the bias of \hat{S}_l^{MT} may not be a particularly critical issue because the ultimate objective (e.g. Jungman *et al.* 1996) is to determine ~ 10 cosmological parameters that characterize the inflationary universe (the baryonic-matter, cold-dark-matter and dark-energy densities Ω_b , Ω_c , Ω_{Λ} ; the Hubble constant H_0 , etc.) and this downstream estimation can be grounded upon estimates of either S_l or $\sum_{l'} M_{ll'} S_{l'}$ as long as the coupling matrix $M_{ll'}$ is known. In writing this we do not mean to suggest that a two-step procedure (estimating the spectrum before fitting parametrized models to it) is necessarily the optimal way of recovering cosmological information (a point well acknowledged by many authors, e.g. Bond *et al.* 1994; Jaffe *et al.* 1999; Oh *et al.* 1999; Christensen *et al.* 2001), nor to deny in any way that fundamental assumptions (e.g. of Gaussianity and isotropy, as in eq. 26) remain constantly under scrutiny and subject to revision. The ‘standard’ cosmological concordance model relies on a very simple inflationary scenario, with initial fluctuations seeding homogeneous and isotropic Gaussian temperature fluctuations in the CMB that are fully described by its angular power spectrum; in this paper we have written expressions for the bias and variance of various estimators of the power spectrum that are conditional on these assumptions. Thus, we have confined ourselves to dealing with the ‘stochastic uncertainty’ arising from variability within the chosen class of models, but ignored the ‘inductive uncertainty’ by which one might question whether the models apply at all (Pawitan 2001), knowing that this would make for a very different paper altogether.

8 MODERATELY COLOURED SPECTRA

Eq. (150) and the analogous expression for the periodogram covariance, eq. (68), are lengthy and difficult to evaluate numerically, which would be required repeatedly when constructing adaptively weighted multitaper estimates (Thomson 1982; Percival & Walden 1993; Wieczorek & Simons 2007). In this section we derive simpler expressions for $\Sigma_{ll'}^{\text{SP}}$, $\Sigma_{ll'}^{\text{MT}}$ and the Fisher matrix $F_{ll'}$ that should be good approximations for ‘moderately coloured’ spectra, for which it is permissible to replace

$$S_p + N_p \leftrightarrow \sqrt{(S_l + N_l)(S_{l'} + N_{l'})} \quad (154)$$

in equations such as (67) and (149). We write the resulting approximations using an $=$ sign rather than an \approx sign, even though they are all strictly valid only in the case of a white signal contaminated by white noise: $S_l = S$ and $N_l = N$.

8.1 Periodogram covariance

Upon making the substitution (154) into eq. (67) and making use of the first of the identities in eq. (24), we obtain

$$\Sigma_{ll'}^{\text{SP}} = \frac{2(4\pi/A)^2}{(2l+1)(2l'+1)} (S_l + N_l)(S_{l'} + N_{l'}) \sum_{mm'} |D_{lm,l'm'}|^2, \quad (155)$$

or, via eq. (57), equivalently,

$$\Sigma_{ll'}^{\text{SP}} = \frac{1}{2\pi} \left(\frac{4\pi}{A} \right)^2 (S_l + N_l)(S_{l'} + N_{l'}) \sum_p (2p+1) B_p \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2 = \frac{8\pi}{A} (S_l + N_l)(S_{l'} + N_{l'})(2l'+1)^{-1} K_{ll'}. \quad (156)$$

The covariance (156) for a moderately coloured spectrum will be a better approximation for a large region, $A \approx 4\pi$, than for a small one, $A \ll 4\pi$, because the extent of the coupling $K_{ll'}$ and thus the bandwidth over which the variation of the spectrum must be regarded as moderate increases as the size of the region R shrinks (see Fig. 4). In the limit (63) of a vanishingly small region, the signal and noise must be completely white, $S_l = S$ and $N_l = N$, in order for eq. (156) to be useful, and in that limit $B_p \rightarrow A^2/(4\pi)$ so that $\Sigma_{ll'}^{\text{SP}} \rightarrow 2(S + N)^2 \delta_{ll'}$, following eq. (60).

8.2 Fisher matrix

The inverse of the pixel-basis data covariance matrix \mathbf{C} can be approximated in the case of a moderately coloured spectrum (154) by a simple generalization of the exact result (112) for a white spectrum, that is, by making the change

$$\mathbf{C}^{-1} \leftrightarrow \frac{\Delta \Omega \mathbf{I}}{\sqrt{(S_l + N_l)(S_{l'} + N_{l'})}} \quad (157)$$

in eq. (79) or—as can be derived from eq. (92) with eqs (154) and (24) or, alternatively, via eqs (88) and (23)—upon inserting the equivalent spectral-basis approximation

$$V_{lm,l'm'} = \frac{D_{lm,l'm'}}{\sqrt{(S_l + N_l)(S_{l'} + N_{l'})}} \quad (158)$$

into eq. (87), we obtain a compact approximate formula for the Fisher matrix:

$$F_{ll'} = \frac{1}{2} (S_l + N_l)^{-1} (S_{l'} + N_{l'})^{-1} \sum_{mm'} |D_{lm,l'm'}|^2, \quad (159)$$

or, equivalently,

$$F_{ll'} = \frac{1}{8\pi} \frac{(2l+1)(2l'+1)}{(S_l + N_l)(S_{l'} + N_{l'})} \sum_p (2p+1) B_p \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2 = \frac{A}{8\pi} (2l+1)(S_l + N_l)^{-1} (S_{l'} + N_{l'})^{-1} K_{ll'}. \quad (160)$$

The result (160), which is due to Hinshaw *et al.* (2003), will also be more accurate for a large region than for a small one; in the limit of whole-sphere coverage, $B_p \rightarrow 4\pi \delta_{p0}$ and $K_{ll'} \rightarrow \delta_{ll'}$ so that $\Sigma_{ll'}^{\text{SP}} \rightarrow 2(2l+1)^{-1} (S_l + N_l)^2 \delta_{ll'}$ and $F_{ll'} \rightarrow \frac{1}{2} (2l+1)(S_l + N_l)^{-2} \delta_{ll'}$, in agreement with eqs (46) and (93). Per (86), the maximum-likelihood covariance $\Sigma_{ll'}^{\text{ML}} = F_{ll'}^{-1}$.

8.3 Multitaper covariance

The assumption that the spectrum is moderately coloured is less restrictive for a multitaper spectral estimate \hat{S}_l^{MT} than for a periodogram estimate \hat{S}_l^{SP} , because the coupling $M_{ll'}$ is confined to a narrow band, of width $L + \min(l, l') + 1$, that is independent of the size, shape or connectivity of the region R . Upon modifying eq. (149) with eq. (154) and using eq. (10) we can write the cross-covariance of two single-taper estimates in the form

$$\Sigma_{ll'}^{\alpha\beta} = \frac{2(S_l + N_l)(S_{l'} + N_{l'})}{(2l+1)(2l'+1)} \sum_{mm'} \left| \int_{\Omega} g_{\alpha}(\mathbf{r}) g_{\beta}(\mathbf{r}) Y_{lm}(\mathbf{r}) Y_{l'm'}^*(\mathbf{r}) d\Omega \right|^2, \quad (161)$$

where we have used the representation (10) of the Dirac delta function to reduce the two integrals inside the absolute value signs to one. Upon utilizing the spherical harmonic product identity (13) and evaluating the sum over m and m' using eq. (16) as in the derivation (52)–(55), we can reduce eq. (161) to

$$\Sigma_{ll'}^{\alpha\beta} = \frac{1}{2\pi} (S_l + N_l)(S_{l'} + N_{l'}) \sum_{pq} \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2 \left| \int_{\Omega} g_{\alpha}(\mathbf{r}) g_{\beta}(\mathbf{r}) Y_{pq}(\mathbf{r}) d\Omega \right|^2. \quad (162)$$

Substituting the representation (115) of $g_{\alpha}(\mathbf{r})$ and $g_{\beta}(\mathbf{r})$ and using eq. (12) we can write eq. (162) in the convenient form

$$\Sigma_{ll'}^{\alpha\beta} = \frac{1}{2\pi} (S_l + N_l)(S_{l'} + N_{l'}) \sum_p (2p+1) \Gamma_p^{\alpha\beta} \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2, \quad (163)$$

where we have defined the quantities

$$\Gamma_p^{\alpha\beta} = \frac{1}{4\pi} \sum_q \left| \sum_{st} \sum_{uv} \sqrt{(2s+1)(2u+1)} g_{\alpha,st} g_{\beta,uv} \begin{pmatrix} s & p & u \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} s & p & u \\ t & q & v \end{pmatrix} \right|^2. \quad (164)$$

It is noteworthy that all the symmetries $\Sigma_{ll'}^{\alpha\beta} = \Sigma_{l'l}^{\alpha\beta} = \Sigma_{ll'}^{\beta\alpha}$ pertain in this moderately coloured approximation. For azimuthally symmetric tapers $g_{\alpha}(\mathbf{r})$, $\alpha = 1, 2, \dots, (L+1)^2$, the matrix $\Sigma_{ll'}^{\alpha\beta}$ in this approximation is almost, but not quite, diagonal (see, e.g. Wiecek & Simons 2007), showing that the individual estimates \hat{S}_l^{α} , $\alpha = 1, 2, \dots, (L+1)^2$, are almost uncorrelated statistically. This is a highly desirable feature that embodies the very essence of the multitaper method: to dramatically reduce the estimation variance at the cost of small increases of bias.

In principle, we could now envisage weighting the single-taper estimates adaptively to minimize quality measures such as estimation variance or mean-squared error (Thomson 1982; Wiecek & Simons 2007). In practice, these methods tend to be rather computationally demanding. We thus propose to retain the eigenvalue-weighted multitaper procedure of eqs (139) and (142). The eigenvalue-weighted multitaper covariance is given by a formula analogous to eq. (163), namely

$$\Sigma_{ll'}^{\text{MT}} = \frac{1}{2\pi} (S_l + N_l)(S_{l'} + N_{l'}) \sum_p (2p+1) \Gamma_p \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2, \quad (165)$$

where

$$\Gamma_p = \frac{1}{K^2} \sum_{\alpha\beta}^{(L+1)^2} \lambda_\alpha \Gamma_p^{\alpha\beta} \lambda_\beta. \quad (166)$$

Upon using the identity (126) to express the double sum in eq. (166) in terms of $D_{st,s't'}$ and $D_{uv,u'v'}$ and then using the boxcar window function (48) to express these matrix elements as integrals of three spherical harmonics over the whole sphere Ω , to be reduced using eq. (12), we obtain a fivefold sum over the order indices t, t', v, v' and q , which can be reduced with the aid of eq. (17), leading to the relatively simple result

$$\Gamma_p = \frac{1}{K^2} \sum_{ss'}^L \sum_{uu'}^L (2s+1)(2s'+1)(2u+1)(2u'+1) \sum_e^{2L} (-1)^{p+e} (2e+1) B_e \times \begin{Bmatrix} s & e & s' \\ u & p & u' \end{Bmatrix} \begin{pmatrix} s & e & s' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u & e & u' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} s & p & u' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} u & p & s' \\ 0 & 0 & 0 \end{pmatrix}, \quad (167)$$

where B_e is the boxcar power, which depends on the shape of the region of interest, summed over angular degrees limited by 3- j selection rules to $0 \leq e \leq 2L$. The sums in eqs (163) and (165) are likewise limited to degrees $0 \leq p \leq 2L$, inasmuch as $\Gamma_p^{\alpha\beta} = 0$ and $\Gamma_p = 0$ when $p > 2L$. The effect of tapering with windows bandlimited to L is to introduce covariance between the estimates at any two different degrees l and l' that are separated by fewer than $2L + 1$ degrees. Eqs (165) and (167) together are our final most succinct formula for the covariance $\Sigma_{ll'}^{\text{MT}}$ of a multitaper estimate in the case of a moderately coloured spectrum. To our benefit, they are very efficiently computable, which should make them competitive with jackknifed estimates of the estimation variance (Chave *et al.* 1987; Thomson & Chave 1991; Thomson 2007).

8.4 Whole-sphere and infinitesimal-area limits

It would obviously be perverse to contemplate using the multitaper method in the case of whole-sphere coverage; we nevertheless present an analysis of the $A \rightarrow 4\pi$ limit of the covariance $\Sigma_{ll'}^{\text{MT}}$ in the interest of completeness. In that limit $B_e \rightarrow 4\pi \delta_{e0}$, and both eqs (19) can be used to reduce eq. (167) to

$$\Gamma_p^{A=4\pi} = \frac{4\pi}{(L+1)^4} \sum_{ss'}^L (2s+1)(2s'+1) \begin{pmatrix} s & p & s' \\ 0 & 0 & 0 \end{pmatrix}^2, \quad (168)$$

and thereby the multitaper covariance (165) to

$$\Sigma_{ll'}^{\text{MT}} = \frac{2(S_l + N_l)(S_{l'} + N_{l'})}{(L+1)^4} \sum_{ss'}^L (2s+1)(2s'+1) \sum_p^{2L} (2p+1) \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2 \begin{pmatrix} s & p & s' \\ 0 & 0 & 0 \end{pmatrix}^2. \quad (169)$$

If the band-averaged quantities $\sum_{l'} M_{ll'} S_{l'}$ of eq. (140) were to be estimated using the maximum-likelihood method with whole-sphere coverage, the covariance in the moderately coloured approximation (154) is

$$\begin{aligned} \text{cov} \left(\sum_p M_{lp} \hat{S}_p^{\text{ML}}, \sum_{p'} M_{l'p'} \hat{S}_{p'}^{\text{ML}} \right) &= \sum_{pp'} M_{lp} \Sigma_{pp'}^{\text{WS}} M_{p'l'}^T \\ &= \frac{2(S_l + N_l)(S_{l'} + N_{l'})}{(L+1)^4} \sum_{ss'}^L (2s+1)(2s'+1) \sum_p^{2L} (2p+1) \begin{pmatrix} l & p & s \\ 0 & 0 & 0 \end{pmatrix}^2 \begin{pmatrix} l' & p & s' \\ 0 & 0 & 0 \end{pmatrix}^2. \end{aligned} \quad (170)$$

In fact, eqs (169) and (170) are identical by virtue of the 3- j identity

$$\begin{aligned} \sum_p (2p+1) \begin{pmatrix} l & p & l' \\ 0 & 0 & 0 \end{pmatrix}^2 \begin{pmatrix} s & p & s' \\ 0 & 0 & 0 \end{pmatrix}^2 &= \frac{1}{2} \iint_{-1}^1 P_l(\mu) P_{l'}(\mu) \left[\sum_p \left(\frac{2p+1}{2} \right) P_p(\mu) P_p(\mu') \right] P_s(\mu') P_{s'}(\mu') d\mu d\mu' \\ &= \frac{1}{2} \int_{-1}^1 P_l(\mu) P_{l'}(\mu) P_s(\mu) P_{s'}(\mu) d\mu \\ &= \frac{1}{2} \sum_p (2p+1) \begin{pmatrix} l & p & s \\ 0 & 0 & 0 \end{pmatrix}^2 \int_{-1}^1 P_{l'}(\mu) P_p(\mu) P_{s'}(\mu) d\mu \\ &= \sum_p (2p+1) \begin{pmatrix} l & p & s \\ 0 & 0 & 0 \end{pmatrix}^2 \begin{pmatrix} l' & p & s' \\ 0 & 0 & 0 \end{pmatrix}^2, \end{aligned} \quad (171)$$

where we have used the Legendre product identity (14), and the representation (11) of the Dirac delta function $\delta(\mu - \mu')$ to reduce the double integral in the second line. The above argument shows that the eigenvalue-weighted multitaper estimate \hat{S}_l^{MT} is the minimum-variance unbiased estimate of the averaged spectrum $\sum_{l'} M_{ll'} S_{l'}$ in the limit $R = \Omega$. In practice, if we should ever be blessed with whole-sphere coverage, it would be easiest to compute this minimum-variance spectral estimate by simply forming a weighted average of the whole-sphere estimates (38)–(39). As we have just shown, eq. (170) specifies the covariance of such an estimate.

Recalling that $B_\epsilon \rightarrow A^2/(4\pi)$ in the opposite limit of an infinitesimally small region and making use of the identity (18), we find that eq. (167) reduces to

$$\Gamma_p^{A \rightarrow 0} = \frac{4\pi}{(L+1)^4} \left[\sum_{ss'}^L (2s+1)(2s'+1) \begin{pmatrix} s & p & s' \\ 0 & 0 & 0 \end{pmatrix}^2 \right]^2, \quad (172)$$

where we note that $\Gamma_0^{A \rightarrow 0} = 4\pi$. The resulting infinitesimal-area limit of the multitaper covariance $\Sigma_{ll'}^{\text{MT}}$ for a fixed bandwidth L is again of the form (165), with Γ_p replaced by its limiting value (172). If the Shannon number $K = (A/4\pi)(L+1)^2$ rather than the bandwidth L is held constant in taking the limit $A \rightarrow 0$, then the multitaper coupling matrix (145) tends to $M_{ll'} \rightarrow K^{-1}(A/4\pi)(2l'+1)$, that is, all degrees across the entire spectrum are coupled. Both the signal and the noise must then be white for the limiting covariance, $\Sigma_{ll'}^{\text{MT}} \rightarrow 2(S+N)^2$, to be a reasonable approximation. The latter can be derived by noting that, in taking the limit as prescribed by eq. (63) and using eq. (60), the fixed- K result is $\Gamma_p^{A \rightarrow 0} = 4\pi$ rather than (172).

9 SPECTRAL SHOOTOUT

In this section we conduct a numerical variance analysis of the various estimates \hat{S}_l^{SP} , \hat{S}_l^{DP} , \hat{S}_l^{ML} and \hat{S}_l^{MT} . We use the variance (47) of the whole-sphere estimate \hat{S}_l^{WS} as a standard of comparison, computing the variance ratio

$$(\sigma_l^2)^{\text{XX}} = \text{var}(\hat{S}_l^{\text{XX}})/\text{var}(\hat{S}_l^{\text{WS}}) = \Sigma_{ll}^{\text{XX}} / \Sigma_{ll}^{\text{WS}}, \quad (173)$$

where XX stands for any of the acronyms SP, DP, ML or MT. The numerators in eq. (173) are computed using the moderately coloured approximations for Σ_{ll}^{XX} derived in Section 8. This has the advantage that a common factor of $(S_l + N_l)^2$ cancels, leading to ratios $(\sigma_l^2)^{\text{XX}}$ that are independent of the signal and noise spectra S_l , N_l . Although the results we exhibit should be reasonable approximations for moderately coloured spectra, they are only strictly correct in the case of a white signal, $S_l = S$, contaminated by white noise, $N_l = N$.

9.1 Variance of a periodogram estimate

Fig. 8 shows the variation with degree l of the spherical-periodogram variance ratio,

$$(\sigma_l^2)^{\text{SP}} = \left(\frac{2l+1}{4\pi} \right) \left(\frac{4\pi}{A} \right)^2 \sum_p (2p+1) B_p \begin{pmatrix} l & p & l \\ 0 & 0 & 0 \end{pmatrix}^2, \quad (174)$$

for single and double polar caps of radii $\Theta = 3^\circ, 4^\circ, 5^\circ, 7^\circ, 10^\circ, 20^\circ, 60^\circ$. The summation index p is limited by 3- j selection rules to even values, with the result that eq. (174) yields identical results for a single and double cap of the same radius Θ , by virtue of the relations (51) and $A^{\text{cut}} = 2A^{\text{cap}}$; stated another way, each double-cap estimate \hat{S}_l^{SP} averages over half as many adjacent degrees l' with a weighting $K_{ll'}$ that is twice as large. The monopole variance ratio is $(\sigma_0^2)^{\text{SP}} = 1$ regardless of the cap size Θ , but as the harmonic degree increases the variance ratio does as well, reaching a maximum at $l \approx 60^\circ/\Theta$ and then oscillating mildly before eventually leveling off at a large- l limit given by

$$(\sigma_\infty^2)^{\text{SP}} = \frac{4\pi}{A^2} \sum_p (2p+1) B_p [P_p(0)]^2, \quad (175)$$

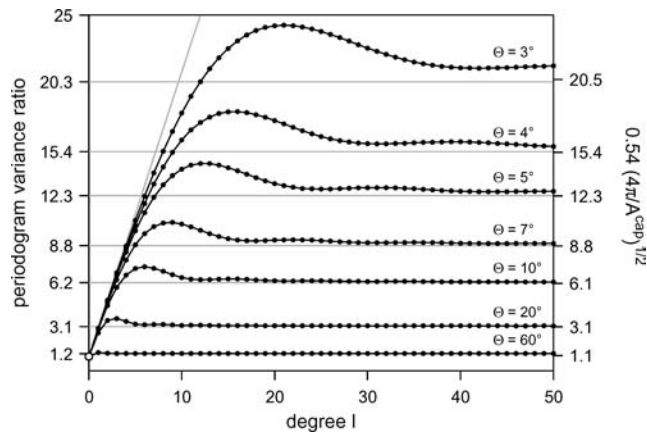


Figure 8. Black dots connected by black lines show the periodogram variance ratio $(\sigma_l^2)^{\text{SP}}$ as a function of degree $0 \leq l \leq 50$ for single and double polar caps of radii $\Theta = 3^\circ, 4^\circ, 5^\circ, 7^\circ, 10^\circ, 20^\circ, 60^\circ$. Grey horizontal lines labeled along the left-hand vertical axis show the large- l limits $(\sigma_\infty^2)^{\text{SP}}$. The open circle is the common monopole variance ratio $(\sigma_0^2)^{\text{SP}} = 1$; the diagonal grey line is the infinitesimal-area limit $(\sigma_l^2)^{\text{SP}} \rightarrow 2l+1$. Labeled tick marks on the right-hand side show the approximation $(\sigma_\infty^2)^{\text{SP}} \approx 0.54(4\pi/A^{\text{cap}})^{1/2}$. It is noteworthy that $(\sigma_l^2)^{\text{SP}} \geq (\sigma_0^2)^{\text{SP}}$ for all Θ , with equality prevailing only in the limit $\Theta = 90^\circ$: half-sphere coverage with a single cap yields the same variance as whole-sphere coverage.

where

$$P_p(0) = \begin{cases} 0 & \text{if } p \text{ is odd,} \\ p! 2^{-p} [(p/2)!]^{-2} & \text{if } p \text{ is even} \end{cases} \quad (176)$$

is the value of the Legendre polynomial of degree p at the argument $\mu = 0$. The oscillatory interval is wider for small regions, $A \ll 4\pi$, than for large ones, $A \approx 4\pi$. As expected, the high-degree variance (175) is greater for a smaller single or double cap, for example, $(\sigma_\infty^2)^{\text{SP}} = 12.3$ for $\Theta = 5^\circ$ versus $(\sigma_\infty^2)^{\text{SP}} = 6.2$ for $\Theta = 10^\circ$, because there are fewer pixelized data available to constrain the estimate \hat{S}_l^{SP} . A useful empirical approximation to eq. (175) for $\Theta \lesssim 65^\circ$ is $(\sigma_\infty^2)^{\text{SP}} \approx 0.54(4\pi/A^{\text{cap}})^{1/2}$, which can be read off the right-hand axis. In the limiting case of an infinitesimally small area, $A \rightarrow 0$, the variance is divergent; in fact, letting $B_p \rightarrow A^2/(4\pi)$ in eq. (174) we find that $(\sigma_l^2)^{\text{SP}} \rightarrow 2l + 1$ for all $0 \leq l \leq \infty$.

9.2 Variance of a maximum-likelihood estimate

The maximum-likelihood estimate \hat{S}_l^{ML} and the deconvolved periodogram estimate \hat{S}_l^{DP} coincide in the case $S_l = S$ and $N_l = N$, as we showed in Section 6.7, and their common variance ratio is given by

$$(\sigma_l^2)^{\text{ML}} = (\sigma_l^2)^{\text{DP}} = \left(\frac{4\pi}{A}\right) K_{ll}^{-1}. \quad (177)$$

To evaluate the ratio (177) we must compute and invert the boxcar coupling matrix $K_{ll'}$ of eq. (57), taking care to avoid truncation effects from large values of l and l' . Fig. 9 shows the variation of $(\sigma_l^2)^{\text{ML}} = (\sigma_l^2)^{\text{DP}}$ with degree l for four double polar caps with radii $\Theta \geq 75^\circ$. For double caps that cover less of the sphere, the matrix $K_{ll'}$ is too ill-conditioned to be invertible, and neither maximum-likelihood estimation (83) nor deconvolution (69) of the periodogram estimate \hat{S}_l^{SP} is numerically feasible. As expected, the maximum-likelihood variance is larger than the undeconvolved periodogram variance, e.g. $(\sigma_\infty^2)^{\text{ML}} = (\sigma_\infty^2)^{\text{DP}} \approx 1.75$ versus $(\sigma_\infty^2)^{\text{SP}} \approx 1.05$ for a double cap of radius $\Theta = 75^\circ$, because the averaging of the periodogram degrades the spectral resolution but improves the variance. In the limit of nearly whole-sphere coverage the maximum likelihood variance ratio can be approximated by $(\sigma_l^2)^{\text{ML}} = (\sigma_l^2)^{\text{DP}} \approx (4\pi/A)^2$ shown on the right-hand axis; that is, the standard error is increased relative to that of a whole-sphere estimate by roughly the reciprocal of the fractional area of the region where there is data. This result can be derived by substituting the approximation $B_p \approx (A^2/4\pi)\delta_{p0}$ in eq. (57) and using eq. (19). At whole-sphere coverage, $A = 4\pi$, and we obtain $(\sigma_l^2)^{\text{ML}} = (\sigma_l^2)^{\text{DP}} = 1$, as expected.

9.3 Variance of a multitaper estimate

Fig. 10 shows the variation with harmonic degree l of the eigenvalue-weighted multitaper variance ratio,

$$(\sigma_l^2)^{\text{MT}} = \left(\frac{2l+1}{4\pi}\right) \sum_p^{2L} (2p+1) \Gamma_p \begin{pmatrix} l & p & l \\ 0 & 0 & 0 \end{pmatrix}^2, \quad (178)$$

for single polar and double polar caps of various radii and for two different bandwidths, $L = 10$ and 20 . The lowest variance for any region R and any bandwidth L is that of the monopole or $l = 0$ harmonic, given by any of the three equivalent expressions that are easily derived from

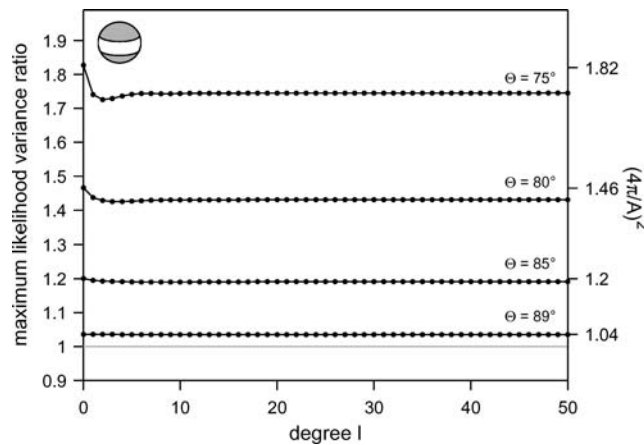


Figure 9. Black dots connected by black lines show the maximum likelihood variance ratio $(\sigma_l^2)^{\text{ML}} = (\sigma_l^2)^{\text{DP}}$ as a function of angular degree $0 \leq l \leq 50$ for double polar caps of radii $\Theta = 89^\circ, 85^\circ, 80^\circ, 75^\circ$. The ratio for a $\Theta = 90^\circ$ double ‘cap’ is obviously unity (grey horizontal line). Labelled tick marks on the right-hand side show the nearly-whole-sphere approximation $(\sigma_l^2)^{\text{ML}} \approx (\sigma_l^2)^{\text{DP}} \approx (4\pi/A)^2$. The slight downward ‘dimple’ between $l = 1$ – 5 for $\Theta = 75^\circ$ is possibly an incipient numerical instability; attempts to invert the matrix $K_{ll'}$ for wider equatorial cuts lead to increasingly unstable results.

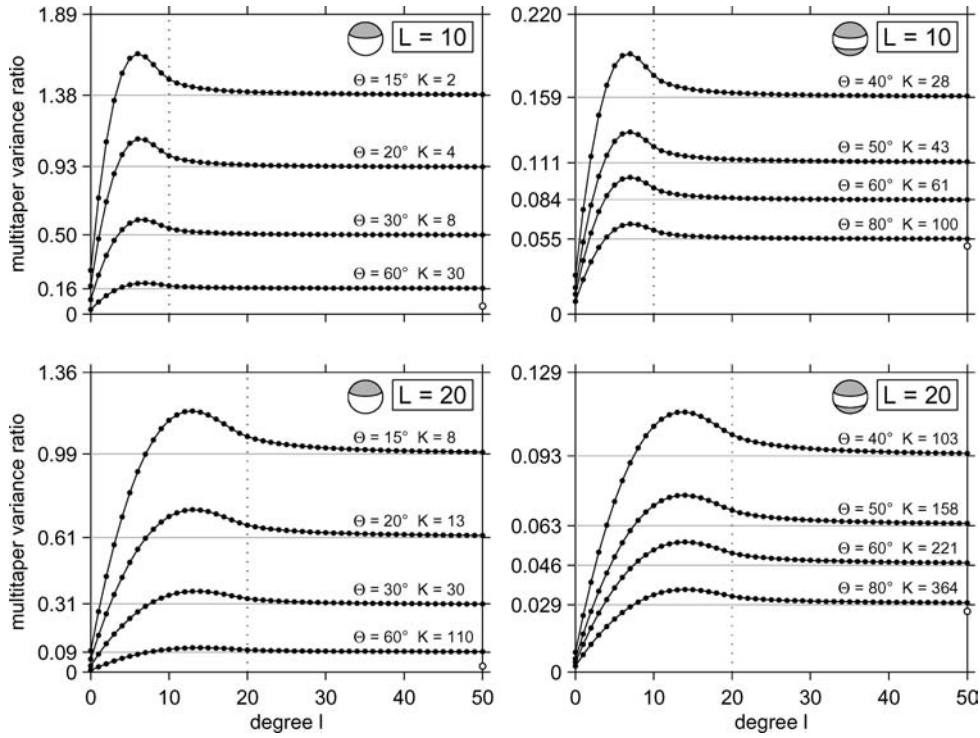


Figure 10. Black dots connected by black lines show the variation of the multitaper variance ratio $(\sigma_l^2)^{\text{MT}}$ with degree $0 \leq l \leq 50$ for single polar caps of radii $\Theta = 15^\circ, 20^\circ, 30^\circ, 60^\circ$ (left-hand side two plots) and double polar caps of common radii $\Theta = 40^\circ, 50^\circ, 60^\circ, 80^\circ$ (right-hand side two plots). Top two plots are for a bandwidth $L = 10$ and bottom two plots are for a bandwidth $L = 20$; the rounded Shannon numbers $K = (A/4\pi)(L+1)^2$ are indicated. Vertical dotted lines at $L = 10$ and $L = 20$ show that above $l = L$ the variance ratio $(\sigma_l^2)^{\text{MT}}$ quickly reaches a large- l asymptotic limit $(\sigma_\infty^2)^{\text{MT}}$, given by eq. (180) and depicted by the grey horizontal lines labeled along the left-hand vertical axis. Open circles on the right-hand vertical axis are the whole-sphere, large- l limits, obtained via eq. (168).

eqs (178) and (167) using eqs (19), (57) and (125):

$$(\sigma_0^2)^{\text{MT}} = \frac{\Gamma_0}{4\pi} = \frac{1}{4\pi K^2} \sum_e (2e+1) B_e \sum_{ss'} (2s+1)(2s'+1) \begin{pmatrix} s & e & s' \\ 0 & 0 & 0 \end{pmatrix}^2 = \frac{1}{K^2} \sum_{st} \sum_{s't'} |D_{st,s't'}|^2 = \frac{1}{K^2} \sum_{\alpha}^{(L+1)^2} \lambda_{\alpha}^2. \quad (179)$$

In the limit of whole-sphere coverage $(\sigma_0^2)^{\text{MT}} = 1/(L+1)^2$, which is easiest to see by noting that in that case, eqs (23) and (8) show that $D_{st,s't'} = \delta_{ss'}\delta_{tt'}$. In the opposite limit of an infinitesimal area, $\Gamma_0^{A \rightarrow 0} = 4\pi$ due to eq. (172), and $(\sigma_0^2)^{\text{MT}} = 1$, the largest possible monopole variance ratio. No matter where it starts, the variance ratio $(\sigma_l^2)^{\text{MT}}$ increases as the target degree l increases, always reaching a maximum at $l \approx 0.65L$ before decreasing equally quickly to an $l \gg L$ asymptotic limit given by

$$(\sigma_\infty^2)^{\text{MT}} = \frac{1}{4\pi} \sum_p^{2L} (2p+1) \Gamma_p [P_p(0)]^2. \quad (180)$$

The whole-sphere limit of eq. (180) is indicated by the four open circles in Fig. 10. Both this and the infinitesimal-area limit, which is off-scale in all four plots, are easily computed by, respectively, substituting $\Gamma_p^{A=4\pi}$ from eq. (168) and $\Gamma_p^{A \rightarrow 0}$ from eq. (172) into eq. (180), thereby avoiding the computation of the Wigner 6- j symbols needed for the more general Γ_p in eq. (167) or the even more cumbersome route through eqs (164) and (166).

Fig. 11 shows the large- l variance ratio $(\sigma_\infty^2)^{\text{MT}}$ plotted versus the bandwidths $0 \leq L \leq 20$ for single polar caps of radii $0^\circ \leq \Theta \leq 180^\circ$ and double polar caps of various radii $0^\circ \leq \Theta \leq 90^\circ$. In the degenerate case $L = 0$, bandlimited ‘multitaper’ estimation is tantamount to whole-sphere estimation so $(\sigma_\infty^2)^{\text{MT}} = 1$ regardless of the ‘cap’ size Θ . Indeed, in that case, the estimate is unbiased, $M_{ll'} = \delta_{ll'}$, and at $L = 0$, the single possible taper of the form eq. (115) is a constant over the entire sphere. For sufficiently large regions ($\Theta \gtrsim 30^\circ$ for a single cap and $\Theta \gtrsim 15^\circ$ for a double cap) the large- l variance ratio is a monotonically decreasing function of the bandwidth L ; for smaller regions the ratio attains a maximum value $(\sigma_\infty^2)^{\text{MT}} > 1$ before decreasing. The grey curves are isolines of fixed Shannon number $K = (A/4\pi)(L+1)^2$; it is noteworthy that the $K = 1$ isoline passes roughly through the maxima of $(\sigma_\infty^2)^{\text{MT}}$, so that for $K \geq 2-3$ the variance ratio is a decreasing function of the bandwidth L regardless of the cap size. Since K is the number of retained tapers, it will always be greater than 2–3 in a realistic multitaper analysis. For large Shannon numbers, above $K \approx 10$, the dependence upon the bandwidth L and area A for both a single or double cap can be approximated by the empirical relation $(\sigma_\infty^2)^{\text{MT}} \approx (4\pi/A)^{0.88}/(2L+1)$. In particular, if $A = 4\pi$, the large- l variance ratio is to a very good approximation equal to one divided by the number of adjacent degrees $l-L \leq l' \leq l+L$ that are averaged over by the coupling matrix $M_{ll'}$. As noted in Section 8.4, a whole-sphere multitaper estimate \hat{S}_l^{MT} can be regarded as a weighted

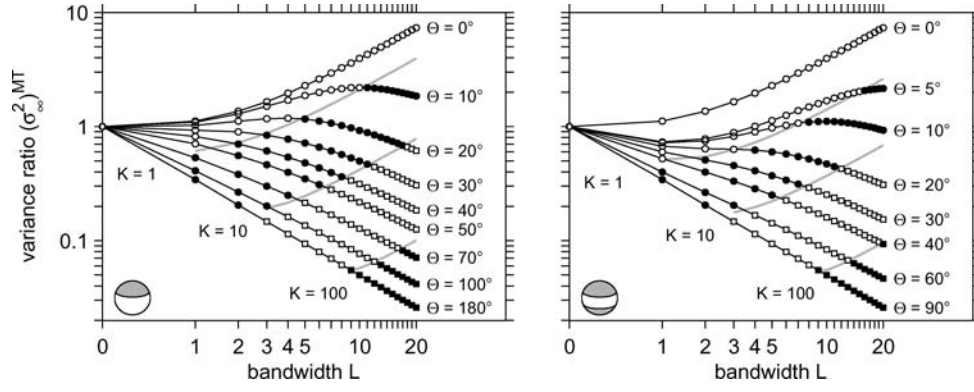


Figure 11. Variation of the large- l multitaper variance ratio $(\sigma_\infty^2)^{\text{MT}}$ with bandwidth $0 \leq L \leq 20$ for single polar caps of radii $\Theta = 0^\circ, 10^\circ, 20^\circ, 30^\circ, 40^\circ, 50^\circ, 70^\circ, 100^\circ, 180^\circ$ (left-hand panel) and double polar caps of common radii $\Theta = 0^\circ, 5^\circ, 10^\circ, 20^\circ, 30^\circ, 40^\circ, 60^\circ, 90^\circ$ (right-hand panel). Ranges of the Shannon number $K = (A/4\pi)(L+1)^2$ are distinguished by different symbols: open circles $0 \leq K < 1$, closed circles $1 \leq K < 10$, open squares $10 \leq K < 100$, closed squares $K \geq 100$. Grey curves labeled $K = 1, 10, 100$ are Shannon number isolines. Axes are logarithmic to illustrate the $1/(2L+1)$ bandwidth scaling above $K \approx 10$.

linear combination of whole-sphere estimates of the form $\sum_l M_{ll'} \hat{S}_{ll'}^{\text{WS}}$, so the variance is reduced by the number of independent random variates $\hat{S}_{l-L}^{\text{WS}}, \dots, \hat{S}_l^{\text{WS}}, \dots, \hat{S}_{l+L}^{\text{WS}}$ that contribute to the estimate. For smaller regions of $A \approx 4\pi$ the whole-sphere variance ratio $1/(2L+1)$ is empirically found to be increased by a factor $(4\pi/A)^{0.88}$. In fact, it is very reasonable to approximate the nearly-whole-sphere variance ratio at large Shannon numbers by $(\sigma_l^2)^{\text{MT}} \approx (4\pi/A)^{0.88} (\sigma_l^2)^{\text{MT}}_{A=4\pi}$ for all spherical harmonic degrees $0 \leq l \leq \infty$.

Finally, it is interesting to compare the large- l variance ratio of a multitaper estimate $(\sigma_\infty^2)^{\text{MT}}$ with that of a spherical periodogram estimate $(\sigma_\infty^2)^{\text{SP}}$, in the case that the coupling to adjacent harmonic degrees l' is roughly the same. Referring to Figs 5 and 7, for example, we see that the widths of the periodogram coupling matrices $K_{ll'}$ for single polar caps of radii $\Theta = 10^\circ, 20^\circ, 30^\circ$ are comparable to the widths of the multitaper coupling matrices $M_{ll'}$ for bandwidths $L = 20, 10, 5$, respectively. In such cases the multitaper variance ratio is always less than the periodogram variance ratio by a factor that is close to the reciprocal of the Shannon number, that is, $(\sigma_\infty^2)^{\text{MT}} \approx K^{-1} (\sigma_\infty^2)^{\text{SP}}$. This empirical approximation is reminiscent of the analogous situation in 1-D (Percival & Walden 1993).

10 RESOLUTION VERSUS VARIANCE: AN EXAMPLE

To illustrate the ease with which a multitaper spectral analyst can control the fundamental trade-off between spectral resolution and variance by altering the bandwidth L or Shannon number $K = (A/4\pi)(L+1)^2$, we consider a specific example in this penultimate section. We choose a cosmological rather than a geophysical example primarily because the CMB temperature spectrum S_l has a readily computable theoretical shape for a specified set of cosmological parameters (Seljak & Zaldarriaga 1996; Zaldarriaga *et al.* 1998; Zaldarriaga & Seljak 2000). Like many geophysical spectra the CMB spectrum is red, varying as $S_l \sim l^{-2}$, with a number of interesting secondary features that one would like to resolve, including acoustic peaks at $l \approx 220, 550, 800$ and higher. To counteract the redness it is conventional in CMB cosmology to plot not S_l but rather the ‘whitened’ spectrum $\mathcal{S}_l = l(l+1)S_l/(2\pi)$, which is shown as the heavy black line in each of the panels of Fig. 12. The theoretical values of \mathcal{S}_l versus harmonic degree $2 \leq l \leq 900$ have been computed for a set of nominal cosmic input parameters, including $\Omega_b = 0.046$, $\Omega_c = 0.224$, $\Omega_\Lambda = 0.730$ and $H_0 = 72 \text{ km s}^{-1} \text{ Mpc}^{-1}$, using the CMBFAST code that is publicly available at <http://lambda.gsfc.nasa.gov>. The monopole term \mathcal{S}_0 , which is a measure of the average CMB temperature $T_0 = 2.725 \text{ K}$ (Mather *et al.* 1999), and the dipole term \mathcal{S}_1 , which is strongly influenced by the proper motion of our galaxy relative to the CMB, are commonly omitted. The slight fluctuations from point to point in the sky about the all-sky mean T_0 are measured in μK so the units of power \mathcal{S}_l are μK^2 . The grey band surrounding the theoretical \mathcal{S}_l -versus- l curve is the standard error $[\text{var}(\hat{\mathcal{S}}_l^{\text{WS}})]^{1/2} = [2/(2l+1)]^{1/2}[\mathcal{S}_l + l(l+1)N_l/(2\pi)]$ of a hypothetical whole-sky spectral estimate $\hat{\mathcal{S}}_l^{\text{WS}} = l(l+1)\hat{S}_l^{\text{WS}}/(2\pi)$. The noise power N_l is assumed to be of the form (37) with pixelization, detector and beamwidth specifications that roughly correspond to those used in the WMAP spacecraft mapping experiment, namely $\Delta\Omega = 4 \times 10^{-6} \text{ sr}$, $\sigma = 100 \mu\text{K pixel}^{-1}$ and $\theta_{\text{fwhm}} = 20 \text{ arcmin}$. The thinning of the grey band at $l \approx 350$ represents the transition between the low-degree region where the uncertainty in a hypothetical whole-sphere WMAP estimate $\hat{\mathcal{S}}_l^{\text{WS}}$ is dominated by cosmic variance and the high-degree region where it is dominated by noise variance. The rapid increase in the whole-sky uncertainty above this transition is due to the exponential increase in the noise power (37) for harmonics that are below the angular resolution of the WMAP antennae. The total uncertainty $[\text{var}(\hat{\mathcal{S}}_l^{\text{WS}})]^{1/2}$ due to both cosmic and noise variance represents the best we can ever do, if we insist upon estimating individual values of the spectrum \mathcal{S}_l , even if we had uncontaminated whole-sky data. The elimination of contaminated data by a sky cut will always increase the variance; the only way to reduce it is to sacrifice spectral resolution. The six panels of Fig. 12 illustrate the effect of making a multitaper estimate of the whitened spectrum \mathcal{S}_l , using tapers of increasing bandwidth $L = 10, 20, 30, 40, 50, 60$. The analysis region in every case is a double polar cap of common radius $\Theta = 80^\circ$, corresponding to an equatorial cut of width 20° , needed to mask the strong foreground contamination from the galactic plane. As we have seen, the bandwidth alone controls the amount of bias deliberately introduced in this way, and not the size or shape of the analysis region—but the latter does influence the variance of the estimate. The open circles

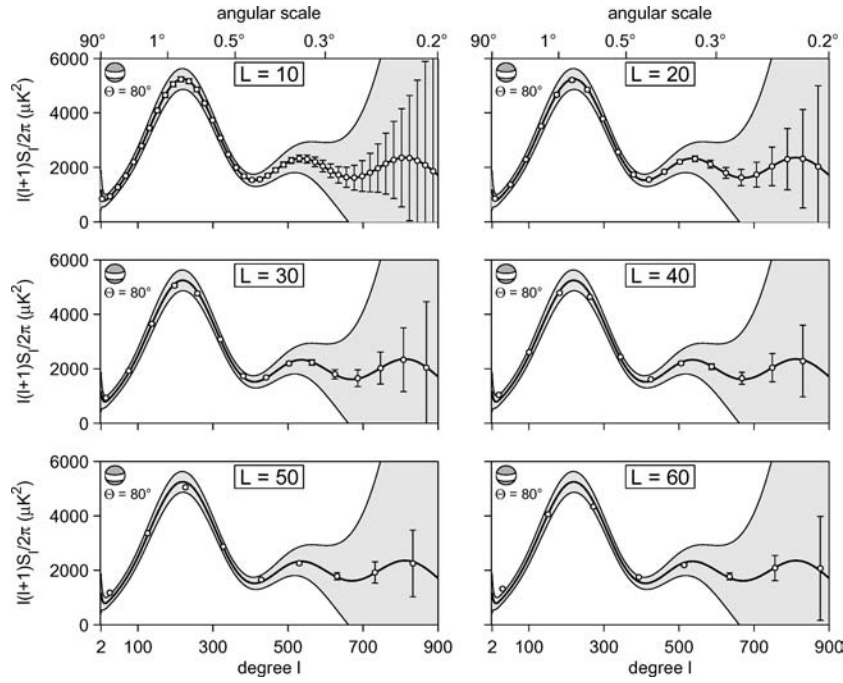


Figure 12. Heavy black lines and surrounding grey band depict the theoretical whitened CMB spectrum $S_l = l(l+1)S_l/(2\pi)$ and hypothetical whole-sphere WMAP estimation error $[2/(2l+1)]^{1/2}[\hat{S}_l + l(l+1)N_l/(2\pi)]$ as a function of angular degree in the range $2 \leq l \leq 900$. Open circles with attached error bars show the expected value and associated standard error $\langle \hat{S}_l^{\text{MT}} \rangle \pm [\text{var}(\hat{S}_l^{\text{MT}})]^{1/2}$ of hypothetical multitaper estimates of the whitened spectrum S_l for various bandwidths, ranging from $L = 10$ (upper left-hand panel) to $L = 60$ (lower right-hand panel). The multitaper analysis region consists of two axisymmetric caps separated by an equatorial cut of width 20° . The angular scale $180^\circ/[l(l+1)]^{1/2}$ of the CMB temperature fluctuations is delineated along the top.

show the expected values of a multitaper estimate $\langle \hat{S}_l^{\text{MT}} \rangle = \sum_{l'} M_{ll'} S_{l'}$, and the accompanying error bars show the associated standard error $[\text{var}(\hat{S}_l^{\text{MT}})]^{1/2}$ under the moderately coloured approximation. The multitaper method yields a band-averaged spectral estimate at every spherical harmonic degree l , but we have only plotted values $\langle \hat{S}_l^{\text{MT}} \rangle \pm [\text{var}(\hat{S}_l^{\text{MT}})]^{1/2}$ whose coupling bands do not overlap, so that they are statistically uncorrelated. The spacing between the open-circle estimates is thus indicative of the spectral resolution. The discrepancy between the open circles and the heavy black S_l -versus- l curve is a measure of the local bias (151) induced by the averaging over adjacent degrees $|l' - l| \leq L$. As expected, the bias $\langle \hat{S}_l^{\text{MT}} \rangle - S_l$ is most pronounced in strongly coloured regions of the spectrum, and it is an increasing function of the bandwidth L and thus the spectral extent of the averaging. For moderate values of the bandwidth, $10 \leq L \leq 40$, the bias is acceptably small in the sense $|\langle \hat{S}_l^{\text{MT}} \rangle - S_l| \ll S_l$; in addition, the spacing between statistically independent estimates $\langle \hat{S}_l^{\text{MT}} \rangle$ and the error bars $\pm [\text{var}(\hat{S}_l^{\text{MT}})]^{1/2}$ are sufficiently small to enable resolution of the first two spectral peaks at $l \approx 220$ and 550 . Bandwidths in this range are, therefore, suitable for multitaper spectral analysis of WMAP temperature data on the cut sky. One can either opt for finer spectral resolution with larger error bars ($L = 10$) or for coarser resolution with somewhat smaller error bars ($L = 40$); to a good approximation the standard error $[\text{var}(\hat{S}_l^{\text{MT}})]^{1/2}$ scales with the bandwidth L as $(2L+1)^{-1/2}$, as we have seen. Because multitaper spectral analysis does not require iteration or large-scale matrix inversion, it is easy to perform analyses for a variety of bandwidths in the range $10 \leq L \leq 40$ and compare the results. In all cases the multitaper errors are significantly smaller than the uncertainty of a hypothetical whole-sky estimate of S_l , with no band averaging. Resolution of the CMB spectral features at higher degrees, above $l \approx 700$, will require a narrowing of the beamwidth θ_{fwhm} and/or a reduction in the instrument noise σ ; motivated by this need and a number of other astrophysical considerations, both ground-based and space-based CMB experiments with narrower-aperture antennae and more sensitive detectors are in advanced stages of development (e.g. Kosowsky 2003; Efsthathiou *et al.* 2005).

11 OVERVIEW AND CONCLUSION

Each of the spectral estimators that we have reviewed or introduced in this paper can be expressed in the general, noise-corrected quadratic form (95), which we repeat here for convenience:

$$\hat{S}_l = \mathbf{d}^T \mathbf{Z}_l \mathbf{d} - \text{tr}(\mathbf{N} \mathbf{Z}_l). \quad (181)$$

We remind the reader that \mathbf{d} is the pixelized data vector (35), \mathbf{C} and \mathbf{N} are the signal and noise covariances, and that the expected value and the covariance of such a quadratic estimator are given by eqs (96) and (97):

$$\langle \hat{S}_l \rangle = \sum_{l'} Z_{ll'} S_{l'}, \quad \text{where} \quad Z_{ll'} = \text{tr}(\mathbf{Z}_l \mathbf{P}_{l'}), \quad \text{and} \quad \Sigma_{ll'} = \text{cov}(\hat{S}_l, \hat{S}_{l'}) = 2 \text{tr}(\mathbf{C} \mathbf{Z}_l \mathbf{C} \mathbf{Z}_{l'}). \quad (182)$$

The specific forms of the symmetric, $J \times J$ pixel-basis matrix \mathbf{Z}_l in the various instances are

whole sphere: $\mathbf{Z}_l = \frac{(\Delta\Omega)^2}{2l+1} \mathbf{P}_l$, where \mathbf{P}_l , given by eq. (28), covers all of Ω ,

spherical periodogram: $\mathbf{Z}_l = \left(\frac{4\pi}{A}\right) \frac{(\Delta\Omega)^2}{2l+1} \mathbf{P}_l$, where \mathbf{P}_l , given by eq. (28), only covers R ,

maximum-likelihood: $\mathbf{Z}_l = \frac{1}{2} \sum_{l'} F_{ll'}^{-1} (\mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1})$, where $F_{ll'} = \frac{1}{2} \text{tr}(\mathbf{C}^{-1} \mathbf{P}_l \mathbf{C}^{-1} \mathbf{P}_{l'})$,

multitaper: $\mathbf{Z}_l = \frac{(\Delta\Omega)^2}{2l+1} \mathbf{G}_l$, where $\mathbf{G}_l = \frac{1}{K} \sum_{\alpha} \lambda_{\alpha} \mathbf{G}_l^{\alpha}$ with \mathbf{G}_l^{α} given by eq. (134). (183)

In writing the final relation in eq. (183) we have assumed that the individual tapers are weighted by the normalized eigenvalues λ_{α} of the spatial concentration problem *sensu* Slepian, eqs (116)–(117). The whole-sphere and maximum-likelihood estimates are unbiased, i.e. $Z_{ll'} = \delta_{ll'}$, whereas the periodogram, with $Z_{ll'} = K_{ll'}$ given by eq. (57), and the eigenvalue-weighted multitaper estimate, with $Z_{ll'} = M_{ll'}$ given by eq. (145), are biased by spectral leakage from neighboring degrees $l' \neq l$. The leakage bias of the periodogram is uncontrollable and can be extensive, particularly for small regions of area $A \ll 4\pi$, rendering the method unsuitable in applications. The extent of the multitaper coupling is in contrast confined to a narrow bandwidth interval $|l' - l| \leq L$ that is specified by the analyst.

The covariance of a whole-sphere estimate is $\Sigma_{ll'}^{\text{WS}} = 2(2l+1)^{-1}(S_l + N_l)^2 \delta_{ll'}$ and the covariance of a maximum-likelihood estimate is the inverse of the Fisher matrix of eq. (87), $\Sigma_{ll'}^{\text{ML}} = F_{ll'}^{-1}$. In the limit of whole-sphere coverage, $A = 4\pi$, the two methods coincide and $\text{var}(\hat{S}_l^{\text{WS}}) = 2(2l+1)^{-1}(S_l + N_l)^2$ is the minimum possible variance achievable for any unbiased spherical spectral estimator. The covariance of a periodogram estimate is given by eq. (68) whereas that of a multitaper estimate is given by eqs (147) and (150). For moderately coloured spectra these cumbersome expressions for $\Sigma_{ll'}^{\text{SP}}$ and $\Sigma_{ll'}^{\text{MT}}$ can be approximated by eqs (156) and (165)–(167), and the Fisher matrix $F_{ll'}$ can be approximated by eq. (160).

The maximum-likelihood method is attractive and has received widespread use in CMB cosmology, because it provides the best (i.e. minimum-variance) unbiased estimate \hat{S}_l^{ML} of the spectrum S_l . This desirable feature is offset by a number of disadvantages that we enumerate in Section 6.8; specifically, it is only feasible without binning for nearly-whole-sphere analyses, $A \approx 4\pi$, and even then it requires a good initial estimate of the spectrum S_l , non-linear iteration to converge to the minimum-variance solution \hat{S}_l^{ML} , and large-scale computation to find the inverse matrices \mathbf{C}^{-1} and $F_{ll'}^{-1}$. For smaller regions, of area $A \not\approx 4\pi$, it is possible to obtain minimum-variance, unbiased estimates \hat{S}_B^{ML} of a binned spectrum $S_B = \sum_l W_{Bl} S_l$ using eqs (107)–(110); however, this requires the somewhat artificial assumption that the true spectrum S_l can be adequately approximated by a coarse-grained spectrum $S_l^{\dagger} = \sum_B W_{lB}^{\dagger} S_B$, where $\sum_l W_{Bl} W_{lB}^{\dagger} = \delta_{BB'}$.

The multitaper method is distinguished by its ease of use, requiring neither iteration nor large-scale matrix inversion. Unlike the unbinned maximum-likelihood method, it yields a smoothed and, therefore, biased estimate of the spectrum, $\langle \hat{S}_l^{\text{MT}} \rangle = \sum_{l'} M_{ll'} S_{l'}$; however, the bias is generally small because it is strictly local, provided that one uses bandlimited rather than spacelimited spherical tapers, and the sacrifice of spectral resolution comes with an auxiliary benefit, namely a reduction by a factor of order $(2L+1)$ in the variance of the smoothed estimate, $\text{var}(\hat{S}_l^{\text{MT}})$. By varying the bandwidth L or the Shannon number $K = (A/4\pi)(L+1)^2$, a multitaper analyst can quickly navigate to any subjectively desirable point on the resolution-versus-variance trade-off curve. The only slight disadvantage of the method is that the shape of the matrix $M_{ll'}$ within the coupling band $|l' - l| \leq L$, and thus the character of the smoothed spectrum $\sum_{l'} M_{ll'} S_{l'}$ that one is estimating, cannot be arbitrarily specified. The coupling matrix $M_{ll'}$ for an eigenvalue-weighted multitaper estimate is illustrated in Figs 6 and 7. In geophysical, geodetic and planetary science applications the objective is generally to obtain a spatially localized estimate of the spectrum S_l of a signal $s(\mathbf{r})$ within a pre-selected region R of area $A \ll 4\pi$. The multitaper method with spatially well-concentrated, bandlimited tapers $g_{\alpha}(\mathbf{r})$ is ideally suited for this purpose, and can be easily extended to estimate cross spectra of two signals such as gravity and topography, enabling admittance and coherence analyses. The spatial leakage from data outside of the target region R can be quelled and the analysis expedited by averaging only the first K tapered estimates \hat{S}_l^{α} , as in eq. (143).

ACKNOWLEDGMENTS

Financial support for this work has been provided by the U.S. National Science Foundation under Grants EAR-0105387 awarded to FAD and EAR-0710860 to FJS, and by a UK Natural Environmental Research Council New Investigator Award (NE/D521449/1) and a Nuffield Foundation Grant for Newly Appointed Lecturers (NAL/01087/G) awarded to FJS at University College London. We thank Mark Wieczorek for a critical first reading of the manuscript, Richard Holme, David Thomson and an anonymous reviewer for constructive comments, and Amir Hajian, Jason McEwen and Sofia Olhede for some last-minute ‘philosophical’ advice. Computer algorithms are made available on www.frederik.net.

REFERENCES

- Albertella, A., Sansò, F. & Sneeuw, N., 1999. Band-limited functions on a bounded spherical domain: the Slepian problem on the sphere, *J. Geodesy*, **73**, 436–447.
- Appourchaux, T., Gizon, L. & Rabello-Soares, M.-C., 1998. The art of fitting p-mode spectra. I. Maximum likelihood estimation, *Astron. Astrophys. Suppl. Ser.*, **132**, 107–119.
- Bendat, J.S. & Piersol, A.G., 2000. *Random Data: Analysis and Measurement Procedures*, 3rd edn, John Wiley, New York.
- Bennett, C.L. *et al.*, 1996. Four-year COBE DMR cosmic microwave background observations: maps and basic results, *Astrophys. J.*, **464**(1), L1–L4.
- Blakely, R.J., 1995. *Potential Theory in Gravity and Magnetic Applications*, Cambridge Univ. Press, New York.
- Bond, J.R., Crittenden, R., Davis, R.L., Efstathiou, G. & Steinhardt, P.J., 1994. Measuring cosmological parameters with cosmic microwave background experiments, *Phys. Rev. Lett.*, **72**, 13–16.
- Bond, J.R., Jaffe, A.H. & Knox, L., 1998. Estimating the power spectrum of the cosmic microwave background, *Phys. Rev. D*, **57**(4), 2117–2137.
- Bronez, T.P., 1988. Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences, *IEEE Trans. Acoust. Speech Signal Process.*, **36**(12), 1862–1873.
- Brussaard, P.J. & Tolhoek, H.A., 1957. Classical limits of Clebsch-Gordan coefficients, Racah coefficients and $D_{mn}^l(\phi, \theta, \psi)$ -functions, *Physica*, **23**(10), 955–971.
- Byerly, W.E., 1893. *An Elementary Treatise on Fourier's Series and Spherical, Cylindrical, and Ellipsoidal Harmonics*, Ginn & Co., Boston, Mass.
- Chave, A.D., Thomson, D.J. & Ander, M.E., 1987. On the robust estimation of power spectra, coherences, and transfer functions, *J. geophys. Res.*, **92**(B1), 633–648.
- Christensen, N., Meyer, R., Knox, L. & Luey, B., 2001. Bayesian methods for cosmological parameter estimation from cosmic microwave background measurements, *Class. Quantum Grav.*, **18**, 2677–2688.
- Chung, M.K., Dalton, K.M., Evans, A.C. & Davidson, R.J., 2008. Tensor-based cortical surface morphometry via weighted spherical harmonic representation, *IEEE Trans. Med. Imag.*, **27**, 8, 1143–1151, doi:10.1109/TMI.2008.918338.
- Cox, D.R. & Hinkley, D.V., 1974. *Theoretical Statistics*, Chapman and Hall, London, UK.
- Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology*, Princeton Univ. Press, Princeton, NJ.
- Edmonds, A.R., 1996. *Angular Momentum in Quantum Mechanics*, Princeton Univ. Press, Princeton, NJ.
- Efstathiou, G., 2004. Myths and truths concerning estimation of power spectra: the case for a hybrid estimator, *Mon. Not. R. Astron. Soc.*, **349**, 603–626.
- Efstathiou, G., Lawrence, C., Tauber, J. & The Planck Science Team, 2005. *Planck: The Scientific Program*, available at <http://www.rssd.esa.int/Planck>.
- Gilbert, E.N. & Slepian, D., 1977. Doubly orthogonal concentrated polynomials, *SIAM J. Math. Anal.*, **8**(2), 290–319.
- Górski, K.M., 1994. On determining the spectrum of primordial inhomogeneity from the COBE DMR sky maps—method, *Astrophys. J.*, **430**(2), L85–L88.
- Górski, K.M., Hivon, E., Banday, A.J., Wandelt, B.D., Hansen, F.K., Reinecke, M. & Bartelman, M., 2005. HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere, *Astrophys. J.*, **622**, 759–771.
- Grishchuk, L.P. & Martin, J., 1997. Best unbiased estimates for the microwave background anisotropies, *Phys. Rev. D*, **56**(4), 1924–1938.
- Grünbaum, F.A., Longhi, L. & Perlstadt, M., 1982. Differential operators commuting with finite convolution integral operators: some non-Abelian examples, *SIAM J. Appl. Math.*, **42**(5), 941–955.
- Gubbins, D., 2004. *Time Series Analysis and Inverse Theory for Geophysicists*, Cambridge Univ. Press, Cambridge, UK.
- Hauser, M.G. & Peebles, P.J.E., 1973. Statistical analysis of catalogs of extragalactic objects. II. The Abell catalog of rich clusters, *Astrophys. J.*, **185**, 757–785.
- Haykin, S., ed., 1991. *Advances in Spectrum Analysis and Array Processing*, Vol. 1, Prentice-Hall, Englewood Cliffs, NJ.
- Hinshaw, G. *et al.*, 2003. First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: the angular power spectrum, *Astrophys. J. Suppl. Ser.*, **148**, 135–159.
- Hivon, E., Górski, K.M., Netterfield, C.B., Crill, B.P., Prunet, S. & Hansen, F., 2002. MASTER of the cosmic microwave background anisotropy power spectrum: a fast method for statistical analysis of large and complex cosmic microwave background data sets, *Astrophys. J.*, **567**, 2–17.
- Hwang, C., 1993. Spectral analysis using orthonormal functions with a case study on sea surface topography, *Geophys. J. Int.*, **115**, 1148–1160.
- Isserlis, L., 1916. On certain probable errors and correlation coefficients of multiple frequency distributions with skew regression, *Biometrika*, **11**(3), 185–190.
- Jackson, J.D., 1962. *Classical Electrodynamics*, John Wiley, New York.
- Jaffe, A.H., Bond, J.R., Ferreira, P.G. & Knox, L.E., 1999. CMB likelihood functions for beginners and experts, in *3 K Cosmology: EC-TMR Conference*, pp. 249–265, eds Maiani, L., Melchiorri, F. & Vittorio, N., Amer. Inst. Phys.
- Jeans, J., 1923. The propagation of earthquake waves, *Proc. R. Soc. Lond., Ser. A*, **102**(718), 554–574.
- Jones, R.H., 1963. Stochastic processes on a sphere, *Ann. Math. Stat.*, **34**(1), 213–218.
- Jungman, G., Kamionowski, M., Kosowsky, A. & Spergel, D.N., 1996. Cosmological-parameter determination with microwave background maps, *Phys. Rev. D*, **54**(2), 1332–1344.
- Kaula, W.M., 1967. Theory of statistical analysis of data distributed over a sphere, *Rev. Geophys.*, **5**(1), 83–107.
- Kendall, M.G. & Stuart, A., 1969. *Advanced Theory of Statistics*, Griffin, London, UK.
- Knox, L., 1995. Determination of inflationary observables by cosmic microwave background anisotropy experiments, *Phys. Rev. D*, **52**(8), 4307–4318.
- Kosowsky, A., 2003. The Atacama Cosmology Telescope, *New Astron. Rev.*, **47**, 939–943.
- Lambeck, K., 1988. *Geophysical Geodesy*, Oxford Univ. Press, New York.
- Lesur, V., 2006. Introducing localized constraints in global geomagnetic field modelling, *Earth Planets Space*, **58**(4), 477–483.
- Libbrecht, K.G., 1985. Practical considerations for the generation of large-order spherical harmonics, *Solar Phys.*, **99**(1–2), 371–373.
- Luscombe, J.J. & Luban, M., 1998. Simplified recursive algorithm for Wigner 3j and 6j symbols, *Phys. Rev. E*, **57**(6), 7274–7277.
- Maniar, H. & Mitra, P.P., 2004. The concentration problem for vector fields, *Int. J. Bioelectromagn.*, **7**(1), 142–145.
- Masters, G. & Richards-Dinger, K., 1998. On the efficient calculation of ordinary and generalized spherical harmonics, *Geophys. J. Int.*, **135**(1), 307–309.
- Mather, J.C., Fixsen, D.J., Shafer, R.A., Mosier, C. & Wilkinson, D.T., 1999. Calibrator design for the COBE Far Infrared Absolute Spectrophotometer (FIRAS), *Astrophys. J.*, **512**, 511–520.
- McKenzie, D.P. & Bowin, C., 1976. The relationship between bathymetry and gravity in the Atlantic Ocean, *J. geophys. Res.*, **81**(11), 1903–1915.
- Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, Vol. 45 of International Geophysics Series, Rev. edn, Academic Press, San Diego, CA.
- Messiah, A., 2000. *Quantum Mechanics*, Dover, New York.
- Mortlock, D.J., Challinor, A.D. & Hobson, M.P., 2002. Analysis of cosmic microwave background data on an incomplete sky, *Mon. Not. R. Astron. Soc.*, **330**, 405–420.
- Mullis, C.T. & Scharf, L.L., 1991. Quadratic estimators of the power spectrum, in *Advances in Spectrum Analysis and Array Processing*, Vol. 1, chap. 1, pp. 1–57, ed. Haykin, S., Prentice-Hall, Englewood Cliffs, NJ.
- Oh, S.P., Spergel, D.N. & Hinshaw, G., 1999. An efficient technique to determine the power spectrum from cosmic microwave background sky maps, *Astrophys. J.*, **510**, 551–563.
- Pail, R., Plank, G. & Schuh, W.-D., 2001. Spatially restricted data distributions on the sphere: the method of orthonormalized functions and applications, *J. Geodesy*, **75**, 44–56.

- Pawitan, Y., 2001. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford Univ. Press, New York.
- Peebles, P.J. E., 1973. Statistical analysis of catalogs of extragalactic objects. I. Theory, *Astrophys. J.*, **185**, 413–440.
- Percival, D.B. & Walden, A.T., 1993. *Spectral Analysis for Physical Applications, Multitaper and Conventional Univariate Techniques*, Cambridge Univ. Press, New York.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P., 1992. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd edn, Cambridge Univ. Press, New York, NY.
- Rasch, J. & Yu, A.C. H., 2003. Efficient storage scheme for precalculated Wigner 3j, 6j, and Gaunt coefficients, *SIAM J. Sci. Comput.*, **25**(4), 1416–1428.
- Schou, J. & Brown, T.M., 1994. Generation of artificial helioseismic time-series, *Astron. Astrophys. Suppl. Ser.*, **107**, 541–550.
- Schulten, K. & Gordon, R.G., 1975. Exact recursive evaluation of 3j-coefficients and 6j-coefficients for quantum-mechanical coupling of angular momenta, *J. Math. Phys.*, **16**(10), 1961–1970.
- Schuster, A., 1898. An investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomena, *Terr. Magn.*, **3**, 13–41.
- Seljak, U. & Zaldarriaga, M., 1996. A line-of-sight integration approach to cosmic microwave background anisotropies, *Astrophys. J.*, **469**(2), 437–444.
- Simons, F.J. & Dahlen, F.A., 2006. Spherical Slepian functions and the polar gap in geodesy, *Geophys. J. Int.*, **166**, 1039–1061, doi:10.1111/j.1365–246X.2006.03065.x.
- Simons, F.J. & Dahlen, F.A., 2007. A spatio-spectral localization approach to estimating potential fields on the surface of a sphere from noisy, incomplete data taken at satellite altitudes, in *Wavelets XII*, Vol. 6701, pp. 670117, eds Van de Ville, D., Goyal, V.K. & Papadakis, M., doi:10.1117/12.732406, Proc. SPIE.
- Simons, F.J., Dahlen, F.A. & Wicczorek, M.A., 2006. Spatio-spectral concentration on a sphere, *SIAM Rev.*, **48**(3), 504–536, doi:10.1137/S0036144504445765.
- Simons, M., Solomon, S.C. & Hager, B.H., 1997. Localization of gravity and topography: constraints on the tectonics and mantle dynamics of Venus, *Geophys. J. Int.*, **131**, 24–44.
- Slepian, D., 1983. Some comments on Fourier analysis, uncertainty and modeling, *SIAM Rev.*, **25**(3), 379–393.
- Sneeuw, N., 1994. Global spherical harmonic-analysis by least-squares and numerical quadrature methods in historical perspective, *Geophys. J. Int.*, **118**(3), 707–716.
- Stoica, P. & Sundin, T., 1999. On nonparametric spectral estimation, *Circ., Syst. Signal Process.*, **18**(2), 169–181, doi:10.1007/BF01206681.
- Strang, G., 1986. *Introduction to Applied Mathematics*, Wellesley-Cambridge, Wellesley, MA.
- Strang, G., 1998. *Linear Algebra and its Applications*, 3rd edn, Harcourt Brace Jovanovich, San Diego, CA.
- Swenson, S. & Wahr, J.M., 2002. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity, *J. geophys. Res.*, **107**(B9), 2193, doi:10.1029/2001JB000576.
- Tegmark, M., 1995. A method for extracting maximum resolution power spectra from galaxy surveys, *Astrophys. J.*, **455**, 429–438.
- Tegmark, M., 1996. A method for extracting maximum resolution power spectra from microwave sky maps, *Mon. Not. R. Astron. Soc.*, **280**, 299–308.
- Tegmark, M., 1997. How to measure CMB power spectra without losing information, *Phys. Rev. D*, **55**(10), 5895–5907.
- Tegmark, M., Taylor, A.N. & Heavens, A.F., 1997. Karhunen-Loève eigenvalue problems in cosmology: how should we tackle large data sets? *Astrophys. J.*, **480**(1), 22–35.
- Thébault, E., Schott, J.J. & Manda, M., 2006. Revised spherical cap harmonic analysis (R-SCHA): validation and properties, *J. geophys. Res.*, **111**(B1), B01102, doi:10.1029/2005JB003836.
- Thomson, D.J., 1982. Spectrum estimation and harmonic analysis, *Proc. IEEE*, **70**(9), 1055–1096.
- Thomson, D.J., 1990. Quadratic-inverse spectrum estimates: applications to paleoclimatology, *Phil. Trans. R. Soc. Lond., Ser. A*, **332**(1627), 539–597.
- Thomson, D.J., 2007. Jackknifing multitaper spectrum estimates, *IEEE Signal Process. Mag.*, **20**, 20–30, doi:10.1109/MSP.2007.4286561.
- Thomson, D.J. & Chave, A.D., 1991. Jackknifed error estimates for spectra, coherences, and transfer functions, in *Advances in Spectrum Analysis and Array Processing*, Vol. 1, chap. 2, pp. 58–113, ed. Haykin, S., Prentice-Hall, Englewood Cliffs, NJ.
- Turcotte, D.L., Willemann, R.J., Haxby, W.F. & Norberry, J., 1981. Role of membrane stresses in the support of planetary topography, *J. geophys. Res.*, **86**, 3951–3959.
- Varshalovich, D.A., Moskalev, A.N. & Khersonskii, V.K., 1988. *Quantum Theory of Angular Momentum*, World Scientific, Singapore.
- Wandelt, B.D., Hivon, E. & Górski, K.M., 2001. Cosmic microwave background anisotropy power spectrum statistics for high precision cosmology, *Phys. Rev. D*, **64**(8), 083003.
- Wicczorek, M.A., 2007. The gravity and topography of the terrestrial planets, in *Treatise on Geophysics*, Vol. 10, pp. 165–206, ed. Spohn, T., doi:10.1016/B978-044452748-6/00156-5, Elsevier, Amsterdam.
- Wicczorek, M.A. & Simons, F.J., 2005. Localized spectral analysis on the sphere, *Geophys. J. Int.*, **162**(3), 655–675, doi:10.1111/j.1365–246X.2005.02687.x.
- Wicczorek, M.A. & Simons, F.J., 2007. Minimum-variance spectral analysis on the sphere, *J. Fourier Anal. Appl.*, **13**(6), 665–692, doi:10.1007/s00041-006-6904-1.
- Wingham, D.J., 1992. The reconstruction of a band-limited function and its Fourier transform from a finite number of samples at arbitrary locations by singular value decomposition, *IEEE Trans. Signal Process.*, **40**(3), 559–570, doi:10.1109/78.120799.
- Xu, P., 1992a. Determination of surface gravity anomalies using gradiometric observables, *Geophys. J. Int.*, **110**, 321–332.
- Xu, P., 1992b. The value of minimum norm estimation of geopotential fields, *Geophys. J. Int.*, **111**, 170–178.
- Xu, P., 1998. Truncated SVD methods for discrete linear ill-posed problems, *Geophys. J. Int.*, **135**(2), 505–541.
- Zaldarriaga, M. & Seljak, U., 2000. CMBFAST for spatially closed universes, *Astrophys. J.*, **129**(2), 431–434.
- Zaldarriaga, M., Seljak, U. & Bertschinger, E., 1998. Integral solution for the microwave background anisotropies in nonflat universes, *Astrophys. J.*, **494**(2), 491–502.