Mon. Not. R. Astron. Soc. **336,** 979–991 (2002)

Peak finding at low signal-to-noise ratio: low-ℓ solar acoustic eigenmodes at $n \leqslant 9$ from the analysis of BiSON data

W. J. Chaplin,[1]★ Y. Elsworth,[1] G. R. Isaak,[1] K. I. Marchenkov,[1] B. A. Miller,[1] R. New,[2] B. Pinter[2] and T. Appourchaux[3]

[1]*School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT*
[2]*School of Science & Mathematics, Sheffield Hallam University, Sheffield S1 1WB*
[3]*ESA Research and Science Support Department, ESTEC, 2200 AG Noordwijk, the Netherlands*

Accepted 2002 July 3. Received 2002 June 25; in original form 2002 March 29

ABSTRACT

We make use of 9 yr of full-disc helioseismic data – as collected by the ground-based Birmingham Solar-Oscillations Network (BiSON) – to search for low-frequency, low-angular-degree (low-ℓ) acoustic modes. A range of tests are applied to the power spectrum of the observations that search for prominent mode-like structure: strong spikes, structure spanning several bins signifying the presence of width (from damping), and the occurrence of prominent multiplet structure at $\ell \geqslant 1$ arising principally from the solar rotation and made from several spikes separated suitably in frequency. For each test we present analytical expressions that allow the probability that the uncovered structure is part of the broad-band noise background to be assessed. These make use of the cumulative binomial (Bernoulli) distribution and serve to provide an objective measure of the significance of the detections. This work has to date uncovered nine significant detections of non-broad-band origin that we have identified as low-ℓ modes with radial overtone numbers $n \leqslant 9$.

Key words: methods: data analysis – methods: statistical – Sun: interior – Sun: oscillations.

1 INTRODUCTION

The properties of low-angular-degree (low-ℓ) solar acoustic (p) modes are affected by conditions in the deep radiative interior of the Sun. As such, the frequencies (and splittings) of these modes act as a probe of the core regions. Their diagnostic potential is enhanced at low frequencies by the increase in mode lifetime that accompanies any decrease in radial order. In accord with the Uncertainty Principle this implies that the frequencies of such modes can be determined to very high precision and accuracy (Chaplin et al. 2002a), with consequent benefits for tests of both the hydrostatic and dynamic interior structure. The cavity properties of the low-frequency modes do, however, contrive to give only very weak oscillatory signals in the photosphere where observational techniques detect the periodic Doppler or intensity variations that are the visible manifestation of the resonant modes. This reduction in the potential signal-to-noise ratio (S/N) at which successively lower frequency modes can be observed arises largely from the increase in depth (beneath the photosphere) at which the tops of their acoustic cavities are located. The observed photospheric signal – which is evanescent – is therefore increasingly attenuated.

Gravity (g) modes can be sustained only in regions that are stable against convection. While this means that those confined in the radiative interior would provide a more sensitive probe of core conditions than the p modes, a substantial attenuation of the signal is observed (as above) but of a magnitude that is greater since the tops of the g-mode cavities are to be found much deeper at the base of the convection zone.

Against this backdrop, helioseismologists have been forced to adopt ever-more-careful methods of analysis in attempts to uncover low-frequency p and g modes. It is now 21 years since the first table of low-ℓ eigenfrequency estimates was published (Claverie et al. 1981) with values given down to $\sim$2400 μHz. The low-frequency threshold at which low-ℓ modes can be unambiguously detected – and their frequencies clearly determined – is currently $\sim$1000 μHz and possibly lower (Bertello et al. 2000; Garcia et al. 2001). Claimed detections at $\nu < 972$ μHz await independent confirmation and are therefore uncertain. An unambiguous, universally accepted g-mode detection remains elusive (e.g. Appourchaux et al. 2000, 2001; Gabriel et al. 2002). Observations at higher ℓ have provided access to both p and f (surface) modes at frequencies well below 1000 μHz. Schou (1998) demonstrates that by making use of data collected by the MDI instrument on board the *ESA/NASA* SOHO satellite some modes can be seen at frequencies as low as $\sim$600 μHz ($\ell \sim 20$). Here, we concentrate our efforts on an attempt to uncover low-ℓ p modes and leave a search for low-ℓ g

★E-mail: wjc@bison.ph.bham.ac.uk

modes – which demand slightly different search criteria in the very low frequency g- and mixed-mode regime where $\nu < 350$ μHz – to another paper. To effect this study we have made use of some 9 yr of data collected by the ground-based Birmingham Solar-Oscillations Network (BiSON). The eigenfrequencies of the best available solar models have been used to guide the search.

In what follows we describe first in Section 2 how the raw data were prepared in order to reduce contributions from unwanted artefacts at low frequencies. We then describe in Section 2 the basic attributes of the power spectrum of the data (in a statistical context), introducing those ideas that are developed in Section 3 into several mode-detection tests with associated formal test probabilities that are fixed by analytical expressions. The tests are applied to search for: (i) prominent spikes (power in one bin only); (ii) the occurrence of prominent multiplet structure at $\ell \geqslant 1$ arising principally from the solar rotation and suitably separated in frequency; (iii) the appearance of two or more prominent spikes distributed over a range of spectrum bins corresponding to two estimated modal linewidths; and (iv) the appearance of two or more prominent spikes in consecutive frequency bins. For each test we present analytical expressions that allow the significance of structure to be assessed against the probability that it is the result of broad-band background noise. These make use of the cumulative binomial (Bernoulli) distribution.

In Section 5 we discuss the manner in which the tests are applied to the BiSON data, in particular how the chances of uncovering modes can be increased by analysing also spectra made by removing different (small) numbers of data from the original series.

We list the detection probabilities and frequencies of uncovered mode-like structure in Section 6 and discuss at length in Section 7 the plausibility of the nine claimed detections, where our results are also compared with those of other observers.

2 DATA PREPARATION

We have used Doppler velocity observations of the visible solar disc made by BiSON over the period 1992 January 1 through 2000 December 31. This extensive epoch spans the declining phase of solar activity cycle 22, and the rising phase of cycle 23. The data collected at each of the six active observatories were first processed on a day-by-day basis in the manner described by Elsworth et al. (1995) to yield daily calibrated velocity residuals. The resulting $\approx 2 \times 10^4$ individual daily sets were then subjected to two further stages of preparatory analysis in order to optimize our sensitivity to the weak, low-frequency p modes.

(i) First, a suitable model was fitted to the daily data in an attempt to remove low-frequency artefacts (or 'footprints') of instrumental origin. This is discussed in Appendix A.

(ii) A quantitative assessment was then made of the quality of the footprint-corrected data over the frequency range of interest, with those data that failed to meet objectively set criteria being discarded prior to the construction of the final, multisite time series (as described in Appendix B). In essence, this involved performing a multiparameter maximum-likelihood minimization over the complete daily residual set to find power thresholds – for a chosen frequency range – which if exceeded on a given day would demand the data be rejected. We found optimized thresholds over two ranges: $200 \leqslant \nu \leqslant 800$ μHz and $800 \leqslant \nu \leqslant 1300$ μHz. From this stage of processing we then collated several sets of daily residuals (from all sites), these being optimized for each range, made with thresholds set to both the maximum-likelihood values and levels slightly above this.

Each of the sets were then combined coherently (Chaplin et al. 1997) to yield multiyear time series ready for subsequent analysis in the frequency domain.

3 THE POWER SPECTRUM

3.1 Characteristics of the p-mode spectrum

Over the well-studied part of the p-mode spectrum the modes are manifested in frequency space as prominent peaks at high signal-to-noise (S/N) ratios; they possess width in frequency that varies in a well defined manner with both radial order and angular degree; they appear at discrete, well-defined locations and exhibit a regular structure in frequency both within each multiplet (i.e. across the azimuthal order m) and across many modes. These are attributes that allow the peaks to be identified as resonant modes and they must form the foundation of any mode-finding strategy at low frequency. However, the very nature of the observed p-mode spectrum denies us the first of these attributes at low radial order, i.e. a high S/N ratio. Furthermore, as the modal width decreases with decreasing frequency this restricts the range of the spectrum at each radial order over which the signature of the p modes is expected to be found. Any coherent strategy designed for low frequencies must accommodate these changing characteristics. Under low S/N conditions – where the power spectrum is dominated by broad-band noise – a proper assessment of the probability of occurrence of structure that might pass for a mode is essential and this depends upon the manner in which power is distributed in the frequency domain.

3.2 Power distribution in the frequency domain

If the background arises from a 'white' noise source that is distributed in a normal manner in the time domain its power spectrum will follow a negative exponential distribution (χ^2, 2 d.o.f.) in the frequency domain. While individual sources that contribute to the noise background may not be strictly Gaussian in character in the time domain, if several do so then the central limit theorem suggests that the resulting combination will tend to a normal distribution. Above the lowest-frequency part of the BiSON spectrum at which diurnal sideband contamination is present from the window function of the observations this is indeed found to be the case, i.e. the observed power follows closely a negative exponential distribution. With regard to the modes themselves, the power across the resonant peaks is distributed also in a negative exponential manner, this on account of the stochastic nature of the excitation mechanism.

In what follows, we use the terms 'spike' and 'peak' when discussing the appearance of power in the spectrum. We regard a spike as being confined to one bin only; a peak is made of several spikes and as such covers more than one bin. So, consider the appearance of a spike with power ξ_ν at frequency ν in the spectrum, where the mean level is $\langle \xi \rangle_\nu$. In a region dominated by broad-band 'white' noise, $\langle \xi \rangle_\nu$ will correspond to the mean level of the resulting flat spectrum. Let the relative height of any given spike be

$$s_\nu = \xi_\nu / \langle \xi \rangle_\nu. \quad (1)$$

The probability of observing, in any given bin, a relative height greater than or equal to s_ν – under the assumption that all bins are independent in the frequency domain – is then

$$p(s_\nu) = \exp(-s_\nu). \quad (2)$$

To determine the probability of observing at least r such spikes across a range of N bins in the spectrum we may make use of the

cumulative binomial (Bernoulli) distribution. The required formalism is

$$P[r; p(s_\nu), N] = \sum_{r=r}^{N} p(s_\nu)^r [1 - p(s_\nu)]^{N-r} \frac{N!}{r!(N-r)!}. \quad (3)$$

If the N bins cover a range Δ_p in frequency, and the spectrum has been made from a time series of length T then

$$N = \Delta_p \cdot T. \quad (4)$$

Owing to the finite length of a real window, and the presence of gap structure in the BiSON time string, all bins in the frequency domain are not strictly independent and leakage of power into bins adjacent to the mode (if present) will occur. Implicit in the use of N in equation (3) is the assumption that all bins are, in fact, independent. We have used extensive Monte Carlo simulations to assess the extent to which the window function of the BiSON observations modifies the statistical distribution of power through correlations between bins. We find that for the BiSON window we can still adequately adopt the assumption of statistical independence.

3.3 Frequency range to be searched

3.3.1 Extreme limits

A preferred lower frequency limit to the range to be searched for p modes would of course be that of the fundamental radial mode (i.e. about ∼250 μHz). However, we have in practice imposed a somewhat conservative limit of ∼400 μHz. At frequencies above this the harmonics of the diurnal frequency that are present on account of the BiSON window function are very small in amplitude; furthermore, one is also above the *g*-mode and mixed-mode regime where different characteristics (e.g. rotational splitting) would demand an altered set of test parameters. Our chosen lower limit aims to ensure, therefore, that the spectral range to be searched contains no sharp, narrow-band structure other than the sought-for p modes. As such we concentrate here upon the development of a suite of tests designed to assess the likelihood of p-mode-like structure emerging by chance from the noise background; we intend to deal with the more complicated regime below 400 μHz in a future paper following developments to the analyses described here.

At frequencies above about ∼1600 μHz the modes are prominent and easily detectable. This leaves us therefore with a range $400 \leqslant \nu \leqslant 1600$ μHz to be searched, within which there are 33 multiplets covering $0 \leqslant \ell \leqslant 3$. The spectrum here is dominated by a broad-band noise background and our aim is then one of testing whether prominent spikes are likely to be part of this background and as such possible modes if this is found not to be the case.

3.3.2 Targeting the expected locations of the modes

The p modes are to be found at discrete frequencies and as a result there are large stretches of the 400- to 1600-μHz range where we would not expect to find the sought-for acoustic signatures. In practical terms this means that the search can be targeted on those 'live' ranges of the spectrum that bracket the expected locations of the modes. The extent to which the search might be narrowed at each location depends, of course, upon the accuracy of the predicted frequencies and nature of the structure one hopes to find. Here one may be guided by a suitable smooth extrapolation of the observed mode frequencies from the well-studied part of the spectrum or comparison with theoretical predictions of the eigenfrequencies.

For the first method one can envisage performing an extrapolation of either the observed frequencies as a function of radial order or their differences (first, second and possibly higher order, depending upon the frequency uncertainties). In principle this purely empirical approach should yield results that are independent of theory. However, in practice any smooth trend present will have imposed upon it a quasi-oscillatory signal that is the signature of discontinuities in the sound speed within the Sun. The most prominent of these arises from the He II ionization zone (with a smaller contribution from the base of the convection zone) and to allow adequately for it requires some theoretical guidance regarding its basic form. With no such allowance made, any prediction is accurate only to ≈1 μHz.

We are currently working upon two approaches to improve this accuracy. In the first one aims to extract the He II-dominated signal from the observed frequencies, fit it to a suitable function, e.g. that proposed by Monteiro & Thompson (1998) and then use the fit and that made to the smooth, underlying trend to predict the uncovered frequencies. In the second, one seeks to use an eigenfrequency equation of the form developed by Roxburgh & Vorontsov (2000) but where maximum-likelihood estimates of its various parameters are first obtained by fitting to the observed frequencies. The best-fitting function is then extrapolated accordingly.

In this paper we make use predominantly of model eigenfrequencies to guide our search. The magnitude of the difference between observed and model frequencies decreases from ∼10 μHz at high frequencies to typically less than ∼1 μHz below ≈2000 μHz. However, it is the smoothness of the trend in the difference – rather than the absolute difference itself – that is of most relevance to us here. This is because we would hope to use an extrapolation of the observed differences to predict the uncovered eigenfrequencies; any departures from a smooth trend will lead to a reduction in accuracy.

By studying the difference between the observed frequencies and those of three solar models – model 'S' of Christensen-Dalsgaard et al. (1996); model 'M1' of Provost, Berthomieu & Morel (2000); and the seismic Saclay model applied in Turck-Chiéze et al. (2001) (described more fully in Couvidat, Turck-Chiéze & Kosovichev, in preparation) – it is clear that rms deviations from a smooth trend are of a magnitude that is ≈0.1 μHz below 2000 μHz. We shall return to this point below in Section 7 (see also Fig. 3 later). Here, we assume therefore that we could in principle locate the central frequency of a given mode to within 0.4 μHz, i.e. within the magnitude of the rotational splitting between adjacent modal components; see Section 4.2). This allows us potentially to label the m components of a partially uncovered multiplet.

The frequency of structure uncovered by tests that search for mode-like signatures (e.g. a peak with width over several bins, or a rotational splitting pattern) can then be compared with the predicted location to help judge whether it is indeed a mode. In particular, the predicted locations can assist in 'bootstrapping' to successively lower frequencies; if a detection has been made with a sufficient degree of confidence the additional frequency difference can be 'added' to those from higher frequencies and a new, more accurate, extrapolated prediction made.

Here, we divided our search into two sets of ranges. First, a 'mode target' set centred on the predicted locations of each multiplet (see above). The outer boundary of each was fixed to lie some ±0.4 μHz beyond the predicted location of the extremities of the multiplet. For the radial ($\ell = 0$) targets this covered ±0.4 μHz around the expected frequency, giving a test bandwidth of 0.8 μHz. We assumed that at $\ell > 0$ the sought-for adjacent m were separated by ∼0.4 μHz (see Section 4.2 below), so that other search ranges were fixed at: 1.6 μHz at $\ell = 1$; 2.4 μHz at $\ell = 2$; and 3.2 μHz at $\ell = 3$. Secondly,

we also tested those ranges outside the above in order to ascertain properly the occurrence of any other structure of non-broad-band origin.

4 STATISTICAL MODE DETECTION TESTS: WHEN IS A SPIKE OR PEAK JUST NOISE?

4.1 Detection on account of the prominence of a spike

First, we consider the most straightforward criterion for detection – that of the appearance of a prominent spike in the spectrum. This basic test requires that we search the spectrum for spikes that lie significantly above the local background. Over the range of frequencies to be searched here, the background is dominated by broad-band noise. The mean level does increase at lower frequencies; however, this is gradual (giving a 'pink' spectrum) and locally the spectrum can be regarded as being flat. The mean level of this, $\langle \xi \rangle_\nu$, is estimated by computing a running average over a suitable (restricted) window of width $\sim$70 μHz.

Formally, the probability of observing at least one spike of relative height s_ν or greater (i.e. relative to the mean background level) across a certain range of N bins of the spectrum is (cf. equation 3)

$$P_s(s_\nu) = P[1; p(s_\nu), N]. \quad (5)$$

This serves as a measure of the significance of a candidate detection, i.e. the lower the value the less likely it is to be part of the noise. Whether or not a prominent spike might be considered as a possible p mode will then depend as such not only upon the value of $P_s(s_\nu)$ but also the location and characteristics of the spike, i.e. whether or not it is a 'plausible' candidate.

For a spike to be considered as the potential signature of a mode on account of its strength alone it must be prominent enough to return a low $P_s(s_\nu)$. A value of $P_s(s_\nu) \leqslant 0.1$ is usually taken as being sufficient to merit consideration (Appourchaux et al. 2000). The relative height threshold required to achieve this follows straightforwardly in the limit where $p(s_\nu) = \exp(-s_\nu)$ is small. Under these circumstances equation (5) is well approximated by (cf. equations 3 and 4)

$$P_s(s_\nu) \approx N p(s_\nu) \approx \Delta_p T \exp(-s_\nu).$$

Re-arrangement gives

$$s_\nu \approx \log_e \Delta_p T - \log_e P_s(s_\nu). \quad (6)$$

So, the value required to give, say, $P_s(s_\nu) \leqslant 0.1$ (i.e. $s_\nu^{0.1}$) will depend upon the number of bins N, and therefore the range Δ_p, over which the probability of the spike appearing has been determined. This need not necessarily be the same range as that used to determine the mean background $\langle \xi \rangle_\nu$ (see above), provided that locally the mean is well determined and that the distribution of the background remains the same at all frequencies (i.e. following χ^2 2-d.o.f. statistics).

The smaller N – and, therefore, the corresponding range in frequency, Δ_p – the lower, and less demanding, the required 10 per cent threshold. We indicated in Section 3.3.2 that the expected locations of the modes can probably be pinned down to very narrow ranges, a fact reflected by our choice to concentrate our search on confined bands of the spectrum. This might suggest therefore that an equivalent range Δ_p of at most a few μHz (i.e. a small N) should be used to calculate $s_\nu^{0.1}$ (equation 6) and the test statistic of a given spike, $P_s(s_\nu)$ (equation 5). At face value this seems a reasonable strategy, i.e. to take a Δ_p that corresponds to each mode range that is searched. However, here we argue that while one can probably predict the frequencies of the modes to a similar accuracy, adopting the same small range N in the probability calculation would increase to an unacceptable level the risk of making a misidentification (i.e. taking a noise spike to be a mode).

To illustrate why, take the case of us having fixed Δ_p at 0.8 μHz. This corresponds to the search band applied over $\ell = 0$ target ranges. When used in equation (6) the 10-per cent probability threshold would then be fixed at $s_\nu^{0.1} \sim 7.7$. Within *each* $\ell = 0$ window there would, by definition, be only a 10 per cent probability of a spike appearing by chance at or above this level. However, if we were to search all *all* 33 mode windows ($0 \leqslant \ell \leqslant 3$) – which cover a total of $\sim$67 μHz in frequency (summation of the search ranges given at the end of Section 3.3.2) – we would expect to find (on average) a total of $\sim$8 spikes above this level on account of the background alone (with a near 100 per cent probability of seeing at least one of these rogue spikes somewhere in the 33 windows). In order to reduce the risk of a misidentification the need for a larger Δ_p is strongly indicated.

We have opted for $\Delta_p = 70$ μHz (the approximate mean spacing between adjacent low-ℓ pairs), a value chosen also by Appourchaux et al. (2000). Coincidentally, this matches closely the full target window of $\sim$67 μHz. If the full 3276-d BiSON time series is analysed in one coherent piece the new Δ_p covers $N = 19818$ bins and the relative height required to give the threshold probability is $s_\nu^{0.1} \sim 12.2$. To follow on from above, there would now be only a $\sim$0.1 per cent chance of finding a spike from the broad-band noise at or above this level in each individual 0.8-μHz wide $\ell = 0$ window, with only a $\sim$10 per cent chance of a spike appearing somewhere in all the mode windows.

The bottom line is that while in practice one will choose to home in on and test those narrow ranges that are expected to contain the modes, with additional guidance from theory or extrapolation used to help judge any mode candidate, such a narrow range should not be used also for the Δ_p needed to calculate the test probability.

4.2 Detection on account of an ordered multiplet structure

At $\ell > 0$ the modes are split into $\ell + 1$[1] observable components in unimaged, full-disc helioseismic data and the ordered structure that results can be searched for as another means of uncovering candidate modes. Any test for the presence of a multiplet demands that the spectrum be scrutinized for appropriate, prominent structure. In order to derive a quantitative test probability – again through the use of the binomial distribution – the nature of this structure must be clearly defined.

The predominant contribution to the splitting arises from the rotation of the solar interior and this lifts the frequency degeneracy in ℓ to give a symmetric pattern where the observed (synodic) separation between adjacent m (i.e. for $|\Delta m| = 1$) is $\sim$400 nHz. At mode frequencies above $\sim$400 μHz theoretical modelling (Provost et al. 2000) indicates that the magnitude of the frequency splitting within p-mode multiplets is very insensitive to the poorly understood rotation at $r/R_\odot < 0.2$. As such, we assume that the synodic splittings of the sought-for modes will be similar to those already measured at higher frequencies, i.e. $\sim$400 nHz between adjacent m (Chaplin et al. 2001).

Magnetic fields can also affect the modes in such a way as to introduce asymmetries in the frequency splittings within each multiplet. In full-disc data this effect may become apparent when $\ell \geqslant 2$. If the field is of near-surface origin the magnitude of the expected

[1] Only $\ell + 1$, rather than the full $2\ell + 1$, components are seen because of our near-equatorial view of the Sun.

$\ell = 2$ and 3 asymmetry should be very small at low frequencies (of the order of a few nHz). However, a deep-seated field would impose an asymmetry, the magnitude of which would vary in a quasi-oscillatory manner with frequency. Chaplin et al. (2001) found small asymmetries at low frequencies from fitting an 8-yr BiSON power spectrum. More recent work (Chaplin et al. 2002b) appears to reinforce these results. While it may be that the asymmetries of the undetected modes are also negligible in extent we nevertheless allow for their possible presence here in our multiplet 'description' for low radial orders.

Our definition of a multiplet is therefore as follows: if a prominent spike is present in the spectrum other spikes must lie at multiples of 400 nHz from it, with an allowed uncertainty on their location of $\delta\nu = \pm 100$ nHz. This uncertainty allows for possible variations in the splitting with mode frequency and any asymmetry in the structure. In what follows we assume that the uncertainty $\delta\nu = 100$ nHz covers N_δ bins in the spectrum (as determined by $N_\delta = \delta\nu T$). In the regime to be searched this is always larger than the expected mode width, $\Delta\nu$. We do not make use of the potential width information in the test, although it could of course be developed to include this. We do, however, in the tests outlined in Section 4.3 below.

Here, the spectrum is searched for prominent multiplet structure that satisfies the definition above, subject to the requirement that all components have relative heights that exceed some threshold, s_ν^m, the magnitude of which is chosen so as to ensure that the probability of the structure appearing by chance (i.e. as part of the background) is very small. To fix the probabilities for a given threshold – that will vary depending upon the value of ℓ – our task is then one of calculating the likelihood of observing multiplet structure by chance.

To proceed we must first calculate the probability of observing one particular pair, triplet or quadruplet arrangement of spikes that lies within the constraints imposed by the chosen N_δ. This corresponds to determining the probability of observing at least one prominent spike above s_ν^m in any bin and at least one other such prominent spike in either one, two or three ranges each of width ± 100 nHz. The probability of observing a spike above the threshold s_ν^m in any bin is just $\exp(-s_\nu^m)$; that of finding at least one spike by chance above s_ν^m over the full allowed uncertainty, that is $2N_\delta$, is therefore

$$P_\delta = P\left[1; \exp\left(-s_\nu^m\right), 2N_\delta\right]. \tag{7}$$

The probability of observing k spikes above s_ν^m which satisfy the requirements for a multiplet configuration is then

$$P_{\rm m}(k) = P[k-1; P_\delta, k-1] \exp\left(-s_\nu^m\right), \tag{8}$$

which, given the form of equation (3), reduces to

$$P_{\rm m}(k) = P_\delta^{k-1} \exp\left(-s_\nu^m\right). \tag{9}$$

The probability that multiplet structure appears by chance at least once over the full range Δ_p will depend upon $P_{\rm m}(k)$ and the number of possible arrangements of the structure over this range. The allowed arrangements are constrained by: (i) the number of components in the multiplet; (ii) the bin spacing, $N_{\rm syn}$, that separates adjacent components (i.e. the synodic splitting of 400 nHz); and (iii) the uncertainty, $2N_\delta$.

In a well-resolved spectrum both $N_{\rm syn}$ and N_δ are much smaller than N – here by more than two orders of magnitude – and as such the number of arrangements is well approximated as N. The required probability is therefore

$$P_{\rm mult} \approx P[1; P_{\rm m}(k), N] \tag{10}$$

where $k = 2$, 3 and 4 for a pair ($P_{\rm pair}$), triplet ($P_{\rm trip}$) and quadruplet ($P_{\rm quad}$) respectively. More complete expressions are given in Appendix C; however, the corrections to equation (10) are small and the approximate expression above more than suffices in practice.

To illustrate the threshold levels set, consider when the full BiSON time series was analysed in one piece. With $\Delta_p = 70$ μHz, the relative height thresholds, s_ν^m, required to give test values $P_{\rm mult} \leqslant 0.1$ were 8.1 for a pair; 6.7 for a triplet; and 6.0 for a quadruplet. Were we to have imposed a more restrictive (i.e. less conservative) splitting uncertainty of $\delta\nu = \pm 50$ nHz this would have, for example, lowered the pair threshold to ∼7.8.

4.3 Detection on account of the width of modes: searching for prominent concentrations of power

The acoustic modes are damped and provided the time series is of a sufficient length the resonant peaks may be resolved in the frequency domain. The presence of a prominent concentration of spikes over a range of bins similar to the modal linewidth, $\Delta\nu$, can therefore be used as another detection criterion. Here, one aims to search for such structure in the vicinity of the expected locations of the various azimuthal (m) components of a given mode. When compared with the simple $P_{\rm s}$ test in Section 4.1 this lowers the relative height s_ν required to give a significant detection; however, more than one spike must lie above this over the restricted width range.

The linewidths of the modes decrease rapidly in magnitude with decreasing radial order in a manner consistent with a power law in frequency. Lengths T of the order of many years are required to give access to the domain where low-frequency modes are resolved (Chaplin et al. 2002a). In Fig. 1 we show the lowest reliable widths extracted from fits to radial modes in a spectrum made from the 9-yr BiSON time series. These data are plotted as a function of frequency (points with error bars returned by the fitting procedure) in units of the frequency-domain bin width, i.e. we plot $\Delta\nu \cdot T$. The solid line (with the accompanying dashed 1σ error envelope) is the extrapolated best power-law fit to the reliable low-frequency widths. We should add that modelled widths (e.g. Houdek et al. 1999) follow closely a power-law-type decrease at low frequencies and this guided our choice of function to fit.

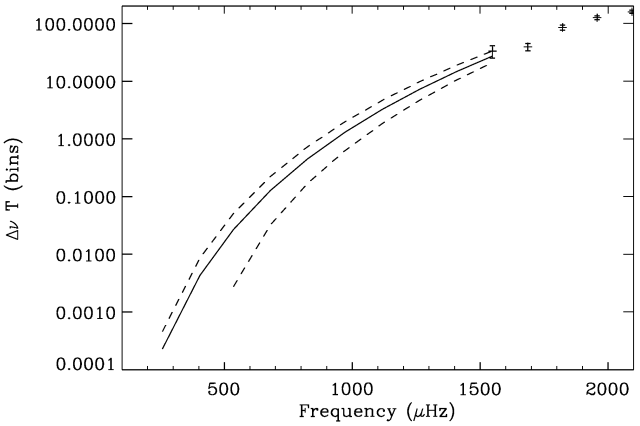

Figure 1. The projected widths of the low-ℓ acoustic modes at low frequencies, plotted in units of the bin width of the 9-yr BiSON time series, i.e. the product of the width and observing time, $\Delta\nu T$. The dashed lines follow the 1σ error envelope on the best power-law fit (solid line) made to reliable measures of the width (data with errors) above ∼1500 μHz. The lower bounds of the error envelopes on the lowest projected frequencies are such that they lie well outside the plotted range.

The plot demonstrates clearly that for $\nu < 1000$ μHz the bin width is too large to allow the modes to be resolved in frequency. One would therefore expect that tests for concentrations of prominent power would be of use only in the higher-frequency part of our target range where the mode peaks would be expected to straddle several bins. However, since the acoustic signals may not be commensurate with T there will be occasions where power from the mode lies in more than one bin even though it is not resolved. In addition, any solar-cycle-related variation of the low-n eigenfrequencies present, although expected to be small in magnitude, could nevertheless be large enough to split the signal between bins because of the fine resolution afforded by the analysis of a long data set. Consequently, tests may still be made at very low radial orders but with the proviso that here one is testing against the probability that prominent power may be present in two consecutive bins. We return to the issue of the non-commensurate nature of the signal more fully in Section 5 below.

Here, we have designed and applied two tests. In the first, we searched for the presence of a 'cluster' of (at least two) prominent spikes that (i) all had relative heights greater than some chosen threshold; and (ii) were all located over a range of the spectrum corresponding to twice the expected width of the mode, i.e. $\pm N_{\Delta\nu}$. The width range was fixed for the undetected modes by the extrapolation of the power-law fit in Fig. 1; as such, the chosen ranges are vulnerable to the appropriateness (or otherwise) of the extrapolation.

While prominent spikes need not lie in consecutive bins they must nevertheless all fall within the stipulated width range to pass for mode-like structure. The threshold must be chosen appropriately (i.e. be high enough) so that the probability of observing two or more spikes over the range on account of the broad-band background is very low; a positive test result will therefore flag the strong possibility of there being a mode present in the tested region.

In practice, a sliding window of width $2N_{\Delta\nu}$ bins was run across the spectrum and at every location the number of spikes r with relative height greater than some chosen threshold, $s_\nu^{\rm d}$, was recorded. The probability that r such spikes are part of the background in the range $2N_{\Delta\nu}$ is

$$P_{\Delta\nu}\left(s_\nu^{\rm d}\right) = P\left[r; \exp\left(-s_\nu^{\rm d}\right), 2N_{\Delta\nu}\right]. \quad (11)$$

One can then calculate the probability that at least one such concentration will appear by chance over the full Δ_p of 70 μHz – which we recall covers N bins in the spectrum – on account of the background alone. This depends upon $P_{\Delta\nu}(s_\nu^{\rm d})$ and the number of possible arrangements of the width window of $2N_{\Delta\nu}$ bins across the N bins of Δ_p. Provided $N_{\Delta\nu} \ll N$ and $P_{\Delta\nu}(s_\nu^{\rm d})$ is small, the test probability we seek is well approximated by

$$P_{\rm d}\left(s_\nu^{\rm d}\right) \approx P\left[1; P_{\Delta\nu}\left(s_\nu^{\rm d}\right), N\right]. \quad (12)$$

The threshold required to return a detection probability that might flag the possible presence of a mode – i.e. $P_{\rm d}(s_\nu^{\rm d}) \leqslant 0.1$ – will vary according to the number of spikes r found above it over a given width range. The value needed when $r = 2$ is higher than for $r = 3$. This is only to be expected, i.e. the greater the number of strong spikes over a narrow range of bins the lower the likelihood that these are the result of the broad-band noise.

To allow for the fact that the threshold $s_\nu^{\rm d}$ need not be as large (i.e. demanding) if a greater number of spikes r are found above it, we tested against two values of the threshold over each mode-target range. One was fixed so that when at least two spikes were found above it a result with $P_{\rm d}(s_\nu^{\rm d}) \sim 0.1$ was returned; the second was set to return a similar probability for the case of $r \geqslant 3$. For a given N (fixed by the length T of the time series) the levels required are set by the changing $N_{\Delta\nu}$ (mode width). When the 3276-d time series was analysed in one piece, the adopted thresholds ranged from ~6.1 to ~9.2 (for $r \geqslant 2$) and ~4.6 to ~7.5 (for $r \geqslant 3$).

In the second test we searched for power being present above a particular threshold in two or more consecutive bins. Formally, we tested therefore against the probability that k consecutive bins all had spikes with relative heights in excess of some threshold $s_\nu^{\rm c}$. Provided that the probability of prominent spikes appearing in any two consecutive bins is small, the probability of this occurring at least once somewhere across the full Δ_p is given approximately (for $k \ll N$) by

$$P_{\rm c}\left(s_\nu^{\rm c}\right) \approx P\left[1; \exp\left(-s_\nu^{\rm c} \cdot k\right), N\right] \quad (13)$$

(The exponent in the expression is as shown since the probability of obtaining a single spike must be raised to the power k to obtain that for k spikes.)

To facilitate this test each modal target range was searched for events where the power in $k \geqslant 2$ consecutive bins exceeded the threshold. This was fixed to give $P_{\rm c}(s_\nu^{\rm c}) \sim 0.1$ when $k = 2$ consecutive bins exceeded the level. The level required to satisfy this is just $s_\nu^{\rm c} = 0.5 s_\nu^{0.1}$.

5 UNCOVERING THE MODES: SHIFTING THE FREQUENCY-DOMAIN BINS

5.1 Shifting the locations of bins in the frequency-domain

As discussed above in Section 4.3, and illustrated in Fig. 1, many of the modes we seek to uncover are expected to be unresolved in the frequency domain. Whether the signal is commensurate with the window function then has a crucial bearing on the observed signal-to-noise ratio. The Fourier transform of a coherent sine wave is realized in the frequency domain as the sinc-squared (power) function of the window. In the worst-case scenario the phase of the signal with respect to the window can be such as to reduce the peak power observed to ~40 per cent of maximum value expected for a fully commensurate arrangement (this for an idealized, noise-free environment). When other sources of noise are present, interference with the signal – be it constructive or destructive – will lead to additional variation. One possible approach to minimizing the uncertainty in the height of any signal present is to oversample the spectrum by zero-padding the time series. This strategy was adopted by Schou (1998) and Gabriel et al. (2002). Here, we adopt a different approach that has the added benefit of giving access to different 'realizations' of the noise background (see below).

We shifted systematically the absolute frequency locations of the bins of a spectrum by removing successively appropriate small numbers of data from the end of the main time series. In this way we would hope to scan across the frequency at which the signal is commensurate. In practice we computed from the main time series spectra at twenty five lengths with a Fast Fourier Transform (FFT) algorithm. Each length (number of samples) was chosen so that the sum of its prime factors was not excessive, this having implications for the run-time of the FFT. For a given main time series the first shorter set analysed from the full length of just over 7 million points required the removal of 363 40-s samples; while the 25th set required the removal of 3744 (with the other sets computed from numbers in between). The numbers of data removed were therefore only a small fraction (1 part in ~2000) of the base length. The 25-set ensemble then provided a fairly even sampling across a range of fractional bin shifts (with respect to the bin locations in the full-length set). Our decision to test spectra at the whole range of lengths (i.e. not just

that closest to the 'commensurate' length), and to settle on an upper limit for this of ∼25 shifts, was based on the following reasoning.

As the bin locations are altered by the removal of data so the spectrum is evaluated at slightly different, discrete locations. If a sufficient number of such bin-shifted spectra are made (each with a different shift) the resulting ensemble will in effect act to give the equivalent of a few different realizations of the noise background. The greater the number of shifts, the greater the 'number' of equivalent realizations obtained. Through the use of Monte Carlo simulations we have been able to demonstrate that a 25-shift set will act to give the equivalent of ≈3 'realizations'. There is little further gain to be had by increasing the number of shifts above this.

To obtain N truly independent spectra would of course demand strict independence of each data set in the time domain. This we clearly do not have; no two bin-shifted spectra are independent. This means that there will be correlation between, for example, the locations of some of the prominent noise spikes in the bin-shifted sets. There is also a marked tendency for spikes to turn into contiguous structures and visa versa as one shifts across the sets. So, while truly independent data offers the advantage of being able to uncover modes (or at least structure that is unlikely to be part of the noise background) by searching for coincidences between sets, the nature of the bin-shifted ensemble denies us this strategy. However, we retain the advantage of being able to observe any signal present against a slightly different realization of the noise background in its vicinity. Therefore, if the likelihood of our uncovering a mode in the base-length set is small, by searching all 25 bin-shifted sets we would expect to increase by a factor of ∼3 the probability of making a detection. This we have verified with simulations of artificial data that have low intrinsic S/N.

5.2 Search strategy over bin-shifted ensemble

In order to take into account the above we adopted the following search strategy for our data. From a given optimized time series we computed a bin-shifted ensemble. These 25 spectra were then searched for mode-like structure around the expected locations, and each detection test was applied (as appropriate).

(i) If uncovered candidate structure passes more than one test in the *same* bin-shifted spectrum (i.e. $P \leqslant 0.1$ in two or more tests) it is most unlikely to be part of the broad-band noise background. Multiple, 'simultaneous' test passes of this type were therefore marked as being significant detections of non-broad-band origin.

To illustrate the expected rare occurrence of such structure, consider a doublet, in which the strength of the weaker component is such that it just passes the pair test (i.e. $P_{\rm pair} = 0.1$) in one of the full-length BiSON spectra; if the stronger peak is also of sufficient prominence to just pass the spike test (i.e. $P_s = 0.1$) the probability that the structure appears by chance – giving a simultaneous pass in both tests – is of the order of $\sim 1.7 \times 10^{-3}$. (This follows from the application of equations (5) and (10) in Section 4.)

(ii) If uncovered candidate structure passed no more than one test in any of the bin-shifted spectra (but possibly the same test in several) we regarded this as being insufficient to award a detection flag. We settled on this conservative strategy for two reasons. First, as previously indicated, the correlation between sets means that structure will tend to persist across several bin-shifted spectra if present already in one, i.e. coincidence *between* shifts is to be expected. Secondly, in a spectrum dominated by broad-band noise the 25-set ensemble will give rise to, on average, about three times as many non-coincident rare and prominent noise structures (e.g. spikes, multiplets etc) as expected from a single set. (This is just a statement of the fact that the ensemble acts like ≈3 separate realizations.)

We therefore allowed for the greater chance occurrence of prominent structure when the full ensemble was taken into account by demanding that the single test-pass in question be re-passed at a higher level of significance. This was done by re-setting Δ_p from 70 to 210 μHz in the detection tests (to offset the expected threefold increase of rare noise structure). To illustrate the effect on the tests, this increased the relative height threshold needed to give $P_s \leqslant 0.1$ in the spike test from $s_\nu^{0.1} = 12.2$ to 13.3 (full 9-yr time series). If the single test passed at the new threshold (still for $P \leqslant 0.1$) it was marked as a significant detection.

6 RESULTS

Table 1 records the detected modes uncovered by our complete analyses while details of the tests that revealed them are given in Table 2. Above ∼1300 μHz the uncovered modes are very prominent. Over the range $400 \leqslant \nu \leqslant 1300$ μHz we found that, other than the structure flagged in the tables, what remained in each tested spectrum was consistent with the expected broad-band noise distribution. As such, the uncovered structure profiled in Table 2 is indeed rare in origin.

The number of tests that were passed by the same structure in the *same* bin-shifted spectrum is indicated in the second column of Table 2 (i.e. coincident test passes), while the individual probabilities associated with each test are given separately in columns three to seven inclusive. Bracketed probabilities denote tests that were passed in isolation (i.e. not simultaneously with others) at the higher threshold demanded by the strategy outlined above in Section 5.2. Some examples of parts of spectra that contain structure that passed the tests are shown in Fig. 2. The $\ell = 2$, $n = 8$ mode is shown in a shorter spectrum, made from a 971-d segment, in which it was more conspicuous (and hence returned the most convincing test probabilities). The dashed line in each panel indicates the 10-per cent threshold, $s_\nu^{0.1}$, for the spike test (Section 4.1); the

Table 1. BiSON low-ℓ mode detections at radial order $n \leqslant 9$ uncovered by the analyses described in the text.

Radial order n	$\ell = 0$	Mode frequency (μHz) $\ell = 1$	$\ell = 2$	$\ell = 3$
6...	972.613 ± 0.002	–	–	–
7...	–	–	1250.564 ± 0.025[1]	–
8...	1263.162 ± 0.012	1329.630 ± 0.004	1394.683 ± 0.011	–
9...	1407.476 ± 0.011	1472.841 ± 0.007	1535.869 ± 0.012	1591.578 ± 0.014
		Other 'mode-like' structure		
3...	535.601 ± 0.002	–	–	718.505 ± 0.001

[1] $m = +2$ only. See text for discussion of estimation of frequencies and errors.

Table 2. Test probabilities associated with Table 1. Unbracketed probabilities were returned in the same spectrum; the number of these coincident test passes is totalled in the second column. The bracketed values were passed in isolation (non-coincident).

Radial order n	Number of coincident test passes	Test probabilities P_s	P_d	P_c	P_{pair}	P_{trip}
			$\ell = 0$			
3...	1	$[6.6 \times 10^{-3}]$	–	–	–	–
6...	1	$[1.9 \times 10^{-3}]$	–	–	–	–
8...	1	$[7.0 \times 10^{-2}]$	$[9.2 \times 10^{-2}]$	–	–	–
9...	2	$[8.7 \times 10^{-2}]$	9.2×10^{-2}	2.2×10^{-4}	–	–
			$\ell = 1$			
8...	4	4.2×10^{-3}	8.0×10^{-5}	2.4×10^{-4}	2.0×10^{-3}	–
9...	4	4.0×10^{-5}	2.2×10^{-4}	8.0×10^{-5}	1.8×10^{-3}	–
			$\ell = 2$			
7...	1	$[4.2 \times 10^{-2}]$	–	–	–	–
8...	5	2.2×10^{-2}	8.3×10^{-2}	9.5×10^{-2}	5.6×10^{-4}	3.6×10^{-3}
9...	5	4.1×10^{-2}	6.2×10^{-4}	9.6×10^{-2}	9.8×10^{-3}	3.8×10^{-2}
			$\ell = 3$			
3...	4	1.7×10^{-2}	9.5×10^{-2}	9.6×10^{-2}	5.0×10^{-3}	$[3.1 \times 10^{-2}]$
9...	2	1.9×10^{-3}	–	–	1.6×10^{-3}	–

P_s: test against relative height (prominence) of spike.
P_d: test against occurrence of prominent power distributed over two estimated linewidths.
P_c: test against appearance of prominent power in contiguous bins.
P_{pair}: test against occurrence of suitably separated pair.
P_{trip}: test against occurrence of suitably separated triplet.

dotted and dot–dashed lines those for the pair and triplet multiplet tests respectively (Section 4.2). The square symbols mark the predicted locations of individual modal components from the Saclay seismic model of Turck-Chiéze et al. (2001), assuming an observed (synodic) rotational splitting of 400 nHz. The spectra have been scaled to show twice the mean-square power.

For the moment, we note that we have flagged the uncovered 'structure' near the expected location of the $n = 3$, $\ell = 0$ and 3 overtones to indicate that we are highly sceptical that these are in fact the signatures of p modes. We discuss these potential candidates in Section 7 below.

The frequencies listed in the first table were estimated as follows. Where modes were resolved, i.e. had width across several bins, and such prominent structure was apparent we fitted their peaks – following the general prescription outlined in Chaplin et al. (1999) – to either the asymmetric function of Nigam & Kosovichev (1998) or a simple symmetric Lorentzian. The likelihood-ratio test (Appourchaux et al. 1995) was used to assess quantitatively whether the use of the asymmetric function led to a significant improvement in the fit. Because of the small widths of the modes this tended not to be the case and under such circumstances the symmetric-fitted frequency was adopted as the estimate of the mode frequency (the difference between the symmetric and asymmetric frequencies is well below the measurement uncertainty). All modes at $n = 9$, in addition to those at $n = 8$, $\ell = 1$ and 2 were analysed in this manner, the quoted errors in each case being the formal uncertainties from the fit. The other detections were dealt with as follows.

At $\ell = 0$, $n = 3$ and 6 the modes were predicted to be unresolved and the frequencies of the uncovered spikes were therefore estimated from the location of the bin in which prominent power was revealed. The uncertainty was taken to be half of one binwidth [i.e. an upper limit estimate of $\pm 1/(2T)$] of the full length of the time series. A similar strategy was adopted for the 'structure' found near the expected location of the $\ell = 3$, $n = 3$ multiplet: here, the frequency was taken as the mean of the locations of those bins that gave a significant P_{pair}. The error was estimated as $1/(\sqrt{2} \times 2T)$.

At higher n all detections arose from the analysis of time-series segments in which the modes were oversampled (i.e. resolved in frequency). Those that remain to be discussed here were made on account of the presence of power in 1 or 2 bins only. Clearly, such spikes need not necessarily lie at the centre of the mode and, depending upon the number of bins straddling one linewidth, the location of 'prominent' power may be more than one bin width from the central frequency. We therefore adopted the more conservative estimate of half of one estimated linewidth [i.e. $\pm 1/(2\Delta\nu T)$] as the uncertainty on such detections, e.g. for the radial-mode detection at $n = 8$ (made in a 971-d time-series slice). The claimed detection at $\ell = 2$, $n = 7$ offered the additional complication of there being only one m value uncovered (here, the upper sectoral component). To estimate the frequency we determined the mean synodic splitting of the other detections in Table 1 ($\sim$400 nHz) and subtracted an appropriate multiple of this from the frequency of the uncovered component. We have of course assumed that the accuracy with which we can predict both the eigenfrequencies (cf. Section 3.3.2) and splittings (cf. Section 4.2) allows us to reliably label the m value uncovered. The listed uncertainty was derived through the combination in quadrature of (i) the estimated error on the individual m-component frequency; and (ii) the error on the synodic splitting value expected from a single detection. For the latter we could have taken the error on the mean of the well-measured splittings but instead chose to adopt the more cautious approach of choosing the larger sample uncertainty.

7 DISCUSSION

The test probabilities listed in columns three to seven of Table 2 provide an objective measure of the likely significance of each claimed detection. Additional guidance regarding their plausibility is given by (i) how far the estimated frequency lies from modelled theoretical

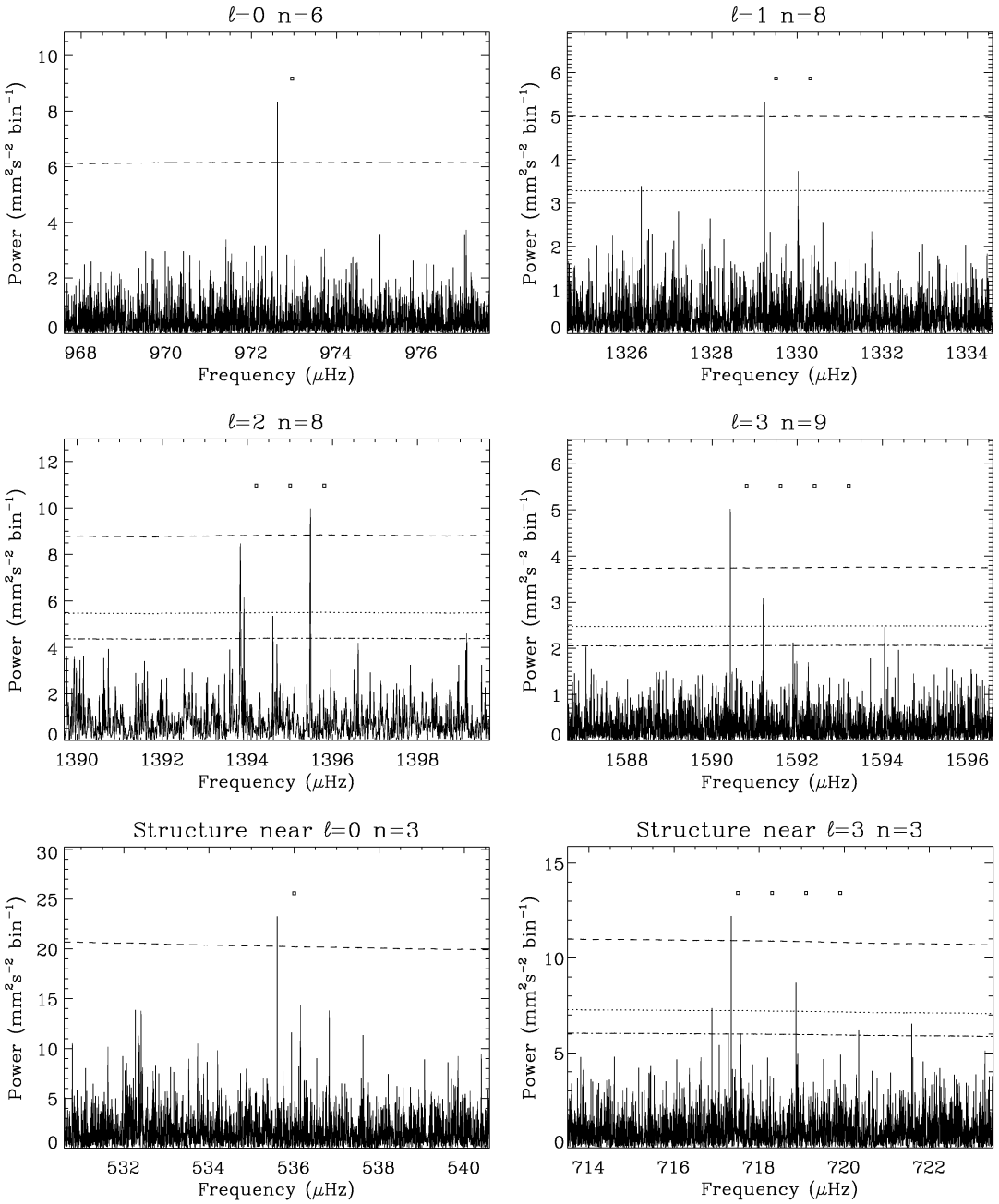

Figure 2. Some examples of detected modes in power spectra made from the BiSON time series, in addition to the mode-like structure present at the expected locations of $n = 3$, $\ell = 0$ and 3 (two lower panels). The $\ell = 2$, $n = 8$ mode is shown in a shorter spectrum, made from a 971-d segment, in which it was more conspicuous (and hence returned the most convincing test probabilities). In each case the dashed line indicates the 10-per cent threshold, $s_v^{0.1}$, for the spike test (Section 4.1); and the dotted and dot–dashed lines those for the pair and triplet multiplet tests respectively (Section 4.2). All levels were calculated against a range $\Delta_p = 70$ μHz (see text). The square symbols mark the predicted locations of individual modal components from the Saclay seismic model of Turck-Chiéze et al. (2001), assuming an observed (synodic) rotational splitting of 400 nHz. The spectra have been scaled to show twice the mean-square power.

values (Section 3.3.2); and (ii) the likelihood of uncovering the mode given a reasonable estimate of the underlying S/N at which it appears in the BiSON spectrum.

With regard first to point (ii), we are currently working on methods to predict reliably the S/N at low frequencies on the basis of extrapolations of the modal energy supply rate and damping rate measured at higher frequencies. The use of these two quantities is appealing since they are independent of one another. Our work to date indicates that the extrapolation is extremely sensitive to the range over which the observed mode characteristics are fitted to derive the extrapolation, and the form of the function that is used. This work is still in progress and will be the subject of another paper.

The differences suggested by point (i) above are shown in Fig. 3. Here, we plot the observed BiSON frequencies minus the Saclay seismic model predictions (Turck-Chiéze et al. 2001). The differences generated with either model 'S' (Christensen-Dalsgaard et al. 1996) or model 'M1' (Provost et al. 2000), while typically of a slightly smaller magnitude, are noticeably more 'noisy' in the sense

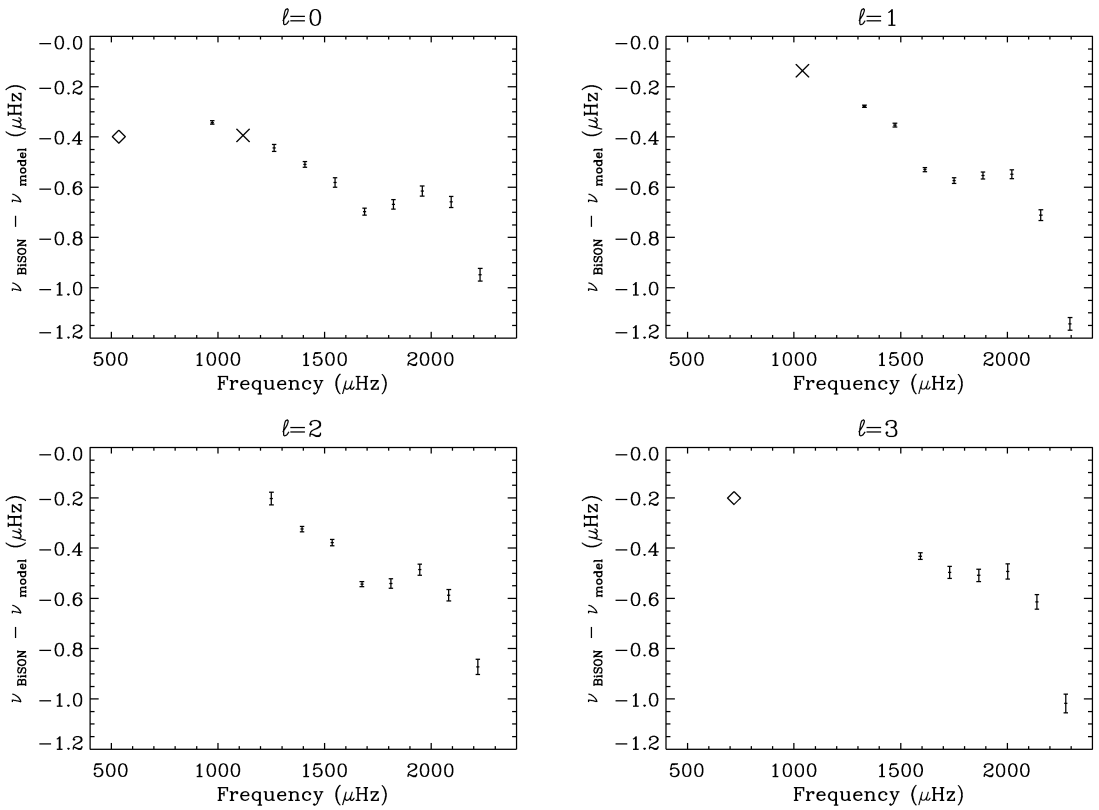

Figure 3. Difference in frequency between the BiSON detections listed in Table 1 and those of the Saclay 'seismic' solar model (Turck-Chiéze et al. 2001), plotted in the sense observation minus model. The BiSON frequencies at $n > 9$ come from maximum-likelihood fits to the 9-yr spectrum (Chaplin et al. in preparation). The locations of the mode-like structures uncovered near the model-predicted $n = 3$, $\ell = 0$ and 3 frequencies are marked by diamond symbols on the appropriate panels. The crosses indicate where mode-like structure was found on account of a positive result in the contiguous power test (near $\ell = 0$, $n = 7$) and pair test (near $\ell = 1$, $n = 6$), but with the test thresholds fixed for a single spectrum; when the levels were raised to the three-set equivalent (because in each case the initial analysis returned a single positive test result only) the repeated tests failed.

that they are not as smooth with frequency (although as noted previously the RMS deviations from a smooth trend are still only of size $\sim \pm 0.1$ μHz). The departures from a smooth trend implied by Fig. 3 are so small as to emphasize the need for numerical accuracy in the model calculations. The fact that our claimed detections all lie on such a trend is interesting, although of course by no means conclusive proof of their validity.

In addition to the frequencies listed in the main part of Table 1, we have plotted also two other sets of points. The crosses mark structure near $\ell = 0$, $n = 6$ and $\ell = 1$, $n = 7$ that in each case passed a single test only (the contiguous and pair tests respectively). Both however failed when the thresholds were reset to a higher level (as demanded by the strategy outlined in Section 5.2) and were therefore not marked as firm detections. With reference again to Fig. 3 each frequency does lie on its respective trend-line; this may suggest that mode structure is present in the BiSON spectrum but at a level that is insufficient to return a significant test result.

The diamonds mark the location of the structures uncovered near $\ell = 0$, $n = 3$ and $\ell = 3$, $n = 3$ (frequencies in Table 1, spectra in Fig. 2). The '$\ell = 3$' structure is particularly noteworthy in that it passes four tests simultaneously; this singles it out as undoubtedly being of non-broad-band origin. Our aversion to identifying both detections as modes is based largely upon the power uncovered in the prominent components. We have indicated already that obtaining a reliable extrapolation of the powers expected at low frequencies is difficult; nevertheless, both would demand an upturn in mode power in this region (drastically so for the $\ell = 3$ mode) which rather opposes the trend suggested by modelling of the acoustic spectrum (e.g. Houdek 2002). We have nevertheless been unable to duplicate similar structure in a variety of simulations of varying complexity, and so the origin of both features is still open to question (e.g. possibly some form of instrumental artefact). Without independent corroboration from other observers we prefer to flag their presence but not list them as uncovered modes.

A comparison of the frequencies listed in Table 1 with those in Toutain et al. (1998), Bertello et al. (2000) and Garcia et al. (2001) gives a satisfactory level of agreement. Only one difference – made with the Bertello et al. and Garcia et al. frequencies for $\ell = 0$, $n = 8$ – is greater than 3σ (combined uncertainty). Our structure at $\ell = 0$, $n = 3$ also lies many standard deviations beyond the claimed detection of Bertello et al. That there are not a greater number of such disparate comparisons is, perhaps, surprising. This is because under conditions where the S/N is low a detection may be made on account of the presence of prominent power in one bin only, whilst the underlying mode peak may be several bins wide. Under these circumstances there is a tendency to underestimate the true uncertainty associated with the claimed detection, since a value for the centroid corresponding to – or heavily dependent upon – the location of the lone, 'prominent' bin may be biased. This indicates that a great deal of care must be taken in arriving at suitable estimates of the errors.

However, one must remember also that since we are all observing one and the same star the stochastic signature of the modes, and much of the solar background (which may show some variation depending upon the observational technique), will be correlated strongly in the independent data sets. So, the potentially 'incomplete' appearance of a mode may be similar in the various data. This condition will hold provided that the S/N is not so low that interference with the background – a proportion of which will not be common in the different data – becomes the dominant factor.

ACKNOWLEDGMENTS

We are indebted to H. K. Williams, J. Allison and R. Bryan for their technical and analysis support in Birmingham and to former colleagues, in particular C. P. McLeod, J. Litherland and R. Lines. We would also like to thank P. Fourie and the late R. Stobie at SAAO; the Carnegie Institution of Washington; the Australia Telescope National Facility (CSIRO); E. J. Rhodes (Mt Wilson, California); and members (past and present) of the IAC, Tenerife. We thank J. Christensen-Dalsgaard, S. Turck-Chiéze and S. Couvidat for providing model frequencies. BiSON is funded by the UK Particle Physics and Astronomy Research Council. We acknowledge the stimulus provided by the PHOEBUS collaboration.

REFERENCES

Appourchaux T., Toutain T., Gough D. O., Kosovichev A. G., 1995, in Ulrich R. K., Rhodes E. J., Däppen W., eds, Proc. GONG 94. Astron. Soc. Pac., San Francisco, p. 314
Appourchaux T. et al., 2000, ApJ, 538, 401
Appourchaux T. et al., 2001, in Eff-Darwich A., Palle P., Wilson A., eds, SOHO10/GONG 2000 Workshop: Helio- and Asteroseismology at the Dawn of the Millennium. ESA SP-464, Noordwijk, p. 467
Bertello L., Varadi F., Ulrich R. K., Henney C. J., Kosovichev A. G., Garcia R. A., Turck-Chiéze S., 2000, ApJ, 537, L143
Chaplin W. J., Elsworth Y., Howe R., Isaak G. R., McLeod C. P., Miller B. A., New R., 1997, A&AS, 125, 195
Chaplin W. J., Elsworth Y., Isaak G. R., Miller B. A., New R., 1999, MNRAS, 308, 424
Chaplin W. J., Elsworth Y., Isaak G. R., Marchenkov K. I., Miller B. A., New R., 2001, MNRAS, 327, 1127
Chaplin W. J., Elsworth Y., Isaak G. R., Miller B. A., New R., 2002a, MNRAS, 330, 731
Chaplin W. J., Elsworth Y., Isaak G. R., Miller B. A., New R., Thiery S., 2002b, MNRAS, submitted
Christensen-Dalsgaard J. et al., 1996, Sci, 272, 1286
Claverie A., Isaak G. R., McLeod C. P., van der Raay H. B., Roca-Cortes T., 1981, Nat, 293, 443
Elsworth Y., Howe R., Isaak G. R., McLeod C. P., Miller B. A., New R., Wheeler S. J., 1995, A&AS, 113, 379
Gabriel A. et al., 2002, A&A, 390, 1119
Garcia R. et al., 2001, Sol. Phys., 200, 361
Houdek G., 2002, in Aerts C., Bedding T., Christensen-Dalsgaard J., eds, Proc. IAU Symp. 185, Radial and non-Radial Pulsations as Probes of Stellar Physics. Astron. Soc. Pac., San Francisco, p. 447
Houdek G., Balmforth N. J., Christensen-Dalsgaard J., Gough D. O., 1999, A&A, 351, 582
Monteiro M. J. P. F. G., Thompson M. J., 1998, in Deubner F.-L., Christensen-Dalsgaard J., Kurtz D., eds, Proc. IAU Symp. 185, New Eyes to See inside the Sun and Stars. Pushing the Limits of Helio- and Asteroseismology with New Observations from Ground and from Space. Kluwer, Dordrecht, p. 317
Nigam R., Kosovichev A. G., 1998, ApJ, 505, L51
Provost J., Berthomieu G., Morel P., 2000, A&A, 353, 775
Roxburgh I. W., Vorontsov S. V., 2000, MNRAS, 317, 141
Schou J., 1998, in Korzennik S., Wilson A., eds, SOHO6/GONG98 Workshop: Structure and dynamics of the interior of the Sun and Sun-like stars. ESA SP-418, Noordwijk, p. 47
Toutain T., Appourchaux T., Fröhlich C., Kosovichev A. G., Nigam R., Scherrer P. H., 1998, ApJ, 506, L147
Turck-Chiéze S. et al., 2001, ApJ, 555, L69

APPENDIX A: 'FOOTPRINT' REMOVAL

A number of the BiSON observatories support equatorially mounted instrumentation. While this has simplified the implementation of the automated operation on which BiSON depends, these instruments are prone to giving quasi-oscillatory outputs at low frequencies ($\leqslant$1000 μHz). In particular, our instruments at Sutherland (South Africa) and Narrabri (NSW, Australia) consistently produce such artefacts, with the amplitudes of these spurious signals typically of the order of $\sim$2 m s^{-1}.

In an effort to increase the quality of our data in the low-frequency regime we have worked on improving the stability of the instruments with the aim of removing the artefact at source. However, despite improvements to temperature stability and the reproducibility of timing, plus careful modelling of the instrument output and thorough tests of the optical design, the cause of the artefact remains uncertain. We have, therefore, pursued a line of investigation whereby appropriate low-frequency functions are fitted to the data archive. This has allowed us to partially remove much of the unwanted low-frequency drift in our data thereby allowing more of our otherwise high-quality archive to be used when we generate power spectra (see below in Section B). Furthermore, this programme is informing our investigation of the origin of the effect.

Fig. A1 shows a typical set of calibrated velocity residuals (plotted as a function of universal time) for the Narrabri instrument. In addition to the strong, 5-min oscillation signal slow oscillations with a period of about 1 h are seen, with the most prominent structure present at the start of the day. The surmise that this signal is non-solar in origin is proved by its near repeatability from day to day (sites not supporting equatorially mounted instrumentation, but which collect simultaneous data, do not show such effects). Slight variations in the signal from site to site give each a characteristic 'footprint'.

The artefact, while not being perfectly reproducible, has a number of characteristic features that make the task of constructing a suitable

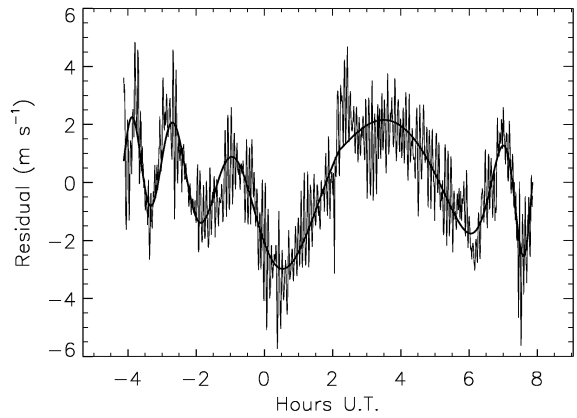

Figure A1. Calibrated velocity residuals collected by the BiSON instrument in Narrabri (NSW, Australia). In addition to the strong, 5-min oscillation signal slow oscillations with a period of about 1 h are seen, with the most prominent structure present at the start of the day. This instrumental 'footprint' has been fitted to an 11-component model (see text), with the best fit indicated by the dark, continuous line.

fitting function a tractable problem. The most obvious component comprises a few oscillations that occur through the morning. The amplitude of the component remains approximately constant (at about $\sim 2\,\mathrm{m\,s^{-1}}$) whilst the period gradually increases. There are few, if any, of the oscillations characteristic of this component present in the afternoon. The second, weaker component of the artefact has a period of ~ 1 d. Its amplitude passes through zero approximately at local noon.

A range of fitting functions have been tested in attempts to match the artefact structure. Functions and variables related to physical processes in the instrument have been chosen wherever possible. The change in period of the stronger component is governed in our function by the time constant of an exponential, suggesting an instrumental time constant; the form of the weaker component depends on the hour angle (HA) of the Sun. The structure of the artefact in the afternoon is sufficiently different from that of the morning (particularly so for the stronger component) that some elements of the fitting function are allowed to change at noon. However, we impose the restriction that the weaker component remain continuous through the day. Since an equatorially mounted instrument could suffer a step change at noon, the composite fitting function and its gradients are not constrained to be continuous at noon.

The function, $s(t)$, applied to a given day of data collected over the epoch $t_{\mathrm{start}} \leqslant t \leqslant t_{\mathrm{end}}$ then takes the form

$$s(t) = \begin{cases} s_{\mathrm{am}}(t) & \text{for} \quad t \leqslant \text{local noon}, \\ s_{\mathrm{pm}}(t) & \text{for} \quad t \geqslant \text{local noon}, \end{cases} \tag{A1}$$

where

$$s_{\mathrm{am}}(t) = s_1 \cos\left\{2\pi\nu_1 \exp\left[-\frac{(t - t_{\mathrm{start}})}{\tau_{\mathrm{am}}}\right] + \phi_1\right\} + s_2 \cos(2\pi\nu_2 \sin[\mathrm{HA}] + \phi_2), \tag{A2}$$

and

$$s_{\mathrm{pm}}(t) = s_3 \cos\left\{2\pi\nu_3 \exp\left[-\frac{(t_{\mathrm{end}} - t)}{\tau_{\mathrm{pm}}}\right] + \phi_3\right\} + s_2 \cos(2\pi\nu_2 \sin[\mathrm{HA}] + \phi_2). \tag{A3}$$

The function has 11 parameters in all. The amplitudes s_1 and s_3, frequencies ν_1 and ν_3 and phases ϕ_1 and ϕ_3, characterize the stronger component in the morning and afternoon respectively. This component decays with time constant τ_{am} prior to local noon and τ_{pm} after local noon. The elements s_2, ν_2 and ϕ_2 characterize the weaker component throughout the day. The dark, continuous line in Fig. A1 shows the best fit of the model to the data. A substantial fraction of the unwanted artefact can clearly be removed by subtracting this from the residuals.

APPENDIX B: QUALITY ASSESSMENT FOR TIME SERIES CONSTRUCTION

The next stage of processing comprises an objective assessment of the quality of all daily residuals available from all sites (with those from the equatorially mounted stations having been processed to suppress the 'footprint' artefact).

The generation of a well-filled BiSON time series demands that an appropriate (coherent) combination of data collected by the network instruments be made. The quality of the data will inevitably vary, with some stations better suited to the study of low-frequency phenomena than others. In order to maximize one's ability to detect low-frequency modes, data selection criteria are required that take into account the quality of the observations made at each site over the frequency range of interest. The need to maximize the duty cycle of the network implies that data should be used, where available, from a given station. However, there is a trade-off between (i) the introduction of these data to the final time series; and (ii) the possibility that – if they are of poor quality – their use may drive up the noise power level of the combined network set to such an extent that this negates the apparent advantage of using the data in the first place.

The above can be expressed quantitatively as follows. First, consider a time series made from single site data. Let d be the fractional duty cycle of the time series, and $\langle\xi\rangle_\nu$ be the average over all contributing days of the mean spectral noise power over the targeted frequency range (say, $\nu \rightarrow \nu + \delta\nu$). A simple 'figure of merit' that encapsulates the trade-off is then

$$\mathcal{F} = d/\langle\xi\rangle_\nu. \tag{B1}$$

For $d = 1.0$, i.e. a 100 per cent fill in the time domain, $\mathcal{F}$ is simply the inverse of the mean power over the frequency band of interest. If data are to be combined from several stations, this expression must be generalized somewhat. If the characteristics of the constituent sites are each tagged by the index i, such that for N_{s} stations, $1 \leqslant i \leqslant N_{\mathrm{s}}$, the expanded expression becomes

$$\mathcal{F} = \frac{D^2}{\sum_{i=1}^{N_{\mathrm{s}}} d_i \langle\xi\rangle_{\nu,i}}, \tag{B2}$$

where

$$D = \sum_{i=1}^{N_{\mathrm{s}}} d_i. \tag{B3}$$

We have implemented an algorithm that makes use of the above in the following manner. Let $\langle\varsigma\rangle_{\nu,i,t}$ be the mean power, calculated over the target frequency range, on a day-by-day basis. Next consider a daily power rejection threshold for each site $T_{\nu,i,t}$, again appropriate to the targeted frequency range. If we choose to reject individual days for which the the power $\langle\varsigma\rangle_{\nu,i,t} > T_{\nu,i,t}$, this will alter (i) $\langle\xi\rangle_{\nu,i}$, i.e. the overall mean power contribution from site i, and (ii) d_i, the fractional fill contribution from the site. Clearly, we wish to find some combination of site thresholds, $T_{\nu,i,t}$ that will give the optimal overall combination, i.e. that which maximizes the figure of merit $\mathcal{F}$. This can be realized by performing a multiparameter minimization where one varies, and then seeks to find at convergence, optimal estimates of the power rejection thresholds of each site.

The fit is performed over a particular epoch to yield the thresholds; those daily data the noise power of which exceeds the appropriate threshold are then discarded and a coherent time series constructed from the remaining 'good' data.

When applied to the 9-yr BiSON data base, the application of the maximum-likelihood rejection thresholds resulted in the effective fill of the data set being reduced from ~ 75 to ~ 60 per cent. However, the mean noise power in the low-frequency band of interest was halved giving an improvement in the figure of merit, $\mathcal{F}$, of ~ 1.6.

APPENDIX C: FULL PROBABILITIES FOR MODE-MULTIPLET TESTS

To arrive at full expressions for the probabilities of observing multiplet-like structure (pair, triplet or quadruplet) we must determine the number of possible arrangements of the structure across the range Δ_p (Section 4.1). To illustrate, we take the example of an $\ell = 1$ doublet.

With reference to Section 4.2, if a prominent spike is present in the spectrum other spikes must lie at multiples of 400 nHz from it (this covering $N_{\rm syn}$ bins), with an allowed uncertainty on their location of $\delta\nu = \pm 100$ nHz (this covering $\pm N_\delta$ bins). In the case of a doublet, the second component must therefore lie somewhere between $2N_{\rm syn} - N_\delta$ and $2N_{\rm syn} + N_\delta$ bins from the first (the mid separation being just $2N_{\rm syn}$).

When the window in which the second spike lies is the full $2N_\delta$ bins wide, the number of possible arrangements of this across the N bins that cover Δ_p will be $N - 2N_{\rm syn} - N_{\delta s} + 1$. The probability of this structure appearing by chance over the range Δ_p – assuming both spikes that make up the pair have relative heights in excess of the relative height threshold s_ν^m – is then

$$P_{\rm full} = P[1; P_{\rm m}(2), N - 2N_{\rm syn} - N_{\delta s} + 1]. \tag{C1}$$

There are, however, a series of further allowed arrangements for which the second spike lies in a window that is narrower than $2N_\delta$. These are to be found when the second spike is at the right-hand frequency edge of the full N-bin window; here, we assume the structure has been moved from left to right across Δ_p to assess the number of possible arrangements. As the structure is shifted progressively to the right, so the width of the second window will decrease from $2N_\delta$ down to just one bin. The structure at each shifted location will have a probability

$$P_{\rm m}(2) = P[1; P_\delta, 1] \exp\left(-s_\nu^m\right), \tag{C2}$$

where

$$P_\delta = P\left[1; \exp\left(-s_\nu^m\right), 2N_\delta - u + 1\right], \tag{C3}$$

and u – which tags each shifted location – runs from 1 to $2N_\delta$. The combined probability of these 'edge' arrangements will be the sum

$$P_{\rm edge} = \sum_{u=1}^{2N_\delta} P[1; P_\delta, 1] \exp\left(-s_\nu^m\right). \tag{C4}$$

For small $P_{\rm full}$ and $P_{\rm edge}$ the overall probability that the structure is part of the broad-band noise background will be

$$P_{\rm pair} \approx P_{\rm full} + P_{\rm edge}. \tag{C5}$$

The expressions for all the various P (pair, triplet and quadruplet) follow in a similar manner, and are

$$P_{\rm pair} \approx \begin{cases} \text{at } \ell = 1; \text{and } \ell = 2,\ 3 \text{ when } |\Delta m| = 2: \\ P[1; P_{\rm m}(2), N - 2N_{\rm syn} - N_\delta + 1] + P_{\rm edge}. \\ \text{at } \ell = 2(\text{sectoral components}): \\ P[1; P_{\rm m}(2), N - 4N_{\rm syn} - N_\delta + 1] + P_{\rm edge}. \\ \text{at } \ell = 3(\text{sectoral components}): \\ P[1; P_{\rm m}(2), N - 6N_{\rm syn} - N_\delta + 1] + P_{\rm edge}. \end{cases} \tag{C6}$$

$$P_{\rm trip} \approx P[1; P_{\rm m}(3), N - 4N_{\rm syn} - N_\delta + 1] + P_{\rm edge}. \tag{C7}$$

$$P_{\rm quad} \approx P[1; P_{\rm m}(4), N - 6N_{\rm syn} - N_\delta + 1] + P_{\rm edge}. \tag{C8}$$

For the triplet and quadruplet tests, the $P_{\rm edge}$ are calculated by summing over all $P_{\rm m}(3)$ and $P_{\rm m}(4)$ (cf. equations C2 and C4) respectively. For the full BiSON data set, the chosen Δ_p of 70 μHz covered $N = 19818$ bins. Since both $N_{\rm syn}$ and N_δ are at least two orders of magnitude smaller, $P_{\rm mult}$ is dominated by $P_{\rm main}$ and furthermore all pair probabilities are similar.

This paper has been typeset from a TEX/LATEX file prepared by the author.