The Dissertation Committee for Aubra Elan Anthony
certifies that this is the approved version of the following dissertation:

A Search for Time Dependence in Astrophysical Neutrino Sources with the Sudbury Neutrino Observatory

Committee:

_____
Joshua Klein, Supervisor

_____
Duane Dicus

_____
Pawan Kumar

_____
Karol Lang

_____
Jack Ritchie

A Search for Time Dependence in Astrophysical Neutrino Sources with the Sudbury Neutrino Observatory

by

Aubra Elan Anthony, B.S., B.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2008

Acknowledgments

I am beyond grateful to have had Josh Klein as my graduate adviser. I have known few people in my life who have been more enthusiastic about physics or more willing to share their appreciation of it. He has truly taken his role of adviser to heart, and has helped me to make some very difficult decisions, from thesis topics to career choices to font choices. He ensured that my time in graduate school was really about learning, even when that required significant effort and patience on his part! Somehow he always made time to take my 'un-solvable' problems to the white board with me, and in so doing he really taught me what it means to do research, and how much fun research can be if you ask the right questions.

I also am incredibly fortunate to have had such a great group of fellow students to work with on SNO and in the dungeon that was ENS. First and foremost, Stan Seibert was by far one of the best officemates I could have hoped for, especially with his never-ending patience for my coding conundrums, his willingness to tackle all sorts of disk-hoggage and CPU tyranny that I would throw his way, and the fact that he never let on when my venting sessions threatened to 'destroy up' his day. By the same token, Chris Tunnell, Julia Majors and Melissa Jerkins all helped make ENS a brighter place, and I prob-

know you'd have preferred that I had gone into something other than physics, since you were so worried about being too strong of an influence. I hope you know that because of that, I now consider myself the true rebel of the family. Sorry to be so difficult, but thanks for helping to make me who I am.

And maybe most of all, thanks to Paul, for putting up with me, encouraging me, and taking care of me. I am so lucky to have found you. After three long years of sticking this out, it's looking like we can finally stop being so patient. Thanks for understanding how important this was to me.

<div align="right">

AUBRA ELAN ANTHONY

</div>

The University of Texas at Austin

December 2008

A Search for Time Dependence in Astrophysical Neutrino Sources with the Sudbury Neutrino Observatory

Aubra Elan Anthony, Ph.D.

The University of Texas at Austin, 2008

Supervisor: Joshua Klein

In this thesis I present the results of a periodicity study on the neutrino data collected over the span of the first two phases of SNO, at both low (1 day^{-1} - 0.1 yr^{-1}) and high (1 day^{-1} - 0.1 min^{-1}) frequency ranges. The high frequency study is the first of its kind, and is of particular interest in that it opens a window into the detection of solar g-mode oscillations, which have never been conclusively experimentally verified. In a data set with 7,646 neutrino candidates over a period of 698.29 live days, there was no detected high-frequency periodic signal.

In addition to a wide-range, single-peak high-frequency search, I have performed a directed-region frequency analysis, and a noise-motivated broad-band

analysis. All searches indicate an absence of periodicity in the ^{8}B solar neutrino signal as measured by SNO. I have also carried out an analysis of time dependence in the context of a trigger-less burst search, with the motivation of either observing neutrinos from an optically occluded supernova, or setting an upper limit on the senstitivity of our detector for such an observation. I include discussions of backgrounds to such a search that are specific to a heavy-water Cherenkov detector such as SNO.

Contents

List of Tables

List of Figures

xviii

xxiv

Chapter 1

Neutrino Physics, Solar Physics, and Solar Neutrino Physics

1.1 Overview

Neutrinos were first postulated to exist by W. Pauli in 1930[10] as a means of explaining the results of β-decay experiments: conservation of energy required the presence of this small, neutral particle to explain the continuous energy spectrum of the electron in nuclear β-decay. Over twenty years later, in 1956, neutrinos were experimentally observed by Reines and Cowan[11]. Reines (*et al.*) detected electron anti-neutrinos (thought to be neutrinos at the time) emitted from fission processes in a nearby nuclear reactor. At this point it was known that neutrinos were not only associated with electrons (*i.e.* electron neutrinos, ν_e) but also with the only other known lepton family member at the time, the muon (*i.e.* muon neutrinos, ν_μ). It was not until the 1970's that the

third charged lepton, the tau, τ, was detected[12], and with it, the existence of a third flavor of neutrino, the tau neutrino, ν_τ was inferred (and ultimately detected in 2000).

Neutrinos play a central role in the Standard Model. However, the SM was developed with the assumption that neutrinos are massless. In contradiction to the Standard Model framework, it has been demonstrated by independent experimental evidence from multiple sources (SNO, Super-Kamiokande, Kam-LAND, MINOS)[13][14][15][16][12] that neutrinos, distinguished by their lepton flavor, can oscillate from one flavor state to another. A neutrino is created in a weak eigenstate but propagates in the mass basis; it is the difference between mass eigenstates (Δm^2) which allows a neutrino that is produced in one flavor state to be detected in a different flavor state after propagation. If the neutrino were massless, there could be no difference in mass eigenstates (*i.e.* Δm^2 would be zero), and therefore oscillation would not be possible. Therefore, the myriad results confirming neutrino oscillations also indicate physics beyond the standard model.

1.2   Neutrinos in the Standard Model

As any system of structure must, the world of particle physics relies on a framework of 'laws', which has been dubbed The Standard Model. The Standard Model is a predictive description of the interactions of particles and fields. In Weinberg's 1967 paper[17] laying out what is considered to be the most correct incorporation of leptons into the Standard Model, an important assumption

was made: due to the lack of experimental observation of chirally right-handed neutrinos, one should develop the Standard Model with the assumption that they (right-handed neutrinos) do not exist at all. Working with this assumption, the Dirac mass term in the Standard Model Lagrangian must be zero, since the neutrino mass component of the general Lagrangian is of the form [1]

$$\mathcal{L}_\nu = m\left(\overline{\psi_R}\psi_L + \overline{\psi_L}\psi_R\right) \tag{1.1}$$

where ψ_L and ψ_R would be SU(2) doublets:

$$\psi_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \text{ and } \psi_R = \begin{pmatrix} \nu_R \\ e_R \end{pmatrix}$$

The above Lagrangian term therefore indicates that for the neutrino to have mass, there must exist a right-handed neutrino, ν_R, in addition to the already-detected left-handed neutrino. The right-handed neutrino has never been experimentally observed (and in its current form in the SM, ψ_R is actually a singlet of U(1): $\psi_R = (e_R)$). However, the weak interaction couples only to chirally left-handed particles, and as such the only method for physicists to observe chirally right-handed neutrinos would be through the interaction of the neutrino with the Higgs boson, which has been predicted but never observed, since neutrinos only interact via the weak force and gravity. It is

---

[1]Here I am dealing with only the Dirac mass term; a Majorana picture differs in its treatment of ν_L and ν_R; they are different 'versions' of the same particle. The mass term in the Lagrangian for the Majorana treatment is of the form $M(\overline{\psi_L}^c \psi_L + h.c.)$, but this relation also requires the existence of a right-handed chiral state (if neutrinos have nonzero mass) and as such, we are safe to use only the Dirac picture in this argument.

therefore important that we recognize that although the Standard Model did not predict that the neutrino have mass, it did not exclude the possibility.

1.3 The Standard Solar Model

Much like the standard model for particle physics, a standard solar model (SSM) describes the fundamental processes at work in our sun, and allows for predictions about physical conditions outside of our means of direct detection. By taking into account measurable parameters like luminosity, size, mass and spectral signature, the conditions throughout the sun can be modeled using energy transport equations and standard conservation laws. The current SSM begins with a zero-age sun given all 'known' inputs, and evolves the solar model with the constraint that the evolved model match present-day observed restrictions.

1.3.1 Our Sun: The Basics

The sun is a fairly standard main-sequence star, with a diameter of roughly 1.4×10^6 km and a total mass of 2×10^{30} kg. The sun is typically divided into different regions, or zones, classified primarily by the physical processes that are dominant in each region (radiative transfer, nuclear fusion, etc.) The main zones of the sun are, starting from the innermost region and moving outward: the core, out to $\sim 0.2 R_\odot$; the radiative zone, from $\sim 0.2 R_\odot - \sim 0.7 R_\odot$, where heat energy from the core is transferred via radiation; the convective

zone, from $\sim 0.7 R_\odot$ out to the surface, where conditions are not hot enough for radiation to occur, and therefore energy is transferred via convection; the photosphere, or surface; and the corona, or solar atmosphere.

The temperature at the surface (5,800 K) differs tremendously from the calculated temperature at the solar core (15,600,000 K) due to the different conditions in the different regions. The core, of course, is where nuclear fusion occurs, generating the energy and pressure which keeps the sun in hydrostatic equilibrium. The SSM predicts that the dominant fusion process which occurs in the sun, the *pp* chain, produces over 98% of solar energy[12]. The *pp* chain is based on the conversion of hydrogen into helium:

$$4p \rightarrow^4 \mathrm{He} + 2e^+ + 2\nu_e \tag{1.2}$$
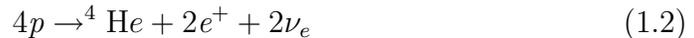
While the *pp* chain is the most prolific of the nuclear reactions in the present solar conditions, there are other higher-energy processes at work as well. The reactions and relative prevalences are shown in Figure 1.1. SNO is primarily sensitive to the higher-energy ^{8}B neutrinos, and also has some sensitivity to neutrinos produced in the *hep* chain. The different reactions' neutrino fluxes as a function of energy are shown in Figure 1.2, the result of the most recent version (2005) of the Bahcall-Serenelli-Basu SSM[2].
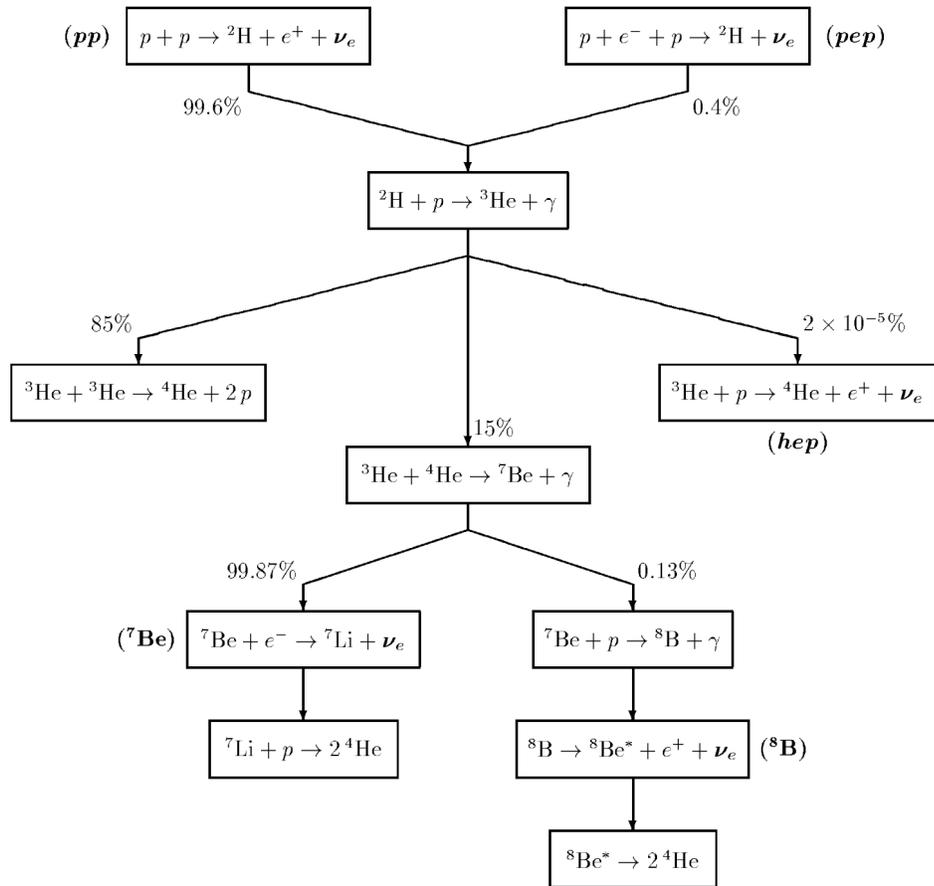
Figure 1.1: Solar nuclear fusion reactions and their relative branching ratios. SNO is sensitive to neutrinos produced in both the *hep* and ^{8}B chains. Taken from [1]
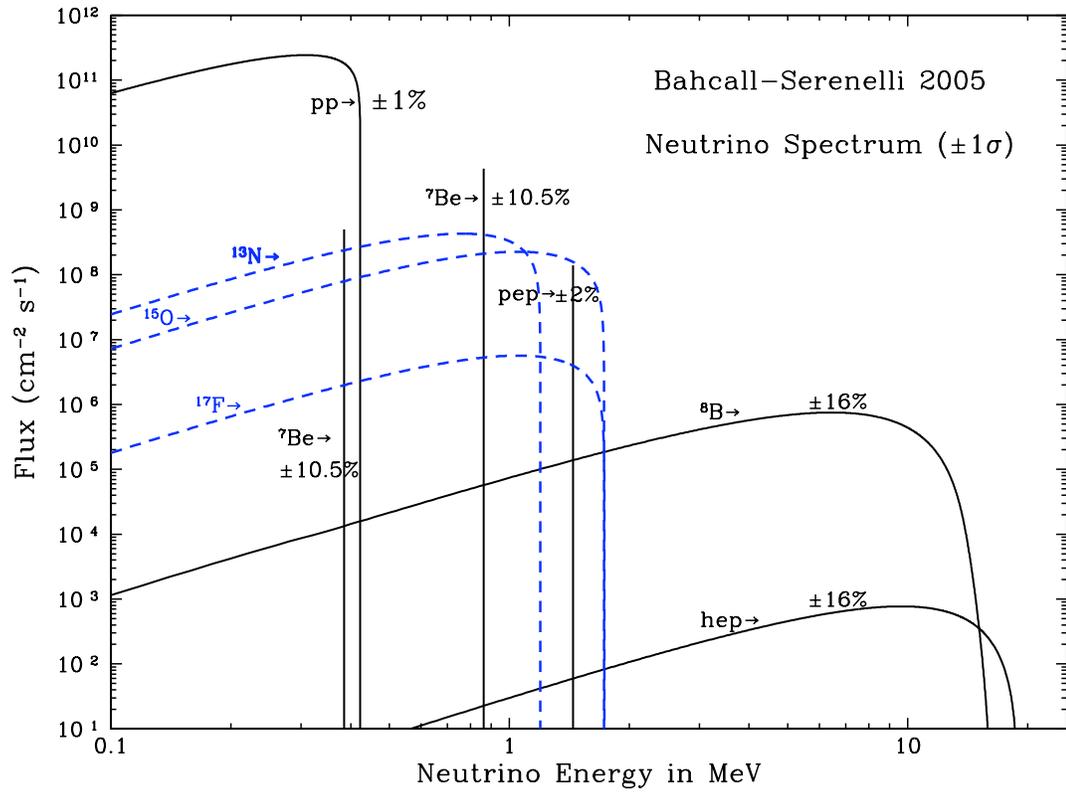
Figure 1.2: Predicted solar neutrino flux as a function of energy. SNO is sensitive to neutrinos produced in both the *hep* and ^{8}B chains. Figure taken from [2]

1.3.2 Inputs of a Standard Solar Model

There are a couple of key equations of stellar evolution and stellar structure which can be used to infer a great deal about a star. In the case of our sun, we have a larger pool of data to feed into our stellar evolution model, and as such we are left with a very complicated set of inter-dependent constraints and predictions. I will by no means show an exhaustive treatment of the generation of a standard solar model; this section should serve only to briefly inform the reader of the basic precepts which come into play when structuring the theory of a system as complex as our sun.

The fundamental required condition for a stellar model concerns hydro-static equilibrium; this is the governing principle for all stars, requiring that any pressure differential (directed radially outward) must be exactly balanced by an equivalent gravitational pressure (directed inward). This concept of dynamical balance is what keeps a star 'alive' in a sense.

$$\frac{dp}{dr} = -\frac{Gm\rho}{r^2} \tag{1.3}$$

where here r is the distance to the center of the sun, p is the pressure, ρ is the density, m is the mass contained inside a radius r, and G is the gravitational constant.

A second constraint is simply the relation that the mass within a given radius should be obtainable by considering the radial density distribution of

the sun:

$$m(r) = \int_0^r dr 4\pi r^2 \rho(r) \qquad (1.4)$$

or in differential form,

$$\frac{dm}{dr} = 4\pi \rho r^2 \qquad (1.5)$$

Having considered already the basic dynamics involved in building a working solar model, we should now look at a basic thermodynamic constraint which governs stellar structure; the relationship between radiative luminosity and energy generation,

$$\frac{dL}{dr} = 4\pi \rho r^2 \epsilon \qquad (1.6)$$

Here L is the luminosity at a radius r and ϵ is the rate of energy generation per unit mass. This is essentially a condition of conservation of energy, in a sense, among successive layers of the sun; while the surface layers of the sun radiate away energy in the forms of light and heat, they are fed an equal amount of energy from layers beneath them. This is all ensuring a stability of the sun; the sun would quickly become unstable if Equation 1.6 did not hold, and more energy were radiated out from the sun than was produced inside it.

The solar structure is also, of course, highly dependent on the chemical composition, and a standard, straightforward constraint on elemental abundances is simply that $X + Y + Z = 1$, where X is the abundance by mass of hydrogen, Y is the abundance of helium, and Z includes all elements heavier than helium. We can relate the pressure, from 1.3, to mass density with the

ideal gas law, giving an approximate equation of state:

$$Pg = Nk_BT \tag{1.7}$$

$$P \simeq \frac{\rho k_B T}{\mu m_u} \tag{1.8}$$

where k_B is Boltzmann's constant, μ is the mean molecular weight of the sun and m_u is the atomic mass unit.

These general equations are no more than a starting point for the development of a standard solar model. To generate a working SSM, they must be backed up with a great deal of additional details concerning the microphysics of the solar interior; equations of state, for example, relating nuclear fusion rates with pressure, temperature, and opacity. Different solar models use somewhat different formulations for equations of state, but the nuclear reaction parameters remain fairly consistent among the different models, and are generally based on the numbers given most recently in [2].

A lot of information about the solar structure is also gleaned from the efforts of the helioseismology community. Much like terrestrial seismic waves can give us information about the conditions and different regions of the inner earth, the inner makeup of the sun can be exposed through the study of acoustic solar waves. More will be said on this subject in Section 2.3, but it is important to note that the knowledge and study of pressure waves, or p-waves, at the solar surface have advanced our understanding of solar structure tremendously, and results from helioseismic experiments invariably have a central role in the solar

model calculations [18][19].

1.4   The Solar Neutrino Problem

Soon after Cowan and Reine's reactor experiment, the idea of detecting neutrinos from solar fusion reactions became a reasonable pursuit. As has been discussed in the previous section, and shown in Figure 1.2, the SSM predicts that a significant number of electron neutrinos are produced in solar core reactions. This prediction led to the experiment of Ray Davis, in 1968[20], which set out to determine the actual flux of electron neutrinos from the sun. Quite surprisingly, the results from Davis' experiment showed a marked deficit of electron neutrinos. In subsequent experiments aimed at measuring the solar neutrino flux, such as SAGE and GALLEX, the results were similar: there was a discrepancy between the numbers expected for the ν_e flux and the numbers being measured. This deficit came to be known as the Solar Neutrino Problem, and was ultimately a pivotal reason that the theory of neutrino oscillations took hold (though they had been postulated as early as 1957 by B. Pontecorvo[21]).

The Solar Neutrino Problem could have been caused by three primary explanations:

1. The Standard Solar Model incorrectly modeled the interior of the sun, and thus predicted more ν_es than were actually being produced,

2. The experiments detecting a deficit of solar νs were wrong, or

11

3. There was some mechanism by which ν_es were being transformed into ν_μs (or some other type of charge-less lepton).

Though none of these explanations was desirable, the first initially seemed the most plausible. However, after much re-evaluation of realistic models, the astrophysical community was only more certain that the SSM had accurately predicted the fusion rates in and composition of the sun. The second possibility was not really seen as reasonable as it would have required multiple independent experiments to be incorrect; a highly improbable solution to the problem. The most plausible remaining explanation was the mechanism of oscillation from one flavor state to another.

The solar sector was not alone in its neutrino deficit: the atmospheric sector had a similar problem, but concerning muon neutrinos. In the atmospheric sector, neutrinos are produced when energetic cosmic rays interact with molecules in the atmosphere to create a sizable amount of ν_μs (and $\overline{\nu_\mu}$s) in the processes:

$$p +^{14} N \rightarrow \pi^+ + X \tag{1.9}$$

giving the end decay products:

$$\pi^+ \rightarrow \mu^+ \nu_\mu \tag{1.10}$$

$$\mu^+ \rightarrow e^+ \nu_e \overline{\nu_\mu} \tag{1.11}$$

With this process, there is a predicted ratio of two to one (muon neutrinos to electron neutrinos). The actual observed ratio is more like 1.3:1, an obvious

discrepancy[12]. Furthermore, it was found that ν_μs which had traversed the earth (upward-going) on their way to being detected were disappearing at a larger rate than those traversing shorter distances (downward-going), which agreed with predictions from an oscillation model. Given the deficit of neutrinos in both the atmospheric and solar neutrino sectors, with both supporting a difference in expected flux to measured flux, it is clear that strong evidence points towards an oscillation theory.

In Figure 1.3, I show the results of a number of different experiments' measurements of the solar neutrino flux. The predicted values, indicated by the black markers, disagree with the experimentally measured values, indicated by the colored markers, except in the case of SNO's neutral current (NC) measurement. Because SNO's NC measurement sampled all three flavors of neutrinos equally, it is the only evidence which conclusively demonstrates the *appearance* of other flavors of neutrinos in addition to the *disappearance* of the ν_es. The combination of results from all experiments shown here demonstrates the measurement of a discrepancy of ν_es coming from the sun, and also demonstrates empirical evidence for the theory of oscillations in the appearance of ν_μs and ν_τs in the measured solar neutrino flux.

1.5   Neutrino Oscillation

Neutrino oscillations occur because neutrinos are created in flavor eigenstates, but propagate in the mass basis. This is similar to the quark sector mixing, in that the eigenstates of the free particle Hamiltonian differ from the flavor

13
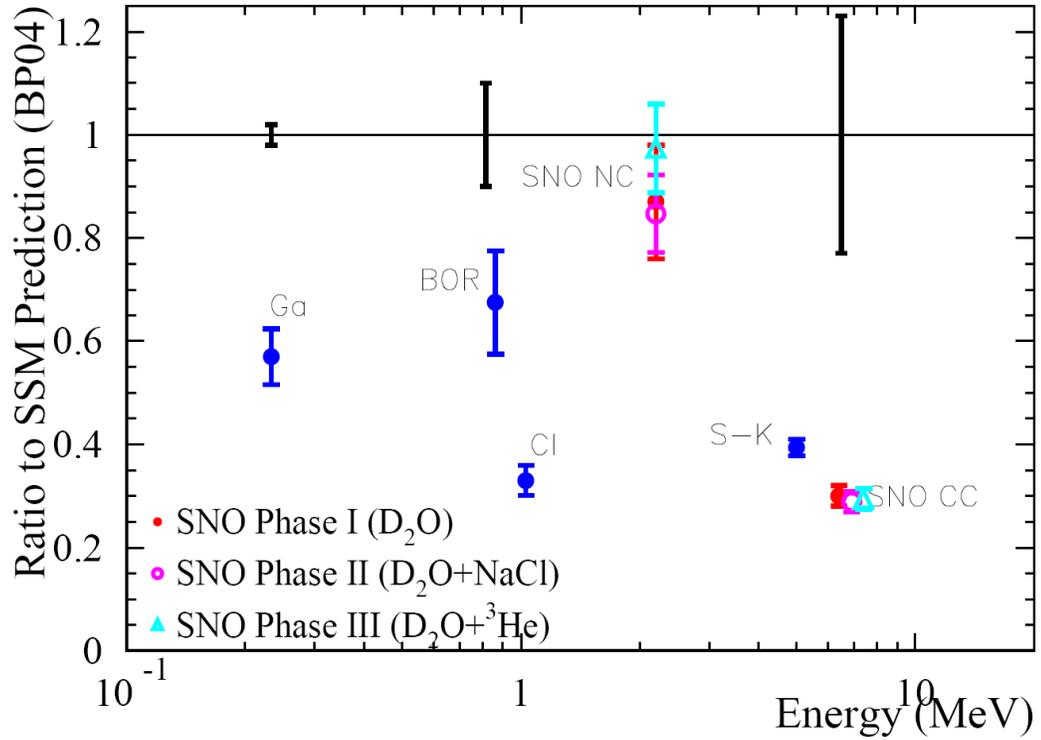
Figure 1.3: The current experimental contributions to the solution of the Solar Neutrino Problem. With SNO's neutral current results, the community was able to demonstrate the agreement between the Standard Solar Model's predicted neutrino flux values and the total detected neutrino flux of all flavors (as opposed to electron neutrinos only). Figure courtesy of J.R. Klein.

eigenstates; the free particle Hamiltonian and the weak interaction Hamiltonian do not commute. While the flavor states couple to the weak interaction, diagonalizing the weak interaction Hamiltonian,

$$
H_{weak} \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \lambda_\nu \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix},
\tag{1.12}
$$

another basis, the mass basis, is required to fully describe the system. The mass states diagonalize the free particle Hamiltonian ($H_{free} = p^2/2m$):

$$
H_{free} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = \alpha_\nu \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}
\tag{1.13}
$$

In the case of no oscillation, these two bases would be equivalent. However, in the case of oscillations there is no longer a direct one-to-one relationship between the two bases. The phenomenon of oscillation requires a more complicated descriptive picture, which is provided in the form of the Maki-Nakagawa-Sakata-Pontecorvo, **U**, the unitary mixing matrix which relates the flavor and mass states in the neutrino sector:

$$
\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\nu1} & U_{\nu2} & U_{\nu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}
\tag{1.14}
$$

The theory of oscillations points to the existence of non-zero off-diagonal elements of the MNSP mixing matrix, **U**, that connect these two groups of states.The three-flavor neutrino mixing matrix, **U** can also be written in a more 'leading' manner:

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \times \begin{pmatrix} c_{13} & 0 & s_{13}\exp i\delta \\ 0 & 1 & 0 \\ -s_{13}\exp -i\delta & 0 & c_{13} \end{pmatrix} \times \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(1.15)

where here, $c_{ij} = \cos\theta_{ij}$, $s_{ij} = \sin\theta_{ij}$, θ_{ij} is the mixing angle and δ is the CP violating phase. The division of **U**into three separate matrices is purposeful, in that we now have grouped **U** into an atmospheric (U_{23}), reactor (U_{13}) and solar (U_{12}) piece. It should be noted here that though there are three known lepton generations, neutrino mixing is commonly treated only in terms of the first or last two groups, *i.e.* mixing between (ν_e,ν_μ) or between (ν_μ,ν_τ). This can be done without loss of generality due to the fact that $\Delta m_{23}^2 \gg \Delta m_{12}^2$ and θ_{13} is small; the two-flavor treatment of oscillations is physically equivalent to a reduced parameter three-flavor approach.

1.5.1 The Mechanics of Oscillations

Reducing now to the use of a two-flavor mixing picture, the neutrino can be represented by two wave packets (different flavors) traveling at different speeds. These two propagating wave packets create an interference pattern with each other which will vary with the speed of propagation (and thus a different ratio

of flavors) at any given point. We can represent the flavor eigenstates by their relationship (through the unitary mixing matrix, $\mathbf{U}$) to the mass eigenstates:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \tag{1.16}$$

From the time-dependent Schrödinger equation, we can see the time evolution[21] of the mass states[2]

$$\begin{pmatrix} \nu_1\left(\vec{x}, t\right) \\ \nu_2\left(\vec{x}, t\right) \end{pmatrix} = \begin{pmatrix} e^{-iE_1 t}|\nu_1(0)\rangle \\ e^{-iE_2 t}|\nu_2(0)\rangle \end{pmatrix} \tag{1.17}$$

which can be written in matrix form,

$$\begin{pmatrix} \nu_1\left(\vec{x}, t\right) \\ \nu_2\left(\vec{x}, t\right) \end{pmatrix} = \begin{pmatrix} e^{-iE_1 t} & 0 \\ 0 & e^{-iE_2 t} \end{pmatrix} \begin{pmatrix} \nu_1(0) \\ \nu_2(0) \end{pmatrix} \tag{1.18}$$

Now, we have a method of expressing our flavor eigenstates as a function of the mass eigenstates, and a method of expressing the mass eigenstates as a function of time, so we can determine what the evolution of the flavor eigenstates would be, assuming they are propagating through a potential-free medium (free-particle):

$$\begin{pmatrix} \nu_e\left(\vec{x}, t\right) \\ \nu_\mu\left(\vec{x}, t\right) \end{pmatrix} = \mathbf{U} \begin{pmatrix} e^{-iE_1 t} & 0 \\ 0 & e^{-iE_2 t} \end{pmatrix} \mathbf{U}^\dagger \begin{pmatrix} \nu_e(0) \\ \nu_\mu(0) \end{pmatrix} \tag{1.19}$$

[2]given that *i.e.* $m_1 \neq m_2$

$$= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} e^{-iE_1 t} & 0 \\ 0 & e^{-iE_2 t} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \nu_e(0) \\ \nu_\mu(0) \end{pmatrix}$$

$$(1.20)$$

If we now take the mass of the neutrino, m_i to be much less than the momentum, p, we can expand the energy, E_i as:

$$E_i = \sqrt{p^2 + m_i^2} \approx p + m_i^2/2p \approx p + m_i^2/2E \qquad (1.21)$$

which then allows us to write Equation 1.19 as:

$$\begin{pmatrix} \nu_e(\vec{x}, t) \\ \nu_\mu(\vec{x}, t) \end{pmatrix} \approx e^{-ipt} \mathbf{U} \begin{pmatrix} e^{-im_1^2 t/2E} & 0 \\ 0 & e^{-im_2^2 t/2E} \end{pmatrix} \mathbf{U}^\dagger \begin{pmatrix} \nu_e(0) \\ \nu_\mu(0) \end{pmatrix} \qquad (1.22)$$

and expanding the exponential we see:

$$\begin{pmatrix} \nu_e(\vec{x}, t) \\ \nu_\mu(\vec{x}, t) \end{pmatrix} \approx e^{-ipt} \mathbf{U} \begin{pmatrix} 1 - im_1^2 t/2E & 0 \\ 0 & 1 - im_2^2 t/2E \end{pmatrix} \mathbf{U}^\dagger \begin{pmatrix} \nu_e(0) \\ \nu_\mu(0) \end{pmatrix}$$

$$(1.23)$$

given the unitarity of $\mathbf{U}$, we can write this as

$$|\nu_\alpha(\vec{x}, t)\rangle \approx e^{-ipt} \left(e_{\alpha\beta}^{-im^\dagger m t/2E} \right) |\nu_\beta\rangle \qquad (1.24)$$

18

This gives us a solution of the Schrödinger equation,

$$id/dt \mid \nu_\alpha\,(\vec{x},t)\rangle \approx (m^\dagger m/2E) \mid \nu_\alpha\rangle \tag{1.25}$$

with the omission of the phase factor, e^{-ipt}. We can further extend this setup to determine propagation probabilities for remaining in any initial flavor state.

We then find, for two-generation mixing (for either Dirac or Majorana mass pictures):

$$|\nu_\alpha\,(\vec{x},t)\rangle =$$

$$
\begin{pmatrix}
\cos(\frac{\Delta m^2}{4E})t - i\sin(\frac{\Delta m^2}{4E})t\cos 2\theta & -i\sin(\frac{\Delta m^2}{4E})t\sin 2\theta \\
-i\sin(\frac{\Delta m^2}{4E})t\sin 2\theta & \cos(\frac{\Delta m^2}{4E})t + i\sin(\frac{\Delta m^2}{4E})t\cos 2\theta
\end{pmatrix}
|\nu_\beta\rangle
$$
$$\tag{1.26}$$

With this description for the evolution of flavor eigenstates we can deduce various properties of the neutrino's state as a function of time as it travels through vacuum. If we know initially that a neutrino is of the electron variety, *i.e.* at time $t = 0$, we start with $|\nu_e\,(0)\rangle = (1,0)^T$, we find its transition rate to be

$$\langle \nu_e\,(x,t)\,|\nu_e\rangle = \cos(\Delta m^2/4E)t - i\sin(\Delta m^2/4E)t\cos 2\theta \tag{1.27}$$

and similarly, we see that the probability for the electron neutrino to still be in that same flavor eigenstate after a time t has passed is

$$P_{\nu_e \to \nu_e} = |\langle \nu_e(x,t)|\nu_e(x,0)\rangle|^2 = 1 - \sin^2 2\theta \sin^2(\Delta m^2/4E)t \qquad (1.28)$$

Conversely, the probability for an electron neutrino to propagate to a ν_μ state after some time t is

$$P_{\nu_e \to \nu_\mu} = |\langle \nu_\mu(x,t)|\nu_e(x,0)\rangle|^2 = \sin^2 2\theta \sin^2(\Delta m^2/4E)t \qquad (1.29)$$

This equation puts oscillation in the context of a mixing angle, θ, which will tell us the degree to which mixing occurs among different flavors . If we were to write the above equation, making the substitution[22] that t is on the order of L, the distance between the source of the neutrinos and the detector (since ν_is propagate very near the speed of light, c), we see that the probability of transition is then dependent upon $sin^2(\Delta m^2/4E)L$. Here we see the importance of the parameter of the distance between a neutrino source, and a detector of neutrinos, L, where here L is the vacuum oscillation length, in km, or $L = 4\pi E/\Delta m^2$, as well as the energy of the neutrino, E, in MeV, and the squared mass difference between the two mass eigenvalues, or $\Delta m^2 = (m_2{}^2 - m_1{}^2)eV^2$.

1.6 The MSW Effect

The discussion up until now of neutrino oscillation has been focused solely on vacuum oscillations, but an equally important treatment of oscillations comes from considering the behavior of neutrinos as they pass through electron-dense regions. The survival probability of electron neutrinos is dramatically affected by the presence of matter; in the context of the sun, this matter effect is dubbed the MSW effect, for its three creators, Mikheev and Smirnov[23], and Wolfenstein [24].

When traveling in matter, electron neutrinos experience the potential for enhanced oscillations due to the fact that in the presence of electrons, electron neutrinos can interact via exchange of both W bosons and Z bosons – that is, electron neutrinos can undergo charged current interactions as well as neutral current interactions – while muon and tau neutrinos are restricted to neutral current interactions only. The inclusion of the possible charged current interaction for the electron neutrino adds an effective potential term, V_{eff}, to the system's Hamiltonian, where $V_{eff} = \sqrt{2}G_F N_e$, and G_F is the Fermi constant and N_e the electron number density. Our Hamiltonian can now be written as

$$H_{matter} = \Delta m^2/4E \begin{pmatrix} -\cos 2\theta + V_{eff} & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix} \tag{1.30}$$

This, then, is the appropriate manner of time-evolving the neutrino wave

function;

$$i\frac{d}{dt}\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = H_{matter}\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} \tag{1.31}$$

The working MSW Hamiltonian is then diagonalized as

$$\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = \tilde{\mathbf{U}}\begin{pmatrix} \tilde{\nu}_1 \\ \tilde{\nu}_2 \end{pmatrix} \tag{1.32}$$

with

$$\tilde{\mathbf{U}} = \begin{pmatrix} \cos\tilde{\theta} & \sin\tilde{\theta} \\ -\sin\tilde{\theta} & \cos\tilde{\theta} \end{pmatrix} \tag{1.33}$$

and $\tilde{\nu}$ is the energy eigenstate in matter. The new matter mixing angle, $\tilde{\theta}$ is given by

$$\cos 2\tilde{\theta} = \frac{-A/\Delta m^2 + \cos 2\theta}{\sqrt{(A/\Delta m^2 - \cos 2\theta)^2 + \sin^2 2\theta}} \tag{1.34}$$

$$\sin 2\tilde{\theta} = \frac{\sin 2\theta}{\sqrt{(A/\Delta m^2 - \cos 2\theta)^2 + \sin^2 2\theta}} \tag{1.35}$$

where

$$A = 2\sqrt{2}EG_F N_e \tag{1.36}$$

In conditions of high electron density, N_e, such as those that exist at the center of the sun, MSW effects are expected to greatly impact neutrino oscillations. MSW 'resonance' occurs at $A/\Delta m^2 = \cos 2\theta$, indicating a critical

22

electron density:

$$N_e = N_{e,crit} \equiv \frac{1}{2\sqrt{2}G_F} \frac{\Delta m^2}{E} \cos 2\theta \qquad (1.37)$$

For electron neutrinos propagating in a medium which satisfies the MSW condition of resonance, oscillations are thus greatly enhanced.

Chapter 2

Solar Periodicity

The sun is known to have inherent periodic behavior; in addition to the simplistic 'periodicity' of the sun's differential rotation, in which the equator completes one rotation cycle over a period of roughly 25 days, and the poles over roughly 35 days, there is also the sunspot cycle, or solar magnetic activity cycle, with an 11-year period (the period actually varies between 9.5 and 12.5 years). This 11-year sunspot cycle is related to the 22-year period cycle in which the sun's magnetic field reverses.

Though periodicities are known to exist in the sun, there is currently no indication that any of the known periodicities could affect the solar neutrino flux. As the nuclear fusion rates at the solar core are expected to be stable, the existence of periodicity in the solar neutrino flux would likely point to new physics.

2.1 Searching for Periodicity in the Solar Neutrino Flux

One possible theory which has been put forward as motivation for searching for a periodic signal in the solar neutrino sector is the phenomenon of Spin Flavor Precession (SFP). SFP would require that the neutrino have a large magnetic moment[1], which would interact with solar magnetic fields, causing the neutrino to flip into a right-handed helicity state. If the neutrino is a Dirac particle, this would result in a sterile neutrino; if the neutrino is a Majorana particle, this coupling to the solar magnetic fields could cause a simultaneous flip in chirality and flavor, thus changing a ν_e to a $\overline{\nu_\mu}$ or $\overline{\nu_\tau}$. This hypothetical Spin Flavor Precession could be resonantly enhanced by matter (Resonant Spin Flavor Precession, RSFP), which would result in an energy-dependent supression of the neutrino flux. It would then be possible that the RSFP mechanism (coupled to the rotating solar magnetic fields) would be detectable as a periodicity in the solar neutrino flux.

In line with this motivation, claims of detection of periodicity in the solar neutrino flux have been made [25], [26], [27], [28], [29], [30], [31], largely in contradiction with the results of analyses done by the experimental collaborators themselves [32] [33] [8]. The main proponent of the existence of this periodicity, P. Sturrock, presented multiple analyses on the topic including Lomb-Scargle periodogram and maximum likelihood methods. Among several

---

[1] "large" = several orders of magnitude more than the SM prediction

claims of periodicity in the solar neutrino flux made by Sturrock, the most persistent claim was of periodicity in the Super-K 5-day binned data set, at a frequency of 9.43 yr^{-1}, and with a strength of 7% amplitude. Sturrock also claimed the existence of a signal at 13.6 yr^{-1} in the Gallex data.

Most recently, when faced with the results of Pandola [33], on behalf of the GNO collaboration, which cited no evidence for periodicity in the combined Gallex-GNO radiochemical data set, Sturrock claimed that the lack of a signal was due to having lengthened the data set (thereby washing out evidence paired with solar magnetic cycles), and searching in too broad of a frequency band (*i.e.* taking into consideration too many frequencies) [34]. Sturrock then claimed that in order to gain access to the periodic signal, one must only look at the data for the period of time when Gallex was running, since this was when evidence for rotational modulation of the solar magnetic fields was concentrated. He also claims that the frequency band for such a search must be less inclusive, in order to avoid the trials penalty (described in Chapter 5). This is a fundamentally flawed argument, in that it amounts to 'hand-picking' a data set, such that it produces the desired results. Unfortunately, this was also his response when faced with the absence of a signal in SNO's solar neutrino data set [8]; in [29], Sturrock discards an entire phase of the SNO neutrino data set in order to match data with his findings (and incidentally is still only able to find 'weak' modulation in the restricted SNO data set).

KamLAND's recent anti-neutrino experimental results [35] [36] probe a similar region in L/E to SNO, and result in a precise Δm_{12}^2 measurement.

This, paired with the almost-orthogonal sampling of mixing parameter space by SNO (resulting in a precise θ_{12} measurement), allow for a highly-constrained formulation of the best-fit mixing parameters, giving, with very high confidence, evidence for the Large Mixing Angle(LMA)-MSW oscillation scenario in the solar sector, and ruling out RSFP as a contributing mechanism. Although Sturrock does concede [28] that the KAMLand results rule out RSFP as a viable means of supression of the solar neutrino flux in the context of the solution to the solar neutrino problem, he persists in claiming that RSFP may still be the cause for the periodicities which he finds, just in a different region of the sun than had been initially predicted, in a 'subdominant' RSFP process.

Though Sturrock's methods of analysis may in themselves be sound, in that he appears to appropriately employ both the Lomb-Scargle and likelihood analyses, his practice of searching for signals (and claiming to find them), with a less-than-rigorous statistical interpretation, has led many in the community to be doubtful of his claims. This has served as ample motivation for SNO to carry out independent analyses of periodicity in our data set, as will be detailed in the remainder of this thesis.

2.2 A Low Frequency Search: Published SNO Results

In light of the motivations laid out in Section 2.1, SNO has performed a low-frequency periodicity search using three independent analyses: the Lomb-Scargle periodogram and a maximum likelihood method, as well as the slightly less conventional Rayleigh power test. Here I briefly explain the first two methods and give our published results from these methods. In Chapter 6 I will fully describe the low-frequency Rayleigh power analysis, and the results from that analysis on the SNO D_2O-phase data.

2.2.1 Lomb-Scargle Periodogram

The Lomb-Scargle periodogram provides a means for searching for periodicity in an unevenly sampled data set. The Lomb-Scargle method generates a power $P(f)$ for different frequencies, f, from the measured flux values $y(t_i)$ in N independent time bins, as follows:

$$P(f) =$$

$$\frac{1}{2\sigma^2} \left(\frac{\left[\sum_{i=1}^{N} w_i (y(t_i) - \overline{y}) \cos(2\pi f(t_i - \tau)) \right]^2}{\sum_{i=1}^{N} w_i \cos^2(2\pi f(t_i - \tau))} + \frac{\left[\sum_{i=1}^{N} w_i (y(t_i) - \overline{y}) \sin(2\pi f(t_i - \tau)) \right]^2}{\sum_{i=1}^{N} w_i \sin^2(2\pi f(t_i - \tau))} \right)$$

$$(2.1)$$

with the phase factor, τ, satisfying $\tan(4\pi f \tau) = \dfrac{\sum\limits_{i=1}^{N} w_i \sin(4\pi f t_i}{\sum\limits_{i=1}^{N} w_i \cos(4\pi f t_i)}$. Here t_i is

the livetime-weighted mean time for the ith bin, and $\overline{y}$ and σ^2 are the weighted mean and variance for all the bins. Each bin is weighted according to that bin's squared uncertainty ($w_i = 1/\sigma_i{}^2$).

For this method, the SNO data is binned in 1-day increments, which reduces the sensitivity of the Lomb-Scargle method for any frequencies higher than 0.5 days^{-1}. If any bin was expected to have fewer than five events, that bin would be combined with the next bin, to ensure that the expected number of events in all bins was larger than 5 (to avoid biases coming from the fact that we have assumed Gaussian statistics in our bins). To verify the robustness of the Lomb-Scargle method in SNO, a variety of bin sizes (from 1-day to 5-day) were substituted into the analysis, as well as a range of starting times of the first bin and different values of the separation between sampled frequencies. The results of the Lomb-Scargle periodogram were not statistically different for any of the tested scenarios.

2.2.2 Unbinned Maximum Likelihood Method

The second main analysis in SNO's published low-frequency periodicity search was an unbinned maximum likelihood method. The likelihood method tests the hypothesis that the events in the SNO data set are drawn from a periodic rate distribution, $\phi(t)$, relative to the hypothesis that the events are drawn

from a constant distribution ($A = 0$), and a 'likelihood' spectrum is generated from this test. Here, $\phi(t)$ is defined by

$$\phi(t) = N(1 + A\cos{(2\pi ft + \delta)}) \tag{2.2}$$

where A is the amplitude of the hypothetical periodicity, N is a normalization constant, and δ is a phase offset. The PDF $\phi(t)$ only exists at times when SNO was operating (*i.e.* within run boundaries); otherwise it is set to be zero. For any given frequency, f, the maximum of the extended likelihood $L(N, A, \delta|f)$ is found via

$$\ln L(N, A, \delta|f) = -\sum_{j=1}^{runs} \int_{t_i{}^j}^{t_f{}^j} \phi(t)dt + \sum_{k=1}^{events} \ln(\phi(t_k)) \tag{2.3}$$

In the first term, summed over all runs, the integral of $\phi(t)$ is evaluated between each run's start and stop times, t_i and t_f; this term accounts for Poisson fluctuations in the signal amplitude. In the second term, the sum is over all events in the data set. This method accounts for deadtime in the detector, given that $\phi(t)$ is only non-zero for times that SNO was running. The parameters A and N are allowed to vary as free parameters (at very low frequencies, A and N become degenerate). This method does not require binning of the data set, allowing greater sensitivity than the Lomb-Scargle method at higher frequencies, due to the 1-day binning used in the Lomb-Scargle approach.

2.2.3 Results

SNO's initial low-frequency periodicity study was performed over the range of periods of 1 day to 10 years, a span which encompassed the periods of all previous claims of 'detected' signals. The results from this study showed no detectable periodicity in the solar neutrino flux on time scales between periods of 10 years and of 1 day (roughly 0.0002/day to 1/day), or at the specific period at which the most persistent experimental detection had been claimed ($9.43yr^{-1}$).

The most significant peak height for the combined D_2O and Salt data sets from the Lomb-Scargle periodogram has a power of 8.7, and is at a period of 2.42 days (f =0.413 days^{-1}). The most significant peak height for the combined data sets from the maximum likelihood approach has a power of 8.8 at a period of 2.40 days (f=0.417 days^{-1}). In Monte Carlo null-hypothesis trials, powers were seen at this level or higher 27% of the time for the Lomb-Scargle periodograms, and 35% of the time for the maximum likelihood power spectra. SNO therefore claims to have detected no significant periodicity in the solar neutrino flux at low frequencies.

2.3 Helioseismology and High-Frequency Oscillations

The sun, although thermally stable, has various modes of oscillation present in its internal structure. These oscillations are present in many stars to vary-

ing degrees, but in the sun, there is a rich spectrum of modes ranging from pressure-mode (p-mode) oscillations, with pressure as the restoring force, to gravity-mode (g-mode) oscillations, where buoyancy serves as the restoring force.

The understanding necessary for the advancement of helioseismology emerged in the 1970s[37][38], when granules on the surface of the sun were resolved to be oscillating vertically, with a consistent speed of a few hundred meters per second with a 5-minute period. These oscillations arise as a result of standing sound waves in the deep convection zone of the sun: certain frequencies are trapped in this region, and their complicated interference pattern is observed via Doppler velocity measurements of the solar surface. These 5-minute period waves are p-modes, thousands of which exist in the sun, at frequencies ranging from hundreds to thousands of μHz.

I have mentioned that certain frequencies are trapped in the convection zone. In any 'edge' of a layer of the sun, where a steep density gradient is present, the boundary between layers will act as a reflective mirror, turning sound waves back in, to propagate until they hit the next steep density or temperature gradient. This refraction of waves depends on the speed of sound in any given space; the speed of sound in a region will therefore tell us a great deal about that region due to its dependence on surrounding physical parameters such as temperature (T), density (ρ), pressure (P), and composition of the medium of propagation:

$$s = \sqrt{\frac{\partial p}{\partial \rho}} = \sqrt{\frac{\gamma k_B T}{\mu m_u}} \sim \sqrt{\frac{T}{\mu}} \qquad (2.4)$$

where γ is the adiabatic index (the ratio of specific heats for constant pressure vs constant volume) of the medium, and μ the mean molecular weight of the sun, with m_u representing the atomic mass unit.

It is largely through studying the behavior of propagation of different frequencies of p-mode waves traveling through different regions of the sun that the helioseismology community has been able to significantly enhance the current understanding of the sun's composition and physical properties.

For the purposes of this thesis, however, g-mode oscillations merit a slightly more in-depth inspection than p-mode oscillations. Gravity mode oscillations are strictly nonradial oscillations, which differ from the standard radial pulsations that come to mind when discussing stellar variability. Taking the model that the sun is simply 'ringing' at different frequencies, these oscillations can be described by the spherical harmonics $Y_l^m(\theta, \phi)e^{-i\sigma t}$. Here, l indicates the total number of nodes along a line of longitude on the solar disk. In the case of non-radial oscillations, the spherical harmonic degree l is required to be nonzero, $l \neq 0$, whereas the radial modes require that $l = 0$ [39]. The other quantum number referenced here, m, describes the number of nodal lines that cross the solar equator. The parameter m is of course related to l, with the allowed range of m spanning from $-l$ to $+l$. Examples of visualizations of some of these modes are shown in Figure 2.1.

For g-mode oscillations to exist two critical conditions must be met: the
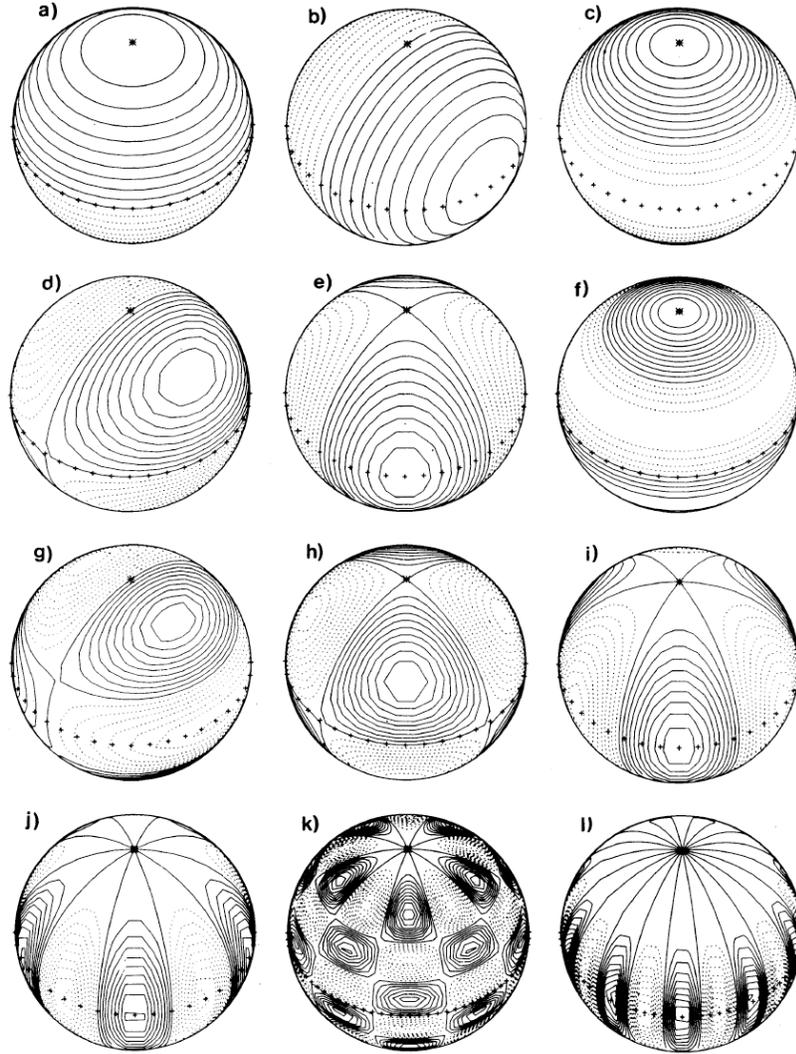
Figure 2.1: Contour plots of the real part of spherical harmonics Y_l^m. Positive contours are represented by continuous lines and negative by dashed lines. The equator is shown by "++++". The cases illustrated are as follows: a)$l = 1, m = 0$, b)$l = 1, m = 1$, c)$l = 2, m = 0$, d)$l = 2, m = 1$, e)$l = 2, m = 2$, f)$l = 3, m = 0$, g)$l = 3, m = 1$, h)$l = 3, m = 2$, i)$l = 3, m = 3$, j)$l = 5, m = 5$, k)$l = 10, m = 5$, l)$l = 10, m = 10$. Taken from [3].

frequency of oscillation must be lower than the minimum of the critical acoustic frequency, S_l, and N^2, the Brunt-Väisälä frequency. The critical acoustic frequency is also known as the Lamb frequency, and is defined as:

$$S_l{}^2 = \frac{l(l+1)\tilde{c}^2}{r^2} \tag{2.5}$$

(here, $\tilde{c}$ is the speed of sound). Physically, the inverse of S_l represents the time it would take for a sound wave to traverse one horizontal wavelength $(2\pi r/[l(l+1)]^{1/2})$ along the circumference of a circle of radius r (concentric with the solar center).

The second condition requires that the Brunt-Väisälä frequency must be greater than zero, or $N^2 > 0$, where N^2 is given by:

$$N^2 = g\left(\frac{1}{\Gamma_1}\frac{d\ln p_0}{dr} - \frac{d\ln\rho_0}{dr}\right) \tag{2.6}$$

where Γ_1 is the adiabatic exponent, $\Gamma_1 = (d\ln p/d\ln\rho)_{ad}$, and p_0 and ρ_0 are the pressure and density of the unperturbed (non-oscillating state).

The Brunt-Väisälä frequency can be best understood by considering the case of a stratified fluid, with P, ρ, and Γ_1 depending only on radial position, r. If we imagine a small element of fluid at a radial position r_0, which is displaced to $r_0 + \delta r$ (as in Figure 2.2), we can take the buoyant force, per unit volume, acting on this small fluid element to be

$$f_B = -g\Delta\rho = \rho g A\delta r \tag{2.7}$$

35
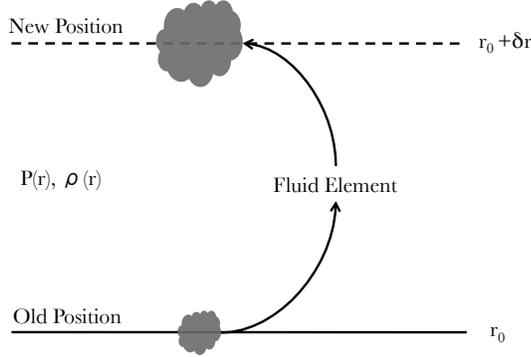
Figure 2.2: Example of a fluid element in a stratified fluid being displaced by an amount δr.

Here A is a parameter describing the radial component of the density disturbance due to the element being moved;

$$\Delta\rho = -\rho \left(\frac{1}{\rho}\frac{d\rho}{dr} - \frac{1}{\Gamma_1 P}\frac{dP}{dr} \right) \delta r \tag{2.8}$$

$$= -\rho A \delta r \tag{2.9}$$

We can see in Equation 2.7 that if f_B and δr have opposite signs, the fluid element will be returned to its original position, r_0, resulting in a *convectively stable* condition. If, however, the buoyant force and direction of displacement have the same sign, we are left with an instability, as the fluid will have no restoring force. In this vein, we can understand local convective stability in terms of A; for $A < 0$, we are in a range of convective stability and for $A > 0$, we are in a range of instability. The Brunt-Väisälä frequency is related to this parameter by $N^2 = -Ag$, so it is clear that the conditions of convective
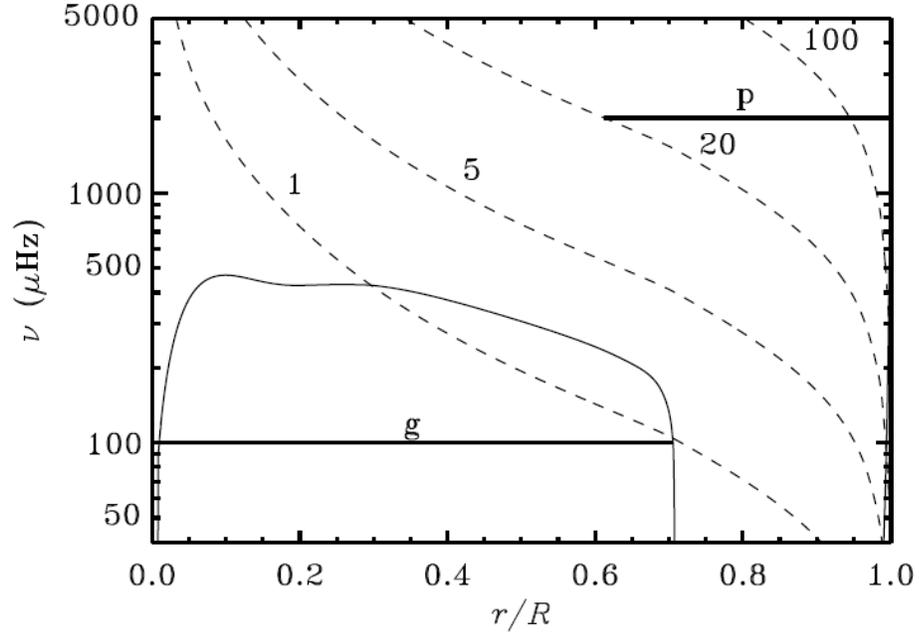
Figure 2.3: Buoyancy frequency, N (solid line), and critical acoustic frequency, S_l (dashed lines, labeled by the corresponding values of l), shown in terms of their corresponding oscillatory frequencies, against fractional solar radius $(r/R_\odot)$. The heavy horizontal lines indicate the trapping regions for a g-mode with frequency $\nu = 100\mu$Hz and for a p-mode with frequency $\nu = 2000\mu$Hz. Taken from [3].

stability and instability will be similarly represented by the sign of N^2.

So we have the condition that the frequency of oscillation, ν, for the g-mode range is limited, $\nu < \min(S_l, N)$. Due to the physical tendencies of both S_l and N, the g-mode frequencies are thus restricted to the interior of the sun (within approximately $0.2R_\odot$ of the core). A plot of the behavior of S_l and N is shown in Figure 2.3, in terms of the corresponding cyclic frequencies.

These restrictions put us in a frequency range with an upper limit of about 500-600 μHz[3], corresponding to a period length of roughly half an hour. This is well within the scope of the Rayleigh Power test for high frequency signals, as will be described in Chapter 7.

Chapter 3

The Sudbury Neutrino

Observatory

3.1   The Detector

The Sudbury Neutrino Observatory was located inside INCO's Creighton mine, an active nickel mine just outside of Sudbury, Ontario. The detector (shown in Figure 3.1) sat at a depth of 6,800 ft. below surface (6,020 meters water equivalent), which allowed a significant reduction in the number of background cosmic ray events which reach SNO compared to surface-level contamination. The cavity housing the detector was 34m tall and 22m wide, and was surrounded by magnetic 'compensation coils', used to offset the vertical component of the earth's magnetic field and thus increase the collection efficiency of the photomultiplier tubes (PMTs). The magnetic field value of the compensation coils sat at 19 μT, and the measured increase in photon detection

efficiency with the coils enabled was roughly 10%. The PMTs used in SNO were R1408 Hamamatsu tubes, each with a 20cm diameter, and operated at roughly 2,000 V. Each PMT was also surrounded by a 27cm-diameter reflective concentrator, designed to increase the effective photocathode coverage in the detector; including reflectors, the photocathode coverage in SNO was roughly 55%.

The target volume of SNO was 1,000 metric tons of heavy water, or D_2O, encased in a 12m-diameter spherical acrylic vessel (AV) with 5.5 cm-thick walls. The AV was suspended by nylon ropes, and was roughly concentric within a spherical support called the PSUP, the PMT support structure. The volume between the AV and the PSUP was filled with 1700 metric tons of ultra-pure H_2O, to help shield the D_2O target from contamination by external radioactivity from the PMTs and the PSUP. The PSUP was 18m in diameter, and housed the 9,438 inward-looking PMTs, facing the D_2O target as well as 91 outward-looking PMTs, or OWLs, facing the outer water volume. The OWLs were designed to act as a cosmic ray muon detector, tagging events which created light in the H_2O volume. Surrounding the PSUP was 5,300 metric tons of H_2O to provide further shielding from radioactivity in the cavity walls. More specific details concerning the SNO detector can be found in [4],[40].

3.2 Neutrino Interactions in SNO

The SNO detector is unique from other water Cherenkov detectors due to the use of heavy water (D_2O) as its target volume. Because of the D_2O target
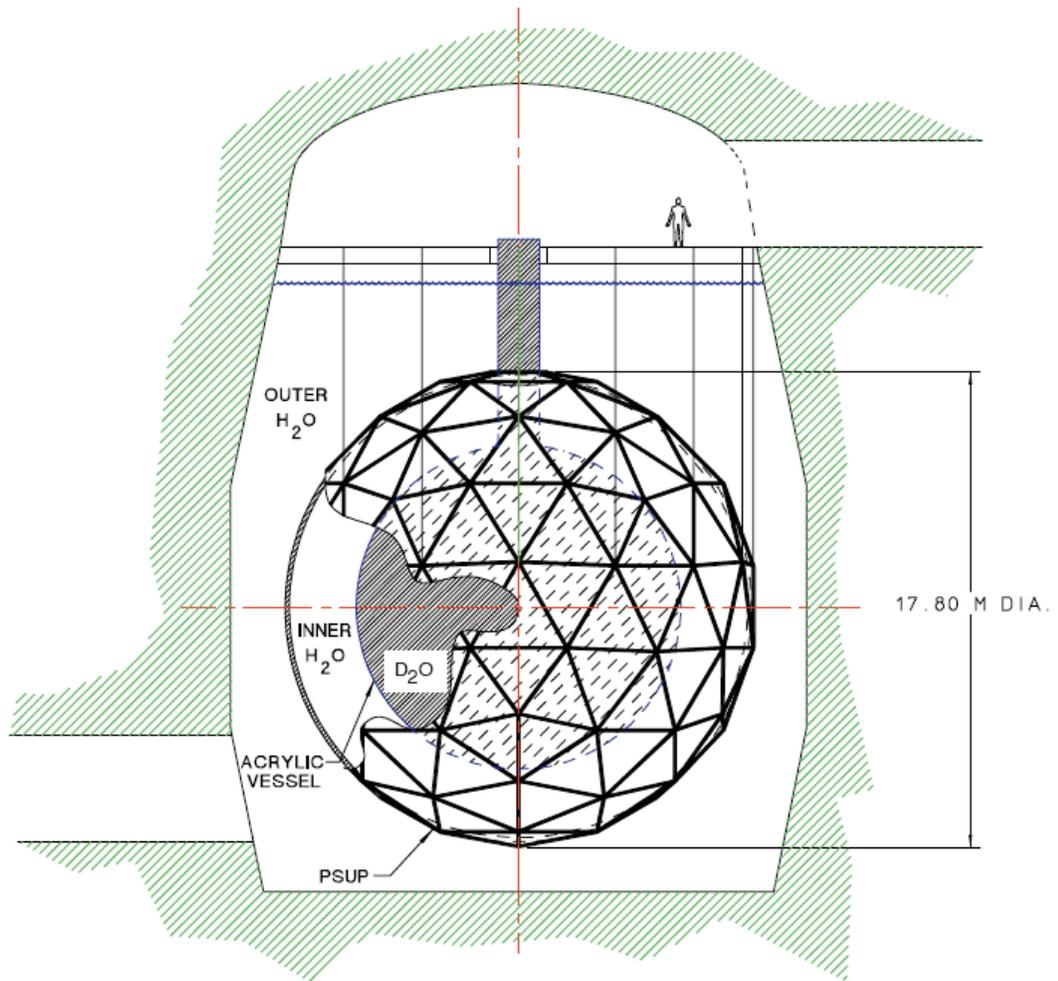
Figure 3.1: Cutaway diagram of the SNO detector. Figure originally published in [4].

volume, SNO can detect neutrinos via not only the standard elastic scattering (ES) interaction,

$$\nu_{e,(\mu,\tau)} + e^- \rightarrow \nu_{e,(\mu,\tau)} + e^- \tag{3.1}$$

but also the neutral current (NC), in which a neutral Z boson is exchanged between the neutrino and deuteron,

$$\nu_x + d \rightarrow n + p + \nu_x \tag{3.2}$$

and Charged current (CC) interactions, in which a charged W boson is exchanged between the neutrino and deuteron,

$$\nu_e + d \rightarrow p + p + e^- \tag{3.3}$$

where ν_x represents all three flavors of neutrino – ν_e, ν_μ, ν_τ. SNO is also able to detect electron antineutrinos via the Charged current $(\overline{CC})$ reaction on deuterons;

$$\overline{\nu_e} + d \rightarrow e^+ n + n - 4.03 MeV \tag{3.4}$$

Over the course of its experimental lifetime, SNO has undergone three different phases of neutrino detection. The first phase, referred to as the D_2O phase, consisted of the conditions which have already been described - namely 9,500 PMTs and a target volume of 1,000 metric tons of D_2O. The D_2O phase lasted from November 1999 to May 2001. The second phase, often referred to as the Salt phase, involved the addition of 2,000 kg of NaCl to the D_2O

volume in order to increase the capture efficiency of neutrons produced in the NC interaction. Data for the second phase was taken between July 2001 until August 2003. The final phase of SNO was termed the NCD phase, for 'neutral current detectors' (a somewhat inaproppriate designation), and consisted of reverting back to a pure D_2O target volume while adding an array of 40 ^3He proportional counters for an enhanced neutral capture efficiency, incorporating a systematically different approach from the Salt phase. The analyses done for this thesis involve data from the D_2O and Salt phases of SNO only; for more information on Phase III, consult [41].

3.3   Backgrounds in SNO

As SNO's primary purpose was the detection of ^{8}B solar neutrinos via Cherenkov radiation, background removal is an essential part of SNO analyses. Although SNO was housed in an extremely clean environment, and the collaboration took significant steps to remove radioactive contaminants from the detector materials and surroundings, there remained a low level of background contamination, primarily due to the decay products of ^{238}U and ^{232}Th. The two specific daughter decays which caused the greatest amount of trouble are ^{214}Bi and ^{208}Tl. These two elements are unfortunate backgrounds for SNO, in that their dominant decay products result in γ rays above the deuteron photo-disintegration energy threshold of 2.2 MeV (for ^{214}B, 2.4 MeV and for ^{208}Tl, 2.6 MeV). This reaction obviously interfered with the solar neutrino NC signal. Another unfortunate artifact of the ^{238}U and ^{232}Th decay chains is low

energy $\beta - \gamma$ events, resulting in Cherenkov light that could lead to the mis-reconstruction of these low-energy events to higher energies, above the solar neutrino analysis thresholds. The expected background contamination of the solar neutrino signal as a function of energy is shown in Figure 3.2.

Energy Distribution for Signal and Background MC
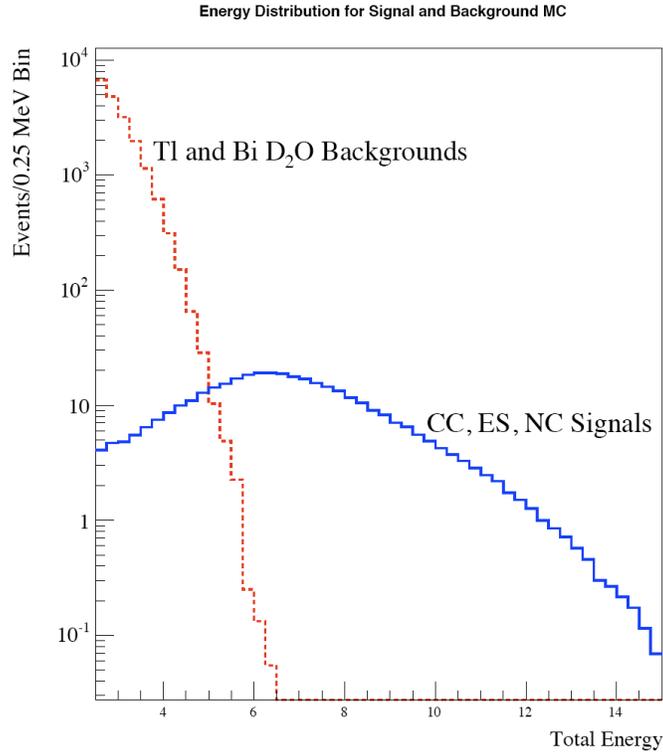
Figure 3.2: Expected backgrounds from the ^{238}U and ^{232}Th decay chains, and MC distribution of CC, ES, and NC signal, as a function of total energy. Figure taken from [5].

Another source of backgrounds in SNO was cosmic rays. SNOLAB is situated very deep underground, as is shown in Figure 3.3; however, we still had some contamination of our data set by cosmic ray muons, at the rate of 3 per

44

hour. To remove these events from our data set, we relied on the previously mentioned OWL tubes, which alerted us to Cherenkov radiation in the light water, thereby giving a reliable 'muon tag'. In SNO, any time an event had an associated muon tag, the event was taken to be a muon, and the event itself was removed from the data set, as well as all events occurring within a period of time following the tagged event (20 sec with the current tagging scheme). This ensured that not only would the muon itself be removed, but also all 'muon follower' events, such as spallation products.

3.4   Data Acquisition: Signal and Electronics

SNO was a water Cherenkov detector: neutrino interactions inside the AV were detected via Cherenkov radiation emitted by the propagation of corresponding relativistic charged leptons through the heavy water. If a photon created in this Cherenkov process was incident on one of the 9,500 PMTs in SNO, it could then release a photoelectron from the photocathode of the PMT (depending on the quantum efficiency of the PMT), starting a cascade down the dynode chain. The resultant pulse, roughly 12 ns wide and no more than a few mV in amplitude, then travelled through a ~30-meter long 75Ω RG59-like custom coaxial cable from the PSUP up to the deck, where the cable was attached to a PMT Interface Card (PMTIC). The PMTICs provided high voltage to the PMTs, and also provided the connection to the Front End Cards (FECs, or motherboards), where the signal was processed, digitized and temporarily stored.
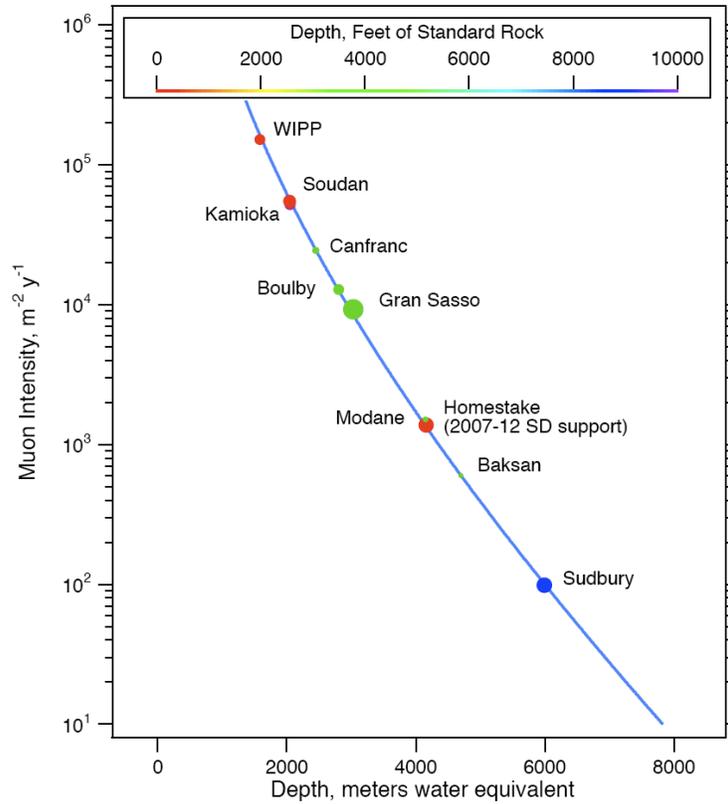
Figure 3.3: Integrated muon flux as a function of depth, with several underground laboratories' depths overlaid. The relative size of each laboratory is indicated by its corresponding circle's size. Figure taken from [6].

The PMT cables were grouped together, such that for every group of 32 cables there was 1 PMTIC. Each PMTIC had 4 HV relays, called paddle cards, which allowed the HV to be enabled or disabled for 8 PMTs at a time. For every PMTIC, there was a corresponding FEC, and for every FEC there were 4 daughterboards that received 8 PMT signals each from the PMTIC. Each PMT signal was routed thorugh a separate *channel* in the electronics, and the channels were organized in groups of 32 (8 channels per daughterboard, 4 daughterboards per motherboard) on the FEC, which we often refer to as a *card*. The cards were all arranged in 1 of 19 *crates*, with 16 cards per crate. This modular organization of the electronics is often internally referred to using a Crate-Card-Channel scheme, with 32 channels grouped per card, and 16 cards grouped per crate.

Each crate had a Crate Trigger Card (CTC) which communicated with a detector-wide trigger system (a more detailed treatment of the SNO trigger system than is given in this thesis can be found in [42]). The SNO trigger system was designed for 'traffic control', in a sense; the rate of PMT hits in SNO due to low-energy backgrounds alone was enough to swamp any reasonably sophisticated electronics readout setup. Collecting and then processing all of the data was less desirable than filtering the data in real time, to rid the data of obvious backgrounds in a low-level manner, and then processing a reduced data set. The trigger system for SNO can therefore be thought of as the first step in background removal for the data set, determining which events should be processed or discarded. A key trigger used in SNO analysis data was the

NHIT 100, which required that a certain number of PMTs (N) were hit within a given time (100 ns); for the D_2O phase, for instance, the requirement was 16 PMTs firing within a period of 100 ns of one another. Other triggers used in the detector depended on the total charge deposited in the detector, or in the case of the Pulsed Global Trigger (PulseGT), simply fired at a constant rate (5Hz) with the purpose of sampling ambient detector activity with no correlation to event occurrences.

Starting with each channel in the detector, the size of an incoming PMT pulse determined whether that channel was registered as 'hit' or not. If a particular channel's signal had an amplitude larger than that channel's discriminator threshold (typically 1/4 of a photoelectron), the discriminator was fired, and the PMT pulse would be integrated to give the charge, and various channel-specific triggers would be sent to the CTC in the form of analog sums. Each crate's CTC would then feed that crate's cards' summed trigger information to a Master Trigger Card (Analog), or MTC/A. The detector-wide sum of all channels' information was then compared to a pre-determined threshold for each type of trigger, and if the signal was above threshold, it was sent to the Master Trigger Card (Digital), or the MTC/D, to be stored, and a Universal Time was assigned to each event using a 10 MHz GPS-synchronized clock, with a backup time stamp from a local 50 MHz clock. The MTC/D also sent out a global trigger (GT) to all crates which forced the collection of each channel's data for that particular event.

If the detector-wide trigger system was triggered, each pulse read in through

the FECs would be processed; otherwise, the event would be ignored. After crossing the discriminator threshold, the PMT pulse was integrated in a variety of ways (short and long integration times, and low and high gains), and stored until a detector-wide trigger arrived. The time at which the signal crossed the discriminator threshold was also recorded using a Time-Amplitude Converter, or TAC. This data was then digitized and put in memory to be read out by the Builder, which combined each event's PMTs' data with relevant run header information, to then be stored and analyzed as part of a run. Though SNO often operated for continuous stretches of time, the data-taking was broken up into runs, generally no longer than 24 hours duration.

3.5 Calibrations

In order to extract a physics result from the SNO detector, we must have an understanding of how well the signals read out by the electronics relate to actual physical quantities. The SNO Monte Carlo is built to represent as accurately as possible the SNO detector conditions; in order to provide SNO-specific physical parameters to the Monte Carlo, like optical constants and energy scales, we need to have some sort of way to measure these quantities. The detector response must be accurately built in to our Monte Carlo, and similarly, we need a means of testing how well the Monte Carlo truly represents SNO conditions.

For these purposes, we rely heavily on a rigorous calibration scheme in SNO. To calibrate the basic charge and timing information which we get from

the PMTs, we must calibrate the electronics and the phototube readouts. To calibrate the detector response to various types of signal, we use a number of calibration sources.

3.5.1 Calibration Sources

The acrylic vessel was built with a 1.46-m wide opening, called the 'neck', through which calibration sources could be passed, and access to the internal target volume was possible. A diagram of the calibration system is shown in Figure 3.4. The majority of calibration sources were deployed via the setup shown here, which allowed sources to be placed at multiple points along not only the central z-axis (up and down), but also along the x and y axes (horizontal plane). There were also 6 external guide tubes, which allowed deployment of sources in the light water, external to the D_2O target volume (only 2 are shown in Figure 3.4).

The sources used for calibration of the SNO detector are itemized in Table 3.5.1; more detail can be found in [43],[44], or [45].

3.5.2 Electronics Calibration: ECA

To calibrate the electronics, we performed two types of specific Electronic Calibration runs, or ECA runs; one run to determine the charge "pedestal" and one run to determine the "TAC slopes"'. These ECA runs focused on extracting the zero offsets, or pedestals, for channels when no signal is present. This allowed us to formulate the electronics 'baseline' of the incoming charge
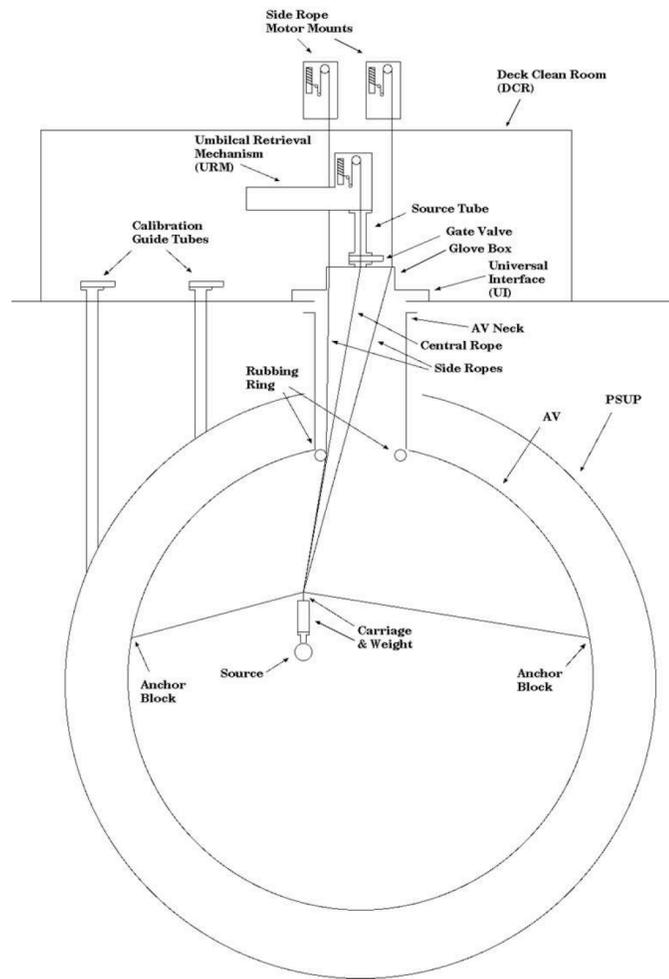
Figure 3.4: Overview of the SNO Calibration System. Figure taken from [7].

Table 3.1: Calibration sources used in SNO

Source	Description	Purpose
Laserball	Triggered, isotropic, multi-wavelength laser source	Main calibration for PMT timing, angular response, efficiency variations, attenuation lengths
^{16}N	Triggered, ^{16}N 6.13 MeV γ source	Main calibration for PMT collection efficiency and energy systematic uncertainties
^{8}**Li**	Triggered, β source from decay of ^{8}Li with endpoint ~ 14 MeV	Confirm ^{16}N calibrations for energy uncertainties and event reconstruction
^{252}**Cf**	Nonn-triggered, ^{252}Cf neutron source	Determine neutron capture efficiency
Acrylic sources, ^{238}**U and** ^{232}**Th**	Non-triggered, U and Th sources in acrylic	Measure detector response to low energy backgrounds
Radon spike	Controlled injections of Rn into D_2O or H_2O	Low energy background calibration
AmBe	"Triggered" source, 4.14 MeV γ and neutron in coincidence	Determine neutron capture efficiency Not used in D_2O phase
pT	Non-triggered γ source $p +^3 H \rightarrow^4 He + 19.8 MeV\gamma$	High energy calibration Not used in Salt phase

and time signals.

The charge pedestal measurement was done by sending a 'pedestal' signal to each channel's discriminator, forcing it to fire. The input signal would then be integrating according to standard data-taking methods (low gain, high gain, short time integration, long time integration). The timing profile measurement fired the pedestals at 31 separate time intervals relative to the global trigger, generating a slope of the Time-to-Amplitude-Converter, or TAC slope.

As the condition of the electronics changed on a regular (weekly) basis, with crates being powered down for repairs, etc., these ECA runs were performed on a bi-weekly basis throughout the running of SNO, as well as any time that the detector underwent any type of power cycling or downtime.

3.5.3   PMT Calibration: PCA

The photomultiplier tube calibration, or PCA, is performed to keep track of the individual PMTs' different timing responses to an incoming signal. Each PMT has a characteristic risetime, or 'walk', which determines the point at which that PMT will cross its discriminator's threshold. An example of this 'walk' is shown in Figure 3.5.

The walk is important in SNO for many reasons. As is evident in the diagram, even if a signal arrives at two PMTs at the same moment in time, if the amplitude of the pulse varies significantly between the two PMTs, the PMT with the lower-amplitude pulse will register the event as occurring at a later time relative to the PMT with a higher-amplitude pulse. This is troublesome
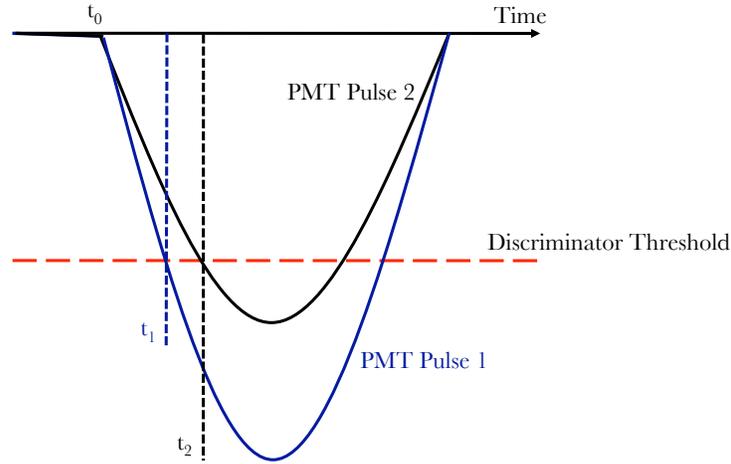
Figure 3.5: Time Walk of PMT signals. Two PMTs with signals originating at the same time, t_0, will cross the discriminator threshold at different times, t_1 and t_2, if their amplitudes are appreciably different. This causes a difference to be registered between the times each PMT was hit, even if the two were actually hit simultaneously.

for a detector like SNO, which depends heavily on the times registered by the PMTs to reconstruct event positions.

The purpose of the PMT calibration is to correct for this walk, or risetime, behavior. By measuring the magnitude of this charge-based variation for each channel, we can correct for it. To measure the discriminator risetime differences, *i.e.* the relationship of charge deposit in a tube and the discriminator firing time, the laserball source was deployed on a monthly basis during SNO's running time.

54

Chapter 4

QRC: The Rate-Dependent Charge Drift Correction

4.1   Charge in SNO

The ~9,500 PMTs used in SNO all read out two important pieces of information upon being hit: first, the precise time at which a PMT fired relative to other PMTs in the detector; and second, the charge of a particular PMT's signal. The single-photoelectron spectrum of the SNO PMTs is well-known and well-modeled, and therefore the measurement of charge from a PMT is thought to be a good indicator of the number of photoelectrons released when a photon hits the photocathode of the PMT. Charge is also used to correct each PMT's recorded hit time; as was described in Chapter 3, each PMT has an associated 'walk' or risetime, such that for a given discriminator threshold, a PMT with a smaller pulse will cross the threshold later than a PMT with a

larger pulse. To provide optimal PMT timing, calibration PCA runs are carried out with a diffuse laser source (the Laserball) to determine the correction factor necessary for each PMT hit time as a function of its integrated charge. In this way, the PMT charge is used to remove the PMT 'walk', or risetime delay, allowing for more accurate timing.

4.2   The Problem of 'Charge-Since-Last-Hit'

During the early running of the SNO detector, it was discovered that the recorded PMT hit times appeared to shift, depending on the time since the last hit on a channel: the longer a channel had gone without a hit, the bigger the shift in its next recorded time. This 'time-since-last-hit' (TSLH) effect was a significant problem, because it affected position reconstruction, which is highly dependent on the relative times each PMT in an event was hit. A new calibration was developed to correct this problem by shifting times according to the gap between events in each channel in the detector[46].

While researching and developing a fix for the TSLH effect, SNO discovered a separate but similar effect on the PMT charges: the charge pedestals appeared to vary depending on detector rate. The dependence, though initially dubbed 'Charge-Since-Last-Hit', or QSLH, in the same fashion as TSLH, was ultimately found not to be a dependence on rate, as much as on the 'hit rate', or the number of recorded hits in an crate or front end card per time [43]. It is believed that this charge pedestal drift is related to the loading of the front end electronics: the more hits per unit time, the more front-end activ-

ity, causing more current to be drawn through the system, which may affect some of the front-end voltages. Though the exact origin of this rate-dependent charge remains unknown, the knowledge of the effect's cause is not necessary to create a fix for it, as will be shown in this chapter.

Although a correction was found fairly quickly for the TSLH time drift, multiple attempts at developing a correction for the rate-dependent charge drift fell short. No straightforward calibration could be developed for the charge drift because it required a dynamic measurement of the charge pedestals under various loading conditions, including the history (recent, and on the scale of minutes) of front-end activity. An early attempt at such a calbration was made using an external pulser to generate pedestal-measuring events at different rates and under differ loading conditions, with the intent of creating a model from this information (for instance, information regarding pedestal shift vs rate, or vs hits, or vs rate*hits). While the method was potentially quite useful, the dynamic behavior of the rate-dependent charge drift, which changes significantly over short periods of time, would have required many of these types of pulser runs throughout different stages of the detector (in addition to SNO's standard electronics calibration runs which were done twice weekly, as described in Chapter 1). Similarly, the statistics of the pulser events were fairly low in this setup, and in the case of such a sensitive effect, more precision was needed to actually implement a correction.

For much of SNO's data analysis, therefore, the information contained in the charge measurements was never used. For example, SNO's primary

energy reconstruction algorithm placed cuts on the charges of PMTs that it included in its measure of energy, but these cuts had to be removed because the rate-dependence of the pedestals created a rate-dependent energy scale. In addition, the energy estimator could not use risetime-corrected PMT times, because as described above and in Chapter 3, the risetime, or walk, correction depended on the charge measurement. Lastly, cuts that could remove events based on the measured charges on the PMTs were never developed, because the charge measurement was considered so unreliable.

However, a new analysis aimed at lowering SNO's energy threshold (called 'LETA' for Low Energy Threshold Analysis), required much lower backgrounds and systematic uncertainties, and therefore needed to use the information in the charge measurement. Furthermore, and more critically for the work presented in this thesis, any rate dependence of the charge measurement could lead to time-correlated effects that could create a background for a periodicity study of the solar neutrino flux.

4.3   Crosstalk as an In-Situ Monitor

In order to correct for the rate dependence of the charge pedestal, an in-situ monitor was needed to record the dynamic variation in the pedestals. This was a daunting request: we needed something that provided a pedestal measurement throughout the SNO data set, and yet did not interfere with data-taking or add to the overall volume of data. Also, the correction would ideally be retro-actively applicable; something that could be used to correct

data which had already been taken.

Fortunately, such an in-situ monitor existed: the electronic crosstalk[1] between neighboring channels that occurred when one of the channels recorded a very large charge. As shown by M. Dunford[47], a crosstalk hit in a channel can be isolated from other background/noise hits by three simple characteristics:

1. The charge on the tube in question must be "low", or near the value of the channel's pedestal. Here, "low" charge means that the channel must have a charge value less than 5 counts above pedestal.

2. The second condition is that the crate+card+channel configuration, within each crate, must be adjacent to that of the potential 'crosstalk channel' (the 'nearest neighbor' effect).

3. A third criterion which must be met has to do with the timing of the potential crosstalk channel. The signal in the crosstalk channel must occur within a certain time window (12ns to 25ns) after the high-charge signal to which it corresponds.

Using these three characteristics, Dunford created a 'crosstalk tag' which identified crosstalk hits in each event, so they could be discarded from the standard SNO analyses.

For this study, however, the crosstalk hits are in *no* way a background to be discarded: the crosstalk hits are extremely useful because they provide an instantaneous, dynamic measure of a channel's charge pedestal. Since crosstalk

---

[1]Not to be confused with optical crosstalk

is a short-timescale (high-frequency), bipolar phenomenon, and SNO uses a (comparatively) long integration gate of roughly 60ns, crosstalk hits will tend to integrate to zero charge. This means that we are able to look at the variation of the charge measured in crosstalk hits as a function of time, or rate, or any other front-end activity-related parameter, and in so doing we can get a true in-situ monitor of how the pedestals are drifting. Figure 4.1 shows a profile histogram of the crosstalk pedestal in a standard data run (a 'neutrino run'), as a function of card hit rate. It is obvious from this plot that the pedestal value changes as a result of a hit having occurred during low-rate or high-rate (relatively speaking) activity within the run. This particular behavior is key to understanding the rest of the approach described in this chapter.

4.4 Generation of the crosstalk-based correction

To generate the charge drift correction, we essentially use the charge measured in crosstalk hits as an indicator of the degree to which the rate-dependent charge drift is affecting any given run's data. The advantage of using crosstalk is not just that it provides a measurement that is close to the charge pedestal itself, but also because it is a purely electronics effect. For the sake of argument, we could use the upper edge of a single photoelectron charge spectrum (the 'high half point', or HHP) to measure the drift in charge, but any drift measured this way could easily be caused by a variation in a PMT's gain, and
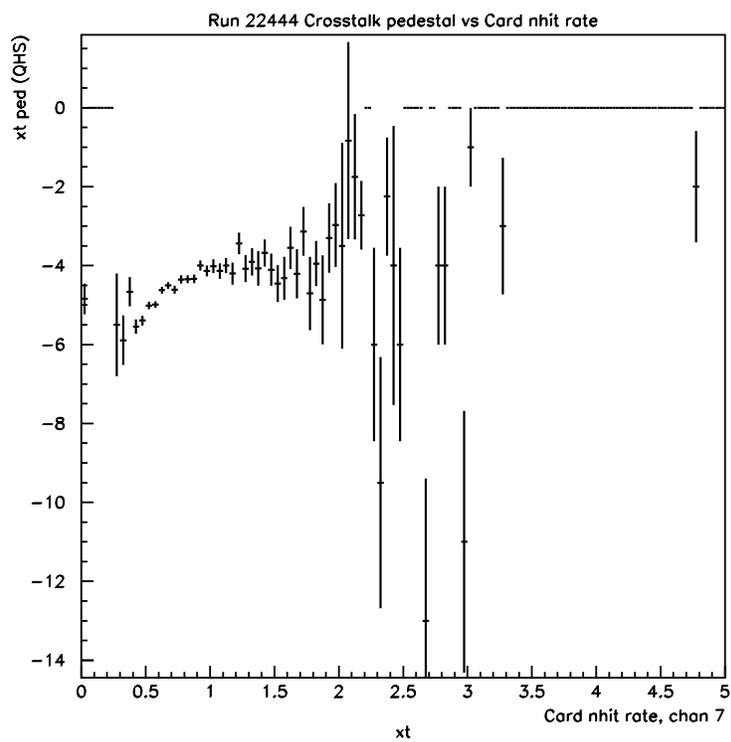
Figure 4.1: Sample crosstalk pedestal as a function of card hit rate

61

not necessarily because of a local loading effect. Similarly, using the low edge of the single photoelectron charge spectrum (the 'low half point', or LHP) could show variations caused by changes in the channel threshold, again not necessarily giving a true measurement of the drift itself.

Thus the amount by which the crosstalk charge spectrum is displaced from the (ideal) peak value of zero should be a direct reflection of how much the rate-dependent charge drift is shifting all other data in that run. This is assuming, of course, that rate dependence is the dominant reason for this pedestal shift. Taking this to be the case, we can divide the crosstalk charge spectrum according to corresponding time within a run, and then relate each charge value to the card hit rate for its time period.

At this point it is important to devote a bit of attention to this last statement. The degree of 'differential-ness' with rate here is somewhat arbitrary. Though the charge drift is known to depend on detector rate, it also is seen to be a loading effect, and so it depends on card hit rate and channel hit rate, with varying degrees of correlation. In this study, we use card hit rate (hits per second per card), as it appears to be more accurate in determining charge rate dependence than simply using detector-wide hit rate. In addition, using card hit rate allows us to create one correction per (approximate) card hit rate, which is not so limiting (or cumbersome in its application) as using channel hit rate (one correction per channel - 9500 total).

This can be more easily shown by looking at a collection of plots. The first group, shown in Figures 4.2, and 4.3, is a cross-sampling of the same type of

profile histogram as shown in Figure 4.1. It shows the distribution of crosstalk pedestal charge value as a function of card hit rate for a group of 2 different channels in the same card (and crate). It is fairly obvious from looking at these plots together that there is no correlation between rate-dependence of the pedestal value among different channels on the same card. In light of this, different channels will therefore have to be assigned different corrections.

On the other hand, in Figures 4.4 and 4.5, the plots are grouped differently: here, each histogram represents the same channel number, but each is from a different run (and does not depend strongly on crate/card number).

From these figures it is evident that there is a strong correlation here, among different runs, but with channel placement (location on a card) the same. Due to this similarity of channel behavior from card to card, it appears to be sensible to approach the charge drift problem as a function of card hit rate, and to develop a correction that depends on channel location but not on which card the channel is on.

4.5 Implementation of the Crosstalk-based Correction

4.5.1 Generation of the Correction

In order to generate the actual crosstalk-based correction, we have up until this point considered each run individually, since there is a fairly noticeable variation, run-to-run, in rate and thus in associated charge drift behavior. So
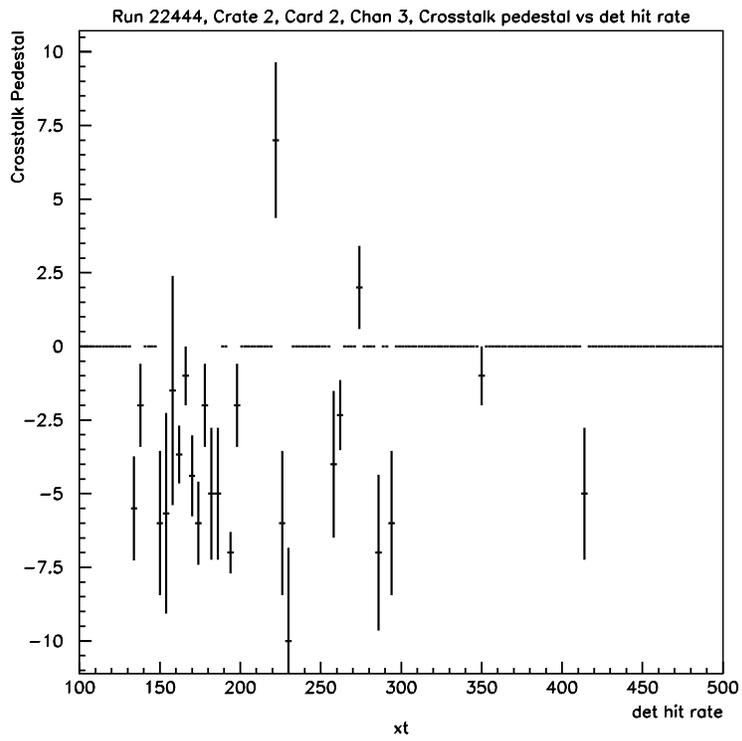
Figure 4.2: Run 22444 Crosstalk Pedestal as a Function of Card hit rate, Channel 3
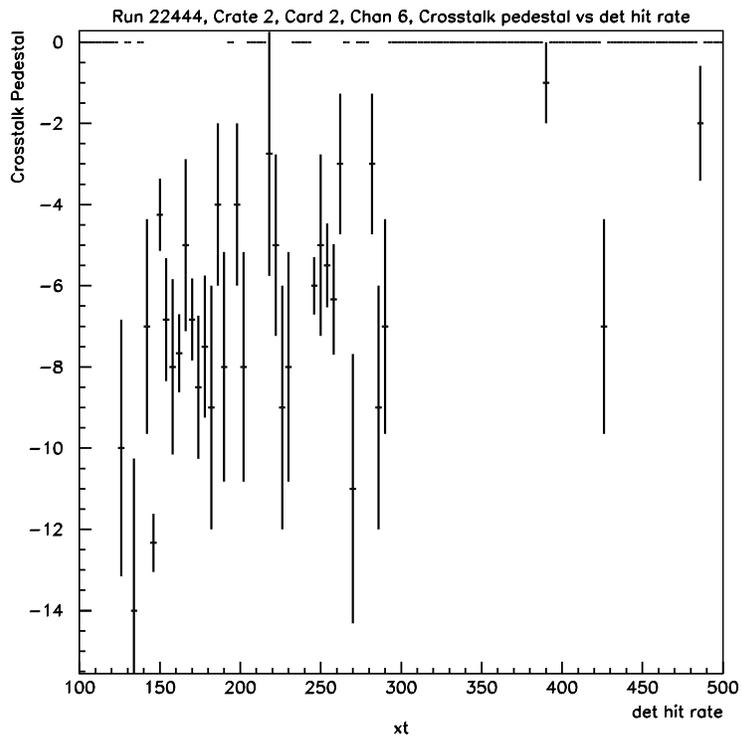
Figure 4.3: Run 22444 Crosstalk Pedestal as a Function of Card hit rate, Channel 7
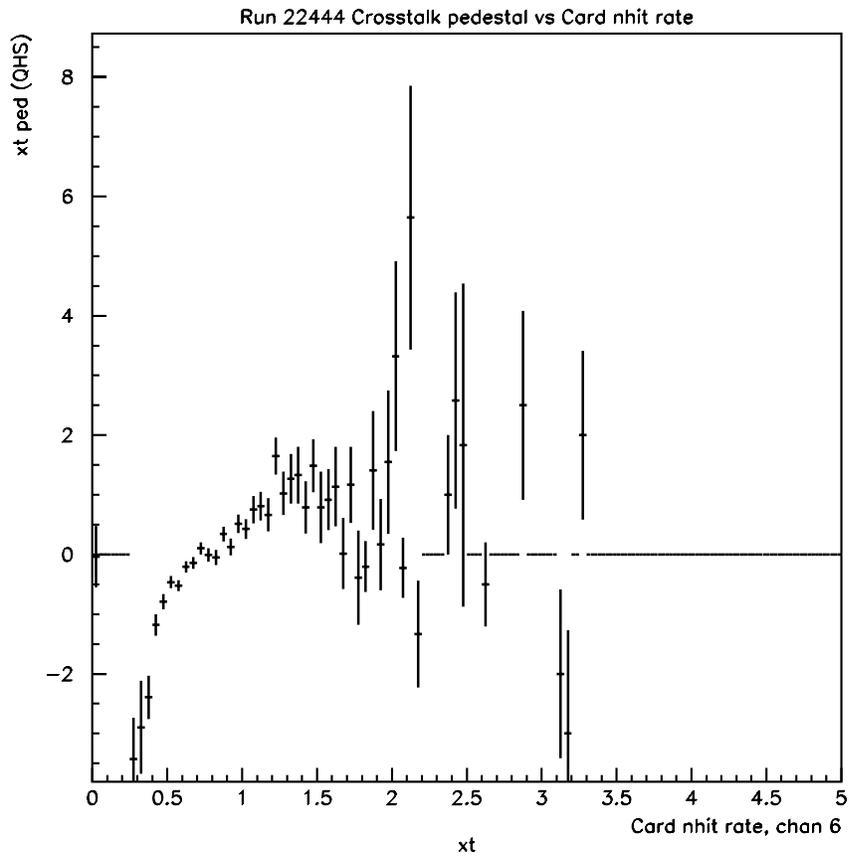
Figure 4.4: Run 22444 Crosstalk Pedestal as a Function of Card hit rate, Channel 6 (all cards)

66
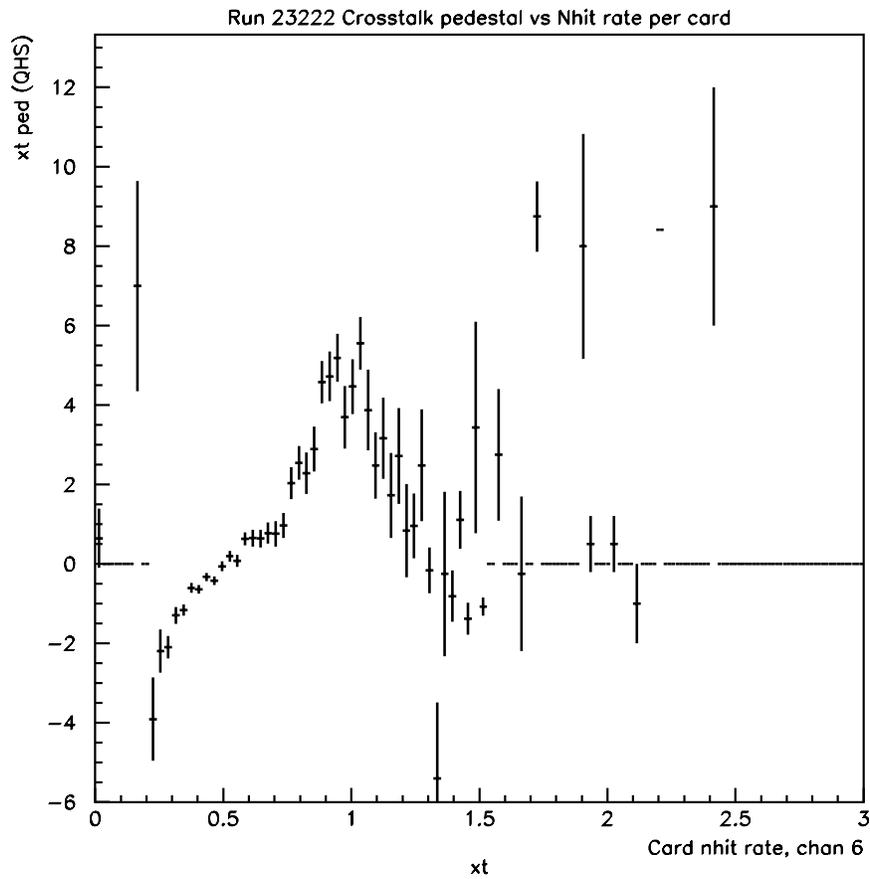
Figure 4.5: Run 23222 Crosstalk Pedestal as a Function of Card hit rate, Channel 6 (all cards)

the approach I describe should be understood as being applied run by run, with a unique correction for each run.

Using the fact that we would like the crosstalk pedestal to have a corrected charge value of roughly zero, generation of a crosstalk-based correction is fairly straightforward. The overall procedure is outlined in the Flow Chart, Figures 4.6 and 4.7. First, we run through a particular run, say Run 22444, flagging the crosstalk bit and breaking the run into different periods, each of x minutes length. We can then group each crosstalk pedestal according to its associated period, and determine what the card hit rate was for each period as well. This then allows us to group all event information in the run according to the range of card hit rates in that run. This way we can fill an associated titles-type array which stores information on period, card hit rate (for each period), and average crosstalk pedestal as a function of card rate. This is considered the corrections array, and is then read in during the second pass of the charge drift correction.

4.5.2   Application of the Correction

At this point, we start the second stage of the correction, which consists of applying the correction file that was generated in the first stage. The main function of this 'second pass' is to group all events - not just crosstalk - into their associated periods (still of length x minutes), and use the setup of the correction array to pair each PMT hit with its associated average crosstalk pedestal (as determined by card hit rate during that period). This average

First Pass: Correction Generation

Run 22444:

Period Division 1	Period Division 2	Period Division 3	Period Division 4	Period Division 5...

What are the average
Card Hit Rates
During this Period?

Create Output
Corrections
File, 'corr.out'

What is the Average
Crosstalk Pedestal
per Card and Channel
During this Period?

CORR.OUT →

Record Period Number, Card Hit Rates, Average
Crosstalk Pedestal per channel per card in
corr.out.

Figure 4.6: Flow Chart: first pass of charge drift correction.

crosstalk pedestal is then used as a correction value for each hit, and is sub-
tracted from the charge of the particular PMT which fired. To reiterate; the
correction for each PMT is a dynamic value, which changes depending on what
the associated card hit rate is during a given time period for that particular
hit.

For example, say in the first pass through a given run, I have calculated
that for all channels, in the 7th period-division of the run, the average card hit

Second Pass: Correction Application

Run 22444:

Period Division 1	Period Division 2	Period Division 3	Period Division 4	Period Division 5...

CORR.OUT

What is the QHS for each event in this period?

What are the associated Card Hit Rates for this period?

Chan. 0	Chan. 1	Chan. 2	Chan. 3	Chan. 4	Chan. 5	Chan. 6	Chan. 7...

What is the associated average crosstalk pedestal for this channel and this card hit rate?

What is the QHS for each event in this channel and in this period?

Subtract off average crosstalk pedestal value from QHS for this event

QSLH-Corrected Charge!

Figure 4.7: Flow Chart: second pass of charge drift correction.

rate for crate 1, card 1 was 2.3 hits/card/sec. For channel 14 in crate 1, card 1, we calculate an average crosstalk pedestal for this period (and thus this card hit rate) of +1.3 counts. We will then take all channel 14's at this card hit rate - throughout the run, in all periods - and calculate the average crosstalk pedestal for that channel, at that card hit rate. This allows us to group our information according to channel and local card hit rate, and then apply a global 'if (card hit rate = w) and (channel = y) then (crosstalk pedestal = z)' statement. With this method we are able to apply our correction (subtract crosstalk-pedestal value from the actual charge value) depending on what the local rate behavior is.

So following the previous example, for ALL channel 14's in the detector (regardless of crate/card) with a card hit rate of 2.3 hits/card/sec, the associated average crosstalk charge value is +1.5 counts. This means that in our second pass, once we get to an event that occurs with an associated card hit rate of 2.3 h/ca/s, and is read by any 'channel 14' in the detector, we will know this corresponds to an average crosstalk pedestal of +1.5. This +1.5 counts is treated as a shift in each channel's charge pedestal, and is thus subtracted from a channel's charge recorded for the current hit. What I am left with, then, is a rate-corrected charge (charge - 1.5 = corrected charge). The correction was given the name 'Charge-Rate Correction', abbreviated by QRC.

4.6   Results of the QRC Correction

4.6.1   Results of QRC on Crosstalk Data

The first approach at verifying that QRC has done what we intended it to do was to apply the correction to the crosstalk data only. Though the mean of the crosstalk peak would necessarily be more centered around zero, we wanted to ensure that the rate dependence of the crosstalk pedestal decreased. In Figures 4.8 and 4.9 one can see that indeed, after applying the correction (channel by channel) to crosstalk data alone, the pedestal value becomes significantly less dependent upon rate, as we would hope.

4.6.2   Results of QRC on ^{16}N Calibration Data

Given that the rate-dependent charge drift correction modifies the charge spectrum of crosstalk data for neutrino runs in the way we would expect, it is important to see if it affects all other data similarly. The same rate-dependence plots were generated for all events within four different types of runs: neutrino (already shown), low-rate, medium-rate, and high-rate ^{16}N calibration runs. Figures 4.10 through 4.15 show the effects of the charge drift correction on ^{16}N crosstalk data. The correction itself appears to have varying degrees of accuracy depending on rates, probably caused by the fact that we have not created a separate correction for each channel, or perhaps because we cannot use periods that are very short to develop the correction due to limited statistics. We can see, for example, the central role that statistics have here; the
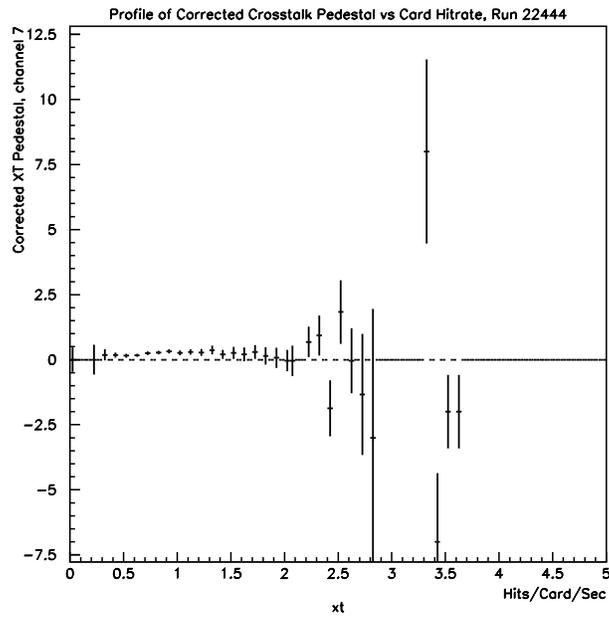
Figure 4.8: Drift-corrected crosstalk pedestal as a function of card hit rate, run 22444, channel 7
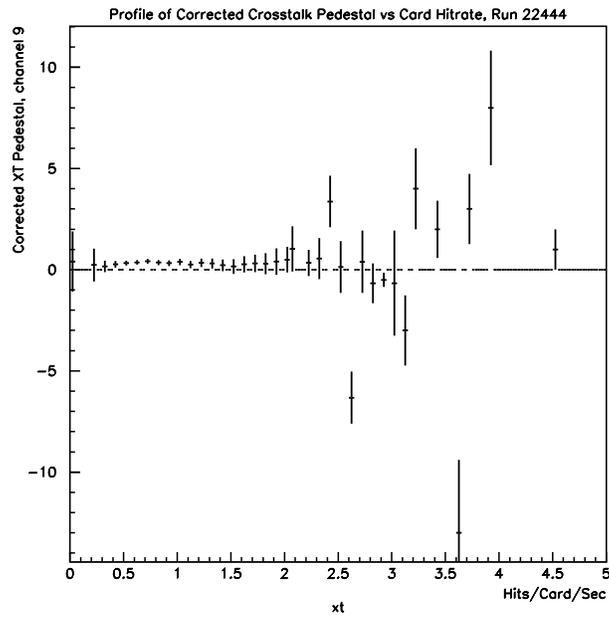
Figure 4.9: Drift-corrected crosstalk pedestal as a function of card hit rate, run 22444, channel 9

value of the corrected crosstalk pedestal is almost invariably closer to zero for points with higher statistics. This reinforces the need for a correction which comes from a large pool of data for each run.

4.6.3   Results of QRC on Charge Spectra

Another telling test for the charge drift correction comes from investigating how the actual charge spectra for various types of runs are impacted. Figure 4.16 represents the entire uncorrected charge spectrum for low-rate, medium-rate, and high-rate ^{16}N calibration data, and neutrino data. Figure 4.17 represents the rate-corrected version of this plot. One can see that before the correction is applied, the four sets of data do not line up well, but that after the correction is applied, there is very good agreement. This is, of course, what we would optimistically expect; by taking the rate-dependence out (or at least minimizing it), we should expect ^{16}N calibration data, taken at differing rates, and neutrino data (inherently very low-rate) to look much more similar.

Figure 4.18 shows a distribution that follows directly from the previous plots. Since the charge spectrum for all of the run types listed above is altered by this correction, so should be the high half point (the upper edge) of the single photoelectron spectrum. Taking a plot of high half point versus detector rate[48] for all uncorrected data, and overlaying the high half points of this new, corrected neutrino and ^{16}N calibration data, shows that this correction does compensate for rate dependence. There is a very visible re-grouping, over
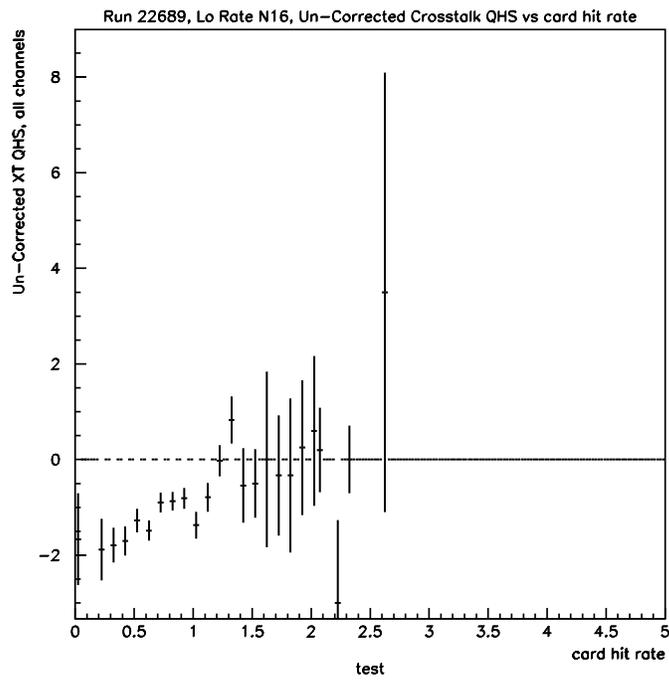
Figure 4.10: Low-rate uncorrected ^{16}N calibration data. Crosstalk pedestal as a function of card hit rate.
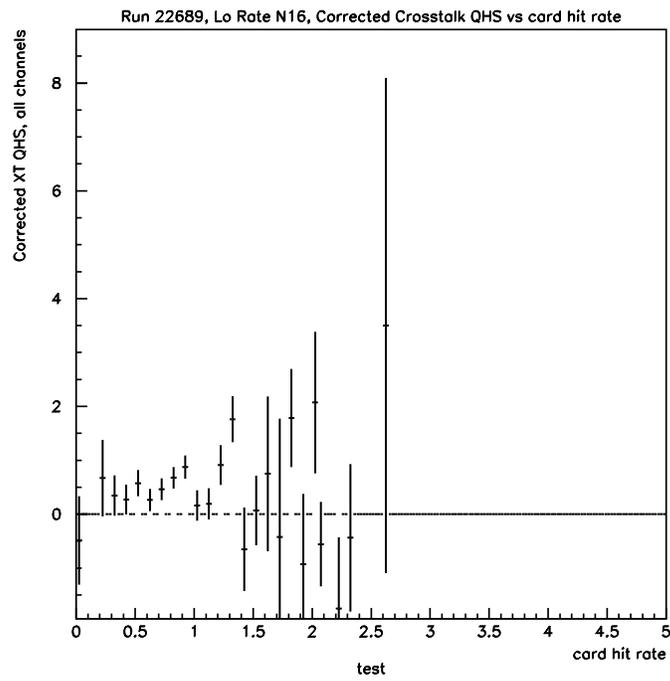
Figure 4.11: Low-rate corrected ^{16}N calibration data. Crosstalk pedestal as a function of card hit rate.
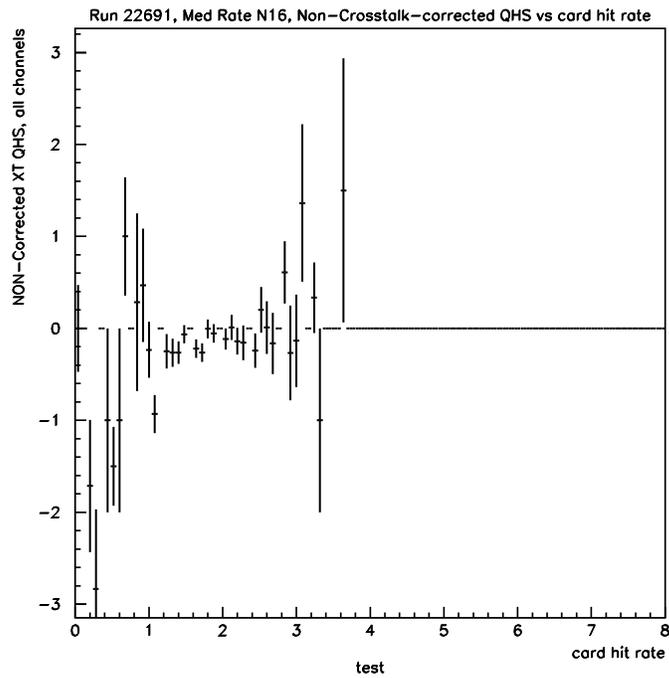
Figure 4.12: Medium-rate uncorrected ^{16}N calibration data. Crosstalk pedestal as a function of card hit rate.
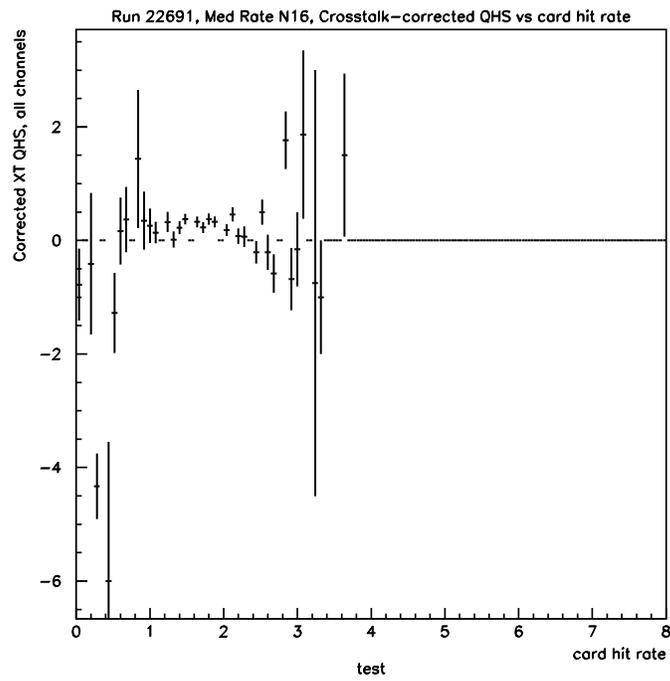
Figure 4.13: Medium-rate corrected ^{16}N calibration data. Crosstalk pedestal as a function of card hit rate.
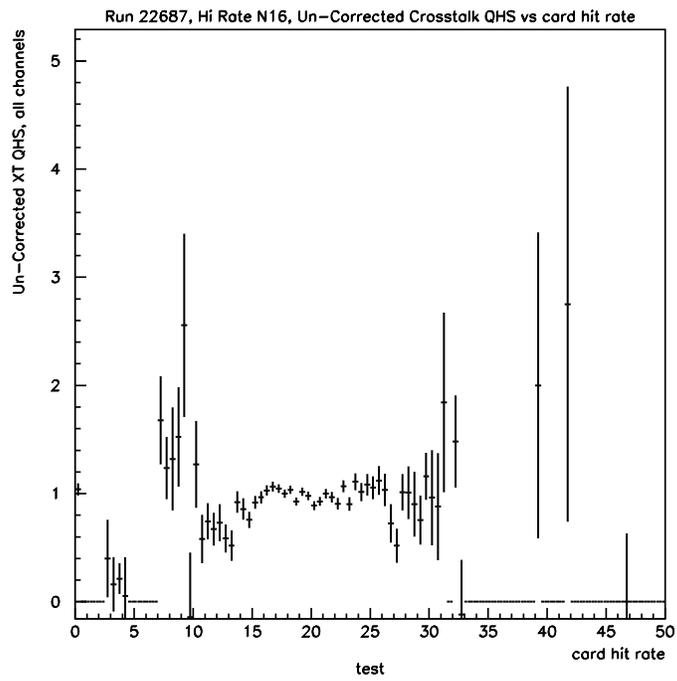
Figure 4.14: High-rate uncorrected ^{16}N calibration data. Crosstalk pedestal as a function of card hit rate.
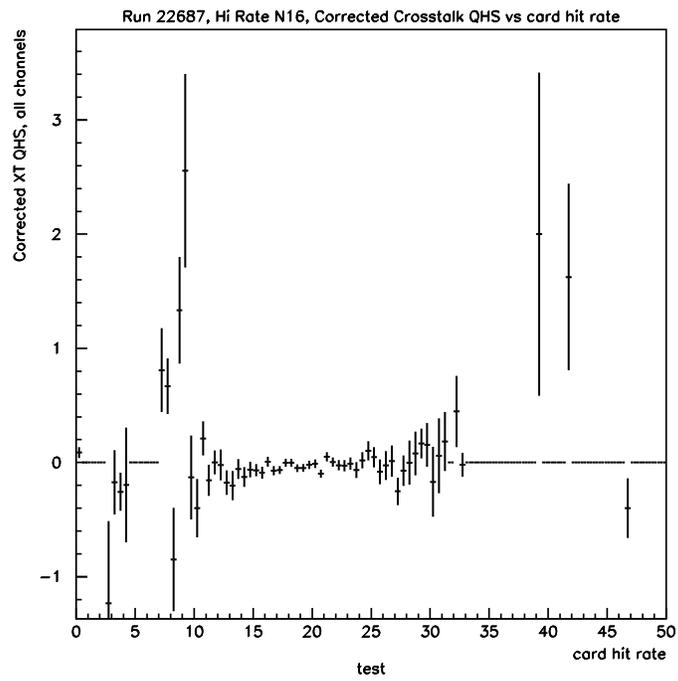
Figure 4.15: High-rate corrected ^{16}N calibration data. Crosstalk pedestal as a function of card hit rate.
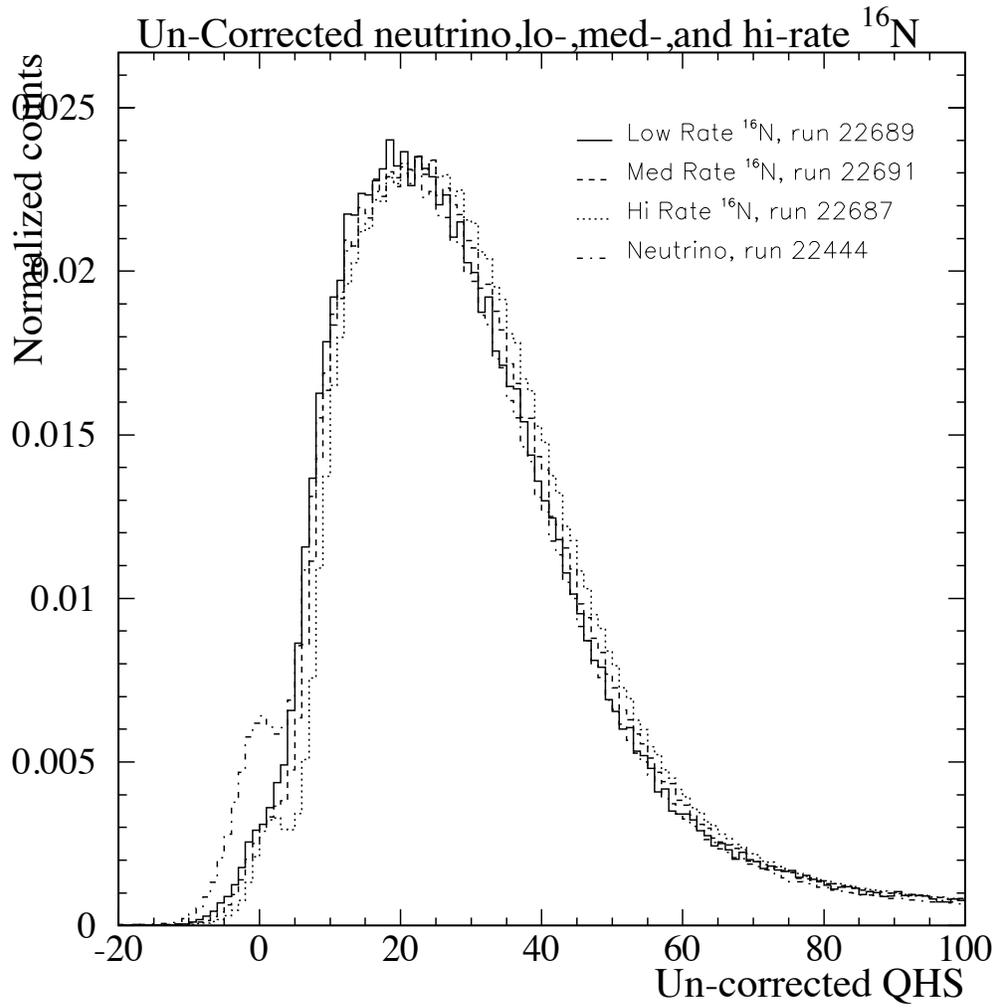
Figure 4.16: Superposition of charge spectra for low-rate, medium-rate, and high-rate ^{16}N data, and neutrino data, with NO rate-dependent drift correction
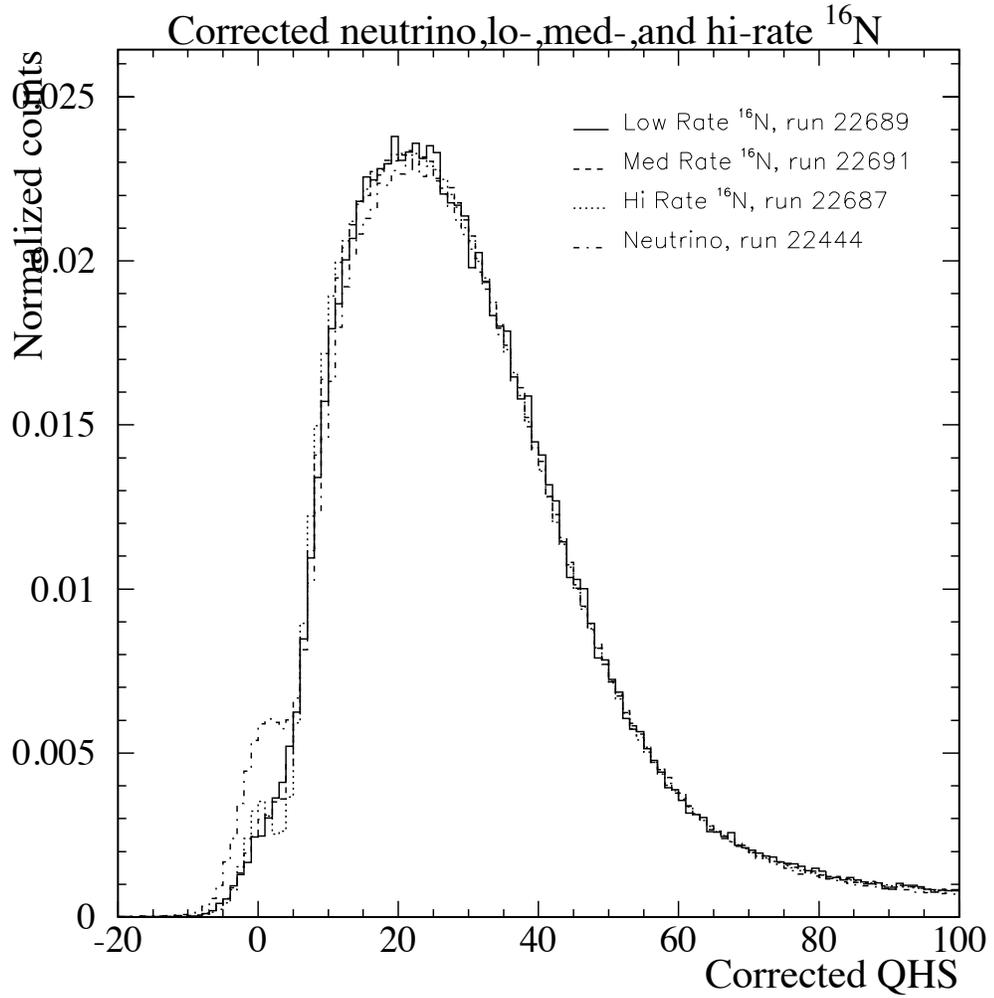
Figure 4.17: Superposition of rate-corrected charge spectra for low-rate, medium-rate, and high-Rate ^{16}N data, and neutrino data.

a range of one count (43 to 44), of this new data, whereas before there was a much broader span of the uncorrected high half points' grouping (five counts).

4.6.4 Results of QRC on Fitted Times

Though the preceding plots are reassuring, it is also important to determine how this rate-dependent charge drift correction alters the time residuals for any given run, since as we mention at the start of this chapter, the charge is uesd to correct the PMT hit times (the 'walk' correction). The time residual can be thought of simply as the difference between the actual time an event from a calibration source triggers the detector (*i.e.* the true time of the event) vs the time the detector reconstructs the event as having occurred. To test this, we can look at both low-rate and high-rate ^{16}N calibration data and compare pre- and post-correction time residuals. These comparisons are shown in Figures 4.19 and 4.20. There does not appear to be any great difference for either low-rate or high-rate data.

However, to get a better understanding of how the correction actually affects the fitted times, it is perhaps more enlightening to look at plots of the number of PMTs fired within a window of $\pm$ 10ns of the time residual peak (a window called 'Nwin') for these two runs, as this window plays an important role in energy reconstruction for SNO. These Nwin values are shown in Figures 4.21 through 4.25. All Nwin 'corrected' values here are generated using time which has already been risetime-corrected, and the time offset has been accounted for.
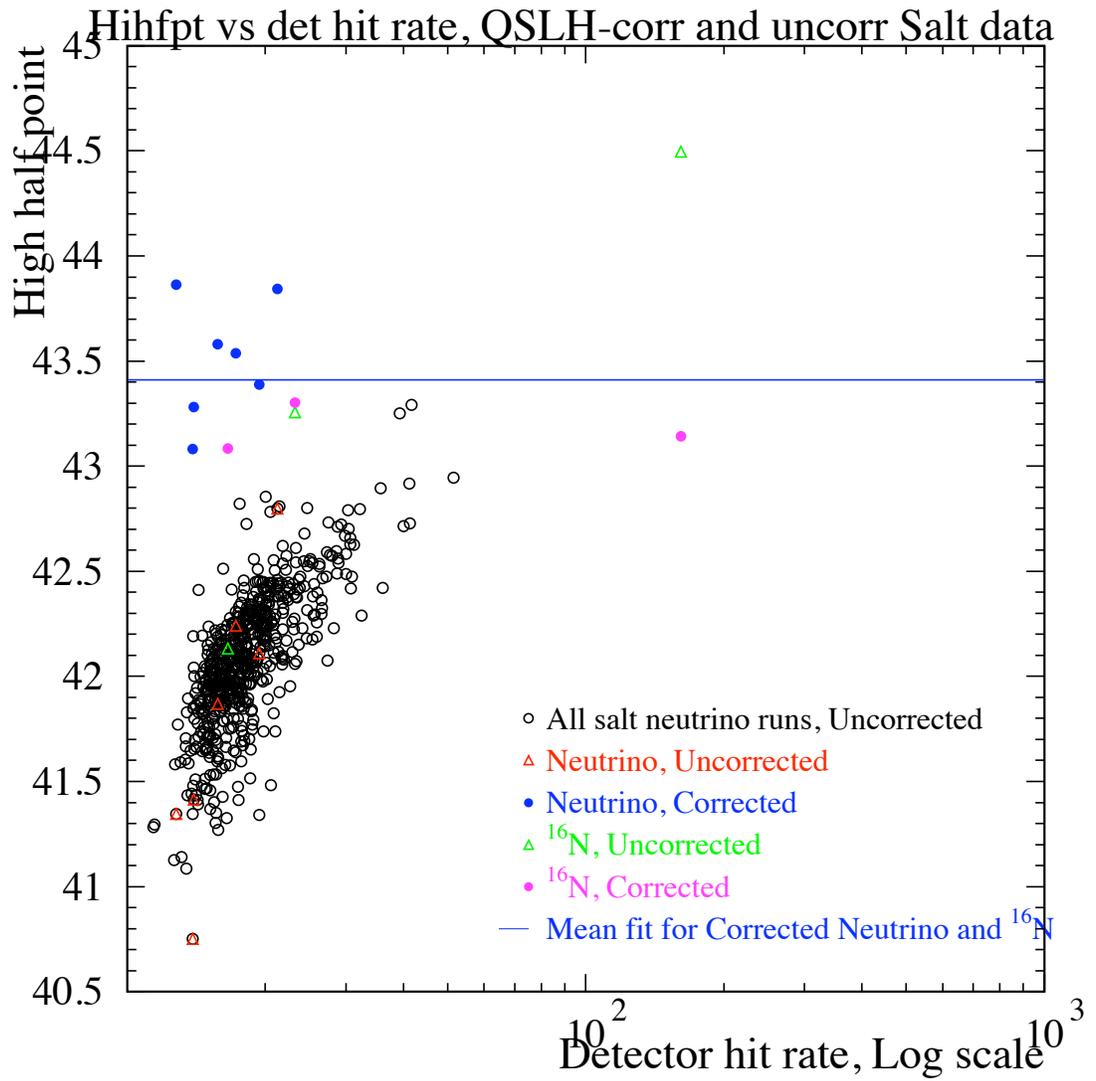
Figure 4.18: High half point for Salt-phase data, versus detector hit-rate. High half points with and without rate-dependent charge drift correction.
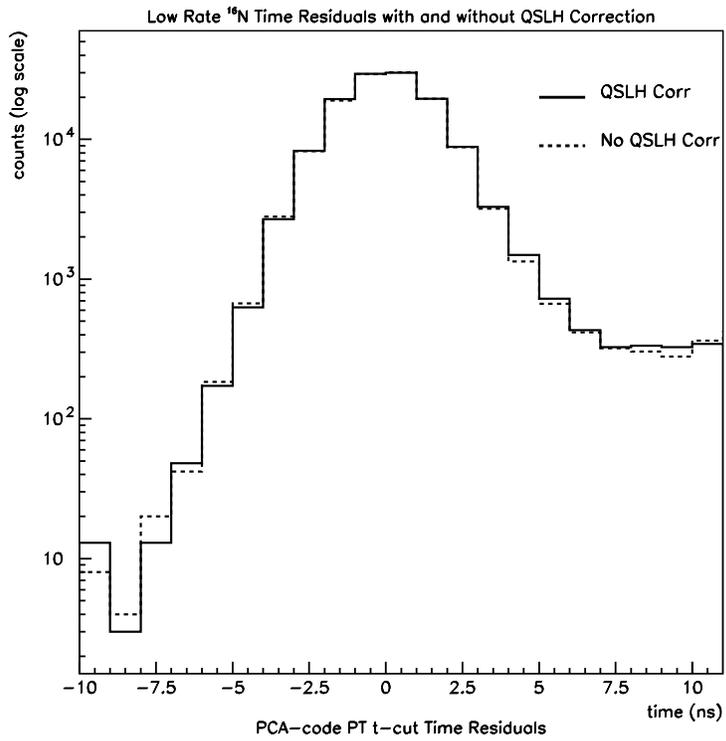
Figure 4.19: Time residuals for both corrected and uncorrected low-rate ^{16}N calibration data (run 22689, enlarged time window)
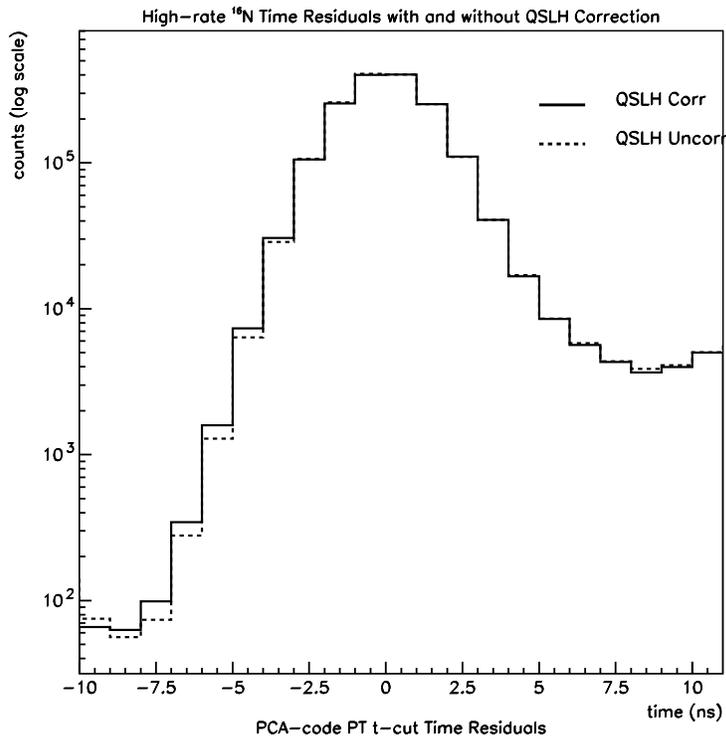
Figure 4.20: Time residuals for both corrected and uncorrected high-rate ^{16}N calibration data (run 22687, enlarged time window)

4.7 Note about using Risetime Corrections

Figure 4.25 shows that the actual value of Nwin after correcting for the rate-dependent charge drift becomes slightly higher for low-rate ^{16}N calibration data, but decreases for high-rate ^{16}N calibration data. This could be an argument against using this type of correction: why would we want to reduce the number of PMTs that fall into this 'good' region? The explanation for this comes from the fact that the basis for our correction is the idea that we should shift crosstalk pedestal data to center around zero. However, to calibrate high-rate ^{16}N data (indeed, all data), we use high-rate PCA calibration data, which could be leading us astray. The pedestals generated from the PCA calibration data, which are used to generate the risetime corrections, will be shifted away from zero: the PCA calibration runs are affected by the rate-dependent charge drift like all other types of runs.

In order to better agree with the high-rate risetimt calibration data, our implementation of the QRC correction has been modified to force the crosstalk charge pedestal to better agree with the high-rate risetime-correction data. Therefore, the crosstalk charge offset is now no longer set to force agreement with a pedestal value of zero; instead the offset value is dynamically calculated whenever SNO performs PCA risetime calibrations, and the values determined from a QRC analysis of the PCA risetime data are stored in files to be applied to all data taken following that date, until the next PCA calibration is performed.
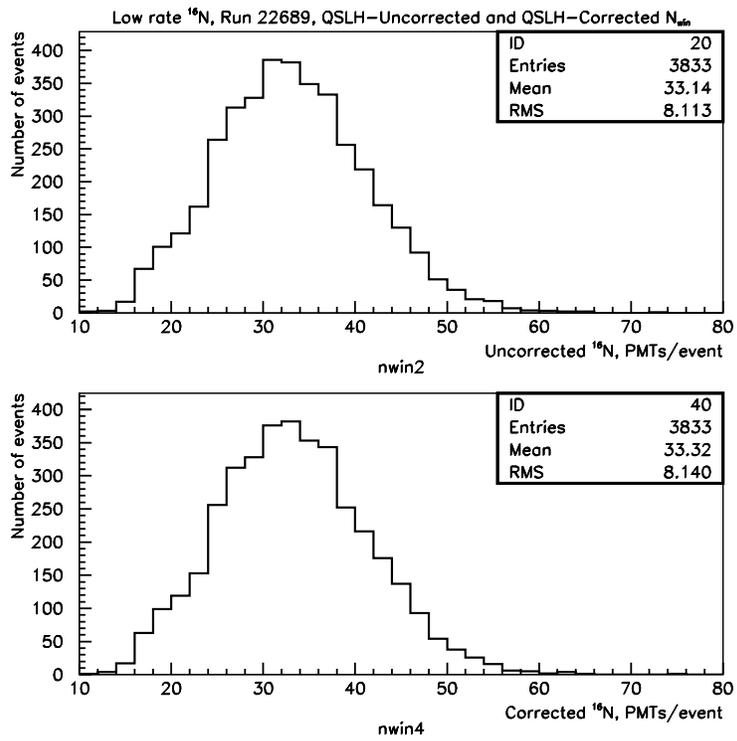
Figure 4.21: Nwin for low-rate ^{16}N calibration data (run 22689), before and after rate-dependent charge drift correction.

High rate ^{16}N, Run 22687, QSLH−Uncorrected and QSLH−Corrected N_{win}

ID	10
Entries	50408
Mean	33.44
RMS	8.071

Number of events

Uncorrected ^{16}N, PMTs/event

nwin1

ID	30
Entries	50408
Mean	33.28
RMS	8.038

Number of events

Corrected ^{16}N, PMTs/event

nwin3
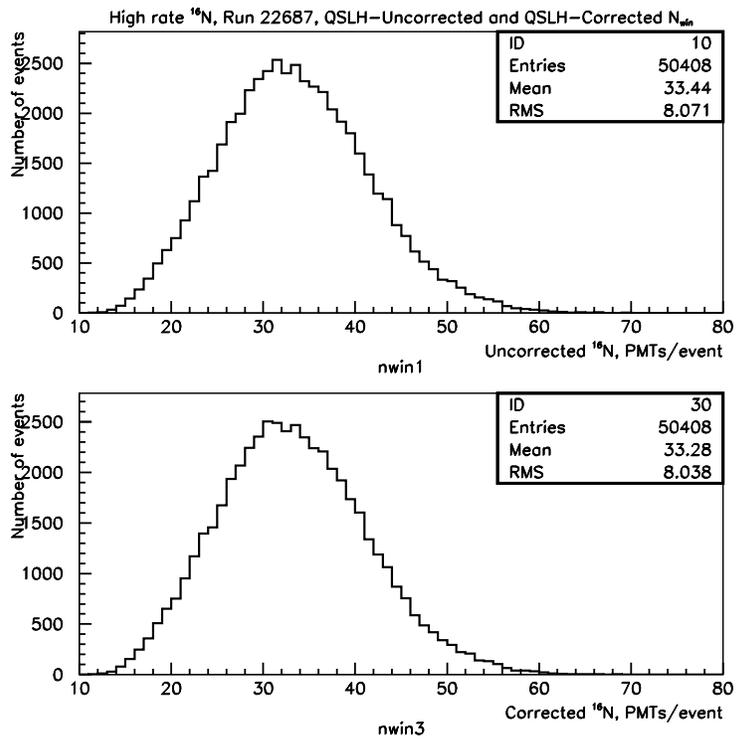
Figure 4.22: Nwin for high-rate ^{16}N calibration data (run 22687), before and after rate-dependent charge drift correction.
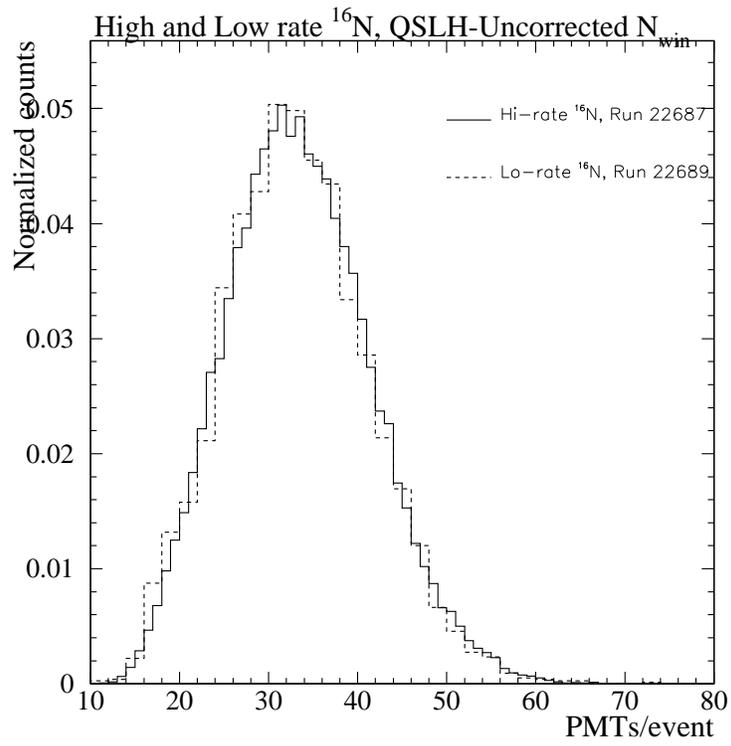
Figure 4.23: Nwin of rate-uncorrected high- and low-rate ^{16}N calibration data
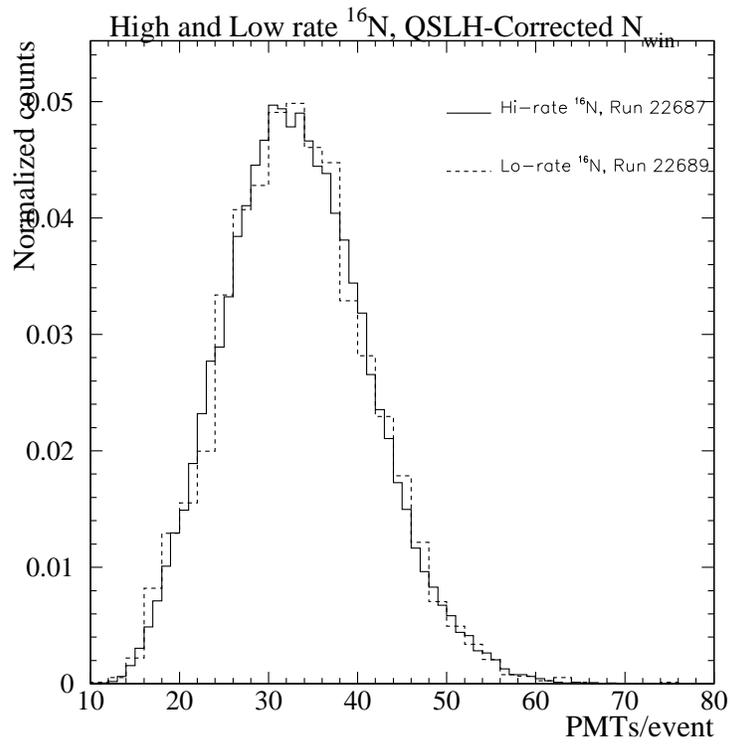
Figure 4.24: Nwin of rate-corrected high- and low-rate ^{16}N calibration data
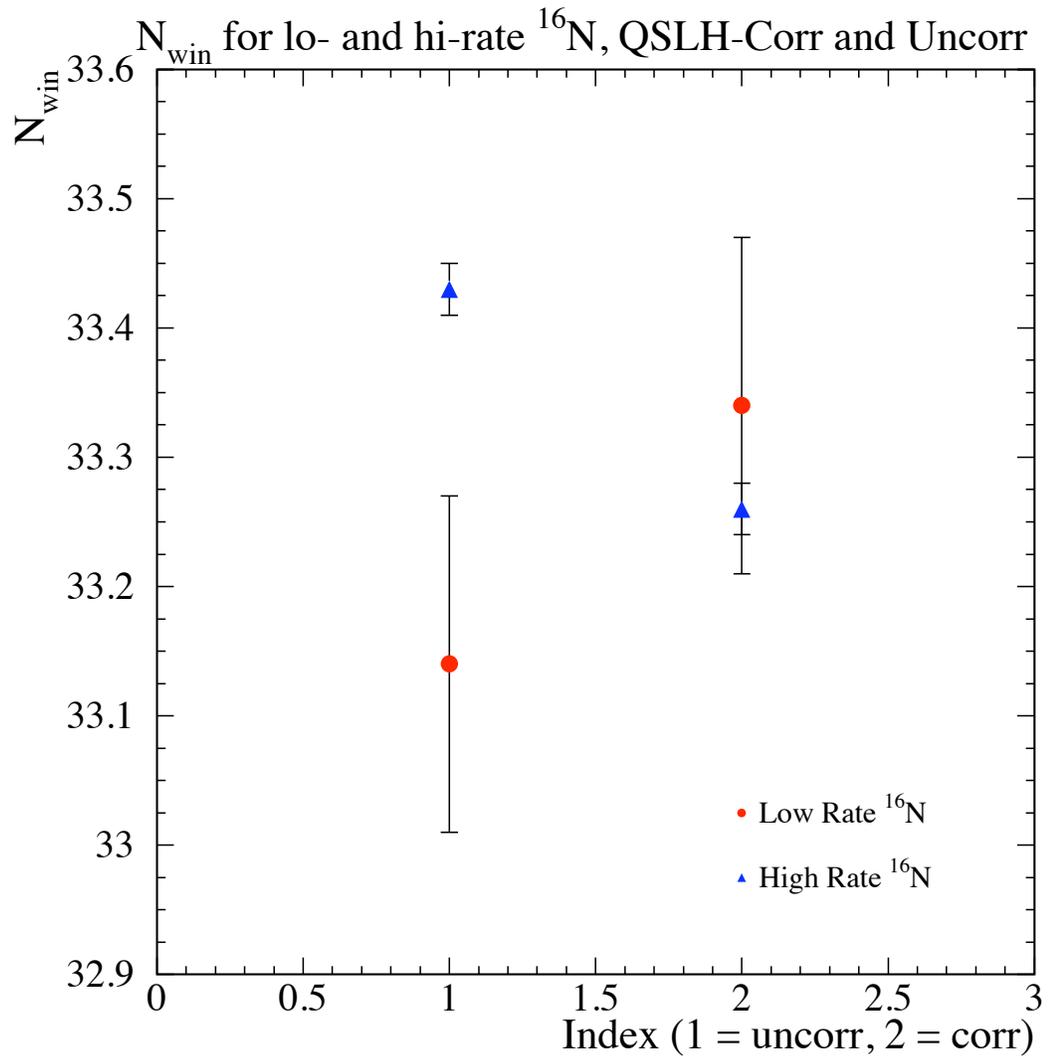
Figure 4.25: Nwin of high-rate and low-rate ^{16}N calibration data, both rate-corrected and uncorrected values

Chapter 5

The Rayleigh Power Method

5.1 A Test for Periodicity

With a QRC-corrected data set, we are now able to rely on charge-corrected times for a time-dependent periodicity analysis. There are a number of approaches one can take to search for periodicity in a data set. One can bin events in time and construct a periodogram based on the behavior of each bin's collective qualities (like average flux) or one can force the data set to remain unbinned. There are obvious benefits to requiring the data to remain unbinned (truer representation of the data) but this technique poses a challenge when dealing with unevenly sampled data. In the case of SNO, the detector had gaps in its livetime due to standard run-time issues (calibrations, maintenance of the electronics, etc.). This caused the data set to have gaps in it, due to SNO deadtime, and this must be accounted for in a time-dependent analysis like a periodicity search.

94

Initially, for the purposes of a 'low-frequency' periodicity search, SNO used two primary methods of searching for periodicity in the solar neutrino flux: a Lomb-Scargle analysis [49], which constructs a periodogram with 1-day-binned SNO data, and an unbinned maximum likelihood approach [50], as detailed in Chapter 6. A third method, called the Rayleigh power method, has also been used to search for periodicity in the SNO data. The Rayleigh analysis is an unbinned method for finding periodicity in a data set of discrete points. In this respect, the Rayleigh power method is much like the maximum likelihood approach, however the Rayleigh analysis also creates a 'true' power spectrum, similar to the Lomb-Scargle method[1]. In using a Rayleigh power approach, we perform steps similar to the previous analyses to generate power spectra describing time dependence in the SNO data.

5.2   The Rayleigh Power

The Rayleigh power is often used in a directional statistics capacity, for instance, mapping the migration of homing pigeons, or determining a favored stopping point for a roulette wheel. However, it is not limited to systems constrained to a 360° circle. It can also be used for more general treatments of periodicity, as described here.

The general approach of the Rayleigh power method for this thesis is as follows: given a periodic signal, of frequency ν, we can sample the event times

---

[1]The maximum likelihood approach does not provide a power spectrum in the 'true' sense of the word in that it doesn't strictly obey Parseval's Identity; it produces a 'significance' rather than a power

at which a signal is detected, t_i, and map these individual event times around a unit circle. The angular mapping of the circle is, of course, determined by the period we are sampling (*i.e.* 0 to 2π corresponds to 0 seconds to T seconds, with $T = \frac{1}{\nu}$). For each event occurring around the unit circle, we assign a radial vector to designate the angular position of the event, starting with $\theta = 0$ up to $\theta = 2\pi$, and allowing multiple revolutions around this circle, depending on the frequency being sampled and the length of the data set.

Each radial vector around the unit circle can be broken into its two components along the x- and y-axes: $u(i) = \cos\theta_i \hat{\mathbf{e}}_{\mathbf{x}} + \sin\theta_i \hat{\mathbf{e}}_{\mathbf{y}}$.

Summing over all individual vector events should tell us what 'direction' the resultant summed vector points towards, thus giving an angular director of the resultant signal direction:

$$U^2 = \left(\Sigma_i(\cos\theta_i)\right)^2 + \left(\Sigma_i(\sin\theta_i)\right)^2 \qquad (5.1)$$

We therefore determine the Rayleigh power, z, at this particular frequency with the formula:

$$z = U^2/N = 1/N\left(\left(\Sigma_i(\cos\theta_i)\right)^2 + \left(\Sigma_i(\sin\theta_i)\right)^2\right) \qquad (5.2)$$

where $\theta_i = 2\pi\nu t_i$.

This is essentially a measure of the nonuniformity of a signal. Given a signal with roughly isotropic distribution, one can construct a similar unit circle and find the result $U^2 \sim N$, or $z \sim 1$. For a highly nonuniform (i.e.

periodic) signal, one would expect a much higher value of Rayleigh power, with the maximum possible Rayleigh power being $z = N$. It is based on this framework that one can find the Rayleigh power for a particular system, working with the understanding that any variation from a nonuniform signal will give us a Rayleigh power of $z > 1$. To construct a power spectrum, then, we simply calculate the Rayleigh power for many 'unit circles' with different periods.

The isolated probability that the Rayleigh power, z, will be greater than any given value, K, for uniformly sampled data is [51]:

$$Probability(z > K) = e^{-K} \qquad (5.3)$$

A simple representation of the Rayleigh power method is shown in Figure 5.1, where two different periods are being sampled (only considering the sine component of the Rayleigh power here for illustrative purposes).

One can see from this rather simplistic model of the Rayleigh power how a specific frequency can 'match' a data set well (blue circle) or not at all (red circle). Here the blue circle represents a period of 18 hours, as is shown by the circumference (0 to 2π corresponds to 0 to 18 hours), and the red circle's circumference similarly represents a period of 24 hours. As can be seen in the toy model's time distribution, the sine wave corresponding to the blue circle (with a shorter period/higher frequency) better describes the time behavior of the data, and this causes the Rayleigh power to be greater for the frequency corresponding to the blue circle than that corresponding to the red circle.
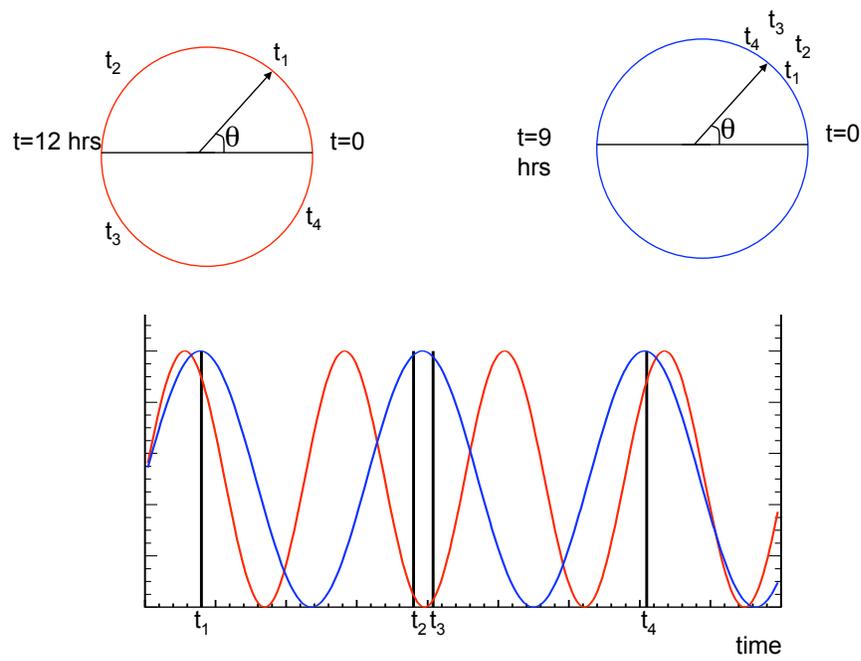
Figure 5.1: Cartoon representation of the y-component of the Rayleigh power method for two different periods, tested against the same data set of four events in time.
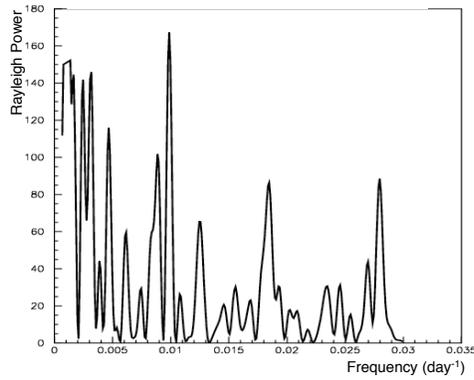
Figure 5.2: Sample Rayleigh power spectrum.

5.2.1 The Rayleigh Power Spectrum

As was just described, one can calculate the Rayleigh power z for any frequency ν of interest. The power spectrum itself is constructed simply by determining the range of frequencies of interest, and calculating the individual Rayleigh power for all values of frequencies included in this range. Depending on the range of frequencies in question, a power spectrum is then generated by collating all of the individually calculated Rayleigh powers as a function of their input frequencies. A sample Rayleigh power spectrum is shown in Figure 5.2.

5.3 The Rayleigh Power for an Unevenly Sampled Data Set

A Rayleigh analysis, at first glance, seems to have great advantage over the previously mentioned approaches used in SNO: since the Rayleigh power is no more than a simple combination of sines and cosines, the processing power required to run this type of analysis is expected to be much lower than that required for the more traditional fit-dependent methods. However, the relatively naive assumption that one will be able to simply compute the Rayleigh power, z, and determine a given Rayleigh power's significance by way of Equation 5.3 proves to be misleading and inaccurate for data sets such as SNO's, which have uneven sampling of a (presumably) continuous signal.

As it happens, for a data set which is discontinuous in time such as SNO's, it is necessary to reformulate our expectations of how to interpret the Rayleigh power's significance for a particular frequency. This is because the Rayleigh power is sensitive to periodicity, or non-uniformity in time, in a data set, be it due to the presence of a physical signal or due to the absence of events at detector 'down' times. So a data set with zero true periodic behavior could still give a (naive) statistical significance, following Equation 5.3, if the data was recorded intermittently in time.

To account for this, we must modify our expression for the probability of achieving a specific Rayleigh power in the presence of a non-uniform data set. In an attempt to build this behavior into the expected probability for different

strengths of Rayleigh power, SNO collaborator Scott Oser proposed an analytic function for the Rayleigh power distribution based simply on a random walk model[52]. The math behind the analytic form is initially straightforward. Taking each event time, with its associated angle θ_i, and breaking this into its two components, $X(=\cos\omega t)$ and $Y(=\sin\omega t)$, we can look at the event times as a random walk in two dimensions, or two simultaneous walks in X and Y.

Now, using the central limit theorem, we can assume that the distribution of these random walks in our data set is Gaussian. For each frequency, then, the data set should have a characteristic mean and variance, according to this distribution. With these parameters from each Gaussian, we can then find a functional form for the total displacement in X and Y. For a simplistic model of a purely no-signal data set, with uniform phase coverage, the equation for the functional form would simply be:

$$f(X,Y) = \frac{1}{2\pi\sigma_x\sigma_y}exp(-X^2/2N\sigma_x{}^2)exp(-Y^2/2N\sigma_y{}^2) \qquad (5.4)$$

However, for a data set such as SNO's, which has inherent 'windowing' (on/off times due to the inherent livetime/deadtime of the detector, as shown in Figure 5.3) this basic theoretical form will not be sufficient to truly model the expected deadtime-induced periodic behavior of the data. So with the intent of incorporating detector-specific properties which necessarily affect the overall phase coverage, I needed to develop a more complete version of this functional form.
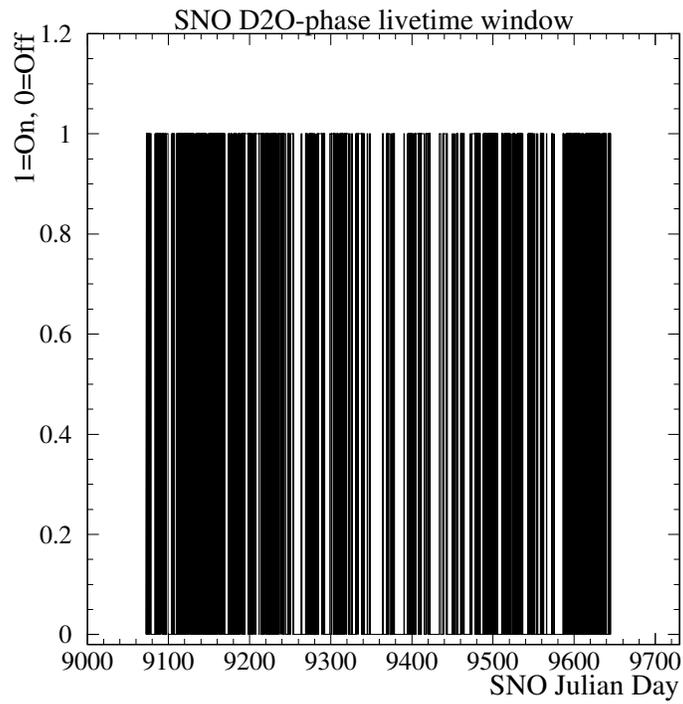
Figure 5.3: SNO livetime/deadtime window function, for the D_2O phase of data-taking. Times when SNO was running and taking data are black, the gaps are when SNO was not taking data or in non-neutrino running conditions.

5.3.1 Phase Coverage

Again, first taking into consideration the simplistic situation of uniform phase coverage, the values of our 2-D Gaussian would be straightforwardly defined as follows. Considering our two Gaussian variables, x and y, and phase $\phi = \omega t$, we define the mean, μ and the variance σ^2 as

$$\mu_x = \frac{1}{2\pi} \int_0^{2\pi} d\phi \cos\phi = 0 \tag{5.5}$$

$$\mu_y = \frac{1}{2\pi} \int_0^{2\pi} d\phi \sin\phi = 0 \tag{5.6}$$

$$\sigma_x{}^2 = \frac{1}{2\pi} \int_0^{2\pi} d\phi \cos^2\phi = \frac{1}{2} \tag{5.7}$$

$$\sigma_y{}^2 = \frac{1}{2\pi} \int_0^{2\pi} d\phi \sin^2\phi = \frac{1}{2} \tag{5.8}$$

and the covariance between x and y, for uniform phase coverage, as

$$cov(x, y) = \frac{1}{2\pi} \int_0^{2\pi} d\phi \cos\phi \sin\phi = 0 \tag{5.9}$$

In the case of non-uniform phase coverage, however, which occurs when there are breaks in the livetime of the detector (*i.e.* SNO's inherent deadtime window, shown in Figure 5.3 for the D_2O phase), we get deviations from these predicted values. Any non-uniformity in phase coverage must be accounted for in our calculation of Gaussian parameters; if a signal is not allowed to occur

103

for certain windows of time, the distribution of resultant Rayleigh vectors will almost certainly no longer have a mean of zero, for instance, and this will affect our calculation of the analytic form for the Rayleigh power.

To account for non-uniform phase coverage, we can include a phase-dependent weighting factor, $g(\phi)$, to allow for the inclusion of information about SNO's deadtime window. The Gaussian parameters then become

$$\mu_x = \frac{1}{2\pi} \int_0^{2\pi} d\phi g(\phi) \cos \phi \tag{5.10}$$

$$\mu_y = \frac{1}{2\pi} \int_0^{2\pi} d\phi g(\phi) \sin \phi \tag{5.11}$$

$$\sigma_x{}^2 = \frac{1}{2\pi} \int_0^{2\pi} d\phi g(\phi) \cos \phi - \mu_x{}^2 \tag{5.12}$$

$$\sigma_y{}^2 = \frac{1}{2\pi} \int_0^{2\pi} d\phi g(\phi) \sin \phi - \mu_y{}^2 \tag{5.13}$$

and the covariance between x and y, for non-uniform phase coverage, is now

$$cov(x, y) = \frac{1}{2\pi} \int_0^{2\pi} d\phi g(\phi)(\cos \phi - \mu_x)(\sin \phi - \mu_y) \tag{5.14}$$

with the assumption that $g(\phi)$ is normalized to unity. For SNO, an obvious implementation of this weighting scheme comes in the form of weighting by deadtime/livetime; remembering that our phase, ϕ, is simply ωt, we can now

write out our parameters and calculate them in the context of SNO's livetime. For example, the mean will now be

$$\mu_x = \frac{1}{T} \sum_{j=1}^{runs} \int_{t_{start,j}}^{t_{stop,j}} dt \cos \omega t \qquad (5.15)$$

where here, T is the total livetime, and we are summing over all runs in the data set, integrating over the start and stop times for each run. It is in this manner that we incorporate the non-uniform phase coverage of the SNO data set in the calculation of our Gaussian parameters for use in the analytic PDF for the Rayleigh analysis.

5.3.2 Calculation of the Full Analytic Form

The full functional form is built to incorporate SNO-specific variations in livetime, and therefore gaps in phase coverage. With the inclusion of the deadtime-determined 'windowing' function, the form ultimately depends on the parameters which we have calculated from the Gaussian distribution of the Rayleigh vectors: $\mu_x, \mu_y, \sigma_x, \sigma_y$, and $cov(x,y)$. Once we have these values, we can build them into the most general functional form, which is found by calculating a χ^2-type variable, and evaluating $exp(-\chi^2/2)$.

$$\chi^2(X,Y) = (X - N\mu_x, Y - N\mu_y) \, V_{xy}^{-1} \begin{pmatrix} X - N\mu_x \\ Y - N\mu_y \end{pmatrix} \qquad (5.16)$$

$$= (X - N\mu_x, Y - N\mu_y) \begin{pmatrix} N\sigma_x{}^2 & Ncov(x,y) \\ Ncov(x,y) & N\sigma_y{}^2 \end{pmatrix}^{-1} \begin{pmatrix} X - N\mu_x \\ Y - N\mu_y \end{pmatrix}$$
$$(5.17)$$

To evaluate the above equation, we must first find the inverse of the co-variance matrix:

$$V_{xy}{}^{-1} = \begin{pmatrix} N\sigma_x{}^2 & Ncov(x,y) \\ Ncov(x,y) & N\sigma_y{}^2 \end{pmatrix}^{-1}$$

$$= \frac{1}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)} \begin{pmatrix} \sigma_y{}^2 & -cov(x,y) \\ -cov(x,y) & \sigma_x{}^2 \end{pmatrix} \qquad (5.18)$$

Carrying out the multiplication:

$$(X - \mu_x, Y - N\mu_y) \begin{pmatrix} \frac{(X-N\mu_x)\sigma_y{}^2}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)} - \frac{(Y-N\mu_y)cov(x,y)}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)} \\ -\frac{(X-N\mu_x)cov(x,y)}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)} + \frac{(Y-N\mu_y)\sigma_x{}^2}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)} \end{pmatrix}$$
$$= \frac{(X - N\mu_x)^2\sigma_y{}^2}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)} - \frac{(X - N\mu_x)(Y - N\mu_y)cov(x,y)}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)}$$
$$- \frac{(X - N\mu_x)(Y - N\mu_y)cov(x,y)}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)} + \frac{(Y - N\mu_y)^2\sigma_x{}^2}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)}$$

So χ^2 is now

$$\chi^2 = \frac{(X - N\mu_x)^2\sigma_y{}^2 + (Y - N\mu_y)^2\sigma_x{}^2 - 2(X - N\mu_x)(Y - N\mu_y)cov(x,y)}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)}$$
$$(5.19)$$

Now that we have an expression for χ^2 we are able to calculate the value for the PDF. However, we would like the PDF to be in terms of the polar

coordinates z and ψ, rather than what we've been using up until now, X and Y. To do this we must make the following substitutions:

$$X = U\cos\psi,\ Y = U\sin\psi,\ z = \frac{X^2+Y^2}{N},\ \text{and}\ U^2 = Nz.$$

We can now convert our equation for χ^2 to be (first) expressed in terms of U and ψ

$$\chi^2(U,\psi) = \frac{(U\cos\psi - N\mu_x)^2\sigma_y{}^2 + (U\sin\psi - N\mu_y)^2\sigma_x{}^2}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)} \tag{5.20}$$

$$-\frac{2(U\cos\psi - N\mu_x)(U\sin\psi - N\mu_y)cov(x,y)}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)}$$

$$= \frac{U^2\cos^2\psi\sigma_y{}^2 - 2NU\cos\psi\mu_x\sigma_y{}^2 + N^2\mu_x{}^2\sigma_y{}^2 + U^2\sin^2\psi\sigma_x{}^2 - 2NU\sin\psi\mu_y\sigma_x{}^2}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)}$$
$$+ \frac{N^2\mu_y{}^2\sigma_x{}^2 - 2cov(x,y)(U^2\cos\psi\sin\psi - NU\mu_x\sin\psi - NU\mu_y\cos\psi + N^2\mu_x\mu_y)}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)}$$

then putting in terms of z and ψ:

$$\chi^2(z,\psi) = \frac{1}{N\sigma_x{}^2\sigma_y{}^2 - Ncov^2(x,y)}\left(Nz\cos^2\psi\sigma_y{}^2 - 2N\sqrt{Nz}\cos\psi\mu_x\sigma_y{}^2\right.$$

$$+ N^2\mu_x{}^2\sigma_y{}^2 + Nz\sin^2\psi\sigma_x{}^2 - 2N\sqrt{Nz}\sin\psi\mu_y\sigma_x{}^2 + N^2\mu_y{}^2\sigma_x{}^2$$

$$\left.- 2cov(x,y)(Nz\cos\psi\sin\psi - N\sqrt{Nz}\mu_x\sin\psi - N\sqrt{Nz}\mu_y\cos\psi + N^2\mu_x\mu_y\right)$$

$$= \frac{1}{N\sigma_x{}^2\sigma_y{}^2 - N cov^2(x,y)} \left((\sigma_y{}^2\cos^2\psi + \sigma_x{}^2\sin^2\psi - 2cov(x,y)\cos\psi\sin\psi))Nz \right.$$

$$- 2((\mu_x\sigma_y{}^2 + \mu_y cov(x,y))\cos\psi + (\mu_y\sigma_x{}^2 + \mu_x cov(x,y))\sin\psi)N\sqrt{Nz}$$

$$\left. + (\mu_x{}^2\sigma_y{}^2 + \mu_y{}^2\sigma_x{}^2 + \mu_x\mu_y cov(x,y))N^2 \right)$$

As a side note, we should make sure to keep track of our integrands here; when making the change of variables between (X,Y) and (U,ψ) we gain a factor of U: $dXdY \rightarrow UdUd\psi$ and upon making the change from (U,ψ) to (z,ψ) we obtain a factor of $\frac{dU}{dz}$, or $\frac{N}{2\sqrt{Nz}}$. Taking all of this together with the relation $U = Nz$, we see that the overall value reduces to N/2 for the change of variables, or $dXdY \rightarrow UdUd\psi \rightarrow \frac{N}{2}dzd\psi$.

Now, in order to get the Rayleigh power PDF we simply need to integrate $e^{(-\chi^2/2)}$, which turns out to be easier said than done, given the complicated form of our χ^2. For simplicity, we will express the exponential argument in three groups, α_1, α_2, and α_3.

$$\alpha_1(\psi) = \cos^2\psi\sigma_y{}^2 + \sin^2\psi\sigma_x{}^2 - 2\cos\psi\sin\psi cov(x,y)$$

$$\alpha_2(\psi) = -2(\cos\psi\mu_x\sigma_y{}^2 + \sin\psi\mu_y\sigma_x{}^2 + (\mu_x\sin\psi + \mu_y\cos\psi)cov(x,y))$$

$$\alpha_3(\psi) = \mu_x{}^2\sigma_y{}^2 + \mu_y{}^2\sigma_x{}^2 + \mu_x\mu_y cov(x,y)$$

And the PDF for the Rayleigh power becomes

$$f(z)dz = \frac{1}{C}dz \int e^{-(\alpha_1(\psi)/2)Nz - (\alpha_2(\psi)/2)N\sqrt{Nz} - (\alpha_3(\psi)/2)N^2} d\psi \qquad (5.21)$$

where C is a normalization constant.

To obtain the actual distribution for each frequency, we integrate the above function with respect to ψ. Since this is obviously not a simple feat, we rely on numerical integration, finding values for $\mu_x, \mu_y, \sigma_x, \sigma_y$, and $cov(x,y)$ for each frequency sampled. We can then feed these values to the final integral to get the full Rayleigh power probability distribution for each frequency in our data set. So in effect, what started out as an analytic 'fit' is no longer fitting *any* parameter aside from a normalization constant. This is a very strong point to be made: the analytic form is now a prediction of what the Rayleigh power distribution should be, and as will soon be shown, it is a remarkably accurate prediction.

5.4 Aside: Simplified Analytic Form

Before proceeding, I should mention that the initial attempts at fitting the Rayleigh power distribution with the analytic form did not employ the full form as described above. There is an elegant simplification of Equation 5.21, suggested by SNO collaborator Scott Oser, which relies on two assumptions. The first assumption is that there should be no difference between the variance in X and the variance in Y, or $\sigma_x = \sigma_y$. In theory, this seems to be a fair assumption since we would not expect a bias of the variance in either direction. The second assumption postulated that there should also be no covariance between x and y, or $cov(x,y) = 0$. If these two (presumably reasonable) assumptions are made, the analytic form simplifies considerably, and there is

no longer a need for numerical integration, etc. As has been displayed in [52], the form of the function for this simplified case is

$$f(z, \psi)dzd\psi = \frac{1}{C}exp(\frac{-z}{2\sigma^2})exp(\frac{-N\mu^2}{2\sigma^2})exp(\frac{\mu\sqrt{Nz}}{\sigma^2}\cos\psi)dzd\psi \qquad (5.22)$$

This version of the form actually would in theory work as a fit: once integrated (w.r.t. $d\psi$) it can be expressed in the very general form

$$f(z)dz = z^a exp(b + cz + d\sqrt{z})dz \qquad (5.23)$$

Obviously this formula would make things much easier to work with, but unfortunately the assumptions that led to this form don't actually work for the SNO data set. An example of the simplified fit, applied to SNO MC, is shown in Figure 5.4:

Though the covariance between x and y is *approximately* zero, and the variances are *approximately* equal, these assumptions obviously don't strictly hold. Only one demonstration of the poorness of fit is shown here; it should be noted that this is among the 'best' of the fits using the simplified analytic form. For a study which ultimately requires strict agreement between the analytic form and actual data distributions (for confidence level generation, etc., as will be seen in later sections), this is enough to make the simplified analytic fit useless for our purposes. It therefore turns out that the full analytic form is necessary to get reasonable agreement between predictions and the actual data.
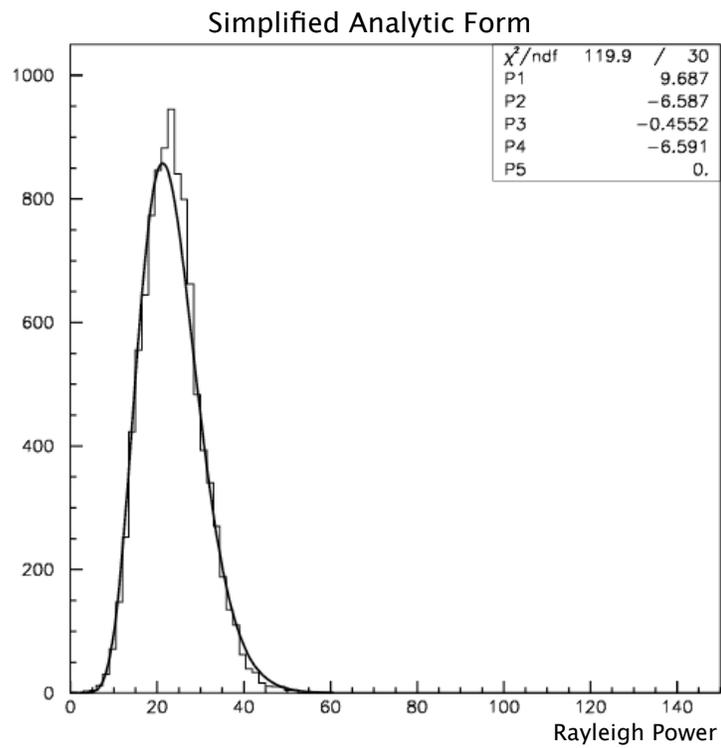
Figure 5.4: Example of Monte Carlo comparison with simplified analytic form.

5.5 Generation of Rayleigh Power Monte Carlo

In order to model and understand the behavior of the Rayleigh power for different frequencies, a first step should be to generate sample power spectra with no signal present other than experimental deadtime. For SNO, this is only possible with the creation of a deadtime window function, which incorporates all of the times that SNO was operating and taking data, as well as the times when SNO was operating but not taking neutrino data (calibration runs, maintenance runs, etc.), or shut down altogether (INCO strikes, power outages, etc.). The window function is simply a binary 'on/off' function which indicates as a function of time when SNO was or was not in neutrino data-taking mode. An example of a SNO window function, taking into account detector deadtimes and livetimes for the D_2O phase is shown in Figure 5.3.

The Monte Carlo data sets are generated by randomly throwing events in time, and only allowing events to occupy those times that are designated 'acceptable' by the SNO deadtime window function. With this accept-and-reject method, we ultimately are left with a data set that has randomly generated event times, only occurring during calendar times which coincide with actual SNO run times. Aside from the built-in signal of the deadtime window function, the SNO no-signal Monte Carlo is thus completely random, and should represent a truly signal-free SNO data set.

Once we have generated Monte Carlo data sets for our analyses, we can evaluate them just as we evaluate data, and run the Rayleigh power analysis on the Monte Carlo. By having fake data sets with absolutely no signal

present other than the underlying SNO deadtime structure, we are equipped with a versatile means of understanding and accounting for the behavior of this inherent periodicity of SNO. A sample SNO Monte Carlo Rayleigh power spectrum is shown in Figure 5.5.

5.6   Confidence Levels

With the existence of a predictive analytic form for the distribution of 'null-hypothesis' Rayleigh powers at all sampled frequencies, we are now armed with the ability to evaluate the statistical significance of the Rayleigh power spectrum for SNO data. Because the analytic form predicts the behavior of the Rayleigh power in the presence of zero external signal, it enables us to generate confidence levels for Rayleigh powers at any sampled frequency. The method of determining confidence levels for an individual frequency is as follows.

Assuming the analytic PDF correctly describes the expected distribution of null-hypothesis Rayleigh powers for a given frequency, we can determine where on this probability curve a Rayleigh power, $z(f)$, describing the periodicity of data at a frequency, f, will sit. For any particular frequency f, we can then find the fractional confidence level of this 'detected' Rayleigh power, $z(f)$. We do this by integrating over the entire probability curve to find a normalization factor, or norm, and then integrating the curve with the detected Rayleigh power, z, as our upper limit, to find the fraction of the PDF which lies below this particular Rayleigh power. We therefore find the confidence
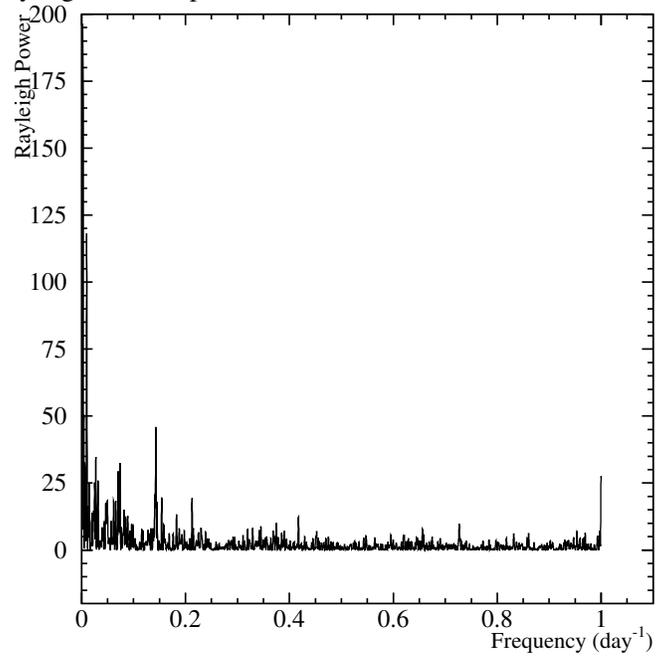
Figure 5.5: Sample combined-phase (D$_2$O and Salt phases) no-signal Monte Carlo Rayleigh power spectrum. The built-in structure which is present in the power spectrum is due to inherent periodicities due to deadtime ('windowing') in SNO's data set.

level by comparing the two results, such that for any given frequency,

$$CL(f) = \frac{\int_0^{z(f)}(PDF)dz}{\int (PDF)dz} \qquad (5.24)$$

If $z(f)$ is found to be on the higher end of the predicted distribution, we will have a higher confidence level, and similarly if $z(f)$ is on the lower end of the predicted distribution, we should end up with a low confidence level. In this way, we are able to decipher which of the many peaks in a data set's Rayleigh power spectrum are truly significant (*i.e.* not statistically likely to be a random fluctuation), and which of the peaks are in agreement with expectations for any particular frequency's built-in structure.

5.7   The Trials Penalty

Now that we have a fully descriptive analytic form with which to model the Rayleigh power behavior for different frequencies, we can determine the statistical significance of a power for individual frequencies through individual peaks' confidence levels. However, in determining the significance of Rayleigh power peaks for the power spectrum as a whole, we run into a problem.

There is obviously a difference between the confidence level for all frequencies sampled and the confidence level for one single frequency. For example, the probability of seeing one peak with confidence level greater than 90% for *one* frequency is much lower than the probability of seeing one peak with confidence level greater than 90% in *any* of the frequencies sampled, due simply

to the increased number of trials, or measurements, being done. This is commonly referred to as the trials penalty, in that the statistical significance of a result must incorporate the number of trials, or rolls of the dice, which led to that result. (For example: roll a pair of dice once, get double sixes, and you're slightly impressed. Roll the pair of dice 100 times, and get double sixes only 1 time out of 100, and you're much more impressed).

Because of the trials problem, in order to determine the false alarm probability for an entire data set (the likelihood that a 'discovery' at a particular CL threshold are just statistical deviations from the norm), we must include in the CL calculation the number of independent frequencies which we are scanning in our Rayleigh power analysis. Here the false alarm probability is defined as being the chance that any given peak at a certain Rayleigh power strength is a false alarm, or $FAP = 1 - CL$.

In order to extend the false alarm probability for one frequency to the entire sampled set (all frequencies) we simply take the complement of the product of seeing a peak above CL for any one frequency, or: $FAP_{bin} = 1 - CL_{bin}$ becomes $FAP_{total} = 1 - (CL_{bin})^f$, where f is the number of independent frequencies we are sampling (an empirically-determined value for SNO's periodicity studies). As is evident in this expression for FAP_{total}, the overall false alarm probability is a much lower value than the FAP for one sampled frequency. If, for instance, we want to find the probability that any peak be above the 90% confidence level, (or in other words, the confidence level for a data set spanning 9,998 frequencies is 90%), then the corresponding confidence level for each frequency

is actually 99.962% (and similarly, the FAP for each sampled frequency would now be 0.038% rather than 10%).

We can easily see that to attain the level of precision needed to assign confidence levels at the 99% and 3σ level, one would need to produce tremendous amounts of zero-periodicity Monte Carlo data to determine what the actual frequency-specific confidence levels are. Because of this trials problem, the Rayleigh power test goes from being an unusually fast and straightforward method of analyzing periodic signals to being a very cumbersome, CPU-heavy analysis due to the potential need for production of hundreds of thousands of Monte Carlo runs.

By correctly modeling the behavior of Rayleigh powers for all sampled frequencies in a data set, one can avoid relying strictly on null-hypothesis Monte Carlo to get around the trials problem. We will still need a high degree of accuracy to be able to determine trials-weighted frequency-specific confidence levels, but having a predictive analytic form for each frequency that we sample greatly reduces the amount of time (both calendar time and CPU time) needed to produce a true description of what occurs at the tails of our Rayleigh power distributions.

5.8 Results of Full Analytic Form on Monte Carlo

In order to test the basic premise that our random-walk-inspired analytic form correctly models a data set with deadtimes in it, I constructed a set of Monte Carlo with a very straightforward 'window function' built in, which models reasonable, simplistic on/off times of 6 days with incoming signal, 1 day without (a weekly periodicity). The Monte Carlo described here is loosely based on SNO's calendar livetime, in that its length has been determined by the length of the D_2O data set (with start and stop times dictated by the first and last dates of data taking during the D_2O phase), however, the actual weekly structure of the window function involved here has no information from SNO (and therefore no additional inherent signal).

5.8.1 Results

The first test of how well the analytic form describes our data set is a comparison between the predicted distribution (full analytic form) and the actual distribution as given by Monte Carlo. The purpose of running this check is to verify that for frequencies which have no signal, the analytic form accurately predicts Rayleigh power behavior in that frequency. Similarly, we can test any deviation from the analytic form for the frequency where a signal is present. We would expect no deviation at the signal-specific frequency, since the inherent frequency in the data set is fed, via the window function, to the analytic

formula. So in order to check agreement between the analytic prediction and the actual Rayleigh power distributions, we can simply look at comparisons, frequency to frequency, as are shown in the following figures.

These results are quite encouraging; the agreement between the actual Monte Carlo Rayleigh powers and the analytic form predictions is remarkable. With this ammunition, we can proceed with confidence to a more complicated SNO-based Monte Carlo and analytic form.
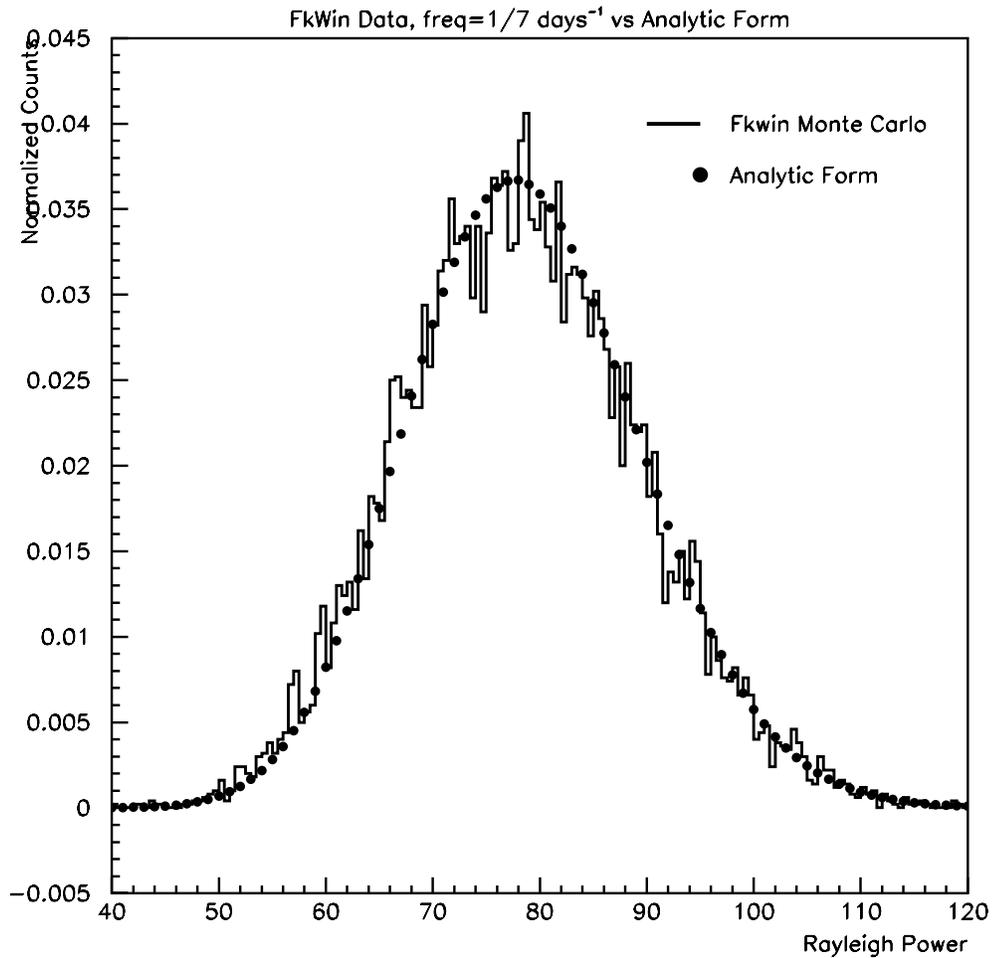
Figure 5.6: Fake Window Monte Carlo comparison with analytic form. Frequency bin 1426, corresponding to the fake window signal frequency of 0.14287/day, or 1/7 days. As expected, there is good agreement between predicted Rayleigh power behavior for this frequency and actual Monte Carlo Rayleigh power distribution
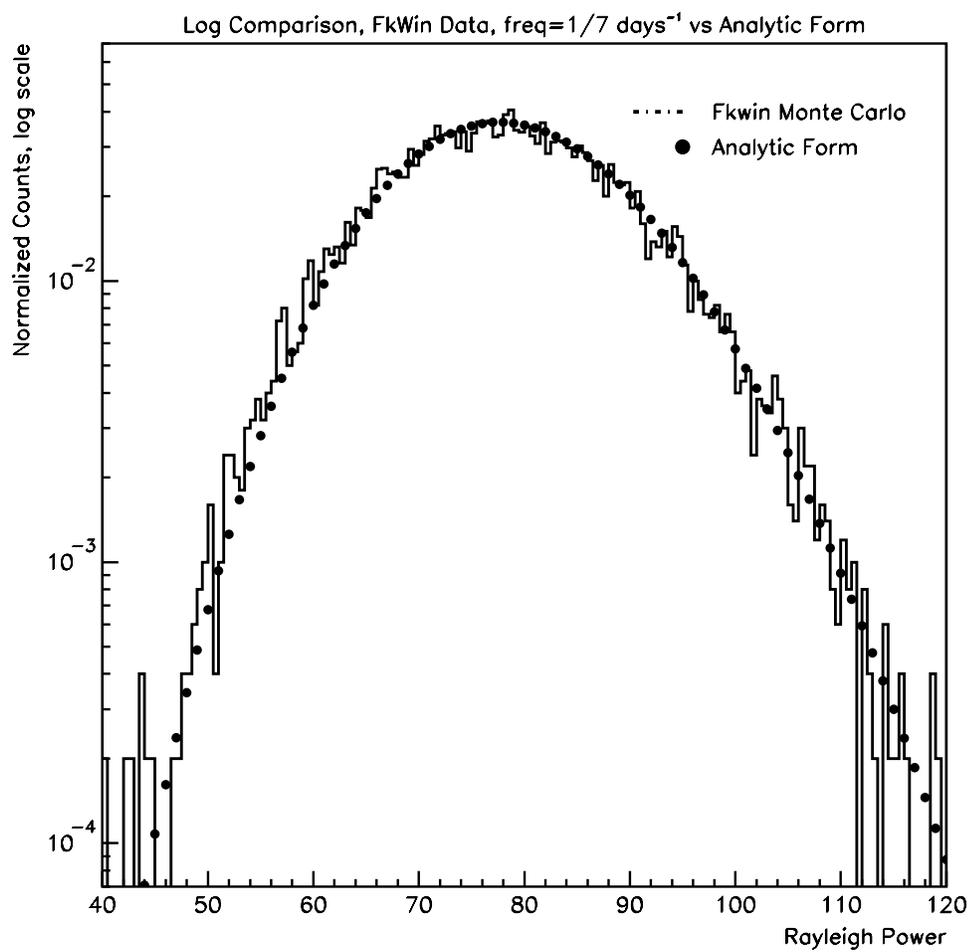
Figure 5.7: Fake Window Monte Carlo comparison with analytic form, log scale. Frequency bin 1426, corresponding to the fake window signal frequency of 0.14287/day, or 1/7 days.
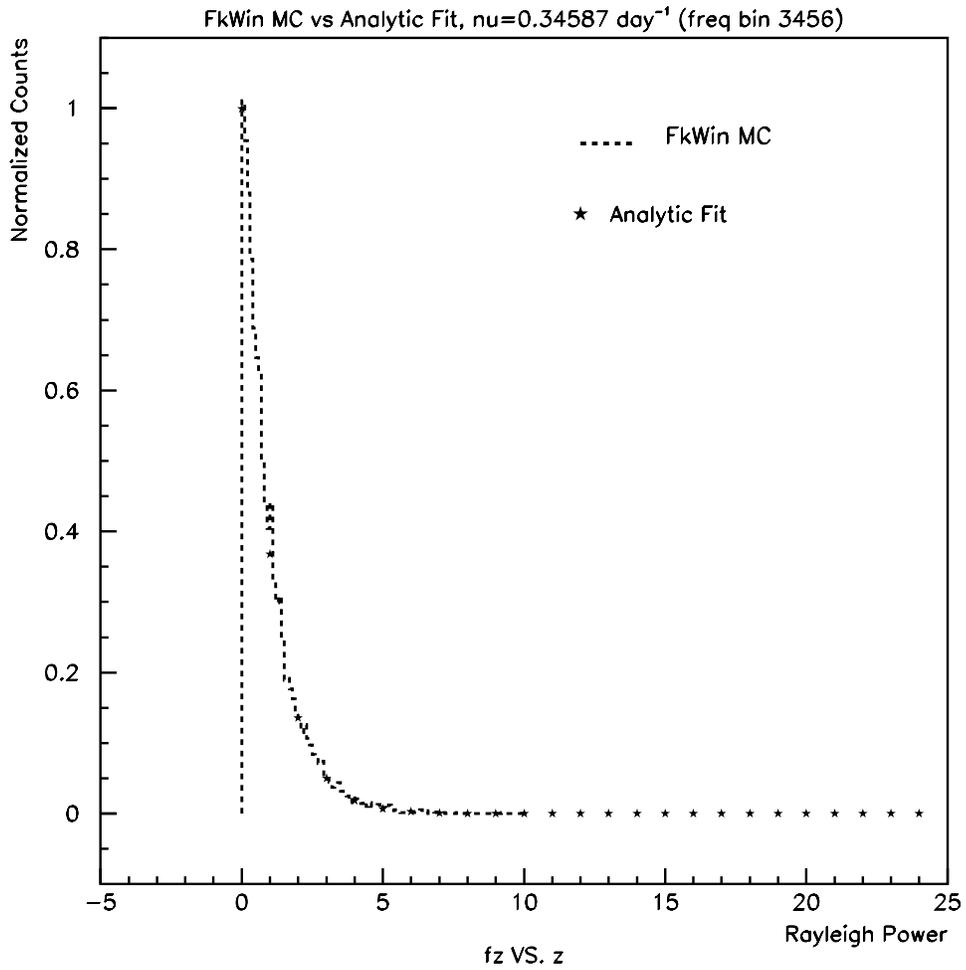
Figure 5.8: Fake Window Monte Carlo comparison with analytic form. Frequency bin 3456, corresponding to frequency of 0.34587/day
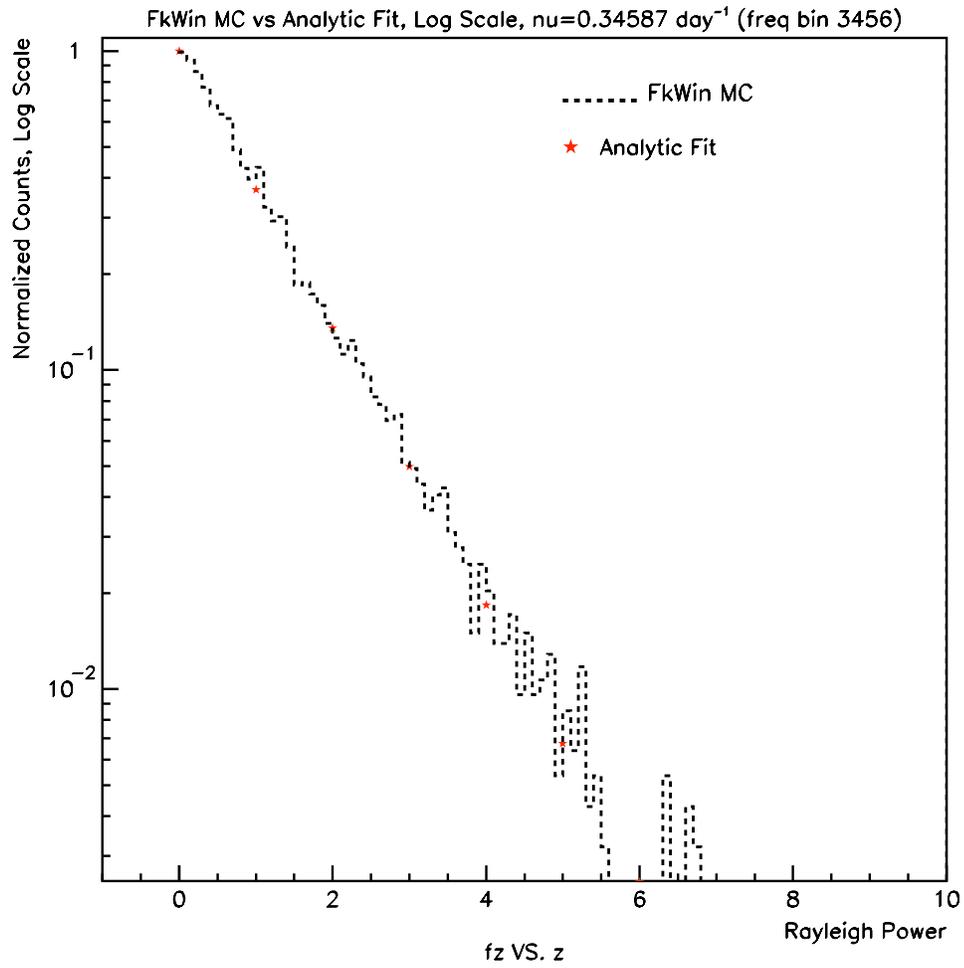
Figure 5.9: Fake Window Monte Carlo comparison with analytic form, log scale. Frequency bin 3456, corresponding to frequency of 0.34587/day

Chapter 6

Low-Frequency Periodicity Search

As has been mentioned in Section 2.1, there have been claims of detection of a periodicity in the solar neutrino flux [25], [26], [27], [28], [29], [30], [31], as well as claims to the contrary [32], [33]. As a critical contribution to the periodicity argument, SNO performed our own search for low-frequency periodicity in the solar neutrino flux, and found no statistically significant signal [8]. A Lomb-Scargle periodogram was constructed from the binned SNO data for both the D_2O and Salt phases of data-taking. An unbinned approach was used as well, which employed the maximum likelihood method to construct a 'likelihood spectrum' of the data. With these two methods, and a 'cross-check' unbinned analysis using the Rayleigh power method, SNO was able to survey the region of interest with enough independence between analyses to ensure an unbiased outcome.

6.1 Low-Frequency Rayleigh Power Methods

Although the published low-frequency periodicity analysis was based primarily on the two methods discussed in Section 2.2, the idea of using the Rayleigh power test as a cross-check for these two methods was proposed and undertaken in the hopes of having a third analysis for verification purposes (the methods used in a Rayleigh power analysis have been detailed in Chapter 5). However, an initial analysis of standard SNO Monte Carlo sets resulted in Rayleigh power spectra with a great deal of built-in structure. As was discovered after some time, this structure is due to the 'windowing' of the SNO data set, which was described in Section 5.3. The presence of windowing structure in the Rayleigh power spectrum ultimately led us to develop the analytic form and analysis which has been outlined in Chapter 5, and recognize the potential applications of the Rayleigh power to new frequency regimes, which will be covered in Chapter 7. Though the work required to extract the correct formula for the Rayleigh power distribution was more involved than had been anticipated for a 'cursory' cross-check analysis, the application of the Rayleigh power method in the low-frequency regime gave us a platform for testing our methods before embarking on the far trickier high-frequency analysis.

6.1.1 Creation of SNO Low-Frequency No-Signal Monte Carlo

As has been detailed in Section 5.5, SNO Monte Carlo are generated using an accept-and-reject method, based on the SNO deadtime window. For the low-frequency search, several thousands of null-hypothesis Monte Carlo were generated, with no signal present, to address the statistical significance of any peaks found in the data's Rayleigh power spectrum. This method of statistical interpretation will be detailed in the following sections.

6.2 Analytic form for Low-Frequency Search

The analytic form described in the previous chapter (Chapter 5) is now to be used in the context of SNO for the low-frequency analysis, specifically, by incorporating the SNO deadtimes in the integration of the analytic PDF. As was detailed in Chapter 5, the Rayleigh power analytic form is based on the Gaussian parameters describing the 'random walk' of events. By structuring our Gaussian parameters with the input of the SNO deadtime window function, as shown in Equation 5.15, we build an analytic form with the foundation of the inherent SNO low-frequency deadtime. In doing this, we expect that the analytic form will have the exact same behavior for any sampled frequency as a collection of no-signal SNO-based Monte Carlo data sets, only if the predictive form that we are using has truly incorporated the necessary phase restrictions.

In the following figures, Figures 6.1- 6.4, I show the distributions of thou-

sands of no-signal SNO Monte Carlo for various sampled frequencies, and the corresponding analytic PDF superimposed on top of these distributions. It is remarkable that even with the inclusion of all of SNO's built-in, detector-specific periodicities (as given by the low-frequency deadtime window function), we get very good agreement between the full analytic form and the Rayleigh power distributions for the Monte Carlo in different frequency bins.

6.3 Confidence Levels

From the figures in the preceding section (Section 6.2) we can easily see that the analytic form serves as an adequate prediction of the Rayleigh power behavior for different frequency bins. This allows us to interpret these predictions in the form of confidence levels for each frequency bin, as previously described in Section 5.6. Though we have analytically determined a formula for the calculation of confidence levels which includes a fix for the trials problem, this formula depends on the number of independent frequencies, f, which we have not been able to make a prediction for. For the low-frequency analysis, we have relied on an empirical value for f, determined by checking our confidence level predictions against an actual Monte Carlo-based ('brute force') count of how many fake data sets pass a confidence level cut versus how many we expect to pass given a trial value of f. This number turns out to be roughly 275 (markedly different from the 9,998 we started off with!),and with this information we can effectively predict the confidence levels for the D_2O-phase data set in the given frequency range. The trials-weighted confidence levels
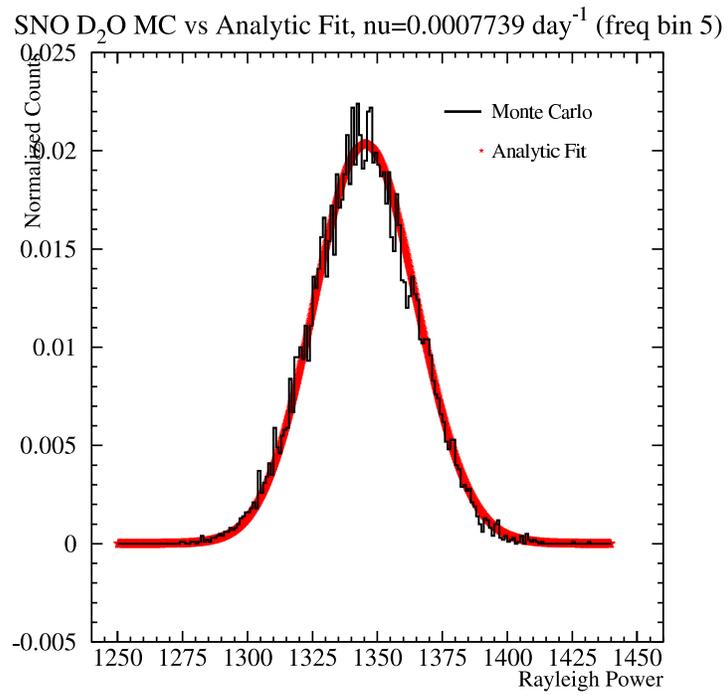
Figure 6.1: SNO Monte Carlo comparison with analytic form. Frequency bin 7, corresponding to frequency of 0.00097/day, or 11.2 nHz.
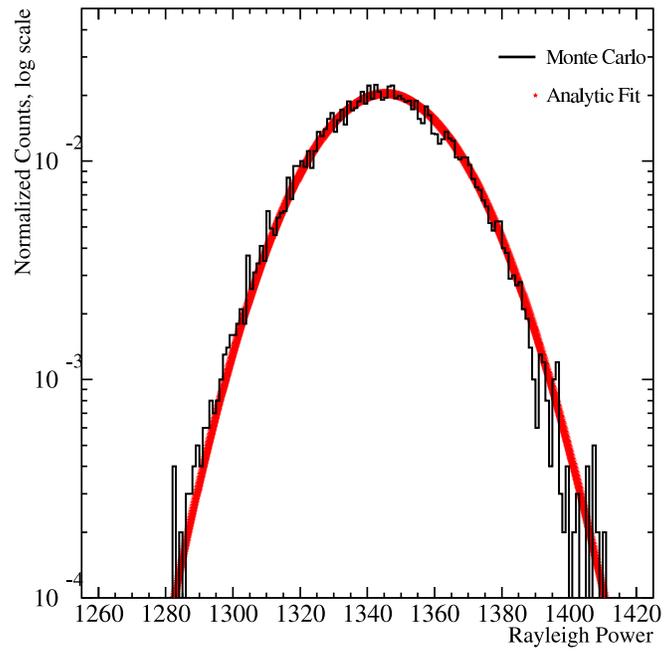
Figure 6.2: SNO Monte Carlo comparison with analytic form, log scale. Frequency bin 7, corresponding to frequency of 0.00097/day, or 11.2 nHz
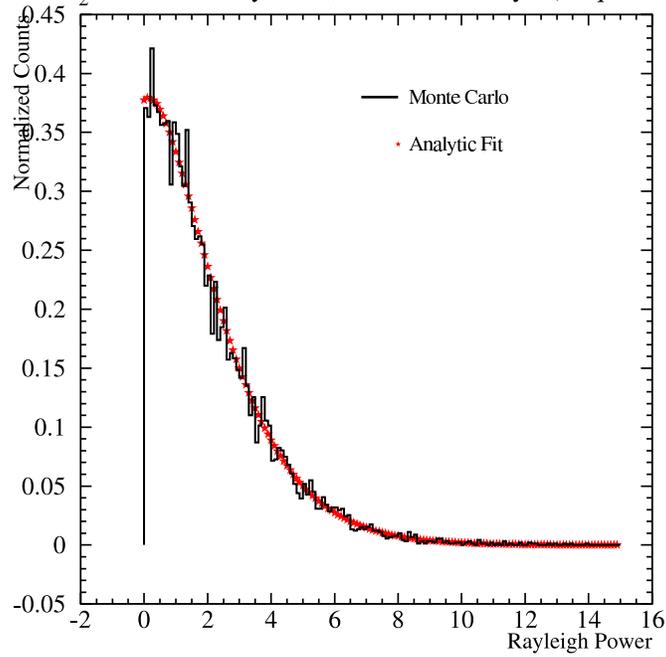
Figure 6.3: SNO Monte Carlo comparison with analytic form. Frequency bin 2537, corresponding to frequency of 0.25397/day, or 2.94 μHz
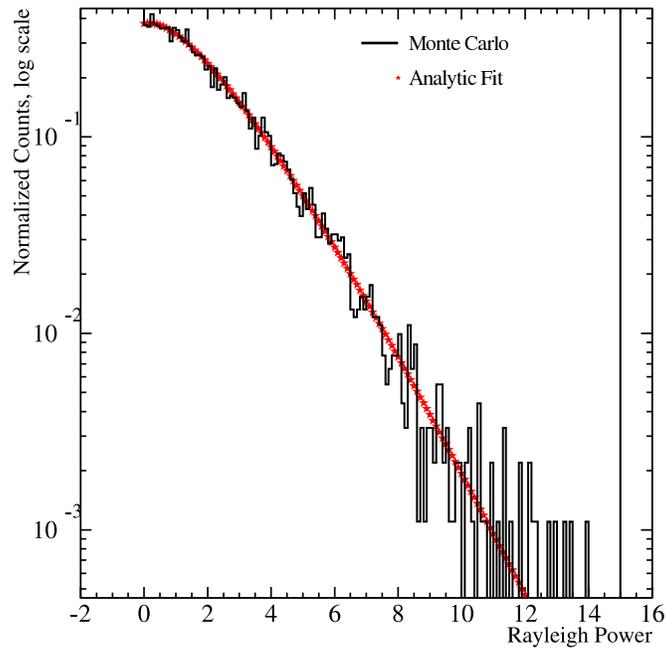
Figure 6.4: SNO Monte Carlo comparison with analytic form, log scale. Frequency bin 2537, corresponding to frequency of 0.25397/day, or 2.94 μHz

for the D_2O phase are shown in Figure 6.5

6.4   Low-Frequency Rayleigh Power Results

Now that we have created the tools to interpret the Rayleigh power spectrum, in light of the inherent SNO deadtime structure which is present at all sampled frequencies, we can use this information to perform a retroactive check of the Rayleigh power values found for the SNO data set in the D_2O phase. By comparing the trials-weighted confidence levels of all frequencies to the SNO D_2O-phase Rayleigh power spectrum, we can attempt to verify the results of the other two methods used in the low-freqeuncy periodicity search, to ensure that we do indeed see no significant periodic signal. Not surprisingly, we see no peaks above a 90% confidence level, and this is in good agreement with our previously published results[8].
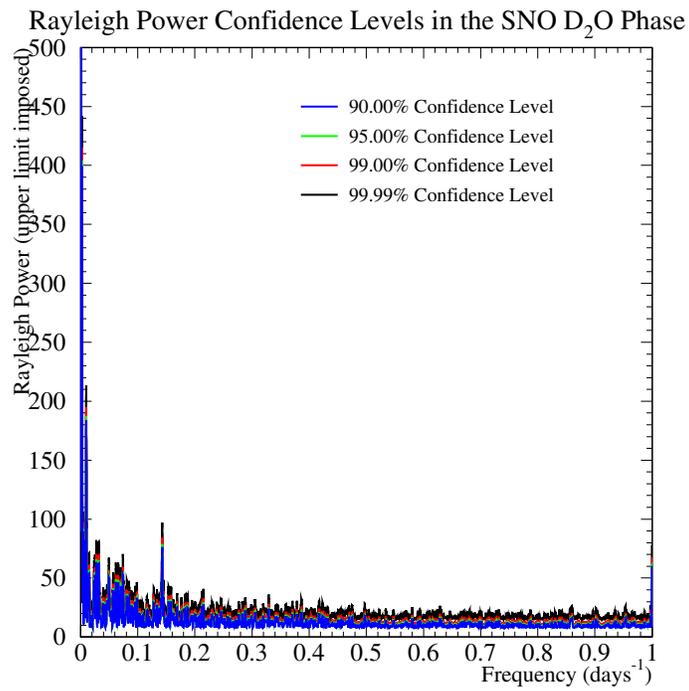
Figure 6.5: Confidence levels for Rayleigh power in the SNO D$_2$O phase.
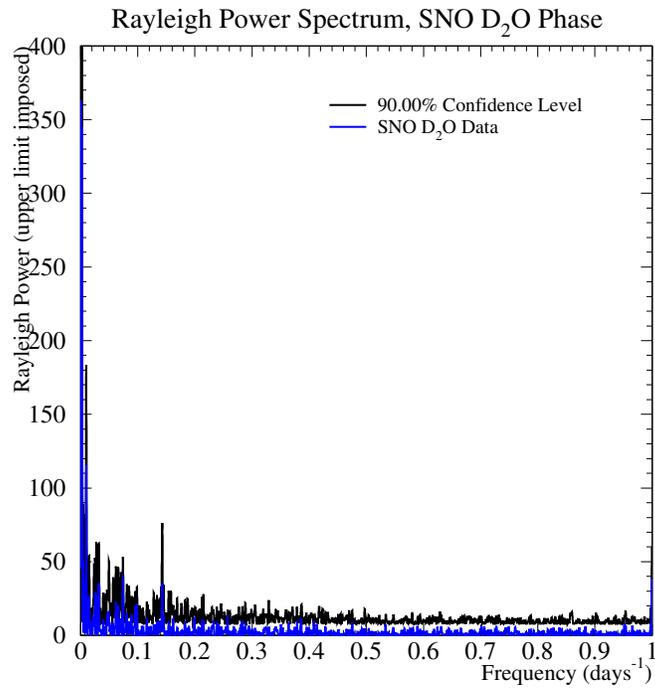
Figure 6.6: SNO D$_2$O-phase data with 90% confidence level imposed. It is clear from the power spectrum that there are no peaks in the D$_2$O-phase data which surpass the 90% confidence level, therefore the Rayleigh power analysis is in agreement with the other two analyses used by SNO for the low-frequency periodicity search.

Chapter 7

High-Frequency Periodicity Search

In this chapter I will detail the markedly different approach of the Rayleigh analysis as it is used for the 'high'-frequency regime, ranging over frequencies of 1/day to 144/day (or 1/10 min). The high-frequency periodicity search has been largely motivated by the possibility of observing effects on the solar neutrino flux by gravity-mode oscillations at the solar core. Due to the magnitude of the region to be searched, the need for minimal backgrounds, and the amount of precision needed to carry out a statistically significant interpretation of the results, this task has not been previously attempted by any other experiment.

7.1 Motivation for a High-Frequency Periodicity Search

For a search of periodicity in the solar neutrino flux at low frequencies, three separate analyses have already been carried out, and a periodicity paper has been published [8], as was discussed in Chapter 6. The idea for a separate periodicity analysis at a higher frequency range (> 1/day) was proposed with the motivation that a search at these frequencies might introduce the possibility of detecting predicted gravity-mode, or g-mode, oscillations' effects on the solar neutrino flux. The solar oscillation modes have been introduced in Section 2.3. The most relevant modes of oscillation for high-frequency studies of the solar neutrino flux are expected to be the g-mode oscillations. These oscillations are inherently difficult to detect on Earth, largely because they are confined to the inner region of the sun, in contrast to the more well-understood pressure, or 'p-mode', oscillations which are present as acoustic waves both in the interior and outer envelope of the sun. However, this property is precisely what makes the g-mode search interesting to SNO, since the solar core is the region in which neutrino production occurs. If g-mode oscillations significantly affect the electron densities or chemical abundances at the solar core, there exists a possibility of seeing the oscillations' effects on the solar neutrino flux by way of the MSW effect, and possibly (though highly doubtful according to the heliosmic community) by way of small fluctuations in the core temperature. Regardless of the predicted mechanism, the Rayleigh power opens

up possibilities of detecting higher-frequency regions due to its fast processing time.

As they are confined to the interior of the sun (they are evanescent in the convection zone), g-mode oscillations have proven to be elusive subjects to study with helioseismology instruments dedicated to observing effects from g-modes at the solar surface. Nevertheless, the GOLF experiment, aboard the SOHO satellite, has undertaken the study of g-modes via Doppler velocity measurements at the solar surface [53], and has made claims for having detected signatures in the power spectrum density (PSD), presumably caused by g-modes [54] [55]. More detail about the specific detected signal will be covered in Chapter 9, but in light of claims from GOLF after over a decade of running and analysis, it appears that g-modes are beginning to enter into the realm of detection.

7.2 SNO at High Frequency

7.2.1 Deadtime Window

As has been described in Chapters 5 and 6, the Rayleigh power suffers from a dependence on inherent deadtime in the data set being considered. One of the main considerations, therefore, in going from a low-frequency Rayleigh power analysis to a high-frequency analysis is the deadtime window function to be used in determining the expected Rayleigh power distribution for different frequencies. For the low-frequency Rayleigh power analysis, we explained the

need for the deadtime window function, which is essentially a binary on/off assignment that specifies the exact calendar times when SNO would be online and able to detect a signal, *i.e.* the detector is on and in 'neutrino' running (not calibration or 'junk' runs), as well as the times when SNO would be offline or unable to detect events, *i.e.* the detector is off, or calibrations are occurring.

For the high-frequency analysis, we again have to create a deadtime window, but we now are including not only low-frequency deadtimes (on the order of days, weeks, and months), but also all the deadtimes which are on the order of the periods which we would like to sample (seconds, minutes and hours). These higher-frequency deadtimes are a result of the 'data cleaning' process SNO uses to rid our data set of non-neutrino backgrounds, as outlined in Chapter 1, and designates short windows of time (on the order of msec and sec) which follow typical background-like events in an effort to clean the data of such events, such as muons, muon followers, malfunctioning PMTs, etc. It should be noted that our high-frequency window function is built to model periods of deadtime, and as such the most relevant information for this study is not the *size* of the gaps in the window so much as the length of time *between* the gaps.

The high-frequency deadtime window was generated by scanning through the D_2O and Salt phase neutrino data sets, and recording all the times (as measured by the 10MHz GPS-synchronized clock) when specific bits of the data-cleaning, or background-removal, mask were flagged. For this study, any short time-scale cuts which were performed on the data set and resulted in

the reduction of deadtime of the data set, be it on the order of milliseconds or minutes, were included in the creation of the high-frequency deadtime window. The specific cuts which affect the deadtime window are

1. Muon cut: designed to tag muon events in the detector

2. Muon Follower Short cut: tags all events following a muon within 20 sec

3. Nhit Burst: designed to cut bursts - defined as 6 events above 40 NHIT (NHIT = Number of PMTs HIT) within 4 sec

4. Missed Muon Follower Short cut: designed to tag neutron followers from atmospheric neutrinos, muon events that were missed by the muon tag, and all other processes that result in neutrons in SNO without a visible precursor. Removes any event within 250 msec of an event with NHIT ≥ 60.

The deadtime window function therefore includes the gaps in phase coverage for the Rayleigh power analysis due to the various background removal cuts which result in a msec-scale or longer deadtime, in addition to the lower-frequency livetime gaps. For each 'gap' in the SNO livetime due to a cut, run boundary, etc., there is a corresponding gap in the livetime window (thus creating the 'deadtime' window function). By including this information in our generation of SNO Monte Carlo files, we are able to model the substructure which is detected with the Rayleigh power test.

7.2.2   Challenges of a High Frequency Search

The trials penalty has already been described in some detail in Chapter 5, and is a significant impediment to a large-scale analysis due to the copious amounts of Monte Carlo which must be generated to offset it. The larger the number of frequencies being sampled, the (presumably) greater the number of independent frequencies which must be accounted for. With this comes the necessity for a greater precision for trials-weighted confidence levels, and therefore the greater the number of Monte Carlo which must be run. This was a frustration for the low-frequency periodicity search, but for the high-frequency search it was initially seen as a deal-breaker; the number of frequencies which must be scanned for a conclusive 'high-frequency' search is formidable in comparison to the already unwieldy number used in the low-frequency analysis. Because the motivation for the search is largely based on g-mode oscillations, which are at the moment very poorly understood in our sun, the range which must be covered in terms of frequency is quite broad, as there is a great deal of uncertainty about the particular frequencies of interest regarding solar g-modes.

Not only must the frequency range for the high-frequency search be a great deal wider than the range for the low-frequency search, we also must use adequately small spacing between frequencies so as to ensure that we do not miss a signal by sampling at too-large intervals. This will be discussed shortly.

In spite of these clear challenges, the Rayleigh test still serves as the most promising approach for such an analysis, as the computing power and time required to evaluate a high-frequency Rayleigh power spectrum still outper-

form other 'fit-heavy' methods such as the previously mentioned Maximum Likelihood and Lomb-Scargle approaches [1].

7.2.3   Unique Abilities of SNO

Having enumerated the challenges facing SNO in this search (and really, any experiment directing its efforts to this frequency regime), we should also pay special attention to the unique abilities held by SNO for such an analysis. Since SNO has such inherently low backgrounds, and effective background-rejection methods, we have a great deal of confidence that events in our neutrino data set are accurately representing solar neutrino behavior. In other words, if we were to find periodicity in our data set, we can be very confident that it is a periodicity in solar neutrino flux rather than a periodicity introduced by contamination of radioactive backgrounds.

Also, SNO's real-time detection of events allows us to have a direct representation of any temporal signal of solar neutrinos. The precision of our 10MHz clock ($\pm 1ns$) is such that for the purposes of a high-frequency periodicity study, there is no appreciable uncertainty on the relative or absolute event times recorded in the SNO detector.

---

[1] For comparison, back-of-the-envelope estimates demonstrate that the 'un-optimized' Rayleigh analysis code is faster than the Maximum Likelihood method by more than three orders of magnitude. Optimizing the code further improved the runtime by another two orders of magnitude.

7.3   High-Frequency Rayleigh Power

As has been detailed in previous chapters, the Rayleigh analysis is an unbinned method for finding periodicity in a data set of discrete points. The general approach for constructing a Rayleigh power is to sample a signal by recording event times, t_i, and use the relation $\theta_i = 2\pi\nu t_i$ to construct the Rayleigh power of any sampled frequency, where the Rayleigh power, z, is (again) defined as $1/N((\Sigma_i(\cos\theta_i))^2 + (\Sigma_i(\sin\theta_i))^2)$.

Though the expected behavior of the Rayleigh power for any given frequency in the presence of zero signal would be exponential, we have found that with the deadtime substructure of SNO, the behavior of Rayleigh powers at all sampled frequencies varies tremendously, from Gaussian-like to exponential-like (and everything in between). Examples of this are shown in the following plots (Figures 7.1- 7.3); the Rayleigh powers are calculated for hundreds of Monte Carlo files (which draw their events from a distribution which is limited by the deadtime window alone), and the distribution of Rayleigh powers for two separate frequencies are shown.

Because of this markedly non-exponential behavior of the Rayleigh power for different sampled frequencies, we must determine exactly what signal the Rayleigh power test measures when there is no physics signal present in a data set, or in other words, determine what 'baseline' is present in the SNO data set. In a similar way as was carried out for the low-frequency search, we can do this by generating the SNO high-frequency deadtime window, and the resulting SNO Monte Carlo and analytic PDF, both of which are dependent

142

Rayleigh Power Distribution for Combined Phase Monte Carlo
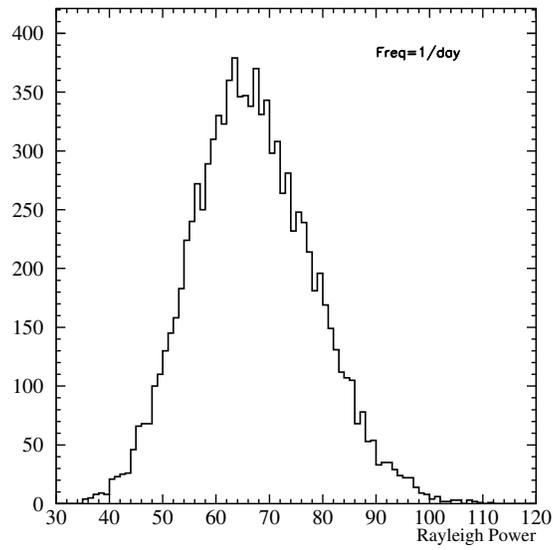
Figure 7.1: Sample distribution of Rayleigh powers for signal-free combined-phase Monte Carlo with sampled frequency = 1.0 day^{-1}.
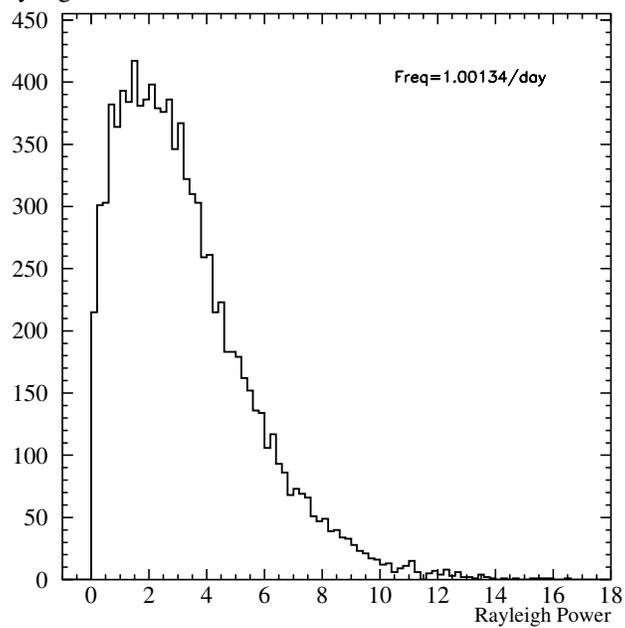
Figure 7.2: Sample distribution of Rayleigh powers for signal-free combined-phase Monte Carlo with sampled frequency = $1.00134 \ day^{-1}$.
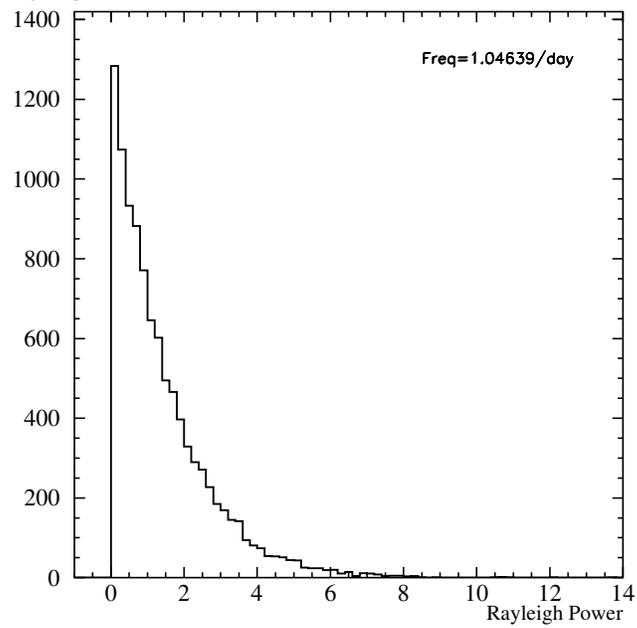
Figure 7.3: Sample distribution of Rayleigh powers for signal-free combined-phase Monte Carlo with sampled frequency = 1.04639 day^{-1}.

on the registered SNO deadtime.

7.3.1 Creation of High-Frequency SNO Periodicity Monte Carlo

With the high-frequency deadtime window function already created, the generation of SNO Monte Carlo for the high frequency regime is fairly straightforward. For purposes of evaluating the confidence levels in the high frequency regime, we must of course have access to null-hypothesis, or zero-signal Monte Carlo data sets. In addition to the zero-signal Monte Carlo sets, we also must create Monte Carlo with a periodic signal built into the data set, with the intent of determining the sensitivities at which the Rayleigh power analysis is capable of detecting such a signal. The methods of constructing both types of Monte Carlo are outlined here.

As was the case in the low-frequency Rayleigh analysis, I have generated all Monte Carlo data sets using an 'accept-and-reject' method, in which event times are randomly generated, and only those occurring within imposed 'deadtime' restrictions (as per the SNO deadtime window function) are accepted in the final event list.

For 'signal-free' Monte Carlo files, the window function is the only restriction present for event acceptance versus event rejection. For the combined-phase analysis, I generate the Monte Carlo separately for the D_2O and Salt phases, and then concatenate the two files to form one 'combined-phase' Monte Carlo. This allows us to correctly model the rates and deadtime of the SNO

detector in the two separate phases.

The method for generating the 'signal' SNO Monte Carlo files is identical to that for the signal-free Monte Carlo files, except that a sinusoidal wave is added on to the (previously flat) window function, therefore making events more likely to occur during times when the sine wave's amplitude is positive, and less likely to occur during times when the sine wave's amplitude is negative (relative to the signal-free baseline). To include a sinusoidal signal in the combined-phase Monte Carlo, one must pay special attention to the phase of the sinusoidal wave being incorporated; since the combined-phase Monte Carlo are generated separately, one must force the starting point for the sine wave in the Salt phase to be the same as that in the D_2O phase so that there is no disagreement in the phase of the sine wave for a combined-phase Monte Carlo file. We can choose at will the input frequencies and input sinusoidal amplitudes which we would like to study, and can therefore simulate a wide range of simulated signal strengths to analyze with the Rayleigh power test.

7.3.2   Creation of Rayleigh Power Files from Monte Carlo

Once the Monte Carlo data sets are created, both with and without built-in periodic signals, we can perform a Rayleigh power analysis on them, as described in Section 5.2. Sample power spectra of fake data both without a signal and with a signal are shown in Figure 7.4 and 7.5.

In Figure 7.5, a periodic signal with an amplitude of 14% modulation has been put into the Monte Carlo at a frequency of 8.63 day^{-1}, or 99.2 μHz. The
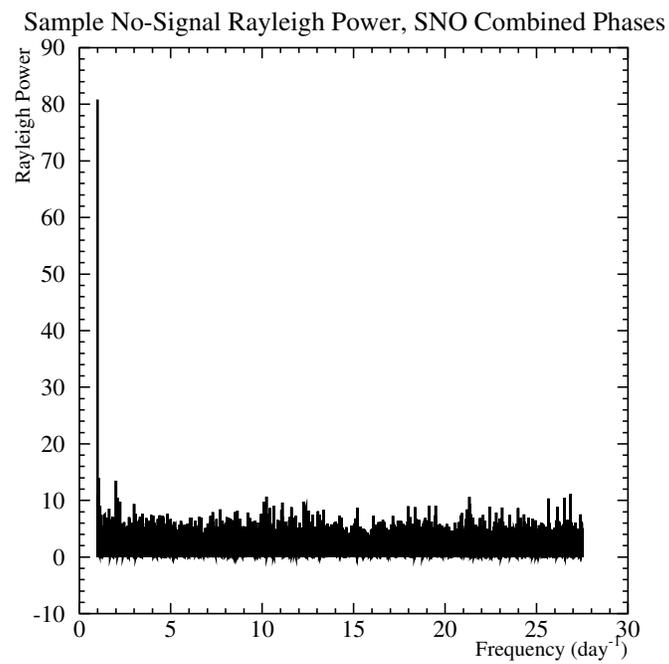
Figure 7.4: Power spectrum of combined-phase Monte Carlo with no input periodic signal.
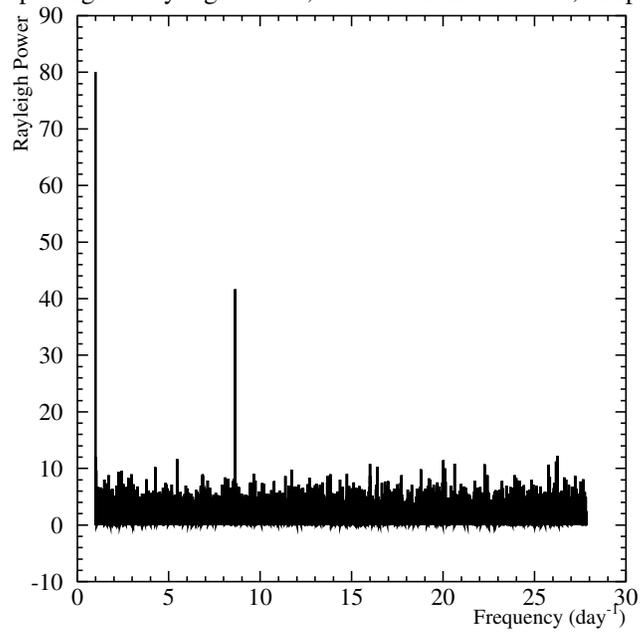
Figure 7.5: Rayleigh power spectrum of combined-phase Monte Carlo with a 14%-amplitude periodic signal at frequency=8.633 day^{-1}, or 99.2 μHz.

presence of this peak is clearly discernible in the power spectrum of the Monte Carlo. The presence of multiple peaks is also visible in Figure 7.4, the power spectrum of the Monte Carlo which has no signal built into it, but includes the inherent SNO deadtime window function. One can see the clear periodic behavior of the SNO detector here. Interesting to note is the obvious peak at 1 day^{-1}, or 11.57 μHz; this is indicative of a periodicity in the SNO deadtime on the order of once per day, which could be due to any number of SNO activities. Nightly blasting, causing a gap in livetime (by default, SNO does not include data taken at times when blasting could trigger a false burst in the data set), could be to blame. The practice of starting a new run daily at 7:00 am upon shift changeover could be to blame; any 'blip' in livetime, if periodic, would be detected here. This is an interesting peak to consider, but for the purposes of our study, the cause is of no consequence; since the power spectra for Monte Carlo with no input signal show this same feature, we are assured that this peak is due to SNO's deadtime-specific periodicities, and not a physics signal.

7.3.3   Frequency Range for High Frequency Search

As the main motivation for performing a search for high frequency variations in the solar neutrino flux comes from the expected g-mode oscillation time scales, we have a guiding hand to help us determine what frequency range in which to search. We also would like to ensure that the Rayleigh power test would be capable of detecting high frequency signals. Though the Rayleigh power test was used successfully for the low frequency periodicity search in

SNO, we wanted to assure ourselves that the test would behave similarly in a high frequency regime, with periods roughly approaching the time scales of SNO's background removal cuts in the data set. Stated plainly, we wanted to determine if there was a maximum frequency which the Rayleigh power was capable of detecting in SNO. To ensure that we were not overlooking any limitations in our methods, I performed a scaling test to verify that the Rayleigh power test retained its sensitivity for signal detection at very high frequencies. Shown in Figures 7.6 and 7.7 are power spectra for increasingly high frequencies. I created Monte Carlo data sets with built-in signals ranging from $1\ day^{-1}$ to $10sec^{-1}$ (far higher than the predicted g-mode frequencies), with a signal amplitude of 100% of the non-periodic allowed amplitude. As is clear from the strengths detected with the power spectra, signals with high frequencies in this regime are not limited by the Rayleigh power analysis. Since we have ruled out a built-in maximum detectable frequency, we have chosen our upper limit based loosely on the safe inclusion of the highest limits for g-mode oscillation frequencies (roughly $43.2\ day^{-1}$, or $500\ \mu$Hz [3]).

7.3.4   Frequency Sampling Size

The spacing used between sampled frequencies for the Rayleigh power spectra is not arbitrarily chosen. The specific spacing used for this analysis was chosen with the intention of maximizing oversampling so as to not risk missing a potential signal, and minimizing redundancy so as to avoid demanding excessive job processing times. This perfect balance between being too coarsely
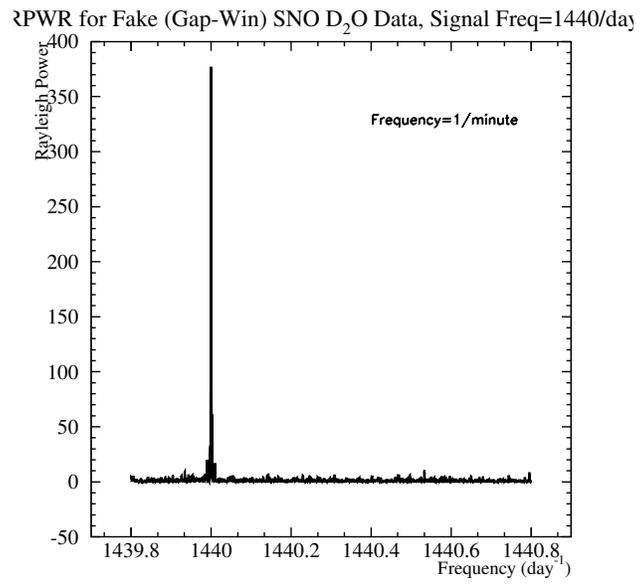
Figure 7.6: Rayleigh power spectrum of D_2O Monte Carlo with input frequency=1/minute, with signal amplitude=100%.
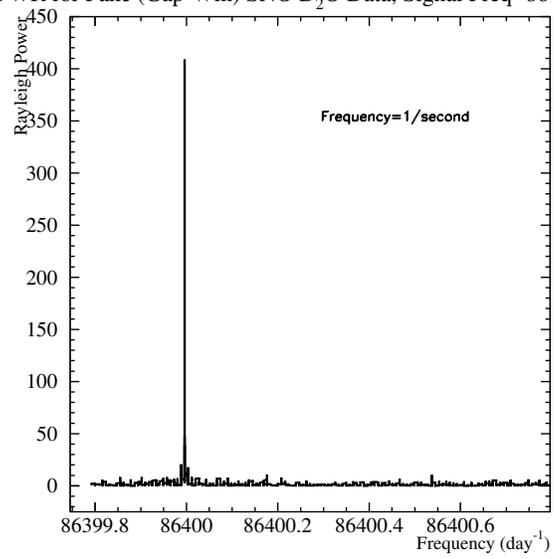
Figure 7.7: Rayleigh power spectrum of D_2O Monte Carlo with input frequency=1/second, with signal amplitude=100%.

sampled and too finely sampled would ideally be determined by the number of independent frequencies in the data set. This is not a number which is readily extracted from the SNO data, since there is a surprisingly large number of unpredicted correlations between different frequencies, as has been seen in the low frequency periodicity search, described in Chapter 6. However, since the high frequency studies span a much larger region in frequency space (more than 100 times larger than the low frequency region), and as such demand a large amount of CPU time, we are not afforded the luxury of running with excessively fine sampling by default, as was the case with the low frequency periodicity search. We must determine the appropriate number of frequencies to sample before performing the search.

While trying to determine if there is an optimal frequency separation, $\delta\nu$, for the high frequency periodicity search, we primarily focused on the impact of a few overarching questions:

1. Will the separation of events in time in the data set be a determining factor; if the separation is larger than the inverse of our chosen $\delta\nu$ do we run into trouble?

2. Is there an absolute maximum component frequency which might restrict our allowed sampling frequency value (possibly analogous to the Nyquist rate)?

3. Should the length in time of the data set have any impact on the smallest allowed spacing in frequency?

The first idea, that the smallest absolute spacing between two events in our data set might affect our sampling ability, turns out to be misleading. As in many other analyses, since the Rayleigh power test takes a data set and 'wraps' it many times around the unit circle, the absolute time difference between two events turns out to be irrelevant, and only the final spacing in phase is of consequence. A toy demonstration of this is if the time of the first event, t_1, occurs at time 09:20:45 on day 1, and the second event, t_2, occurs at time 09:20:46 on day 2 (24 hours and one second later), it can easily be seen that these two events will sit near one another in phase if the period (or the circumference of the Rayleigh power unit circle) corresponds to $\text{T}(0:2\pi) = 1\text{hr}$, for instance, which is large compared to their relative separation of 1 second (or alternately, $2\pi n + 1$ second, where $n = 24$).

The second item taken into consideration was the possibility that we should treat the SNO data as a continuous-time signal, in which case we would be able to infer from the maximum component frequency what the sampling frequency should be, analogous to the Nyquist rate for a system. As was demonstrated above, in Section 7.3.3, there is no detected maximum frequency for the Rayleigh analysis. This is related to the first item in the above list, since the highest possible frequency which we would have sensitivity to would largely depend on the spacing of all of the events in the data set in terms of relative phase, not on the spacing of the events in the standard 'absolute' time. The presence of a 'cutoff' frequency for the Rayleigh analysis was not observed. Additionally, and perhaps more importantly, the treatment of the SNO data

set as a continuous system is likely not the correct way to formulate the problem. Since SNO data is really just discrete samples of a presumably continuous signal, it is clear in retrospect that this method would not have been perfectly applicable even in the presence of a maximum frequency.

The third approach we explored to determine optimal frequency separation for our high frequency studies was focused solely on the length of time of our data set. This is illustrated in Figure 7.8; two sine waves which have only slightly different frequencies of oscillation will begin propagation in a highly correlated state. These two sine waves will be expected to become fully uncorrelated only after some length of time, T, which will depend on the separation of frequencies, $\delta\nu$, between the two oscillations. If the length of time over which we allow two sine (or cosine) waves, separated in frequency by $\delta\nu$, to oscillate (the length of the data set) is short compared to the length of time it would take for the two waves to become completely uncorrelated, then we would expect strong correlations between neighboring frequencies. However, as the length of the data set approaches the length of time it would take for two sine (or cosine) waves to decohere, the weaker the correlations will become.

Since the Rayleigh power is simply a combination of sines and cosines, we would expect to be able to see evidence of this relationship in the correlations between frequencies and the length of the data set which we are analyzing. We essentially want to know if there exists a 'beat' frequency for two Rayleigh powers separated in frequency by $\delta\nu$, as this will tell us at what spacing be-
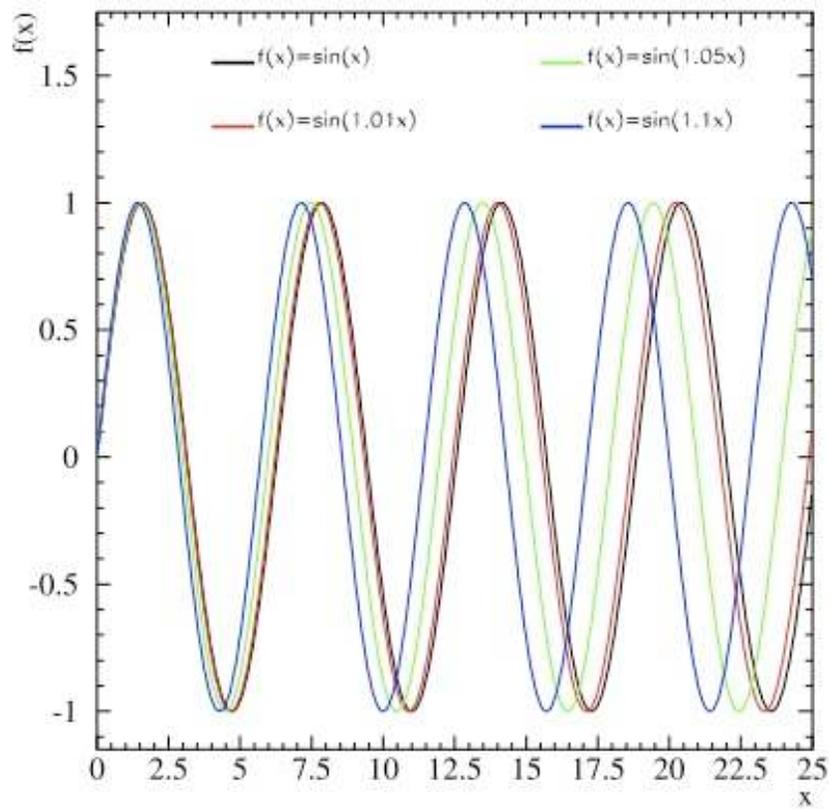
Figure 7.8: Example of the behavior of multiple sine waves sampling neigh-boring frequencies.

tween two frequencies we must be able to live with.

To test this theory, I generated two fake data sets; one spanning a length of calendar time of four days, and another spanning a length of calendar time of 40 days. In Figures 7.9 through 7.12, I show the results of applying the Rayleigh power analysis to these two fake data sets, with the same separation in frequency, $\delta\nu$, for both analyses. It is clear that the data set with shorter calendar time has much larger correlations between neighboring frequencies, as was expected with our 'beat' hypothesis. As the length of time which the data set spans is increased (from 4 to 40 days), the amount of correlation between neighboring frequencies decreases.

This is an encouraging result, and it prompts a more quantitative approach to solving the problem of what separation of frequencies is necessary for there to be zero correlation between one frequency, ν, and its neighbor, $\nu + \delta\nu$ given a certain length data set, T. Looking at this another way, if we assume the presence of 'beats' between two Rayleigh powers over some period of time, T, we can determine the minimum spacing between frequencies necessary for the beat wave to complete one cycle in this time period. We would like to find the $\delta\nu$ necessary such that the calendar length of our SNO data set allows for one beat of two sine (cosine) waves separated by $\delta\nu$, *i.e.* the frequency separation necessary for two Rayleigh powers of neighboring frequencies to be completely uncorrelated over the length of time that SNO is running.

It should be emphasized here that what is important is not the SNO-specific deadtime window (including cuts, etc.), but the total length of calendar time

Figure 7.9: 4-day data set. Correlations between frequencies for Rayleigh power spectrum of Monte Carlo with input frequency=11.99 day^{-1}.
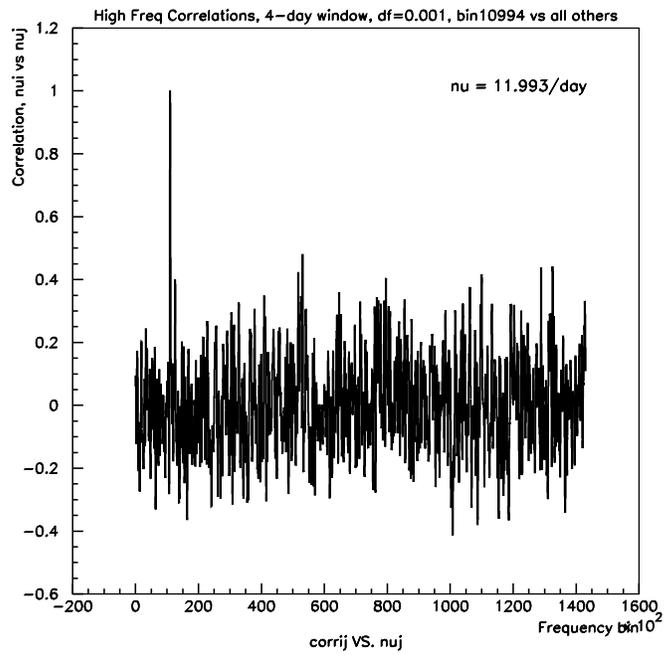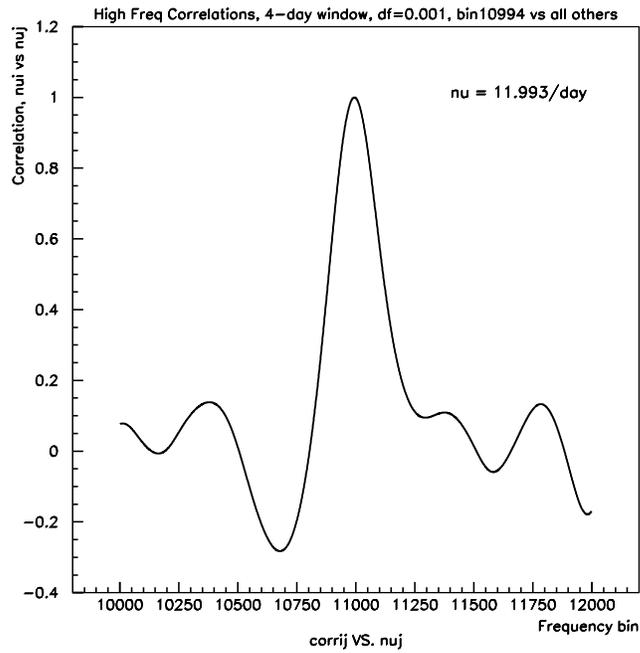
Figure 7.10: 4-day data set. Zoomed view, correlations between frequencies for Rayleigh power spectrum of MC with input freq=11.99 day^{-1}.
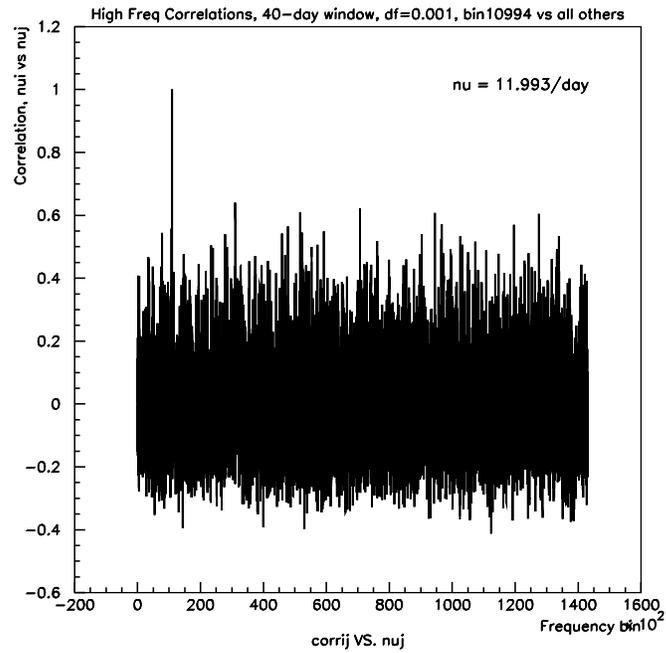
Figure 7.11: 40-day data set. Correlations between frequencies for Rayleigh power spectrum of Monte Carlo with input frequency=11.99 day^{-1}.
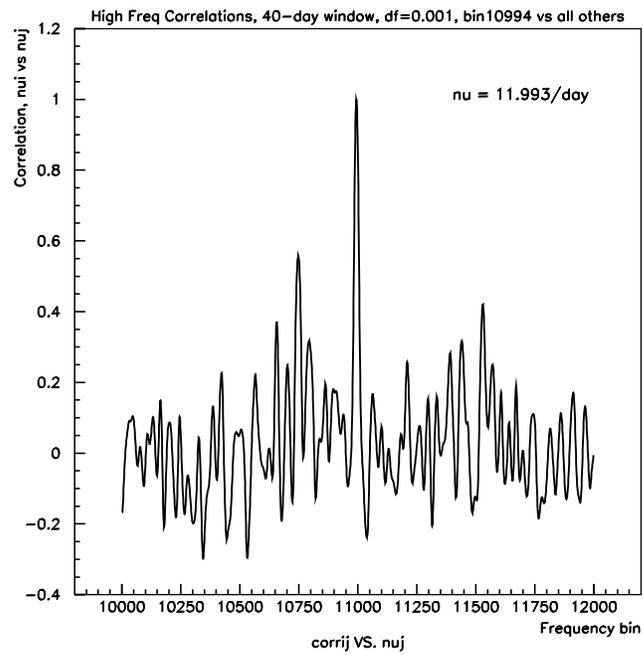
Figure 7.12: 40-day data set. Zoomed view, correlations between frequencies for Rayleigh power spectrum of MC with input freq=11.99 day^{-1}.

over which two neighboring sine waves would have to evolve to be able to fully de-correlate. As the periodic structure of the sine waves will not be changed by the gaps in SNO's detection livetime, we are not required to include that information in our evaluation; what is important is the large-scale sampling time of SNO.

We can represent the degree of correlation by the dot product, where maximally correlated neighbors' sine waves will have a dot product of 1, and completely uncorrelated (orthogonal in phase) sine or cosine waves will give a dot product of 0. With this approach, we can determine what frequency separation is necessary for the products of two neighboring sine waves to integrate to zero over a certain calendar time, T.

By using the integral representation of the dot product between two sine waves and two cosine waves, we develop an approximate equation for the expected de-correlation for two neighboring Rayleigh powers;

$$\int_0^T \sin(2\pi\nu t)\sin(2\pi(\nu + \delta\nu)t_i)dt + \int_0^T \cos(2\pi\nu t)\cos(2\pi(\nu + \delta\nu)t_i)dt = 0$$

(7.1)

Simplifying, we end up with

$$\sin(2\pi\delta\nu t)|_{t_0}^T = 0$$

(7.2)

which gives us the separation that we are searching for, $\delta\nu$. Since t_0, the start of data-taking, and T, the end of data-taking, are determined for us by SNO's calendar running time, we can now very easily find the particular value which

163

is necessary for our specific run-time requirements. For the combined-phase SNO data set, that value is: $\delta\nu = 0.00035817 day^{-1}$, given $T = 1396$ days, with $t_0 = 0$. For the D_2O data set, which spans only 574 days, we find the value necessary for no correlations between frequencies is $\delta\nu_{D_2O} = 0.000871 day^{-1}$. This shows the behavior which we would expect; for a shorter length data set, larger spacing between frequencies is required to avoid correlations. Another way of demonstrating this is shown in Figure 7.13; using the specific value for $\delta\nu$ which we have just calculated (here, for instance, looking at the D_2O phase), we plot the transition from maximal correlation to minimal correlation, which occurs (as expected) after the length of time over which SNO ran (for the D_2O phase). Again: we are showing that by using a specific separation in frequency, $\delta\nu$, for the argument of two neighboring sine waves, the length of time over which the two sine waves become completely de-correlated is exactly the calendar length of time over which SNO took data.

This is no more than a consistency check; the time necessary for two sine waves of slightly different frequencies to evolve to show zero correlation with one another should only equal the length of the data set if we have chosen the correct value for the separation between the neighboring frequencies. As is evidenced in Figure 7.13, this is indeed the case.

Since the value which we have found, $\delta\nu$, is the specific separation in frequency which is necessary to *avoid* correlations between neighboring frequencies, we know the *maximum* acceptable spacing for our search (*i.e.* the *minimum* number of frequencies we can inspect over the range which we have
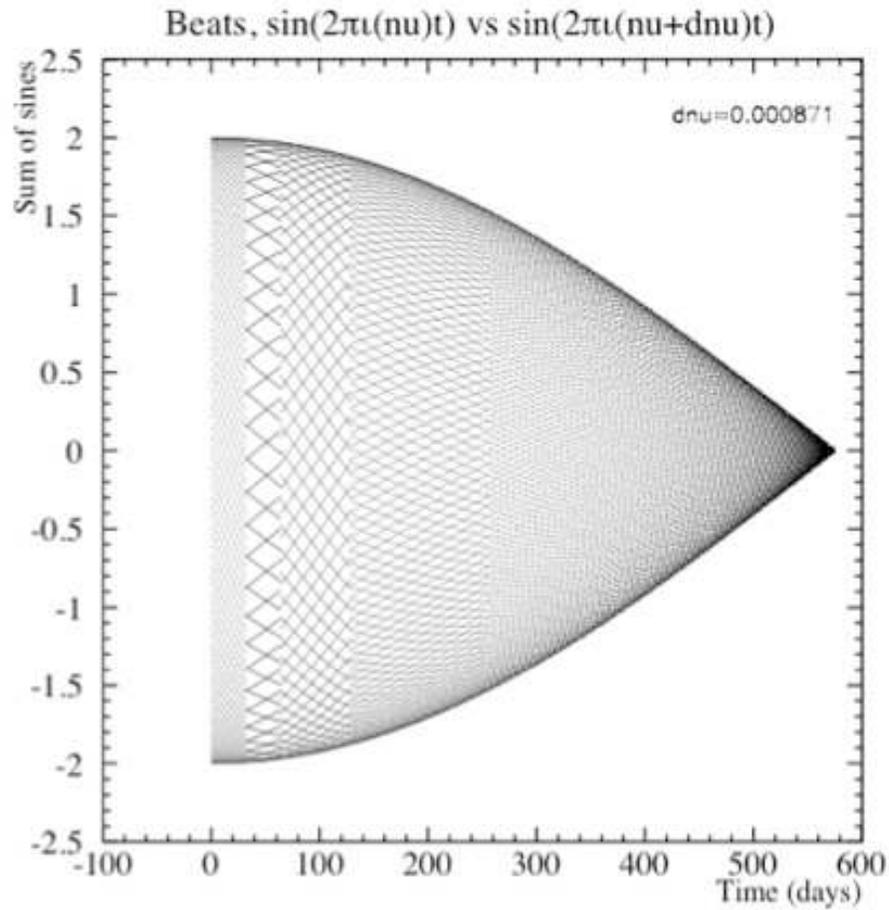
Figure 7.13: Degree of correlation between two neighboring sine waves, separated by the specific value $\delta\nu_{D_2O} =$, as a function of time, for the D_2O phase. As is expected, for the length of the D_2O data set (T_{D_2O}=594 days), and the $\delta\nu$ value found for this particular length of time, the minimal correlation is reached at $t = T_{D_2O}$. (The curves inside the structure are simply artifacts of plotting with reduced data points, for purposes of plotting; original (filled-in, pixel-heavy) figure can be provided upon request).

chosen). If we use spacing sizes any larger than this, we risk washing out a signal with sampling separations that are too large. We want to ensure that there *are* correlations between neighboring frequencies (rather than avoid them) so as to not miss a potential signal by misaligning our sampled frequencies; we want to oversample the frequency space to guarantee that we will not miss a signal.

Ideally we would like to use spacings which are only a small fraction of the $\delta\nu$ value which we have found, but this is not feasible given the limitations we are faced with, namely the necessity for speedy analysis and processing. In Figure 7.14, the Rayleigh power for a particular signal frequency is shown, with different spacings indicated on top of the trial sampling frequency separations (the spacing used to create the figure was orders of magnitude smaller than the value found, $\delta\nu$). It is evident that with different fractions of the 'maximum' allowable spacing, we would be able to oversample a given signal to different degrees. As we have stated, CPU time is a central determining factor, and as such, we have compromised with a value of $\delta\nu$ equal to 1/4 the value found for the combined-phase data set length ($\delta\nu = \delta\nu_{COMB}/4$). This leaves us with 1.6 million frequencies to sample, with a separation of 0.000089 day^{-1} for the combined-phase frequency region.
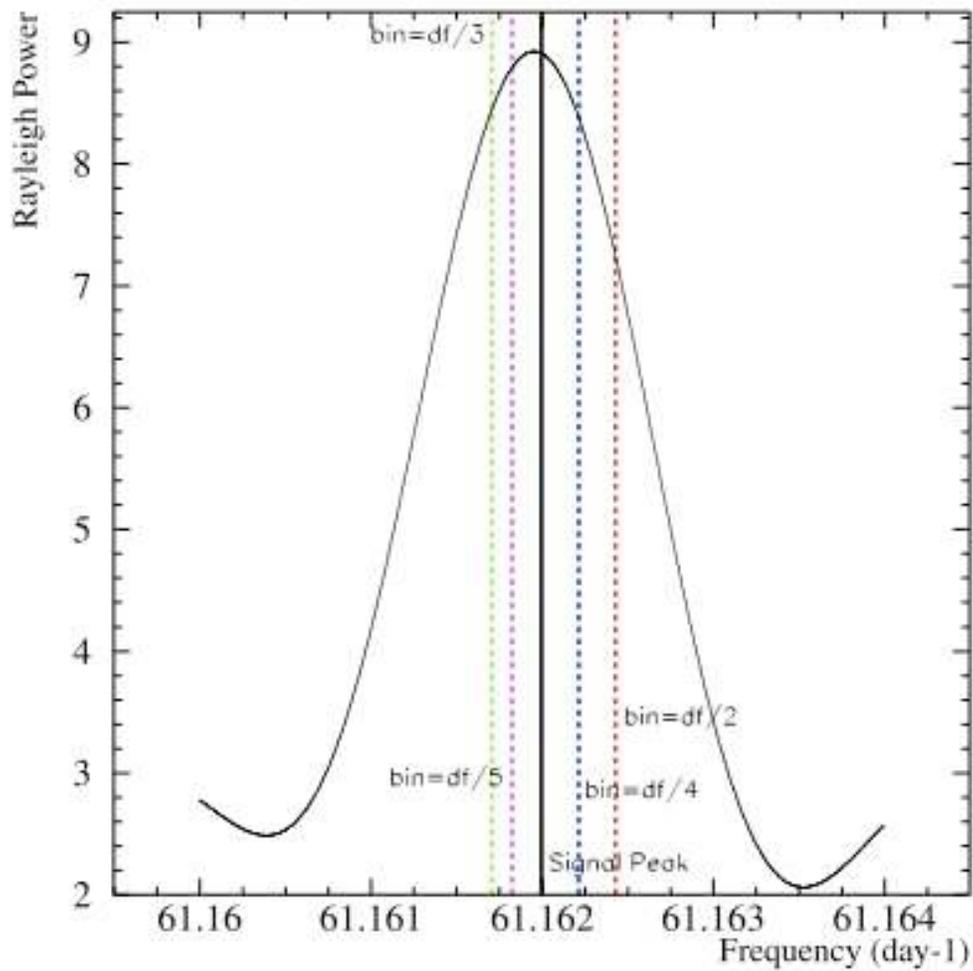
Figure 7.14: Zoomed-in view of a signal peak in the Rayleigh power spectrum, with different frequency sampling-size values superimposed on top of the over-sampled power spectrum.

7.4 Confidence Level Generation for the High-Frequency Search

We have seen that the Rayleigh power behaves differently for different frequencies in the presence of SNO's deadtime window function; for some frequencies, the distribution of Rayleigh powers is exponential, and for others the distribution appears more Gaussian due to the inherent 'periodicity' of the detector's on and off times. This fact makes a simple interpretation of the results of the Rayleigh power test difficult, as there initially appears to be no explicit 'predictable' value for the significance of a detected signal, since the level of significance of a given value for a peak will depend on the typical value of the Rayleigh power for signal-free Monte Carlo at that frequency. This was taken into consideration when performing the Rayleigh power analysis in the low frequency regime, detailed in Chapter 6, and will be treated similarly in the generation of confidence levels for the high frequency regime, with a predictive analytic form describing the expected distributions of Rayleigh powers for individual frequencies.

Another consideration which must be taken into account is that for a search which spans such a large range of frequencies, we are hurt by the trials issue, in that the more frequencies we consider, the more chances we have to get a peak which is larger than the frequency-specific confidence level (the probability for statistical fluctuations up to a 'significant' level is increased as our statistics are increased). This is dealt with in a manner which is different from the

method used for the low frequency analysis, and which avoids the problems encountered in the earlier analysis.

7.4.1   Analytic Form of the Rayleigh power

The derivation and testing of the analytic form for the Rayleigh power was already described in great detail in Chapter 5, but for completeness, I will briefly outline it again here. In this study, the behavior of the Rayleigh power was modelled after a random walk in two dimensions, as discussed earlier, in Chapter 5. From this starting point, and with the full treatment of the two dimensions' variances and covariances, we arrive at an analytical predictive form for the behavior of the Rayleigh power at each frequency sampled, as is outlined in Chapter 5. Because this analytical form is built from our knowledge of the boundaries imposed by the SNO deadtime (in the form of our window function), it is truly a *predictive* form, up to a normalization constant, and as such allows us to attain the degree of accuracy required for the determination of confidence levels for such a large range.

The results from the analytic form and its numerical integration, shown in Figures 7.15 and 7.16, initially allowed us to calculate the confidence level for each frequency in the high-frequency range without including information due to rate difference effects between the two phases of running (D_2O and Salt phases). Having calculated the expected distribution for each frequency with no signal present, we can determine where a given Rayleigh power sits on that distribution, and from this we are determine the confidence level of each
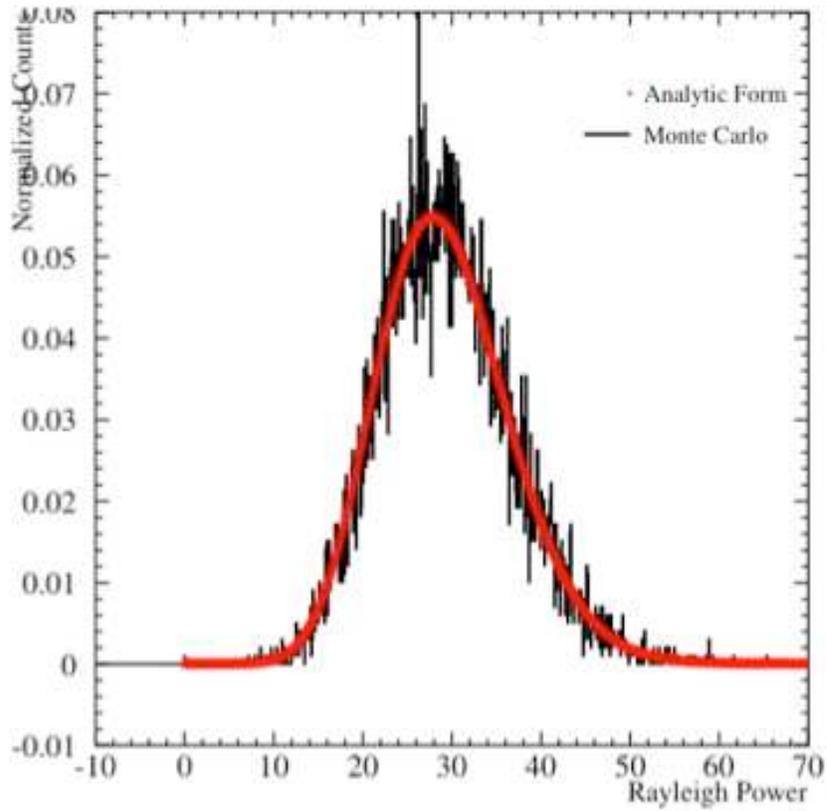
Figure 7.15: Monte Carlo (black) and analytic form (red) without phase-specific rate dependence included, for sampled frequency = 1 day^{-1}.

Rayleigh power for a given data set.

7.4.2 Rate Differences: Effects on the Analytic Form

While the previous discussion demonstrates the agreement between the predictive analytic form and the actual distribution of Monte Carlo Rayleigh power values with the deadtime window function taken into account, we have until
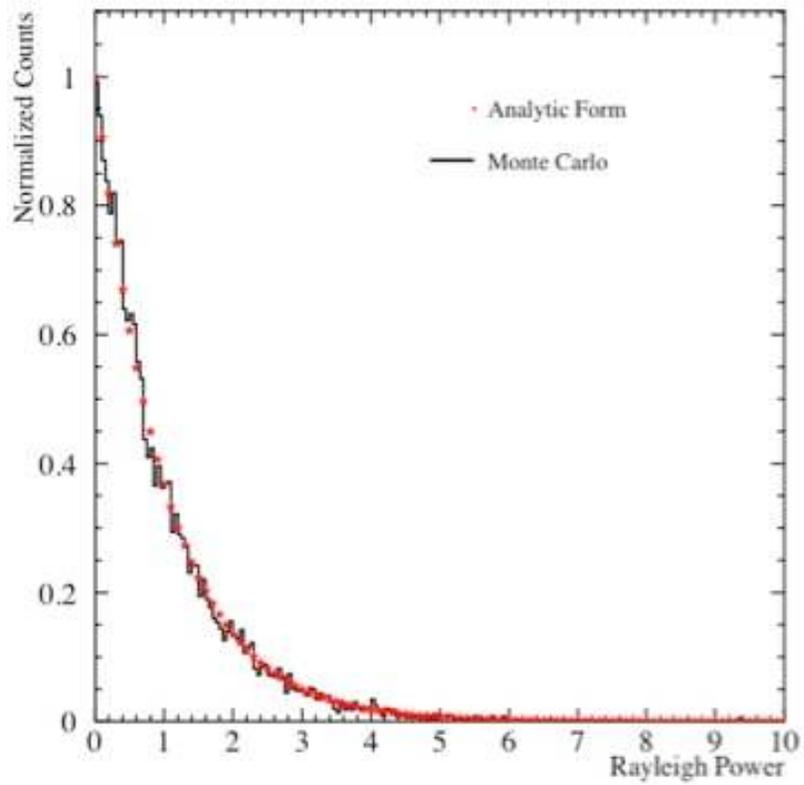
Figure 7.16: Monte Carlo (black) and analytic form (red) without phase-specific rate dependence included, for sampled frequency = 743 day^{-1}.

now neglected to include effects due to the difference in event rate between the D_2O and Salt phases in the analytic form. At first glance, the rate difference was thought to be unimportant for the purposes of this study, the reasoning being that the Rayleigh power test is impacted not by the universal increase in the number of events for the entire Salt phase, but instead by the relative position of events around the unit circle, and the resulting directional vector of the sum of the individual events. As this is dependent solely on the events' occurence relative to the frequency being sampled (*i.e.* the position of each event around the unit circle will depend on the circumference of the circle being sampled), the rate was thought to be of no consequence on the Rayleigh power value. Though to an extent this is a good approximation, for our purposes it is not accurate enough, as will soon be seen.

What was found in exploring this difference in rates between the two phases, and its impact on the analytic function's behavior for different frequencies, was that although by itself a rate change would not be expected to affect the Rayleigh power behavior at high frequencies, when there is a change in rate paired with a difference in the deadtime window function (the acceptance window for event times), an increase in rate causes the deadtime window for the higher-rate time period to be disproportionately sampled. In other words, what is important is not that the rate itself changes, but instead that the deadtime (or phase coverage) is different as the rate is changed. If the deadtime windows were identical between the D_2O and Salt phases (or if there were no deadtime window at all), the difference in rates would not have mattered for

172

our high-frequency analysis. However, since for SNO the deadtime window will always vary with time , we do need to include the presence of this rate difference between the two phases. This has also been evidenced in retrospect by observing the differences in the behavior of the Rayleigh power spectra for Monte Carlo including phase-specific rates vs Monte Carlo assuming constant rates between phases.

In Figure 7.17, I show the non-rate-inclusive analytic form in relation to the new rate-inclusive Monte Carlo. We can see that if we were to assume that the previous non-rate-weighted analytic form would still accurately describe the new rate-inclusive Monte Carlo, though we would have been roughly correct for most frequencies, for frequencies like that shown in Figure 7.17 we could have easily claimed false detection of a signal due to mismodelling of the signal-free Rayleigh power distribution.

So the question, then, is how must we include this effect in our analytic form? The method which we have settled on comes from the initial understanding of the Rayleigh power in terms of a two-dimensional random walk. If we consider the D_2O and Salt phases to each be independent random walks, with unique means and variances resulting from their respective deadtime windows, then we can weight the two phases' means by their respective contributions to the total. The number of steps taken, or in our case the number of events having occurred, during each phase will determine the extent of sampling of the deadtime window, and therefore we must accordingly include this in our analytic description of the Rayleigh power behavior.

173
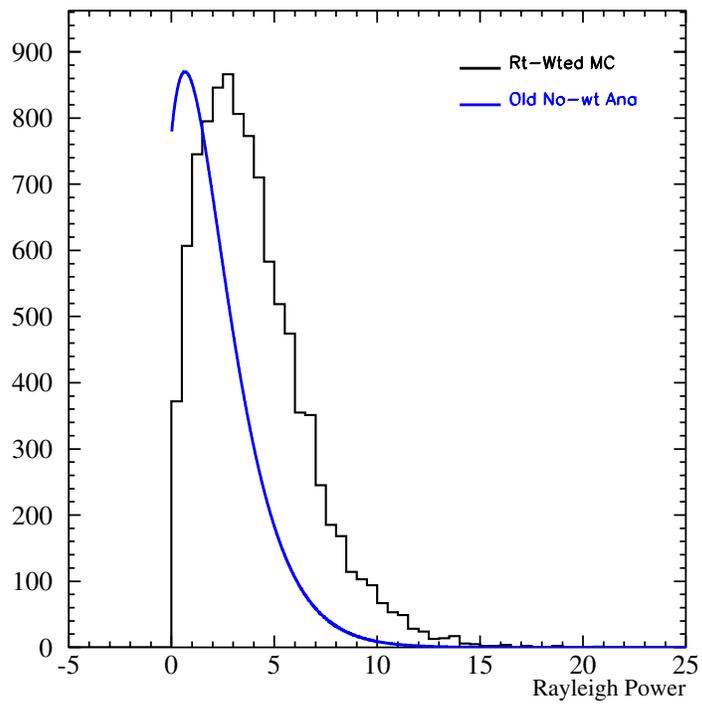
Figure 7.17: Demonstration of the poorness of fit of the analytic form with no phase-dependent rate information included, compared to the rate-inclusive Monte Carlo Rayleigh power distribution.

174

The basic setup of the analytic form will be exactly the same as it was before taking the rate effects into account, with the exception that now, instead of weighting global means, variances, and covariances by the total number of events (*i.e.* equivalent weighting of both deadtime window functions) we will now include relative N-weighting and phase-specific parameter calculations. In other words, any term which initially had a standard weighting of 'N' will now be divided according to phase, and weighted with the appropriate phase-specific 'N' factor:

$$N\mu_{x,y} \rightarrow N_{D_2O}(\mu_{x,y})_{D_2O} + N_{Salt}(\mu_{x,y})_{Salt}$$

Similarly for any term involving variance or covariance:

$$N\sigma_{x,y}{}^2 \rightarrow N_{D_2O}(\sigma_{x,y}{}^2)_{D_2O} + N_{Salt}(\sigma_{x,y}{}^2)_{Salt}$$

and

$$Ncov(x,y) \rightarrow N_{D_2O}(cov(x,y))_{D_2O} + N_{Salt}(cov(x,y))_{Salt}$$

So now, making the above substitutions to include phase-specific N-weighting, the unexpanded χ^2 used to calculate the analytic form of the Rayleigh power is:

$$\chi^2(X,Y) = \left(X - [N_{D_2O}(\mu_x)_{D_2O} + N_{Salt}(\mu_x)_{Salt}], Y - [N_{D_2O}(\mu_y)_{D_2O} + N_{Salt}(\mu_y)_{Salt}]\right)$$

$$\begin{pmatrix} (N_{D_2O}(\sigma_x{}^2)_{D_2O} + N_{Salt}(\sigma_x{}^2)_{Salt}) & (N_{D_2O}(cov(x,y))_{D_2O} + N_{Salt}(cov(x,y))_{Salt}) \\ (N_{D_2O}(cov(x,y))_{D_2O} + N_{Salt}(cov(x,y))_{Salt}) & (N_{D_2O}(\sigma_y{}^2)_{D_2O} + N_{Salt}(\sigma_y{}^2)_{Salt}) \end{pmatrix}^{-1}$$

$$\begin{pmatrix} X - [N_{D_2O}(\mu_x)_{D_2O} + N_{Salt}(\mu_x)_{Salt}] \\ Y - [N_{D_2O}(\mu_y)_{D_2O} + N_{Salt}(\mu_y)_{Salt}] \end{pmatrix} \quad (7.3)$$

From the above expression for χ^2 we can then determine the full analytic form, including the change in rate between the D_2O and Salt phases and the relative deadtime window functions, with the relation

$$f(z, \psi)dz = \frac{1}{C} \int e^{(-\chi^2/2)} \quad (7.4)$$

By numerically evaluating all parameters for each phase, we have modified the analytic form to produce a rate-change-inclusive prediction for the combined-phase Rayleigh Power. The results of these changes in the analytic form and the agreement with Monte Carlo are displayed in Figures 7.18 through 7.20.

In Figures 7.18- 7.20, we can see the agreement between rate-weighted Monte Carlo Rayleigh power distributions and the N-weighted analytic predictive form for standard exponential-style distributions and gaussian-style distributions, as well as an 'in-between' distribution which is not quite gaussian or exponential. As was anticipated, Figure 7.20 shows a marked change from conditions where rate differences were not accounted for in either the Monte Carlo or the analytic form, to the conditions where the rate differences were accounted for in both the Monte Carlo and the analytic form. It is very encouraging to see that the rate-inclusive analytic form agrees well with the rate-inclusive Monte Carlo, and we are now safe to rely on the analytic form as a predictive method of generating confidence levels for the different frequencies

176
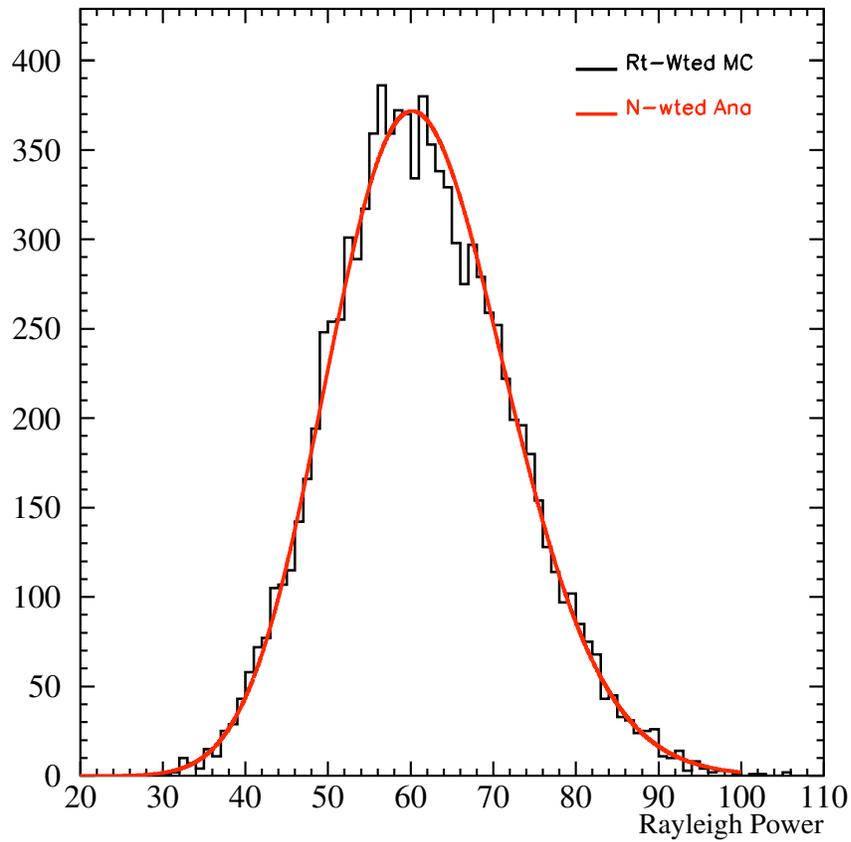
Figure 7.18: Agreement between rate-inclusive combined-phase Monte Carlo and N-weighted analytic form for the distribution of Rayleigh powers at sampled frequency = 1.000089 day^{-1}.
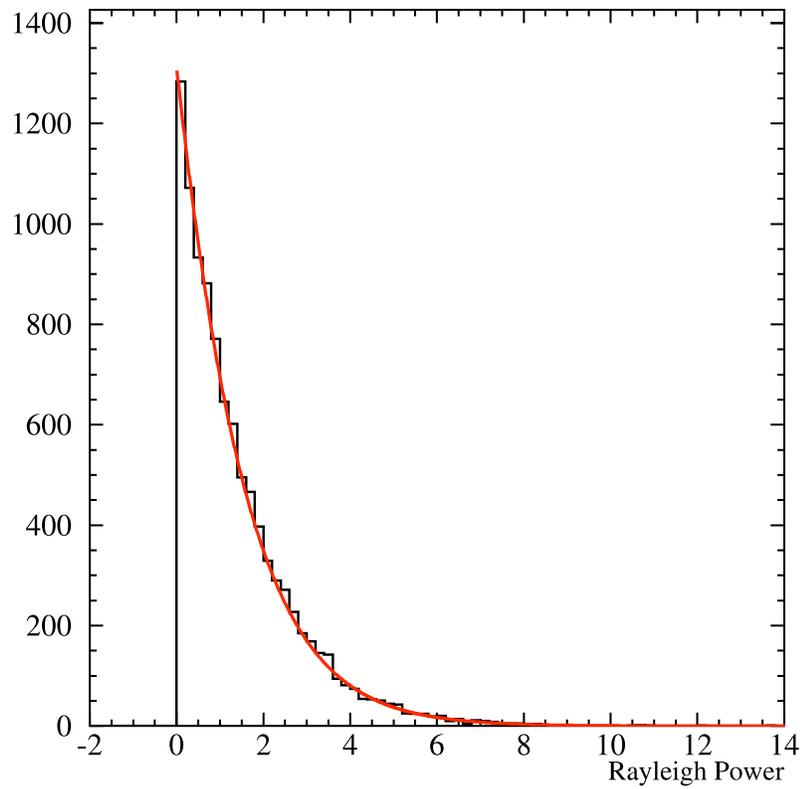
Figure 7.19: Agreement between rate-inclusive combined-phase Monte Carlo and N-weighted analytic form for the distribution of Rayleigh powers at sampled frequency = 1.046359 day^{-1}.
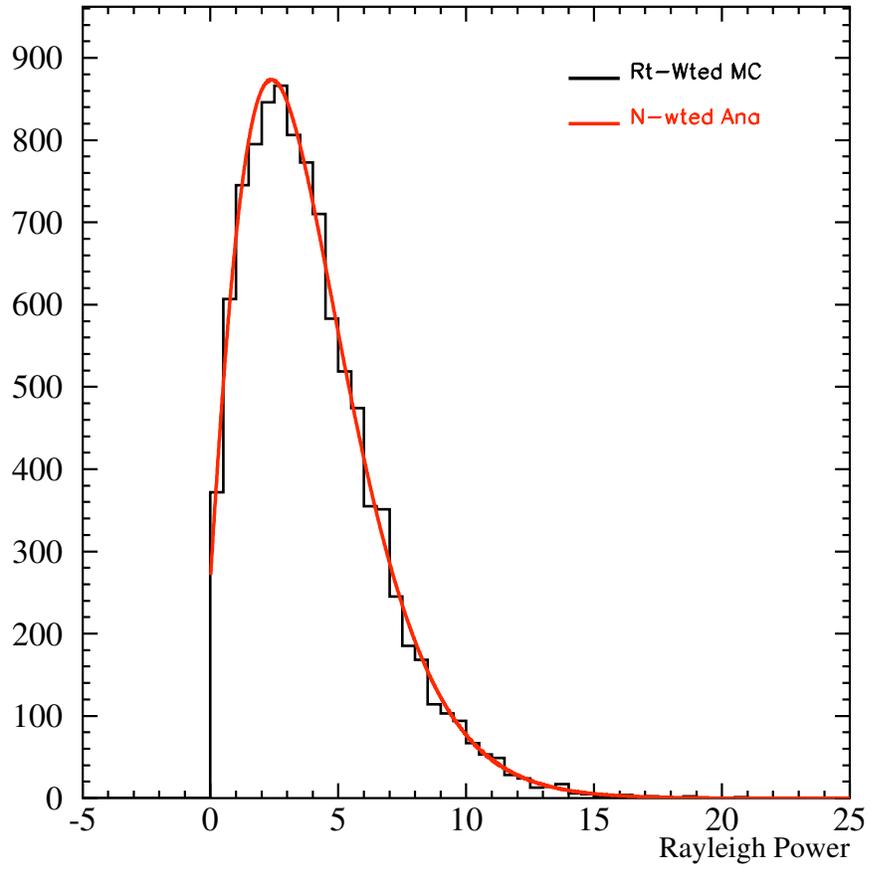
Figure 7.20: Agreement between rate-inclusive combined-phase Monte Carlo and N-weighted analytic form for the distribution of Rayleigh powers at sampled frequency = 1.000626 day^{-1}.

in our analysis.

7.4.3 Calculation of Data-Wide Confidence Levels

The N-weighted analytic form allows us to calculate the confidence level for each individual frequency in our high-frequency region by predicting the distribution of signal-free Rayleigh powers for each frequency sampled, and then allowing us to determine where on that distribution different Rayleigh powers sit. Powers on the higher end of the distribution naturally have a higher probability of being statistically significant, and by this reasoning, we can calculate each frequency's confidence levels by computing a fractional integral (integrate over entire distribution, then determine what Rayleigh power corresponds to 90%, for example, of the total distribution integral, and for this specific frequency, this is the Rayleigh power corresponding to 90% Confidence Level. For questions about this method, refer again to Chapter 5).

However, to find Confidence Levels for the entire range of the high-frequency search (including all individual frequencies, or 'data-wide') we must account for the problem of the trials penalty. This was also a problem in the low-frequency analysis; the trials penalty refers simply to the fact that the chance of any one sampled frequency having a Rayleigh power with a statistical fluctuation above a confidence level of 90% (for example) increases as we increase the number of frequencies which are included in our study. In the low frequency periodicity search, I used a brute-force method of calculating the data-wide vs frequency-specific CL (Chapter 6). This is necessary because the true calcu-

lation of a data-wide CL depends on the number of independent frequencies in the range being sampled, n, a number which is not easily extracted from the SNO data set for either the low or high frequency range. This brute-force method was time-consuming and not very elegant, and did not seem to be an appealing solution to apply to the much larger range of frequencies being treated for the high-frequency search.

In order to avoid using this method for the high-frequency analysis, I used a method of drawing a 'data-wide' confidence level from a distribution of maximum confidence levels. The idea for this approach is to generate 10,000 Monte Carlo data sets with no signal built into them, and run the Rayleigh power analysis on these 10,000 files. For each power spectrum, I determine the Rayleigh power with the maximum frequency-specific confidence level and record the value of this maximum confidence level. I then build a distribution of these maximum confidence levels, until I have 10,000 total max CL values. My data-wide confidence level is then pulled from this distribution; a 90% data-wide confidence level would correspond to the actual frequency-specific confidence level that sits at a point where 90% of this Max-CL distribution falls below it, and 10% sits above it. In this way, we have in a sense built the trials penalty into our confidence level determination. The distribution of maximum confidence levels and the cut off for a data-wide confidence level are shown in Figure 7.21.

This method of creating a 'confidence level of confidence levels' makes life easier, in the sense that the trials penalty is implicitly taken into account.

181
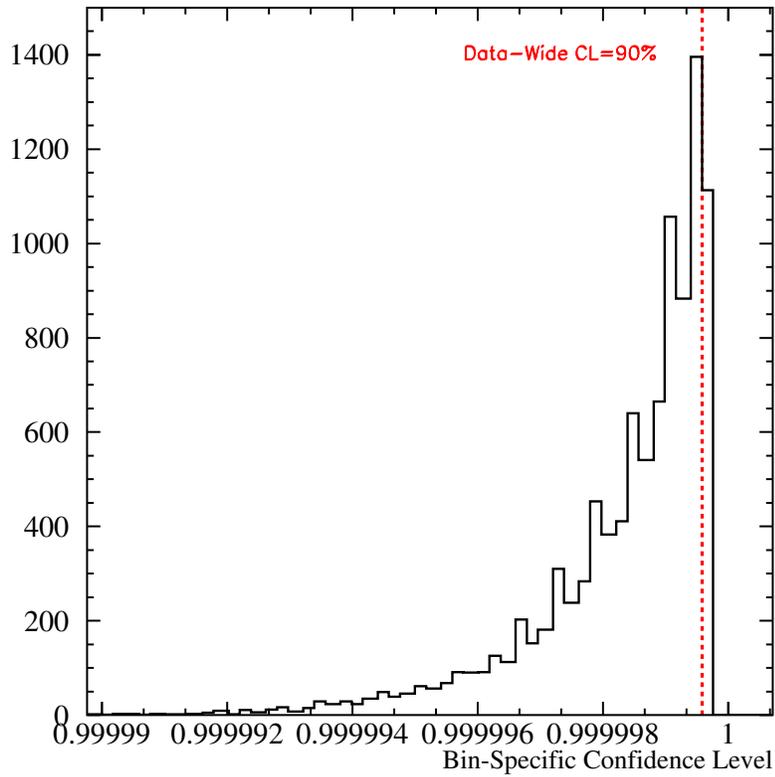
Figure 7.21: Distribution of Maximum Confidence Levels from Rayleigh analysis of 10,000 Signal-free Monte Carlo. By building this distribution of maximum confidence levels, we can determine a 'confidence level of confidence levels' and account for the trials penalty in our generation of the data-wide confidence level. The frequency-specific confidence level corresponding to the data-wide confidence level of 90% is shown with the superimposed red line.

There are, however, some drawbacks to this method, the most serious of which being the way that this method treats a power spectrum with multiple 'high' peaks. If the physical signal is of a non-sinusoidal form, potentially producing many peaks in one power spectrum, this method will not allow for detection, since it scans all the Monte Carlo power spectra for only the one highest peak in the power spectrum, not a combination of multiplets with high confidence-level Rayleigh power values. Therefore, if the actual signal's power is spread amongst several frequencies, this could be statistically significant, but this 'CL-of-CLs' method would not interpret it as such.

It should be noted that in Figure 7.21, the red line indicates the trials-weighted confidence level value that each individual frequency's corresponding Rayleigh power must have to ensure a data-wide confidence level of 90%. This is obviously a much higher value for each frequency than 0.90; the frequency-specific CL=90% value is closer to 0.9999999, due to the inclusion of the trials penalty in this calculation. So although we still are only requiring 90% confidence level for the entire data set (*i.e.* 1 out of every 10 'simulated SNO's' would give us a peak higher than this line due to statistical fluctuations), we are requiring a much higher confidence level for each individual sampled frequency (only something like 1 out of every 10 million frequencies would give us a peak higher than this frequency-specific line due to statistical fluctuations).

7.5 Results of Rayleigh analysis with SNO Data

With the successful inclusion of the trials penalty in our generation of the data-wide confidence levels for the high frequency region, we now have a meaningful way of determining which, if any, peaks in the SNO combined-phase data are significant. In Figure 7.22, the results from the Rayleigh analysis of the entire data set is shown with the Confidence Level=90% line indicating the Rayleigh power for each individual sampled frequency which would have been necessary to have a detection of a periodic signal at the 90% level. For this analysis, we used the same event list which was used in the Low Frequency periodicity study. The event list therefore consisted of 2,924 events in the D_2O phase, where an effective kinetic energy threshold of $T_{eff} > 5$ MeV was used, and 4,722 events in the Salt phase, where the effective kinetic energy threshold was $T_{eff} > 5.5$ MeV. In both phases a fiducial volume of $R < 550cm$ was used, and burst cuts were applied to remove muons and muon followers from the data set. For a more detailed explanation of the specific burst cuts used, consult [8].

The highest peak in the data set for this all-inclusive high frequency region was at a frequency of 103.384 day^{-1}, with a confidence level of only 1.95%. Though this may seem alarmingly small at first glance, we must remember that we are, in essence, sampling a probability distribution here; our confidence level analysis tells us what the expected distribution of many SNO data sets would look like, and the physical (real) SNO data simply sits on the low end of this distribution.
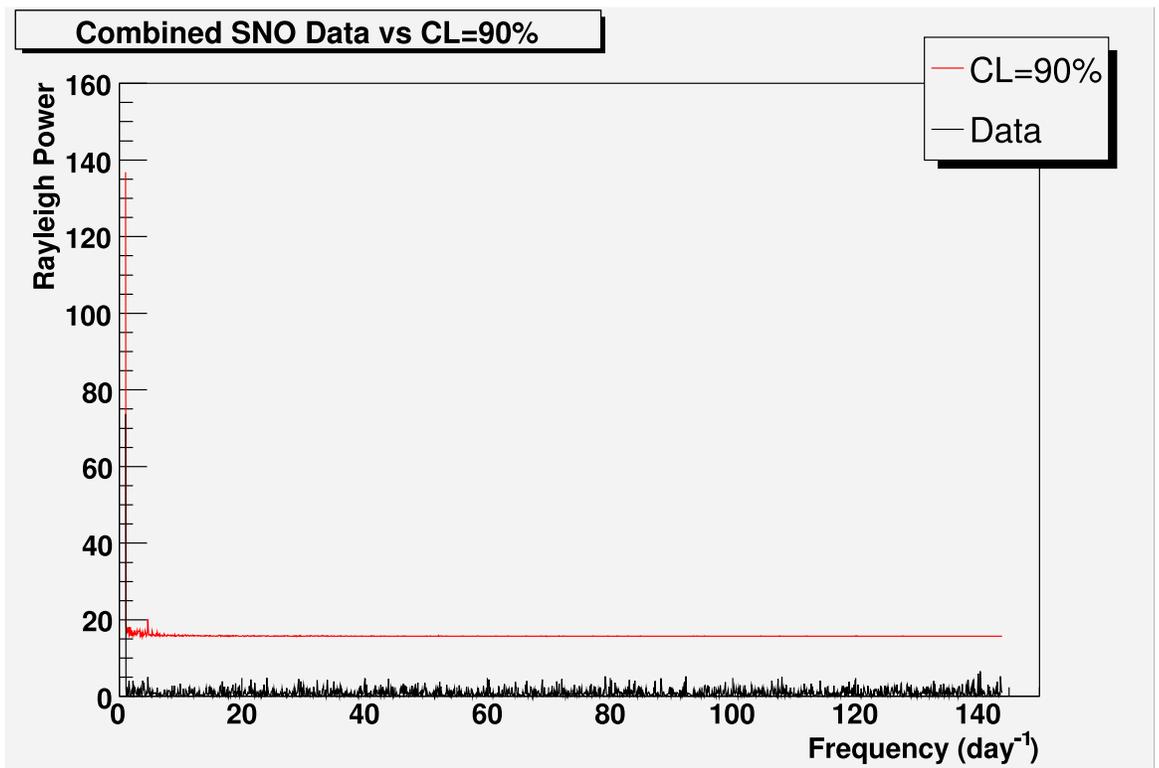
Figure 7.22: Rayleigh power spectrum of SNO combined-phase (Salt+D$_2$O) data (black), with the CL=90% curve showing the levels which would be necessary for a detection at the 90% confidence level (red). This plot shows no statistically significant peaks in the SNO combined-phase data set.

The second-highest peak in the data set sat at a frequency of 58.735 day^{-1}, with a confidence level of 1.75%. These two zoomed-in regions are shown in Figures 7.23 and 7.24.

There is obvious substructure in these power spectra, specifically in the lower frequency regions. This is expected, due to the behavior of the Rayleigh power in the presence of SNO's deadtime windows. The largest built-in structures exist at lower frequencies (typically with periods no smaller than the hour scale), which is consistent with what one might expect; the largest periodicities in the SNO deadtime window come from day-to-day (or week-to-week) activities, such as shift changes, calibration schedules, etc. The dying-off of this substructure at higher frequencies (higher than 12 day^{-1}) is in good standing with our knowledge of any predicted non-physics elements of periodicity in the SNO data set, as there are no predicted high-frequency inherent periodicities in the SNO data set. The region with highest inherent substructure is shown in Figure 7.25.

As we have found no statistically significant signal in our analysis of the combined-phase data set, we would like to be able to set a limit to understand what the lack of detection might be able to tell us about the existence of a high-frequency periodicity in the solar neutrino flux. To do this, we must determine what sensitivity SNO might have to such a signal.
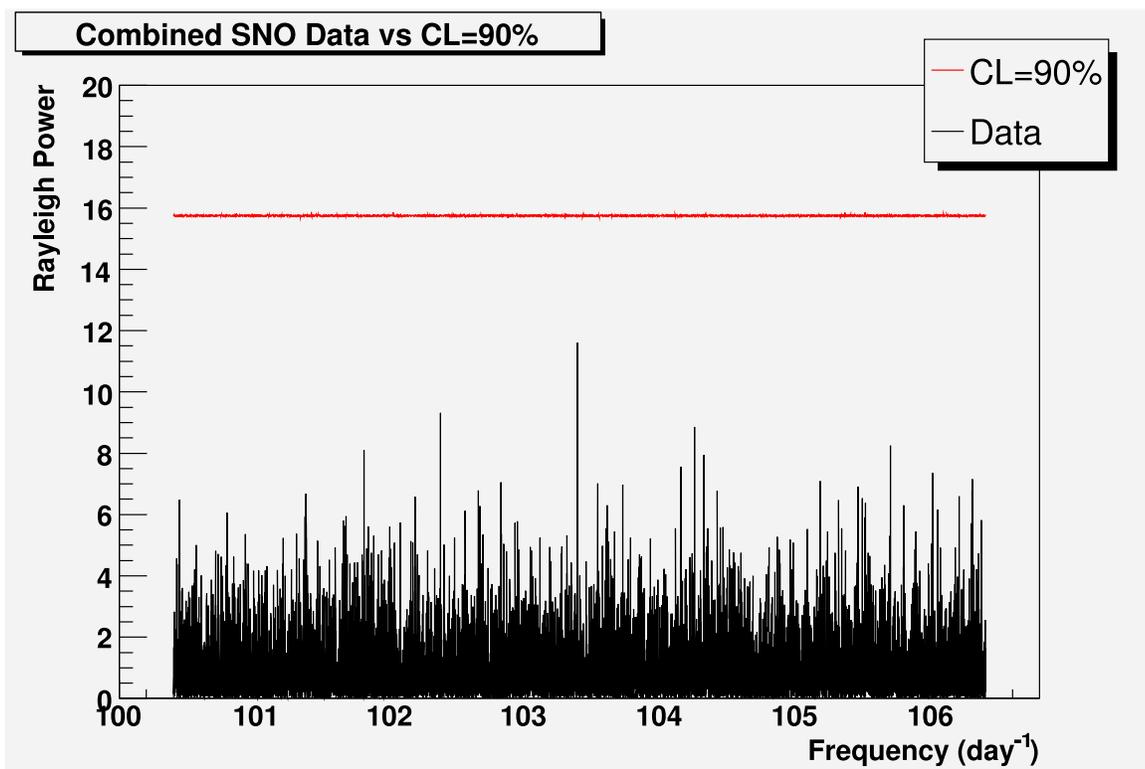
Figure 7.23: Zoomed-in region of Rayleigh power spectrum for highest-significance peak in the combined-phase data set, which was detected at frequency=103.384 day^{-1}, with a confidence level of 1.95%.
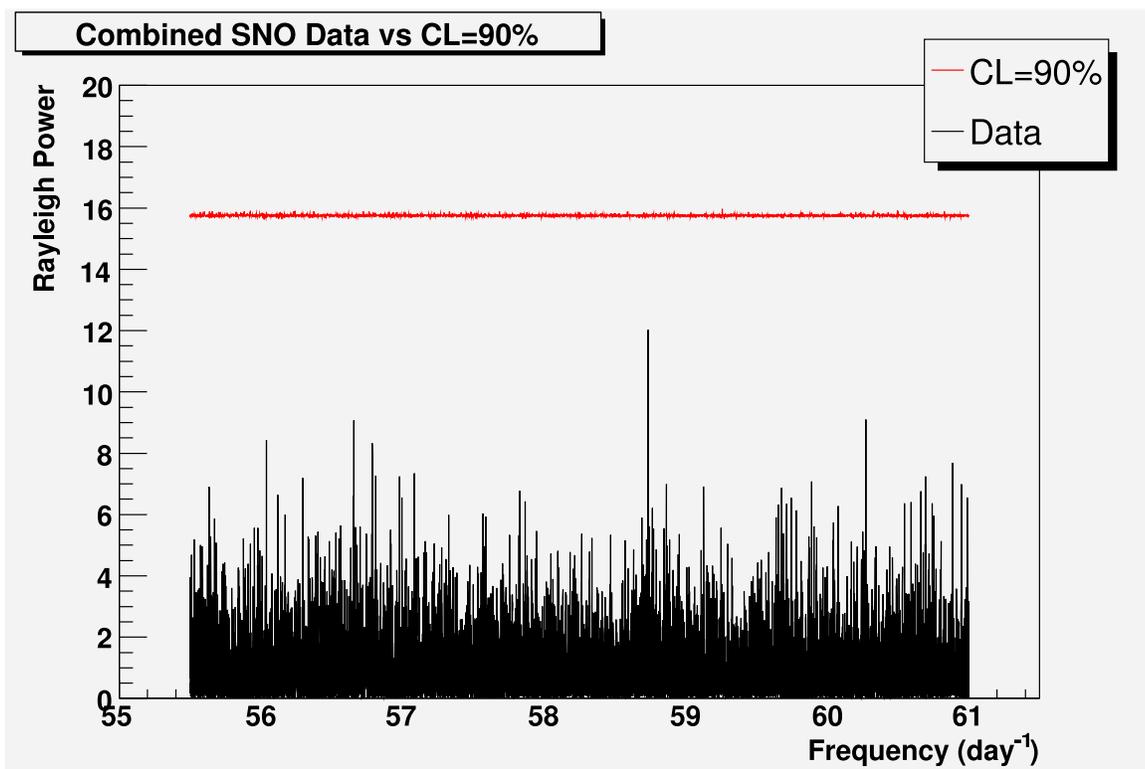
Figure 7.24: Zoomed-in region of Rayleigh power spectrum for the second-highest-significance peak in the combined-phase data set, which was detected at frequency=58.735 day^{-1}, with a confidence level of 1.75%.
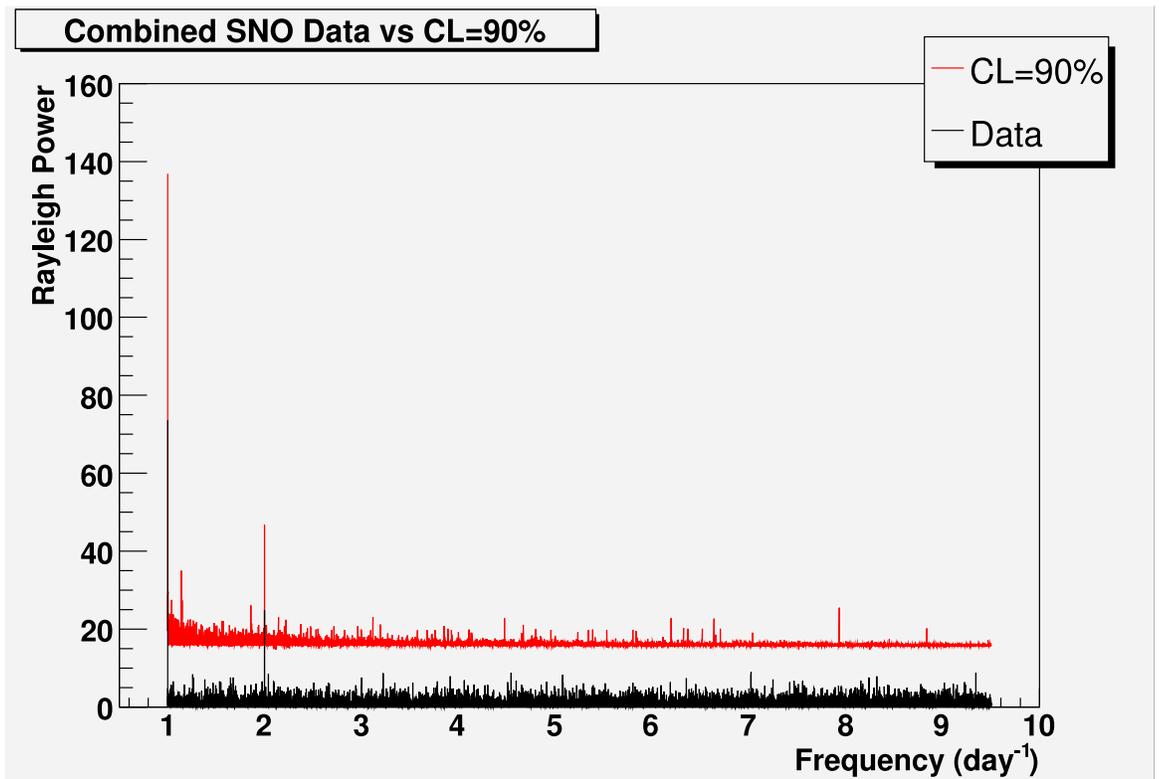
Figure 7.25: Zoomed-in region of Rayleigh power spectrum for the lower frequencies. There is evident substructure in the power spectrum, due to inherent deadtime-based periodicities detected with the Rayleigh power test.

7.6 SNO's Sensitivity to a High-Frequency Periodic Signal

The determination of SNO's sensitivity to a high-frequency sinusoidal signal is done by performing the Rayleigh power analysis on Monte Carlo files with signal built into them, and finding at what strength of signal we are no longer able to detect the presence of periodicity in our data set. I should emphasize that this is not directly telling us anything about the 'size' of the g-mode oscillations, but instead is telling us what magnitude of periodicity in the ^{8}B neutrino flux we might be able to detect. The actual question of the magnitude of g-mode oscillations in relation to their resulting effects on the neutrino flux is a complicated one, and has been explored by Bahcall and Kumar in [56] and Burgess *et al.* in [57], as well as Bamert, Burgess and Michaud from [58], and more indirectly by Burgess and Michaud from [58], and more indirectly by Couvidat, Turck-Chieze and Kosovichev in [59].

As has already been described, the Signal Monte Carlo files are generated in an acceptance/rejection routine which allows a sinusoidal amplitude to be built into the SNO deadtime window function. The Rayleigh power analysis is carried out on these Signal Monte Carlo, and the resulting power spectra are filtered through 'signal-detection' code which uses the confidence levels which we've determined (based on the signal-free Monte Carlo) to evaluate the relative 'strength' of the peaks detected.

The code I use to determine the level of signal which is necessary for the
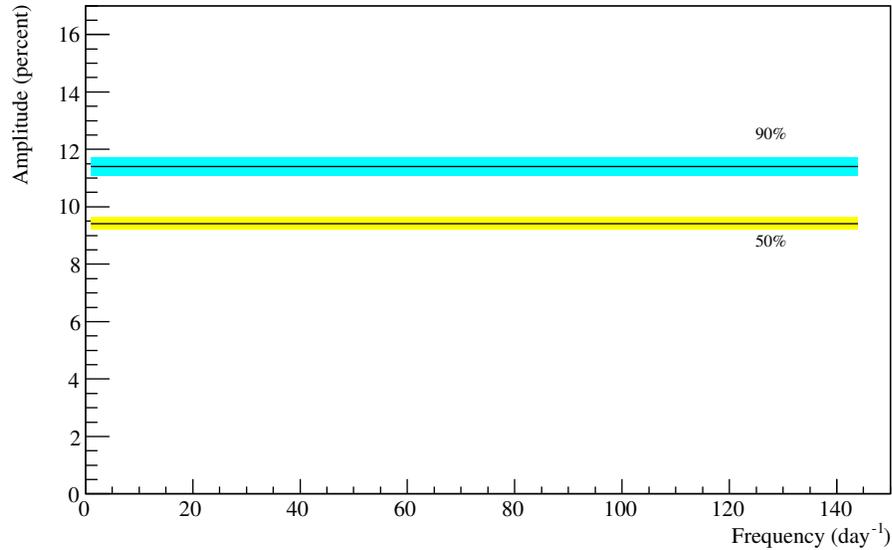
Figure 7.26: SNO's Sensitivities to a High-Frequency Periodic Signal in the combined-phase data sets (D_2O and Salt) for the Full High Frequency Search Region. The two lines shown are the calculated sensitivity at which we detect a signal 50% of the time, with 99% confidence level, and 90% of the time, with 99% confidence level, and the yellow and cyan bands represent the respective measurement uncertainties.

Rayleigh power to detect it scans through all the power spectra and calculates what fraction of data sets have a signal which sits above the 99%-detection level. Figure 7.26 shows the sensitivity curves for 99% detection in 50% of the files, and in 90% of the files. As can be seen, to discover a signal in the full high frequency region at a confidence level of 99%, its amplitude must be at least 9.45% for us to detect it 50% of the time, and at least 11.53% for us to detect it 90% of the time.

191

Chapter 8

A Search for High-Frequency Solar 'Noise'

8.1   A High-Frequency 'Noise' Signal

Searching for a single peak in the frequency domain is the most direct (and optimistic) approach one can use to find periodic behavior in a data set. However, a physical signal may ultimately manifest itself in the form of a multiplet of peaks, spread amongst several frequencies. The methods used in this thesis would likely not be sensitive to such a signal, as our confidence levels and thus our sensitivity estimates have been based on the single-peak approach (requiring a more restrictive statistical interpretation). However, another possible manifestation of a high-frequency periodic signal in the sun is a 'signal' in the form of noise.

A recent paper by C. Burgess[57] addresses the possibility of the presence

of noise in the solar interior, in the form of matter density fluctuations caused by the effect of magnetic fields in the solar radiative zone on g-modes. Burgess claims the possible existence of resonances at radii where the frequencies of magnetic Alfven modes cross those of g-modes, resulting in spikes in the solar density profile at various radii (depending on strengths of magnetic fields present), and thus a predicted 'noise', which could in theory affect the solar neutrino propagation through the radiative zone.

Although no evidence has been found to support Burgess' proposed model, in light of his postulation of noise in the sun, we were prompted to consider not just a search for one single significant peak, but to look if there is a distribution of peaks due to noise. The presence of noise in the solar interior would almost certainly not create a strong enough signal for any particular frequency to register a statistically significant Rayleigh power, but if the noise were distributed over many different frequencies, it is likely that the distribution of Rayleigh powers, and their respective confidence levels, would have a slightly larger mean than the case of no signal. This is the goal of the search described here: to determine if 'noisy' density fluctuations are present in the solar interior, in a form which affects the solar neutrino propagation.

8.2   Rayleigh Power Test with Noise Model

The Rayleigh power method as described in this thesis (in Chapters 5, 6, and 7) would not be sensitive to a noise signal in its current setup. As was just mentioned, the statistical tools with which we interpret our Rayleigh power

spectra (our generation of trials-weighted confidence levels) have been built with the intention of detection of a single-peak signal. However, with very slight modification of our approach, we can use the Rayleigh power test as a method for analyzing

The method of search previously described for the high-frequency regime (Chapter 7) focused on the detection of a significant Rayleigh power peak for one frequency out of 1.6 million possible. Due to the number of frequencies searched, and the trials issue therefore associated with this search, the necessary frequency-specific confidence level which must be passed for a peak to be statistically significant is far more restrictive than the stand-alone confidence level for one individual peak, as was explained in Chapter 5. Because of the limitations due to the trials issue, any periodic signal which might be present in the sun would necessarily have to have a relatively strong Rayleigh power, and be restricted to (or at least strongly represented at) one frequency only, for it to be detected with the method outlined so far. In light of the Burgess paper, however, we now craft our search around a model in which the periodic signal is present in the form of noise, rather than a single-frequency sinusoidal variation.

The idea that the presence of noise could distort the Rayleigh power spectrum just slightly, and therefore not be detected by the previously carried out searches, is a reasonable one. In order to test this, we created a sample SNO Monte Carlo with a noise signal present, in the form of a 2% amplitude Gaussian white noise model. A comparison of the power spectra for this white noise

MC and the standard no-signal MC which has been used for the more general single-frequency signal search is shown in Figure 8.1.

8.3 Rayleigh Power Confidence Levels: A Figure of Merit

As it is difficult to judge the degree of distortion of the Rayleigh power spectrum in Figure 8.1, we will be able to get much more information from inspecting the distribution of confidence levels for the noise MC, since that distribution should account for any statistical fluctuations (which may be harder to rule out by eye). But before we get ahead of ourselves, let us first consider the case of the distribution of confidence levels for a Monte Carlo Rayleigh power file with no noise signal built in. We would expect that for a signal-free Monte Carlo, we would get a uniform distribution of confidence levels between 0% and 100% due to the fact that the Rayleigh powers will be truly randomly distributed in terms of the predictive analytic form for any given frequency (this is the form which is used to generate the confidence levels).

Just to reiterate what has already been described, for clarity: the Monte Carlo are generated with no knowledge of the predictive analytic form, but have been demonstrated to agree remarkably well with it for all frequencies tested. So if the no-signal Monte Carlo are truly signal-free for all frequencies (and include the SNO deadtime in them), this distribution of confidence levels turns out to be not only a means of inspecting our data for the presence of
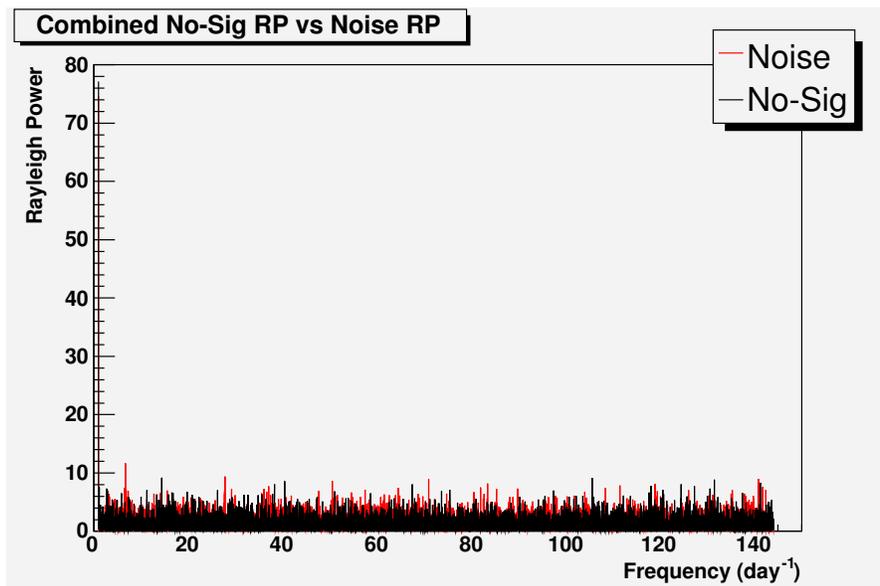
Figure 8.1: Comparison of Rayleigh power spectra for the case of no-signal SNO Monte Carlo (black), and Gaussian white noise signal SNO Monte Carlo (red). To make printing easier, this figure has been 'watered down' to only show the Rayleigh powers for every 100th frequency. Full frequency-inclusive figure can be provided if necessary.

196

solar 'noise', but it can be looked at as a test of the robustness of our analytic form.

If the Monte Carlo does (as we expect it to) agree with the predictions of our random-walk model analytic form, we should see an even distribution of confidence levels. If there were some bias to our form, or if it did not truly predict the behavior of the Rayleigh power at different frequencies, we would expect to have a slightly lower-mean distribution, in the case where we have underestimated confidence levels of the data (i.e. the statistical distribution of our no-signal Monte Carlo caused us to be too restrictive), or we would expect a slightly higher-mean distribution if our analytic form was not restrictive enough. If our predictive analytic form is actually correct, we would expect a mean of 50% (or 0.5), with an equivalent number of confidence levels above 50% and below 50%. In Figure 8.2, we can see that this is indeed the case.

8.4   Rayleigh Power Noise Model: Noise Monte Carlo

Now that we have convinced ourselves that not only is our expectation of a flat confidence level distribution true when no noise signal is present, we also have demonstrated the robustness of our predictive analytic form. The next obvious question is, how does our confidence level distribution look when there *is* a noise signal present? We are basing this confidence level distribution approach on the assumption that the distribution will be distorted if there is indeed a
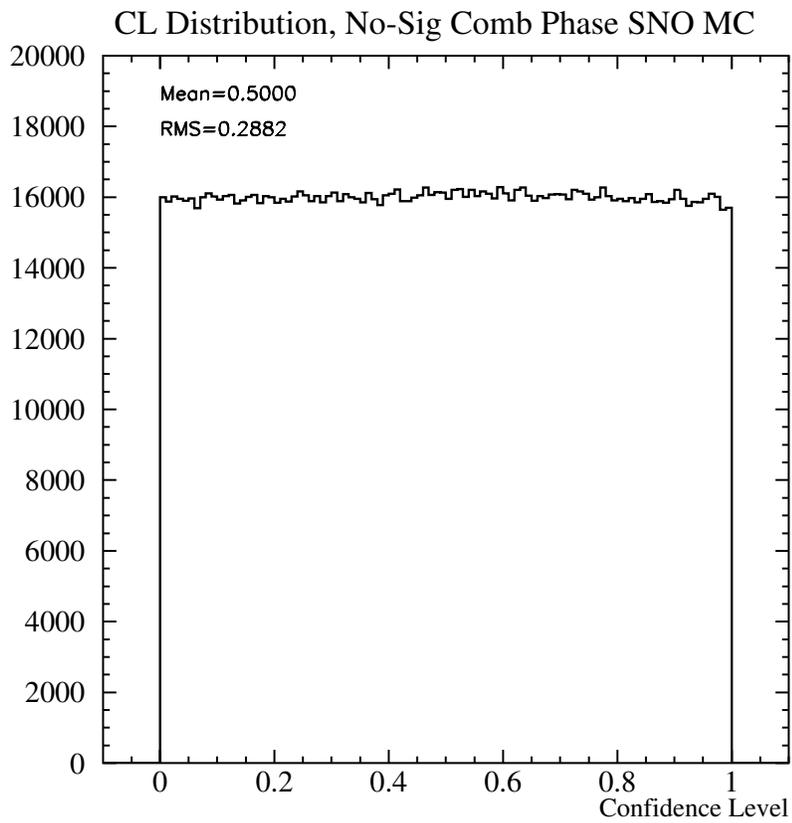
Figure 8.2: Distribution of confidence levels for SNO Monte Carlo with zero signal built in. The distribution not only tells us what a noise-free data set would look like in terms of Rayleigh power confidence levels, but also serves as a goodness-of-fit for our predictive analytic model, for all frequencies in our region.

noise signal in the data. So obviously, we must check that this assumption is correct. In order to do this, I ran the Rayleigh power code on the sample Monte Carlo data set with 2% amplitude Gaussian white noise distributed among all frequencies, and then generated confidence levels for all the frequencies (1.6 million) sampled. The distribution of confidence levels for all frequencies is shown in Figure 8.3.

8.5   Rayleigh Power Noise Model: SNO Data

It is evident that if there were noise present in the solar interior which affected the neutrino production and/or propagation, even at the 2% amplitude, as is shown in Section 8.4, this confidence level check would be sensitive enough to be used for its discovery. I have not performed a rigorous check of the point at which we lose sensitivity for claims of a detection: this should serve more as a proof-of-principle to verify that this type of signal could potentially be detected in our data set if it were indeed present in the solar neutrino signal.

That being said, the next obvious step is to investigate the nature of the confidence level distribution of the actual SNO combined-phase data. The distribution of confidence levels for the SNO combined-phase data is shown in Figure 8.4. The similarities between the data CL distribution and the no-signal Monte Carlo CL distribution are striking. There is apparently no noise present in our data set above at the very most a 2% limit. This is somewhat remarkable, in that this basically tells us that neutrino production in the sun is highly stable.
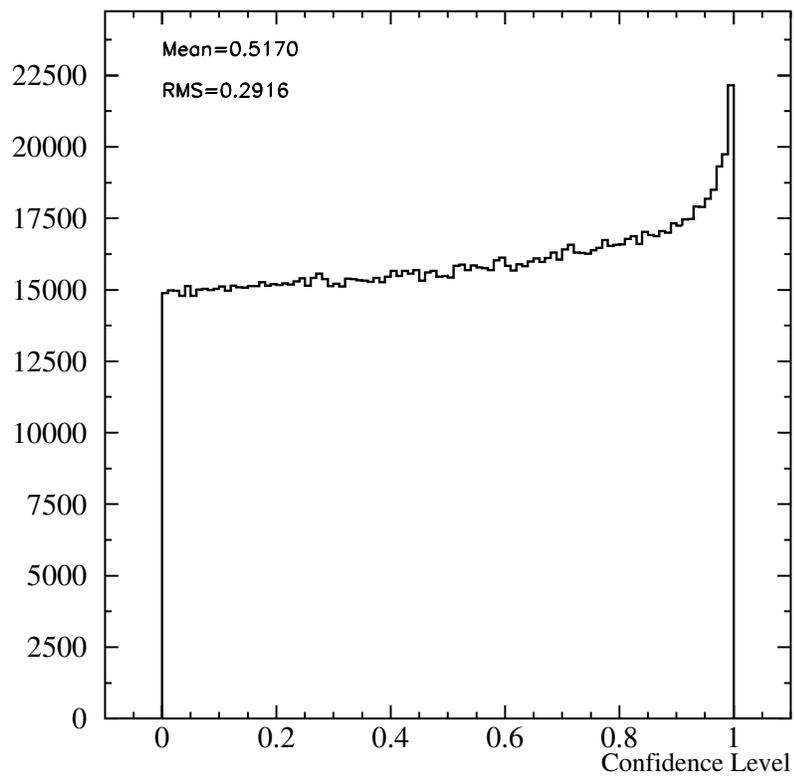
Figure 8.3: Distribution of confidence levels for SNO white noise Monte Carlo (signal of 2% amplitude).
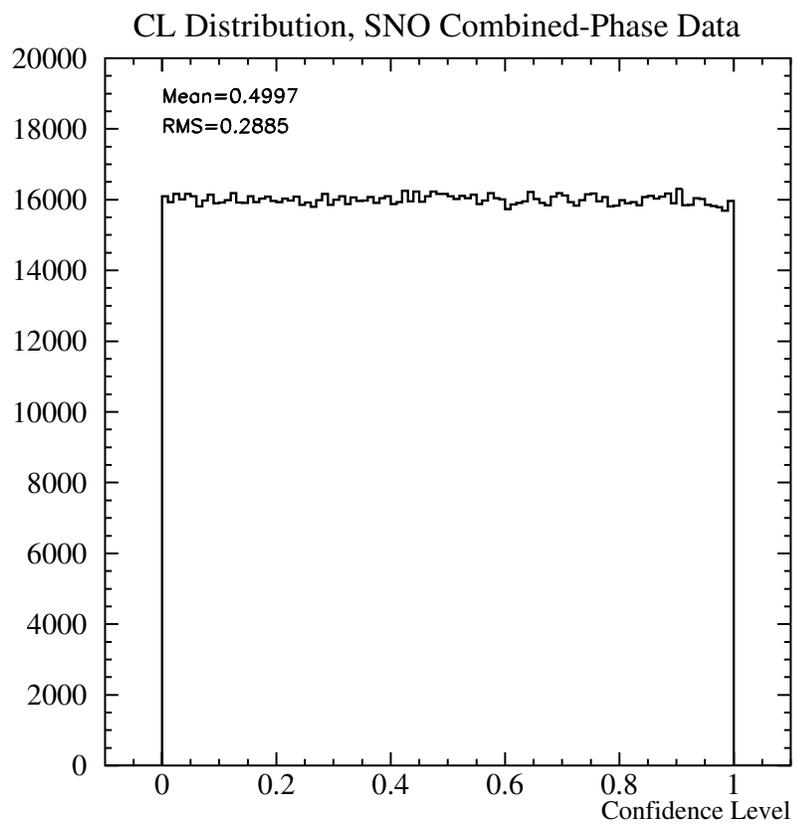
Figure 8.4: Distribution of confidence levels for SNO combined-phase data set.

There is more work that could be done in this regime (namely sensitivity studies), but this analysis, at least for the time being, will remain as it is; a straightforward check to determine whether or not this other potential time dependence in the neutrino flux (here, noise) is detectable in the SNO data set. We have seen that indeed, while it is possible to detect such a signal via the Rayleigh power test, it is not actually detected in the combined-phase SNO data set.

Chapter 9

Directed Search for

High-Frequency Periodicity

9.1   Experimental Motivation (Helioseismology)

The GOLF experiment has consistently reported claims of the detection of g-modes with data taken from the SOHO satellite[60], with a persistent multiplet signal in the region of 220 μHz[61],[53]. Recently, analysis of the ten-year data set has reinforced the case for detection in this region[62],[63], piqueing interest in g-mode searches directed in this region.

Though not yet verified by a truly independent experiment, the GOLF results *have* been verified by two other experiments aboard the SOHO satellite; VIRGO[1], measuring changes in solar irradiance via a full-sun photometer, and

---

[1]Variability of solar IRradiance and Gravity Oscillations

SOI/MDI[2], which measures changes of the solar surface via interferometry. As these experiments are housed by the same satellite as GOLF, their results are only considered nominally independent from those of GOLF. Temperature fluctuations of the SOHO satellite, for instance, have been looked into as a universal cause for this peak - and encouragingly, no correlations have been seen[63]. A sample power spectrum density from each of the three instruments is shown in Figure 9.1.

This tentative signal gives SNO the opportunity to perform a search specifically in the region of the strongest expected g-mode frequency, which greatly improves our sensitivity for detecting or ruling out a g-mode-induced signal in the neutrino flux. Because the high-frequency Rayleigh power analysis is significantly compromised by the trials issue, being provided with a directed-search region (with far fewer frequencies to be sampled) gives us the opportunity of searching for a signal with much greater sensitivity.

9.2   SNO Directed Search

The directed search is carried out in exactly the same manner as the full-range search. To determine confidence levels, signal-free Monte Carlo are generated and the Rayleigh power analysis is run on them. To determine sensitivities, signal Monte Carlo are generated, the Rayleigh power analysis is carried out, and ultimately signal-detection code is run on the resulting power spectra. The only difference for the directed search is that rather than scanning a

---

[2]Solar Oscillations Investigation/Michelson Doppler Imager
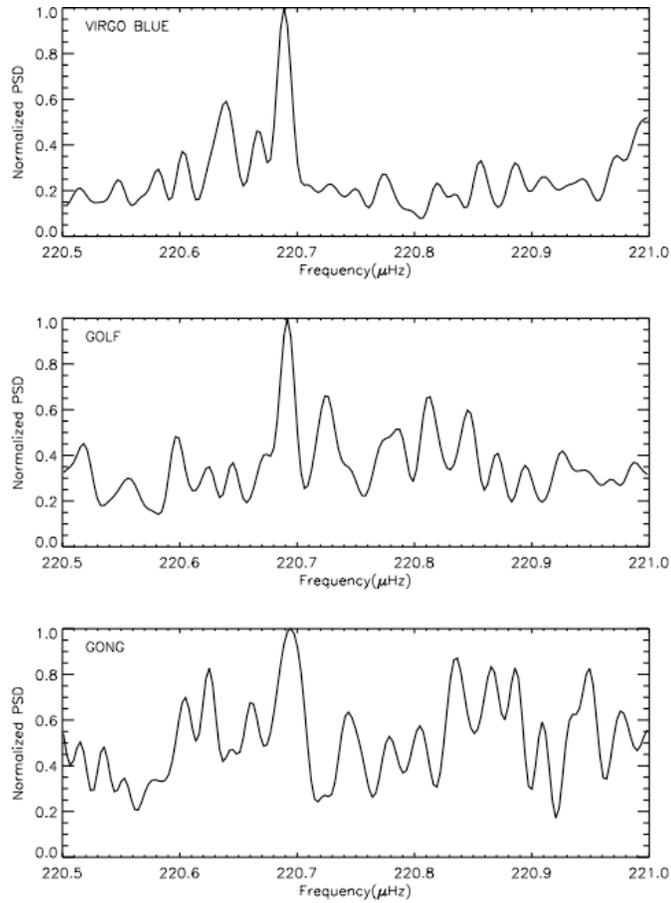
Figure 9.1: Power Spectrum Density (PSD) profiles from three experiments aboard SOHO; GOLF, VIRGO, and SOI/MDI. All three experiments show evidence of possible g-mode oscillation at 220.7 μHz.

region from 1/day to 144/day, which requires the treatment of 1.6 million individual frequencies, we 'zoom in' on a region centered around the GOLF-detected frequency, therefore significantly restricting the number of frequencies being scanned. This means that the bin-specific confidence level necessary for detecting a signal is reduced significantly (the chance of statistical fluctuations to a 'significant' level are much lower if you have 1,000 entries as opposed to 1,000,000).

The frequency at which GOLF has claimed evidence for g-mode oscillations sits at roughly 220 μHz. The analysis methods used in the detection depend on the presence of multiple peaks in the power spectrum for one specific l-mode. This causes the 'true' frequency of the g-mode to be washed out, in a sense, and the values reported as part of the multiplet therefore are interpreted as frequency splittings of the main mode, or possibly mixed modes (due to the theoretically predicted presence of the $l = 3$ and $l = 5$ modes in the same frequency range as the $l = 2$ mode). The actual predicted mode which corresponds to the GOLF 'detected' frequency range ie the $l = 2, n = -3$ mode, which sits at 222.14 μHz. To perform a thorough search of this region, including both the frequency at which g-modes were claimed to have been detected, as well as the corresponding frequency which is predicted for these modes, we designated the 'directed search' region to scan from 214 μHz to 225 μHz (roughly 18.5/day - 19.5/day). Using the value for frequency spacing which we have decided upon for the combined-phase SNO dataset, this leaves us with roughly 11,000 frequencies to sample, in comparison with the previous

value of 1.6 million. This allows us far greater sensitivity, due to the ever-present trials issue.

9.3 Results of a Directed Search

The results of the 'directed' search are shown in Figure 9.2, with the corresponding curve for 90% confidence level. As is clear from the power spectrum, there were no statistically significant peaks in this directed-search range. The highest-significance peak was found at a frequency of 19.2579/day with a confidence level of 58.28%. Though this is more significant than what we have found for the full-range search, it is still not statistically significant enough to merit claims of an independent verification of the GOLF results.

9.4 Sensitivities of SNO in a Directed Search

As there was no detection of a statistically significant signal in the directed-search GOLF region, we also would like to understand what limits can be placed on the impacts of g-mode oscillations on the neutrino flux in the directed GOLF region. In Figure 9.3, the sensitivity for detection of a signal at the 99% Confidence Level is shown, both for detection 50% of the time, and 90% of the time. The amplitudes necessary for detection in this region are lower than for the full search (due to the trials issue), and are similar to those published in the low-frequency periodicity search [8]. This is as expected, since the number of bins used for the GOLF directed search region are comparable to the number

207
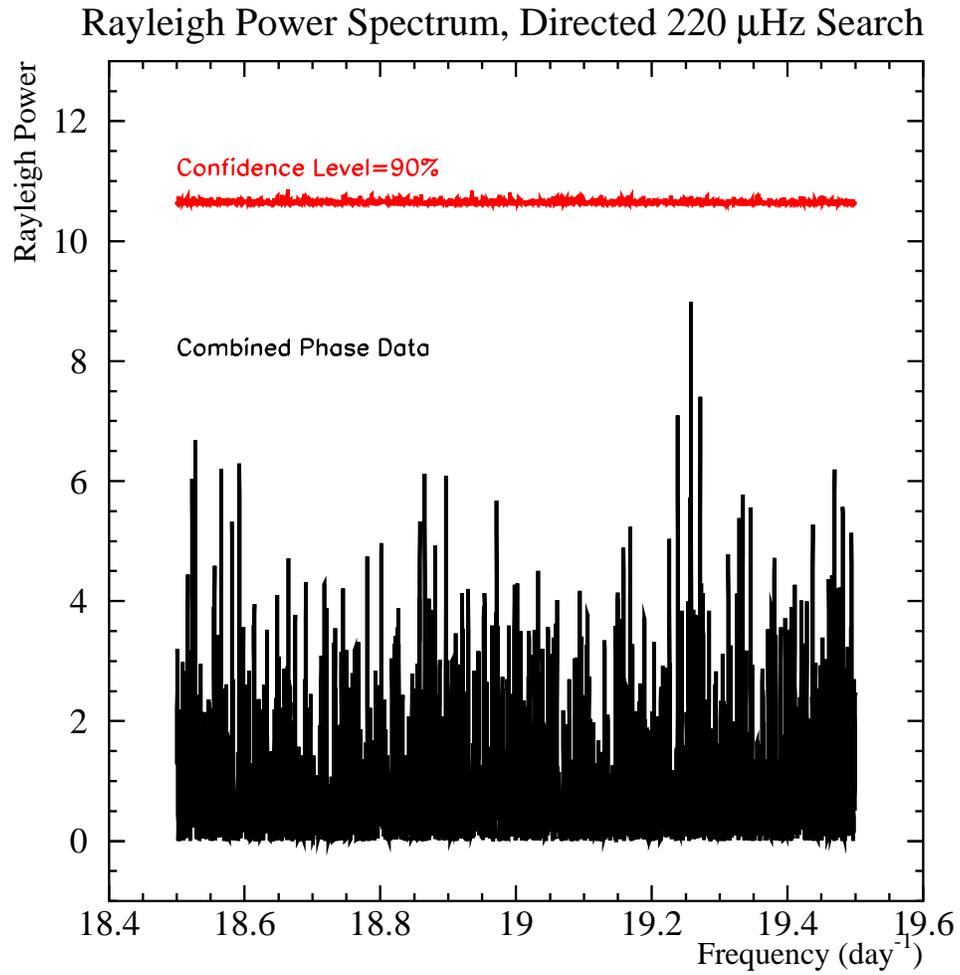
Figure 9.2: Rayleigh power spectrum for 'directed' high-frequency search, in black. The line corresponding to detection with 90% confidence level is shown in red. The highest peak in the power spectrum is found at a frequency of 19.2579/day with a confidence level of 58.28%.

of bins used in the low frequency periodicity search (11,190 vs 9,998). This is encouraging, as it reassures us that the Rayleigh power is competitive with the other two previously-used methods (maximum likelihood and Lomb-Scargle analyses), in terms of sensitivity to an inherent periodic signal in the neutrino flux in SNO.

Again, as is seen in Figure 9.3, for us to discover a signal in the directed high frequency region at a confidence level of 99%, its amplitude must be at least 7.85% for us to detect it 50% of the time, and at least 9.93% for us to detect it 90% of the time.
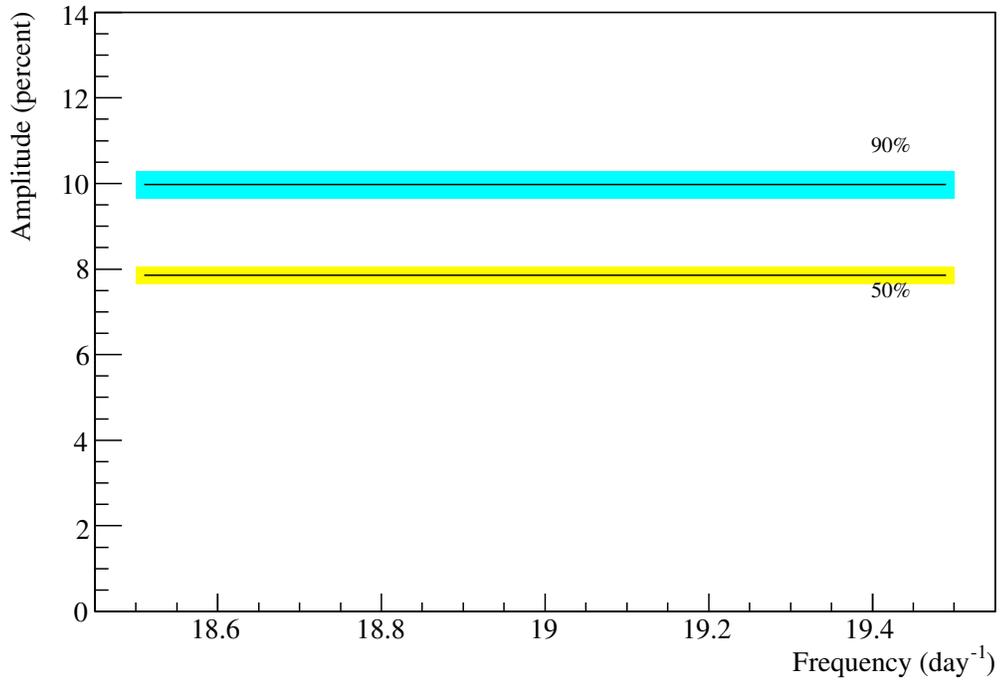
Figure 9.3: SNO's Sensitivities to a High-Frequency Periodic Signal in the Combined Phase Data sets (D$_2$O and Salt) for the Directed High Frequency Search Region. The two lines shown are the calculated sensitivity at which we detect a signal 50% of the time, with 99% confidence level, and 90% of the time, with 99% confidence level, and the yellow and cyan bands represent the respective binomial errors. The amplitudes necessary for detection in this region are lower than for the full search due to the trials issue, and are similar to those published in the low-frequency periodicity search [8].

Chapter 10

Conclusion

The weak interactions that neutrinos have with matter have made the neutrino one of the most elusive elementary particles to study. However, this weakness of interaction means that neutrinos from astrophysical sources can escape environments that trap or delay easier-to-detect photons. Neutrinos produced in fusion reactions at the solar core are able to travel unimpeded through the solar envelope to be detected on Earth. Similarly, neutrinos produced in distant processes such as supernovae can travel through the interstellar medium, retaining valuable information about the conditions which they have left behind. This has allowed neutrino detection to serve as a secondary observation method for astrophysical processes, in addition to the more standard electromagnetic detection.

The Sudbury Neutrino Observatory (SNO) has proven to be an optimal testing ground for studying neutrinos and the sources from which they are emitted; information about the true interaction time of the neutrino, as well

as the neutrino energy, have been gathered and processed, and because the detector target was heavy water, SNO was uniquely able to differentiate between neutrino flavors. These detector attributes, paired with the extremely low levels of backgrounds SNO has been able to attain, allow for unprecedented studies of time dependence in our neutrino signal.

I have developed and applied a statistical treatment of the directional Rayleigh power test, in the context of time-ordered data: as has been evidenced in this thesis, the Rayleigh power analysis, if appropriately interpreted, can be a very powerful tool in the context of a periodicity search.

I have put forth a number of frequency-dependent analyses of the SNO data sets. For the low-frequency periodicity study, three independent methods were used to investigate claims both for and against the existence of a periodic modulation of the solar neutrino signal. All three methods employed by SNO showed an absence of signal in the low-frequency regime.

I have also performed the first-ever analysis of high-frequency periodicity in the solar neutrino flux, with tentative helioseismic evidence for g-mode oscillations as a motivation and guiding tool. In a separate study, I have concentrated on the most empirically promising band of frequencies to determine whether predicted g-mode oscillations could appreciably affect neutrino production in the solar core, or could impact their propagation as they exit through the radiative zone of the sun. The effect of solar 'noise' on the solar neutrino flux has also been studied. In all of these contexts, I have measured no statistically significant modulation in the ^{8}B solar neutrino flux.

As a heavy-water Cherenkov detector, SNO has the potential to carry out a search for bursts of neutrinos of astrophysical origin, which are not accompanied by an optical trigger. I have laid out the groundwork for addressing the backgrounds of such a search, and removing them from the data set, with the motivation of carrying out a blind burst search on SNO combined-phase data.

Though SNO is no longer in production-running, there is still a great deal of exciting physics which can be extracted from the data set, as is shown by the analyses in this thesis. Even more promising, a new generation of neutrino experiments is at the forefront of current sensitivity thresholds, and promises to open up new energy regimes and improve on SNO's low-background and real-time-detection capabilities. Perhaps with these advancements, we can push farther into experimentally untapped realms, and gain a better understanding of the neutrino's potential to illuminate its astrophysical origins and surroundings, from gravity-mode oscillations at the solar core to high-energy astrophysical processes far beyond our solar system.

Appendix A

A Search for Trigger-Free Bursts

Detection of neutrinos from a supernova, gamma ray burst, or any other exotic astrophysical object could give us great insight into processes that are not well understood within the interiors of these bodies. Perhaps more importantly, detection of astrophysical neutrinos also provides us with the chance of further illuminating basic neutrino properties, such as the mass hierarchy for instance (normal or inverted). This appendix should serve as a beginning step for establishing the possibility of a trigger-less burst in SNO, and in particular examining the backgrounds to such a search, and developing cuts for their removal.

One of the expected functions of a large underground Cherenkov detector such as SNO is its ability to detect neutrinos from a galactic core-collapse supernova. Though there were no optically detected galactic supernovae during

214

the lifetime of SNO, it is still beneficial to carry out a study of bursts of neutrino interactions in SNO. This is due to the possibility of a supernova occuring in a 'non-standard' manner, in which the emission of neutrinos occurs without the accompaniment of an optical explosion. Also, as has been previously mentioned, there does exist the possibility that a 'standard' [1] core-collapse supernova would be optically obscured from detection, and it would be expected that the cause of optical obscuration (such as dust, or a dark compact object) might not affect a neutrino signal. As such, it is reasonable to search for bursts in SNO's data set even in the absence of a known astrophysical trigger.

A.1   Supernova Neutrinos

The emission of neutrinos is thought to be one of the chief mechanisms for energy loss in core-collapse supernovae. Since neutrinos are neutral and weakly interacting, they are often able to escape conditions in which photons and charged particles are trapped. Such a situation exists in the death of a massive star, or a supernova. There are two main categories of supernovae, Type I and Type II, so designated because of the absence (Type I) or presence (Type II) of hydrogen in their spectral signatures. The most common type of supernova, and incidentally the type most relevant to our studies, is death by way of core-collapse.

The most recent detection of neutrinos from a non-solar astrophysical

---

[1]I put this in quotes because there is much room for uncertainty in the definition of a 'standard' galactic supernova

source occurred over two decades ago with Supernova 1987A, the result of the death of a star in the Large Magellanic Cloud, roughly 52 kpc away. SN1987A was detected by three independent neutrino observatories; Kamiokande II, in Japan [64], IMB (Irvine-Michigan-Brookhaven), in the U.S. [65], and Baksan, in Russia [66], with 19 neutrino events recorded by K-II and IMB. The Baksan events were not conclusively linked to the supernova (5 events, occurring later in time), and as such their results are not typically cited as contributing to the SN1987A data set. Though statistics from this supernova were limited, many studies were performed with the data in the hopes of learning more about the elusive process of piecing together the mechanisms involved in a stellar explosion, as well as neutrino production and propagation in a supernova.

A.1.1 Gravitational Core Collapse Supernovae

The process of the core-collapse supernova begins with a typical star, with nuclear fusion occurring at the stellar core. In all stars, the first stages of nuclear burning consist of fusing hydrogen to helium. The important concept involved here is that of hydrostatic equilibrium, where gravitational pressure incurred by the attraction of mass in the outer shell of the star is fully balanced by the nuclear pressure being generated at the stellar core in fusion reactions. Once hydrogen at the core has been exhausted, nuclear fusion shuts down and the stellar core begins to collapse under the gravitational pressure from the star. This causes an increase in pressure and temperature at the core, which then creates the conditions necessary for the next stage of fusion to begin,

with the fusing of helium into carbon. This process continues through the elements, successively producing oxygen, then magnesium, etc. Massive stars ($\sim$8 times more massive than our sun) generally are capable of carrying out nuclear fusion reactions at their core up to the end point of iron.

Once iron has been created at the core, the balancing act of hydrostatic equilibrium has reached its end; fusion of iron is an endothermic process, and as such the fusion chain has reached a stopping point. When this happens, energy is no longer created by fusion in the stellar core and the nuclear pressure, which until now had counter-balanced the gravitational pressure, no longer serves as a stabilizer. The star, having spent all of its nuclear fuel, begins to implode. (Although electron degeneracy pressure delays the collapse, once the core's mass exceeds the Chandrasekhar limit ($\sim 1.4 M_\odot$), degeneracy pressure is no longer enough to hold off collapse due to the gravitational pressure).

As the implosion occurs, the core of the star begins to heat up very rapidly, igniting nuclear reactions, with copious amounts of neutrinos and neutrons being produced. High-energy gamma rays produced in the heating of the core decompose iron nuclei, resulting in free neutrons. With the heightened pressure and temperature, a period of 'neutronization' occurs, where density conditions allow for electrons and protons to merge via inverse beta decay, producing a huge release of neutrinos in a neutronization burst. As neutrinos are carried away as part of this neutronization burst, an enormous amount of energy is released from the core (roughly 10^{51} ergs, a substantial portion of the total amount of energy released in a supernova). Once the density of the

217

core is greater than that of nuclear matter, the infalling material rebounds, and generates a pressure wave now propagating outward from the core.

A.1.2   Supernova Neutrino Energies and Times

For a gravitational core-collapse supernova, there are many competing models to consider when determining the expected energy and time signatures. The most comprehensive model that takes into account the evolution of the neutrino flux with regard to both time and energy is the often-cited Burrows model [67], which forms the basis for many supernova simulations. Other models (Beacom and Voegel [68], Bruenn-Mezzacappa [69][70]) were developed for specific studies such as supernova timing or relativistic effects, and as such placed more emphasis on modeling one aspect of supernova behavior than on modeling the supernova system as a whole. The Burrows *et al.* model is the most generally appropriate one for the purposes of SNO supernova studies, and it is this model which we use to understand supernova energy and time signatures.

As has been stated already, the neutronization burst is a period in the very beginnings of the supernova when enormous amounts of neutrinos are released in a burst of energy $\sim 10^{51}$ ergs; Figure A.1 shows this burst in terms of neutrino luminosity as a function of time.

The time scale for this neutronization burst, which releases ν_e's only, is expected to be somewhere between roughly 0.01 and 0.1 second, with fall-off up to 1 second, as is shown in Figure A.1.
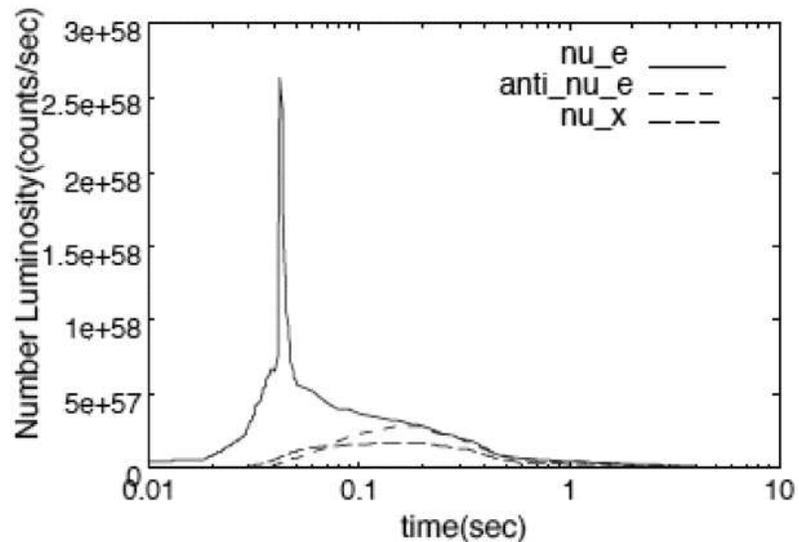
Figure A.1: Neutrino luminosity as a function of time relative to core collapse, assuming no oscillation. Taken from[9]

The original distribution of neutrino energies has been predicted with some accuracy also[2], and are shown in Figure A.2.

These time and energy signatures expected for neutrino emission from a core collapse supernova assure that we make an informed and physically-motivated decision when determining what type of window should be used to scan for bursts in the SNO data set.

[2]I refer to the 'originial' distribution here: in this study we will not be accounting for any oscillation effects, so 'original' should be taken to indicate that matter effects and vacuum oscillations could ultimately have an impact, and the 'original' distribution would therefore be our initial state.
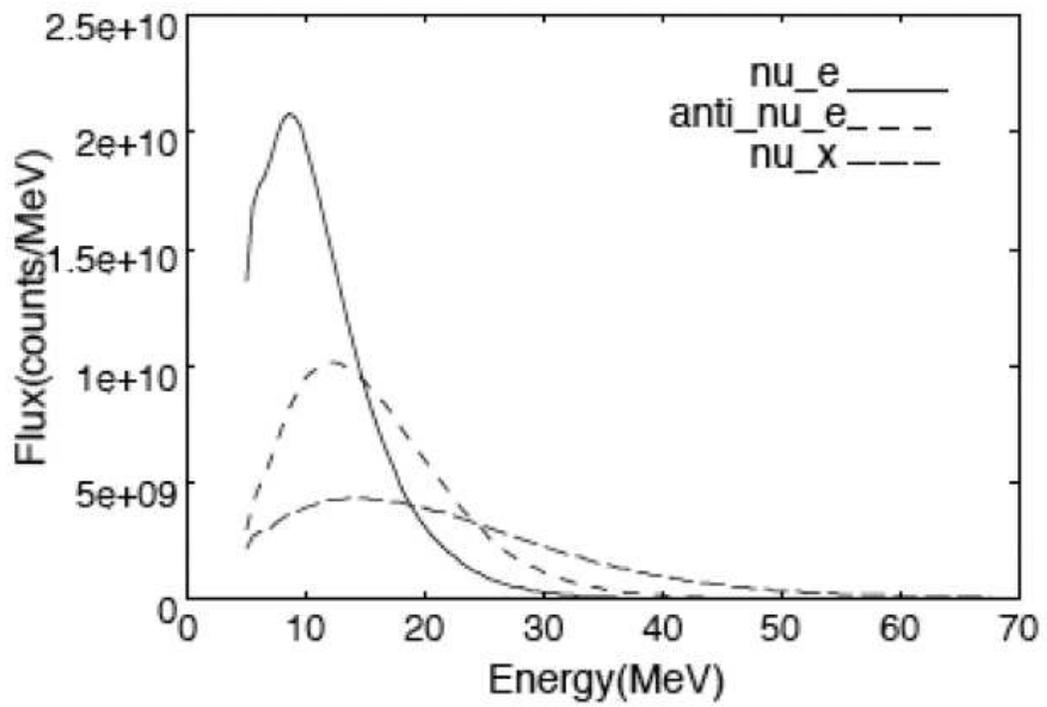
Figure A.2: Expected distribution of original energies for supernova neutrinos, not accounting for oscillation effects. Taken from [9]

A.2 Determining Factors for Window Lengths

In the first steps of determining what region in Energy, window length, and fiducial volume (E, τ, FV) 'space' we are interested in, we are guided by the expected physics of a supernova. We know of the predicted neutronization burst within the first second of supernova explosion, just after core collapse, which occurs over a time scale of less than one second. With such a short time scale, we have the advantage that we can afford to lower the energy threshold substantially and still be able to avoid contamination of our burst search with accidental coincidences.

A more SNO-centric version of the time signature in Figure A.1 is shown in Figure A.3, where the expected number of events in SNO from a Burrows-model supernova is plotted against time since core collapse.

During and directly following the neutronization burst, there is a substantial increase in ν_μ and ν_τ luminosities, in addition to the much larger burst of ν_e's. We are therefore justified in wanting to maximize our ability to detect neutral current events, rather than focusing on a purely charged current signal. Even though the expected supernova neutrino energy is fairly high with regard to SNO's typical solar neutrino analysis energy thresholds, for the neutral current reaction, detection will depend not primarily on the energy of the incoming supernova neutrino, but instead on the neutron capture energy for whichever phase we are inspecting. Because of this, we want to be able to keep the energy threshold for our burst window fairly low, comparable to the neutron capture energy. This is (unfortunately) more important for longer
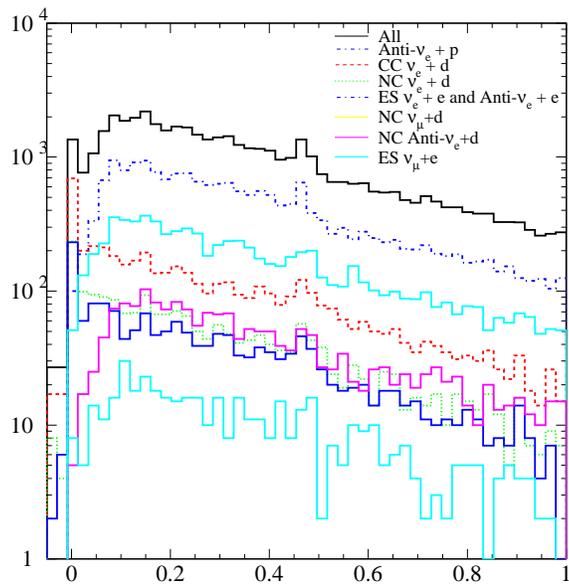
Figure A.3: Expected number of events in SNO from a standard Burrows-model supernova (distance of 10 kpc) as a function of time since core collapse

burst windows (post-neutronization), where ν_μ's and ν_τ's make up a larger fraction of the overall supernova neutrino output.

However, the neutral current signal is not the only reason to drive the energy threshold lower; the charged current signal benefits from a lower energy threshold as well. One reason to try for this comes from the original distribution of neutrino energies, as is shown in Figure A.2.

The energy distribution for ν_e's appears to peak lower than those for the other flavors; this is also not accounting for the fact that in the charged current reaction, we 'lose' 1.442MeV in terms of the energy that we detect;

$$\nu_e + d \to p + p + e^- - 1.442\text{MeV}$$

Thus for SNO we can imagine that the ν_e energy distribution in FigureA.2 are shifted to the right by 1.442MeV, in terms of energy of the detected charged lepton (*i.e.* the detected energy spectrum would be shifted left by 1.442MeV relative to this plot). This all motivates us to try to push the energy threshold as low as is reasonable, so we must find a suitable compromise between low energy and as long of a time window, τ, as we can afford.

A.3 Determining Backgrounds

As there is no astrophysical trigger to guide us in our search for neutrino bursts, we must scan through the SNO data set with great care. The detection of a burst of events in SNO would not in and of itself be reason for claims of a detection of a supernova; there exist multiple sources of backgrounds in and around

the detector which mimic the expected signal of neutrinos from a supernova. One of the main difficulties in performing a trigger-less burst search with SNO, therefore, is the elimination of background burst events. Ultimately, we would like to set a very low threshold in energy, and use the largest possible volume in the detector, and the longest length burst window, in order to optimize our sensitivity to supernovae. The problem we encounter with this approach is obviously that we end up including many more background events, causing an increase in the probability that there will be accidental coincidences, and these accidentals may mimic a supernova burst signal. In order to eliminate false supernova events from our analysis and still have decent sensitivities at low energies, we must find the best compromise between having a low-energy threshold and a long time window in which to test for bursts.

It should be pointed out that in SNO, we have an advantage over other Cherenkov detectors for a trigger-less burst search in that we are able to focus on bursts with low multiplicities (here, multiplicity refers to the number of events comprising a burst; a burst made up of three events would have a multiplicity of three). While Super-Kamiokande (described in Section A.9) required high multiplicities to reduce background contamination in longer time windows, at SNO we can afford to look only at the least restrictive signal condition of two- or three- multiplicity bursts. After deciding on an optimal set of paramaters for a burst search window (low energy, long time window, large fiducial volume), we can calculate what residual backgrounds we will be left with in our search, and then develop effective cuts to help us minimize

these residual backgrounds.

A.3.1 Accidentals

Before doing any fine-tuning of energy thresholds and burst window lengths, the accidental coincidence rates must be calculated. This is done by running through all of the standard SNO data (for the D_2O and Salt phases individually), and simply counting the number of raw events that pass the relevant data cleaning cuts for each phase [3].

This can be translated into the overall rate of events, given the appropriate livetime[4], and then with the equation,

$$R_{coinc} = R_{det}^2 \tau \tag{A.1}$$

where R_{det} is the rate of (clean) events in the detector and τ is the length of the time window in consideration, we find the rate of accidental coincidences, R_{coinc}, for any given event rate in the detector. The event rate will obviously be different for different fiducial volumes, energy thresholds, and window lengths, what the probability of detecting less than one accidental coincidence in the entire span of SNO's data-taking would be, where the number of total expected coincidences in a data set is of course given by

---

[3]For the burst analysis, the standard DAMN mask was used with the exception of the NHIT Burst cut and the Missed Muon Follower Short cut, but including the Muon cut and the Muon Follower Short cut

[4]Since I have used the data processed for the LETA working group, livetime values were taken from the LETA Cheat Sheet, http://deapclean.org/sno/leta_cheat_sheet.html, and were corrected for difference due to the inclusion of NHIT Burst and Missed Muon Follower Short events

$$N_{coinc} = R_{coinc} * T \qquad (A.2)$$

(T is the length of the SNO detector livetime). There are modifications to this equation for higher multiplicity coincidences; specifically, for a three-fold coincidence, the accidental coincidence rate and expected number would be:

$$R_{coinc}{}^{mult=3} = R_{coinc}{}^{mult=2} * R_{det} * \tau = R_{det}^3 * \tau^2 \qquad (A.3)$$

and

$$N_{coinc} = R_{coinc}{}^{mult=3} * T \qquad (A.4)$$

For SNO, our window lengths will be tuned for coincidences with multiplicities of two and three.

Ideally, we are shooting for less than one accidental coincidence in our entire data set, but the difference in numbers of coincidence events passing threshold between a 0.1-second time window and a 1-second time window is significant. To pin down the energy threshold at which we cross over between less than one accidental coincidence event in the entire data set and one or more coincidence events in the entire data set, I scanned through the data and calculated overall event rates, and the resulting accidental event rates for multiple energy and burst window combinations.

In Figures A.4 and A.5, I show the expected number of accidental coincidences (N_{coinc}) for many window-lengths and different energy thresholds (using the standard fiducial volume of r limited to 0-550cm, in agreement with what
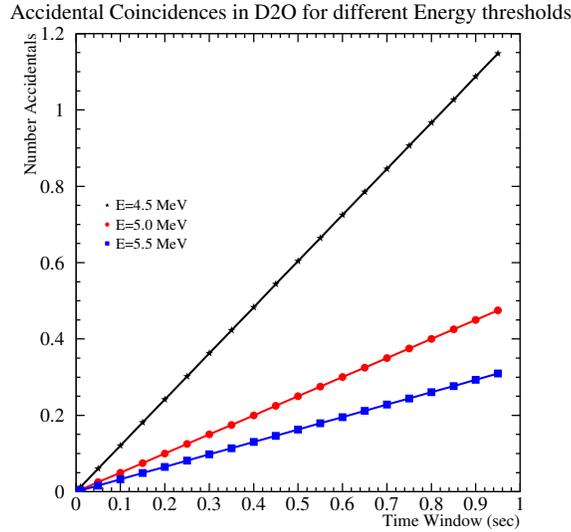
Figure A.4: Expected values of accidental coincidences for the D$_2$O phase at different energy thresholds

is used for the LETA analysis and earlier signal extractions).

If we look initially at the window with fiducial volume $0 - 550$ cm, and energy threshold (total) of 4.5 MeV, we find that in the D$_2$O phase, between t=0.25 sec and t=0.3 sec, for instance, the number of expected accidental coincidences increases from 0.30203 to 0.362432063. If we simply refer again to Figure A.2, we can loosely determine how much ν_x flux we are sacrificing by choosing this energy threshold and in turn, restricting our time window to be less than t=0.3 sec. If the amount that we are sacrificing is significantly more than what we might be gaining by having such a low energy threshold, we should re-evaluate whether the energy threshold should be raised.

One important point to note here is that the numbers which I have cal-
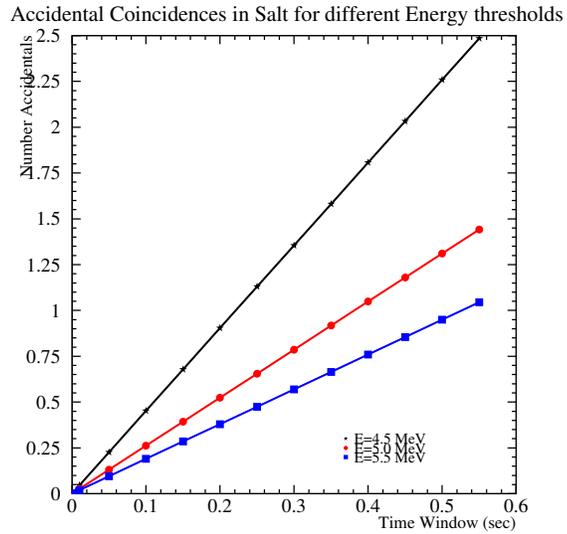
Figure A.5: Expected values of accidental coincidences for the Salt phase at different energy thresholds

culated and plotted in Figures A.4 and A.5 are only the expected values for numbers of accidental coincidences in the data set. The actual upper limit for each of these window lengths will be determined by treating these expected values as a mean in a Poisson process, and extrapolating up to the corresponding upper limit for a Poisson distribution with a mean as given by the N_{coinc}-value we've determined.

If we believe that the values in Figures A.4 and A.5 are the expected number of accidental coincidences for each window, we can determine the probability that there will be exactly k coincidences in the window based on the Poisson

distribution. The probability, of course, is given by

$$f(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{A.5}$$

So if we want less than one coincidence in the entire data set we can start with the first set of conditions (FV 0-550 cm, and $E > 4.5$ MeV, considering only the D_2O-phase data for now), and we calculate the probability that with the value calculated for this set of conditions, and with a window length at t=0.25 sec, again (so λ is 0.302026689) we have $k=0$ coincidences, we find that there is a 73.9% chance of having zero coincidence events in the entire data set. For a higher energy threshold (5.0 MeV), but with that same burst window length (0.25 sec) we find (with λ now at 0.125042841) that there is an 88.25% chance that we will have zero accidental coincidence events in the entire data set. If we push even higher in energy threshold, so we now sit at 5.5 MeV, and with λ now at 0.0814243481, we will have an even higher probability, 92.2%, of having no accidental coincidences. It is clear that the lower the probability of having a single accidental coincidence in our data set, the better. We will use the fairly standard limit of 10% probability of fluctuation as our measure of an acceptable value, which gives us the following values, shown in Table A.3.1 for the D_2O and Salt phases, for multiplicities of both two and three.

Table A.1: Time Windows giving acceptable accidental contamination of SNO data set. (Here, Energy is Total Energy).

Window Length	Multiplicity	Phase	E Thresh
0.2 sec	2	D_2O	5.5 MeV
		Salt	7.0 MeV
0.4 sec	2	D_2O	6.0 MeV
		Salt	7.5 MeV
1.0 sec	2	D_2O	7.0 MeV
20 sec	3	D_2O	4.5 MeV
		Salt	6.0 MeV

A.3.2 Physics Backgrounds

Although accidental coincidences have now been accounted for as a dominant effect on our choice of window length and energy threshold, we still must account for all other 'physics' backgrounds in SNO which could create non-supernova 'burst' events that interfere with the detection of true supernova burst events. The main backgrounds to a trigger-free burst search are events which can create a simulated burst signal, such as a burst of neutrons from fission within the detector, or spallation neutrons following the interaction of an energetic muon in the detector.

Muon spallation is one of the greatest possible backgrounds for the SNO trigger-less burst search: when a muon interacts in the detector, it can break apart a nucleus into small radioactive fragments. These radioactive fragments can decay with the production of a neutron, or a gamma ray above the deuteron photodisintegration threshold (2.2 MeV) which results in a neutron. Regardless of the production mechanism, if more than one neutron is produced as a result of spallation, it can interfere with our search as a burst background.

Although muons generally can be removed from the data set by requiring coincidence with the firing times of the outward-looking tubes (OWLs), there remains a small fraction of muons which are low enough in energy that they do not cause the OWLs to fire. Also, muons due to atmospheric neutrinos interacting in the heavy water will not be tagged by OWLs, and as such we will have no visible precursor to spallation events alerting us to the presence of a muon. The typical muon detected in SNO generally has energies far

greater than what we will be concerned with for our burst search, and as such we can remove a large number of them with a simple energy cut (detailed in Section A.7.2). Also, any muon which has been flagged by our OWL detectors will be removed, and all spallation followers as well. There does still exist a relatively large level of backgrounds due to muons in the detector, and these must be accounted for and removed, as will be discussed in the following sections. The remaining backgrounds we encounter are very much the same as the backgrounds that were dealt with in SNO's anti-neutrino paper[71], and they are expected to come primarily from the following sources:

- Atmospherics/muon spallation

- Spontaneous fission of ^{238}U

- Photodisintegration of the deuteron ($+$ Compton e^-)

- $^{13}C(\alpha, ne^+e^-)^{16}O$

- Instrumental contamination (neck events, Leslies, etc.)

- Reactor antineutrinos

- Geo-antineutrinos

- Diffuse Supernova Background (DSNB)

This is a list of solely coincidence backgrounds, *i.e.* backgrounds with more than one product in a short time window. There are also many other physics backgrounds that do not affect our analysis. Since we are only hurt by fake

bursts, not just background noise, the only way other zero-multiplicity events could hurt us are in the form of accidental coincidences, which have already been outlined in the previous section. Events such as leaked muon spallation ('leaked' = sneaking past the DAMN cuts[5] and atmospherics (since for the burst search we can't use the standard SNO burst-removing 'Missed Muon Follower Short' cut) will be the largest contributors on this list, since they so closely mimic a signal, and as such may be hard to distinguish (and eliminate) from the data.

I have collected in Table A.2 all of these expected number of all backgrounds for roughly our window parameters ($FV < 550cm$, and $E > 5.5MeV$). The sources from which the values are taken are indicated for each field. The numbers for this table come from various different sources, but primarily from [71] and [72]. The estimate for coincidences due to muon spallation after applying the data cleaning cuts is taken from [73], which gives the total amount of muon-induced events in SNO after the Muon-Follower-Short cut (MFLSH), fiducial volume (FV), and energy cuts have been applied. The values I show for atmospheric backgrounds come from an internal analysis done for the LETA group [74].

The fission number for D_2O comes from [71], and we can try to extrapolate a rough number for Salt based on the D_2O value. The method which was used to calculate the value in [71] relied on a measurement of the ^{238}U in the detector after NaCl had already been added, so we can assume the concentration of

_____

[5]DAMN=Data Mask Number, SNO's data cleaning/background removal system

233

^{238}U would be the same for both calculations, however the detection efficiencies would be different, which will significantly change the number of fission backgrounds for Salt. Another source that might be another reliable method for the Salt phase is [75], but here the parameters beng fit are specifically for single-neutron fission numbers. Plots therein seem to indicate, however, roughly 10 fission bursts. I will assume this is a safe upper limit for Salt.

For photodisintegration, we take the coincidence value given in [71] for D_2O, and for Salt, the value comes from Table 21.1 in J. Orrell's thesis [76], where he has provided, based on MC studies of N_{hit} response to 6.25-MeV gamma rays from neutron capture on deuterons, what the number of coincidences due to secondary neutron production ($N_{\gamma,n}$) is. The value taken for this plot is assuming an energy of 5.5 MeV, or NHIT of roughly 45.

The geo- and reactor- neutrino background values come from the analysis in [71] (D_2O), and use as an upper limit (since coincidence events are what would cause a background) the absolute number listed in Table X of [72]. The values calculated for (α, n) in coincidence with a e^+e^- pair come from the analysis done by John Orrell again [76], the results of which are found in Table 20.3 of his thesis. The numbers for $(n, 2n)$ are for now simply assumed to be smaller than or equivalent to those for (α, ne^+e^-), based on the fact that in [72], the (α, ne^+e^-) process is referred to as the 'main source' of coincidence in the calculations of 'Other Backgrounds', counting the $(n, 2n)$ process as a much weaker source of coincidence. However, this only takes into account $(n, 2n)$ induced by fast neutrons due to (α, n) reactions, and presumably does not

Table A.2: Coincidence backgrounds expected for SN Burst Search

Backgrounds	Residuals (D_2O)	Residuals (Salt)
Muon Spallation	(< 0.5)	(< 0.5) [73] [72]
Atmospherics	32.24 [74]	102.34 [74]
Fission (^{238}U)	< 0.79 [71]	< 10 [75] [72]
Photodisintegration	$< 8\text{x}10^{-4}$ [71]	0.41 [72] [76]
geo-ν ($\overline{\nu}$)	0.0 [71]	$< 0.5 \pm 0.1$ [72]
reactor-ν ($\overline{\nu}$)	0.019 ± 0.002 [71]	$< 1.4 \pm 0.3$ [72]
(α, ne^+e^-)	0.02 ± 0.10 [76]	0.07 ± 0.07 [76]
$(n, 2n)$	(< 0.02) [72]	(< 0.07) [72]
DSNB	≤ 0.005 [71]	(≤ 0.005)
Instrumentals	< 0.027 [71]	(< 1) [72]
Total Backgrounds	< 35	< 116

take into account $(n, 2n)$ coincidences induced by photodisintegration neutrons within the detector. This needs to be accounted for, but hopefully will only add a negligible amount of background. The diffuse supernova background (DSNB) for D_2O was taken from [71], and it is assumed that the value given is a high enough upper limit that it is safe to carry over to the Salt phase as well. Lastly, the value for instrumental backgrounds (backgrounds coming from the electronics and PMTs rather than Cherenkov processes in the detector) for D_2O is again taken from [71], and for Salt, the upper limit comes from the total value reported in [72], which will presumably be an upper limit.

The items listed should ultimately have a relatively small impact on the final window decisions, since they will be largely removed by cuts which are developed for this purpose, described in Section A.4.

A.4   Burst Cuts and Search Methods

As was previously mentioned, one difficulty in eliminating backgrounds comes from the fact that the signal that we would expect from supernova neutrinos - two or more events in a short time window - is also present in SNO in the form of unwanted signal, such as fission, or muon spallation products. These events are identified and removed from the SNO published data set, typically with a 'burst' cut, but this cut cannot be used for the burst search, since this would cut *true* bursts in addition to cutting backgrounds.

As was discussed in Section A.1, the predicted energies of neutrinos emitted in core-collapse supernovae vary from the range of ^{8}B solar neutrinos (~ 10 MeV) to much greater energies (~ 70 MeV). Because of this wide range, we cannot afford to use the same energy restrictions as are used in preparing the standard SNO solar neutrino data set. We must therefore come up with a more creative means of removing higher-energy background events such as muons and muon followers which have not been correctly identified as such in SNO's data set.

We can develop alternative cuts, however, which would selectively remove a majority of these backgrounds caused by spallation or fission products (for instance) in the heavy water by following Super-K's example.

A.5   A Δr Cut

Super-K developed a Δr cut which removed events which were localized in a small volume; events from a supernova would be expected to be dispersed throughout the volume of the detector, and therefore we can say with some certainty that events which exhibit a small 'Δr' are not likely to be supernova events. Here, we are defining Δr as a weighted mean of distances between events' reconstructed positions for any candidate 'burst';

$$\Delta r = \frac{\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} |\overrightarrow{r_i} - \overrightarrow{r_j}|}{_MC_2} \tag{A.6}$$

where $|\overrightarrow{r_i} - \overrightarrow{r_j}|$ is the distance between the reconstructed positions of events i and j within a burst, M is the multiplicity of the burst, and $_MC_2$ is the number of non-redundant combinations. An example of expected Δr distributions for supernova events is plotted alongside a sample distribution of Δr for calibration neutrons (Cf source), which we expect to emulate background events.

A fortunate characteristic of these distributions stems from the fact that there is very little overlap in Δr for the expected signal (here, SN MC) and background (here, Cf calibration data). Typically, the backgrounds we would expect will follow this tight distribution in space, and allow us to cut in Δr without losing too large a fraction of our SN signal. I've run a quantitative study on this, to determine where we must place a Δr cut if we are to eliminate 99% of our expected physics backgrounds, (*i.e.* there will be less than one
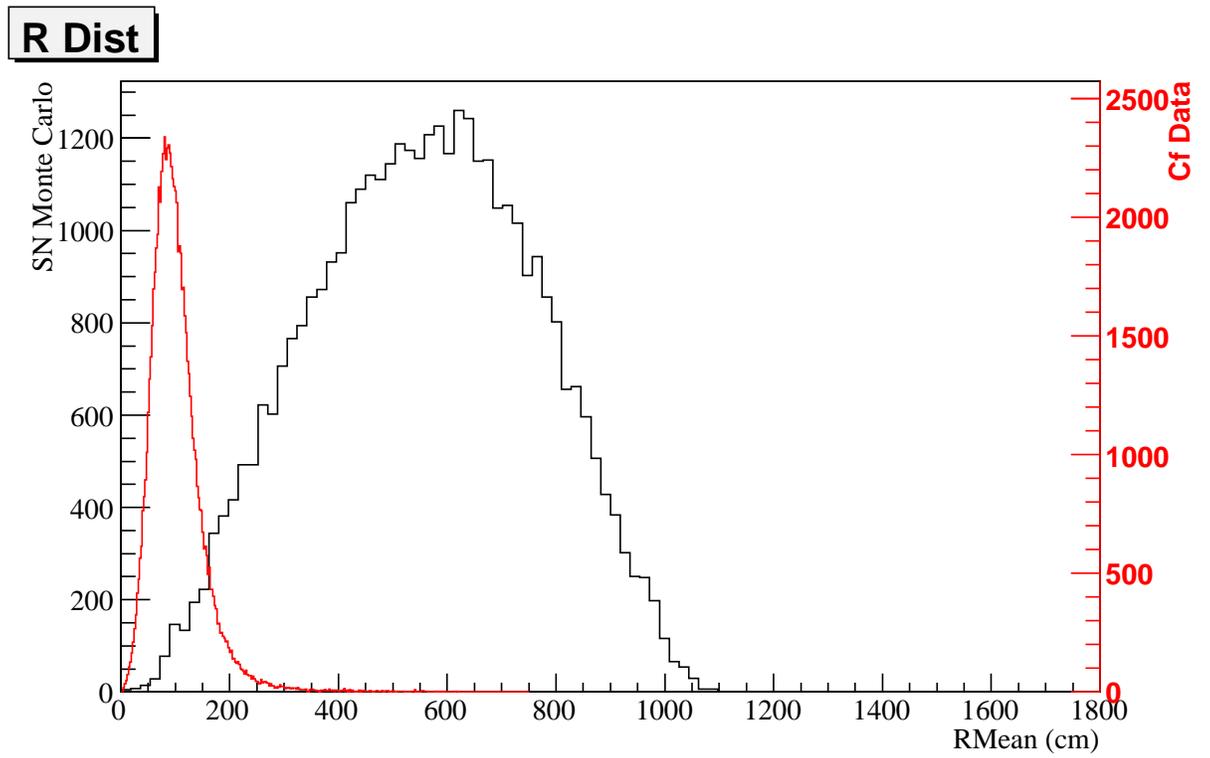
Figure A.6: Sample MC Δr for supernova bursts plotted alongside sample Δr for calibration Cf bursts, Salt phase.
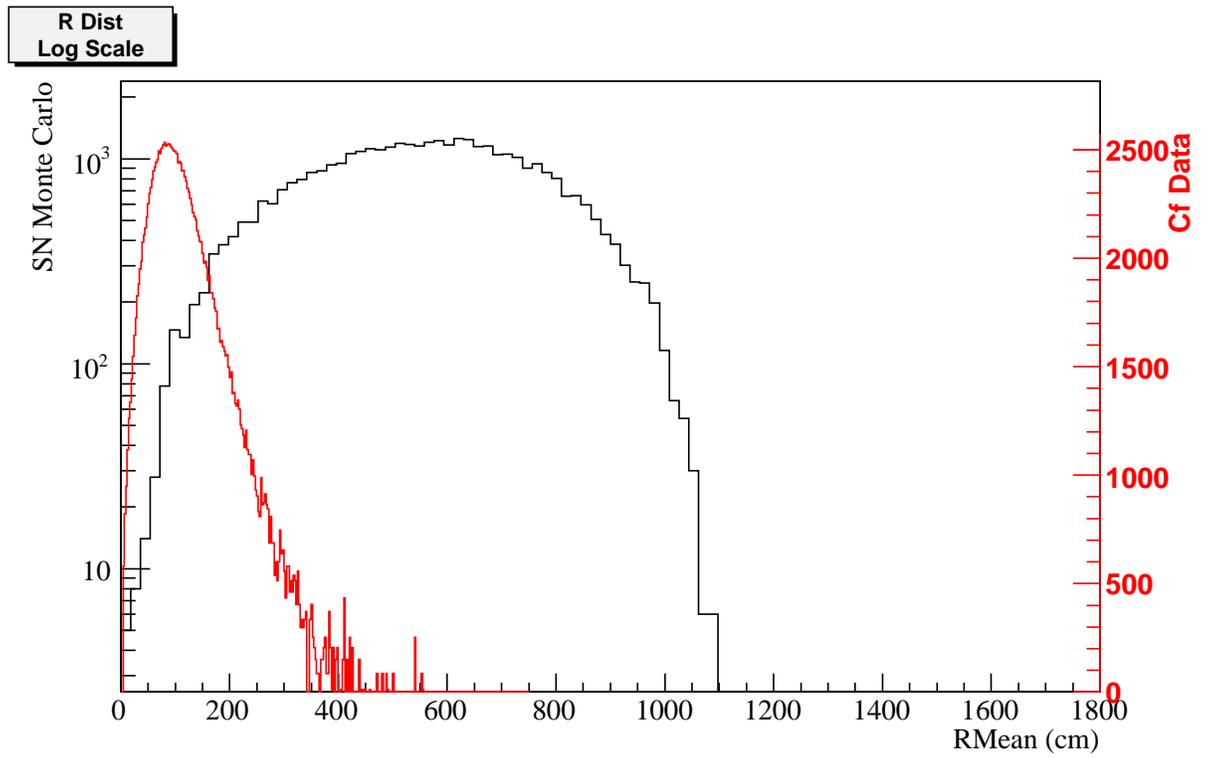
Figure A.7: Sample MC Δr for supernova bursts plotted alongside sample Δr for calibration Cf bursts, Salt phase, Log scale.

Table A.3: Results of signal acceptance (SN MC) and background contamination (Cf Data) for Salt SN MC and Cf Data. (Last column calculated assuming total background number of 116 or lower for Salt).

Δr Cut (cm)	SN Signal Acceptance	Bkgrd Contamination (%)	Bkgrd Contamination (events)
310	87.4%	0.81% ± 0.32	0.93 ± 0.37
320	87.4%	0.72% ± 0.29	0.84 ± 0.34
325	85.4%	0.67% ± 0.26	0.78 ± 0.30
330	85.4%	0.63% ± 0.25	0.74 ± 0.28
335	85.4%	0.60% ± 0.24	0.70 ± 0.28
340	85.4%	0.57% ± 0.24	0.67 ± 0.27

background burst in our final search). I scanned through the SN MC and the Background data (Cf) and tested what percentage of signal made it through a given Δr cut, and what percentage of background contamination remained. The results are in the following table.

If we are anticipating that we must limit the physics backgrounds to produce fewer than one coincidence event in the entire data set, we are restricted to a Δr cut no more relaxed than Δr=330 cm, as this value for Δr sits just at the threshold for letting at most one background event into the signal analysis. However, we also must include in this calculation the contamination from accidental coincidences. We require that the combined background - both accidentals and true physics backgrounds - produce less than one false burst in the entire SNO data set. This appears out to be a fairly restrictive requirement, as has been shown in the previous calculations, and leads us to consider a more restrictive cut.

A.6   A $\Delta r \Delta t$ Cut

It appears that we need to look into more stringent cuts than Δr alone, since the contributions to the background from both accidentals and physics backgrounds are large. One clear option is to restrict events not only by their distribution in space but also by their corresponding differences in time; a Δr-Δt cut, which would allow us to cut more tightly on background-like objects.

$$\Delta t = \frac{\sum_{i=1}^{M} t_i - t_{i-1}}{M} \tag{A.7}$$

The idea behind a Δr-Δt cut comes from the assumption that events which occur as a background product, such as spallation neutrons, will not only be correlated in their relative positions, the amount of distance which exists between them should be correlated with the amount of time difference there is between the events; two neutrons which were released via spallation at the same location in the detector will only have a large Δr value if there has been an appreciable amount of time having elapsed since the time of spallation. Similarly, a very small Δr value with a very small Δt value would accurately describe spallation neutrons, directly following the muon interaction. Conversely, a SN burst will be more likely to have a high Δr value with a small Δt value. Figures A.8 and A.9 show sample distributions of Δr-Δt for background events and simulated supernova events.

The Δr-Δt cut has been developed using a simple parameterization, relat-

RMean vs Delta-t, Mult=2

drvdt2	
Entries	30745
Mean x	103.6
Mean y	5.216
RMS x	65.45
RMS y	6.638
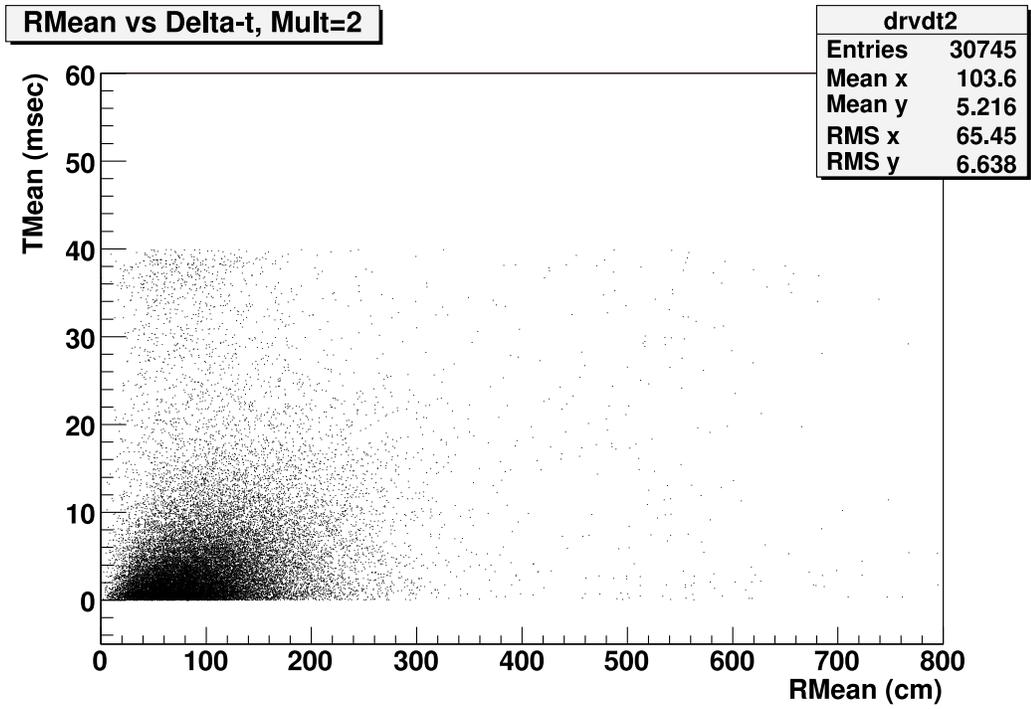
Figure A.8: Sample Δr-Δt distribution for background events.
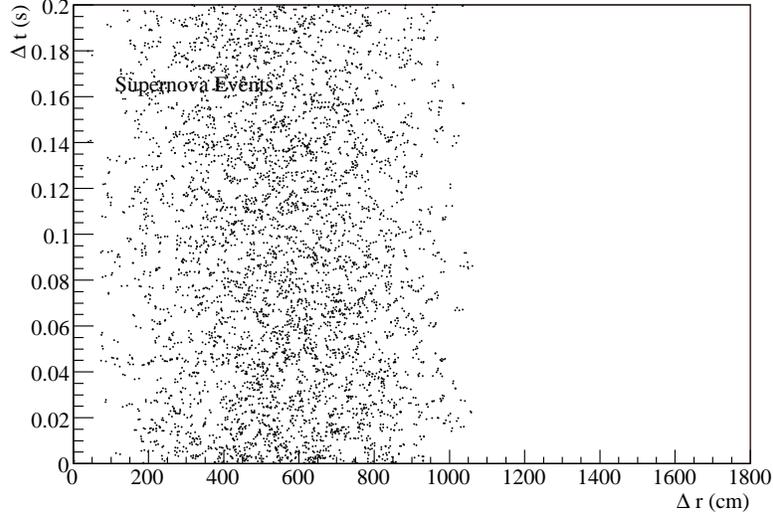
Figure A.9: Sample Δr-Δt distribution for simulated SN events.

ing Δr to Δt by a factor, α, as follows

$$\Delta t = \sqrt{\alpha \Delta r} \qquad \text{(A.8)}$$

For different multiplicities (two and three) and different phases (D_2O and Salt), there will be different relations between the distance the background particles will be separated by, Δr, and their separation in time, Δt, *i.e.* there will be a different value of α for each set of multiplicity and phase we are inspecting, due to the different capture times and detection efficiencies in the relative phases. The values of α for multiplicity-two and multiplicity-three bursts must be fine-tuned to ensure that the percentage of background-type events which are not cut by the Δr-Δt is small enough that we ensure that

243

fewer than one event (out of the phase-specific 35 or 116 estimated in Table A.2) contaminating our final search windows. This means our Δr-Δt cut will have to be more than 99% efficient for the Salt phase, and more than 97% efficient for the D$_2$O phase. Though the Δr-Δt cut still needs tuning, different values for α have been tested, and it appears that we will have to sacrifice a large amount of signal even with this stringent cut. Further analysis on the best implementation of this cut should be explored.

A.7 Other Cuts

A.7.1 Window Length

This section differs from the previous two in that it details the window length, which is not truly a cut, but it is the parameter which is used to eliminate accidentals from the data set; therefore, while the previous cuts (Δr and $\Delta r \Delta t$) were designed to eliminated backgrounds due to physical processes in the detector, the window length is determined with the intention of eliminating backgrounds due to unavoidable accidental coincidences.

A.7.2 High-NHIT Follower (HNF) Cut: Upper Energy Threshold

As has been discussed earlier, the window of expected energies for supernova neutrinos has limits based on our understanding of the conditions in a supernova. Because of this, we are safe in assuming that any event detected in SNO

with an energy above the upper threshold of expected neutrino energy, even if part of a burst, will not be a supernova candidate. This helps us eliminate atmospheric backgrounds from our data set, as the typical energy for atmospherics ($\sim$GeV) will be much greater than the typical energy for supernova neutrinos. The energy cut relies on the relation between the energy of the detected particle and the number of PMTs hit (NHIT) for any event. By using a ratio of roughly 8:1 (PMTs hit: MeV energy - this ratio will differ between phases), we can rely on a very basic high-NHIT follower (HNF) cut to help remove high-energy burst candidates that are too high in energy to be due to a supernova.

For the HNF cut to work, it must not only remove events which are high in energy themselves, but also remove events which follow the high-NHIT primary event, assuming that any followers of a high-energy event are likely to be spallation neutrons or other background-induced followers (photodisintegration gammas, etc.). We will assume that four neutron capture times is adequate to avoid inclusion of high-energy-event followers in our burst search, therefore the cut is structured (much like the muon follower cut) to flag any high-NHIT event and all events following within a window of four neutron capture times (600msec for D_2O phase, 200 msec for Salt phase).

By placing an upper limit on energy as well as the already-present lower limit energy threshold which we have placed on our window (to eliminate accidentals), we have limited the number of contributing backgrounds which must be reduced. One step which should be considered, therefore, before final

245

cut parameters are determined, is what the expected distribution of energies for the atmospheric bursts would be. If the energy (NHIT) of the expected bursts has a high mean, and a significant fraction of these events would be cut by our HNF Cut, we are allowed to be less restrictive in our Δr-Δt cut, which would be highly desirable in the context of our sensitivity and sacrifice measurements. This is something which should be considered before finalization of the Δr-Δt cut.

A.7.3 Non-Electron Follower (NEF) Cut

One other possible background which must be eliminated is that of a high-energy event in SNO, such as a pion, which misreconstructs as a low-energy event due to a non-standard hit pattern. An example of such an event is shown in Figure A.10, in which a high-energy, double-ring event interacted in the heavy water of SNO, and was actually misreconstructed by our path fitters as a single low-energy event, due to its odd hit pattern.

In order to attempt to remove such events from our data set, we can develop a cut, which we call the non-electron follower (NEF) cut (for followers of non-electron-like events), which identifies specific characteristics of non-electron-like events and tags all events within a certain time window of any such non-electron-like event. The manner in which we identify whether an event is electron-like or not is by using some specific high-level cuts (HLCs) which have been developed and tested for the LETA analysis [77]. The cuts which we use for this purpose are described in [77]; they are the β_{14} isotropy cut and
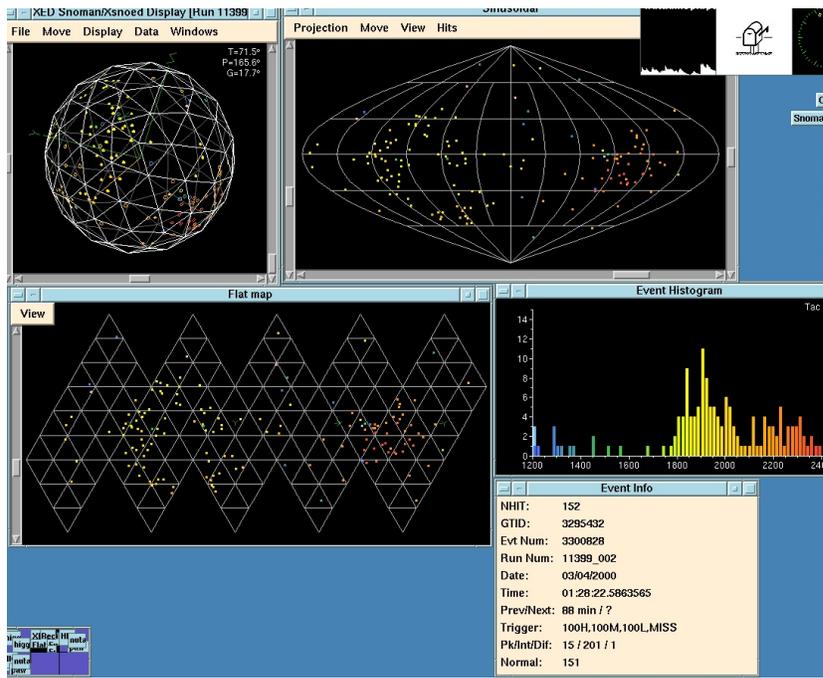
Figure A.10: Example of a high-energy event in SNO which resulted in the fitting algorithms' misreconstruction of the event's energy.

the angular fitter figure-of-merit cuts (FTP FOMs). The β_{14} cut was designed for the separation of electrons from neutrons in the Salt phase, and is a linear combination of the average value of the first and fourth Legendre polynomials of the cosine of the angle between each pair of PMTs hit in an event. This gives us a measure of the degree of isotropy of the PMTs hit in any given event: the β_{14} of electron-like events is somewhat distinguishable from that of neutrons, or non-electron-type events, and as such we use this as one of our criteria for identifying non-electrons for our NEF cut.

The other HLCs which we use in the NEF cut are the angular FTP FOMs, which are an assortment of parameters which are essentially K-S tests on the angular distribution of Cherenkov light of reconstructed events, with the hypothesis that each event is a single Cherenkov electron (with energy of 5 MeV). Events which fail these FTP FOM cuts, obviously, are taken to be not well-described as 5-MeV Cherenkov electrons, and as such they serve as another foundation of our NEF cut.

As the combination of the FTP FOM cuts and the β_{14} cut together should give a fairly good indication of an event which would possibly be misreconstructed as a low-energy event, when in reality it's a background-inducing high-energy event, we have built our NEF cut to trigger on any event which fails either of these two checks, and then we flag (and ultimately cut) any event which follows within 4 neutron capture times of the primary event (similar to the HNF cut). Initial tests show that this cut is only slightly more than 30% effective at eliminating atmospheric bursts.

A.8 Remaining Analyses

Given the array of cuts outlined here, it seems that the next step for the SNO collaboration will be to test the sacrifice of each cut, and determine SNO's ultimate sensitivity for detection after ascertaining the contamination and sacrifice due to the combination of all cuts. The specific cuts listed here must be further tested against Monte Carlo to ensure a complete accounting for, and removal of, all expected backgrounds. Once the cuts have been developed further, the sacrifices must be understood, and a final sensitivity estimate for SNO's trigger-less burst search must be made. Pending final cut-tuning and review from the SNO Topic Committee, this analysis should be carried out on the SNO data sets.

A.9 Super-Kamiokande Search for Bursts with No Trigger

The Super-Kamiokande Collaboration has also carried out a search for bursts in their neutrino data set with no optical trigger, and published their method and results in [78]. The data set for this search spanned the calendar times of May, 1996, to July, 2001, and then December, 2002, to October, 2005. Their method of search consisted of sliding a pre-determined window through the data set sequentially and noting if any window contains a burst. They then revisited any detected bursts to determine whether the bursts were caused by backgrounds or a real signal. This is a markedly 'un-blind' analysis, and SNO

would prefer to avoid using this approach, by instead performing the 'inverse' of the Super-K search; determine a pre-set window, calculate the expected number of backgrounds in this window, develop methods of ensuring the removal of such backgrounds, and then finally rely on our 'blind' background removal techniques while scanning through the data with our pre-set window, with the understanding that any 'burst' detected in this final search would then be classified as due to a non-background signal.

Super-K used three separate time and energy windows in their burst search. The first window was optimized for detecting a distant supernova, where here, 'distant' means hundreds of kpc away from our galaxy. For this search, they used a low-multiplicity threshold (≥ 2 events) and a long time window (20 seconds). Because accidental coincidences and solar neutrino events would swamp such a window, they set a high energy threshold, of 17 MeV, for this search. The second window used in the Super-K study was intended for achieving better sensitivity for lower-energy events from nearby supernovae. Because Super-K would have far higher backgrounds at lower energies than were present in the previous distant-supernova window, they increased the required multiplicity for this window, and decreased the length in time of the window. The resulting combinations of windows for this portion of the search were then ≥ 3 events in 0.5 seconds, ≥ 4 events in 2.0 seconds, and ≥ 8 events in 10 seconds. The energy threshold for the first phase of Super-K (SK-I) was 6.5 MeV and for the second phase (SK-II) was 7.0 MeV. The third and final window used in the Super-K burst search was directed at detection of neutrinos emitted during

the neutronization burst expected in a core-collapse supernova. The window parameters for this were times of 1, 10 and 100 milliseconds, with multiplicity of ≥ 2 events for all times. The Super-K collaboration found no evidence of core-collapse supernovae during any of the windows used, over the period of 2589.2 live days which they considered.

Bibliography

[1] G. D. Orebi Gann, "An Improved Measurement of the ^{8}B Solar Neutrino Energy Spectrum at the Sudbury Neutrino Observatory." PhD thesis, University of Oxford, 2008.

[2] J. N. Bahcall, A. M. Serenelli, and S. Basu, "New solar opacities, abundances, helioseismology, and neutrino fluxes," *Astrophys. J.* **621** (2005) L85–L88, `arXiv:astro-ph/0412440`.

[3] J. Christensen-Dalsgaard, "Lecture Notes on Stellar Oscillations,". 5th edition, on the web.

[4] **SNO** Collaboration, J. Boger *et al.*, "The Sudbury Neutrino Observatory," *Nucl. Instrum. Meth.* **A449** (2000) 172–207, `arXiv:nucl-ex/9910016`.

[5] M. Dunford, "Measurement of the (8)B solar neutrino energy spectrum at the Sudbury Neutrino Observatory,". Ph.D. thesis, University of Pennsylvania.

[6] C. Kyba, "Measurement of the atmospheric neutrino induced muon flux at the Sudbury Neutrino Observatory,". Ph.D. thesis, University of Pennsylvania.

[7] J. Wilson, "A Measurement of the 8B Solar Neutrino Energy Spectrum at the Sudbury Neutrino Observatory,". D.Phil. thesis, University of Oxford.

[8] **SNO** Collaboration, B. Aharmim *et al.*, "A search for periodicities in the B-8 solar neutrino flux measured by the Sudbury Neutrino Observatory," *Phys. Rev.* **D72** (2005) 052010, `arXiv:hep-ex/0507079`.

[9] K. e. Takahashi, "Effects of neutrino oscillation on the supernova neutrino spectrum," *Phys. Rev. D* **64** (2001) 093004.

[10] W. Pauli, "Letter to Physicists at Tubingen," *Translated in Inward Bound, A. Pais.* (1930).

[11] C. L. Cowan, Jr., F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, "Detection of the Free Neutrino: A Confirmation," *Science* **124** (July, 1956) 103–104.

[12] D. Caldwell, *Current Aspects of Neutrino Physics.* Springer Press, NY, 2001.

[13] **SNO** Collaboration, Q. R. Ahmad *et al.*, "Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury

Neutrino Observatory," *Phys. Rev. Lett.* **89** (2002) 011301, `arXiv:nucl-ex/0204008`.

[14] **Super-Kamiokande** Collaboration, Y. Fukuda *et al.*, "Evidence for oscillation of atmospheric neutrinos," *Phys. Rev. Lett.* **81** (1998) 1562–1567, `arXiv:hep-ex/9807003`.

[15] **KamLAND** Collaboration, K. Eguchi *et al.*, "First results from KamLAND: Evidence for reactor anti- neutrino disappearance," *Phys. Rev. Lett.* **90** (2003) 021802, `arXiv:hep-ex/0212021`.

[16] **MINOS** Collaboration, D. G. Michael *et al.*, "Observation of muon neutrino disappearance with the MINOS detectors and the NuMI neutrino beam," *Phys. Rev. Lett.* **97** (2006) 191801, `arXiv:hep-ex/0607088`.

[17] S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.* **19** (1967), no. 21, 1264–1266.

[18] J. N. Bahcall and R. K. Ulrich, "Solar models, neutrino experiments, and helioseismology," *Rev. Mod. Phys.* **60** (Apr, 1988) 297–372.

[19] J. N. Bahcall, M. H. Pinsonneault, and S. Basu, "Solar models: Current epoch and time dependences, neutrinos, and helioseismological properties," *Astrophys. J.* **555** (2001) 990–1012, `arXiv:astro-ph/0010346`.

[20] R. Davis, "Attempt to Detect the Antineutrinos from a Nuclear Reactor by the $CL^{37}(\overline{\nu}, e^-)A^{37}$ Reaction," *Phys. Rev.* **97** (1955) 766–769.

[21] M. Fukugita and T. Yanagida, *Physics of Neutrinos and Applications to Astrophysics.* Springer Press, NY, 2003.

[22] P. V. F. Boehm, *Physics of Massive Neutrinos.* Cambridge Press, NY, 1982.

[23] S. Mikheev and A. Smirnov, "Resonant amplification of neutrino oscillations in matter and solar neutrino spectroscopy," *Nuovo Cim.* **C9** (1986) 17–26.

[24] L. Wolfenstein, "Neutrino oscillations in matter," *Phys. Rev.* **D17** (1978) 2369.

[25] P. Sturrock, "Joint time series analysis of Homestake and GALLEX-GNO solar neutrino data: Evidence for Rieger type periodicities and their interpretation as R mode oscillations," *Astrophys. J.* **594** (2003) 1102.

[26] P. Sturrock, "Analysis of super-Kamiokande five day measurements of the solar neutrino flux," *Astrophys. J.* **605** (2004) 568.

[27] A. Milsztajn, "A search for periodicity in the Super-Kamiokande solar neutrino flux data,". hep-ph/0301252.

[28] D. Caldwell and P. Sturrock, "Evidence for solar neutrino flux variability and its implications," *Astropart. Phys.* **23** (2005) 543–556.

[29] P. Sturrock, "Comparative analysis of super-Kamiokande and SNO solar-neutrino data and the photospheric magnetic field,". e-Print: hep-ph/0610065.

[30] P. Sturrock, "Solar neutrino variability and its implications for solar physics and neutrino physics," *Astrophys. J.* **688** (2008) 12.

[31] P. Sturrock and J. Scargle, "Histogram Analysis of GALLEX, GNO, and SAGE Neutrino Data: Further Evidence for Variability of the Solar Neutrino Flux," *Astrophys. J. Lett.* **550** (2001) L101.

[32] J. Y. *et al.*, "A search for periodic modulations of the solar neutrino flux in super-kamiokande-i," *Phys. Rev. D* **68** (2003) 092002.

[33] L. P. (for the GNO Collaboration), "Search for time modulations in the Gallex/GNO solar neutrino data," *Astropart. Phys.* **22** (2004) 219.

[34] P. S. *et al.*, "Comparative analysis of Gallex and GNO solar neutrino data," *Astropart. Phys.* **26** (2006) 174–185.

[35] T. Araki, S. Enomoto, K. Furuno, Y. Gando, K. Ichimura, H. Ikeda, K. Inoue, Y. Kishimoto, M. Koga, Y. Koseki, T. Maeda, T. Mitsui, M. Motoki, K. Nakajima, H. Ogawa, M. Ogawa, K. Owada, J.-S. Ricol, I. Shimizu, J. Shirai, F. Suekane, A. Suzuki, K. Tada, S. Takeuchi, K. Tamae, Y. Tsuda, H. Watanabe, J. Busenitz, T. Classen, Z. Djurcic, G. Keefer, D. Leonard, A. Piepke, E. Yakushev, B. E. Berger, Y. D. Chan, M. P. Decowski, D. A. Dwyer, S. J. Freedman, B. K. Fujikawa,

J. Goldman, F. Gray, K. M. Heeger, L. Hsu, K. T. Lesko, K.-B. Luk, H. Murayama, T. O'Donnell, A. W. P. Poon, H. M. Steiner, L. A. Winslow, C. Mauger, R. D. McKeown, P. Vogel, C. E. Lane, T. Miletic, G. Guillian, J. G. Learned, J. Maricic, S. Matsuno, S. Pakvasa, G. A. Horton-Smith, S. Dazeley, S. Hatakeyama, A. Rojas, R. Svoboda, B. D. Dieterle, J. Detwiler, G. Gratta, K. Ishii, N. Tolich, Y. Uchida, M. Batygov, W. Bugg, Y. Efremenko, Y. Kamyshkov, A. Kozlov, Y. Nakamura, H. J. Karwowski, D. M. Markoff, K. Nakamura, R. M. Rohm, W. Tornow, R. Wendell, M.-J. Chen, Y.-F. Wang, and F. Piquemal, "Experimental investigation of geologically produced antineutrinos with KamLAND," *Nature* **436** (July, 2005) 499–503.

[36] **KamLAND** Collaboration, S. Abe *et al.*, "Precision Measurement of Neutrino Oscillation Parameters with KamLAND," `arXiv:0801.4589 [hep-ex]`.

[37] R. Ulrich, "The Five-Minute Oscillations on the Solar Surface," *Astrophys. J.* **162** (1970) 993.

[38] J. Leibacher and R. Stein, "A New Description of the Solar Five-Minute Oscillation," *Astrophys. J. Lett.* **7** (1970) 191.

[39] W. Unno *et al.*, *Nonradial Oscillations of Stars.* Univ. of Tokyo Press, 1979.

[40] **SNO** Collaboration, B. Aharmim *et al.*, "Measurement of the nu/e and total B-8 solar neutrino fluxes with the Sudbury Neutrino Observatory

phase I data set," *Phys. Rev.* **C75** (2007) 045502,
`arXiv:nucl-ex/0610020`.

[41] J. F. Amsbaugh *et al.*, "An array of low-background ^{3}He proportional counters for the Sudbury Neutrino Observatory," *Nucl. Instrum. Meth.* **A579** (2007) 1054–1080, `arXiv:0705.3665 [nucl-ex]`.

[42] J. R. Klein, M. S. Neubauer, R. V. Berg, and F. M. Newcomer, "The SNO Trigger System,". SNO technical report SNO-STR-97-036, University of Pennsylvania.

[43] James Cameron, "The Photomultiplier Tube Calibration of the Sudbury Neutrino Observatory,". PhD thesis, University of Oxford.

[44] R. Ford, "Calibration of SNO for the detection of ^{8}B neutrinos." PhD thesis, Queen's University, 1998.

[45] M. R. Dragowsky *et al.*, "The N-16 calibration source for the Sudbury Neutrino Observatory," *Nucl. Instrum. Meth.* **A481** (2002) 284–296, `arXiv:nucl-ex/0109011`.

[46] G. McGregor, "HCA Effects Summary,". Technical Report, Sudbury Neutrino Observatory.

[47] M. Dunford, "A Measure of Threshold Uncertainties,". Technical Report, Sudbury Neutrino Observatory.

[48] A. Anthony, "New Rate-based Studies of Hi/Lo Halfpts,". Technical Report, Sudbury Neutrino Observatory.

[49] L. Heelan, "The Lomb-Scargle Periodicity Analysis of SNO's D2O and Salt Phases,". `http://manhattan.sno.laurentian.ca/sno/ananoteb.nsf/URL/MANN-6ABN8H`. Technical Report, Sudbury Neutrino Observatory.

[50] T. Costin and S. Oser, "Results From An Unbinned Maximum Likelihood Time Series Analysis of SNO's D2O, Salt, and Combined Data Sets,". `http://manhattan.sno.laurentian.ca/sno/ananoteb.nsf/URL/MANN-6AHEY6`. Technical Report, Sudbury Neutrino Observatory.

[51] K. Mardia and P. Jupp., *Directional Statistics*. John Wiley and Sons, Ltd., 2000.

[52] S. Oser, "Analytically Predicting the Rayleigh Power Distribution,". Technical Report, Sudbury Neutrino Observatory.

[53] S. T.-C. *et al.*, "Looking for Gravity-Mode Multiplets with the GOLF Experiment aboard *SOHO*," *Astrophys. J.* **604** (2004) 455–468.

[54] R. G. *et al.*, "Tracking Solar Gravity Modes: The Dynamics of the Solar Core," *Science* **316** (2007) 1591–1593.

[55] R. G. *et al.*, "Detection of periodic signatures in the solar power spectrum: on the track of l=1 gravity modes,". Proceedings SOHO 18/GONG 2006, arXiv:0611806 [astro-ph].

[56] J. Bahcall and P. Kumar, "g-modes and the Solar Neutrino Problem," *Astrophys. J.* **409** (1993) L73–L76.

[57] C. B. *et al.*, " Large-Mixing Oscillations as a Probe of the Deep Solar Interior," *Astrophys. J. Lett.* **588** (2003) L65L68.

[58] P. B. *et al.*, "Neutrino Propagation Through Helioseismic Waves," *Nucl. Phys.* **B513** (1998) 319–342.

[59] S. Couvidat, S. Turck-Chieze, and Kosovichev, "Solar Seismic Models and the Neutrino Predictions," *Astrophys. J.* **599** (2003) 1434–1448.

[60] S. Turck-Chieze, "Solar Gravity Modes: Present and Future," in *Advance Space Research: COSPAR 2004 Proceedings.* 2004. `arxiv:astro-ph/0511126v1`.

[61] S. M. *et al.*, "On the Characteristics of the Solar Gravity Mode Frequencies," *Astrophys. J.* **668** (2007) 594–602.

[62] A. Jimenez and R. Garcia, "On the solar origin of the 220.7 signal,". arXiv:0810.1706 [astro-ph].

[63] R. G. *et al.*, "Update on g-mode Research," *Astron. Nachr. 999* **88** (2006) 1–9.

[64] K. H. *et al.*, "Observation of a neutrino burst from the supernova SN1987A," *Phys. Rev. Lett.* **58** (1987) 1490–1493.

[65] R. B. *et al.* (The IMB Collaboration), "Observation of a neutrino burst in coincidence with Supernova 1987A in the Large Magellanic Cloud," *Phys. Rev. Lett.* **58** (1987) 1494.

[66] E. A. *et al.*, "Possible Detection of a Neutrino Signal on 23 February 1987 at the Baksan Underground Scintillation Telescope of the Institute of Nuclear Research," *JETP Lett.* **45** (1987) 589–592.

[67] R. G. A. Burrows, D. Klein, "The future of supernova neutrino detection," *Phys. Rev.* **D45** (1992) 3361–3386.

[68] J.Beacom and P.Vogel, "Can a supernova be located by its neutrinos," *Phys. Rev.* **D60** (1999) 33007.

[69] S. Bruenn, "Neutrinos from SN1987A and current models of stellar-core collapse," *Physical Review Letters* **59** (1987) 938–941.

[70] A. M. S. Bruenn, K. De Nisco, "General Relativistic effects in the core collapse supernova mechanism," *online* (2001). astro-ph/0101400.

[71] **SNO** Collaboration, B. Aharmim *et al.*, "Electron antineutrino search at the Sudbury Neutrino Observatory," *Phys. Rev.* **D70** (2004) 093014, `arXiv:hep-ex/0407029`.

[72] **SNO** Collaboration, B. Aharmim *et al.*, "Electron energy spectra, fluxes, and day-night asymmetries of B-8 solar neutrinos from the 391-day salt phase SNO data set," *Phys. Rev.* **C72** (2005) 055502, `arXiv:nucl-ex/0502021`.

[73] A. Marino, "Evidence for Neutrino Oscillations in the Sudbury Neutrino Observatory,". Ph.D. Thesis, University of California at Berkeley.

[74] F. Zhang, "A Monte Carlo study of the ν_{atm} background in the full leta salt and d2o dataset,". Technical Report, Sudbury Neutrino Observatory.

[75] A. Hallin, "Limits on Fission,". Technical Report, Sudbury Neutrino Observatory.

[76] J. Orrel, "A Search for an Electron Antineutrino Signal in the Sudbury Neutrino Observatory,". Ph.D. Thesis, University of Washington.

[77] J. R. Klein, "Tuning the High Level Cuts for LETA." Technical report, University of Texas, 2006.

[78] M. I. *et al.* (Super-Kamiokande Collaboration), "Search for Supernova Neutrino Bursts at Super-Kamiokande," *Astrophys. J.* **669** (2007) 519–524.

Vita

Aubra Elan Anthony attended Richardson High School in Richardson, Texas, graduating in the Spring of 1997. In the Fall of 1997, she began undergraduate coursework at the University of Texas at Austin. She earned a Bachelor of Science degree in Physics in 2002, as well as Bachelors of Arts in both Astronomy and French. She enrolled as a physics graduate student at the University of Texas in Fall of 2003.

Permanent Address: 13505 Lone Rider Trl

Austin, TX

78738

This dissertation was typeset with LaTeX 2_ε[6] by the author.

---

[6] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin, and extended by Bert Kay, James A. Bednar, and Ayman El-Khashab.