Scientific Advancements in Analysis of Solar Oscillation Data

by

Steven James Hale

A thesis submitted to
The University of Birmingham
for the degree of
Master of Philosophy

School of Physics and Astronomy
The University of Birmingham
November 2003

Abstract

The main content of this thesis discusses extensive new software designed to improve the quality of long timeseries produced using data from the Birmingham Solar Oscillations Network. This includes an analysis of the effects of rejecting bad data, and seven different algorithms for merging data from overlapping stations. Methods for checking the relative timing between stations are discussed.

Also included is a brief history of the Birmingham Network, and a description of the resonant scattering spectrometers used by the group along with a discussion on the standard calibration for converting from raw data to velocity residuals.

Acknowledgements

I am indebted to two people for giving me the opportunity to produce this thesis. They are Prof. Yvonne Elsworth and Dr. Bill Chaplin. I owe a great deal of thanks to them both for having sufficient confidence in me when no one else did to arrange funding for this project, and also for their extensive help and advice.

I would also like to thank all the other members of the Birmingham Solar Oscillation Network for creating a pleasant working environment and for the nights of beer and curry. Specifically, Graham Verner, Bernie McCarty, John Alison, Rosy Simoniello, Brek Miller, Richard Bryan, and Hugh Williams.

Finally I would like to acknowledge the financial support of the Birmingham Solar Oscillations Network and the School of Physics and Astronomy at the University of Birmingham.

Steve Hale

Birmingham, UK.
November 6, 2003

Preface

The aim of this project was to provide a suite of programs that would deliver an improvement in the quality of long timeseries produced from the daily residual data files from the Birmingham Solar Oscillations Network. The network is made up of six international observing stations distributed around the globe such that as the Sun is setting at one site it is rising at another. Data is sometimes available from more than one site, and historically the overlaps were processed by simply selecting whichever station has produced the better quality data based on a figure of merit, and discarding the rest. However, it is possible to improve the quality of the data by merging any overlapping residuals from other sites before concatenating them into a long timeseries.

Unfortunately, the current concatenation software is old and very difficult to develop further. Before any new overlap merging algorithms could be implemented and tested it was necessary to write new software from scratch. This was not a trivial task. The new software was written in IDL - the Interactive Data Language - and is based around an Object Oriented philosophy with a fully procedural design and a novel internal data structure. If new functionality is required it is simply a matter of writing a new module and plugging it in.

In addition to the main code several pre-processing programs were written. These included an assessment of overall data quality in order to assist in the setting of noise rejection thresholds, and an advanced semi-automatic timing error detection and correction routine with a full graphical user interface. Understandably this work consumed a significant percentage of the total project-time.

Throughout this thesis the following definitions will be used when referring to common frequency bands:

- Low Frequency Noise (LP1) $\rightarrow$ 0.8 to 1.3mHz.
- Low Frequency Noise (LP2) $\rightarrow$ 0.2 to 0.7mHz.
- High Frequency Noise (HF) $\rightarrow$ 5.5 to 12.5mHz.
- Five Minute Signal (FM) $\rightarrow$ 2.0 to 5.0mHz.

There are currently seven different overlap merging algorithms implemented in the software. Five are variations on the basic priority algorithm using different methods of ranking data in order of quality, and the remaining two take an average and a weighted average of all overlapping data respectively.

Chapter 1 gives a brief history of the High Resolution Optical Spectroscopy Group and the foundation of the Birmingham Solar Oscillations Network. Chapters 2, 3, and 4 describe the BiSON instruments and discuss the quality of the data they produce. Finally, Chapters 5 and 6 discuss the concatenation of the daily data from each site into one long timeseries.

Contents

List of Figures

List of Tables

There is a theory which states that if anyone ever discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.

There is another theory which states that this has already happened.

The Restaurant at the End of the Universe
(The Hitch Hiker's Guide to the Galaxy Trilogy)
Douglas Adams 1952 - 2001

Chapter 1

Introduction

The Sun is crucial to life here on the Earth. It radiates an enormous amount of energy - $4x10^{26}$W - which is essentially all in the form of electromagnetic radiation ranging from radio waves through to X-rays. The majority of the light from the Sun is emitted from the photosphere, and analysis of this light has contributed substantially to our understanding of physics and astronomy. Unfortunately, the extent of the photosphere is only a tiny fraction of the solar radius and the rapid increase in optical depth below this layer prevents us looking any deeper into the Sun, at least via conventional methods.

In the last thirty years a technique known as helioseismology has been developed for probing the solar interior, and is a close analogue of terrestrial seismology. Just as the internal structure of the Earth can be determined by analysing the arrival times of sound waves travelling through the Earth (such as those produced by an earthquake) at different locations, similarly it is possible to calculate the internal conditions and structure of the Sun through an analysis of the surface vibrations.

1.1 High Resolution Optical Spectroscopy at Birmingham

In the early 1960's Leighton had found periodic convective motions of solar granules with characteristic timescales of about five minutes and velocity amplitudes of order 1000ms^{-1}, (Leighton et. al. 1960). He thought that since the origin of this process should be stochastic any motion would have a random appearance, and hence considered the periodicity to be an incoherent phenomenon local to only small areas of the Sun. It was not until much later in 1970 when Roger Ulrich (Ulrich, 1970) and Leibacher and Stein (Leibacher and Stein, 1971) suggested independently that the oscillations could be more extended and have some structure in space and in time.

It is the study of these oscillations that formed the foundation of *The High Resolution Optical Spectroscopy* (HiROS) group in 1971 by George R. Isaak, at the University of Birmingham (Izaak 1992/93). The group set out to search for very small Doppler Shifts due to radial oscillations of the Sun as a whole, and the first instrument was in use by 1974.

The instrument was initially deployed at Pic du Midi in the French Pyrenees, and revealed that global averaging of the small scale granular motions seen by Leighton gave rise to noise with a root mean square amplitude of around one metre per second and a timescale of five minutes. From 1975 onwards the instrument was moved to the island of Tenerife, where the group was joined by Andre Claverie in 1976.

The predictions of structure within the five minute region made by Ulrich, and Leibacher and Stein were first verified in a patch on the equator by

Deubner (Deubner, 1975) and later globally by Claverie (Claverie et al., 1979). Claverie realised that the signals were due to both radial and non-radial oscillations of the entire Sun, and were analogous to the seismic waves within the Earth. Sound waves are generated in the turbulent upper regions of the convention zone, where the convective velocities reach their maximum. As the acoustic waves perturb the solar material so gentle oscillations of the solar surface arise, with the strongest observed periods centred around five minutes. Some of the oscillations are formed by sound waves which pass through the core of the Sun where the nuclear reactions take place which power its evolution. As such, they serve to provide an unprecedented probe of the deep solar interior.

The propagation of sound waves through the Sun is governed by the dispersion relation (Lamb 1916) which links the frequency of oscillation, ω, with the wavevector. The wavevector can be split into a horizontal component, κ_h, and a radial component, κ_z. The dispersion relation thus becomes

$$\kappa_z^2 = (\frac{\omega}{c})^2 - \kappa_h^2 - \frac{1}{4H^2} \tag{1.1}$$

where c is the local speed of sound and H is the scale height due to the atmosphere being stratified. These can be described by

$$c = \sqrt{\frac{\Gamma P}{\rho}}$$

$$H = \frac{RT}{\mu g}$$

where Γ is the local adiabatic modulus, ρ the local density, T the local temperature, P the local pressure, and μ the local mean atomic mass.

If waves are to propagate, the condition $\kappa_z^2 > 0$ must be satisfied. This implies the existence of an acoustic cut-off frequency, ω_{ac}^2, where waves of this frequency or below will not propagate and are evanescent in nature. Equation 1.1 can be re-written as

$$\kappa_z^2 = (\frac{\omega^2 - \omega_{ac}^2}{c^2})$$

where ω_{ac}^2 is given by

$$\omega_{ac}^2 = c^2(\kappa_h^2 + \frac{1}{4H^2})$$

The cavity is formed due to the variation of ω_{ac} with depth within the solar interior. Ulrich suggested that such sound waves travelling towards the centre of the Sun would be progressively refracted away from the normal as the temperature increases until it turns around and returns to the surface, since the deepest part of the wavefront travels faster than the shallowest part. At the top of the convection zone near the Sun's surface the scale height H rapidly decreases as the temperature drops. This sharp density gradient causes the cut-off frequency to increase rapidly until it reaches a maximum of about 5.3mHz, and since the oscillations typically have frequencies of about 3mHz they become evanescent as they approach the photosphere and are subsequently reflected back into the Sun. The restoring forces for the oscillations are local pressure gradients, and hence they are often referred to the p-mode oscillations of the Sun.

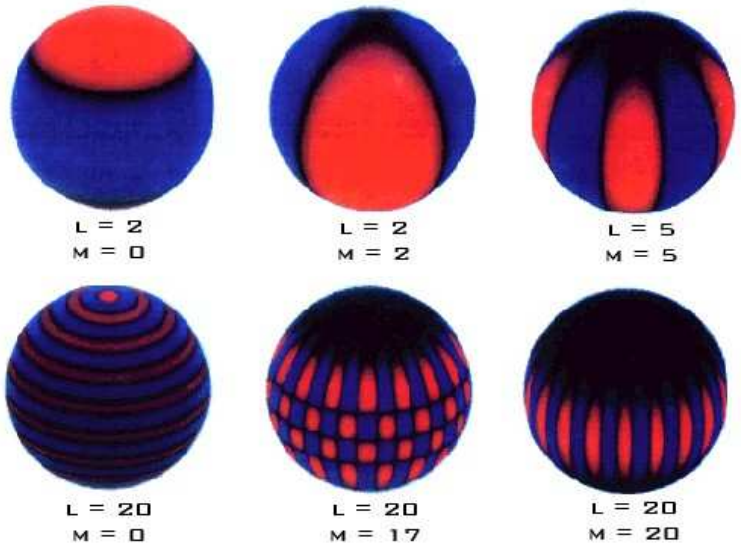

Figure 1.1: Modes of Oscillation. The blue regions are moving towards the observer and the red regions moving away from the observer. If the diagram were animated the coloured regions would be constantly fading between red and blue.

Just as with conventional sound waves in a musical instrument, an acoustic wave in the Sun can constructively or destructively interfere with itself. As a wave progresses around the interior of the Sun if it arrives back where it started in phase with the excitation it will constructively interfere and its amplitude will increase. If not, it will tend to be cancelled out. This results in a number of normal or preferred *modes* of oscillation.

The constructive and destructive interference of the sound waves result in regions of local high and low acoustic pressure within the Sun's interior. As the Sun is a three dimensional object, these pressure nodes are distributed throughout its interior and a numbering convention is used to describe the number of nodes present in a particular standing wave configuration. The letters n, l, and m are used to represent the number of nodes distributed

throughout the spherical geometry of the Sun where n is the number of nodes radially outward from the centre of the Sun, m is the number of nodes found around the equator, and l is the degree of the mode representing the number of nodes found around the azimuth (a great circle through the poles). A particular arrangement of n, m, and l is known as a *mode* of oscillation, figure 1.1. If the Sun were not rotating, modes of the same l and n would be degenerate and have the same frequency, but since it is rotating the degeneracy is lost and modes of the same l and n are split into a multiplet of $2l + 1$ different m values, specifically between $m = -l$ to $m = +l$.

The horizontal component of the wave vector, κ_h, varies with the distance r from the Sun's centre according to

$$\kappa_h = \frac{\sqrt{l(l+1)}}{r} \tag{1.2}$$

where l is the degree of the mode as before. As r decreases and κ_h increases, the temperature, speed of sound, and scale height get very large. The boundary of the cavity is marked by the point at which the wave vector becomes complex. For purely radial modes ($l = 0$), the horizontal wavevector is zero throughout the Sun and so they penetrate to the core and are coherent over large portions of the solar surface. Higher l modes do not penetrate as deep into the Sun, and the highest l modes are essentially just a surface ripple. It is this property that allows the solar interior to be mapped out by observing a large number of modes, since the properties of the sound waves are modified by the conditions within the cavity.

1.2 Data Capture

In order to show the individual frequency components which are responsible for the observed oscillation a standard Fourier analysis is used. The central peak of such an analysis occurs at around 3mHz, or periods of five minutes.

The original discoveries were made by analysing sets of data with durations of several hours only. However, a dataset of about thirty hours is required in order to begin to resolve the 10 μHz spacing between a mode of $l = 0$ and given n, and the mode with $l = 2$, $n - 1$. A dataset of this length can be obtained by adding together contiguous days of data from a single site, however the night-time gaps cause sidebands to appear in the power-spectrum under Fourier analysis. These sidebands are the Fourier transform of the gap structure, and for a spacing of one per day this is approximately 11.6 μHz, which is unfortunately close to the spacing between some of the p modes. For a spectrum to show clearly any rotational splitting requires a dataset of over four months.

It is desirable to have continuous long-term observation of the Sun (Elsworth 1996), not only to eliminate the sidebands caused by the diurnal cycle but to generate the high resolution spectra that only a very long dataset can provide. There are a number of ways this can be achieved.

1.2.1 Observe from the South Pole

At the South Pole the day-light hours are at a maximum, since during the summer the Sun does not set.

However, the Antarctic weather conditions make it difficult and costly to maintain observations. The possible lengths of datasets are limited to only a few months each year, and so it is still impossible to generate the very long sets required to resolve the fine detail in the power spectrum. Despite the problems a French-American expedition in 1980 by Eric Fossat, Gerard Grec, and Martin Pomerantz was able to acquire five days of uninterrupted data and at the time was the longest dataset available (Grec, Fossat, and Pomerantz, 1980).

1.2.2 Observe from Space

From a science perspective, this may appear to be the ideal solution. A spacecraft at the Lagrangian point of the Earth's orbit around the Sun would in principle be free from the diurnal cycle and the Earth's atmosphere. A near one hundred percent duty cycle of low noise data is possible.

From a financial perspective, as with any space project there is enormous expense involved. Equipment operating in such a harsh environment as space is often prone to failure, and since there are very little prospects for repairs to be made even the simplest problem can completely compromise a mission. The mission lifetime is also limited by the efficiency of the craft, with most typically carrying enough fuel to last for only a few years.

This idea has been implemented with the GOLF experiment on the SOHO spacecraft. Despite the nature of the mission they still experience a diurnal effect due to interactions between ground stations.

One of the first large scale space-based missions for detecting global oscillations was the ACRIM experiment (Active Cavity Radiometer Irradiance Monitor) on the Solar Maximum satellite. This detector looked at intensity variations in the photosphere caused by non-adiabatic exchange of energy into the solar plasma from the acoustic waves. The intensity varies at the same frequency as the oscillations as the gas heats and cools. The results showed variations of a few parts per million of the mean intensity (Woodard and Hudson 1983), with peaks in the power spectrum corresponding to modes with periods of oscillation of around five minutes.

1.2.3 Observe using a Terrestrial Network

This is currently a popular option, not only within BiSON but also with other groups involved with solar oscillations such as the Global Oscillations Network Group (GONG 1984, Hill et. al. 1985, Hill 1990), and the International Research of Interior of the Sun (IRIS) network (Fossat 1991).

A network of mid-to-equatorial latitude observatories allows a twenty-four hour view of the Sun as the Earth rotates. Theoretically, assuming no equipment failure or bad weather, this requires only three stations equally separated in longitude. If each station collects eight hours of data per day this completely covers each day with no overlaps and no gaps. However in a not so perfect world a level of redundancy is required to achieve the desired 100% duty cycle.

1.3 The Birmingham Solar Oscillations Network

In 1975 the HiROS group decided to attempt to establish a network of six stations around the Earth. As already discussed, a one hundred percent duty cycle is theoretically possible with just three stations in equally spaced longitudinal bands. However, if we assume that each station only collects data for sixty percent of the day either due to bad weather or breakdowns, then the network duty cycle is also sixty percent. If we now have two stations in each band, separated widely such that losses are uncorrelated, then the total fraction of time that neither station is collecting data is $(1.0 - 0.6)^2$ giving a network duty cycle of eighty four percent. These predictions suggest that obtaining a network duty cycle of over ninety percent would require at least nine stations in total. However, the GONG site-survey indicates a clear-time fraction similar to BiSON and yet they are obtaining duty cycles as high as ninety-three percent from their six-site network (Hill, 1994). It is possible that they are using times when the Sun is lower in the sky whereas BiSON reject the data due to extinction problems; but the most likely cause is that their larger group allows better maintenance of the sites, and hence they have less down-time due to hardware malfunction.

Isaak visited many observatories in the USA to find a suitable site for a second station to complement the existing instrument in Tenerife. Unfortunately, the grant application was rejected by the Science and Engineering Research Council and the group was unable to begin construction.

The group continued collecting data from Tenerife, and during 1978 and 1979 operated a second instrument again at Pic du Midi during minor and

inexpensive expeditions. The independent observations helped to establish that the five minute solar oscillations were indeed a real phenomenon. Later in 1981, after the work of Fossat et. al. at the South Pole, the group secured funding to operate an instrument at Haleakala, Hawaii. Data were collected for up to twenty-two hours per day for eighty-eight days producing the longest timeseries, and more highly resolved power spectrum, than anyone had seen before. The interest in helioseismology started to develop quickly.

Three years later in 1984, almost ten years after the initial concept of a global network was introduced in 1975, the group built the first semi-automatic station in Carnarvon, Western Australia, and was made fully automatic in 1985. Four years later in 1989 a larger instrument based on the Carnarvon prototype was built and based in Birmingham for development, testing, and training.

The six station network of today, figure 1.2, was completed in 1992. There are two stations in each 120-degree longitude band, table 1.1. Moving successively East in longitude: Sutherland and Carnarvon are in band one; Narrabri and Mount Wilson in band two (Mount Wilson sits on the band two - band three boundaries); and Las Campanas and Izaña are in band three. The network yields a fill factor of up to eighty percent (Chaplin 1996). The oldest site in the Birmingham Network has now been collecting data for over twenty five years.

Figure 1.2: The six station Birmingham Solar Oscillations Network.

Location	Longitude (deg E)	Latitude (deg N)	Altitude (metres)	Commissioned (year)
Sutherland, South Africa	+20.82	-32.38	1771	1990
Carnarvon, Western Australia	+113.75	-24.85	10	1985
Narrabri, New South Wales	+149.57	-30.32	217	1992
Mount Wilson, California	-118.08	+34.13	1742	1992
Las Campanas, Chile	-70.70	-29.02	2282	1991
Izaña, Canary Islands	-16.50	+28.30	2368	1975

Table 1.1: The six station Birmingham Solar Oscillations Network

Chapter 2

BiSON Instrumentation

The main aim of any instrument used for detecting solar oscillations is to be able to convert from a measurement of light intensity to a measurement of velocity. Solar oscillations are very low amplitude compared with the 'carrier' signal, and as such detectors need to be very sensitive. The lifetime of modes can be several days, and hence a detector needs to be very stable such that the response of the instrument does not change from day to day.

The method chosen by the Birmingham Solar Oscillations Network is resonant scattering spectroscopy.

2.1 Resonant Scattering Spectroscopy

The fundamental principles of operation have remained unchanged since the network was first commissioned. As the solar surface rises it causes the radial component of velocity to decrease with respect to the observer. Similarly, when it falls the radial velocity will increase. This motion causes absorption

lines in the solar spectrum to be successively blue and red-shifted. The aim is to measure Doppler shifts in a solar Fraunhofer line (averaged over the whole disc) with sufficient precision and stability to study any large scale oscillations of the Sun.

2.1.1 Choice of Fraunhofer Line

The choice of Fraunhofer line falls between one of two elements, with all resonance spectrometers in solar instruments using either potassium or sodium. Ideally we want a line on the Sun which is strong and clean. The Na lines are deeper and are optical, with the 'D1' line having a wavelength of 589.6nm. The K 'D' lines are weaker and in the near infra-red, with the 'D1' line having a wavelength of 769.9nm. Historically, optical lines were preferred since they are in the optimum sensitivity range of photomultiplier tubes. However, newer photo detectors are more sensitive in the infra-red.

There are also a number of other benefits associated with K lines. Firstly, the K 'D' lines are further apart at 3.4 nm compared to the sodium D lines at 0.6 nm, and this makes them easier to separate. Secondly, unlike Na the K D1 lines are not blended with any other solar lines and are not confused with absorption within the Earth's atmosphere. Finally, one of the K lines features steeper sides than either of the Na lines. This makes any instrument using K intrinsically more sensitive than one using Na, since the same line displacement gives a greater change in measured absorption intensity.

For these reasons all BiSON spectrometers observe the D1 transition in neutral potassium. The instruments used by the IRIS network (Fossat 1988, 1991), and those used by the GONG network (GONG 1984) both use the sodium D1 line (Grec et. al. 1991, Harvey J. 1998).

2.1.2 Doppler Imaging

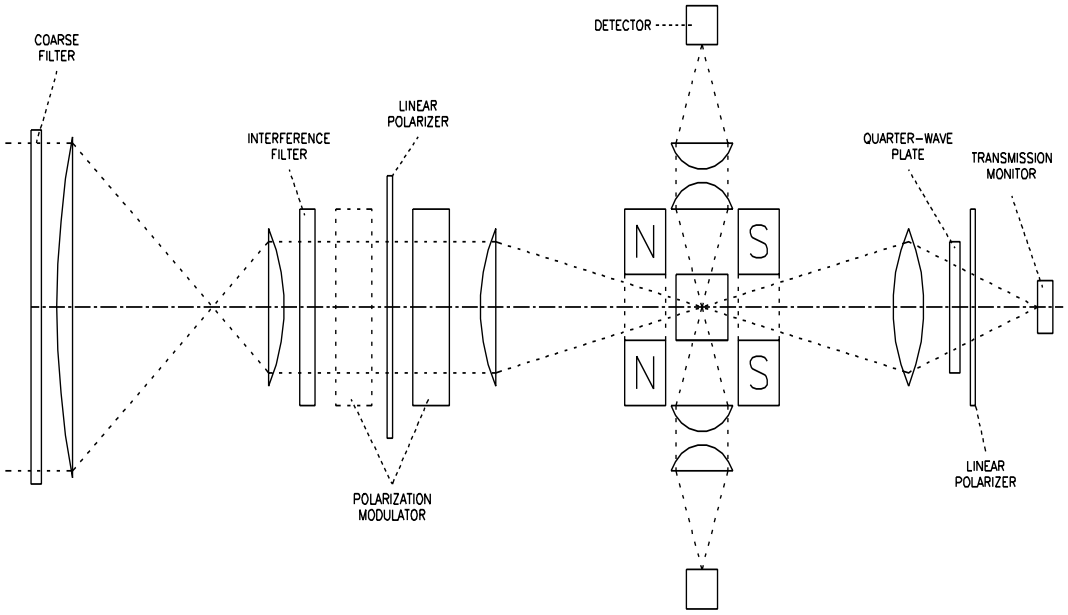

Figure 2.1: Schematic of a BiSON Resonant Scattering Spectrometer.

Sunlight is first passed through a color filter to remove the infra-red and relieve the thermal load on the instrument. An interference filter with a bandpass of approximately 1nm is then used to further reduce the frequency range down to that of the Fraunhofer line of interest.

At the heart of all BiSON instruments is a cell containing potassium vapour heated to 400K in a longitudinal magnetic field. This acts as the wavelength selection device, figure 2.1.

The magnetic field causes Zeeman splitting of the single line into a multiplet, with the two peaks corresponding to both senses of circularly polarised light. The field strength is chosen such that the two Zeeman components sit on the wings of the absorption line. This means that the instrument does not look at the centre of the absorption line since this would provide very little change in measured intensity for a given Doppler shift. Rather, it looks at the wings of the line where the sides are steepest giving the greatest change in intensity.

In order to take advantage of this we need to circularly polarise the light from the Sun. This is accomplished by passing the light through a linear polarizing filter and then onto an electronic implementation of a quarter wave plate, such as a Pockels' cell. The sense of the resulting circularly polarized light is changed by varying the voltage across the cell.

When light within the chosen passbands is incident on the cell, photons are absorbed raising the atoms to their excited state. The excited atoms then spontaneously decay to the ground state by emission of a photon, but since these photons have no directional preference the vapour radiates isotropically into 4π steradians. Two detectors placed on the port and starboard sides of the cell pick up a percentage of the scattered light. By switching the voltage on the Pockels' cell it is possible to observe either the red-wing or the blue-wing depending on whether the cell is acting as either a '$+\frac{\lambda}{4}$ wave plate' or a '$-\frac{\lambda}{4}$ wave plate'. The oldest station in the network modulates the Pockels' cell at a frequency of 0.5Hz. However newer stations modulate the cell at up to 100Hz which gives an improvement in data quality, since the effects of atmospheric seeing variations are greatly reduced.

Some instruments also measure the intensity transmitted through the spectrometer. This is a measure of intensity over a much larger wavelength compared to the width of the line and so gives an indication as to the clarity of the atmosphere.

The BiSON spectrometers integrate sunlight for 3.2 seconds (giving a total of 1.6 seconds per wing) and then pause for 0.8 seconds while the information is read-out from the detector. Thus the differential ratio

$$\rho = \frac{I_R - I_B}{I_R + I_B} \tag{2.1}$$

is calculated and stored every four seconds, where I_R and I_B are the measured intensity of the red and blue wings respectively. The difference between the two intensities is normalised to give stability to the measurement. This reduces variations in noise level and overall sunlight intensity such as when the atmospheric extinction changes during the observing period, for instance at sunrise and sunset.

At the end of each day every ten points of four second data are averaged to produce forty second data in order to reduce storage requirements and download time when the data are transferred back to Birmingham. The nyquist limit states that you must sample a signal at at least twice the frequency you wish to observe. The forty second cadence enables frequencies of up to 12.5mHz to be studied, which is well above the 1mHz to 5mHz band where the solar oscillations peak in amplitude.

2.2 Data Calibration

There are a number of methods for calibration of the residuals (Chaplin 1995). There is no one 'best' method, but rather the method chosen is determined by the use to which the data are to be put. The standard analysis regards the correct calibration of the five-minute residuals as the primary objective. This means that the amplitude of the residuals should be correctly scaled, and should be free from any low frequency drifts or offsets.

This section covers the 'standard method' of BiSON data reduction and analysis (Elsworth 1995, Chaplin 1997). Methods used by other groups can be found in the relevant papers, for example: IRIS - Pantel 1995, Salabert 2002, GONG - Anderson 1989, Trueblood 1991, etc.

2.2.1 The Standard Method

Ideally we would like to convert the ratio ρ (eq. 2.1) directly into line-of-sight velocity of the residual solar oscillations. Unfortunately this is not a simple operation since the relationship is not linear due to the variation of the gradient ($\frac{\delta I}{\delta \lambda}$) of the Fraunhofer line.

The ratio is expanded as a polynomial of the form

$$\rho(v) = \sum_{i=0}^{N} (a_i (v - v_{grs})^i) \tag{2.2}$$

where v is the topocentric solar velocity, and v_{grs} is a variable offset related to the gravitational redshift of General Relativity.

The number of terms used in the series must be chosen with care. If there are not enough terms, the resolution is too poor to follow the change in sensitivity throughout the day. If too many terms are used the fit becomes unstable, and the solar oscillations can be lost along with all the other components when the series is subtracted from the measured values.

There are four components making up the velocity v, such as the rotation of the Earth on its axis and the orbital motion, all of which must be determined before the solar oscillations can be extracted, (Brookes et. al., 1976).

2.2.2 Gravitation Redshift - v_{grs}

This phenomenon exists since a photon must do work in travelling from the Sun to the Earth. In doing work, the photon's associated energy decreases, resulting in a shift to a longer wavelength. For the K D1 transition the magnitude of this apparent velocity is around -632ms^{-1}.

2.2.3 Orbital Velocity - v_{orb}

This component arises from the Earth's motion around the Sun. Since the orbit is not circular, the eccentricity causes a relative line-of-sight velocity as the Sun-Earth distance changes. There is also a small component due to the Sun's motion which is a reaction caused by the orbiting planets. The maximum velocities of $\pm$500ms^{-1} are achieved in April and October. An overall change of 1000ms^{-1} in six months equates to just over 5ms^{-1} per day, and so for any given day the effect is regarded as constant. Values for v_{obs} can be obtained from a standard Almanac. The BiSON uses a lookup table

of values taken from the JPL Ephemeris, using the supplied first and second derivatives to interpolate between daily values.

2.2.4 Rotation Velocity - $v_{rotation}$

It is this effect that dominates the daily signal, and creates the familiar sinusoidal shape of the raw data. As the Earth rotates about its axis, an observer is approaching the Sun in the morning, has no line-of-sight velocity at midday, and is receding from the Sun in the afternoon. The maximum effect of this occurs on the equator with approximately $\pm 460 \mathrm{ms}^{-1}$, but the overall effect also depends on the observer's altitude and the declination of the Sun. It is described by an equation of the form (Elsworth et. al. 1995)

$$v_{rotation} = \omega_E R_E(\lambda) \cos\lambda \cos\delta \sin(\frac{\pi}{12}(t - t_0))$$

where ω_E is the angular velocity of the Earth, $R_E(\lambda)$ the observer's distance from the centre of the Earth, λ the latitude, t_0 the time of local noon, t the time of the observation (both expressed in hours), and δ the declination of the Sun. As with the orbital velocity, this is calculable using standard ephemeris algorithms.

2.2.5 Solar Oscillations - v_{osc}

This is the mission objective.

Individual solar oscillation velocities are typically just a few metres per second at most, which is a tiny fraction of the total observed signal - usually

only 0.1 percent. The oscillations have a much shorter periodicity than the length of the dataset and so average out to zero. This has the advantage of meaning they can simply be regarded as producing the residuals from the fit of the power series to the ratio data. Hence the solar oscillations are known colloquially as simply *the residuals*.

2.2.6 Other - v_{other}

This component comprises solar, terrestrial, and instrumental noise.

Solar noise arises from the random turbulent motions of pockets of gas, and the so-called convective blueshift. Hot gas rises and cool gas falls, but the hot gas is brighter and hence is unfairly weighted in the integrated sunlight resulting in a net blueshift at disc centre.

Terrestrial noise results from atmospheric seeing, and differential extinction gradients across the solar disc at the beginning and end of the day. This is where one part of the disc suffers greater extinction than the rest, resulting in unfair weighting to one part of the Sun and a net velocity signal will arise due to the Sun's rotation.

Instrumental noise can arise for a number of different reasons. If the instrument is misaligned and part of the beam is blocked the solar image will become disrupted, resulting in an effect similar to that of having differential extinction gradients. Noise could also occur from unfair weighting of each absorption wing either optically or electronically, for instance by not having the Zeeman passbands centred correctly. It is for this reason that a single detector is used to look at both wings even though it means discarding half

of the available signal, since it removes any problems arising from different gains in different detectors.

While good site maintenance can help eliminate any instrumental noise there is very little that can be done about solar or terrestrial noise, and these are currently ignored during the calibration.

2.2.7 The Conversion Factor

Once both v_{orb} and $v_{rotation}$ have been calculated, a best fit algorithm is used to determine the coefficients a_i, and v_{grs} in equation 2.2. The power series is then subtracted from the measured ratio

$$\rho_{osc} = \rho_{measured} - \sum_{i=0}^{N}(a_i(v - v_{grs})^i))$$

to leave only the oscillations and any associated noise.

The conversion factor between ratio and velocity, also known as the sensitivity, is given by $(\frac{\delta\rho}{\delta v})_v$. Once this value has been calculated, converting from ratio to velocity is simply a matter of dividing by the sensitivity. As already discussed, the ratio is only approximately a linear measure of the relative line-of-sight velocity of the solar surface. Hence, the sensitivity is a function of the orbital velocity and is typically about 3000ms^{-1} per unit ratio.

Historically a three term quadratic ($N = 3$) has been used, but more recently a four-term cubic ($N = 4$) has been found to provide a more accurate fit and removes many unwanted low frequency effects.

Once this process is complete the residuals have to be combined into an extended timeseries, taking into account gaps in the data and also any multi-station overlaps. The series can then be Fourier transformed into the frequency domain allowing the mode characteristics to be analysed.

Chapter 3

Instrumental Errors

3.1 Theoretical Noise Sensitivity

From equation 2.1 we know that the ratio of intensity in the red wing to the intensity in the blue wing of the Fraunhofer line is defined as

$$\rho = \frac{I_R - I_B}{I_R + I_B}$$

where I_R and I_B are the measured intensity in photons per second of the red and blue wings respectively. However, this is the 'ideal' scenario since realistically these values also contain background signals from non-resonantly scattered light (I_R^{non} and I_B^{non}), systematic errors such as electronic offsets (I_R^{sys} and I_B^{sys}), and other noise (i_R and i_B) as well as the desired contribution from resonantly scattered light (I_R^{res} and I_B^{res}). Hence the ratio is more accurately described by (New 2003)

$$\rho = \frac{I_R^{res} + I_R^{non} + I_R^{sys} + i_R - I_B^{res} - I_B^{non} - I_B^{sys} - i_B}{I_R^{res} + I_R^{non} + I_R^{sys} + i_R + I_B^{res} + I_B^{non} + I_B^{sys} + i_B}$$

It is reasonable to assume that the non-resonantly scattered light and the systematic errors are equal in both the red and blue wings and so the above equation can be simplified to

$$\rho = \frac{I_R^{res} - I_B^{res} + i_R - i_B}{I_R^{res} + I_B^{res} + 2(I^{non} + I^{sys}) + i_R + i_B}$$

It is also reasonable to assume that the signals due to resonantly scattered light are always much greater than the sum of the other contributions, so the denominator can be replaced by a Taylor expansion, thus

$$\rho = (I_R^{res} - I_B^{res} + i_R - i_B)\frac{1}{(I_R^{res} + I_B^{res})}\left(1 - \frac{2(I^{non} + I^{sys})}{(I_R^{res} + I_B^{res})} - \frac{(i_R + i_B)}{(I_R^{res} + I_B^{res})}\right)$$

Rewriting this in terms of the ideal ratio and again making the assumption that the resonant scattering signals are much greater than all others we find

$$\rho = \rho_{ideal}\left[1 - \frac{2(I^{non} + I^{sys})}{(I_R^{res} + I_B^{res})}\right] + \frac{2(I_B^{res} i_R - I_R^{res} i_B)}{(I_R^{res} + I_B^{res})^2} \tag{3.1}$$

where

$$\rho_{ideal} = \frac{I_R^{res} - I_B^{res}}{I_R^{res} + I_B^{res}}$$

The first term in equation 3.1 would be equal to the ideal ratio, equation 2.1, only if there were no systematic errors or non-resonantly scattered light. As these background levels increase the sensitivity of the instrument decreases.

The second term contains the noise dependence, and it is useful to simplify this by defining

$$\delta\rho = \frac{2(I_B^{res} i_R - I_R^{res} i_B)}{(I_R^{res} + I_B^{res})^2} = x i_R + y i_B \tag{3.2}$$

where

$$x = \frac{2I_B^{res}}{(I_R^{res} + I_B^{res})^2}$$

and

$$y = \frac{-2I_R^{res}}{(I_R^{res} + I_B^{res})^2}$$

3.1.1 Uncorrelated Uniform Noise

If i_R and i_B are uncorrelated then the two independent contributions add in quadrature such that

$$\delta\rho = \sqrt{x^2 i_R^2 + y^2 i_B^2} \tag{3.3}$$

and if we assume that the magnitude of i_R and i_B are both equal to i then the above equation can be simplified to

$$\delta\rho_{uncor} = i\sqrt{x^2 + y^2}$$

3.1.2 Correlated Uniform Noise

Alternatively, if i_R and i_B are completely correlated such that $i_R = i_B = i$ then equation 3.2 becomes

$$\delta\rho_{cor} = i(x + y) \tag{3.4}$$

3.1.3 Photon Shot Noise

The effect of photon shot noise is governed by Poisson statistics since it is a random process, and hence the noise contribution to the resonantly scattered signals I_R^{res} and I_B^{res} will be $\sqrt{I_R^{res}}$ and $\sqrt{I_B^{res}}$ respectively. If we assume that the shot noise contributions are dominant then in equation 3.2, $i_R = \sqrt{I_R^{res}}$ and $i_B = \sqrt{I_B^{res}}$. The two values will not be correlated so the expected mean noise obeys a form such as that in equation 3.3, thus

$$\delta\rho_{shot} = \sqrt{x^2 I_R^{res} + y^2 I_B^{res}}$$

Substituting the definitions of x and y from equation 3.2 back into the above expression we obtain

$$\delta\rho_{shot} = 2\sqrt{\frac{I_R^{res} I_B^{res}}{(I_R^{res} + I_B^{res})^3}}$$

If we assume that the magnitude of the contribution from I_R^{res} and I_B^{res} are equal to a value I^{res}, and include an integration time τ and a velocity calibration constant k then this can be simplified to

$$\delta v_{obs} = 2k\sqrt{\frac{1}{8\tau I^{res}}} \tag{3.5}$$

The velocity calibration constant k is approximately 3000ms^{-1} and as already discussed, the effective total integration time is forty seconds such that $\tau = 20$ seconds per wing. The maximum photon counting rate at Tenerife is 10^7 photons per second (Brookes et. al., 1978), and a typical estimate for I^{res} is around 5x10^5 photons per second.

Using these values equation 3.5 gives an approximation for the standard deviation due to shot noise as 0.67ms^{-1}.

3.2 Empirical Noise Sensitivity

We can also determine empirically the standard deviation of the residual data by looking at the total noise power in the frequency domain. In the time domain, the magnitude of the standard deviation, σ, is determined by the distribution of noise-source velocities v_i, such that for N points in the timeseries

$$\sigma^2 = \sum_{i=1}^{N} \frac{(v_i - \overline{v})^2}{(N-1)}$$

If the timeseries of white noise possesses a zero mean level, then for N $\gg$ 1, the right hand side of the above equation is simply the sum of the power in the frequency domain, and this is known as Parseval's theorem.

The content of the residual solar oscillations can be simplified by thinking of them as having three components. These are the low frequency oscillations, the five minute oscillations, and high frequency noise. If we assume that the noise actually permeates the whole spectrum, but that it is masked by the signal power, then we can use Parseval's theorem to calculate the standard deviation.

$$\sigma^2 = \sum Power = 2.P.f_{max}$$

where P is the average noise power per Hz, and f_{max} is the nyquist limit which as previously discussed is 12.5mHz for a forty second sample interval. The factor of two is to take into account the symmetrical nature of the Fourier transform.

Location	Average Noise Power, P $(\mathrm{ms}^{-1})^2\ \mathrm{Hz}^{-1}$	Variance, σ^2 $(\mathrm{ms}^{-1})^2$	Standard Deviation, σ (ms^{-1})
Sutherland	3.21	0.080	0.28
Carnarvon	11.3	0.283	0.53
Narrabri	2.59	0.065	0.25
Mount Wilson	9.95	0.250	0.50
Las Campanas	2.10	0.053	0.23
Izaña	20.8	0.519	0.72

Table 3.1: Average noise power and standard deviation for each station on the network, as computed from the power between 5.5 to 12.5mHz.

This was done for all six stations using one month of data, assuming the region between 5.5mHz and 12.5mHz to be a measure of the average noise power which was then extrapolated to fill the entire spectrum, table 3.1

The worst station is Izaña in Tenerife with a value of $0.72\mathrm{ms}^{-1}$. If we compare this with the theoretical value we calculated earlier of $0.67\mathrm{ms}^{-1}$ it is clear that the Tenerife instrument is shot noise limited.

The newer stations have larger apertures with a photon flux approximately 100 times greater in magnitude giving a theoretical standard deviation from equation 3.5 of $0.06\mathrm{ms}^{-1}$. This is much lower than the noise level of any of the remaining five stations so they must be limited by some other source of noise such as atmospheric effects.

Chapter 4

Data Quality and Rejection Thresholds

Naturally the quality of data from each site on the network is quite variable; no two days have exactly the same weather and atmospheric conditions. In principle the very worst data are excluded at the first stage of the data processing where the residuals are produced from the raw data. However, it is still useful to have some way of quantifying the quality of data from each station, not only as a check on the rejection routines in stage one, but also to help in combining data where output from more than one station is available at the same time. Furthermore if data from one station are continually being flagged as bad over a number of days this can serve as a useful alarm indicating something may have gone wrong.

In setting a data quality rejection threshold we are attempting to balance the undoubted advantage of adding in more signal by using data, and the possibly disproportionate disadvantage of adding more noise due to low data quality. However we need to consider the question of at what point do the

gains of rejecting more bad data using a lower noise threshold become offset by the decrease in fill level (and correspondingly larger gaps) of the long timeseries? The only way of determining this is by analysis of empirical results.

4.1 Methods

There are two methods for determining data quality currently implemented.

4.1.1 Mean Noise Power

There are two regions of noise that can be used in this analysis, either low frequency or high frequency. Low frequency is defined as being below five minutes, usually 0.8 to 1.3mHz. Similarly, high frequency is defined as being above five minutes, usually 5.5 to 12.5mHz. The low frequency regime is significant when signals due to low frequency modes are being sought. The high frequency regime is mainly used as a check on the overall data quality.

4.1.2 Figure of Merit

The figure of merit is basically a signal-to-noise ratio estimate. The typical five-minute-FOM is calculated by comparing the power in the five minute region (2.0 to 5.0mHz) with the high frequency noise power (5.5 to 12.5mHz). Other passbands can be selected depending on requirements. This method is mainly used to rank data in order of quality when combining data from overlapping stations.

4.1.3 Typical Values

In order to determine the 'correct' rejection threshold, we first need to ascertain the typical values produced using both methods for each site on the network.

The mean low frequency noise power and the five minute figure of merit were calculated per day for each station between 1995 and 2001. The results are shown in figures 4.1, 4.2, 4.3, and 4.4.

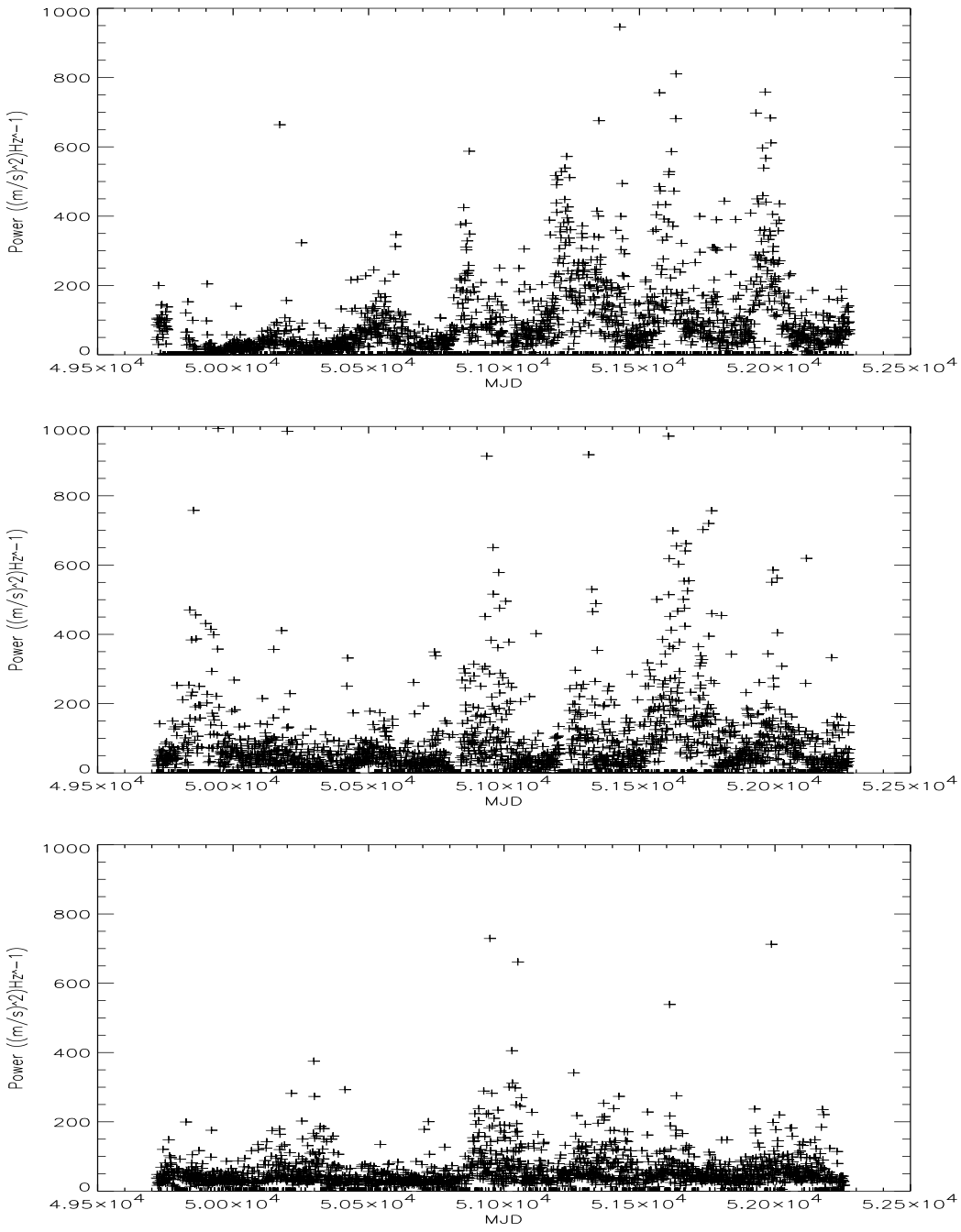

Figure 4.1: Daily low frequency noise power per day for Carnarvon, Sutherland, and Izaña, over the period January 1995 to December 2001. Lower values are better.

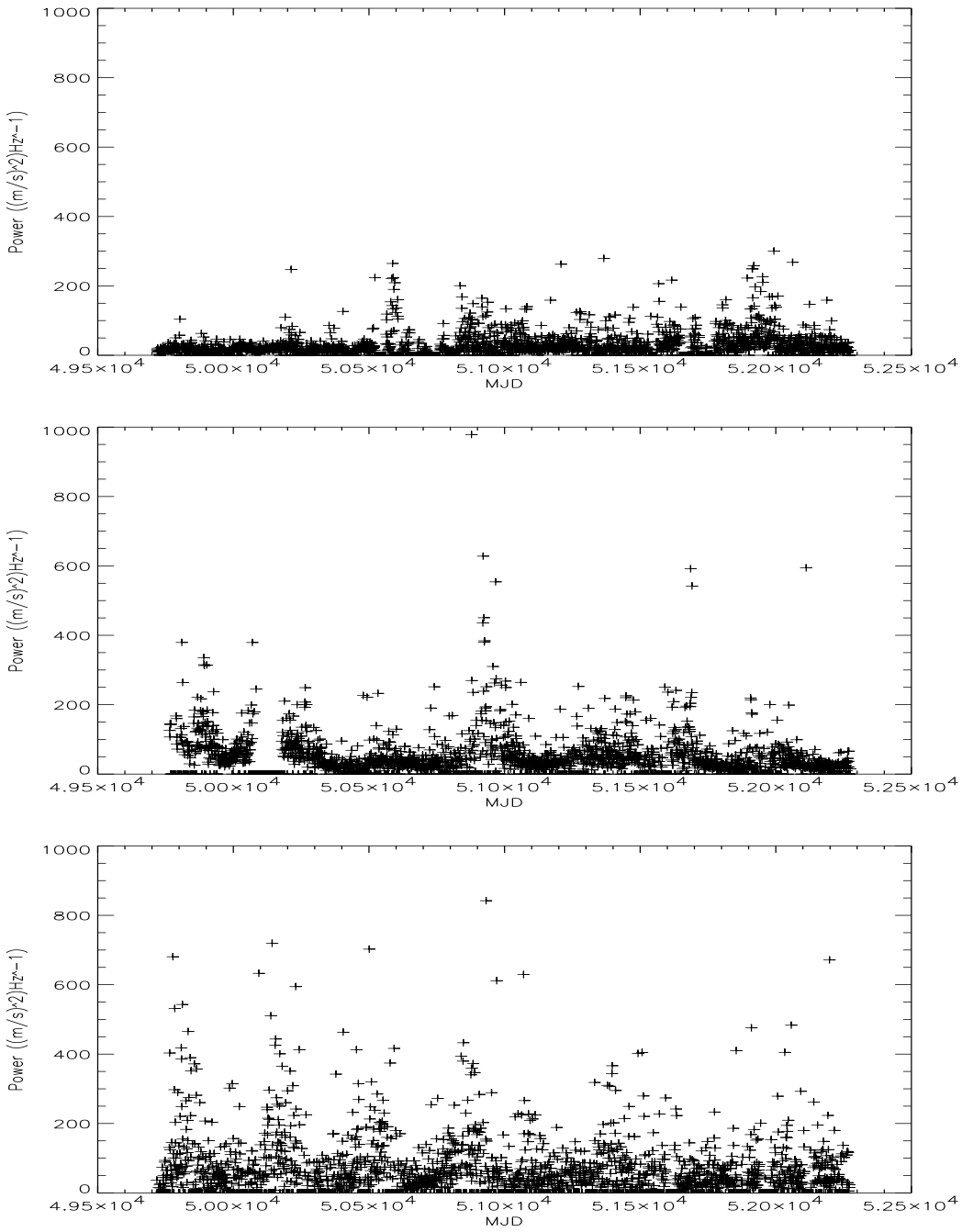

Figure 4.2: Daily low frequency noise power per day for Las Campanas, Mount Wilson, and Narrabri, over the period January 1995 to December 2001. Lower values are better.

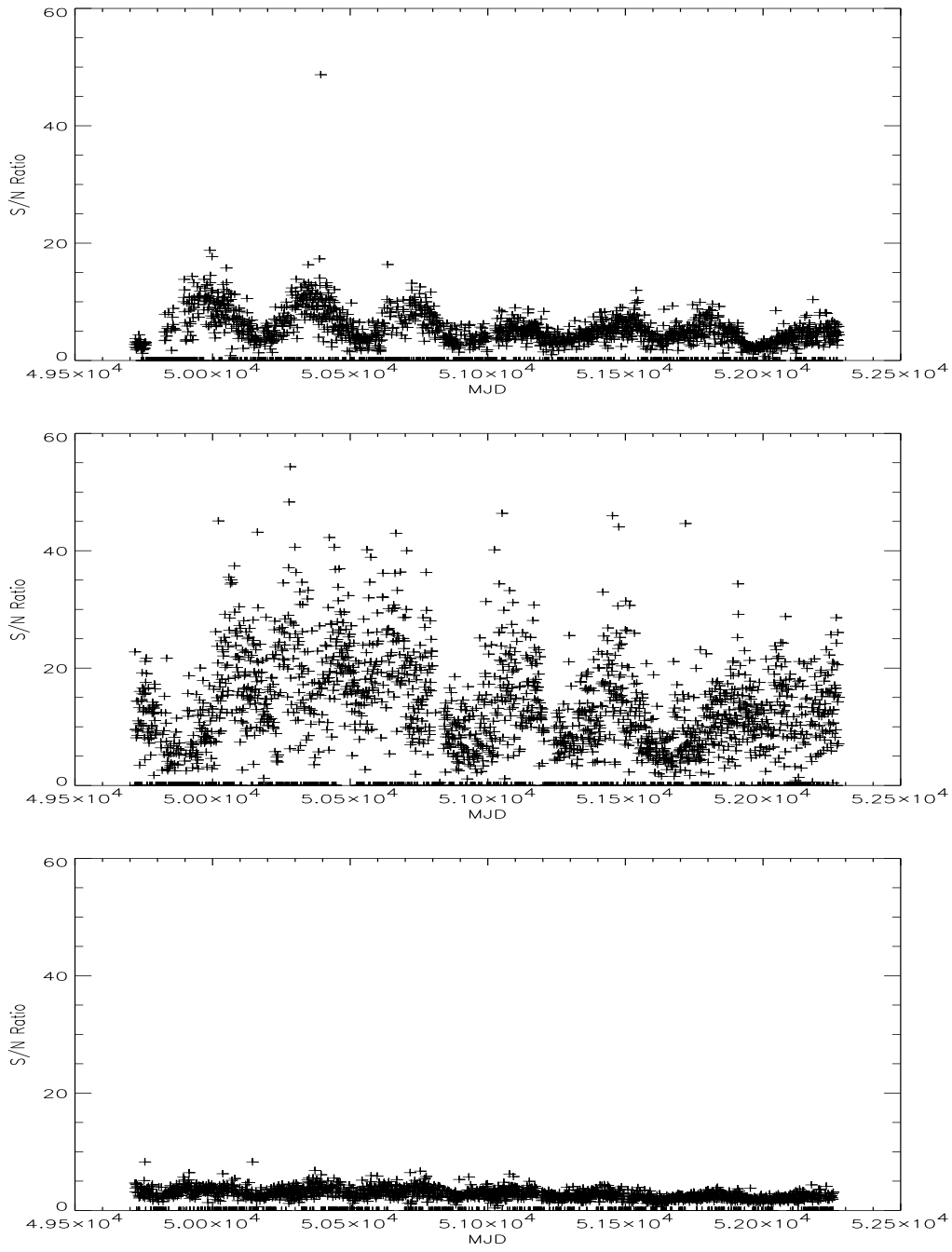

Figure 4.3: Daily five minute figure of merit per day for Carnarvon, Sutherland, and Izaña, over the period January 1995 to December 2001. Higher values are better.

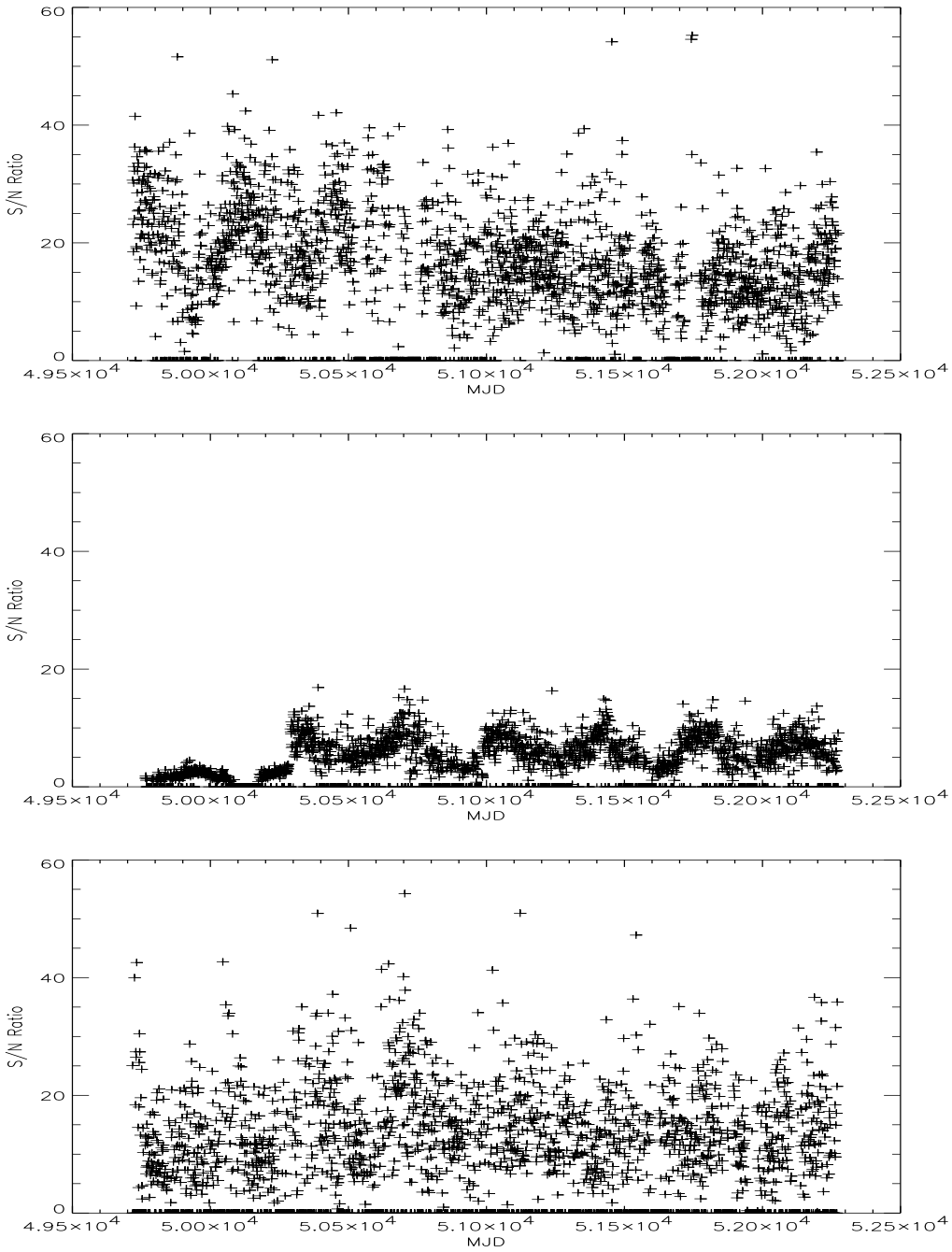

Figure 4.4: Daily five minute figure of merit per day for Las Campanas, Mount Wilson, and Narrabri, over the period January 1995 to December 2001. Higher values are better.

Location	Low Noise Power $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$	High Noise Power $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$	FOM
Sutherland	56	4	13
Carnarvon	65	13	5
Narrabri	61	3	13
Mount Wilson	43	8	6
Las Campanas	23	2	17
Izaña	44	21	3

Table 4.1: Median noise powers and FOM per station.

All stations, most notably Sutherland, Las Campanas, and Izaña show quite clearly a yearly variation in noise level. Noise present in the resonant scattered intensity measurements obviously results in noise in the velocity residuals, but how this noise affects the residuals depends on the operating point on the Fraunhofer line, (New 2003). The mean velocity noise peaks in March every year.

Mount Wilson was refurbished in 1996 giving an improvement in high frequency noise performance. This has produced the improvement in average figure of merit from two to six seen in figure 4.4 from 1996 onwards.

There are a disturbing number of noise peaks way above the median values shown in table 4.1 caused by bad data that should have been rejected during the residual calibration stage. If left unchecked they will have an adverse effect on any long timeseries and associated power spectra produced from these data, as we shall see.

4.2 Analysis of Rejection Thresholds

l	n	Frequency (μHz)
0	6	972.75
0	7	1118.15
0	8	1263.52
0	9	1407.63
0	10	1548.52
1	6	1039.56
1	7	1185.62
1	8	1329.70
1	9	1472.97
1	10	1612.73
2	6	1105.17
2	7	1250.73
2	8	1394.71
2	9	1535.99
3	8	1451.06
3	9	1591.55
3	10	1729.22
3	11	1865.25

Table 4.2: Model predicted frequencies of low-l modes. l=0 have one peak on listed ν. l=1 have two peaks at $\pm 0.4\mu$Hz around listed ν. l=2 have three peaks, one on listed ν, with two others at $\pm 0.8\mu$Hz. l=3 have four peaks, two at $\pm 0.4\mu$Hz, and two at $\pm 1.2\mu$Hz from listed ν.

The most interesting modes of oscillation available to the BiSON are those of low-l, since as we have already discussed these probe the very core of the Sun. It would therefore seem logical to use the clarity of these modes as a measure of the quality of a long timeseries. We explore modes with frequencies in the range of between 1 and 2mHz many of which are on the limit of detectability, and so a computer model of the oscillations was used to produce a list of predicted frequencies for eighteen low-l modes, table 4.2.

These modes could then be studied individually for each timeseries and a qualitative decision made on the number of modes that are resolved clearly against the background noise level. Given that these modes are only theoretical predictions they may not match up completely with the actual modes, but several of the modes have been detected and their frequencies agree to within a few tenths of μHz with these theoretical frequencies (Toutain & Kosovichev 1998). More importantly, from observation and theory there is no reason to believe that any of the modes should not exist at all.

A long timeseries was produced using all available data for the period 1995 to 2001, with no rejections made on the basis of quality. Any overlaps were treated by simply selecting the best site based on the five minute figure of merit. The very bad days noted earlier manifested themselves as delta functions in the timeseries, and caused a high intensity white noise in the power spectrum. The five minute FOM of the entire long timeseries is only slightly above unity, meaning that the power spectrum has been completely destroyed. Clearly in order to ensure only data of a reasonable quality are included in a timeseries a rejection threshold is required.

Twelve further timeseries were produced using varying rejection thresholds based on the mean low frequency noise. The highest threshold of 400 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ was chosen as being considerably higher than the mean level for any one station such that only the very worst days would be rejected. Similarly, the lowest threshold of 15 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ was chosen as being considerably lower than the mean level for any one station such that the majority of data would be rejected.

The results of this analysis are shown in full in appendix A. Here we will look at the effects of the various rejection thresholds on just one mode of oscillation instead of all sixteen in order to simplify the comparison between each threshold. The l=1, n=10, mode at 1612.73μHz is shown in figure 4.5 for each rejection threshold. Some statistics about each timeseries are given in tables 4.3, and 4.4.

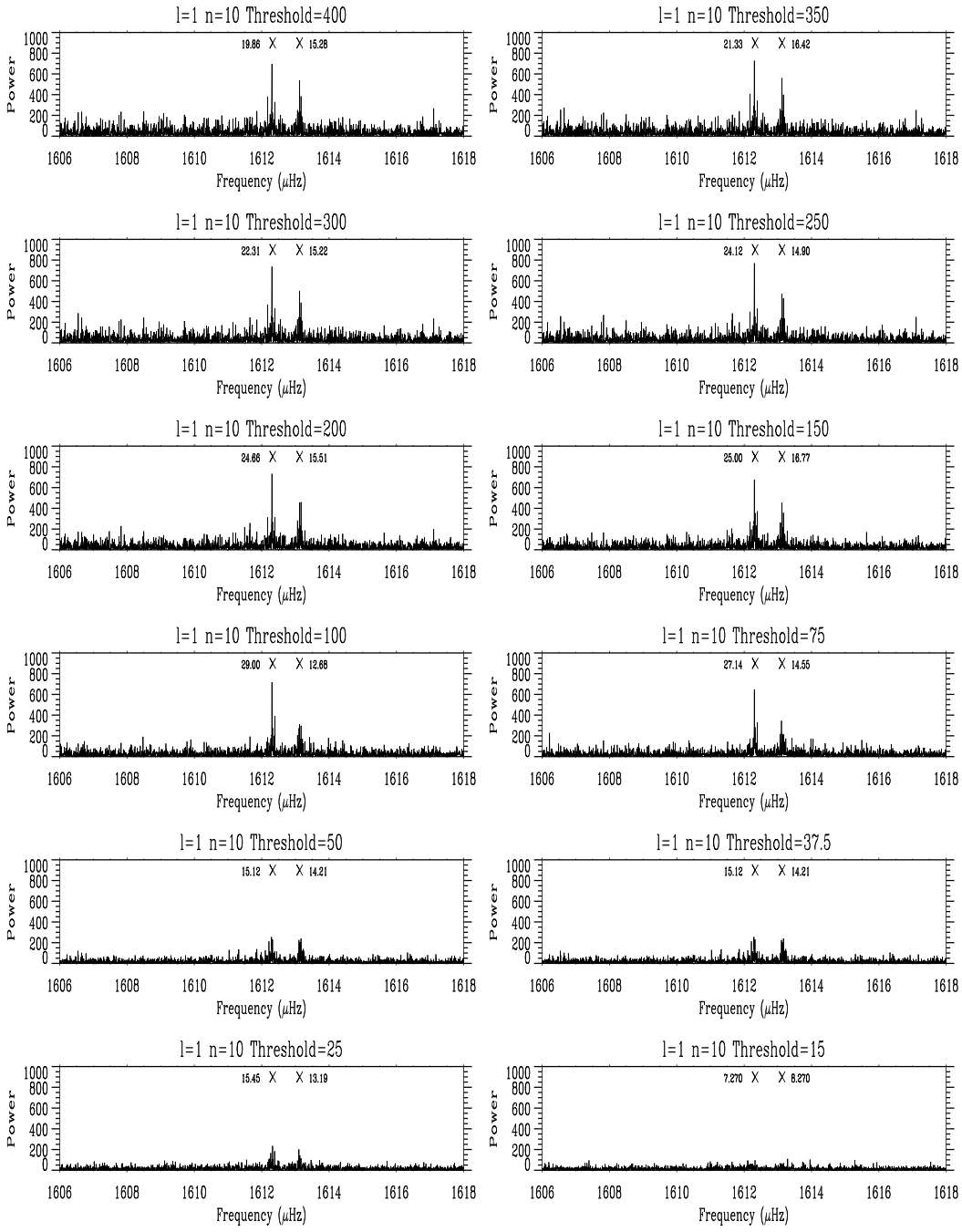

Figure 4.5: The effects on detectability of the l=1, n=10, mode at 1612.73μHz for different rejection thresholds. The theoretical position of each component of the mode is marked with an 'x', and the figures are the signal to noise ratio of each component.

Threshold	Fill	SU	CA	NA	MO	LA	IZ
400	79.4	18.4	14.2	15.1	11.7	30.5	10.1
350	79.1	18.2	14.2	15.0	11.8	30.6	10.2
300	78.8	18.2	14.0	15.0	11.8	30.7	10.3
250	78.3	18.0	14.0	14.9	11.8	30.8	10.5
200	77.4	18.0	13.6	14.7	11.8	31.0	10.9
150	75.5	17.5	13.0	14.4	11.9	31.7	11.5
100	71.5	16.2	12.8	13.3	12.2	32.9	12.6
75	65.9	15.3	12.3	12.3	12.4	34.3	13.4
50	54.6	13.4	11.4	9.8	12.8	38.5	14.1
37.5	44.0	12.3	11.6	8.40	12.7	42.7	12.3
25	26.9	10.0	10.8	6.3	12.3	53.9	6.7
15	8.9	5.6	11.2	4.50	7.90	68.5	2.3

Table 4.3: Percentage fill per station per threshold.

Threshold	2-Site	3-Site	4-Site
400	29.0	6.63	0.187
350	28.8	6.59	0.186
300	28.6	6.53	0.185
250	28.2	6.35	0.182
200	27.4	6.03	0.169
150	25.9	5.45	0.138
100	22.9	4.22	0.097
75	19.5	3.02	0.066
50	12.8	1.46	0.029
37.5	7.67	0.68	0.008
25	2.42	0.07	0.001
15	1.51	0.01	0.000

Table 4.4: Percentage of two, three, and four station overlaps per threshold.

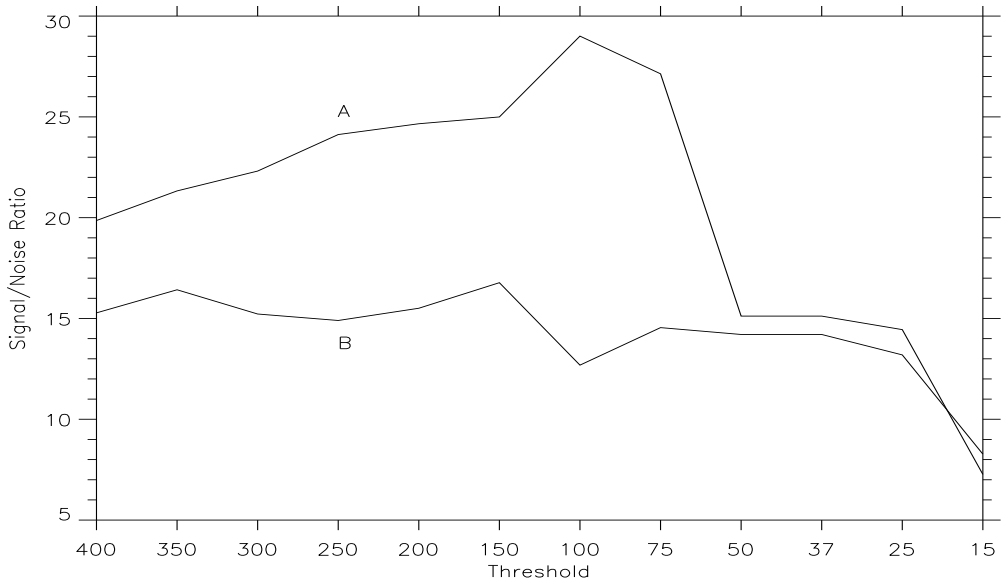

Figure 4.6: The signal to noise ratio of each component of the l=1, n=10, mode at 1612.73μHz for different rejection thresholds. The 'A' and 'B' traces correspond to the m=-1 and m=+1 components respectively.

Using figure 4.6 we can make a quantitative decision as to which rejection threshold provides the best results. The signal to noise ratio of each component was calculated by comparing the power in the peak of the component to that of the background noise level surrounding it. Clearly the best result for the stronger component occurs with a threshold of around 100 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$. The S/N ratio increases up to this point and then decreases as the fill becomes very low, as expected. The lower strength component shows very little change for the higher thresholds, but still drops off rapidly as the fill decreases when using lower thresholds. This is because as the fill decreases the power from the mode is aliased out to other frequencies known as sidebands.

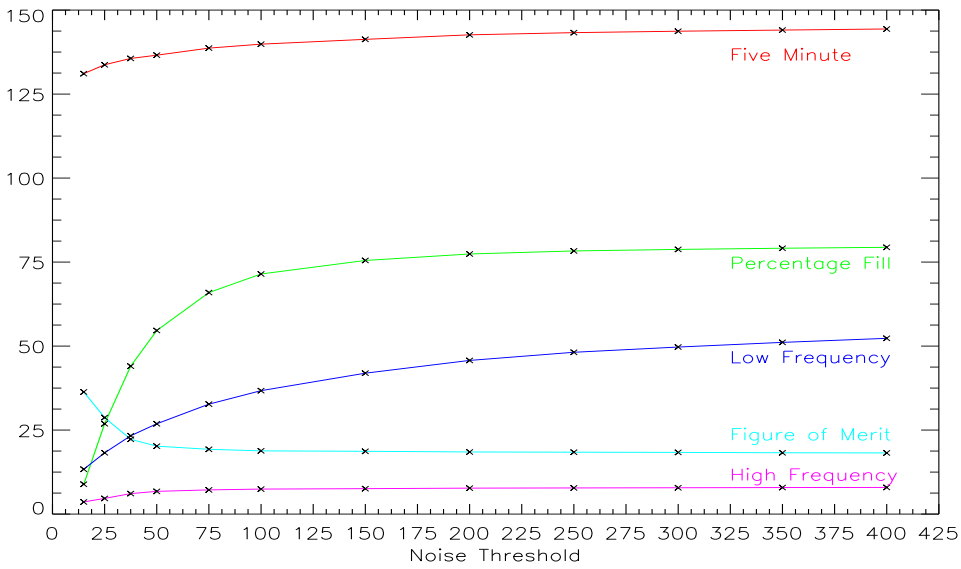

Figure 4.7: Mean five minute, high, and low frequency power, five minute figure of merit ratio, and percentage fill against rejection threshold. All powers are measured in $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$.

This has the double-edged effect of reducing the mode amplitude whilst at the same time increasing the background noise level, to the point where the mode is no longer detectable as in the case with the 15 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ threshold.

By looking at the power spectrum as a whole we can determine if the results found for this particular mode also apply for other modes.

Figure 4.7 shows how the three power bands (LP1, FM, and HF) vary with rejection threshold, along with the fill and the five minute figure of merit for each long timeseries. As expected the mean low frequency power drops as the rejection threshold is reduced since this is the parameter we are controlling. The fill level remains remarkably constant until the rejection threshold starts to drop below around 150 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ and then rapidly falls off as we

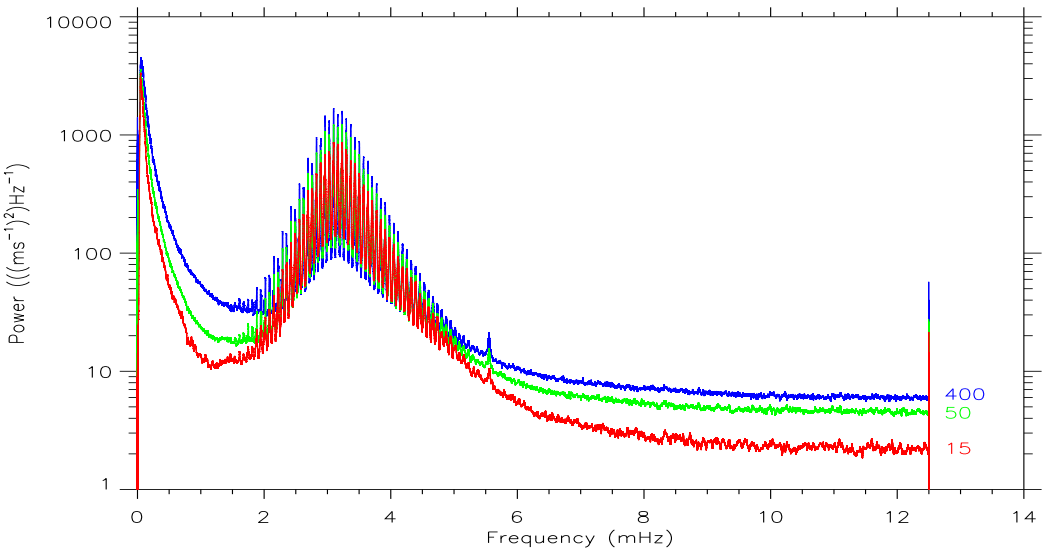

Figure 4.8: Power spectra produced using three different rejection thresholds and smoothed using a 10μHz moving mean, plotted on a log scale. Rejection was based on mean low frequency noise power (0.8 to 1.3mHz) but an improvement in high frequency noise performance has also occurred along with the gains at low frequency.

approach the very low thresholds. Interestingly as the low frequency noise is reduced, we also see an improvement in the high frequency noise performance.

Unexpectedly we have also seen a drop in the mean power in the five minute band. One would assume that this should remain constant since even with a fill of just eight percent, a seven year timeseries still contains over two hundred days of data - more than enough to resolve all five minute modes. However we have to remember to take into account the changing noise levels. Although we treat the high frequency noise region as being between 5.5 and 12.5mHz obviously the noise does not start and end at these boundaries, it permeates the entire spectrum. The mean five minute power is really the sum of the signal and the noise, and as we reduce the noise we see a corresponding reduction in five minute power, figure 4.8.

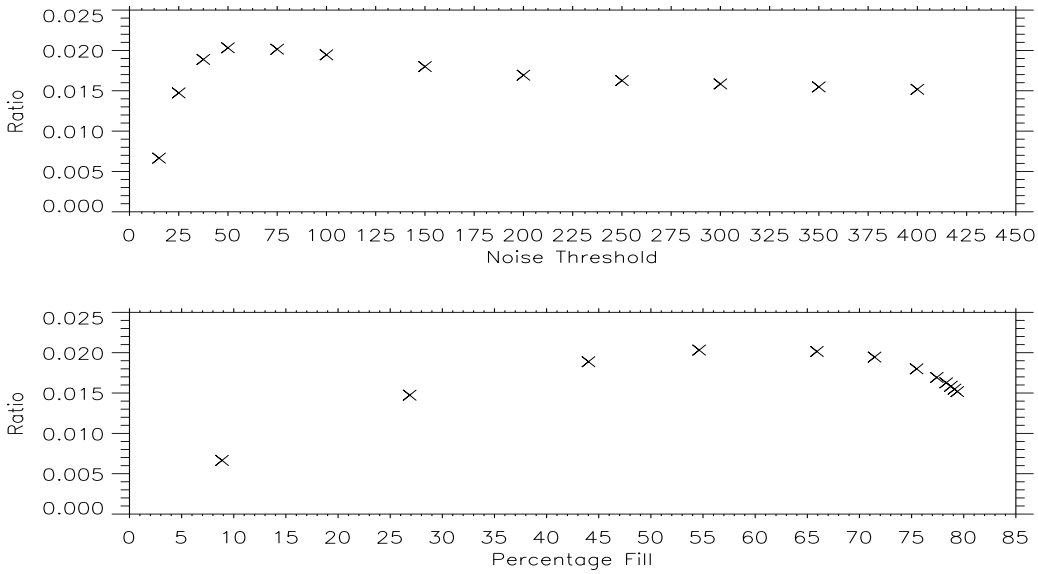

Figure 4.9: The upper chart shows the ratio of fill to low frequency noise against different noise rejection thresholds. The lower chart shows the ratio of fill to low frequency noise against fill. The peak occurs at a rejection threshold of between 50 and 60 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ which corresponds to a fill of around 60 percent.

A further reduction in five minute power has resulted from increased aliasing as the fill decreases, meaning more power is spread throughout the spectrum. The biggest change is for the lowest threshold level of 15 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ where the high frequency noise is down to thirty percent of the original value when using a 400 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ threshold. However, the reduction in noise has only produced a ten percent drop in the five minute power and hence the differential between the two has caused an increase in the figure of merit. This is really a false alarm since as we saw in figure 4.5 a timeseries with a fill this low is useless, despite the higher figure of merit.

In order to answer the question posed earlier - 'At what point do the gains of rejecting more bad data using a lower noise threshold become offset by the

decrease in fill level (and correspondingly larger gaps) of the long timeseries?' - we need to study the variation in fill level with the mean noise power over the frequency band of interest. We want to maximise the fill level whilst at the same time minimising the mean low frequency noise power, hence we need to maximise the relation $fill/noise$.

Figure 4.9 shows the value of this ratio plotted against both threshold and fill. Clearly, the peak occurs at a rejection threshold of between 60 and 70 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ which corresponds to a fill of around sixty percent. Interestingly, looking at table 4.1 Mount Wilson, Las Campanas, and Izaña all have low frequency performance well below this value, whilst Sutherland, Carnarvon, and Narrabri are all slightly higher.

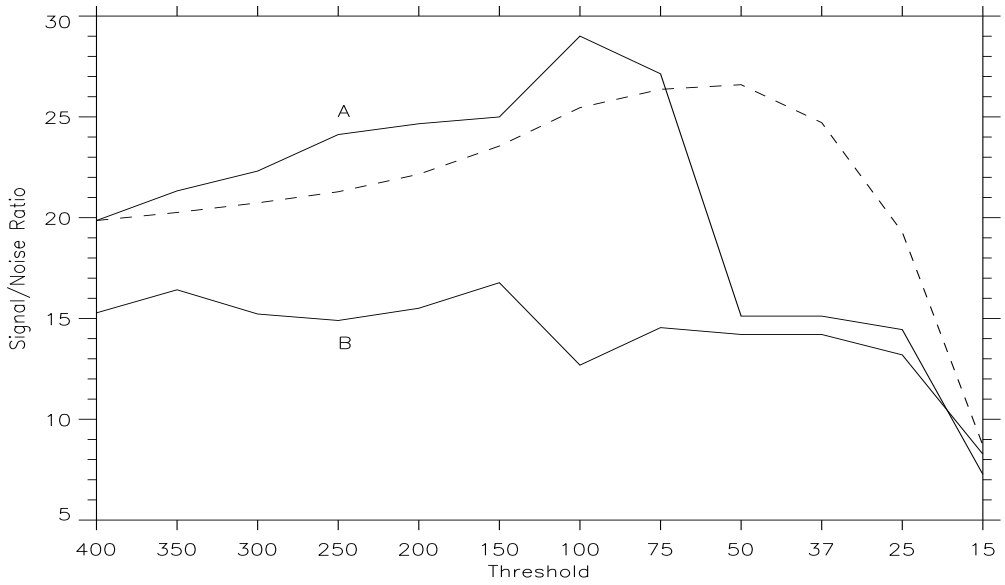

Figure 4.10: The signal to noise ratio of each component of the l=1, n=10, mode at 1612.73μHz for different rejection thresholds. The 'A' and 'B' traces correspond to the m=-1 and m=+1 components respectively. The dashed line is the 'average' result from all modes.

If we take the results from figure 4.9 and overplot them on the first result from figure 4.6 we can compare the results from one individual mode to that of many. Figure 4.10 shows a reasonably good correlation.

Although data are available to provide a fill of almost eighty percent when generating a long timeseries, it is clear from this analysis that the 'working' fill when analysing low frequency modes in the region 0.8 to 1.3mHz (i.e the amount of useful data remaining after all bad data has been rejected) is only approximately sixty percent. However, the mean noise power in the low frequency region of interest has been halved and as such despite the lower fill the ratio $fill/noise$ has shown an improvement of around one third. It is likely that the 'correct' rejection threshold would be different if modes outside the 0.8 to 1.3mHz regime are sought.

Chapter 5

Time Correction

5.1 Introduction

If data from more than one station are to be combined into a long timeseries then it is imperative that all stations have correct relative timing such that the residuals can be combined coherently. This means that the observed phase is consistent from one station to the next. Hardware synchronisation of timing is provided by GPS receivers at Carnarvon, Las Campanas, and Narrabri. The timing signals at Tenerife, Sutherland, and Mount Wilson are provided by the host establishments.

Despite this hardware, timing errors can and do occur. Usually these are through a temporary GPS failure or through site staff simply writing down the wrong time, amongst other reasons, and these errors must be corrected. If the time error is an integer multiple of the sample time then the correction is achieved by simply shifting the time axis by the appropriate number of samples. If the correction is non-integer then interpolation is required.

Before any corrections can be implemented, the timing errors must first be detected. There are two forms of error, either the site is actually sampling at the wrong time or the site is sampling at what appears to be the correct time but the clock is wrong. The former is easy to detect since the time values will not coincide with any of the standard BiSON time indices. The latter is more difficult since the time values are all valid, and unfortunately it is also the most common of the two errors.

Providing there are at least three overlapping stations it should be possible to check, and where necessary correct, any timing problems. While it would be possible to deduce a timing problem from just two stations, there is no way of knowing which station is at fault. Hence comparison with a third station is required in order to produce any actual correction. For timing errors that are not corrected within one day, it is usually possible to extrapolate and correct days in between when the error was first introduced to when it was resolved even if those days do not overlap any other stations.

5.2 Techniques

The most accurate method of determining timing errors is to manually look at every overlap and use human judgment to determine whether the data from each station are co-incident. However, given that with data arriving daily from six sites there can be several thousands of overlaps to analyse per year, some degree of automation is desired. While it is almost impossible to fully automate this task, there are a number of methods that can be applied to relieve some of the burden from the user.

5.2.1 Cross-Correlation

Theory suggests that since the instruments are looking at the same source, when the data are appropriately arranged such that there are no timing errors the cross-correlation coefficient between the two sets should be unity. This is not true in practice partly due to the inherent noise in the datasets, and partly due to the distance between stations on the Earth causing them to observe at slightly different points on the line profile. However, despite these effects it is still possible to check the timing between stations by looking for the *maximum* value of the cross-correlation.

The correlation can be calculated at various time offsets between the two datasets, for example -125 to +125 samples, and if the maximum value does not occur at zero then the overlap can be flagged as requiring a possible time correction.

If this method is to be fully exploited we first need to understand the shape and values of the cross correlations expected for different scenarios. In order to do this artificial data was generated for six stations using the same window function as the real data captured during the year 2001. This artificial data contains a full spectrum of modes between one and five mHz but with a noise power of zero, and hence represents an 'ideal' dataset in that the overlapping data from different stations are identical allowing us to see the nature of the correlations more clearly.

The cross correlation was calculated for every overlapping station using this artificial 2001 dataset. The average of all correlations was then taken in order the find the 'ideal' cross correlation result of two datasets with no

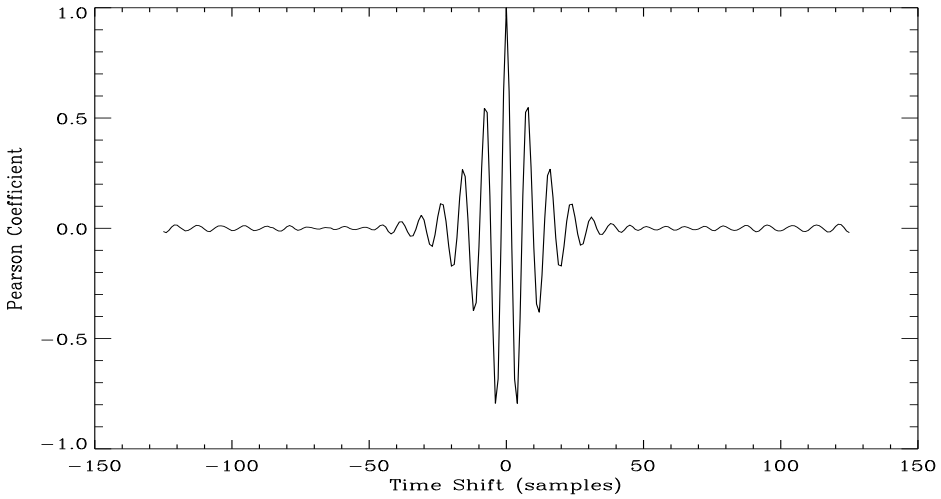

Figure 5.1: The ideal cross correlation result for two correctly timed datasets.

timing errors, figure 5.1.

As expected the position of maximum correlation occurs at zero offset and has a value of unity since with no noise the overlapping timeseries are identical. There are several subsidiary maxima separated by approximately 7.5 samples. Since each sample is forty seconds in length an offset of 7.5 samples is equal to three hundred seconds, or five minutes. This is as expected since from the five minute characteristic timescale of the solar oscillations they will naturally show a tendency to correlate around these periods. The envelope of the correlation gradually decays and oscillates around zero once the offset becomes greater than approximately forty samples indicating that the timeseries are almost completely uncorrelated past this point. The primary maximum has a fifty percent greater magnitude than the secondary maxima and hence a timing error resulting in a shift of this peak from zero could be identified easily.

5.2.2 Power Ratio

Another method is to look at the sum and difference power spectra of the overlapping stations. If correctly synchronised, the power spectrum of the sum of two overlapping timeseries should show a strong five minute peak and the power spectrum of the difference should show only noise, figure 5.2. If there were a timing shift or a scaling problem with the data then this would be indicated by excess power in the difference spectrum, figure 5.3. It should be possible to detect any problems by looking at the ratio of the two powers.

The power ratio was calculated for every overlap using the artificial 2001 data, again with time offsets from -125 to +125 samples, and averaged to produce the 'ideal' correlation, figure 5.4. The results are very similar to the cross-correlation approach with subsidiary maxima every 7.5 samples, centred on the primary peak at zero offset. The maximum has an infinite value since we are dividing by zero noise power when the data are correctly aligned, in reality the magnitude of this point would be defined by the level of noise in the dataset. The envelope of the power ratio decays more quickly than the cross correlation, and oscillates around unity once the offset becomes greater than approximately twenty samples indicating that the timeseries are almost completely uncorrelated past this point. The faster decay and the greater ratio between primary maximum and secondary maxima indicate that this method may be more accurate in detecting time errors than the cross correlation technique.

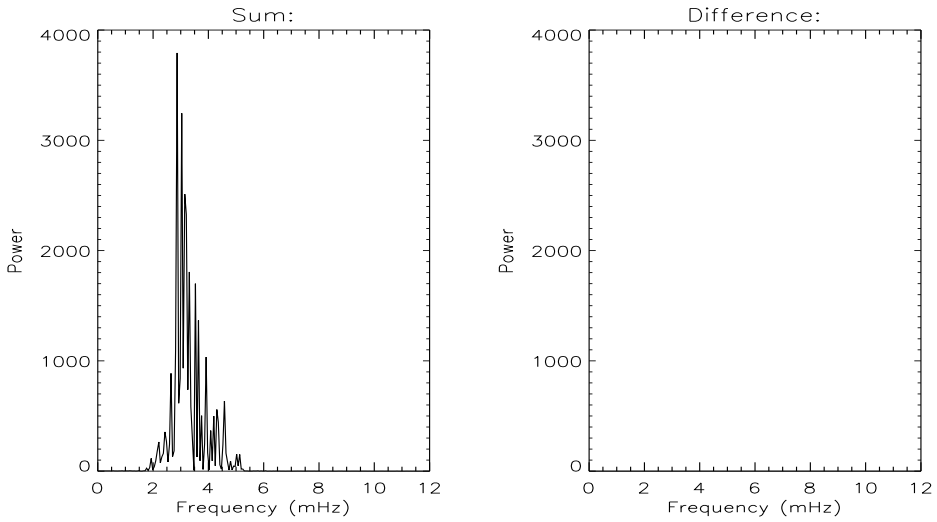

Figure 5.2: Ideal Sum and Difference spectra for two correctly timed overlapping stations. Note the strong five minute peak in the sum as expected, and the zero total power in the difference due to this artificial dataset having a noise level of zero. Clearly if we calculated the sum/difference ratio here the result would be infinite since we are dividing by zero.

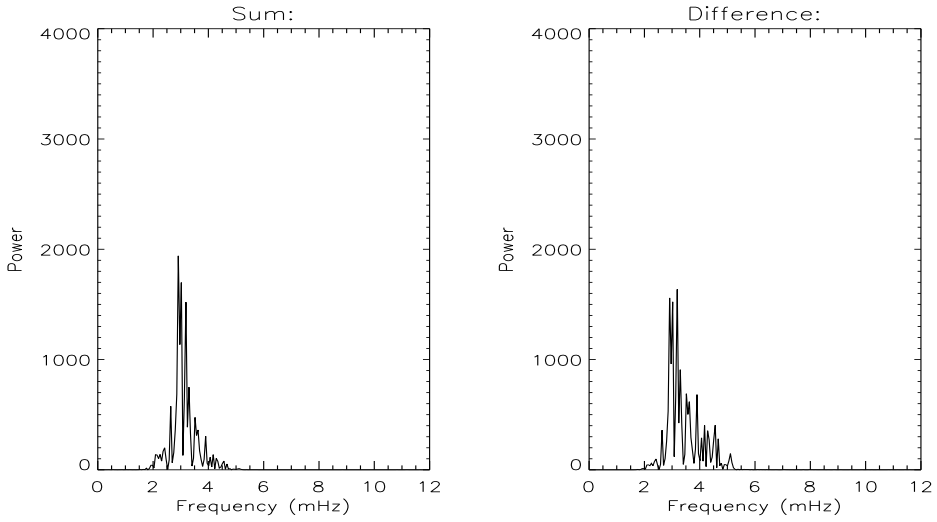

Figure 5.3: Ideal Sum and Difference spectra for two incorrectly timed overlapping stations. Introducing an artificial timing error of two samples we can see that the power in the sum has been reduced and the difference power has increased. This error could be detected by looking for the maximum ratio between the sum and the difference spectra.

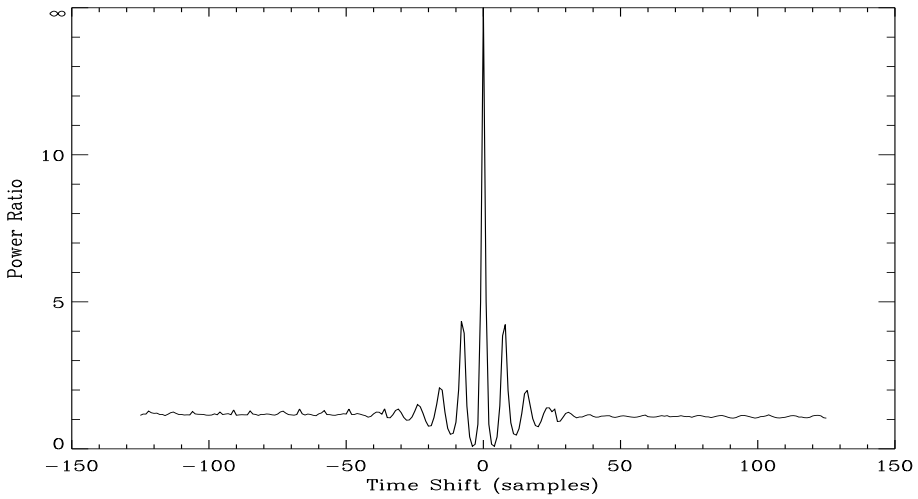

Figure 5.4: The ideal power ratio result for two correctly timed datasets.

Rather than looking at the whole spectrum, it can also be beneficial to look at the power in different pass bands. For example concentrating on the five minute regime should give the largest power ratio and allow the primary maximum to be identified even more easily.

5.3 Assessing Technique Effectiveness using Real Data

5.3.1 Cross-Correlation

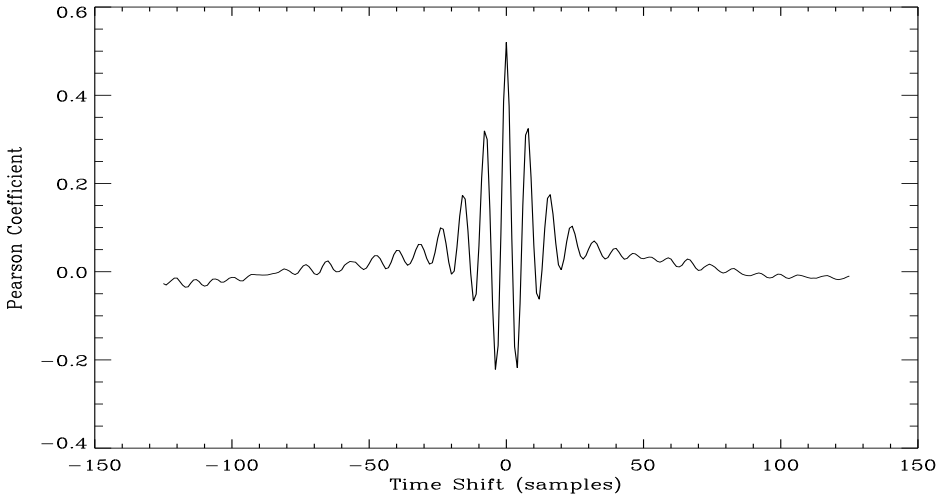

Figure 5.5: The average of all cross correlations for 2002.

The cross correlation was calculated for all overlaps during the year 2002. For the majority, the peak of the correlation was found at zero offset, however there were a significant number of overlaps incorrectly identified as requiring a time correction. Upon inspection it was found that of these most were because the data were too noisy to give a reliable correlation. Many have large gradients caused by low frequencies in the data which results in problems detecting the position of maximum correlation. These low frequencies are caused by two effects, partly due to extinction effects at the beginning and ends of each day, and partly due to an instrumental error where some stations place a characteristic 'footprint' on the data they produce.

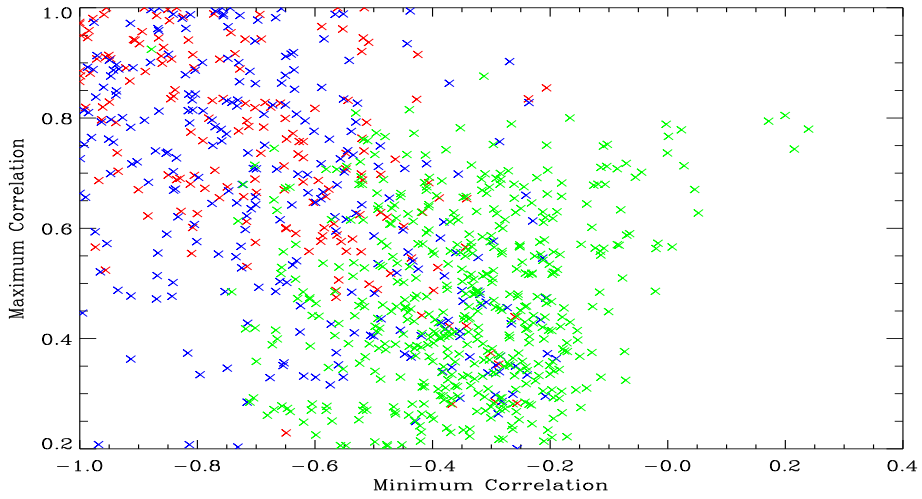

Figure 5.6: Maximum against minimum value of each cross correlation during 2002. Red are overlaps of less than 150 points, blue are between 150 and 300 points, and green are overlaps of more than 300 points.

These low frequencies manifest themselves as a non-zero background or 'bump' in the average correlation, figure 5.5, when compared to the ideal average produced from the artificial dataset, figure 5.1. In an attempt to overcome this problem a 'figure of merit' for the correlation was introduced. This is defined as the difference between the maximum and minimum value of the correlation such that a threshold level can then be defined below which automatic analysis will not take place.

One would naively assume that longer overlaps would result in cleaner cross correlations. Looking at figures 5.6 and 5.7 it would seem that this is not true. Almost all correlations with a high FOM of between 1.5 and 2.0 are produced from overlaps of less than 300 points. This is mainly a statistical effect, where as the number of points increases the inherent noise in the dataset starts to have more of an impact.

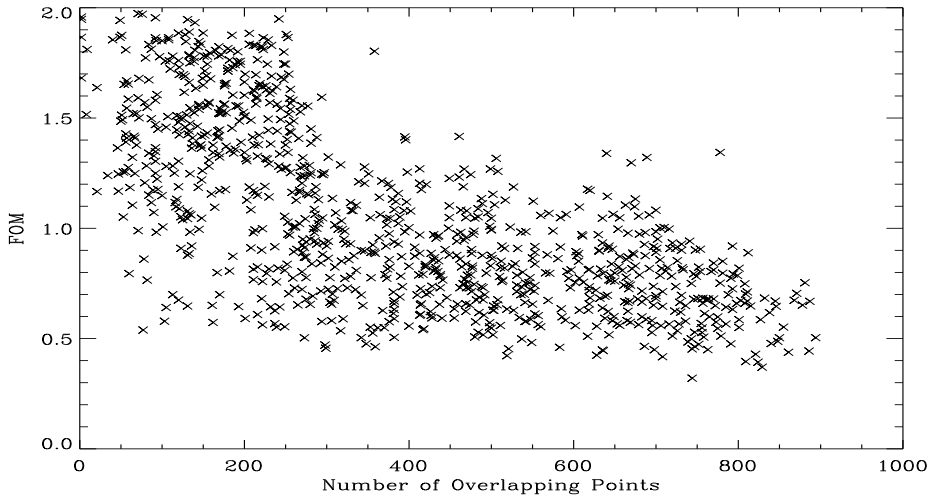

Figure 5.7: Cross Correlation Figure of Merit (maximum - minimum) against number of points for each overlap during 2002.

In order to determine how well each correlation compares with the average a χ^2 goodness-of-fit test was applied to each overlap,

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}(\frac{Observed_i - Model_i}{\sigma_i})^2$$

where the observed data is the correlation of each overlap, the model is the 'ideal' average, and the error, σ, on the observed data is assumed to be equally weighted. If an overlap has a correlation identical to the ideal then it would have a χ^2 value of zero, and hence low values of χ^2 are desired since these indicate a better fit.

The results of this analysis, figure 5.8, show that there is a significant percentage of correlations with a very poor match to the ideal. If we re-scale χ^2 such that a high value (bad fit) corresponds to a light blue and a low value (good fit) corresponds to a dark blue, we can observe the range of correlation coefficients over good and bad fits, figure 5.9.

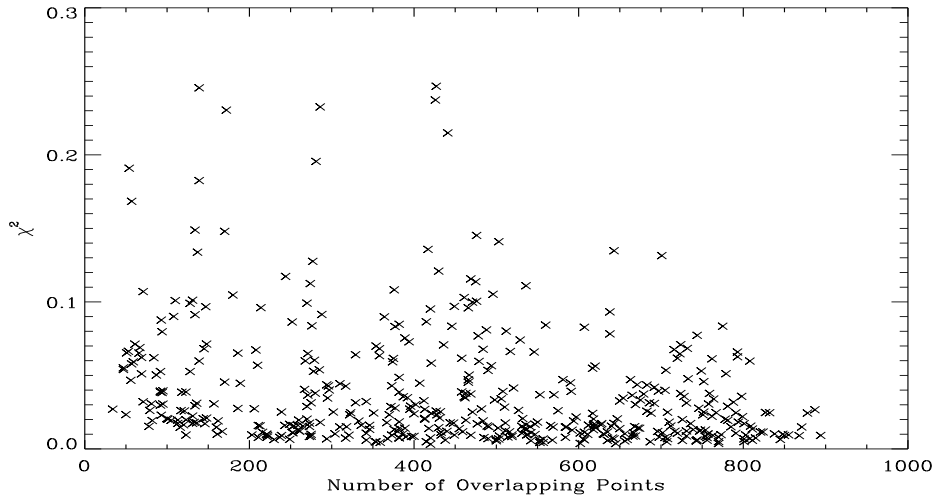

Figure 5.8: Cross Correlation χ^2 against number of points for each overlap. Lower values of χ^2 indicate a better fit.

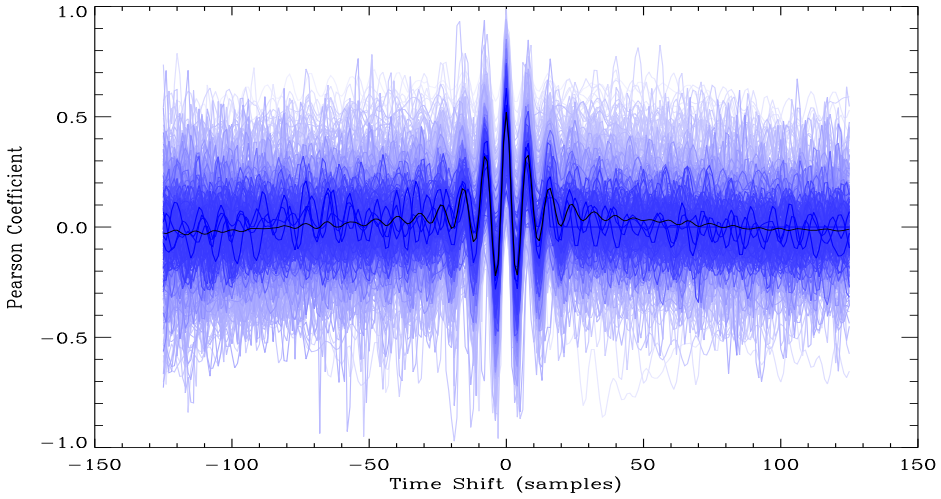

Figure 5.9: Cross Correlations colour coded such that light blues are a bad fit and dark blues are a good fit. The black line is the average of all correlations.

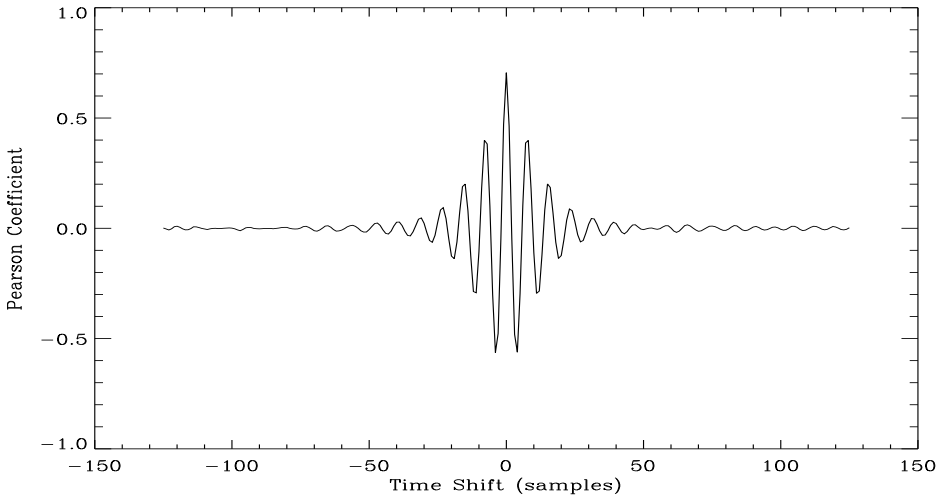

Figure 5.10: The average of all cross correlations for 2002 after moving mean processing.

These problems are causing the high percentage of false alarms when attempting to auto-detect timing errors. As mentioned previously the large gradients in the correlations caused by low frequencies in the data results in problems detecting the position of maximum correlation. If we could filter out these low frequencies it might result is a more stable and reliable autodetection routine.

This was done by subtracting a twenty-five point moving mean from all the residual data. The moving mean has the effect of a high pass filter which cuts in at approximately one mHz, with some slight attenuation of the low frequency end of the five minute region. If we now re-calculate the average of all the correlations to find the 'ideal', figure 5.14, we can see that the result is now much more like that which we obtained from the artificial data, figure 5.1.

Looking at figures 5.11, 5.12, and 5.13, there has been a dramatic improvement in the quality of the correlations. While overlaps of less than approximately 250 points seem unaffected those over 250 points have a much lower χ^2 value indicating a better fit.

Again, if we re-scale χ^2 such that a high value (bad fit) corresponds to a light blue and a low value (good fit) corresponds to a dark blue, figure 5.14, there are many more dark blue correlations. It would seem that an overlap of at least 250 points is required to achieve reliable results using this method.

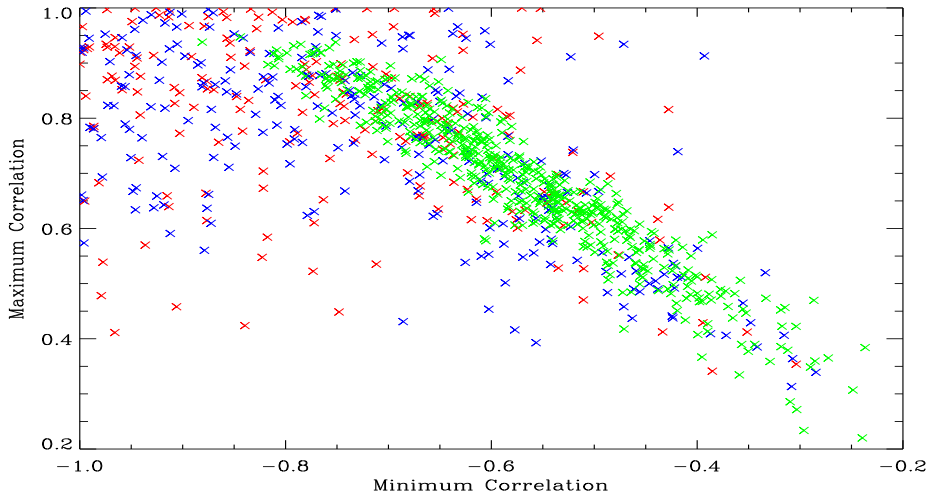

Figure 5.11: Maximum against minimum value of each cross correlation during 2002 after moving mean processing. Red are overlaps of less than 150 points, blue are between 150 and 300 points, and green are overlaps of more than 300 points.

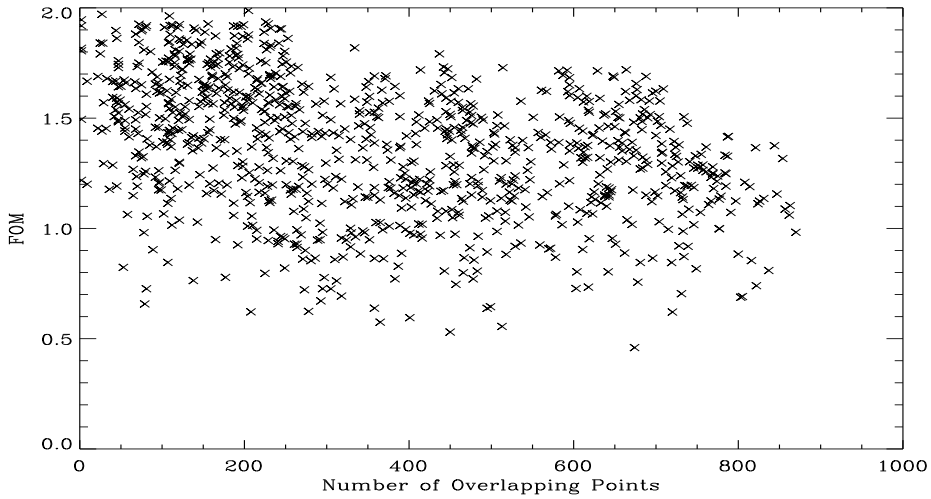

Figure 5.12: Cross Correlation Figure of Merit (maximum - minimum) against number of points for each overlap during 2002 after moving mean processing.

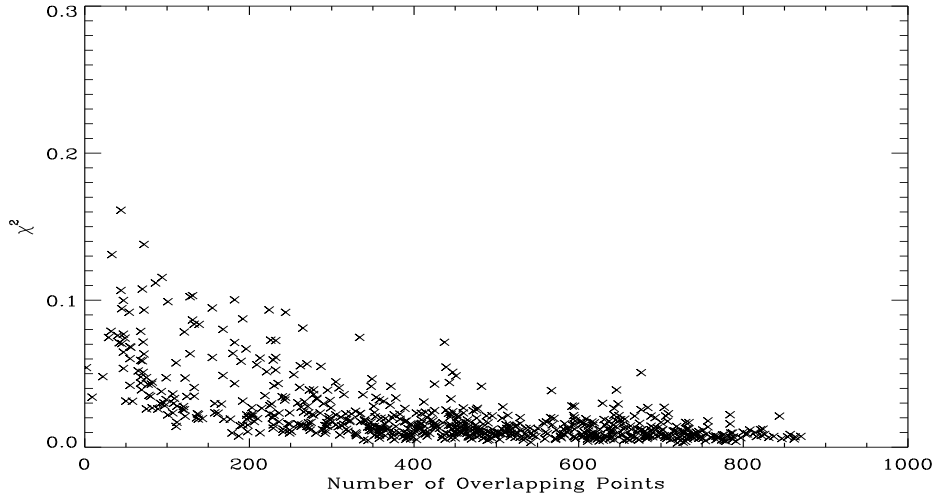

Figure 5.13: Cross Correlation χ^2 against number of points for each overlap during 2002 after moving mean processing. Lower of χ^2 values indicate a better fit.

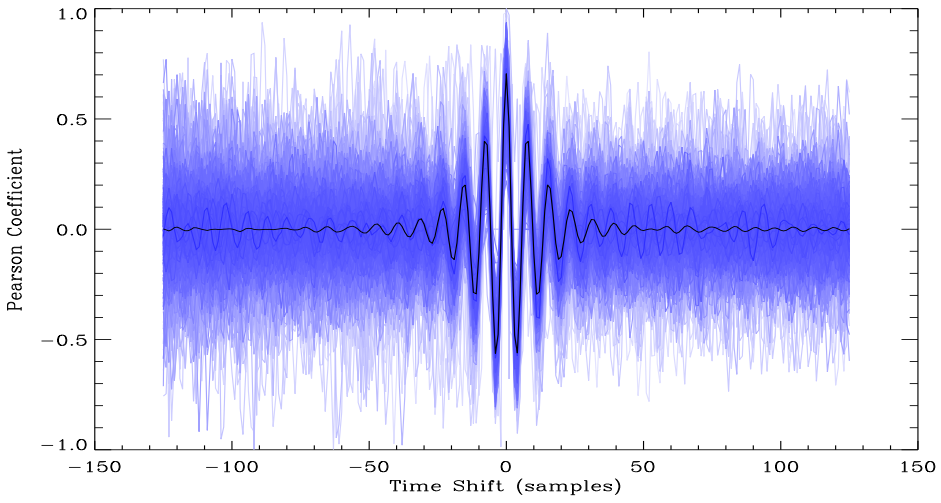

Figure 5.14: Cross Correlations after moving mean processing, colour coded such that light blues are a bad fit and dark blues are a good fit. The black line is the average of all correlations.

5.3.2 Power Ratio

The same process was followed using the sum/differece power ratio method as for the cross correlation technique. The five minute (2mHz to 5mHz) passband was used as opposed to the entire spectrum in order to obtain the maximum signal to noise ratio.

Looking at figures 5.15 and 5.16 we can see that this technique is not as susceptible to low frequencies in the datasets as with the cross correlation. The moving mean average is almost identical to the raw average except for a higher magnitude central maximum.

From figures 5.17, 5.18, 5.19, 5.20, 5.21, and 5.22 it is clear that this technique has another advantage over the cross correlation approach. There is very little correlation between the number of points in the overlap and the quality of the fit of the power ratio compared to the average, and this means the technique should work regardless of how many points make up the overlap. This is as expected since we are only interested in the total power in the power spectrum rather than the resolution.

Finally, from figures 5.23, and 5.24 we can see again that the moving mean has little effect on the shape of the power ratio but it does increase the magnitude difference between the central maximum and the secondary maxima which should improve the reliability of automatically finding the central maximum.

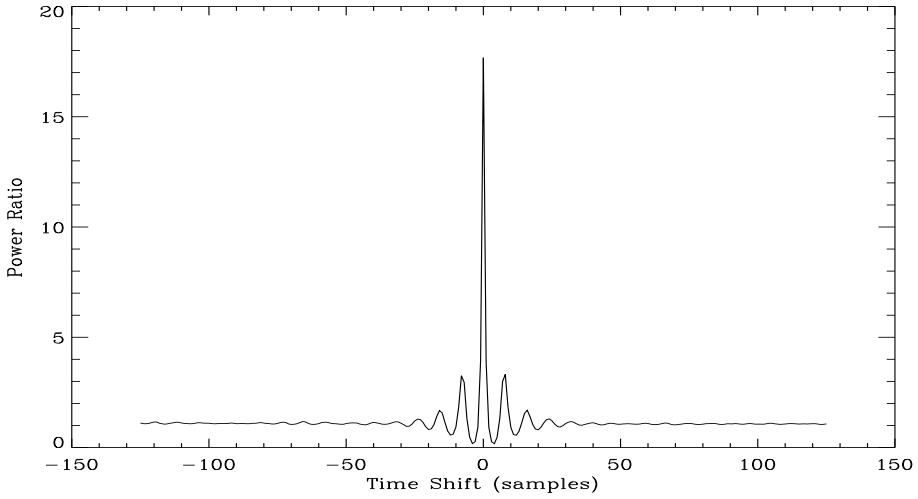

Figure 5.15: The average of all power ratios for 2002.

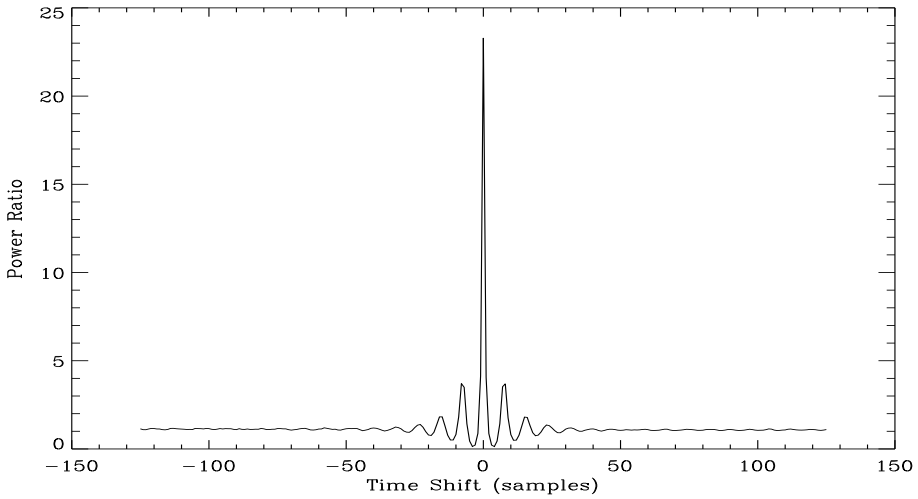

Figure 5.16: The average of all power ratios for 2002 after moving mean processing.

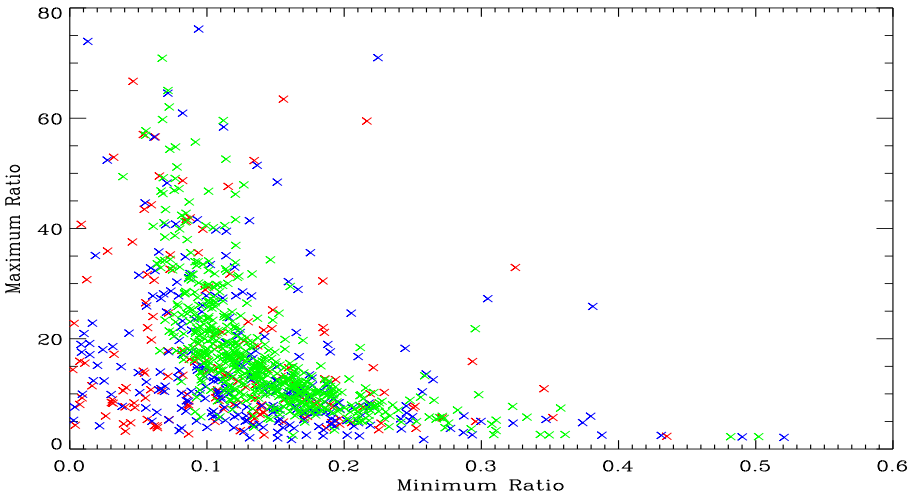

Figure 5.17: Maximum against minimum value of each power ratio during 2002. Red are overlaps of less than 150 points, blue are between 150 and 300 points, and green are overlaps of more than 300 points.

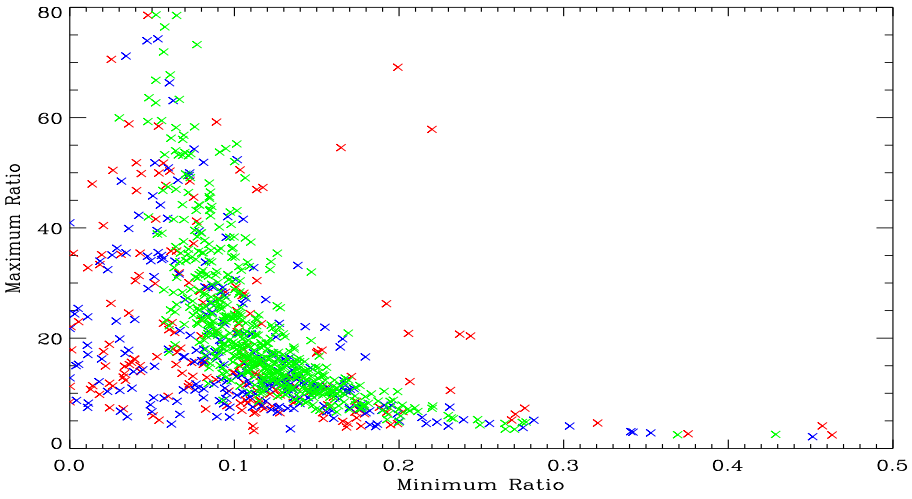

Figure 5.18: Maximum against minimum value of each power ratio during 2002 after moving mean processing. Red are overlaps of less than 150 points, blue are between 150 and 300 points, and green are overlaps of more than 300 points.

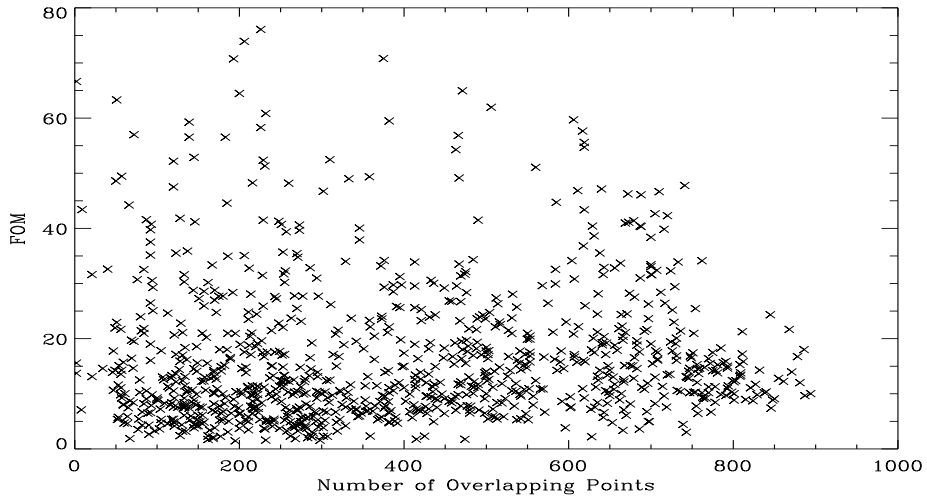

Figure 5.19: Power ratio Figure of Merit (maximum - minimum) against number of points for each overlap during 2002.

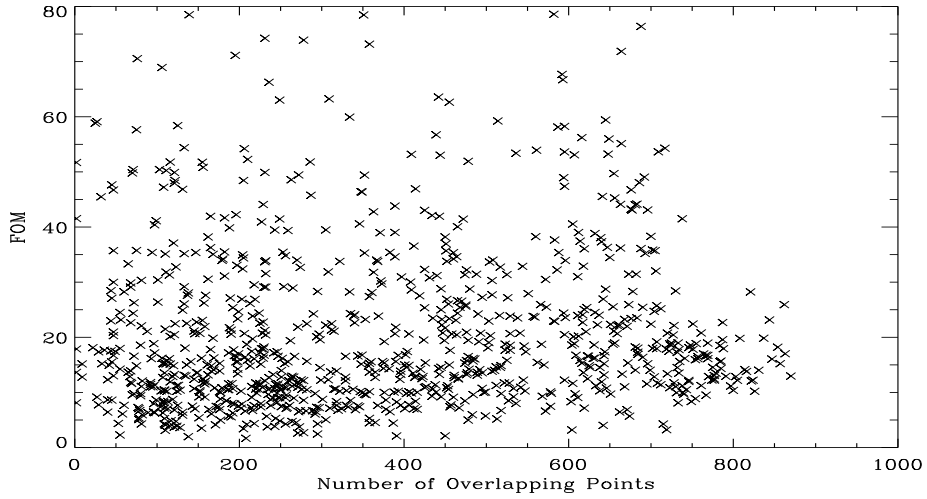

Figure 5.20: Figure of Merit (maximum - minimum) against number of points for each overlap during 2002 after moving mean processing.

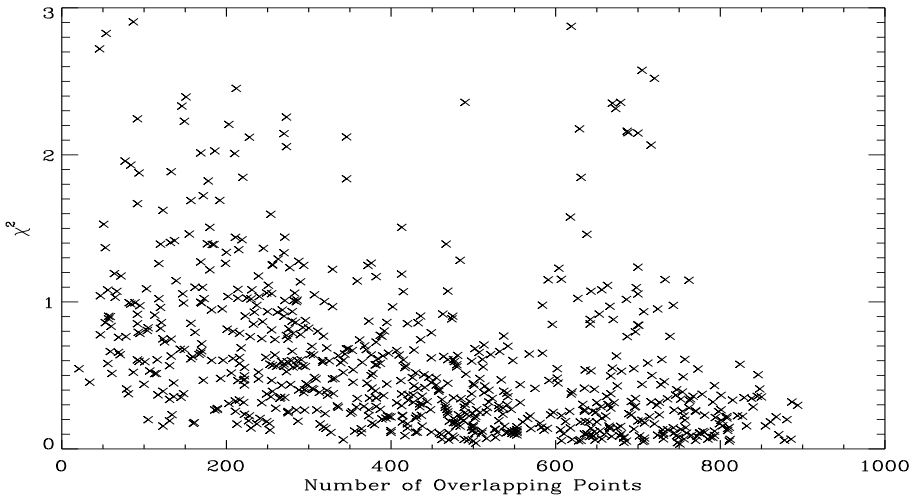

Figure 5.21: Power ratio χ^2 against number of points for each overlap during 2002. Lower values of χ^2 indicate a better fit.

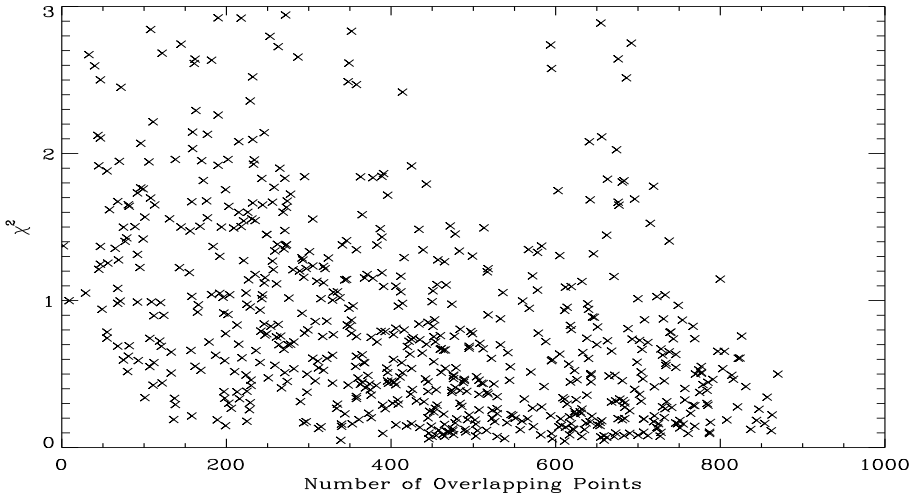

Figure 5.22: Power ratio χ^2 against number of points for each overlap during 2002 after moving mean processing. Lower values of χ^2 indicate a better fit.

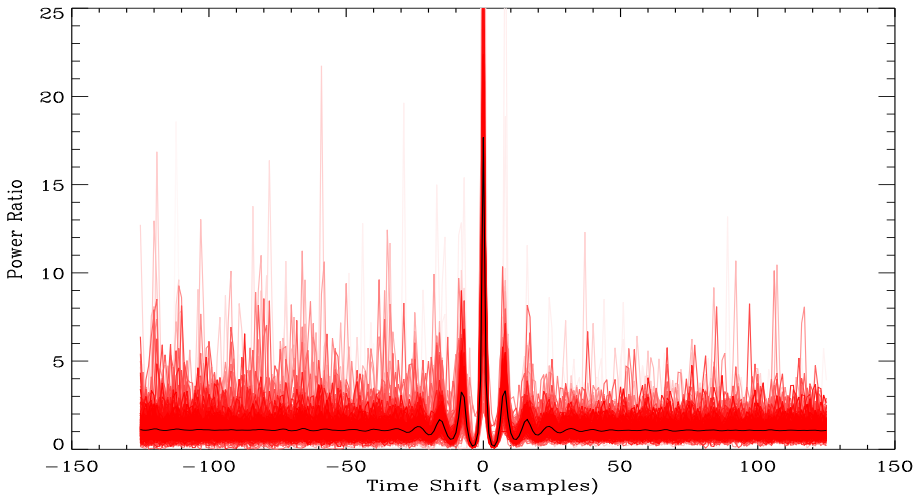

Figure 5.23: Power Ratios colour coded such that light reds are a bad fit and dark reds are a good fit. The black line is the average of all power ratios.

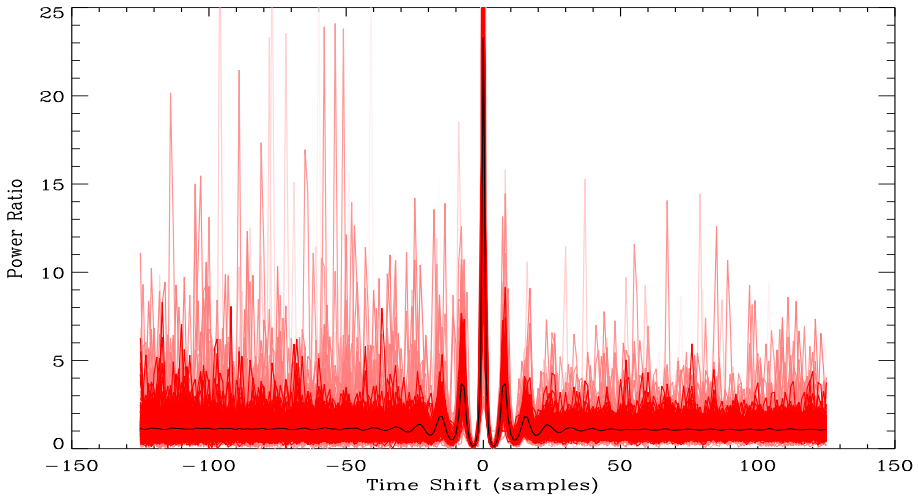

Figure 5.24: Power Ratios after moving mean processing, colour coded such that light reds are a bad fit and dark reds are a good fit. The black line is the average of all correlations.

5.4 Correcting Real Timing Errors

Now that we have an understanding of the advantages and limitations of both methods, we can apply them and attempt to detect some real timing errors.

The two methods were applied to data from all stations for 2002, using both raw residuals and data processed using a moving mean. The possible time correction required was determined using four criteria, these are a 21 sample wide correlation ($\pm$10 samples from zero), a 251 sample wide correlation ($\pm$125 samples from zero), a wideband sum/difference power ratio using the whole of the power spectrum, and finally a narrow band sum/difference power ratio using just the five minute regime.

Two good examples of the results from these methods are detailed below.

5.4.1 13 August 2002

Location	Points	±10 Correlation	±125 Correlation	Wide Power	Five Minute Power
LA-IZ	785	-2	-124	-124	-2
MO-LA	697	0	-120	-120	0
MO-IZ	484	-2	-2	-2	-2

Table 5.1: Autodetected timing errors for 13 August 2002 using raw data.

During 13 August 2002 there was a three station overlap between Izaña, Las Campanas, and Mount Wilson. From the results in table 5.1 we cannot conclude anything regarding the timing of these stations with one hundred percent certainty. It looks as if there may be a problem between Mount Wilson and Izaña with all criterion stating -2 samples, but this is not confirmed conclusively by the Las Campanas and Izaña overlap.

Looking at the data these results were obtained from, figure 5.25, and 5.26, it is easy to see why the results are inconclusive. Although it is obvious to the human eye that the central maximum in the wide correlation is located at -2 samples, as we expected an automatic routine gets confused due to the gradients in the correlation. Both the narrow correlation and the narrowband power ratio have detected the correct result, but the wideband power and wide correlation are having trouble.

As already discussed these gradients are caused by low frequencies in the data, figure 5.27. If we process every station using the twenty-five sample moving mean, figure 5.28, the auto detection routine should provide better results.

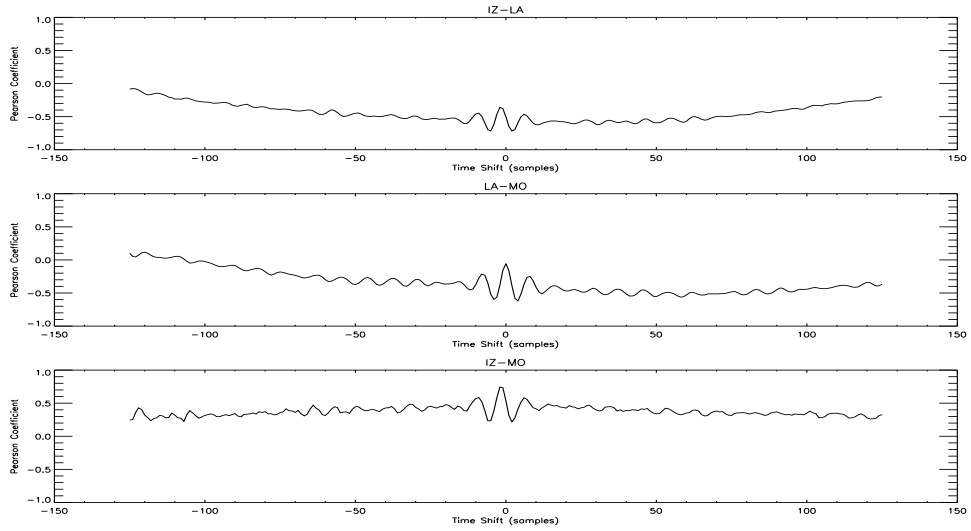

Figure 5.25: 13 August 2002. Cross correlation between Izaña, Las Campanas, and Mount Wilson before moving mean processing. Note how the gradients would prevent successful detection of the central peak over the wide range. If we looked at just ±10 samples then the peak would be detected in this instance but larger timing errors would still be missed.

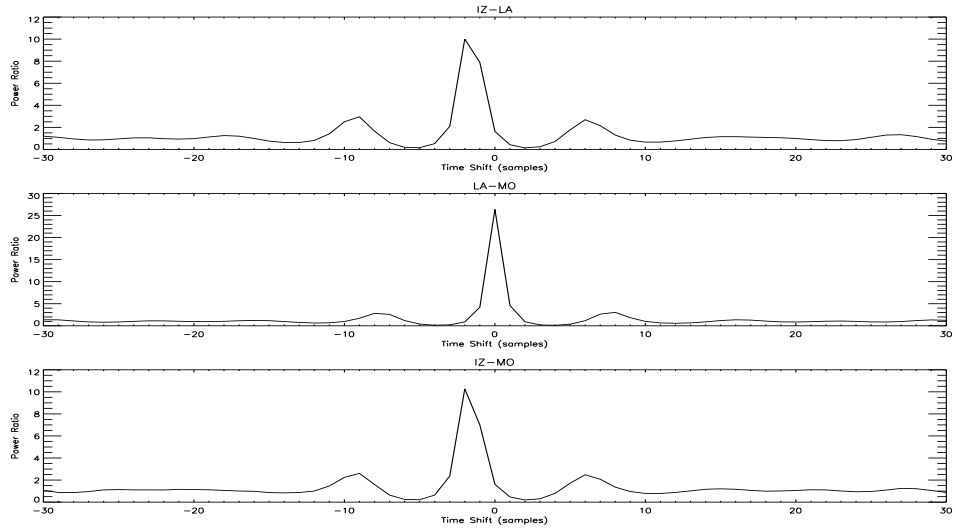

Figure 5.26: 13 August 2002. Sum/Difference power ratio between Izaña, Las Campanas, and Mount Wilson before moving mean processing. This method has produced much better results than the cross-correlation approach. The -2 sample offset is visible clearly with the ratio dropping to unity just a few tens of samples from zero offset.

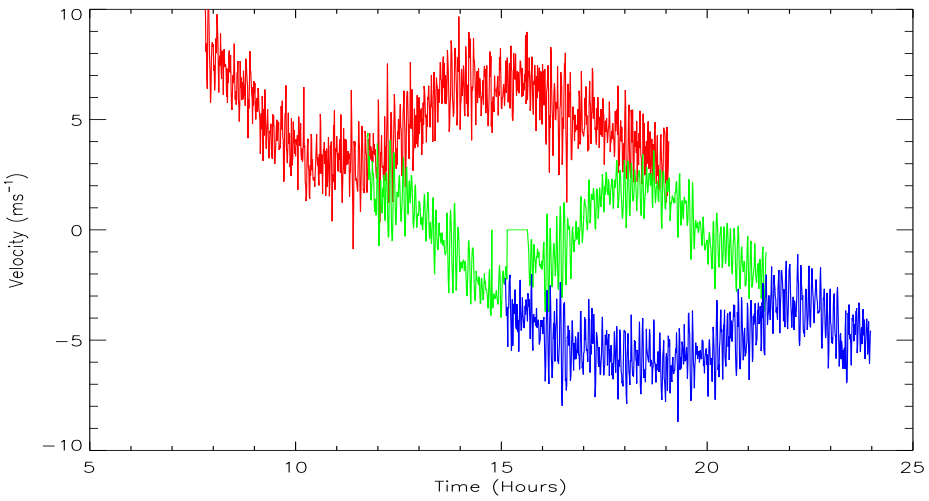

Figure 5.27: 13 August 2002. Timeseries overlap between Izaña, Las Campanas, and Mount Wilson before moving mean processing. Red is Izaña, green is Las Campanas, and blue is Mount Wilson. Izaña and Mount Wilson have been offset by ± 5 ms^{-1} respectively for clarity.

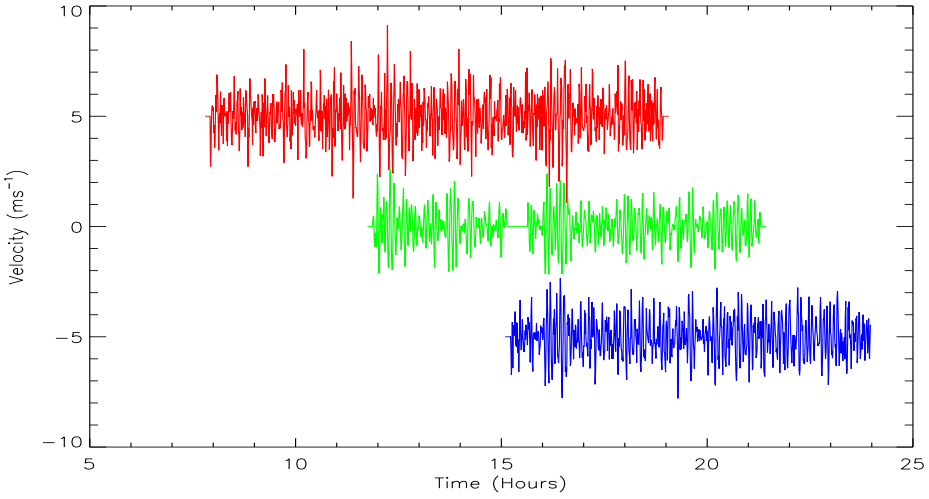

Figure 5.28: 13 August 2002. Timeseries overlap between Izaña, Las Campanas, and Mount Wilson after moving mean processing. Red is Izaña, green is Las Campanas, and blue is Mount Wilson. Izaña and Mount Wilson have been offset by ± 5 ms^{-1} respectively for clarity.

Location	Points	±10 Correlation	±125 Correlation	Wide Power	Five Minute Power
LA-IZ	761	-2	-2	-2	-2
MO-LA	673	0	0	0	0
MO-IZ	460	-2	-2	-2	-2

Table 5.2: Autodetected timing errors for 13 August 2002 using moving mean data.

Table 5.2 shows the results of the automatic detection with moving mean data. Note that the number of points in each overlap have been reduced by twenty-four samples due to the application of the moving mean. All criterion are now showing an offset of -2 on both overlaps involving Izaña, and since it would be highly unlikely for both Mount Wilson and Las Campanas to be out by +2 samples we can safely assume that it is Izaña that is at fault.

If we again look at the data these results were obtained from, figure 5.29, 5.30, and 5.31, the correlations are now much cleaner.

From these results it would appear that the time correction required for Izaña is -2 samples, or -80 seconds. However, if we look closely at figure 5.31 we can see that the central maximum does not actually peak, it has a flat top. This is because we are actually missing the peak by sampling either side of it.

The real peak is probably somewhere in the region marked by the 'x' in figure 5.32. By experimenting with various time offsets between -40 and -80 seconds and looking at the overlap in the time domain it is possible to determine exactly where the two datasets are coincident. For this particular overlap this occurs with an offset of -60 seconds, and involves shifting the timeseries by one sample, and then interpolating a further half a sample.

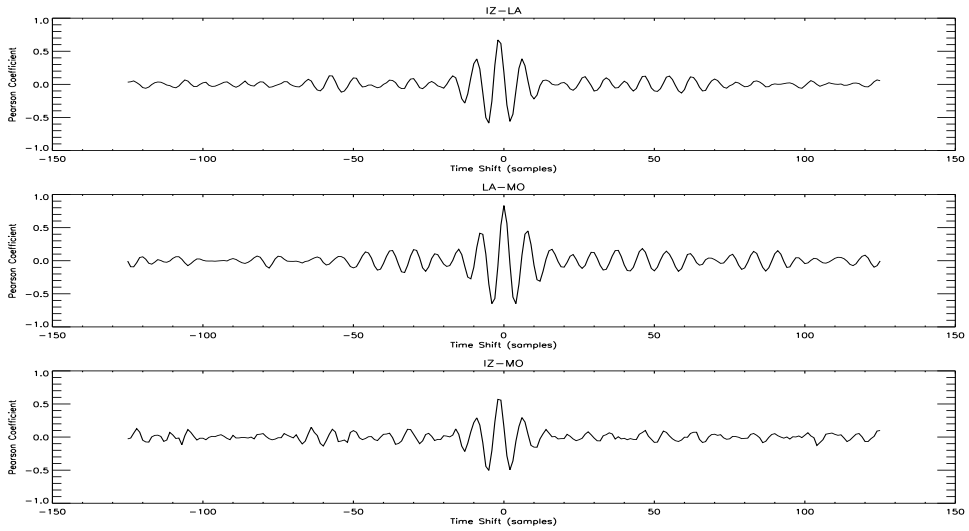

Figure 5.29: 13 August 2002. ±125 sample cross correlation between Izaña, Las Campanas, and Mount Wilson after moving mean processing.

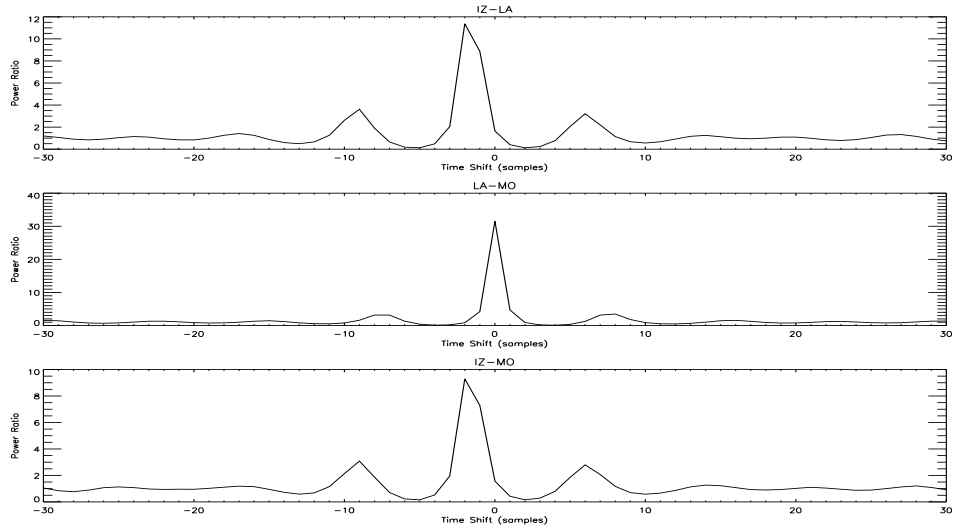

Figure 5.30: 13 August 2002. Sum/Difference power ratio between Izaña, Las Campanas, and Mount Wilson after moving mean processing.

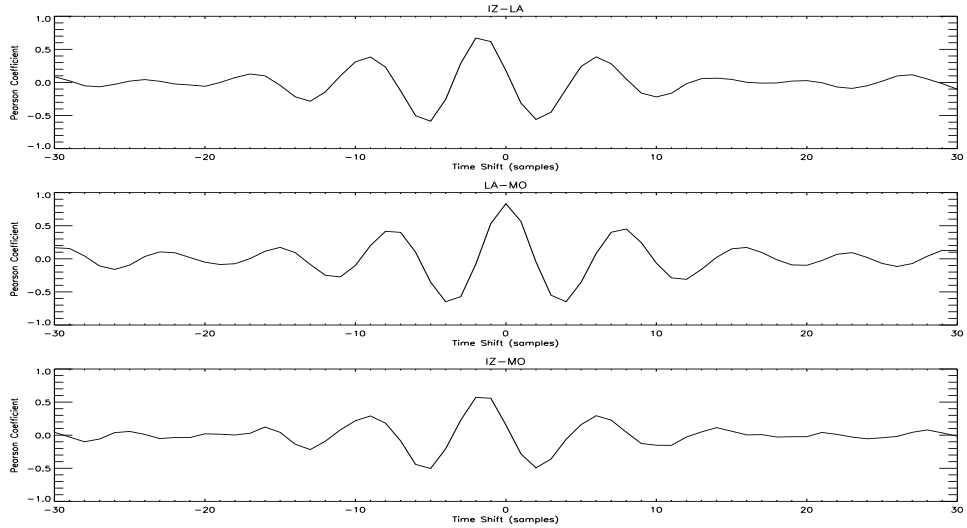

Figure 5.31: 13 August 2002. ±30 sample cross correlation between Izaña, Las Campanas, and Mount Wilson after moving mean processing.

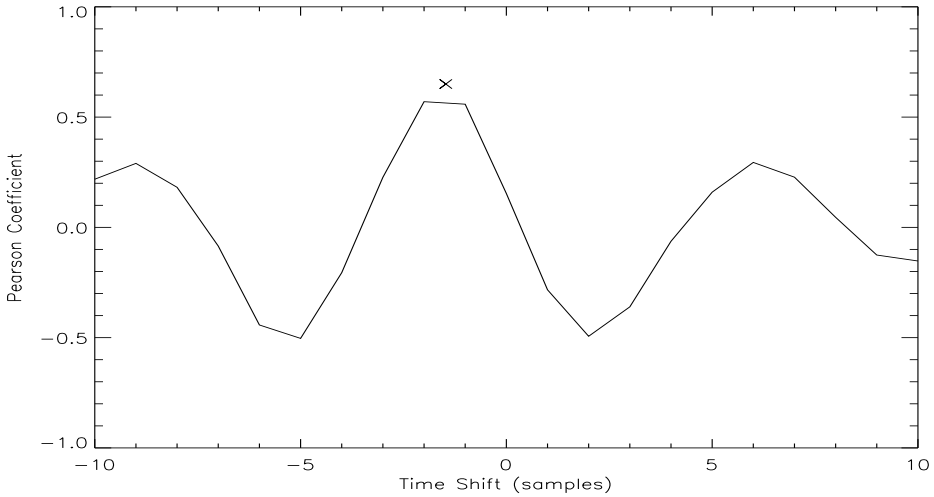

Figure 5.32: 13 August 2002. ±10 sample cross correlation between Izaña and Mount Wilson after moving mean processing. Note how we have sampled either side of the peak, indicated by the 'flat top'. The real peak is probably somewhere in the region marked by the 'x' meaning this is a timing error of around one and a half samples, or 60 seconds.

5.4.2 25 January 2002

Location	Points	±10 Correlation	±125 Correlation	Wide Power	Five Minute Power
IZ-SU	294	10	112	112	112
LA-IZ	746	-5	-113	-113	-113
LA-SU	228	0	-92	0	0

Table 5.3: Autodetected timing errors for 25 January 2002 using raw data.

From the results in table 5.3 we see much more conclusive results using the raw data than for the previous example. Again it looks to be Izaña that is in error. Las Campanas and Sutherland appear to be correctly timed with all criterion except the wide correlation showing zero. The error is likely to be between -112 and -113 samples, since both the Izaña-Sutherland overlap and the Las Campanas-Izaña overlap have produced these figures. The opposite sign is due to Izaña being the first station in the one overlap, but the second station in the other. If this value is correct then it is understandable that the twentyone point wide correlation has detected the wrong figure since the real offset is outside of the detection range.

Looking at the data these results were obtained from, figure 5.33, and 5.34, we can see that while there are still gradients in the correlations they have not prevented us from detecting the correct value this time.

As already discussed these gradients are caused by low frequencies in the data, figure 5.35. If we now process every station using the twenty-five sample moving mean, figure 5.36, we can determine if the auto detection routine can provide any better results.

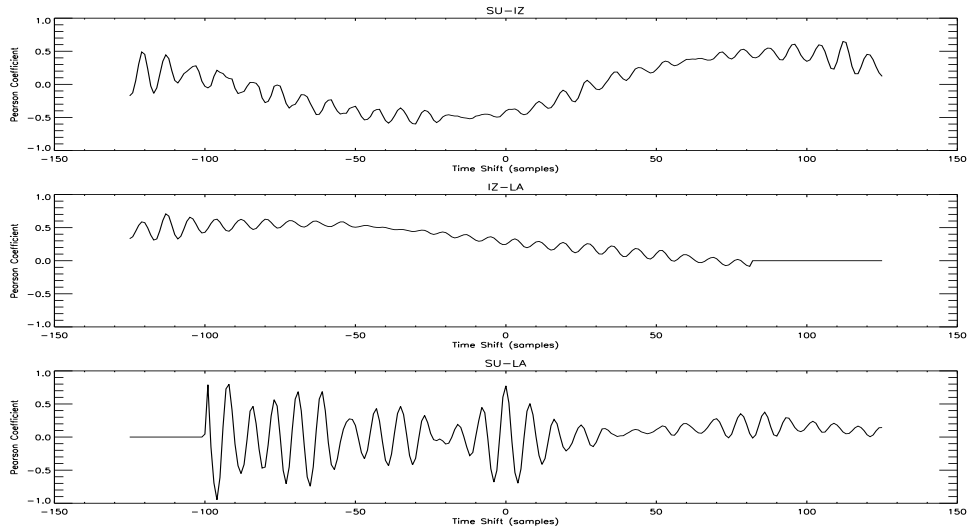

Figure 5.33: 25 January 2002. Cross correlation between Sutherland, Izaña, and Las Campanas before moving mean processing.

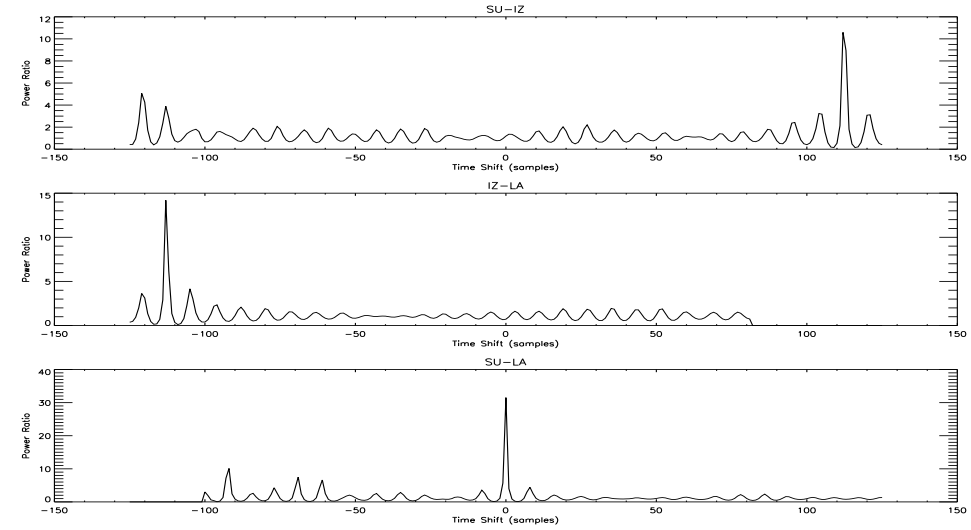

Figure 5.34: 25 January 2002. Sum/Difference power ratio between Sutherland, Izaña, and Las Campanas before moving mean processing.

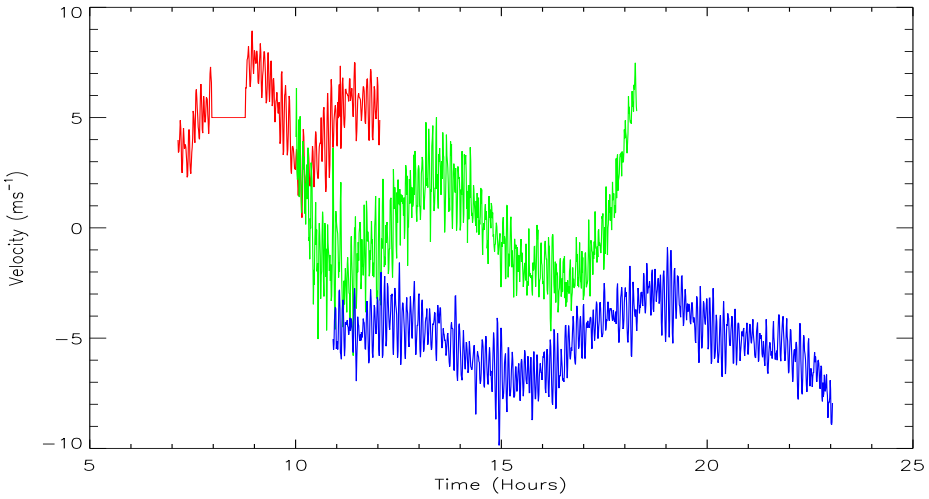

Figure 5.35: 25 January 2002. Timeseries overlap between Sutherland, Izaña, and Las Campanas before moving mean processing. Red is Sutherland, green is Izaña, and blue is Las Campanas. Sutherland and Las Campanas have been offset by ± 5 ms^{-1} respectively for clarity.

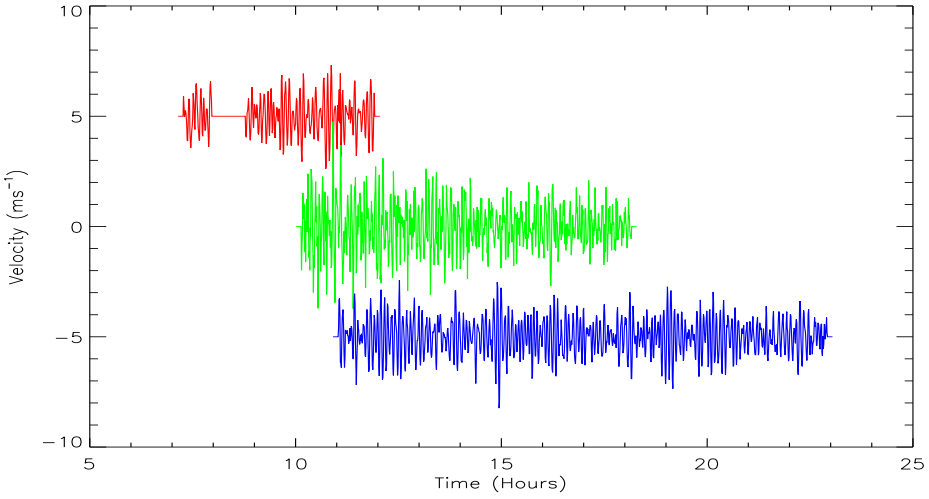

Figure 5.36: 25 January 2002. Timeseries overlap between Sutherland, Izaña, and Las Campanas after moving mean processing. Red is Sutherland, green is Izaña, and blue is Las Campanas. Sutherland and Las Campanas have been offset by ± 5 ms^{-1} respectively for clarity.

Location	Points	±10 Correlation	±125 Correlation	Wide Power	Five Minute Power
IZ-SU	282	10	112	112	112
LA-IZ	722	3	-113	-113	-113
LA-SU	204	0	-77	-69	-69

Table 5.4: Autodetected timing errors for 25 January 2002 using moving mean data.

Table 5.4 shows the results of the automatic detection with moving mean data. Note that the number of points in each overlap have been reduced by twenty-four samples due to the application of the moving mean. The two Izaña overlaps have remained the same at detecting somewhere between -112 and -113 samples. It is interesting to note that the Las Campanas-Sutherland overlap has actually become worse with only the narrow correlation now giving the correct value of zero. If we again look at the data these results were obtained from, figure 5.37, and 5.38, we can see the reason for this problem.

As with the previous example, the detected time correction does not have a peak but rather a flat top and therefore must not be an integer number of samples. This agrees with the detected value of between -112 and -113 samples. By looking at the overlap in the time domain it is possible to determine exactly where the two datasets are coincident. For this particular overlap this occurs with an offset of -112.5 samples or -75 minutes.

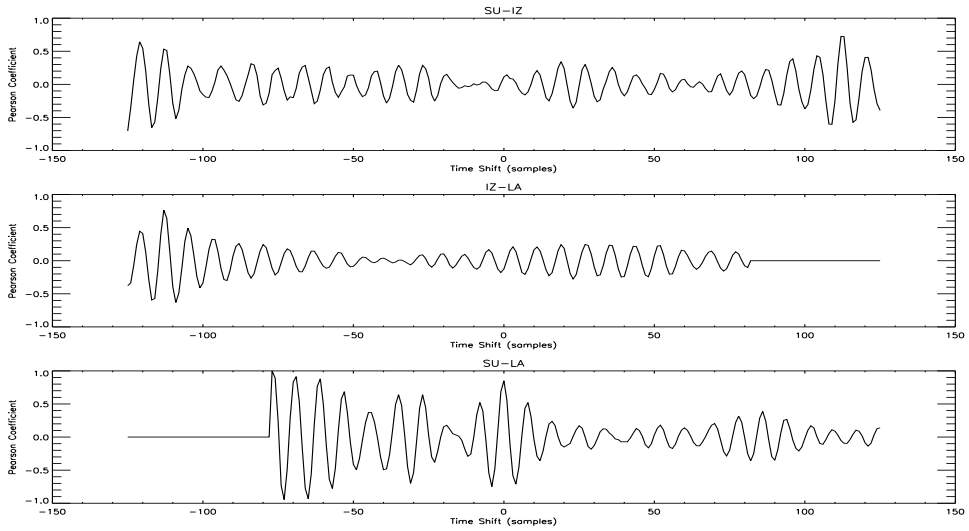

Figure 5.37: 25 January 2002. Cross correlation between Sutherland, Izaña, and Las Campanas after moving mean processing.

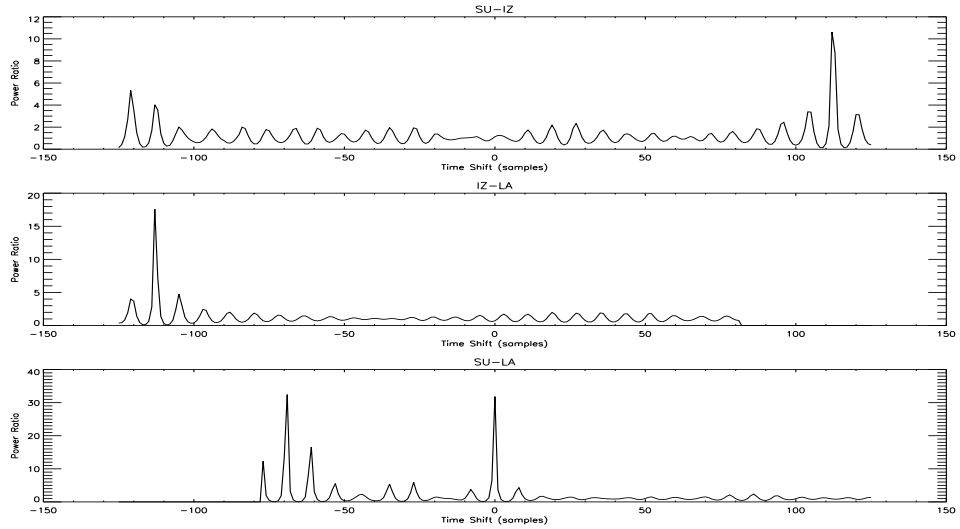

Figure 5.38: 25 January 2002. Sum/Difference power ratio between Sutherland, Izaña, and Las Campanas after moving mean processing.

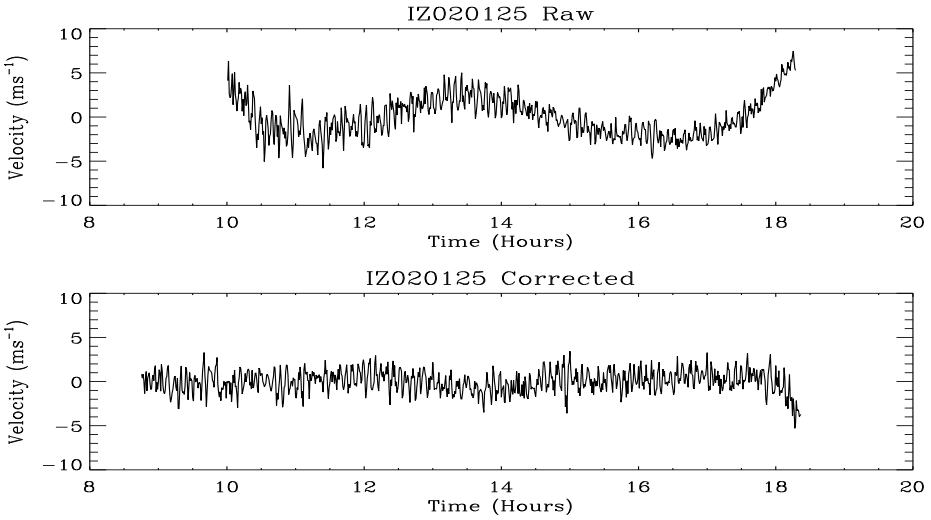

Figure 5.39: Top: 25 January 2002 Izaña uncorrected. Bottom: 25 January 2002 Izaña corrected by -75 minutes.

Timing errors of this magnitude will have implications on the residual calibration since the value of $v_{rotation}$ will be incorrectly calculated. Small errors can safely be corrected by adjusting the residuals but large timing corrections should be implemented at the raw instrumental data stage and the residuals reproduced, figure 5.39

5.5 Conclusions

Clearly this software is not capable of determining time corrections automatically with one hundred percent reliability, but then that was never the aim. As we discussed at the beginning of this chapter that is an almost impossible task. All stations with possible timing errors are flagged but some of these will naturally be false alarms arising mainly from poor data quality, and it is at this point where human intervention is required. However, the software does successfully reduce the time required to correct a year of data from six stations down to just a few hours compared with a few days or even weeks by conventional methods.

These results show that the program is capable, at the very least, of detecting all timing errors. If the error is within -125 to +125 samples it will also make a suggestion as to the magnitude of the error with reasonable reliability and precision to the per sample level. A user switching to interactive mode after examining the automatic analysis can increase this precision to within half a sample for most data, and to within a few seconds if the data are of very high quality.

Chapter 6

Residual Concatenation

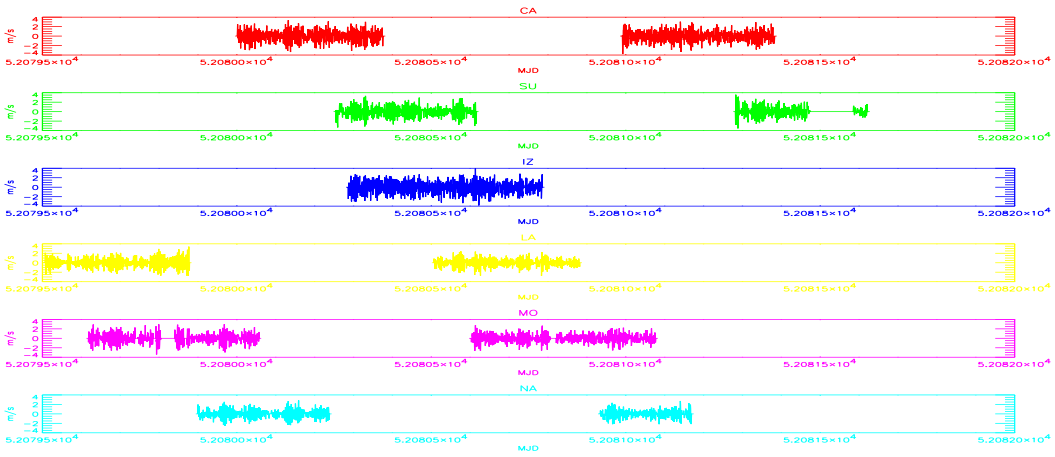

Figure 6.1: Several days of residuals from each station on the network showing the overlapping nature of the data.

The main component of this project was the production of software for concatenating the daily data, such as in figure 6.1, into a long timeseries.

Although some software is currently available for this purpose it was written many years ago in an old version of FORTRAN, and is also very difficult to develop futher. The new software, presently unnamed, is written in IDL - the Interactive Data Language - and is based around an Object Oriented

philosophy. The program is extremely flexible due to a fully procedural design and a novel internal data structure. If new functionality is required it is simply a matter of writing a new module and plugging it in.

Fundamental to the program philosophy is the maximisation of code efficiency and the minimisation of data held in memory. During runtime there is only ever one array passed between procedures, and this holds all data loaded into the program. Element zero of this array always contains the 'DataInfo' object which holds all the information about the residuals such as number of files loaded, number of locations, start and end dates, etc. It also contains the long timeseries when this has been generated. From element one onwards each element contains an instance of the 'BiSONres' object, which is one day or 'block' of data. All of the BiSONres objects have pseudo-intelligence in that they can tell you any of their properties when asked such as their figure of merit, whether they contain a certain date, or with which other block numbers they overlap. The array is always sorted into time order with the earliest time at the beginning and the latest time at the end, such that the start time of block x is always earlier than the start time of block $x+1$. Block number x by definition can only overlap with blocks with numbers less than x since blocks greater than x are in the future and do not 'exist' yet.

With the data organised in this way concatenation is fairly simple. We start at block one and move along the array asking each block in turn which other blocks it overlaps with, the overlap is processed, and we move onto the next block. This does lead to a certain amount of duplicate processing of some data but it is the only reliable way of dealing with the thousands of orientations of blocks and gaps that could arise. However, with the speed of

modern processors this does not cause a problem - seven years of data from six sites can be concatenated easily within half an hour.

6.1 Combination Methods

One of the main requirements of this software was the ability to add new methods of combining data from overlapping stations. There are currently seven methods available.

6.1.1 Priority

If the priority method is selected, the overlaps are combined by selecting data based on the five minute figure of merit (FMFOM), defined as the ratio of power in the five minute band to that in the high frequency band. The overlapping stations are ranked in order of FOM from lowest to highest with higher being better. The lowest FOM is selected first, then data are gradually overwritten by the stations with higher figure of merits until no more data is available.

6.1.2 Low Frequency Optimisation (Mode 1)

This method works in the same way as the Priority method, but overlapping stations are ranked in order of mean power in the 0.8 to 1.3mHz band (LP1) from highest to lowest with lower being better.

6.1.3 Low Frequency Optimisation (Mode 2)

This method works in the same way as the Priority method, but overlapping stations are ranked in order of mean power in the 0.2 to 0.7mHz band (LP2) from highest to lowest with lower being better.

6.1.4 Low Frequency Optimisation (Mode 3)

This method works in the same way as the Priority method, but overlapping stations are ranked in order of low frequency figure of merit (LP1FOM) from lowest to highest with higher being better. The LP1FOM is defined as the ratio of power in the 0.8 to 1.3mHz band to the power in the high frequency band.

6.1.5 Low Frequency Optimisation (Mode 4)

This method works in the same way as the Priority method, but overlapping stations are ranked in order of low frequency figure of merit (LP2FOM) from lowest to highest with higher being better. The LP2FOM is defined as the ratio of power in the 0.2 to 0.7mHz band to the power in the high frequency band.

6.1.6 Basic Average

So far with the five previous methods some data has been discarded. This is not an efficient way of using the data. By taking an average of all overlapping residuals there is the potential for the signal to noise in the long timeseries

to be increased. This method treats all data equally by combining with a basic even-weights average.

6.1.7 Weighted Average

There is a greater potential gain in signal-to-noise if the residuals are merged with a weighted-average. This is where all data from all stations are used, but greater emphasis is placed on those data with a higher measure of quality such as the figure of merit.

Let the appropriate weight for the overlap data from each site be W_j and let there be n sites in the overlap. If combined in the time domain, the combined residual is given according to

$$v_{combined}(t) = \frac{\sum_{j=1}^{n} W_j v_j(t)}{\sum_{j=1}^{n} W_j}$$

In order to take full statistical advantage of the extra data obtained by the overlap, the weights must be set appropriately. The correct weights are proportional to the inverse of the square of the sample standard deviations of the noise sources for each of the overlapping timeseries (Chaplin et. al. 1997)

Hence,

$$W_j = \frac{1}{\sigma_j^2}$$

where j is the station number such that $1 \leq j \leq n$, and σ^2 is as discussed in chapter three.

Chaplin concluded achieving only a few percent improvement in signal to noise ratio of a two month timeseries when using a weighted-merging technique compared to using a best-quality-station-only overlap technique. The results of this analysis will show if similar conclusions can be drawn when using seven years of data.

6.2 Results

Seven timeseries were produced using these concatenation methods. The noise rejection threshold was set to 200 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ over the 0.8 to 1.3mHz frequency band. This value was chosen such that the very worst data would be rejected but would still give a high fill level. Dispite the fact that, as we saw in chapter four, a lower threshold gives a higher quality spectrum this value gives a higher percentage of overlaps and hence should show a greater difference between overlap combination methods.

The results of this analysis are shown in full in appendix B. Here we will look at the effects of the various concatenation methods on just one mode of oscillation instead of all sixteen in order to simplify the comparison between each threshold. The l=1, n=10, mode at 1612.73μHz is shown in figure 6.2 for each concatenation method. Some statistics about each timeseries are given in table 6.1.

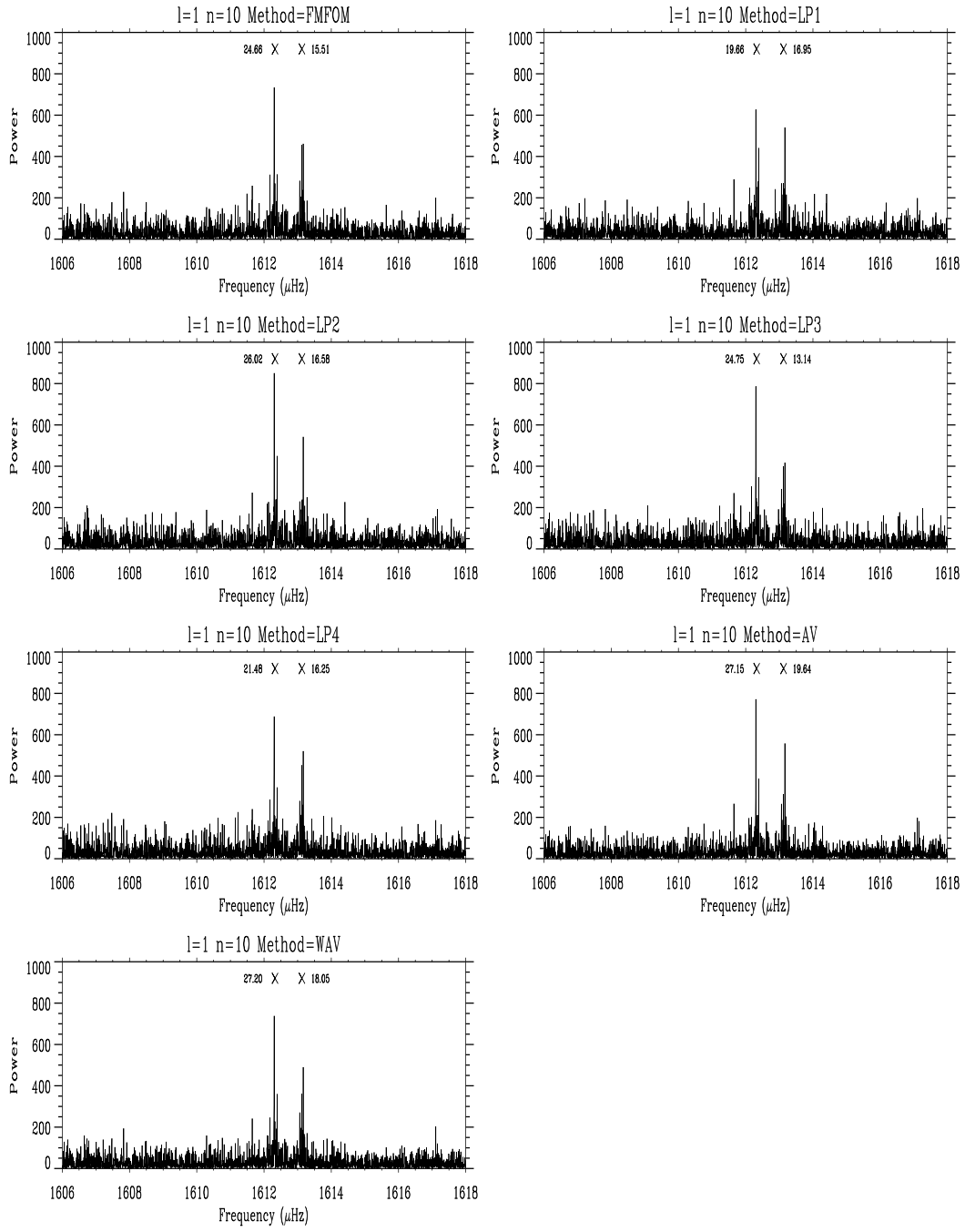

Figure 6.2: The effects on detectability of the l=1, n=10, mode at 1612.73μHz for different concatenation methods. The theoretical position of each component of the mode is marked with an 'x', and the figures are the signal to noise ratio of each component.

Method	SU	CA	NA	MO	LA	IZ
FM Priority	18.0	13.6	14.7	11.8	31.0	10.9
LP1-FOM	19.1	13.6	15.5	13.3	26.9	11.6
LP2-FOM	20.7	13.3	15.9	12.4	26.1	11.6
LP1	12.3	16.0	11.1	14.7	28.3	17.6
LP2	8.4	18.2	9.1	16.4	25.8	22.1
Average	22.1	22.3	16.6	24.4	34.1	32.1
W. Average	22.1	22.3	16.6	24.4	34.1	32.1

Table 6.1: Fill per station for different overlap methods. Threshold was set at 200, giving an overall fill of 77.4 percent for each timeseries.

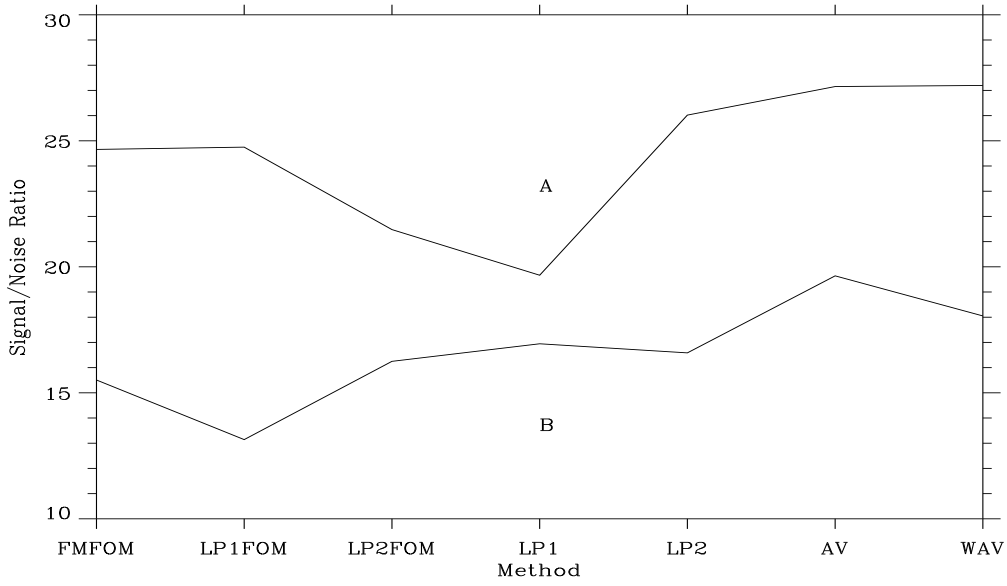

Figure 6.3: The signal to noise ratio of each component of the l=1, n=10, mode at 1612.73μHz for different concatenation methods. The 'A' and 'B' traces correspond to the m=-1 and m=+1 components respectively.

From table 6.1 we can see how the percentage fill from each individual station varies between concatenation methods. Both methods taking an average of each station have exactly the same fill per station since they are using all available data. Each station should be capable of providing about eight hours of data per day given good weather and no mechanical problems. This would provide a 33 percent fill per station, but only Las Campanas and Izaña are anywhere near this level.

Using figure 6.3 we can make a quantitative decision as to which concatenation method provides the best results. The signal to noise ratio of each component was calculated by comparing the power in the peak of the component to that of the background noise level surrounding it. Clearly taking an average of the overlapping data produces the best results, as we expected. The weighted average has produced no improvement on the stronger component, and has actually degraded the weaker component. The low frequency priority mode two has also produced a good result, since we are attempting to minimise the noise power over the whole of the low frequency band. As before when we looked at different rejection thresholds, the methods have really made little difference to the weaker component of this mode since the noise levels are much more dominant.

By looking at the power spectrum as a whole we can determine if the results found for this particular mode also apply for other modes.

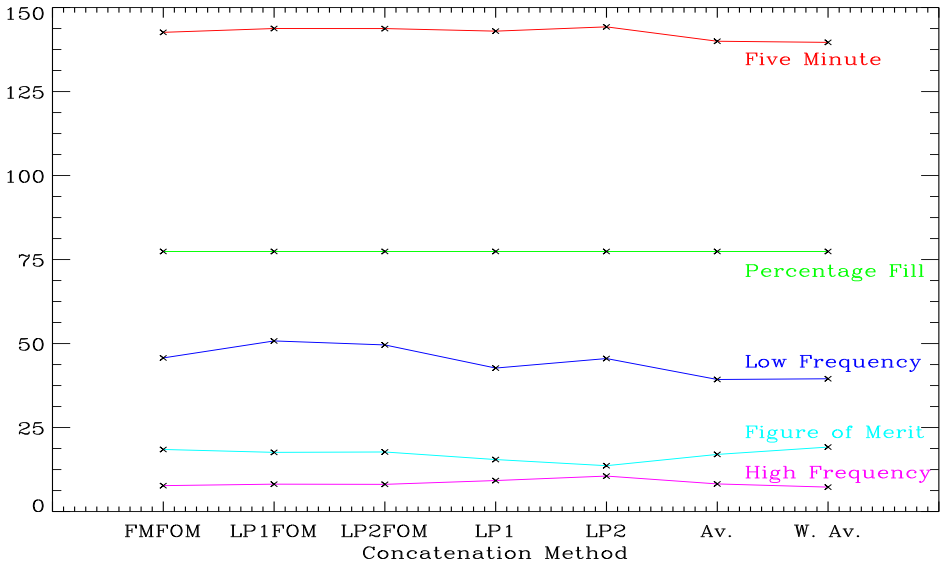

Figure 6.4: Mean five minute, high, and low frequency power, five minute figure of merit ratio, and percentage fill against concatenation methods. All powers are measured in $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$.

Method	Low Noise Power $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$	High Noise Power $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$	FOM
FM Priority	45.7	7.72	18.5
LP1-FOM	50.8	8.16	17.6
LP2-FOM	49.6	8.10	17.7
LP1	42.7	9.2	15.5
LP2	45.5	10.6	13.7
Average	39.3	8.23	17.0
W. Average	39.5	7.27	19.2

Table 6.2: Noise powers and FOM per overlap method.

Table 6.2 shows the variation in low frequency noise, high frequency noise, and the figure of merit ratio between five minute power and high frequency power. Figure 6.4 represents this graphically. The fill level remains constant for each method as expected since the rejection threshold has not been changed. We saw earlier that if residuals selected for a long timeseries have good low frequency noise performance then they usually have good high frequency noise performance as well. From these results we can see that this also works in reverse. The two timeseries using a selection method based on low frequency power, LP1-FOM and LP2-FOM, both have increased power in that region as expected but also slightly increased power in the high frequency and five minute band. Both the average and the weighted average have good low frequency performance, but the weighted average has much better high frequency noise performance and a correspondingly high figure of merit.

When looking at the data quality we had the requirement of maximising the fill level whilst at the same time minimising the mean low frequency noise power. If we now look at the same relationship between $fill/noise$ we can determine how the quality of these timeseries compare with those generated earlier.

Figure 6.5 shows the value of this ratio for each concatenation method. The basic average has produced a slightly higher result than the weighted average due to its lower low frequency noise power. The best result we achieved earlier was 0.0203 with a rejection threshold of 50 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ and a fill of just 55 percent. By taking an average of all station overlaps we have achieved a value of 0.0197 which is just under three percent lower, but with a rejection threshold of 200 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ and a fill of over 77 percent.

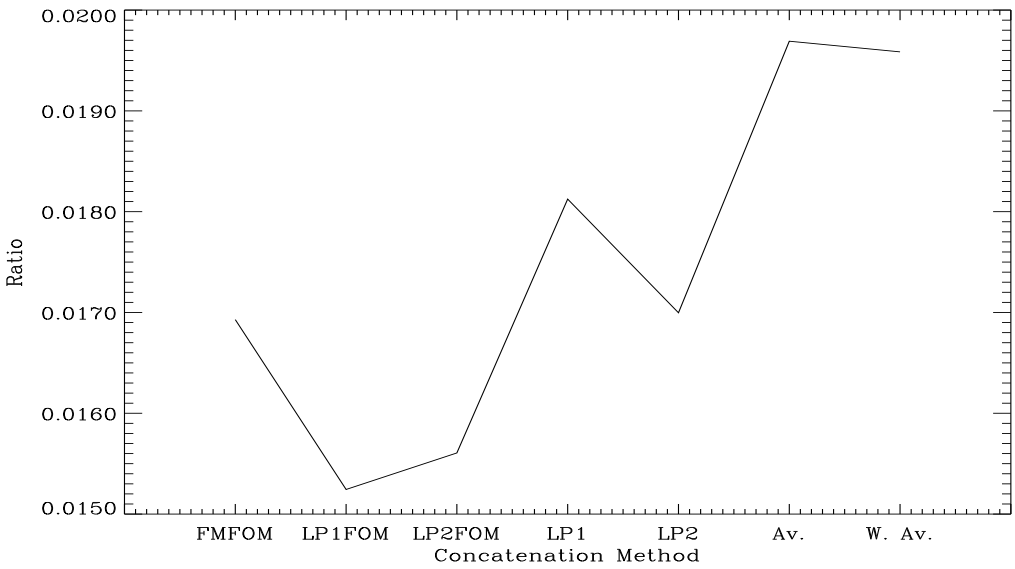

Figure 6.5: The ratio of fill to low frequency noise for different concatenation methods.

Overall the weighted average has produced the best result. Although the low frequency noise is slightly higher then the basic average the weighted average has a much better high frequency noise performance and a correspondingly higher five minute figure of merit.

If we now take these results and produce a final timeseries using a rejection threshold of 60 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ which is where the peak occurred in figure 4.9 on page 47, and an overlap treatment using a weighted average of all stations we achieve a timeseries with a mean low frequency (LP1) noise power of just 26.9 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$, and a mean high frequency (HF) noise power of just 6.7 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$, the lowest noise powers of all the timeseries we have produced so far.

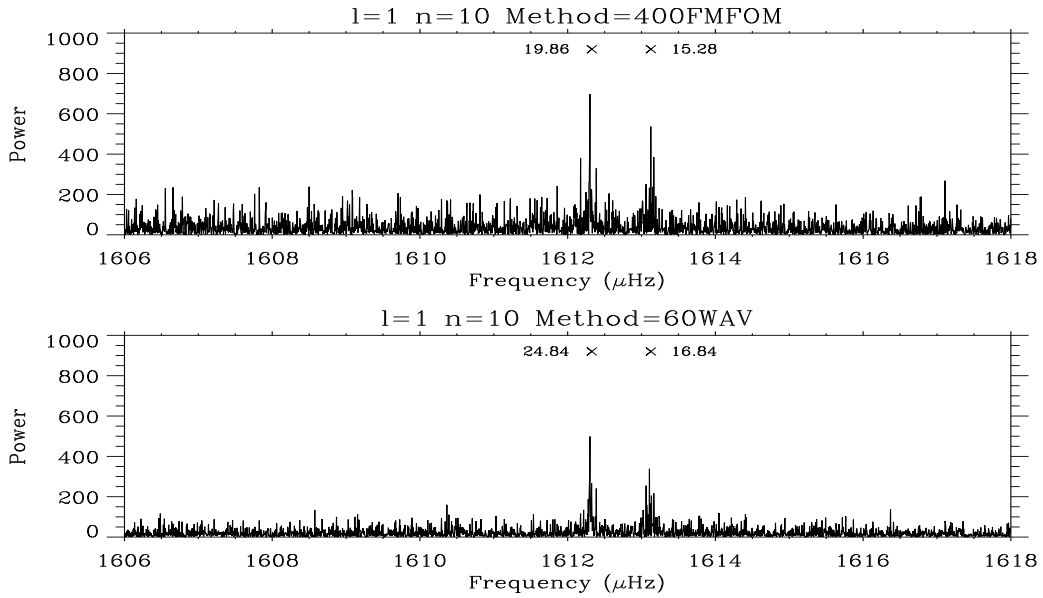

Figure 6.6: Comparing the mode detectability between the 'worst' dataset produced using a 400 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ rejection threshold and a combination method based on the five minute FOM, and the 'best' dataset produced using a 60 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ rejection threshold and a weighted average of each overlapping station.

With a five minute figure of merit ratio of over twenty, and a $fill/noise$ ratio of just over 0.022 one would expect this to be the best of all timeseries produced during this project.

However, looking at figure 6.6 which compares the mode detectability between the 'worst' dataset produced using a 400 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ rejection threshold and a combination method based on the five minute FOM, and the 'best' dataset produced using a 60 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ rejection threshold and a weighted average of each overlapping station, this does not necessarily seem to be the case. While an improvement in signal to noise ratio has been achieved, it is not as high as the result we obtained earlier with a 200 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ rejection threshold and a weighted average of each overlapping

station. This is due to the fact that the higher rejection threshold allows for more overlapping regions, and hence the more sophisticated concatenation routines have more data with which to work.

The results of this analysis show a greater improvement in data quality when using a weighted average of overlapping data than those achieved by Chaplin (Chaplin et. al. 1997), but the increase in signal to noise is still only just over ten percent. The reason for this limited improvement is simple; in order to provide a reasonable overall signal to noise gain the noise level of the overlapping data must be of comparable quality, and as we saw in chapter four the quality from each station is quite variable. Unfortunately, the majority of overlaps occur between sites with an old and a new spectrometer resulting in one 'good' dataset and one 'bad' dataset.

By upgrading the older spectrometers such that all instruments would be of similar quality the signal to noise gain when combining stations using a weighted average would increase considerably.

Chapter 7

Conclusions and Future Work

The aim of this project was to provide a suite of programs that would deliver an improvement in the quality of long timeseries produced from the daily residual data files from the Birmingham Solar Oscillations Network.

In order to achieve this goal it was first necessary to produce an in-depth analysis of the quality of the data obtained from each site on the network. This allowed us to assess the seasonal variation in high and low frequency noise level throughout a year and to assign typical values to measures of quality such as the five minute figure of merit. We then used this assessment to apply various data rejection thresholds based on the low frequency noise power power in an attempt to find the balance between the undoubted advantage of adding in more signal by using data, and the possibly disproportionate disadvantage of adding more noise due to low data quality. The aim was to find the point at which the gains of rejecting more bad data using a lower noise threshold become offset by the decrease in fill level (and correspondingly larger gaps) of the long timeseries. The highest threshold of 400 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ would reject only the very worst data giving the maximum

fill level, while the lowest threshold of 15 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ would cause the majority of data to be rejected resulting in a very low fill level. The optimised point for modes in the 0.8 to 1.3mHz regime was found to be at a rejection threshold of 60 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ giving a fill level of sixty percent, where the ratio of $fill/noise$ was improved by one third.

The next step was to check the timing of the data. If data from more than one station are to be combined into a long timeseries then it is imperative that all stations have correct relative timing such that the residuals can be combined coherently. This means that the observed phase is consistent from one station to the next. The data could be incredibly high quality but they would be useless if they were combined out of phase due to timing errors. Historically this was done manually and would take days or sometimes even weeks to complete one year of data, which is clearly unacceptable. As an improvement to this an advanced semi-automatic timing error detection and correction routine was written with a full graphical user interface. After extensive testing we found that the software was capable of detecting both low magnitude errors of just one or two samples, and also large errors on the scale of hours with a sensitivity of within half a sample. While not fully automatic (some user input is required to confirm each detected timing error), the program does successfully reduces the time taken to correct a year of data from six stations to just a few hours.

Finally, the issue of merging data from overlapping sites was approached. Before any new overlap merging algorithms could be implemented and tested it was necessary to write new software from scratch. This was not a trivial task. However, the completed program is much more flexible than its predecessor. If new functionality is required, such as a new merging algorithm,

it is simply a matter of writing a new module and plugging it in. There are currently seven different overlap merging algorithms implemented in the software. Five are variations on the basic priority algorithm using different methods of ranking data in order of quality, and the remaining two take an average and a weighted average of all overlapping data respectively. The weighted average, combined with a rejection threshold of 60 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$, gave an improvement in the signal to noise ratio of just over ten percent.

Future merging algorithms, such as combining data in the frequency domain, should be able to give a further improvement on this figure. For example, if one station has good high frequency performance and the other has good low frequency performance then the best Fourier components of each could be selected as opposed to simply choosing one station or the other.

This project has laid the foundations for extensive further work. With a suite of flexible new software the BiSON data processing pipeline has received a major refurbishment. We have demonstrated that it is possible to improve the quality of long timeseries produced from BiSON data, but that is only the beginning. Work on more sophisticated overlap merging techniques is an exciting prospect for the future.

Chapter 8

References

Anderson, E. R. 1989, Bulletin of the American Astronomical Society, 21, 786.

Brookes J. R., Isaak G. R., & van der Raay H. B. 1976, Nature 259, 92.

Brookes J. R., Isaak G. R., & van der Raay H. B. 1978, Mon. Not. astr. Soc. 185, 1-17.

Deubner F. L. 1975, Astron. Astroph. 44, 371.

Chaplin W. J., et. al. 1995, Preceedings of Fourth SOHO Workshop: Helioseismology, Pacific Grove, California. ESA SP-376.

Chaplin W. J., et. al. 1996, Sol. Phys. 168, 1-18.

Chaplin W. J., et. al. 1997, Astron. Astrophys. Suppl. Ser. 125, 195-205.

Claverie A., et. al. 1979, Nature 282, 591

Elsworth Y. et. al. 1995, Astron. Astrophys. Suppl. Ser. 113, 379-386.

Elsworth Y. 1996, Bull. Astr. Soc. India 24, 181-188.

Fossat E. 1988, in: Seismology of the Sun and Sun-like Stars, ESA SP-286, 161.

Fossat E. 1991, Sol. Phys. 133, 1.

Fossat E. 1992, Astron. Astrophys. 263, 443-452.

The Global Oscillation Group, 1984, Project, National Solar Observatory. Tucson, AZ.

Grec G., Fossat E., & Pomerantz M. A. 1983, Sol. Phys. 82, 55.

Grec G., Fossat E., Gelly B., Schmider F. X. 1991, Sol. Phys. 133, 13.

Harvey J. 1998, Bull. Astr. Soc. India 26, 135-142.

Hill F. & Newkirk G. A. 1985, Solar Physics 95, 201.

Hill F. 1990, in: Berthomieu G., Cribier M. (eds.) Inside the Sun, IAU Coll. 121, 131.

Hill F. 1994, Sol. Phys. 152, 321.

Hoyng P. 1989, Astroph. J. 345, 1088-1103.

Izaak G. 1992-1993, Carnegie Institution of Washington Year Book, 149-156.

Lamb H. 1916, Hydrodynamics - 4th Edition, Cambridge University Press.

Leibacher J. W. & Stein R. F. 1971, Astroph. Let. 7, 191.

Leighton R. 1960, Proc. IAU Symp. #12, 321.

New R. 2003, 'Noise in resonant scattering spectrometers'. Unpublished, in preparation.

Pantel A. & Fossat E., 1995, Astronomical Society of the Pacific Conference Series, Volume 76.

Salabert D., et. al. 2002, Astron. Astrophys. 390, 717-723.

Toutain & Kosovichev, 1998, Astroph. J. 506.

Tomczyk S., Streander, K., Card, G., et al. 1995, Sol. Phys., 159, 1.

Trueblood M. 1991, International Amateur-Professional Photoelectric Photometry Communication, 45, 45.

Ulrich R. 1970, Astroph. J. 162, 993.

Woodard M., & Hudson H. S. 1983, Nature 305, 589.

Appendix A

Data Quality Results

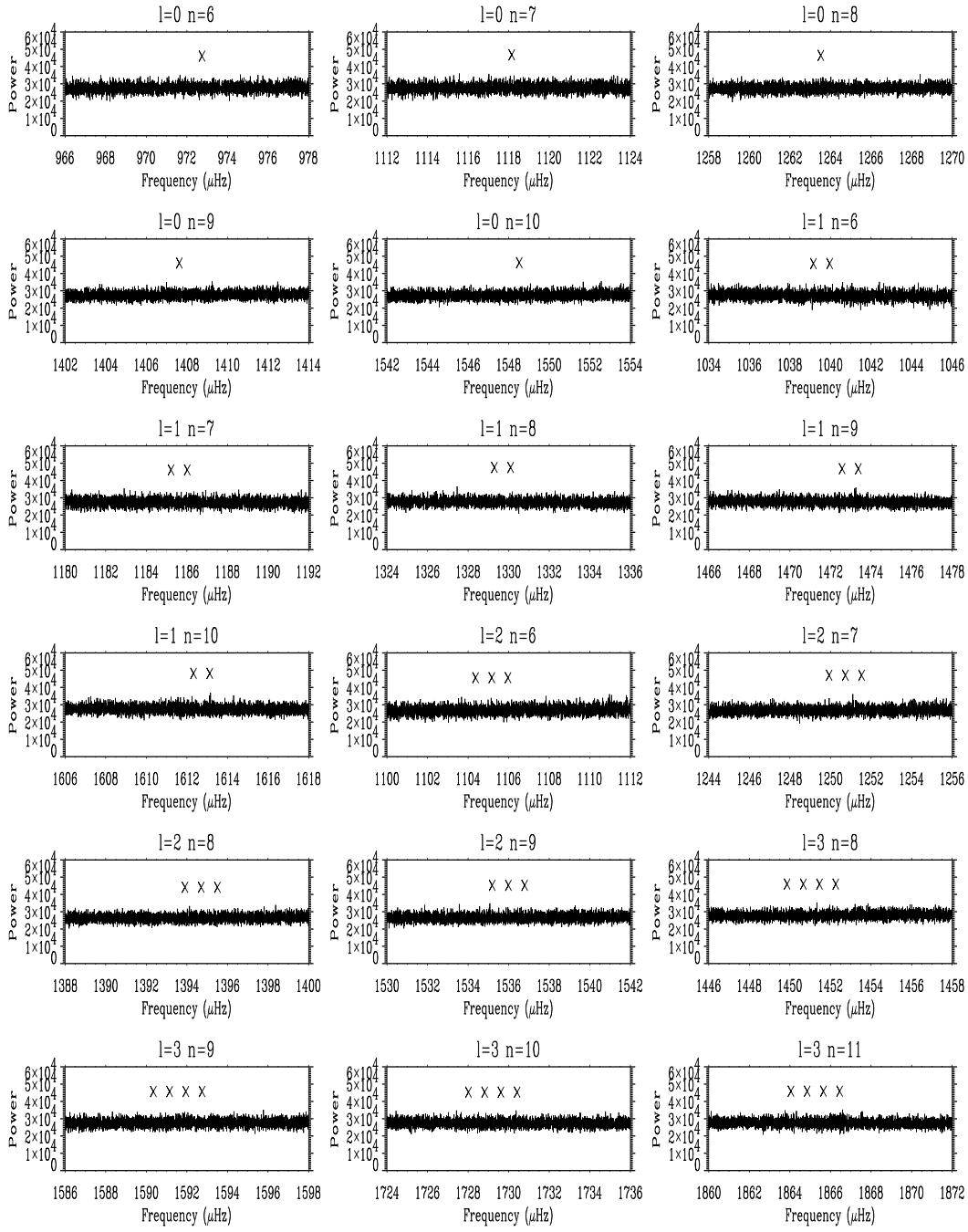

Figure A.1: Power spectrum produced from a timeseries with **no rejection threshold** and a fill of 79.9 percent. Frequencies of predicted low-l modes are marked with an 'x'. Some particularly low quality days of data have manifested themselves as delta functions in the timeseries, and have caused a high intensity white noise in the power spectrum.

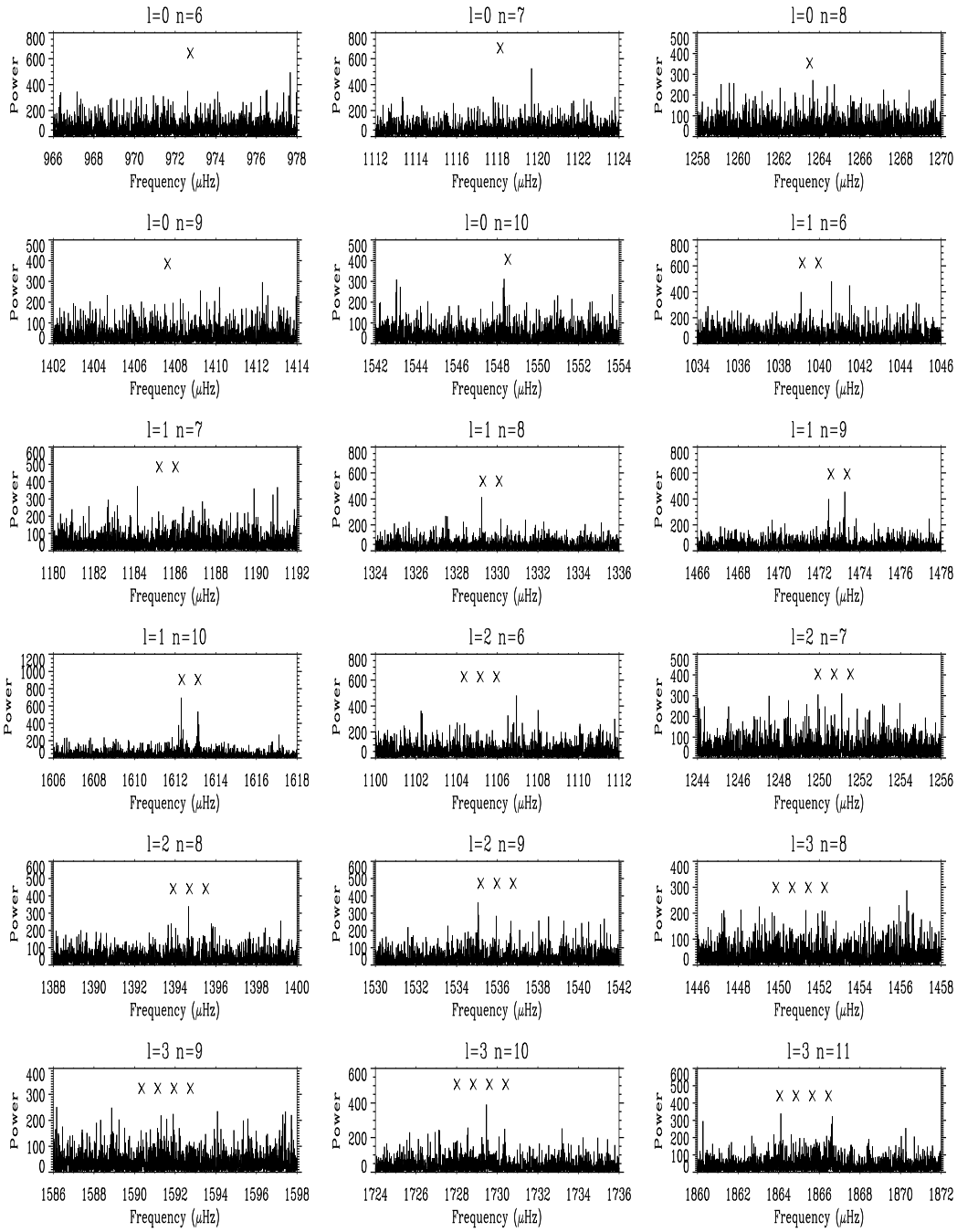

Figure A.2: Power spectrum produced from a timeseries with a **rejection threshold of 400 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$** and a fill of 79.4 percent. Frequencies of predicted low-l modes are marked with an 'x'.

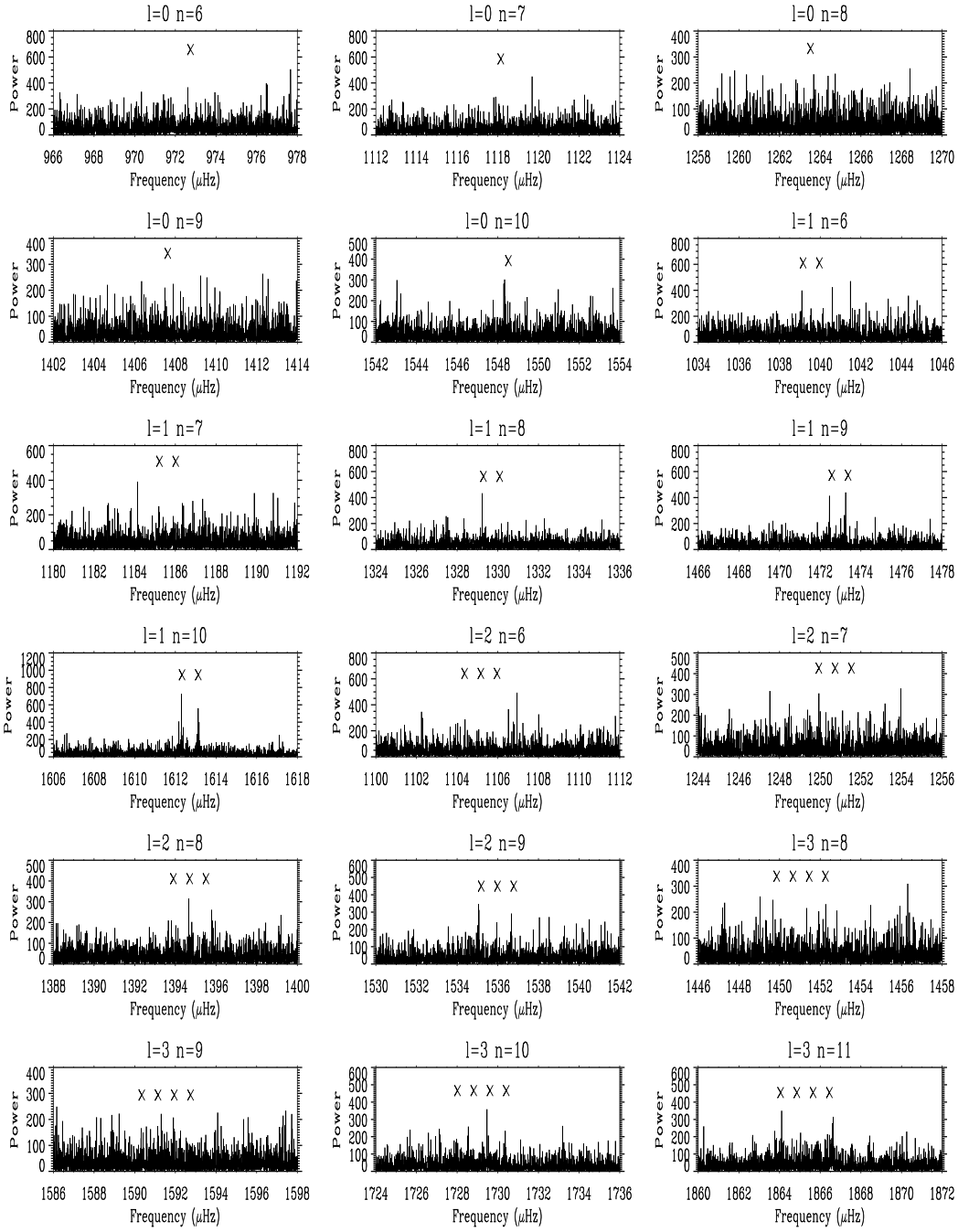

Figure A.3: Power spectrum produced from a timeseries with a **rejection threshold of 350 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$**. Frequencies of predicted low-l modes are marked with an 'x'.

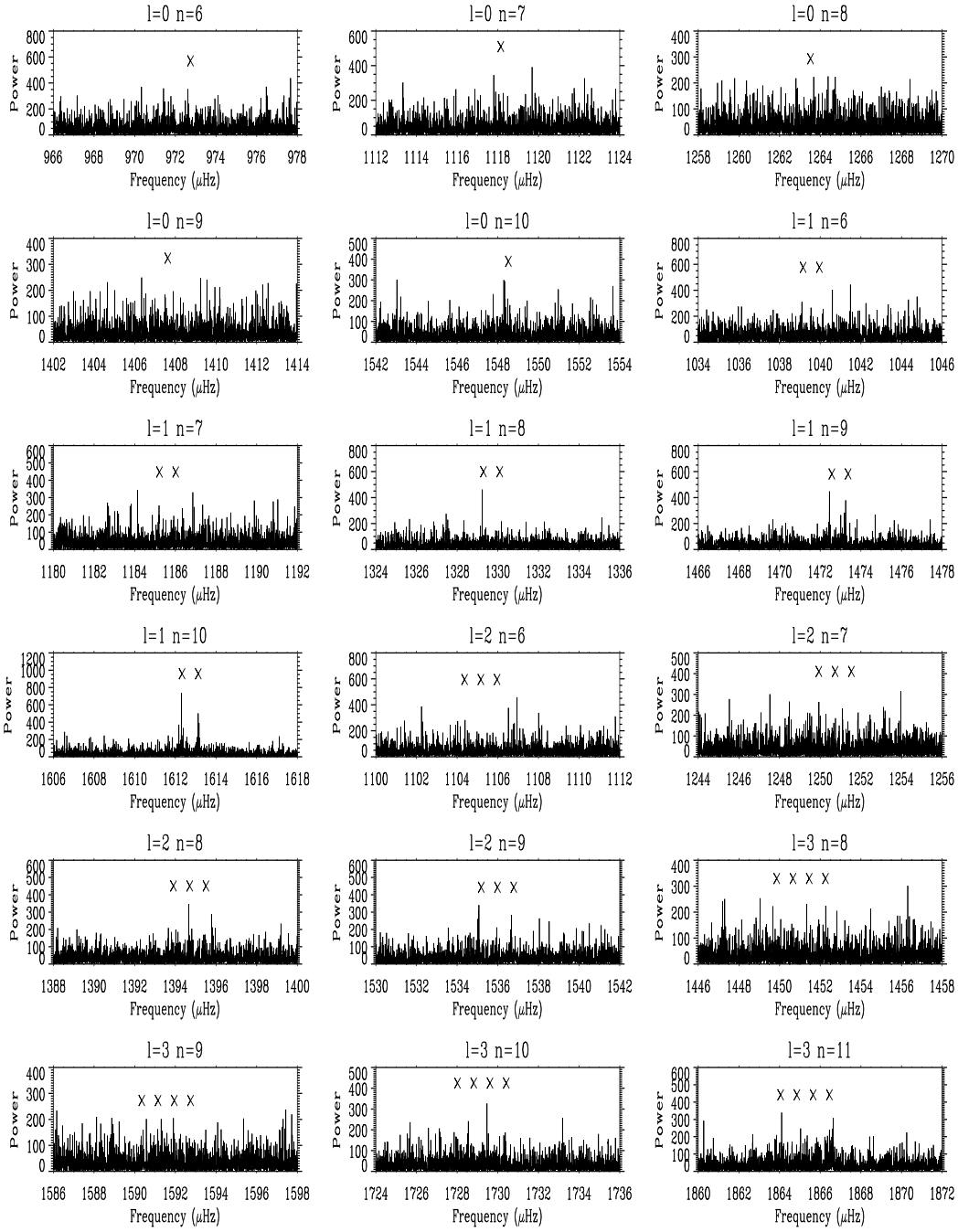

Figure A.4: Power spectrum produced from a timeseries with a **rejection threshold of 300 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$**. Frequencies of predicted low-l modes are marked with an 'x'.

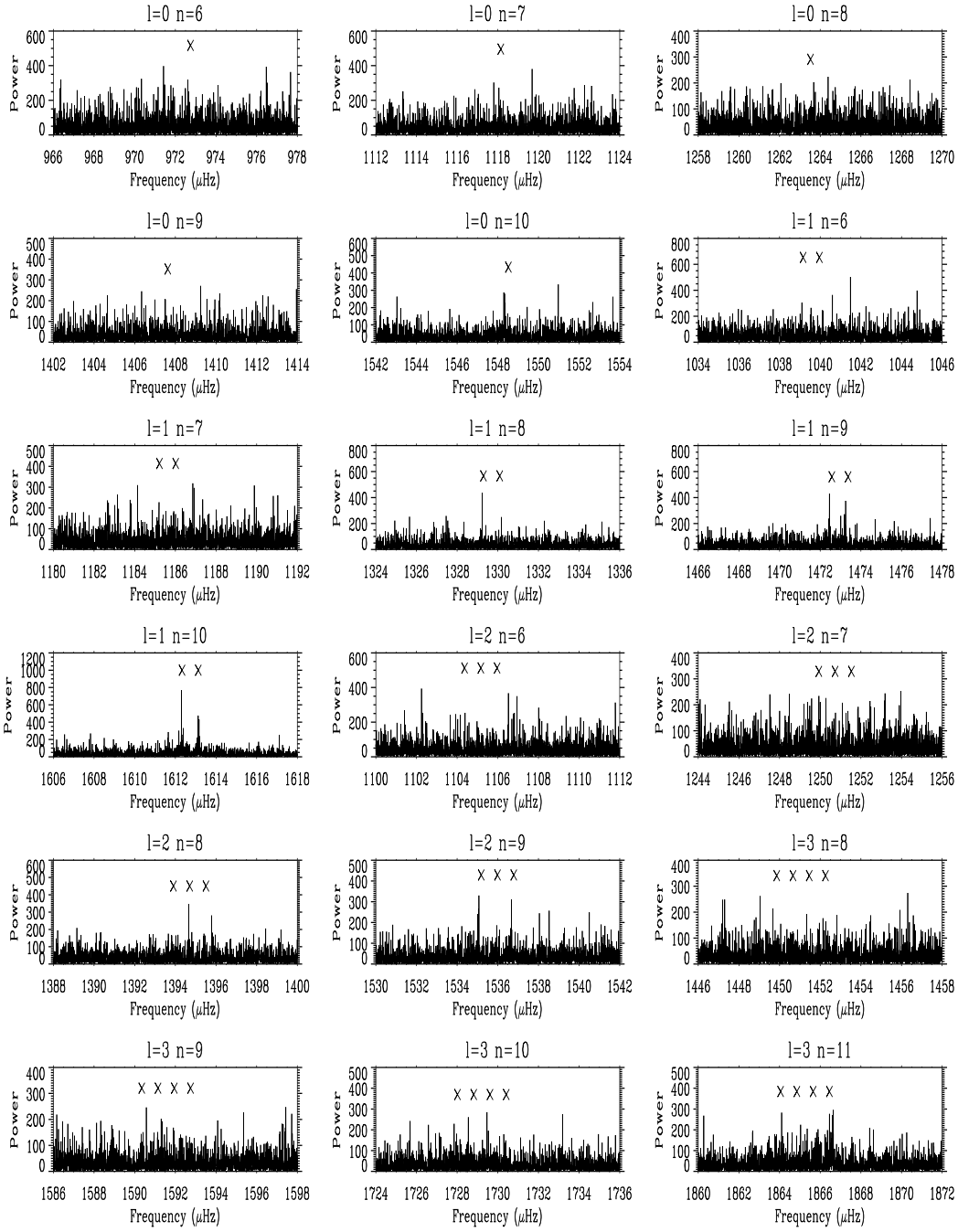

Figure A.5: Power spectrum produced from a timeseries with a **rejection threshold of 250 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$**. Frequencies of predicted low-l modes are marked with an 'x'.

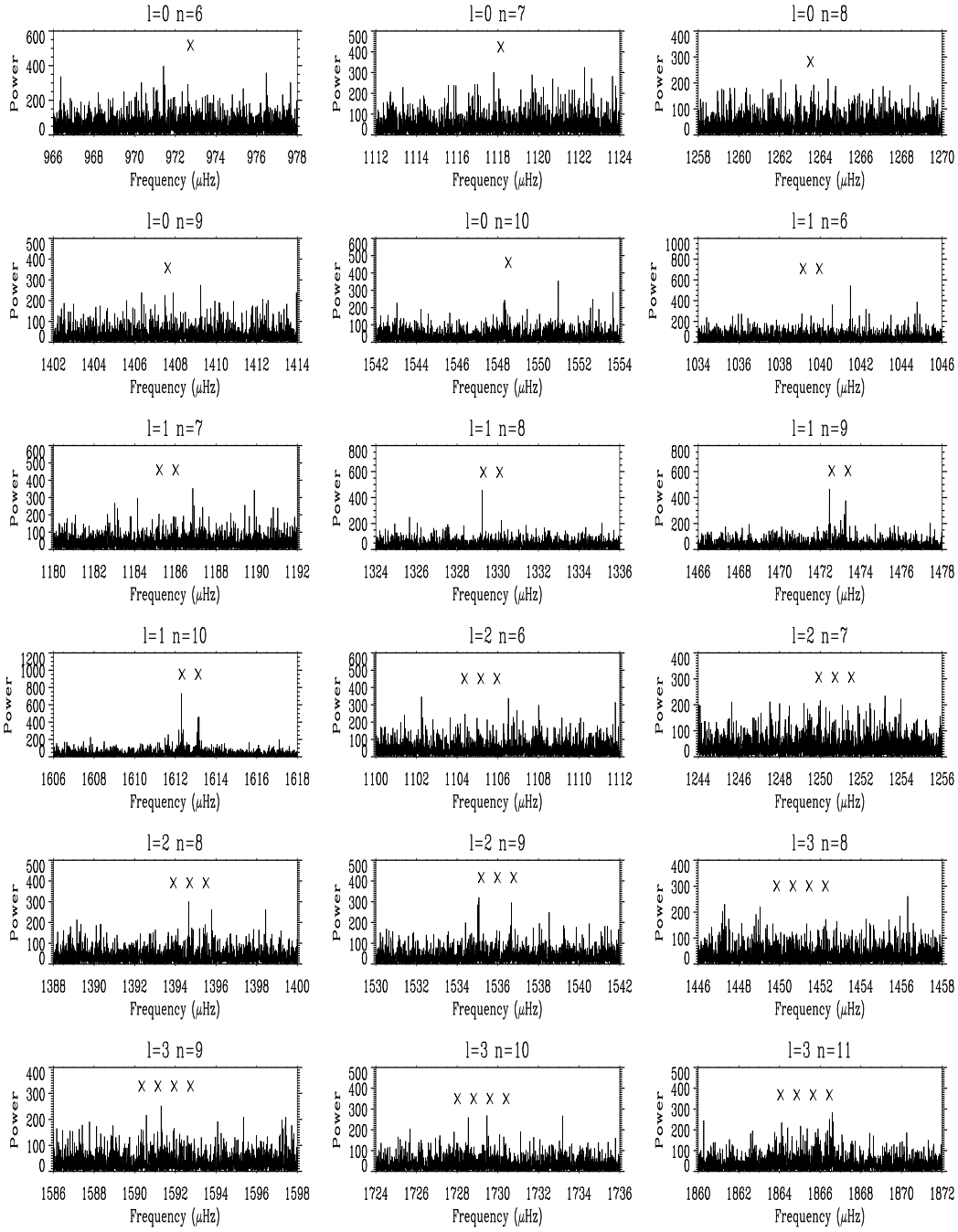

Figure A.6: Power spectrum produced from a timeseries with a **rejection threshold of 200 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$**. Frequencies of predicted low-l modes are marked with an 'x'.

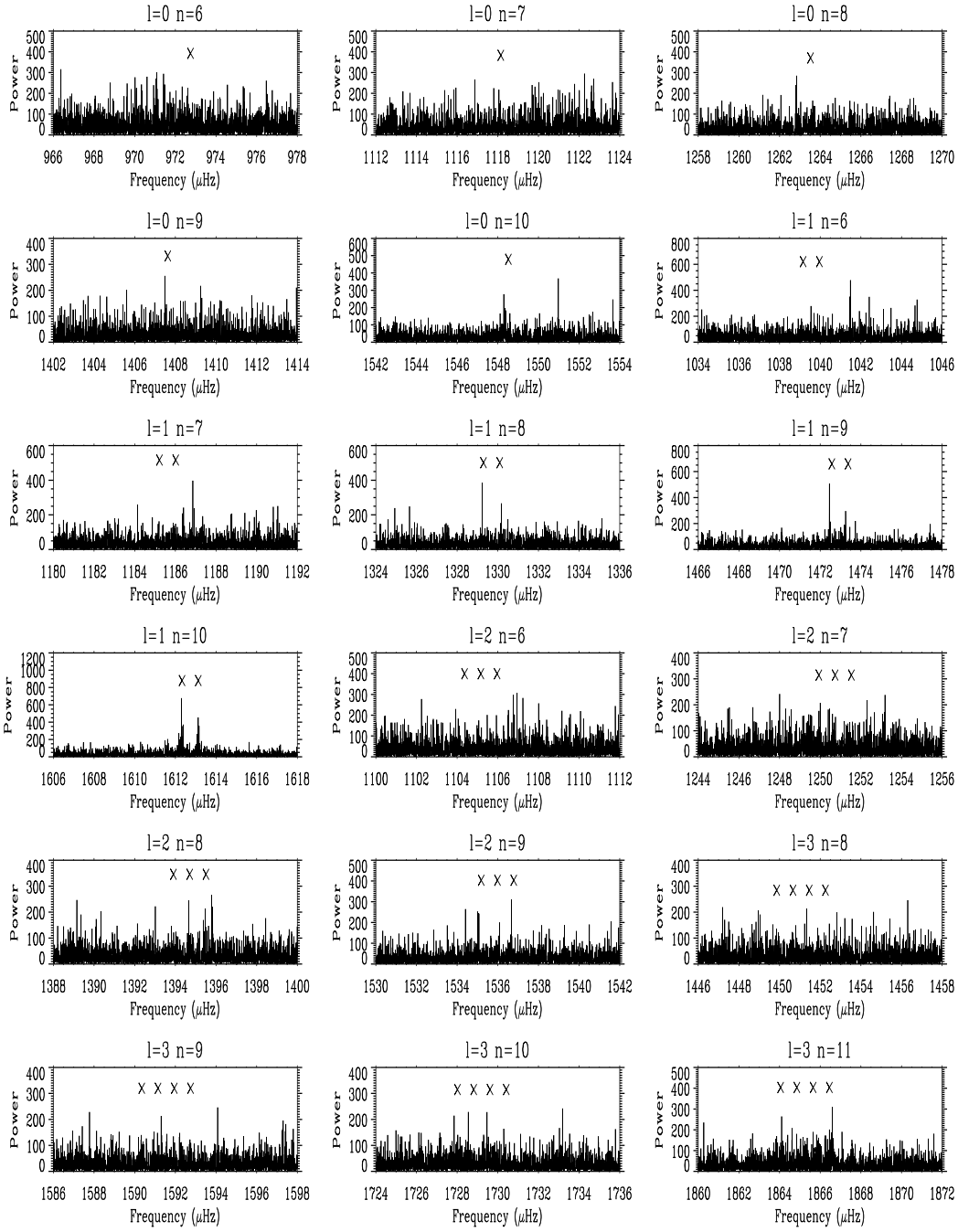

Figure A.7: Power spectrum produced from a timeseries with a **rejection threshold of 150 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$**. Frequencies of predicted low-l modes are marked with an 'x'.

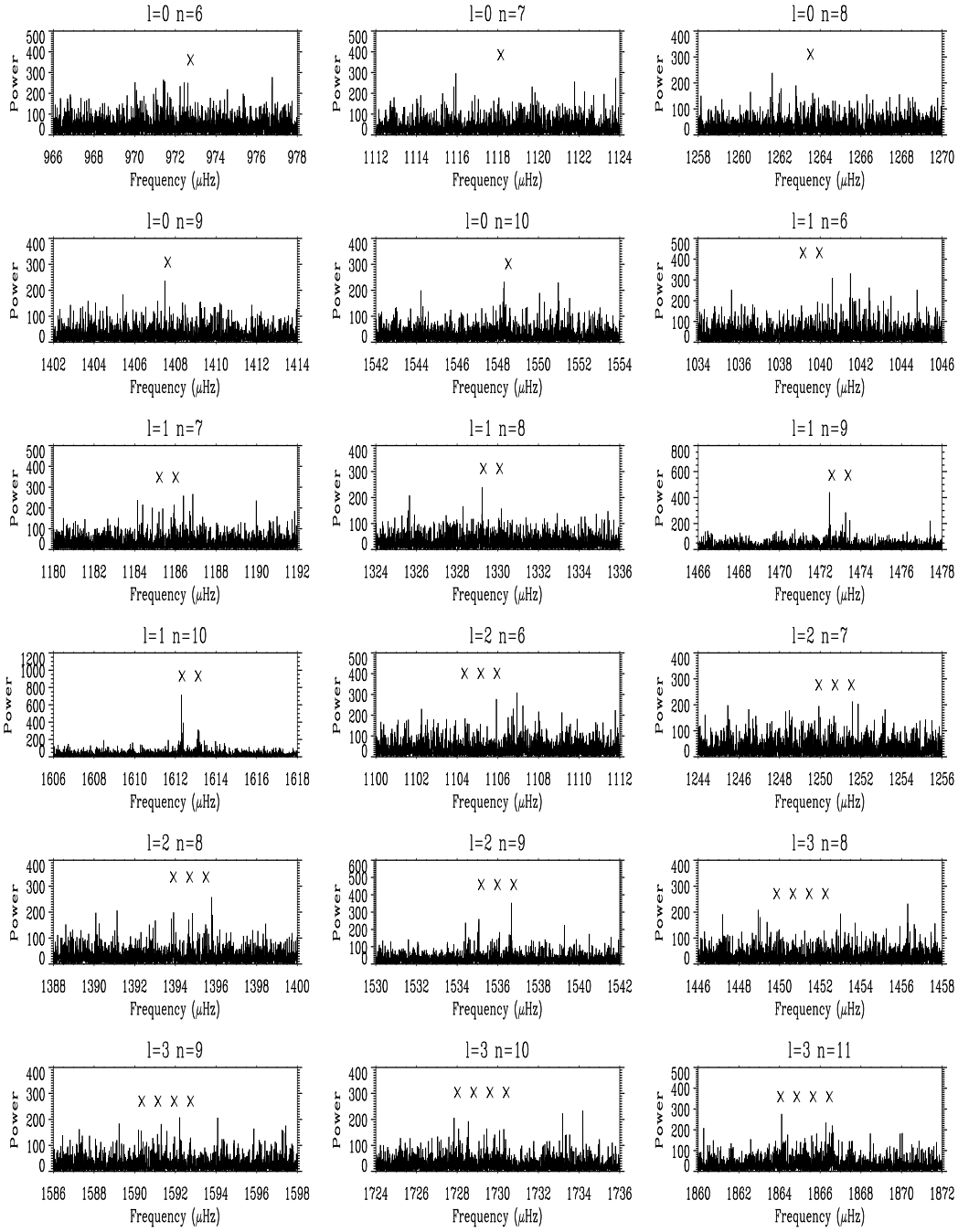

Figure A.8: Power spectrum produced from a timeseries with a **rejection threshold of 100 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$**. Frequencies of predicted low-l modes are marked with an 'x'.

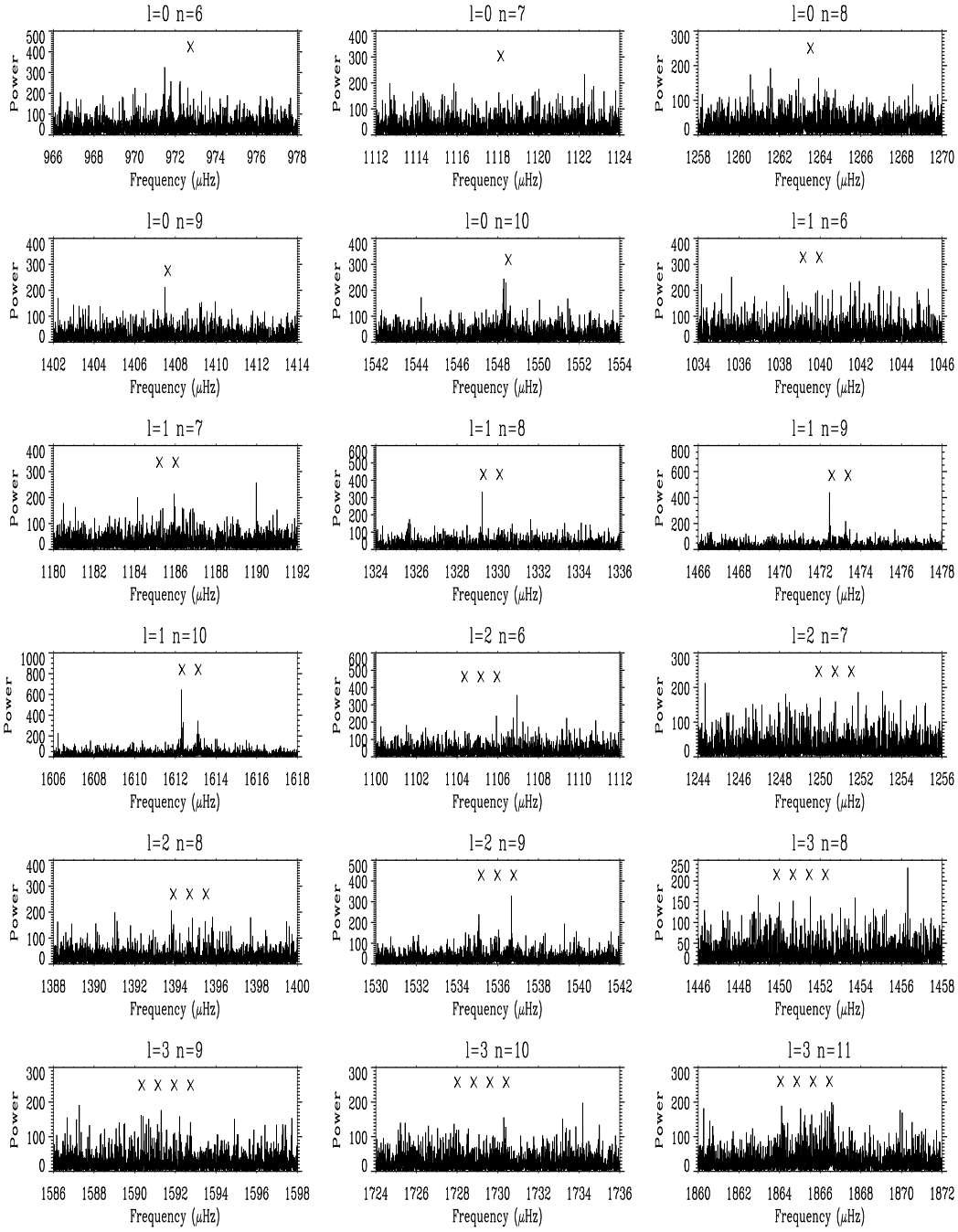

Figure A.9: Power spectrum produced from a timeseries with a **rejection threshold of 75 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$**. Frequencies of predicted low-l modes are marked with an 'x'.

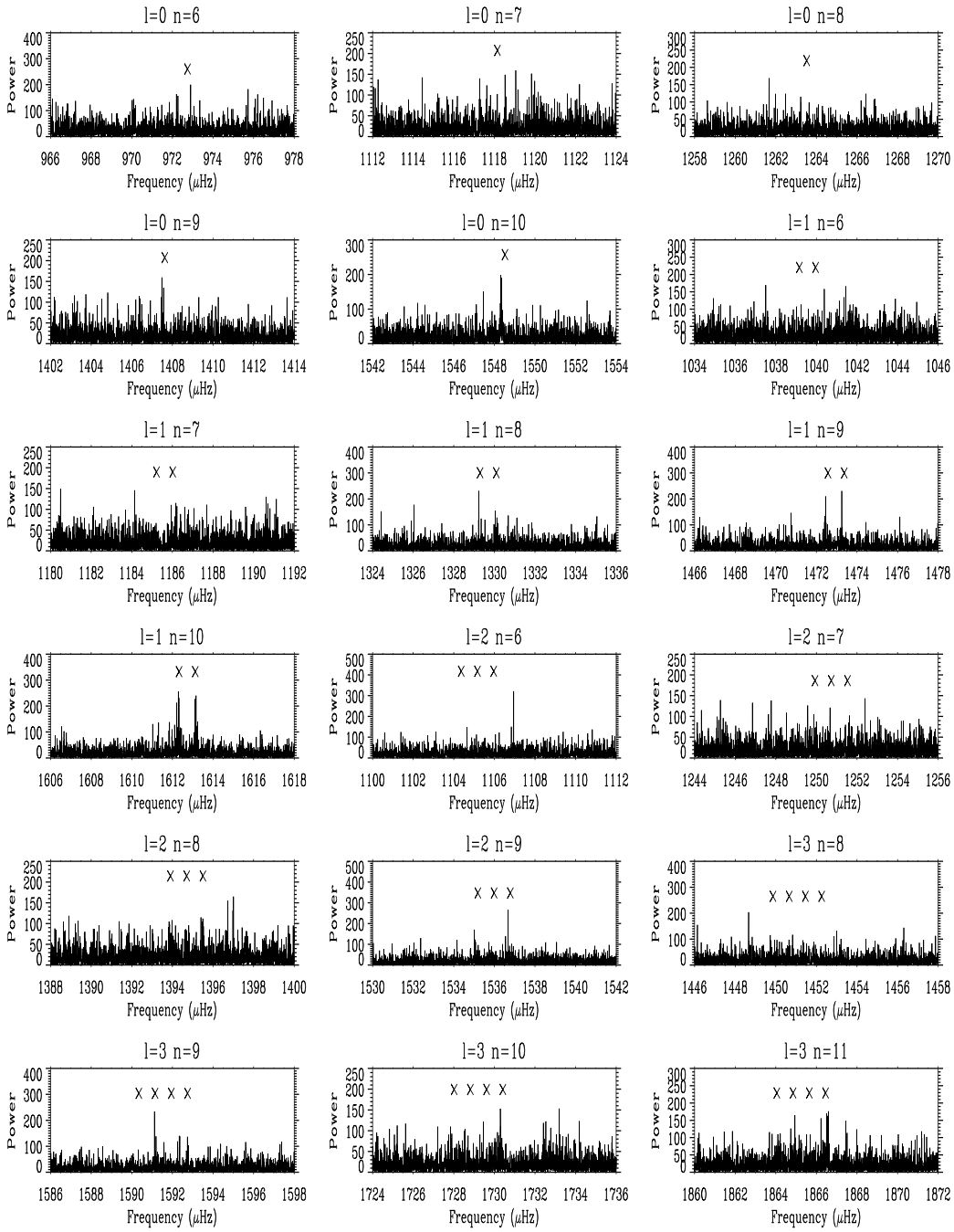

Figure A.10: Power spectrum produced from a timeseries with a **rejection threshold of 50 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$** and a fill of 54.6 percent. Frequencies of predicted low-l modes are marked with an 'x'.

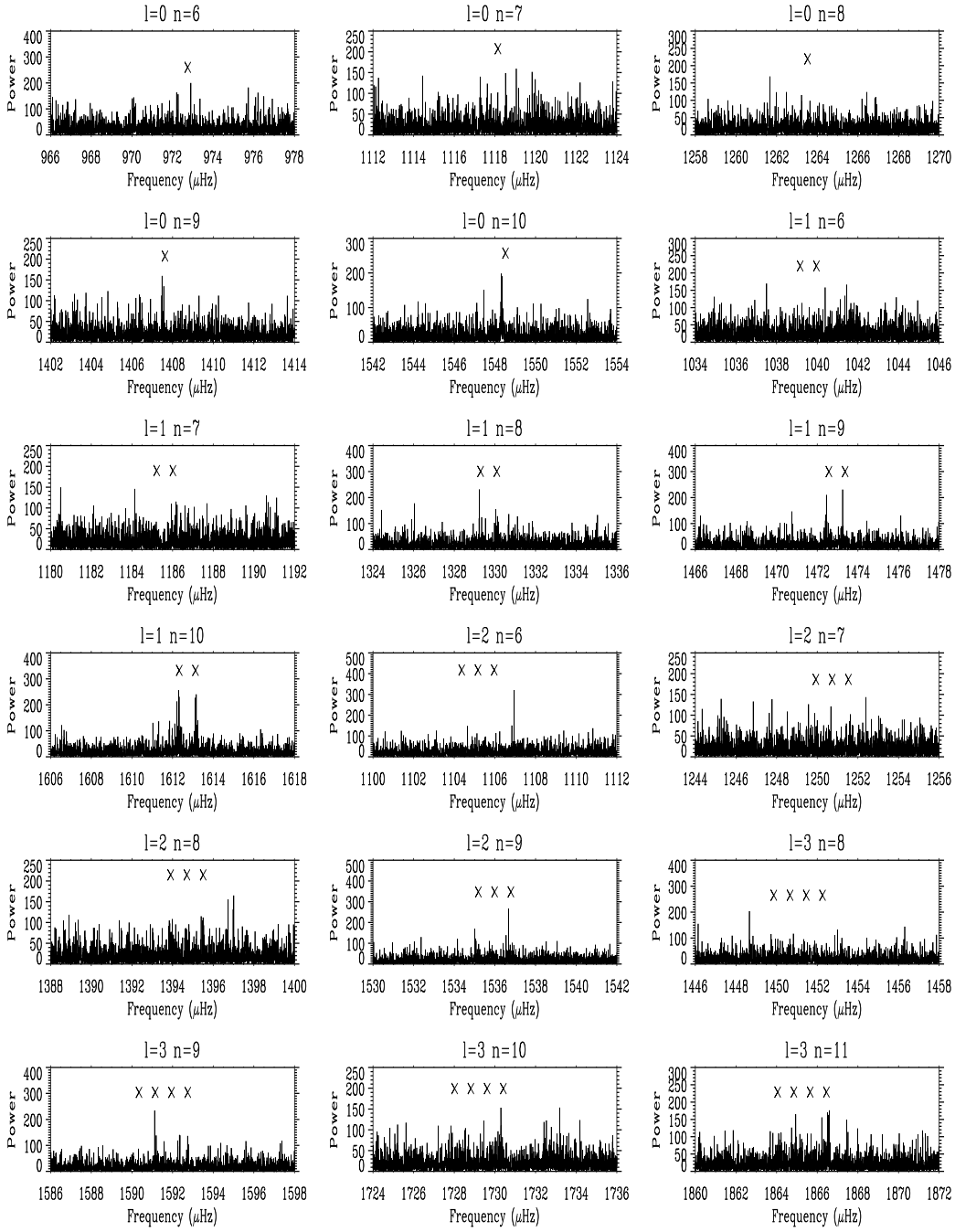

Figure A.11: Power spectrum produced from a timeseries with a **rejection threshold of 37.5 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$**. Frequencies of predicted low-l modes are marked with an 'x'.

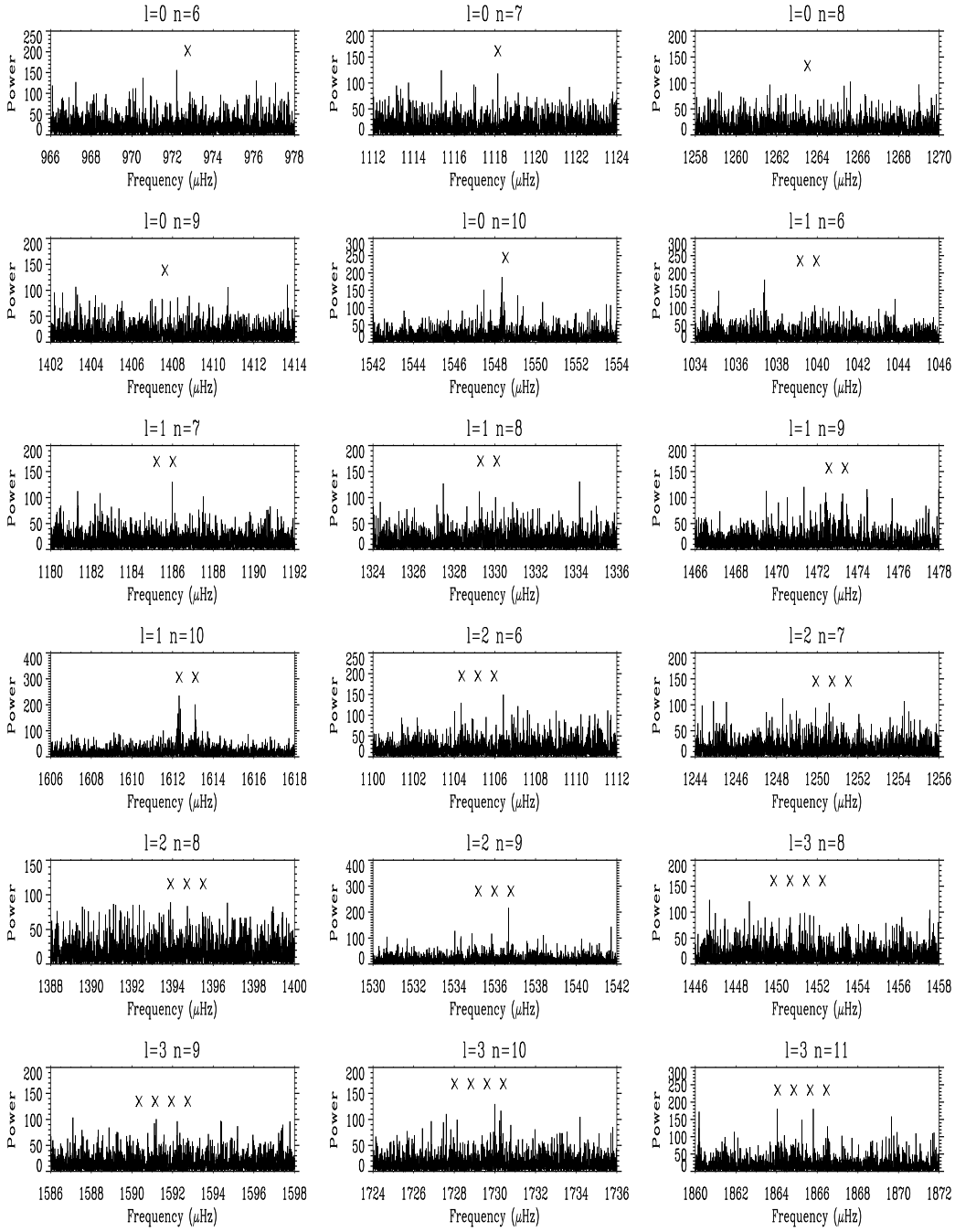

Figure A.12: Power spectrum produced from a timeseries with a **rejection threshold of 25 $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$**. Frequencies of predicted low-l modes are marked with an 'x'.

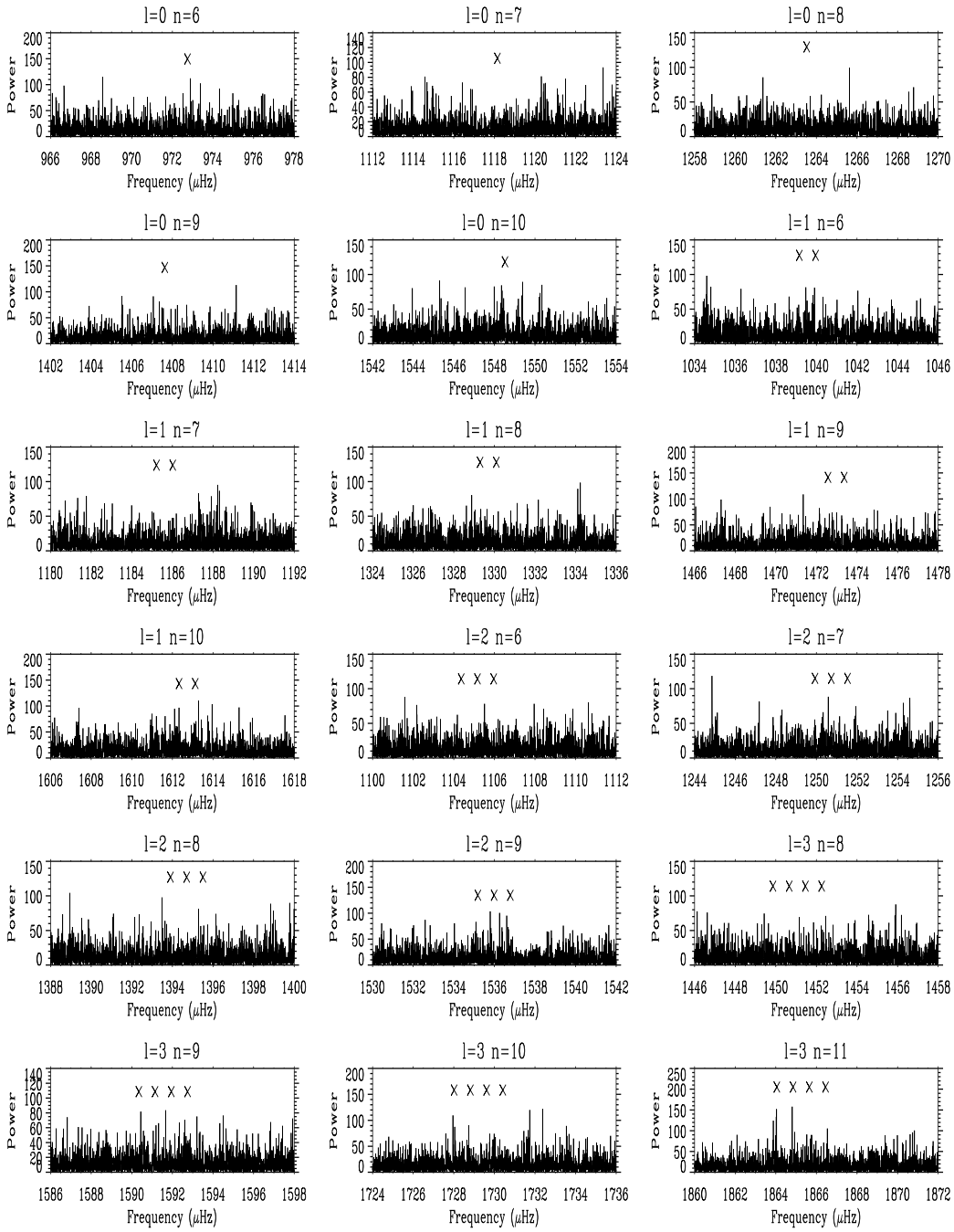

Figure A.13: Power spectrum produced from a timeseries with a **rejection threshold of 15** $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$ and a fill of 8.9 percent. Frequencies of predicted low-l modes are marked with an 'x'.

Appendix B

Concatenation Method Results

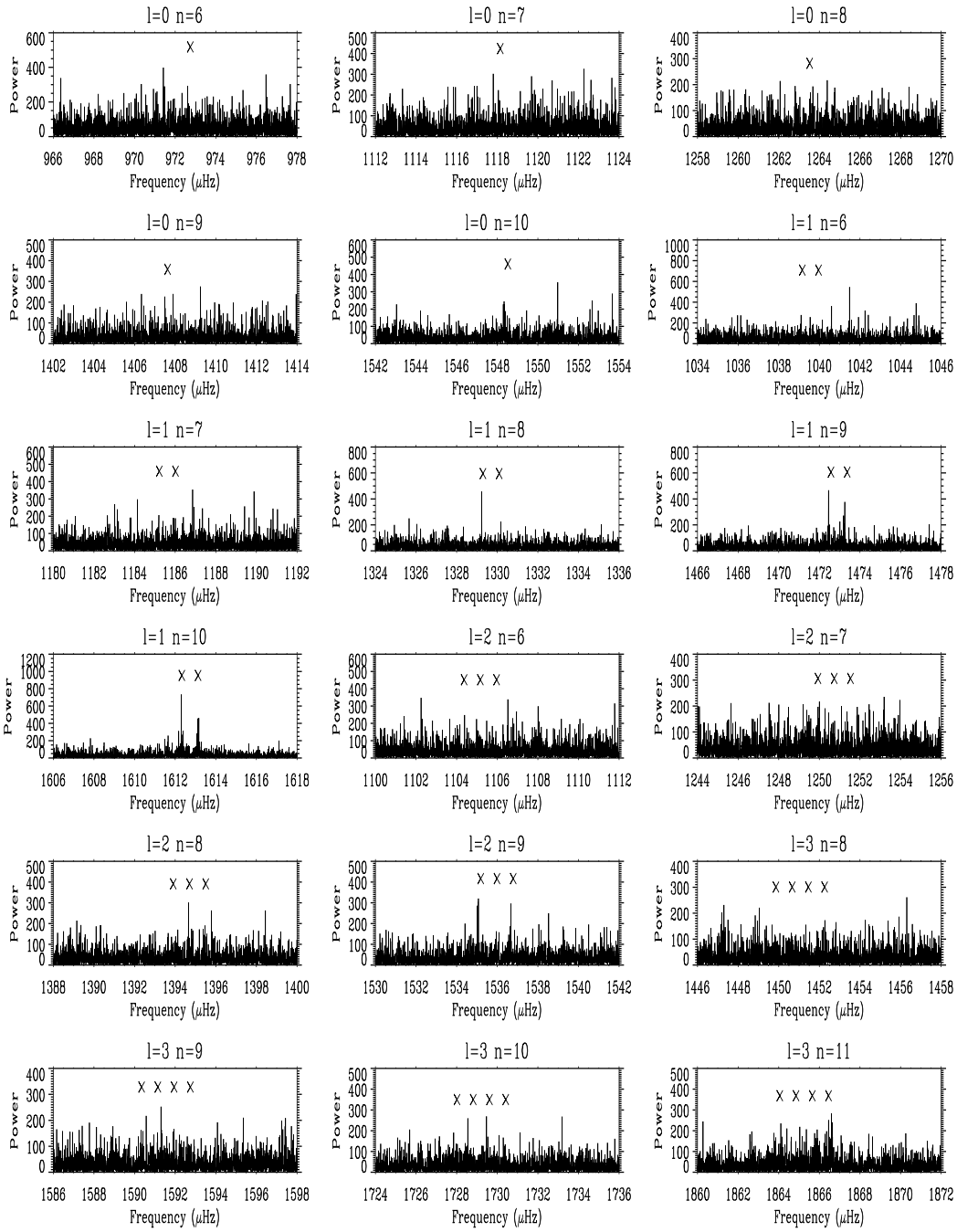

Figure B.1: Power spectrum produced from a timeseries with a rejection threshold of 200 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ with overlap selection based on **five minute priority**. Frequencies of predicted low-l modes are marked with an 'x'.

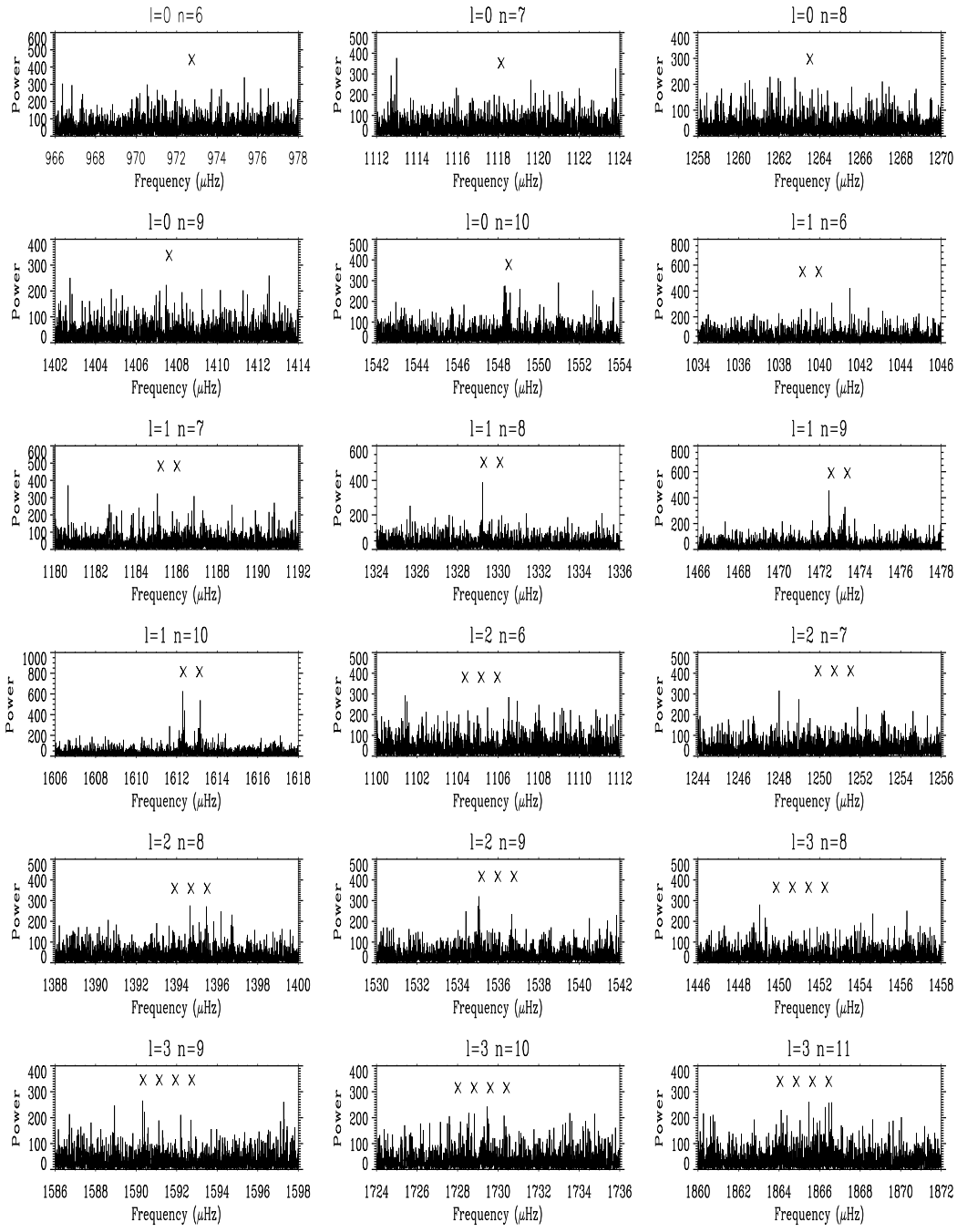

Figure B.2: Power spectrum produced from a timeseries with a rejection threshold of 200 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ with overlap selection based on **low frequency priority mode one**. Frequencies of predicted low-l modes are marked with an 'x'.

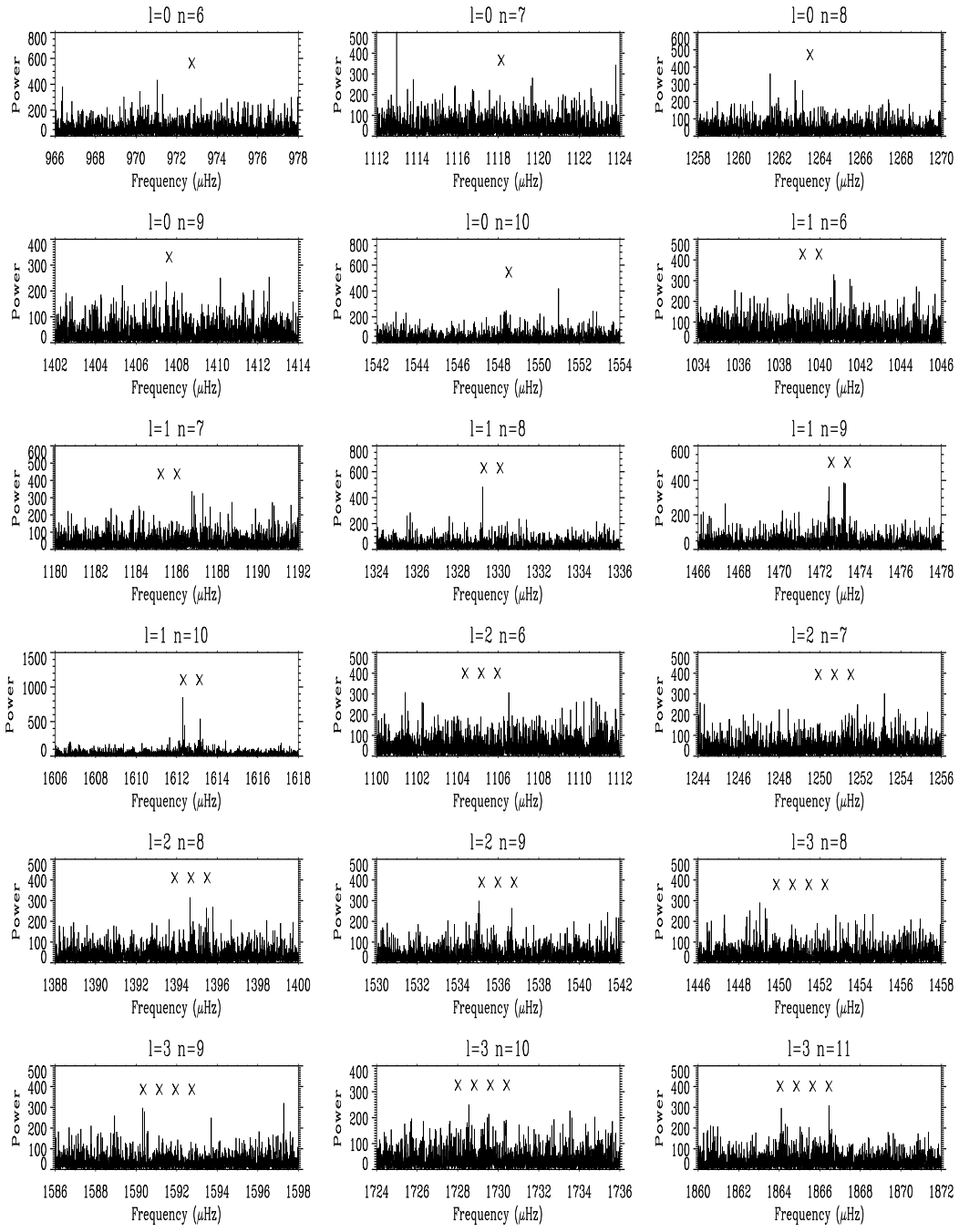

Figure B.3: Power spectrum produced from a timeseries with a rejection threshold of 200 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ with overlap selection based on **low frequency priority mode two**. Frequencies of predicted low-l modes are marked with an 'x'.

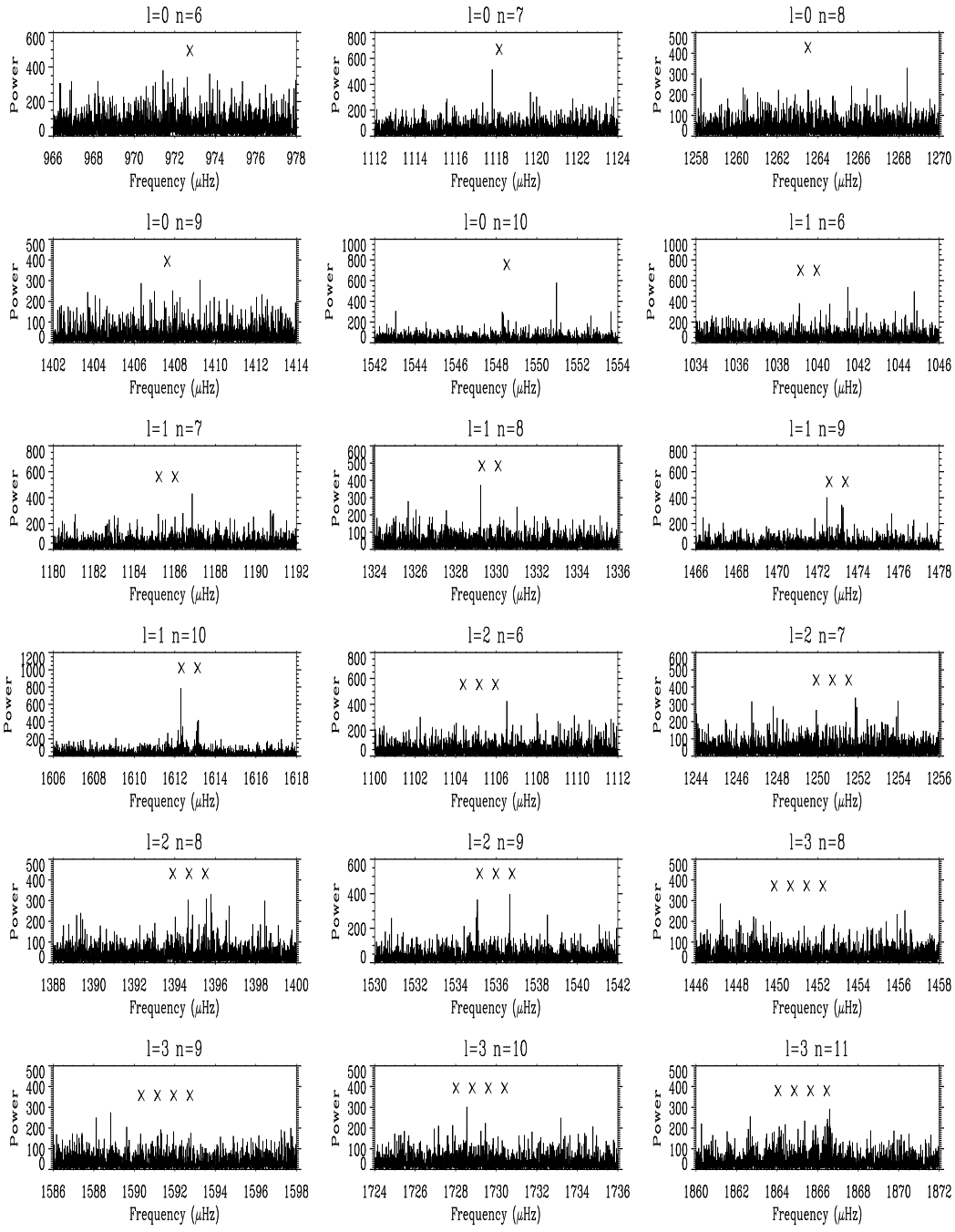

Figure B.4: Power spectrum produced from a timeseries with a rejection threshold of 200 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ with overlap selection based on **low frequency priority mode three**. Frequencies of predicted low-l modes are marked with an 'x'.

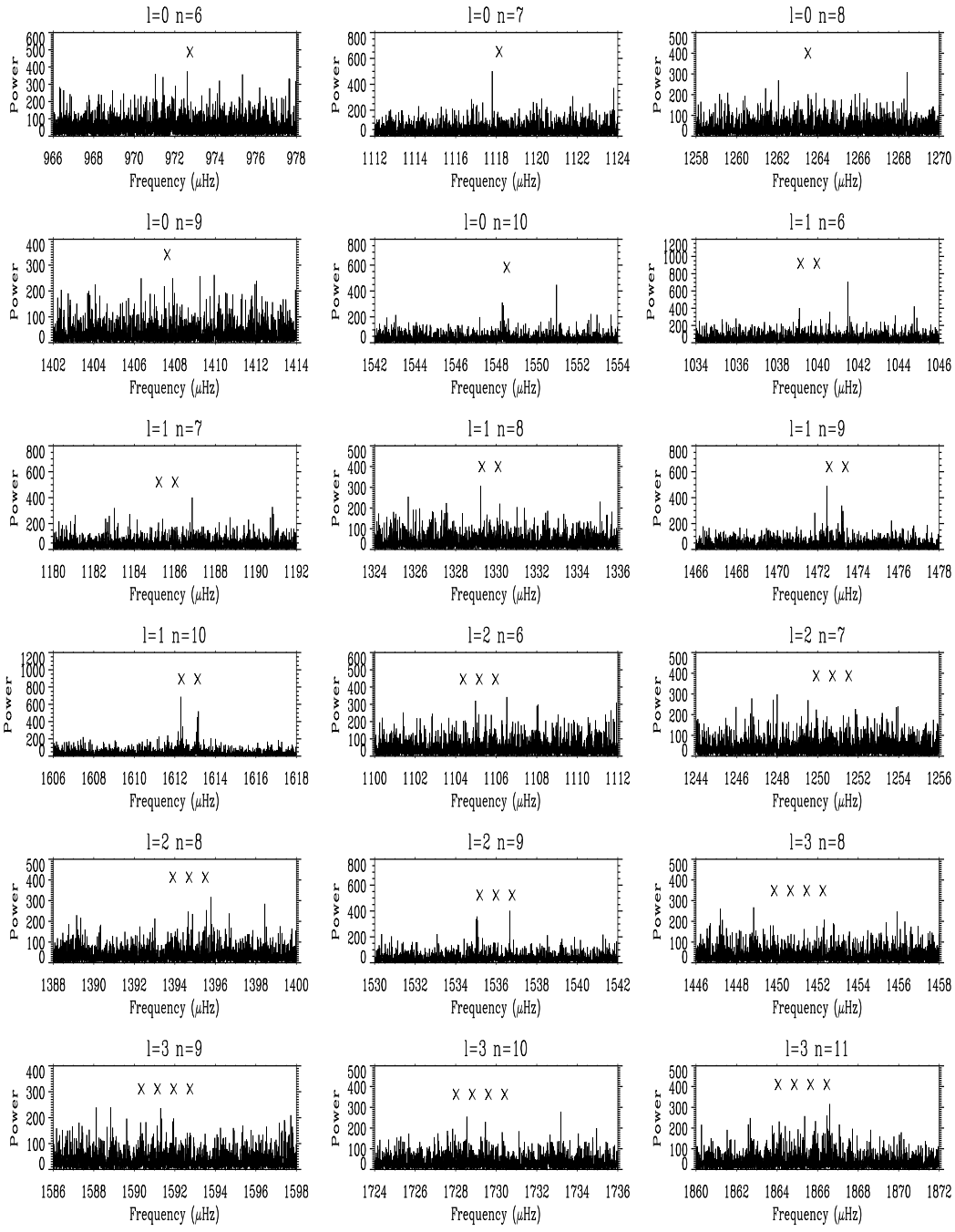

Figure B.5: Power spectrum produced from a timeseries with a rejection threshold of 200 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ with overlap selection based on **low frequency priority mode four**. Frequencies of predicted low-l modes are marked with an 'x'.

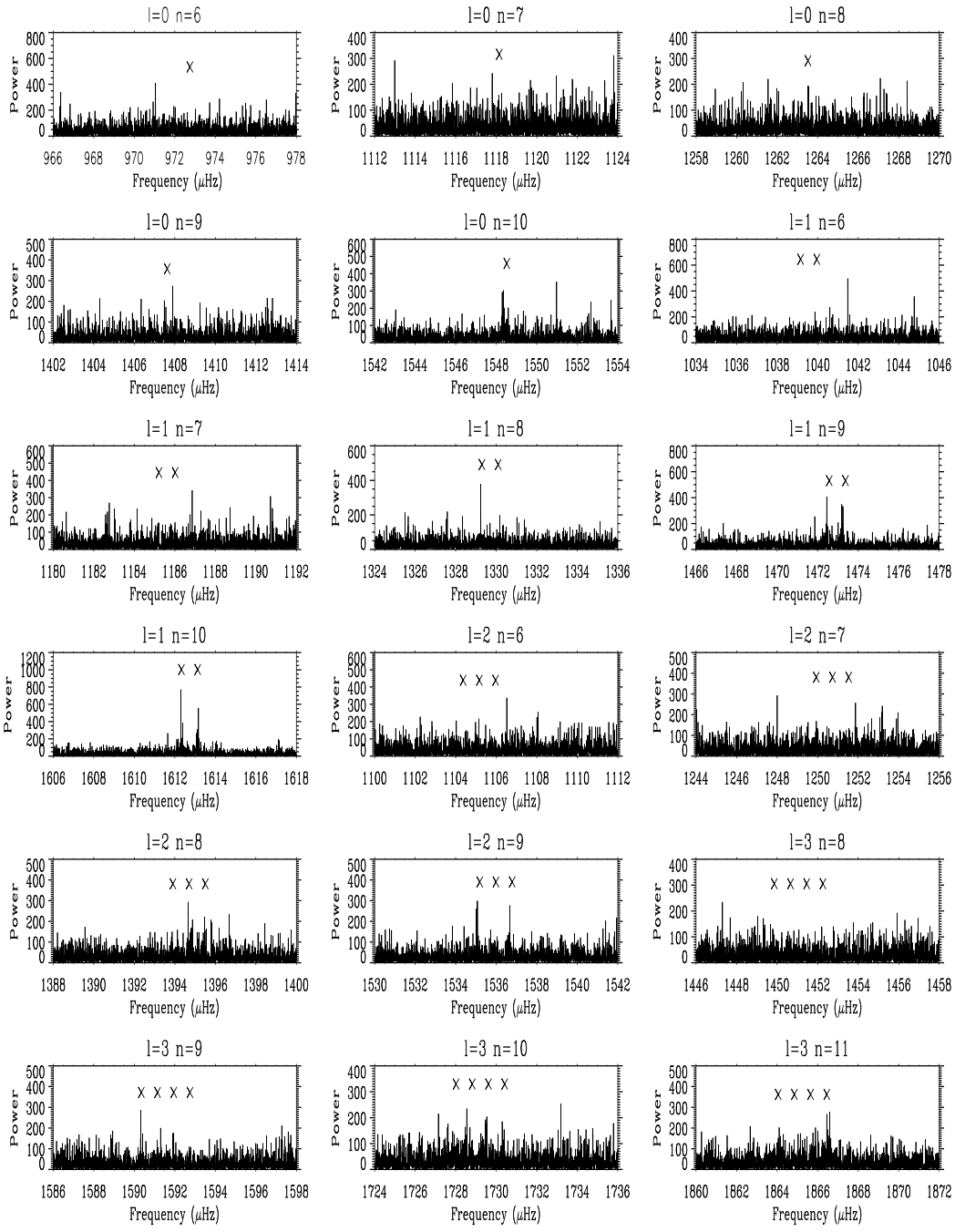

Figure B.6: Power spectrum produced from a timeseries with a rejection threshold of 200 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ with overlap selection based on an **average** of all available stations. Frequencies of predicted low-l modes are marked with an 'x'.

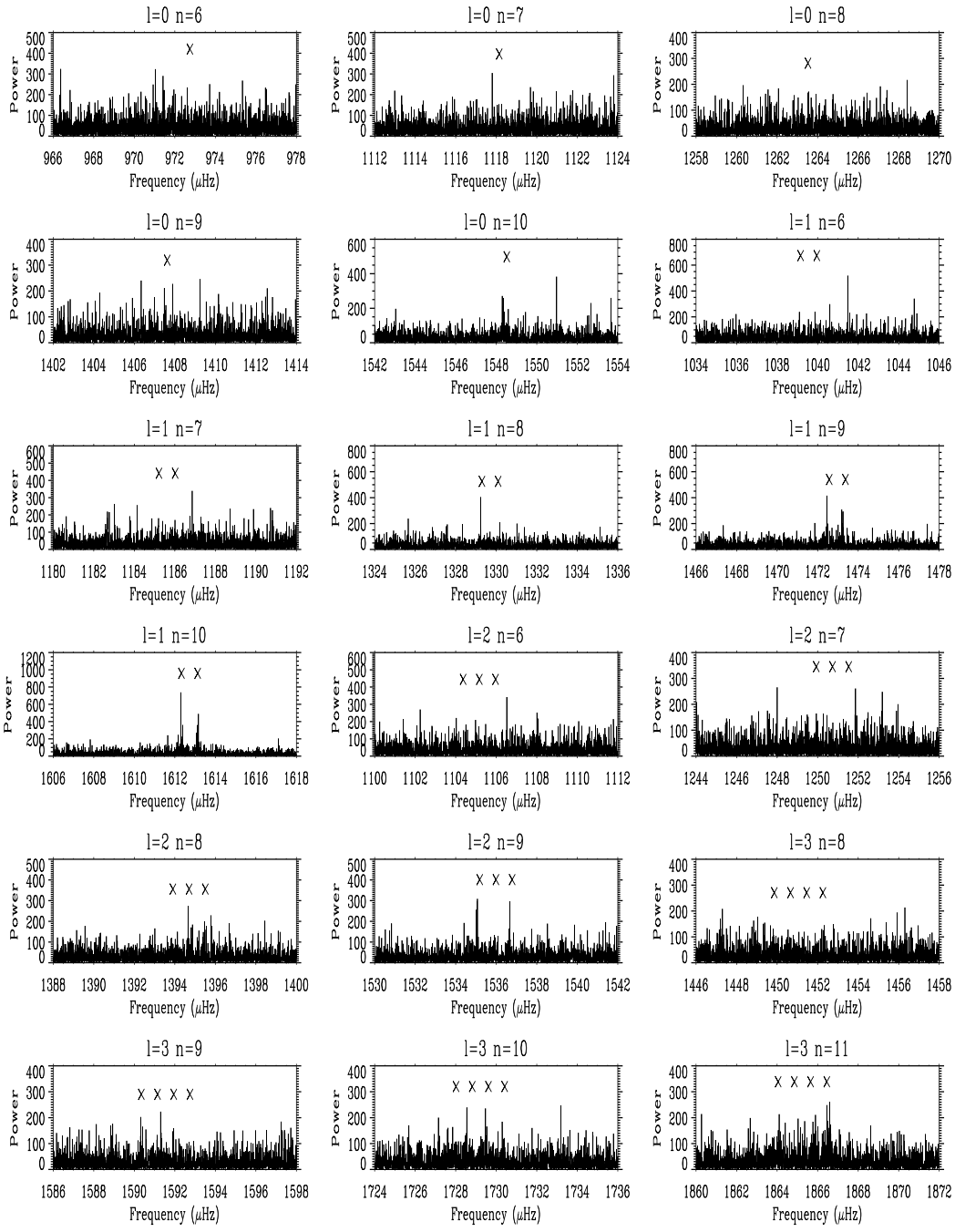

Figure B.7: Power spectrum produced from a timeseries with a rejection threshold of 200 $((\mathrm{ms}^{-1})^2)\mathrm{Hz}^{-1}$ with overlap selection based on a **weighted average** of all available stations. Frequencies of predicted low-l modes are marked with an 'x'.

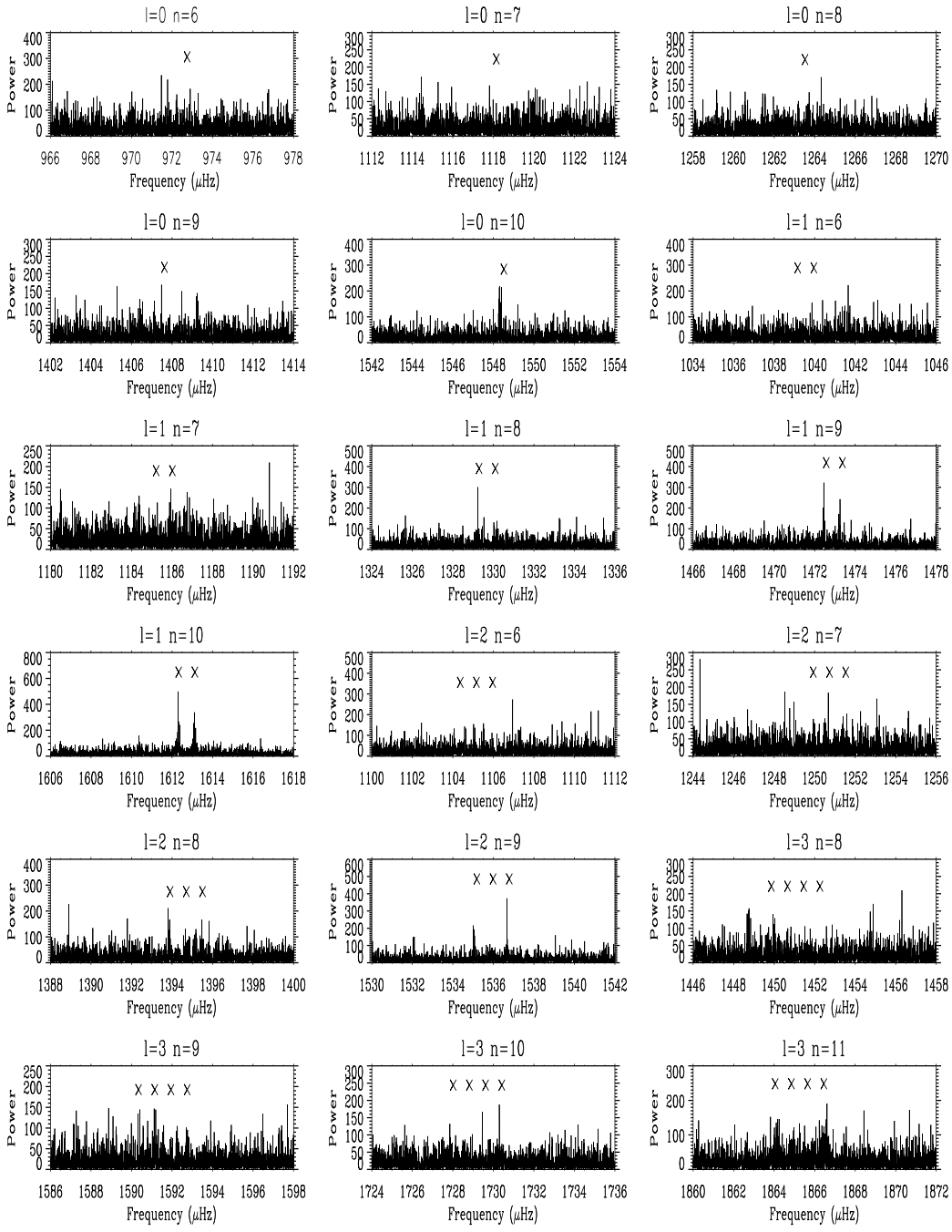

Figure B.8: Power spectrum produced from a time series with a **rejection threshold of 60** $((\mathbf{ms}^{-1})^2)\mathbf{Hz}^{-1}$ with overlap selection based on a **weighted average** of all available stations. Frequencies of predicted low-l modes are marked with an 'x'.