University of Alberta

A Search for *hep* Neutrinos with the Sudbury Neutrino Observatory Detector

by

Christopher William Howard

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Physics

Fall 2010
Edmonton, Alberta

Examining Committee

Dr. Aksel Hallin, Physics

Dr. Doug Gingrich, Physics

Dr. Kevin Beach, Physics

Dr. James McMullin, Electrical and Computer Engineering

Dr. Robert McPherson, Physics and Astronomy, University of Victoria

Abstract

This thesis focuses on the search for neutrinos from the solar *hep* reaction using the combined three phases of the Sudbury Neutrino Observatory (SNO) data. The data were taken over the years 1999–2006, totalling 1,083 days of live neutrino time.

The previous published SNO *hep* neutrino search was completed in 2001 and only included the first phase of data taking. That *hep* search used an event counting approach in one energy bin with no energy spectral information included. This thesis will use a spectral analysis approach.

The *hep* neutrino search will be a Bayesian analysis using Markov Chain Monte Carlo (MCMC), and a Metropolis-Hastings algorithm to sample the likelihood space. The method allows us to determine the best fit values for the parameters. This signal extraction will measure the ^{8}B flux, the atmospheric neutrino background rate in the SNO detector, and the *hep* flux.

This thesis describes the tests used to verify the MCMC algorithm and signal extraction. It defines the systematic uncertainties and how they were accounted for in the fit. It also shows the correlations between all of the parameters and the effect of each systematic uncertainty on the result.

The three phase *hep* signal extraction was completed using only 1/3 of the full data set. With these lowered statistics, this analysis was able to place an upper limit on the *hep* flux of $4.2 \times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ with a 90% confidence limit. It was able to measure a *hep* flux of $(2.40^{+1.19}_{-1.60}) \times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. These numbers can be compared with the previous SNO upper limit of $2.3 \times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ with a 90% confidence limit, and the standard solar model prediction of $(7.970 \pm 1.236) \times 10^3\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

Acknowledgements

I would like to thank all my friends and family, whose contributions are far to great to try and list here. It is so wonderful to finally be here, at the end of this long ride, we call school. I am looking forward to my journey from here on out.

This trip would not have been possible if it wasn't for the guidance of Aksel Hallin. His never ending patience and ability to make the the most difficult of concepts accessible, has been a blessing. He has taught me a whole new way to approach problems, and has been an example of fantastic teacher and a great physicist.

Even though I made the switch from the east to the west during my time as PhD student, I am grateful for the opportunity to meet new people. Of course, that switch led me to meeting, then marrying, Andrea.

Andrea, you have made my time as PhD student much more enjoyable! I really appreciated the support, and advice you provided. Even the conversations we would have about my work, they meant a lot, especially since I am quite sure you had no idea what I was talking about. It is your support that gave me the motivation to find the time to write this thesis, and your encouragement that drove me to find future opportunities.

Finally, I would like to thank my parents, Steve and Cathy, and my brother, Greg. You have given me unending support over the years, and it has always been appreciated. Whether it was a long phone call just to talk, or a funny e-mail that raised my spirits. Also, the constant asking if I was done yet was really helpful...

Thanks everyone, I made it!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

PP chain	Proton-Proton chain
CNO cycle	Carbon-Nitrogen-Oxygen cycle
SNO	Sudbury Neutrino Observatory
AV	Acrylic Vessel
PMT	Photo Multiplier Tube
PSUP	Photo multiplier SUPport structure
SSM	Standard Solar Model
NCD	Neutral Current Detector
RSP	energy ReSponse Processor
SNOMAN	SNO Monte carlo and ANalysis software
MC	Monte Carlo
PDF	Probability Distribution Function
MCMC	Markov Chain Monte Carlo
ACF	Auto-Correlation Function
LETA	Low Energy Threshold Analysis

Chapter 1

Introduction

1.1 The Sun and Solar Neutrinos

The Sun is our closest star, and wondering how it works has always puzzled humanity. There is a large field of scientists who work to unravel the mystery of the Sun. Astrophysicists have made theoretical predictions (or models) of how the Sun behaves, and have compared their predictions to observational information.

Most observations come in the form of optical, radio, and helioseismic data. This information has provided a good basis for comparison to solar models. Another source of information is neutrinos that are created from the solar nuclear fusion.

Neutrinos rarely interact with matter, so those produced in the core contain information about the core, unlike photons which can take thousands of years to travel from the core to the surface of the star, interacting millions of times, leaving them with no information about how, or when they were produced. This means that neutrinos are the best way we have to obtain direct information about the solar interior.

The following section will explain the production of neutrinos that can be used to extract information about the Sun and test the *standard solar model* (SSM).

1.1.1 Neutrino Production in the Sun

The Sun produces most of its energy through a fusion process known as the proton-proton (pp) chain. Some of the reactions in the fusion process produce neutrinos. This section will outline the pp chain and highlight the reactions that produce neutrinos. Figure 1.2 shows a schematic view of the pp chain, including the branching ratios of each interaction.

The pp chain is divided into four branches, where the start of the chain is shown in Equation (1.1), in which two protons combine to form ^{2}H, which subsequently combines another proton to form ^{3}He, using

$$\begin{aligned} {}^1_1\mathrm{H} + {}^1_1\mathrm{H} &\rightarrow {}^2_1\mathrm{H} + e^+ + \nu \\ {}^2_1\mathrm{H} + {}^1_1\mathrm{H} &\rightarrow {}^3_2\mathrm{He} + \gamma. \end{aligned} \tag{1.1}$$

The first branch (pp I) has a termination of 85%, shown in Equation (1.2),

$${}^3_2\mathrm{He} + {}^3_2\mathrm{He} \rightarrow {}^4_2\mathrm{He} + 2\ {}^1_1\mathrm{H} \quad \text{(pp I)}. \tag{1.2}$$

The termination percentage is the branching fraction of the pp chain, originating from Equation (1.1), in which each reaction occurs. The termination percentage is based on the BP2000 solar model [1].

The pp II chain is shown in Equation (1.3),

$$\begin{aligned} {}^3_2\mathrm{He} + {}^4_2\mathrm{He} &\rightarrow {}^7_4\mathrm{Be} + \gamma \\ {}^7_4\mathrm{Be} + e^- &\rightarrow {}^7_3\mathrm{Li} + \nu_e \quad \text{(pp II)} \\ {}^7_3\mathrm{Li} + {}^1_1\mathrm{H} &\rightarrow 2\ {}^4_2\mathrm{He}, \end{aligned} \tag{1.3}$$

and the resulting neutrino spectrum is labelled as ^{7}Be in Figure 1.1. This reaction has a termination of 15%.

After the pp II reactions, a proton can be captured on the ${}^{7}_{4}\mathrm{Be}$ and the pp III chain takes place as shown in Equation (1.4),

$$\begin{aligned} {}^{7}_{4}\mathrm{Be} + {}^{1}_{1}\mathrm{H} &\rightarrow {}^{8}_{5}\mathrm{B} + \gamma \\ {}^{8}_{5}\mathrm{B} &\rightarrow {}^{8}_{4}\mathrm{Be} + e^{+} + \nu_e \quad \text{(pp III)} \\ {}^{8}_{4}\mathrm{Be} &\rightarrow 2\ {}^{4}_{2}\mathrm{He}. \end{aligned} \tag{1.4}$$

The pp III reaction has a termination of 0.02%. The resulting neutrino spectrum is labelled as ${}^{8}\mathrm{B}$ in Figure 1.1.

An even more rare reaction is the pp IV chain, also called the *hep* reaction, shown in Equation (1.5). The *hep* reaction has a termination of 0.00003%. The resulting neutrino spectrum is labelled as *hep* on Figure 1.1. This chain also branches off Equation (1.1).

$${}^{3}\mathrm{He} + p \rightarrow {}^{4}\mathrm{He} + e^{+} + \nu_e \quad \text{(pp IV or } \textit{hep}\ \text{)} \tag{1.5}$$

There is a second independent source of energy from the Sun, called the CNO cycle. Although it dominates in some stars, the cycle is highly temperature dependent and, at the Sun's temperature, only accounts for about 1% of the solar energy production. The neutrino spectra from this cycle are shown as (blue) dashed lines on Figure 1.1.

Another source of neutrinos is from the *pep* reaction, which is the forming of deuterium by electron capture on two protons. This is rare because of the need for three objects to combine, instead of two as in the pp chain. The pep reaction is shown in Equation (1.6) and the resulting neutrino spectrum is labelled *pep* on Figure 1.1.

$${}^{1}_{1}\mathrm{H} + e^{-} + {}^{1}_{1}\mathrm{H} \rightarrow {}^{2}_{1}\mathrm{D} + \nu_e \tag{1.6}$$

The pp III chain and the *hep* reaction create neutrinos that can be measured by terrestrial Čerenkov detectors. The low energy radioactivity in SNO dominates over the low energy neutrinos, and we are only able to see neutrinos with energies greater that 3.5 MeV, which means we are not sensitive to any solar neutrino sources except for those produced by the ^{8}B and *hep* reactions.

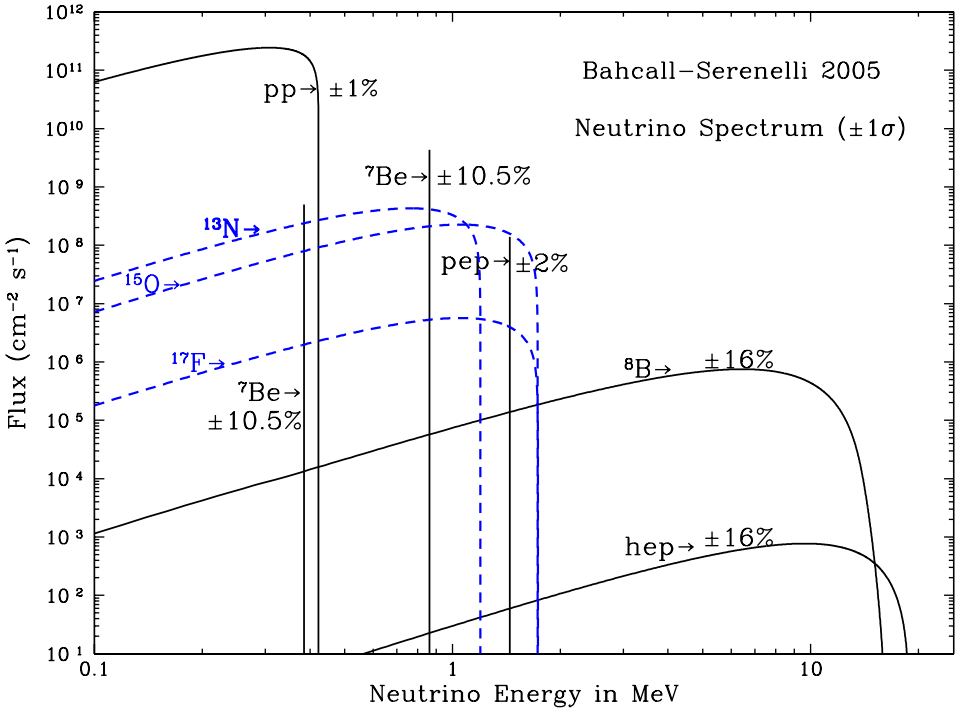

Figure 1.1: Solar neutrino energy spectra. The blue dashed lines are those lines produced in the CNO cycle. The uncertainties shown are the theoretical uncertainties [2].

Measurement of ^{8}B neutrinos by water Čerenkov detectors has already been done by the SNO and Super-Kamiokande experiments, and is described in References [4] and [5]. These experiments have updated their results as they have collected more data, and the most recent results can be found in References [6], [7], and [8].

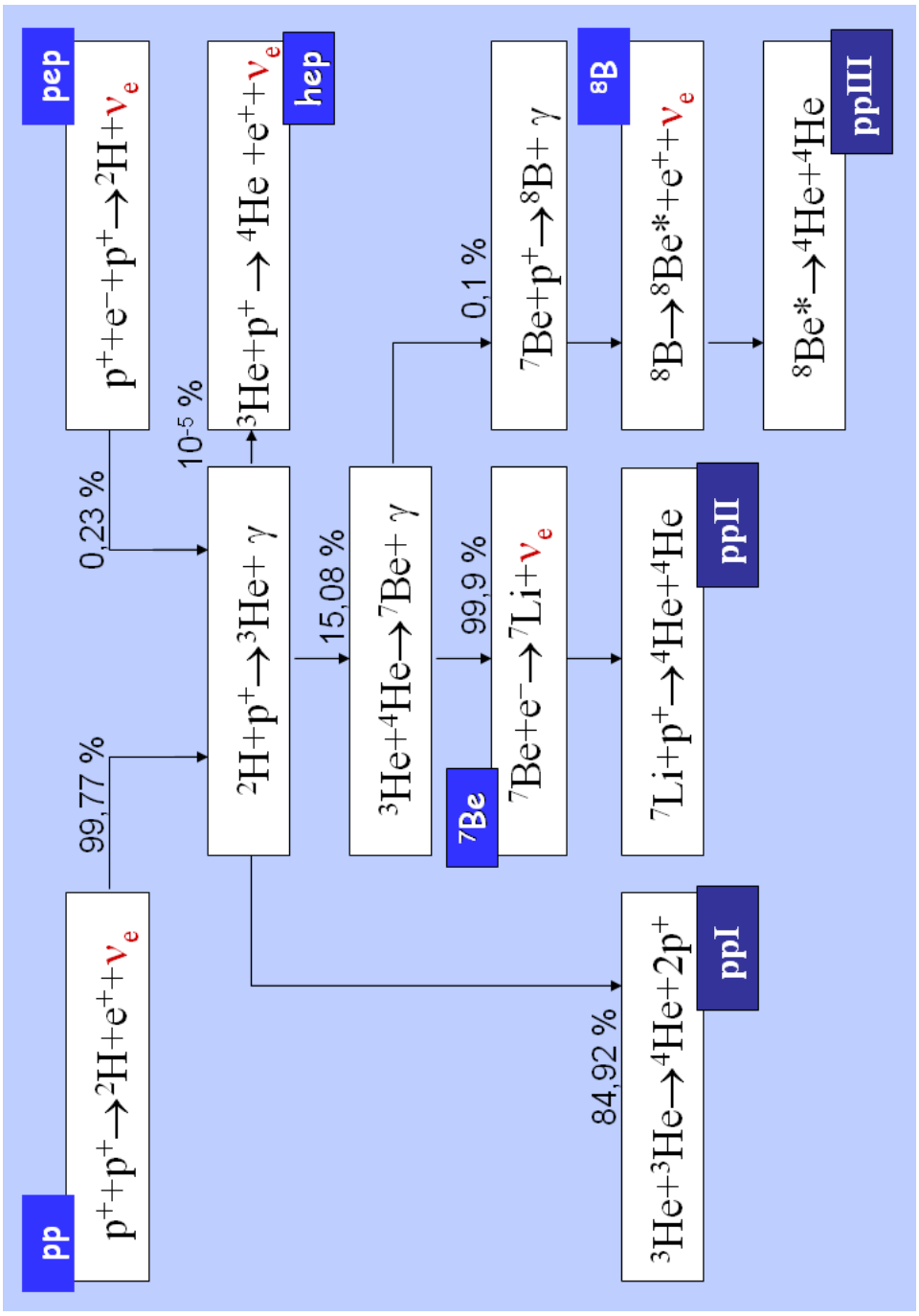

Figure 1.2: A schematic view of the proton-proton chain [3]. The neutrinos have been highlighted in red.

1.1.2 Neutrino Oscillations

Neutrino experiments have been observing solar neutrinos since 1964 when Ray Davis first published his results from a chlorine detector in the Homestake mine [9]. This was the first evidence of non-standard model physics with respect to neutrinos. Davis observed a discrepancy of the flux of neutrinos produced in the Sun when compared to the SSM. This was called the *solar neutrino problem*.

A theoretical framework for the solution to the solar neutrino problem had been in place since 1969 when Pontecorvo first postulated neutrino oscillations [10]. The Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix is a mechanism that allows for massive neutrinos to oscillate. The PMNS matrix is developed in Refernces [11] and [12] and summarized as

$$|\nu_\alpha\rangle = \sum_i U_{\alpha i}|\nu_i\rangle, \tag{1.7}$$

where $|\nu_\alpha >$ is the neutrino flavour eigenstate, $|\nu_i\rangle$ is the neutrino mass eigenstate, and $U_{\alpha i}$ is the matrix that governs the probability of oscillation between states. In the case of two neutrino flavours, which is a very good approximation for the case of solar neutrinos [13], we have:

$$U = \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} \\ -\sin\theta_{12} & \cos\theta_{12} \end{pmatrix}, \tag{1.8}$$

where θ_{12} is the *mixing angle* between neutrino mass eigenstates 1 and 2. If we now calculate the probability of a neutrino oscillating from an electron flavour state, $\alpha = e$, to a muon flavour state, $\alpha = \mu$ we get, from Reference [13],

$$P_{e\to\mu} = \sin^2 2\theta_{12} \sin^2\left(\frac{\Delta m^2 L}{4E}\right), \tag{1.9}$$

where L is the distance travelled, E is the energy of the neutrino, and Δm^2 is the mass difference of the neutrino mass eigenstates squared. In the case of solar neutrinos, the L parameter is fixed, and the θ_{12} and Δm^2 parameters are referred to as the neutrino *oscillation parameters*.

1.1.3 Neutrino Oscillations in Matter

Section 1.1.2 discussed neutrino oscillations in a vacuum, but it is possible for neutrinos to oscillate at a higher frequency when moving through matter. Electron neutrinos produced near the core of the Sun are influenced by the large electron density within the Sun and are much more likely to change flavour because of this.

Mikheyev-Smirnov-Wolfenstein (MSW) postulated an effect which explains how matter acts on neutrinos causing oscillations [14, 15]. The MSW effect can be seen if we consider the propagation of a neutrino (a lepton) to follow the Schrödinger equation. The potential energy of the electron neutrino, V_{ν_e}, is given by,

$$V_{\nu_e} = \sqrt{2} G_F N_e, \tag{1.10}$$

where G_F is the Fermi constant and N_e is the electron density.

The equation of neutrino propagation is given by,

$$i\frac{d}{dx}\begin{pmatrix} \nu_e \\ \nu_\alpha \end{pmatrix} = \frac{1}{2E} M^2 \begin{pmatrix} \nu_e \\ \nu_\alpha \end{pmatrix}, \tag{1.11}$$

where α is a muon or tau neutrino, and the matrix M^2 is,

$$M^2 = \frac{1}{2}\left[U \begin{pmatrix} -\Delta m^2 & 0 \\ 0 & \Delta m^2 \end{pmatrix} U^T + 2E \begin{pmatrix} \sqrt{2} G_F N_e & 0 \\ 0 & -\sqrt{2} G_F N_e \end{pmatrix} \right], \tag{1.12}$$

where U is the same as in Equation (1.8) and Δm^2 is the same as in Equation 1.9.

To summarize Equation (1.11), the neutrino oscillation length decreases for a higher electron density medium. The Sun has a very high electron density at the core so the distortion of the electron neutrinos is significant. As will be mentioned in Section 1.2 we know that *hep* neutrinos are produced with a different radial distribution than ^{8}B neutrinos. Figures 1.3 and 1.4 show the probability that an electron neutrino created in the Sun remains an electron

neutrino at the Earth (P_{ee}) for various values of the neutrino oscillation parameters for the *hep* and ^{8}B neutrinos respectively. The oscillation parameters are represented as $\tan^2\theta_{12}$ and Δm^2_{12}.

The current best fit of the neutrino oscillation parameters is in the large mixing angle (LMA) region. In the LMA region, the survival probability is nearly constant in the high energy range ($13.5 < E < 20$ MeV). It has a mean survival probability of around 0.36.

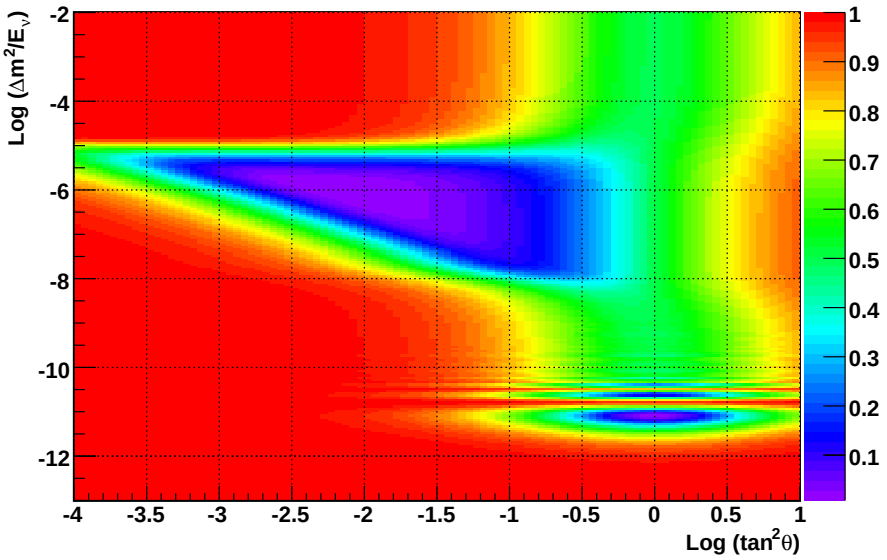

Figure 1.3: Probability of an electron neutrino created by the *hep* reaction remaining an electron neutrino at the SNO detector as a function of the neutrino oscillation parameters.

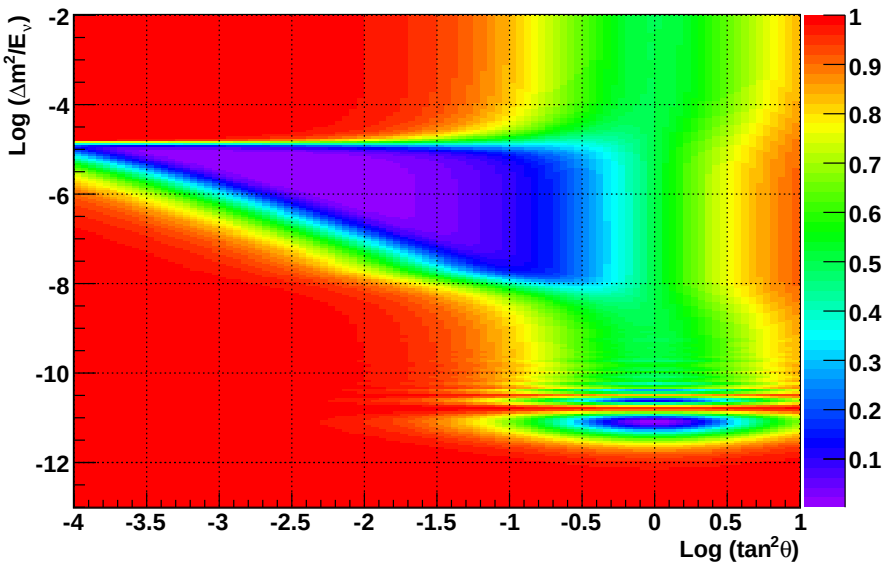

Figure 1.4: Probability of an electron neutrino created by the ^{8}B reaction remaining an electron neutrino at the SNO detector as a function of the neutrino oscillation parameters.

1.2 *hep* Neutrinos

The *hep* reaction occurs according to all standard solar models (SSMs) and therefore *hep* neutrinos (i.e., neutrinos with a *hep* energy spectrum) should be found. The proton-proton solar fusion chain (Section 1.1.1) necessarily produces *hep* neutrinos. As shown in Figure 1.1, the ^{8}B neutrino spectrum is similar (and overwhelms) that of the *hep* up to 15 MeV; the ^{8}B spectrum then ends at 16 MeV, where the *hep* neutrino extends to 18.8 MeV.

The *hep* neutrino is also interesting because of its origin in the Sun. The *hep* neutrino production takes place out to a much larger solar radius than the ^{8}B neutrino. Figure 1.5 shows that the *hep* neutrino production extends out to 0.3 of the solar radius where ^{8}B neutrino production only reaches 0.1 solar radii. Since neutrinos react with matter through matter-induced flavour oscillations (discussed in Section 1.1.3) that depend on the density of electrons, such reactions are quite different for *hep* and ^{8}B neutrinos.

The *hep* neutrino flux is a parameter in the SSM, so the precision in measuring this flux can be important for the accuracy of the model. John Bahcall calculated the value of 21 input parameters to his solar model by running $10,000$ simulations varying each input for each simulation [16]. He then used the models that resulted in a helium abundance within the accepted range. One of the input parameters was the *hep* flux, the distribution of good *hep* fluxes is shown in Figure 1.6. The calculated *hep* flux from this model is $(7.970 \pm 1.236) \times 10^3$ cm^{-1}s^{-1}. This is the best calculation of the *hep* flux. This model states that the theoretical uncertainty of the *hep* flux is as high as 15.1%.

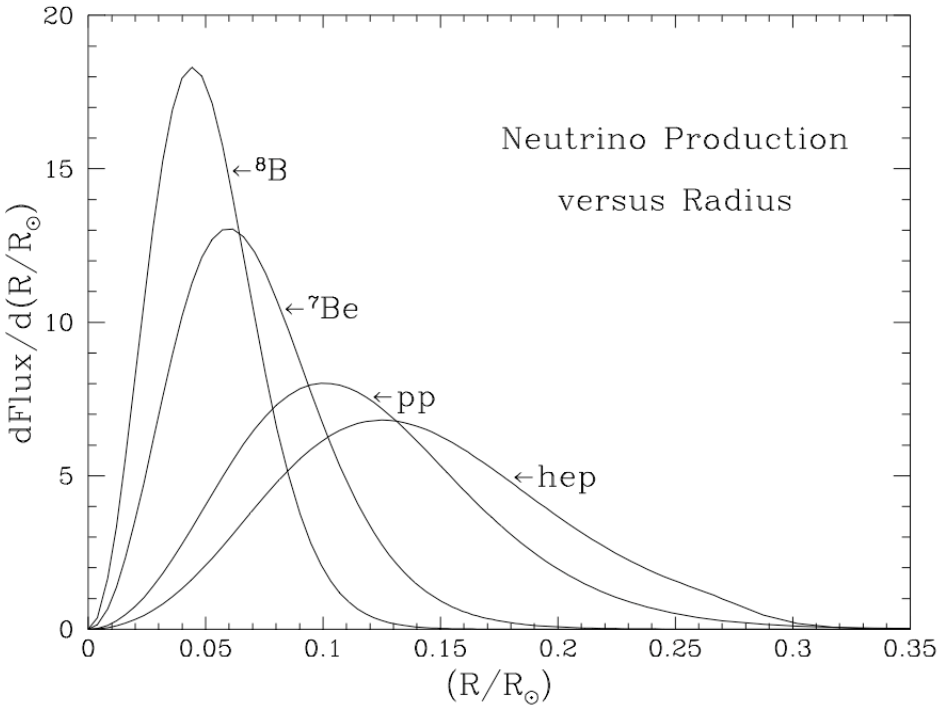

Figure 1.5: Radial production of neutrinos in the Sun [13]. In units of $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

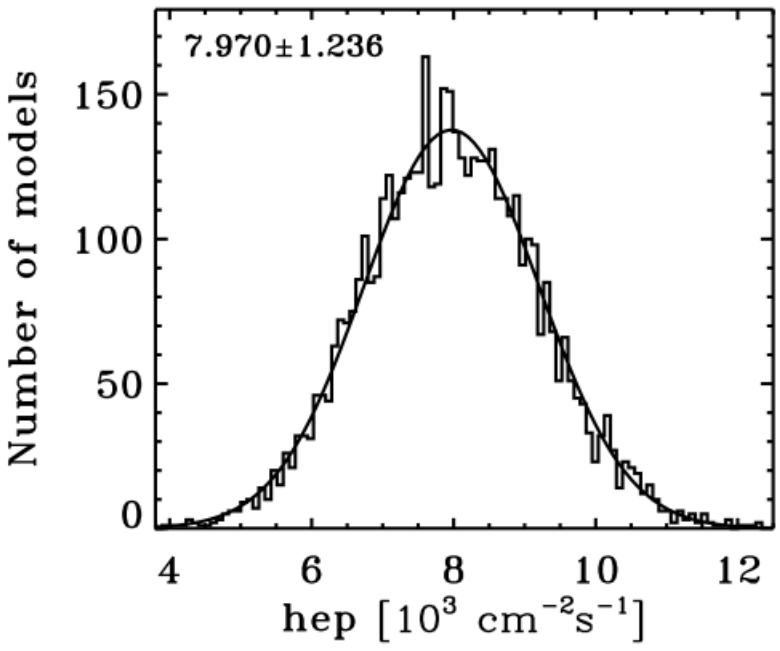

Figure 1.6: Bahcall's calculation of the *hep* flux using 10,000 simulations [16].

1.2.1 The History of *hep* Reaction

From 1999–2001, the Super-Kamiokande collaboration reported that the *hep* flux was 16.7 times larger than the theoretical prediction with a 24% confidence level on their χ^2 fit [17, 18, 19, 20] . This fit is shown in Figure 1.7. The fit was done allowing the *hep* flux to be a free parameter. The *hep* flux was then experimentally determined by looking at the flux above the ^{8}B neutrino endpoint (18 MeV). Super-Kamiokande used a low energy threshold placed at 6.5 MeV, then after 708 days, they moved the threshold to a *super low energy* (SLE) threshold of 5 MeV, corresponding to an additional 524 days. "SLE" in Figure 1.7 refers to this lower energy range. If we look at Figure 1.8 we can see the high energy neutrino spectrum based on events beyond the end point of the ^{8}B neutrino spectrum.

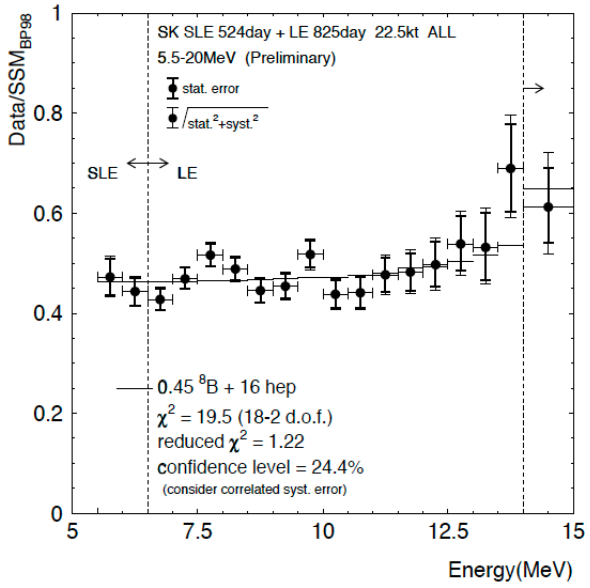

Figure 1.7: Super-Kamiokande recoil energy spectrum. Plotted as a fraction of the Standard Solar Model (BP98) [18].

The Super-Kamiokande result spurred much excitement in the nuclear astrophysics community. This result caused a lot work to be done in the calculation of the *hep* reaction flux. Physicists went to work on their calculations

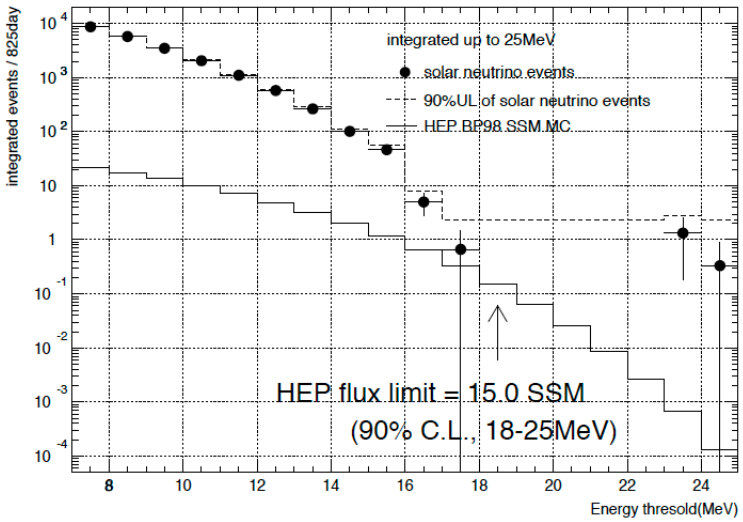

Figure 1.8: Super-Kamiokande high energy spectrum. The *hep* flux is obtained from the data above 18 MeV [20].

of the astrophysical S-factor:

$$S(E) = E\sigma(E)e^{\frac{4\pi\alpha}{v_{rel}}}$$

where $\sigma(E)$ is the cross section at centre of mass energy E, v_{rel} is the relative velocity between the ^{3}He and the proton, and α is the fine structure constant. This S-factor is the main input into the SSM for the *hep* flux, it is basically the ^{3}He-proton cross section, which will almost totally dominate the *hep* rate.

The S-factor has been calculated in References [21], [22], [23], and [24], which found an S-factor 4.5 times larger than the one used in the SSM. It is noted in Reference [22] that this is still not large enough to be consistent with the Super-Kamiokande measurement. In Reference [25], the reaction:

$$n + {}^3\text{He} \rightarrow {}^4\text{He} + e^- + \bar{\nu}_e$$

is proposed to estimate the S-factor for the similar *hep* reaction. This reaction is much easier to work with because there are experiments with high intensity beams of thermal neutrons that can quite accurately measure the cross section. The hope would then be to extrapolate to the *hep* cross section. The goal was

to find an S-factor that would influence the SSM prediction of the *hep* flux to agree with the Super-Kamiokande result.

Other work has also been motivated by the Super-Kamiokande result, such as trying to use the solar luminosity to constrain the *hep* flux [26] and more complex associations of the *hep* S-factor to the s- and p-wave calculations [27]. In Reference [28] the *hep* flux disagreement was tested in models that added variations in the Maxwell-Boltzmann distribution expected in the solar core plasma.

A review of the solar *hep* process was published in 2004 [29] which gives an overview of many of the calculations preformed to predict the *hep* flux, as well as the comparison to the experimental results. The review also states the importance of a *hep* measurement.

The measurement of the *hep* flux can affect the SSM. A *hep* flux that is higher than predicted, as measured by Super-Kamiokande, means that the calculation of the ^{3}He-proton cross section is not accurate. This can also be true if the *hep* flux were lower than predicted, as long as it can be measured. A more interesting result could be the absence of the *hep* flux, as this could imply some sort of new physics.

The Super-Kamiokande result did not stand the test of time: as statistics increased, the limit that Super-Kamiokande was able to put on the *hep* reaction converged toward the SSM prediction. In 2001, Super-Kamiokande was able to bring the *hep* limit down to 4.0×10^4 cm^{-2} s^{-1}, which was only 4.3 times the SSM [30]. In 2006 Super-Kamiokande's *hep* limit was 7.3×10^4 cm^{-2} s^{-1} [31]. Both limits are stated at the 90% confidence limit.

The *hep* flux was placed by SNO in 2006 [32]. This method was done by counting all events with energy above 14.3 MeV and below 20 MeV. This counting experiment is detailed in Section 5.1 and resulted in an upper limit of 2.3×10^4 cm^2 s^{-1} with a 90% confidence level.

1.3 The Sudbury Neutrino Observatory

The Sudbury Neutrino Observatory (SNO) is located near Sudbury, Ontario, Canada, two kilometres underground in Vale INCO's Creighton mine. This location uses the 6,000 metre water equivalent (two km of norite rock) over burden as a shield against cosmic rays. The detector itself is a one-kilotonne ultra-pure heavy water Čerenkov light detector. The heavy water (D_2O) is contained in a 12 metre diameter spherical acrylic vessel (AV), surrounded by a 17.8 metre geodesic photomultiplier support structure (PSUP). This frame houses nearly 10,000 photomultiplier tubes (PMTs or PMT array). The PMT array is how SNO observes the Čerenkov light that is produced by the neutrinos. To shield the heavy water from external backgrounds and help support the AV there is ultra-pure light water (H_2O) between the PSUP and the AV. On the outside of the PSUP the remainder of the cavity is filled with light water. This light water outside the PSUP is also acts as neutron shielding and as support to the PSUP. This structure is shown in Figure 1.9. Full details of the SNO detector can be found in Reference [33].

SNO has the unique ability to detect all three neutrino flavours via the reactions:

$$\nu_e + d \rightarrow p + p + e^- \tag{1.13}$$

$$\nu_x + d \rightarrow p + n + \nu_x \tag{1.14}$$

$$\nu_x + e^- \rightarrow \nu_x + e^-. \tag{1.15}$$

These equations describe the charged current (CC), neutral current (NC) and elastic scattering (ES) reactions, respectively. The charged current reaction is only sensitive to the ν_e flavour. The elastic scattering reaction is sensitive mostly to the ν_e flavour, but has some sensitivity to the ν_μ and ν_τ

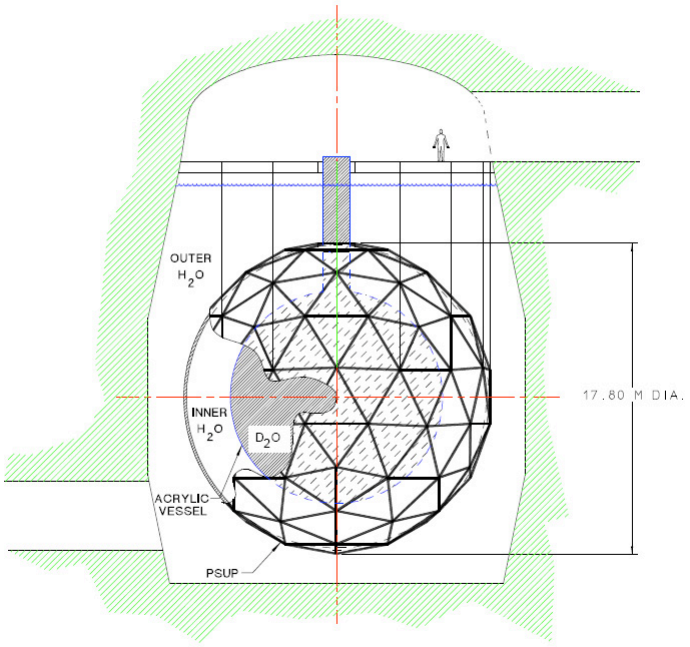

Figure 1.9: An artist's representation of the SNO cavity containing the support structure and AV.

flavours. The neutral current reaction is equally sensitive to any of the three flavours, ν_x, where x can be e, μ, or τ.

All prior solar neutrino experiments showed a deficit of ν_e neutrinos to the rate predicted by the SSM [9, 34, 35, 5], the *solar neutrino problem*. SNO's sensitivity to both NC and CC allowed SNO to distinguish an oscillation scenario $\left(\frac{\nu_e}{\nu_x} = \frac{1}{3}\right)$ from a solar model scenario $\left(\frac{\nu_e}{\nu_x} = 1\right)$. SNO solved the solar neutrino problem by showing that the rate of NC reaction events agreed with the number predicted by the standard solar model [4], while the rates of CC reaction events and ES events were consistent with oscillations in which two-thirds of the neutrinos at the Earth are ν_μ or ν_τ neutrinos.

For SNO to detect a neutrino, the neutrino must react in the SNO detector, and the reaction must produce Čerenkov light. Čerenkov light (or Čerenkov radiation) is produced by a charged particle moving faster than the speed of light in the medium it is moving in. The CC and ES reactions produce elec-

trons, which directly produce Čerenkov light. The NC reaction does not. The NC reaction produces a neutron that will thermalize and may be captured, which produces a gamma particle which can then Compton scatter an electron to produce Čerenkov light. This means that events that are produced by neutrons, including those from the NC reaction have little information about the source of the neutron that caused it. The direction that the gamma particle is emitted after neutron caption is random, and the energy of the gamma particle produced by thermal neutron capture is also independent of the neutron source.

The SNO experiment was done in three phases. Each phase employed a different technique to detect neutrons. This three phase approach gave SNO the ability to self-verify its results. Each phase has different systematic uncertainties allowing the results from each phase to be compared to each other. This also allows the phases to be combined into one global analysis where the different systematic uncertainties are used to break correlations that limit one phase analyses.

1.3.1 Phase I (D_2O)

The first phase of the SNO experiment ran from November 2, 1999 to May 31, 2001 and contained 277.4 days of neutrino live time; that is the amount of time that SNO was able to detect a neutrino. The active volume was made up of ultra-pure heavy water (D_2O). This phase detected the neutrons from the NC reaction by neutron capture on the deuterium, which has a cross-section of 0.5 mb. The capture leads to the production of a single gamma particle, which then Compton scatters off an electron to produce Čerenkov radiation.

1.3.2 Phase II (NaCl)

The second phase of SNO took place from July 26, 2001 to August 28, 2003 resulting in 391.4 live days of neutrino data. In this phase, pure sodium chloride ($^{23}Na^{35}Cl$) was dissolved into the ultra-pure heavy water. We refer to this phase as the *salt* phase. The salt was added to the active volume because the chlorine has a much larger thermal neutron capture cross-section than deuterium. Chlorine has a cross-section of 44 b to neutron capture.

$$n + {}^{35}\mathrm{Cl} \rightarrow {}^{36}\mathrm{Cl} + \gamma \tag{1.16}$$

This large cross-section allows this phase to capture a much higher rate of neutrons produced by the NC reaction. This led to a more precise measurement of the NC flux. Equation (1.16) shows the neutron capture on ^{35}Cl. When the neutron is captured multiple gamma particles are produced. This multiple gamma particle production leads to easier NC identification compared to a single gamma particle produced from neutron capture on deuterium as in phase I.

1.3.3 Phase III (Neutral Current Detector Array)

The third and final phase of SNO collected data from November 27, 2004 to November 28, 2006, accumulating 385.2 live days of neutrino data. In this phase, ^3He proportional neutron counters were installed into the active volume. 36 of these detectors were used as well as four ^4He counters. The purpose of the 36 ^3He detectors is to detect neutrons from the NC reaction. The advantages are to increase the efficiency for the detection of the NC reaction, as compared to pure D_2O, and to allow event-by-event particle identification.

The counters are referred to as neutral current detectors (NCDs). The NCD array is made up of "strings," each of which is divided into three or four

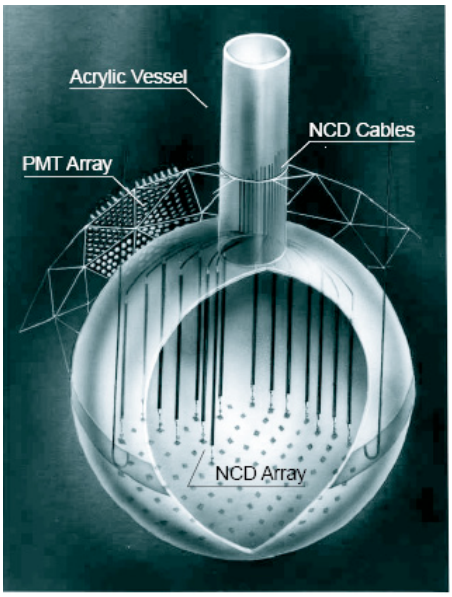

Figure 1.10: An artist's drawing of SNO and the NCD array.

individual counters. Each counter is an ultra-pure nickel tube with a 5.08 cm inner diameter. The tubes are strung with a 50 μm copper anode wire down the centre, and are filled with a ^{3}He-CF_4 gas mix. The NCDs are 9 to 11 metres in length. They are arranged in a one metre rectangular grid inside the AV, as shown in Figure 1.10.

The ^{3}He nucleus has a large cross-section to thermal neutrons (5330 b), which allows the NCD to capture virtually all the neutrons which enter the tube. The detection reaction is

$$n + {}^3\mathrm{He} \rightarrow p + t, \tag{1.17}$$

where p is a proton and t is a triton. The proton and triton are emitted anti-parallel with kinetic energies of 573 and 191 keV, respectively. If both the full energy of the proton and triton are deposited into the NCD a total kinetic energy of 764 keV can be detected. If either or both hit the wall before depositing all their energy, less energy is seen.

1.3.4 Combined Three Phase Analysis

Over the three phases of running, the SNO experiment has accumulated data corresponding to a neutrino live time of 1,054.0 days. Combining the three phases of data allows us to measure a neutrino flux, where the systematic uncertainties are different, and in some cases completely independent, from phase to phase. This allows us to break some of the correlations that may exist in a single phase type analysis. The most important gain for the *hep* analysis is the increased number of events. The number of events in the energy range of interest is less than 10 for any single phase. With so few events, a flux upper limit is all that can be determined. Using all three phases, we expect to observe 15–30 events, meaning that we may be able to, for the first time in any experiment, measure the *hep* flux. At the time for writing the SNO internal review committee has not approve the release of the full data set. The results presented in this thesis are based on a 1/3 of the full data set.

Chapter 2

Characterizing Events in the Sudbury Neutrino Observatory for the Measurement of *hep* Neutrinos

2.1 Calibrations

The SNO detector is a Čerenkov light detector, which measures information about the light that the PMTs observe. The distribution of the light in the detector and timing of the light arriving at the PMTs can be used to identify the light as Čerenkov radiation. The distribution of PMT hits due to Čerenkov radiation (a Čerenkov cone) is shown in Figure 2.1. Events that are more energetic produce more light than those of lower energy.

The relationship between the amount of light detected, or the number of PMTs that are triggered, is complicated by optical effects in the water such as attenuation (see Reference [36] for details on the optical characterization of SNO), PMT efficiencies, and asymmetries in the PMT arrangement.

For the *hep* analysis the most important information about an event is the energy of the event. The calculation of the energy of an event is done in the energy response processor (RSP), which is discussed in Section 2.2.1, and explained in detail in Reference [37].

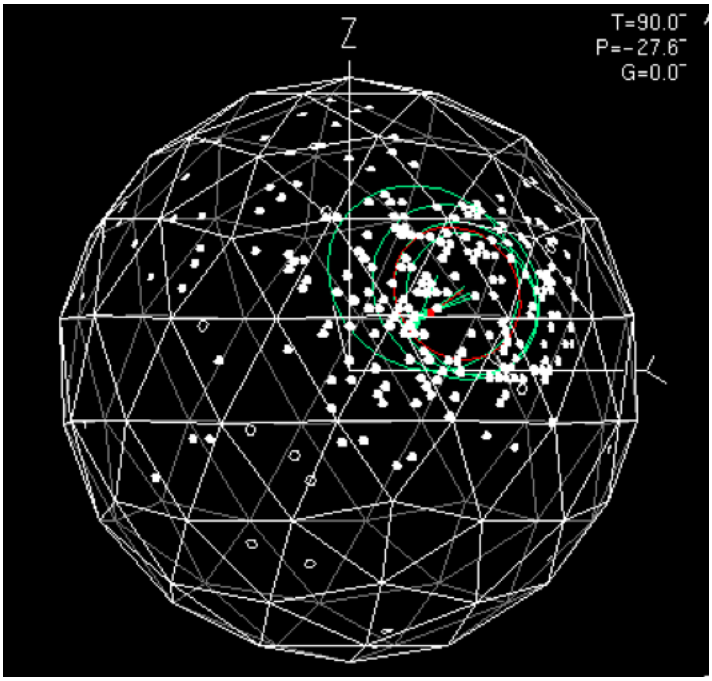

Figure 2.1: A Monte Carlo event (from a Michel Electron) in the Sudbury Neutrino Observatory. The dots represent PMTs triggered by the event.

To be able to estimate the energy of an event we must first see what events of known energies look like in SNO. The *hep* analysis relies on four calibration sources at various energies: a 6.13 MeV ^{16}N gamma particle source (Section 2.1.1), a 19.8 MeV gamma particle from a positron-triton (pT) source (Section 2.1.2), a ^{8}Li beta particle source with an end-point energy of 12.96 MeV (Section 2.1.3), and a Michel electron source which can have energies up to 60 MeV (Section 2.1.4). These calibration sources were used to characterize the RSP algorithm, which allowed us to get an accurate determination of the energy of events in the energy range we chose to look for *hep* neturinos.

2.1.1 ^{16}N Source

The ^{16}N source is the main energy calibration device used in the SNO detector. It produces a 6.13 MeV gamma particle that Compton scatters off electrons producing Čerenkov light. To create the ^{16}N, high energy neutrons

are produced outside of the SNO volume by an external deuteron-triton (DT) generator that uses the fusion reaction:

$$d + t \rightarrow n + {}^4\mathrm{He}, \tag{2.1}$$

where d is a deuteron (^{2}H), t is a triton (^{3}H), and n is a neutron. These neutrons are used to produce the ^{16}N through the (n, p) reaction on ^{16}O where the ^{16}O is in the form of CO_2 gas.

The ^{16}N gas is then brought into the chamber shown in Figure 2.2 through a long tube called the *umbilical*. The umbilical enters through the top O-ring plate and the gas flows into the decay chamber which is enclosed by the plastic scintillator shell.

The ^{16}N decays in the chamber through the beta decay scheme shown in Figure 2.3. Only the decay to the 6.1 MeV branch produces a gamma particle and a beta particle. The beta particle reacts with the scintillator producing light that is observed by the PMT above the decay chamber. The gamma particle escapes into SNO to be used in the energy calibration. The ^{16}N source is encased in stainless steel so that no light from the scintillation is allowed into SNO.

The beta particles produced are mainly stopped in the steel housing; however, there is a chance they may cause bremsstrahlung radiation, or additional gamma particles in the steel. Nevertheless, this effect has been studied using a Monte Carlo simulation, and was determined to be small. Reference [38] is a published Nuclear Instruments and Methods article which contains in-depth details about the source.

Figure 2.4 shows the result from an ^{16}N calibration run. The source was placed in in the centre of the volume; this calibration run was taken in the first phase of SNO. The number of events in this figure has been normalized. The blue histogram represents the data, and the red histogram represents the

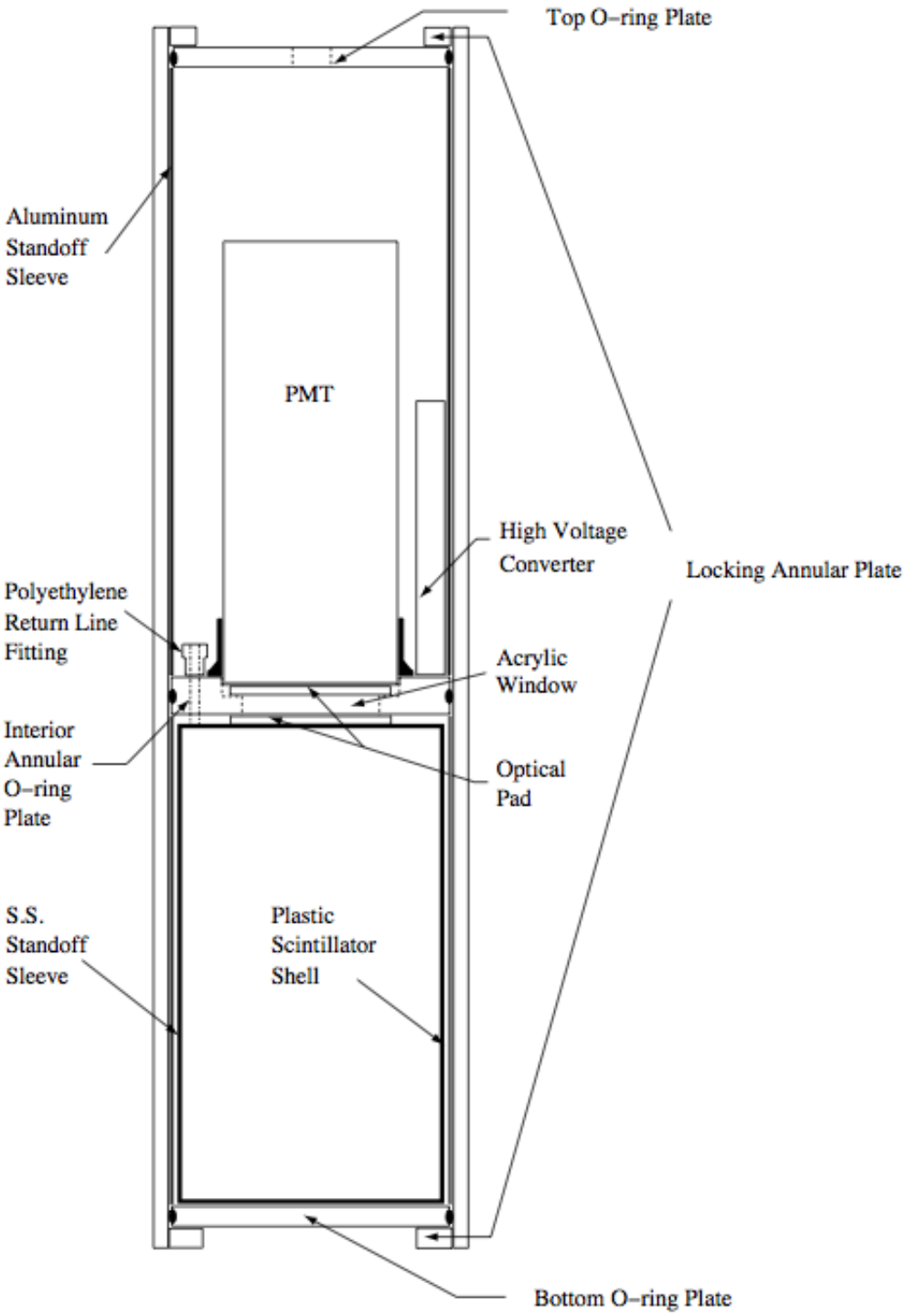

Figure 2.2: Decay chamber of the ^{16}N source as shown in Reference [38].

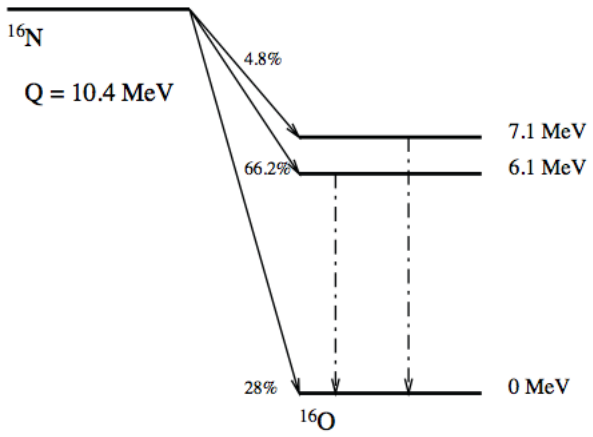

Figure 2.3: The decay scheme of ^{16}N as shown in Reference [38].

MC. Each histogram has a Gaussian fit to determine the resolution (sigma).

It is from a set of plots similar to Figure 2.4, taken from many positions throughout the three phases that we can determine the overall measurement of the level of agreement between the data and Monte Carlo. This characterization, in part, leads to the energy systematic uncertainties that are discussed in Section 6.1.1.

2.1.2 pT Source

The ^{16}N source probes energies near 6 MeV, which is in the heart of the ^{8}B neutrino energy spectrum. The energy response can be assumed to be constant up to higher energies. However, it is alway a good idea to test this sort of assumption. The pT source was designed to do just that. It was designed to test and set limits on the energy non-linear response of the SNO detector. The pT source was made up of three parts, the gas discharge line, the ion acceleration line, and the target chamber. Protons were accelerated through the source and they collided with a high-purity scandium tritide target.

The pT reaction is given by:

$$p + t \rightarrow {}^{4}\mathrm{He} + \gamma \ (19.8 \text{ MeV}). \tag{2.2}$$

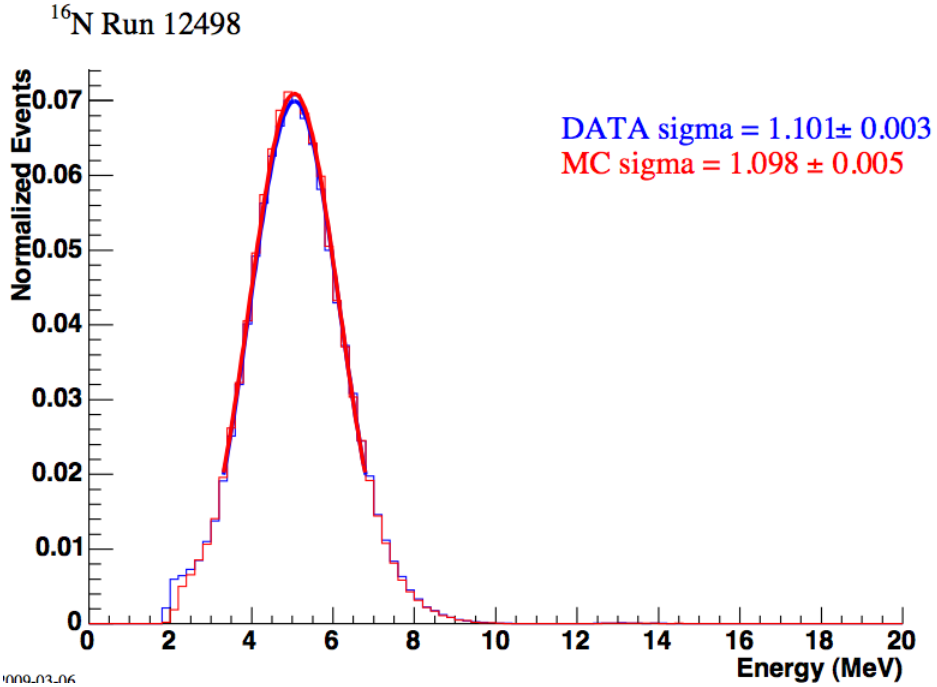

Figure 2.4: The ^{16}N data compared to the ^{16}N MC for a central calibration run in phase I [39]. This figure is used in part to evaluate the energy systematic uncertainties.

The pT source (Figure 2.5) produces a gamma particle with an energy of 19.8 MeV. This is much higher than the 6.1 MeV produced by the ^{16}N source. The pT source allowed us to calibrate the detector near the ^{8}B and *hep* endpoints.

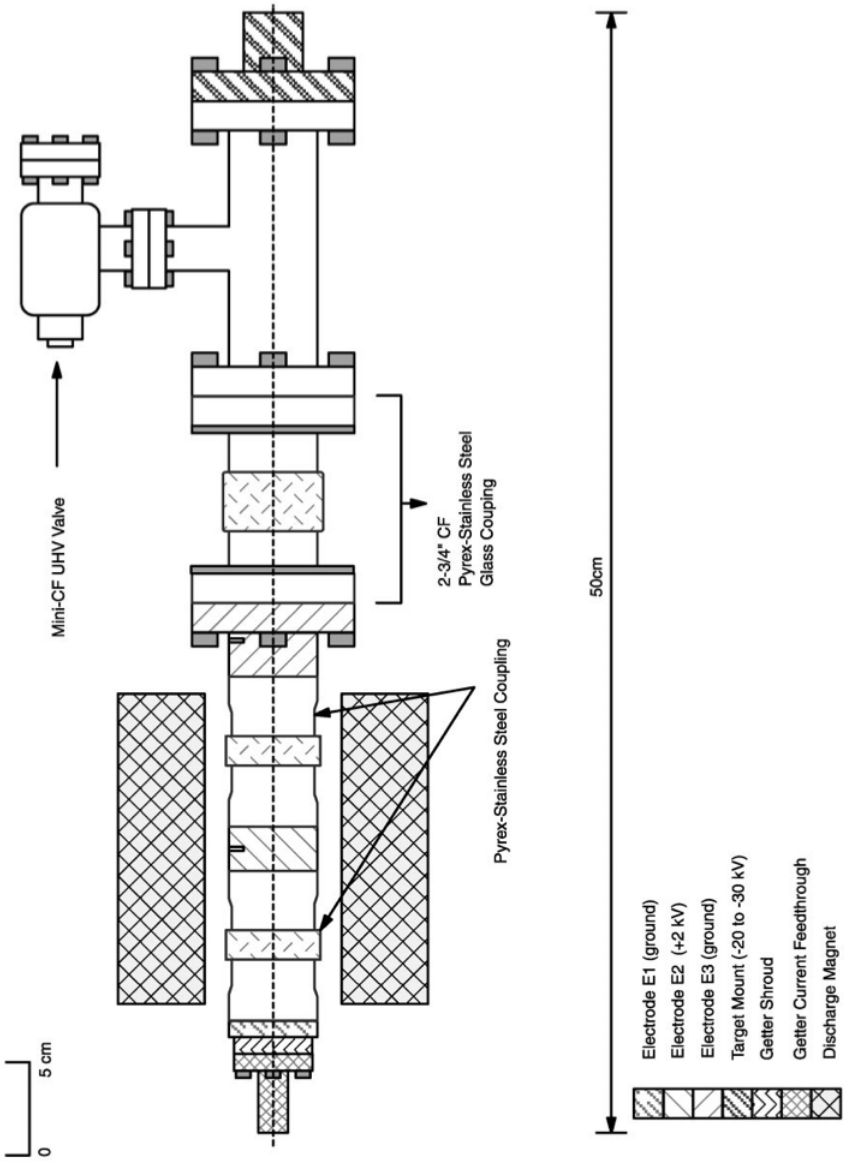

Figure 2.5: Diagram of the pT source [40].

A secondary reaction that takes place is the ^{3}H + ^{3}H reaction, which

produces large numbers of neutrons. This secondary reaction in the pT source produces 1000 times as many neutrons as gamma particles, and if this were to be run in the salt phases, they would all capture and produce gamma particles. It was only then feasible to deploy the source in the pure D_2O phase. Reference [40] is a published Nuclear Instruments and Methods article with contains in-depth details about the source.

Figure 2.6 shows the energy spectrum produced by the pT source. We can see in the top panel the large low-energy peak. This is the peak from the neutron producing $^3H + ^3H$ reaction. A set of data cleaning cuts are applied to this spectrum, resulting in the lower panel of Figure 2.6. It is this lower panel spectrum that we use to verify the energy characterization determined by the ^{16}N source is valid at higher energies.

2.1.3 ^{8}Li Source

The ^{8}Li source was designed to produce a similar energy spectrum to the ^{8}B neutrino energy spectrum. The ^{8}Li source produces beta particles with an end-point energy of 12.96 MeV. The ^{8}Li is produced by the $^{11}B(n,\alpha)$ reaction, where the neutrons are produced by the same DT generator as mentioned in Section 2.1.1. The ^{8}Li is carried from the DT generator to the ^{8}Li source, shown in Figure 2.7, by helium gas with a salt aerosol.

The decay volume is made of spun stainless steel, in the decay volume, the ^{8}Li decays via the scheme in Figure 2.8. The ^{8}Li that decays in the decay volume also produces two alphas, and these alphas cause scintillation in the helium transport gas. There is not much light produced by the scintillation, and most of the light produced is in the UV wavelengths, which is outside of the detectable wavelength range of the PMT.

The inside of the decay volume is coated with a reflective white titanium oxide paint to reflect the light, and then a thin layer of tetraphenyl oxide is

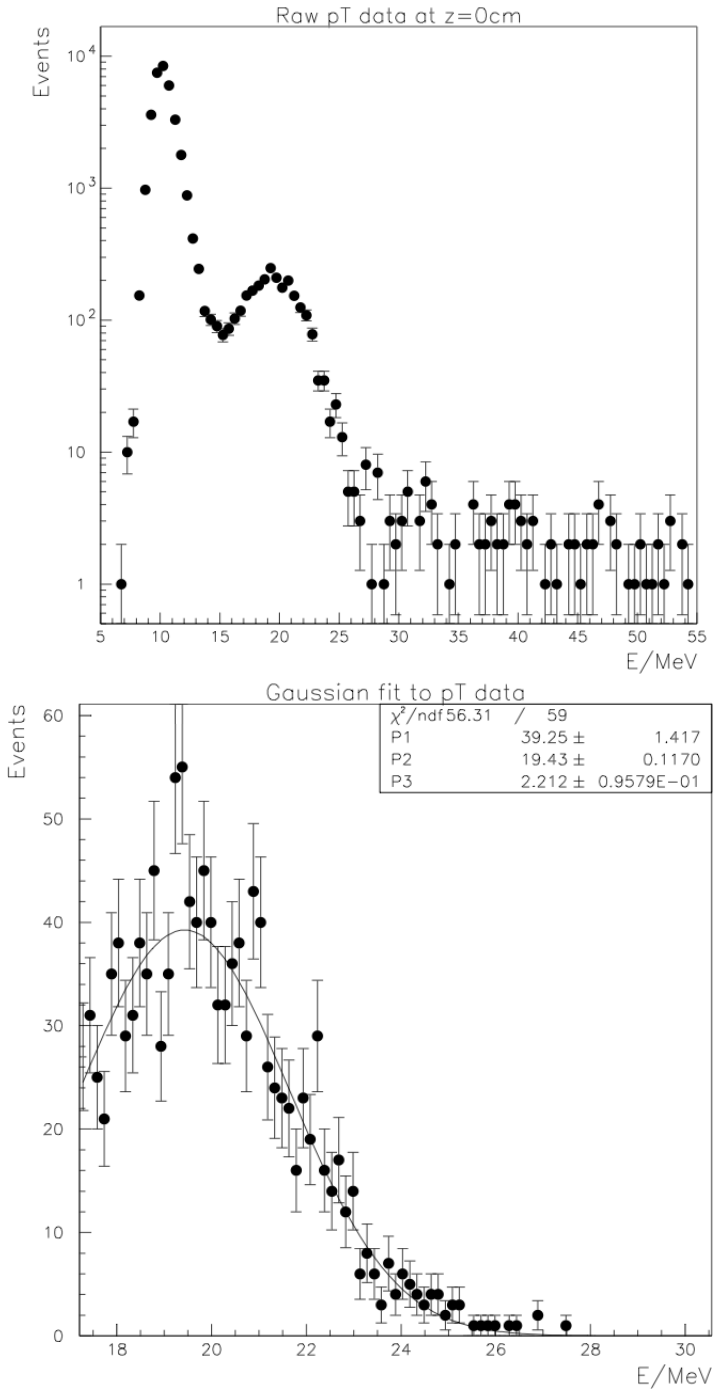

Figure 2.6: The pT data fit to find the resolution at high energy [41].

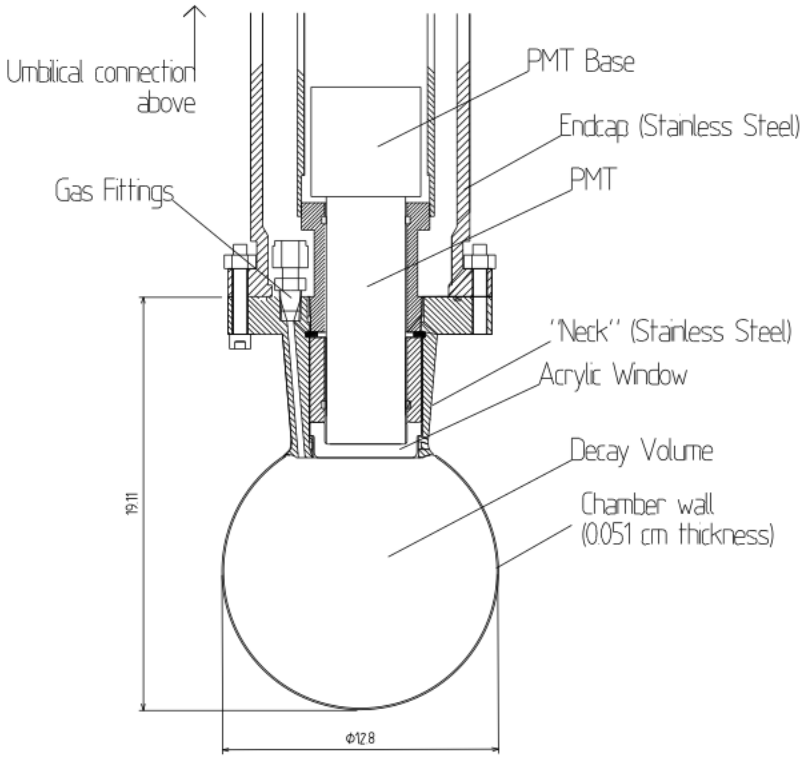

Figure 2.7: Decay chamber of the ^{8}Li source as shown in Reference [42].

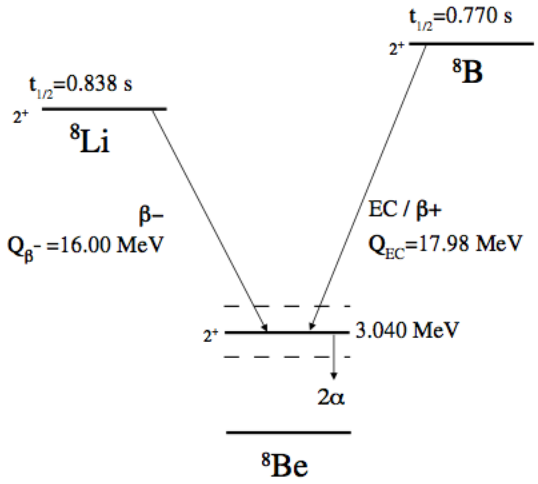

Figure 2.8: Decay scheme of ^{8}Li [42].

applied to act as a wavelength shifter. To aid in the wavelength shifting, a small amount of N_2 gas is also added to the helium gas. This light is then detected by the PMT on the source so that the beta particles can be tagged for analysis. Reference [42] is a published Nuclear Instruments and Methods article which contains in-depth details about the source.

Figure 2.9 shows the comparison of the data taken with the ^{8}Li source, compared to the ^{8}Li Monte Carlo. This agreement is used to verify our characterization of the SNOs energy response to electrons in the ^{8}B neutrino energy region.

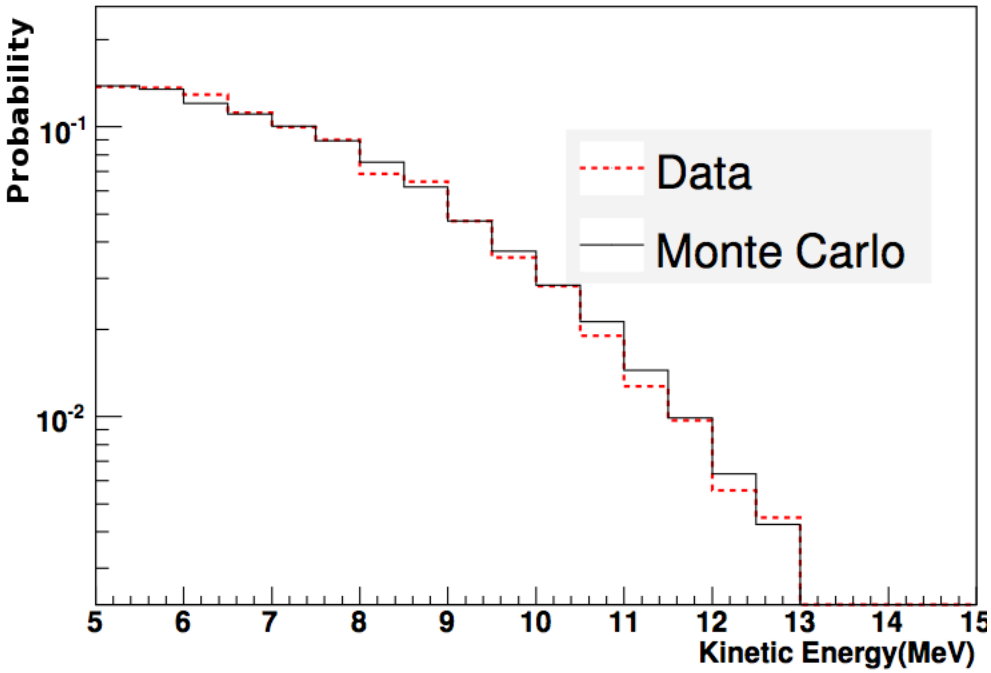

Figure 2.9: The ^{8}Li data compared to the ^{8}Li MC [43]. This comparison is used to verify the detectors energy response up to the end of the ^{8}B energy spectrum.

2.1.4 Michel Electrons

Michel electrons come from cosmic ray muons that decay within the SNO volume. The muon can decay via the modes:

$$\mu^- \rightarrow e^- + \bar{\nu_e} + \nu_\mu \tag{2.3}$$

or

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu_\mu}. \tag{2.4}$$

These decays produce Čerenkov electrons (or positrons) which appear as a high energy event. These events can only be separated out of the data analysis because of their timing. Each event starts with a muon, which creates Čerenkov light that is detected by the PMT array; then any event within a 5 μs window is tagged as a decay electron. This 5 μs window is approximately 2.5 muon lifetimes. If an event is detected in this window we call this event a *re-trigger*. If a re-trigger event is detected, then we start the 5 μs window over again, so in the case of a re-trigger event, the cut window can be much longer than 5 μs.

The Michel electron energy spectrum was calculated using the SNO MC, and compared to the events tagged as mentioned above. Figure 2.10 shows the comparison of the data to the MC. These events are removed from the data analysis, but because they are distributed over such high energies, they become an excellent high energy calibration source. Further study of event selection and MC verification can be found in Reference [41].

2.2 Event Energy Estimation

As stated earlier, the most important measurement that is made with the SNO detector with respect to the *hep* analysis is the energy of an event. The

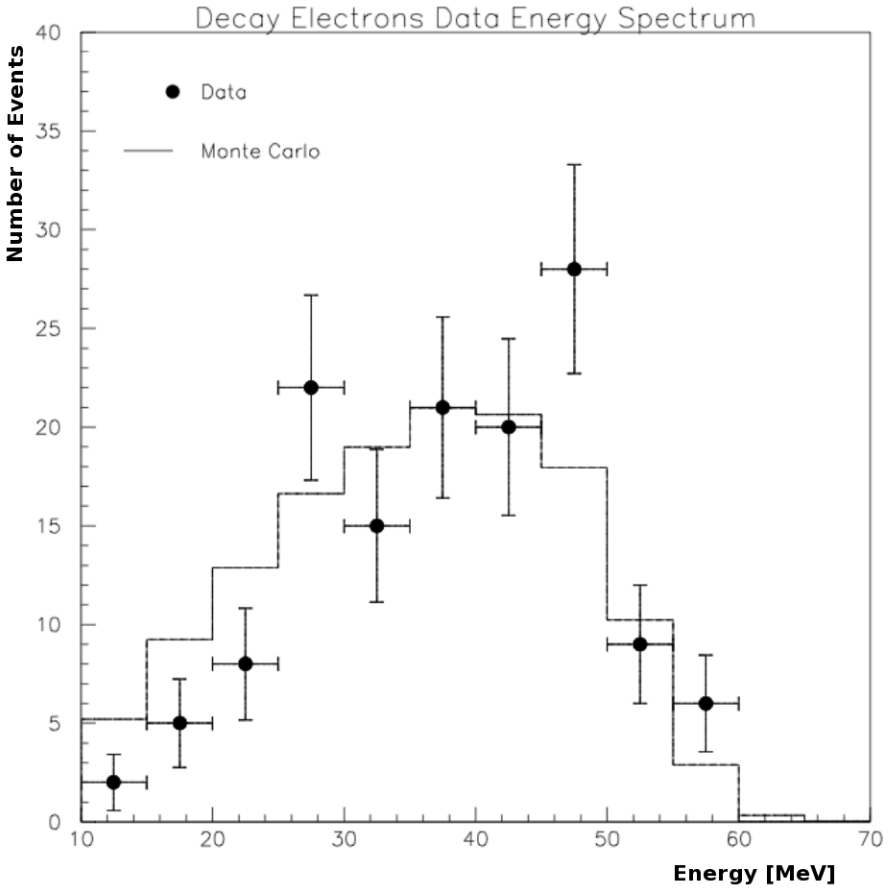

Figure 2.10: Energy spectrum (in MeV) of Michel Electrons in phase I for data and Monte Carlo [41]. The MC is scaled to the number of data events.

previous section described the SNO calibration sources, which show us how events of various energies will be observed. We can now use this as a basis to determine the energy of any event.

The energy of an event at a given position and direction is proportional to the number of PMTs that are triggered by the Čerenkov light caused by the event. Section 2.2.1 will describe how we then turn the number of PMT hits into an energy.

2.2.1 The RSP Energy Estimator

The *hep* analysis uses an algorithm, called RSP as an abbreviation for *energy ReSPonse*, in the SNO analysis software to estimate the energy in all three phases of SNO. The RSP estimator uses information obtained by all of the calibration sources. The *laserball* (explained in Reference [36]) was used to measure the attenuation lengths as a function of wavelength in each of the optical media, the efficiency of the PMT-concentrator module as a function of energy (the so-called angular response of the PMT), and the relative efficiencies of the different PMTs. The ^{16}N and pT sources (Sections 2.1.1 and 2.1.2, respectively) were used to measure the PMT array response to Čerenkov light produced by gamma rays Compton scattering off electrons. The ^{8}Li source allows us to measure the PMT response to beta particles (Section 2.1.3), and a ^{252}Cf source measures the response of PMT array to electrons produced by neutrons captured on deuterium or chlorine.

To estimate the energy of an event in the SNO detector, SNO MC simulations of monoenergetic electrons emitted isotropically from the centre of SNO are used to create a mapping of triggered PMTs to energy. This is done by comparing the number of PMTs triggered to the number of expected to occur at the centre of the detector.

The RSP algorithm starts by correcting the number of PMT triggers,

N_{PMTs}, to the number of PMTs that would have been triggered if that event had happened at the centre of the detector. The corrected number of triggered PMTs, N_{corr} is given by

$$N_{\text{corr}} = N_{\text{eff}} \frac{R_{\text{centre}}}{R} \frac{N_{\text{PMTs}}}{N_{\text{working}}}, \tag{2.5}$$

where $N_{working}$ is the number of PMTs that are turned on in the detector, R is the optical response of the detector, R_{centre} is the optical response at the centre of the detector, and N_{eff} is the number of triggered PMTs in a selected time window (prompt). It is this N_{corr} that is then mapped to the electron energy through MC.

This has been simplified here, as the actual mapping is from the number of Čerenkov photons created by an event to electron energy. The process is essentially the same, except corrections to the response are considered in more detail. The effects of Rayleigh scattering, PMT angular response, and PMT optical and electronic efficiencies must be carefully added and compared to calibration measurements and MC. Further details can be found in Reference [37].

2.3 Three Phase Data and Data Cleaning

The data and MC used in the analysis must pass a series of quality control checks before it is included. These verifications are done on an event by event basis to ensure that an event is a neutrino candidate. There are two types of data cuts:

1. Instrumental or low level

2. Analysis

The *low level*, or *instrumental* cuts, test the raw measurements from the PMTs such as the timing and location of the triggered PMTs. The low level

cuts are used to reject events that are not caused by Čerenkov light. These are predominantly events from *flashers*, large pulses of light occasionally emitted from PMTs, *neck events* which are believed to originate by static electric discharges at the water-acrylic interface, cross-talk and electronic noise (give pulses with anomalous charges, and high correlation inside certain electronic modules or crates), and cosmic-ray muons and muon followers.

The selection of the low level events (events that pass low level cuts) are done on a phase by phase basis. The low level cuts used in phase I and II are the same low level cuts as developed for the low energy threshold analysis (LETA) [6]. The cuts for phase III are the same as those used in the NCD phase analysis [7].

The *analysis* cuts are the cuts that check the calculated values, such as energy, position, and direction, using the events that pass the low level cuts. The specific cuts used in the *hep* analysis include restricting the energy range of interest, the isotropy of the Čerenkov light, and the fiducial volume.

Table 2.1 shows the number of events that pass all the cuts for each phase. The number of events in the *Total* column express how many events pass all the analysis cuts, the number of *Good* events are those that pass all the analysis and low level cuts. These good events that will be used in the *hep* analysis. Figure 2.11 shows the data energy spectrum for the three phases of the analysis.

	Total	Cut	Good
Phase I	22	19	3
Phase II	34	25	9
Phase III	82	62	20

Table 2.1: Number of events passing cuts in the three phase data set.

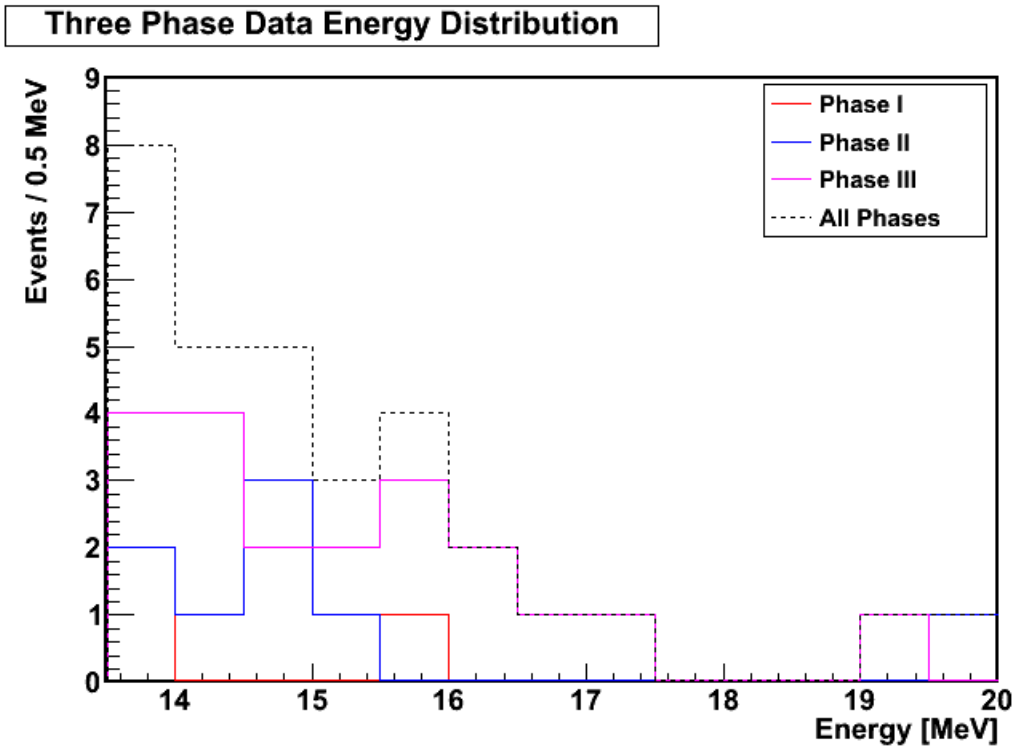

Figure 2.11: Three phase data energy spectrum.

2.3.1 Corrections

Corrections sometimes need to be made to the Monte Carlo or data when discrepancies are found in comparing calibration data to calibration Monte Carlo. Most corrections that are made to the SNO data and MC are not relevant for the *hep* analysis. Most of the effort with respect to the corrections were done at low-energies, to distinguish the signal from the low energy background. The *hep* analysis has a very high low-energy threshold, meaning that these corrections have negligible impact on the *hep* results.

The only correction that needed to be done to the data was to correct a difference observed between the main optical calibration source (laserball) and the event position reconstruction. The data had to be lowered globally 5 cm along the z-axis, which may affect the events that fall into the fiducial volume.

2.4 Backgrounds

The *hep* analysis is only interested in events that occur at much higher energies than the traditional SNO analysis. Most of the backgrounds that occur within the SNO detector originate from low energy sources such as radon. These backgrounds typically have a maximum energy in the 6 MeV range. The *hep* analysis will have a low energy threshold of 13.5 MeV. This threshold makes the *hep* analysis nearly background-free.

The remaining background to the *hep* analysis is the *atmospheric neutrinos*, detailed in Section 2.4.1. The Michel electrons are also in the energy range of interest; however, as mentioned in Section 2.1.4, they are easily identified to be removed from the analysis. It is possible for a Michel electron to be missed if the preceding muon is below the Čerenkov threshold, and therefore missed. This would then look like a high energy electron type event. This happens rarely, and is included in the atmospheric neutrino background, as defined in Section 2.4.1.

The most significant background to the *hep* analysis are the neutrinos produced in the ^{8}B solar reaction, which mis-reconstruct with an anomalously high energy.

2.4.1 Atmospheric Neutrinos

Atmospheric neutrinos, ν_{atm}, are produced when high energy cosmic rays interact with the atmosphere and cause muons and other sources of neutrinos. The neutrinos can be of any flavour of neutrino or anti-neutrino, although they are mostly of the ν_e and ν_μ type. These neutrinos can then travel to and interact in the SNO detector. These neutrinos are at energies that can cause complications in the *hep* analysis, and are included in the signal extraction.

The simulation package NUANCE [44] is used to generate the atmospheric

neutrino flux at the SNO detector and then to create the events inside SNO. These events were then passed to the SNO MC for simulation inside the SNO detector. The NUANCE simulation assumed the atmospheric neutrino model Bartol04 [45] as an input flux for neutrinos produced with an energy above 100 MeV and the Battistoni *et al.* model [46] for neutrinos produced with an energy below 100 MeV.

The resulting NUANCE simulation creates events in the SNO detector. These events are then used as input to the SNO MC, at which point they are propagated to simulate the SNO detector response to the atmospheric neutrino background. The energy spectrum as produced by the SNO MC and NUANCE is shown in Figure 2.12.

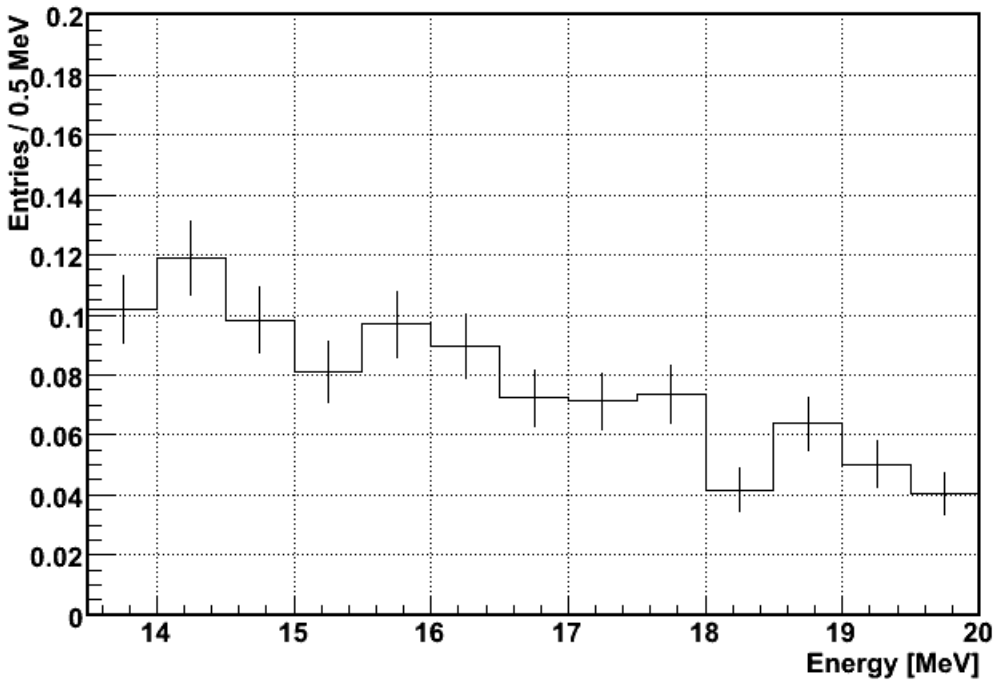

Figure 2.12: The atmospheric neutrino energy spectrum.

The specific sources that make up the atmospheric neutrino background are:

- CC events from the ν_e and $\bar{\nu}_e$ interactions;
- Invisible Michel electrons; that is, Michel electrons where the leading muon is not detected (below Čerekov threshold);
- 15.1 MeV gamma particles from the de-excitation of ^{12}C from an excited state, which was created by a neutrino interaction on ^{16}O;
- Events that were mis-reconstructed as electrons.

The work to characterize and place limits on the flux of atmospheric neutrinos was done for the previous *hep* paper [32][41]. This *hep* analysis uses the previous measurement for phase I, and assumes a constant rate for phase II and III. This implies that the electron (CC and ES signals) energy response is consistent over time in the high energy range. As the detector is the same, the uncertainties on the atmospheric neutrino background is large, and the number of events is expected to be an order of magnitude smaller than the *hep* signal, these approximations are fair.

Section 2.3 discussed many ways in which the SNO data is cleaned. These data cuts remove much of the atmospheric neutrino background. Most events occur close together in time, as most of the events that follow an atmospheric neutrino event are neutron type events. This is cut with the SNO *burst* cut, which looks for events too close together in time.

Section 6.2 will give the measurement of the atmospheric neutrino background, and the corresponding constraint on the signal extraction. Because of the structure of the atmospheric neutrino simulation, the input flux is not straightforward. In this thesis, the atmospheric neutrino result will be measured as a rate in the SNO detector.

Chapter 3

Probability Distribution Functions for *hep* Neutrinos in the Three Phases of SNO

This section will outline the SNO MC software and the probability distribution functions (PDFs) that make up the expected signals. The probability distribution function describes how likely it is that a given event has a given value of an observable (Section 3.2). This section will describe the details of how the probability distribution functions are estimated.

3.1 SNO Monte Carlo

The SNO Monte Carlo and Analysis software (SNOMAN) is a set of simulation and analysis routines that both model the SNO detector and process the data. The modelling uses information such as detector geometry, particle propagation tracking, and particle direction. During the data processing, various parameters are estimated for each event, including the position, direction, energy, and angle to the Sun. The neutrinos originate at the Sun and propagate to the SNO detector. Those which interact via the CC or ES reactions produce electrons directly; the NC reaction produce electrons indirectly. These electrons move faster than the speed of light in heavy water, and consequently produce Čerekov light. This light then travels through the water to reach the

PMT array where it is detected.

The spacial and temporal distributions of the PMT triggers are used to determine the direction and position of the source of the light. Čerenkov light has a characteristic cone shape. SNOMAN has been designed to output parameters that allows easy determination of the event shape. From the data, other information can also be inferred from an event, such as the direction the source of the light was travelling and the energy of the event.

SNOMAN has been designed so that we can simulate, through a Monte Carlo (MC) technique, the expected SNO signals and backgrounds and can generate a data set consisting of simulated events that can be analyzed in exactly the same way as actual data. This allows us to predict the response of the detector to compare to the observed data. Given a type and distribution of particles within SNO, SNOMAN can track all relevant particles that can be observed, or lead to particles that can be observed within SNO. It uses the MCNP [47] and FLUKA [48] packages to track neutrons and hadrons. The MCNP package is a specialized neutron package, and FLUKA is a general package to simulate hadrons. The EGS4 [49] package is used to track gamma particles and electrons.

Starting from a theoretical production of neutrinos in the Sun, as shown in Figure 1.1, the Monte Carlo can produce expected results to compare to our data. In the *hep* analysis we are interested in two solar spectra, from the ^{8}B and *hep* neutrinos. Differences that arise between the SNOMAN simulation and the data are quantified as part of the systematic uncertainties. The implementation of the systematic uncertainties for this analysis is described in Section 6.1

3.2 Observables

An observable is a physical distribution that can be used to distinguish one signal from another. There are two observables used in this analysis: 1) energy, and 2) the angle between the event direction and the Sun. These observables are explained in the following sections.

3.2.1 Energy

The energy of any event in the SNO detector is estimated by the RSP, discussed in Section 2.2.1. In Equations (1.13), 1.14, and 1.15, the three possible interactions of neutrinos with heavy water in SNO are listed. Each of these interactions results in a distinct energy spectrum, which can be used to distinguish *hep* events from other events. Figure 3.1 shows the energy spectrum from these equations from the ^{8}B and *hep* signals.

The neutral current reaction produces neutrons, which are subsequently detected. Since the neutrons are captured as thermal neutrons, the energy distributions from any initial reaction cannot be distinguished and any *hep* event that interacts via the NC reaction will contribute to the overall NC normalization; that is, it is impossible to distinguish any NC contributor from another. The NC reaction has an energy distribution that does not exceed 10 MeV, which is well below the 13.5 MeV low energy threshold chosen for the *hep* analysis. Thus, the NC signal is not considered in this analysis.

The *hep* events that interact via the CC and ES distribution will be combined into one signal. This is done for two reasons: 1) the cross section of each of these reactions is known, so we are not physically free to let their relative normalizations vary with respect to each other; 2) this allows us to increase the number of events we can possibly observe. In the *hep* analysis, we are not trying to determine information about the shape of the neutrino energy

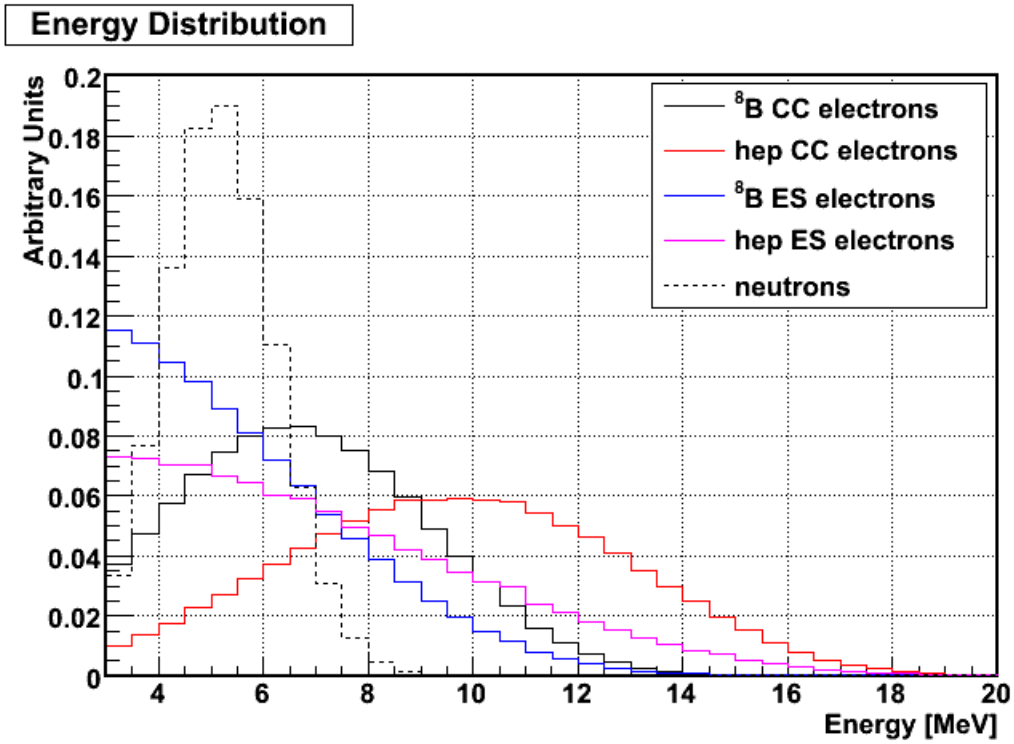

Figure 3.1: ^{8}B and *hep* energy spectra for the CC, ES, and NC reactions in SNO. NC events are captured as thermal neutrons, and thus all neutrons fit into this label.

spectra; rather, we are more interested in the overall number of *hep* events (the *hep* flux). This is called an energy constrained analysis.

Other SNO analyses that focus on the ^{8}B spectrum determine the shape of the energy spectrum (unconstrained analysis). This is possible for a ^{8}B analysis because of the larger solar flux and extended energy range, resulting in many more events and increased sensitivity to changes in survival probability as a function of energy.

SNOMAN does not consider neutrino oscillations as it propagates the neutrinos. This is done so that the distortion due to the neutrino oscillations can be measured by the data directly. The size of the distortion to the data is the difference between the MC energy distribution and the data distribution.

Since the *hep* analysis is constrained we will not attempt to measure this distortion. Instead, this analysis explicitly distorts the MC energy shape by the expected energy distortion due to neutrino oscillations. To do this, we multiply the neutrino energy spectrum used to generate the SNOMAN simulation by the P_{ee} distortion, described in Section 1.1.3, using the best fit neutrino oscillation parameters. Using the SNO best fit parameters in the LMA region from Reference [6], $\tan^2\theta_{12} = 0.457^{+0.038}_{-0.042}$ and $\Delta m^2 = 5.50^{+2.21}_{-1.62} \times 10^{-5}$ eV^2, we obtain the P_{ee} distortion from Figures 1.3 and 1.4 to make Figures 3.2 and 3.3 respectively. Also included in these Figures is the size of change to the distortion with a fluctuation of one standard deviation (1σ) to the best fit parameters.

After combining the CC and ES signals and distorting them by Figures 3.2 and 3.3, the *hep* analysis expects signals from three sources, ^{8}B CC+ES, *hep* CC+ES and atmospheric neutrinos. The expected energy probability distributions functions (PDFs) are shown in Figures 3.4 to 3.6 for the three phases.

It can be seen from Figures 3.4 to 3.6 that the atmospheric neutrino signal is distinctly different from the other signals and therefore easy to separate.

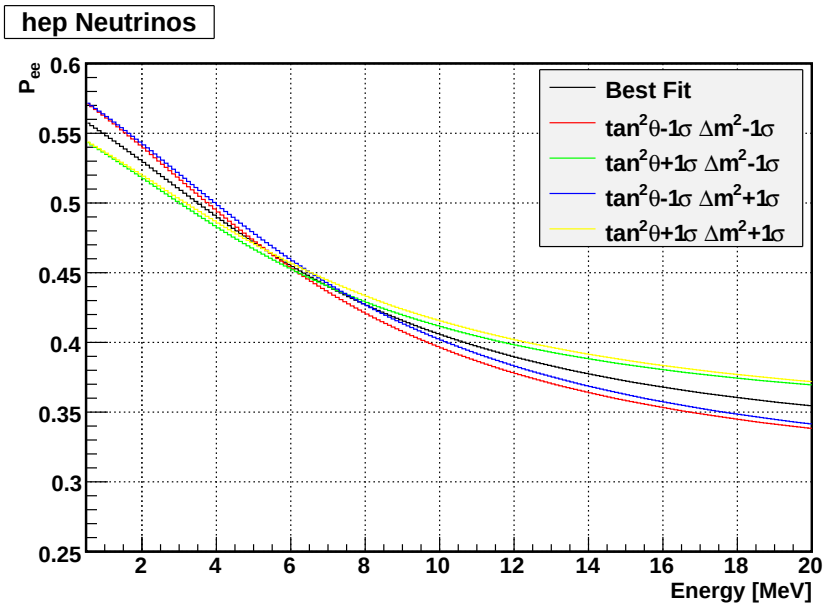

Figure 3.2: Probability that an electron neutrino of a given energy produce by the *hep* reaction will remain an electron neutrino at the SNO detector for the SNO best neutrino oscillation parameters.

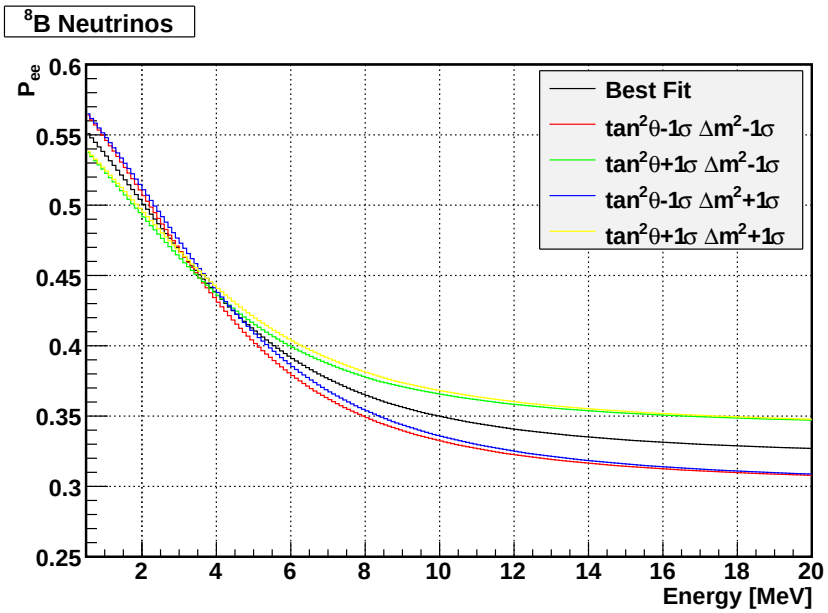

Figure 3.3: Probability that an electron neutrino of a given energy produce by the ^{8}B reaction will remain an electron neutrino at the SNO detector for the SNO best neutrino oscillation parameters.

The ^{8}B CC+ES and *hep* CC+ES signals however are similar. The difference only occurs at higher energies, the focus of this analysis.

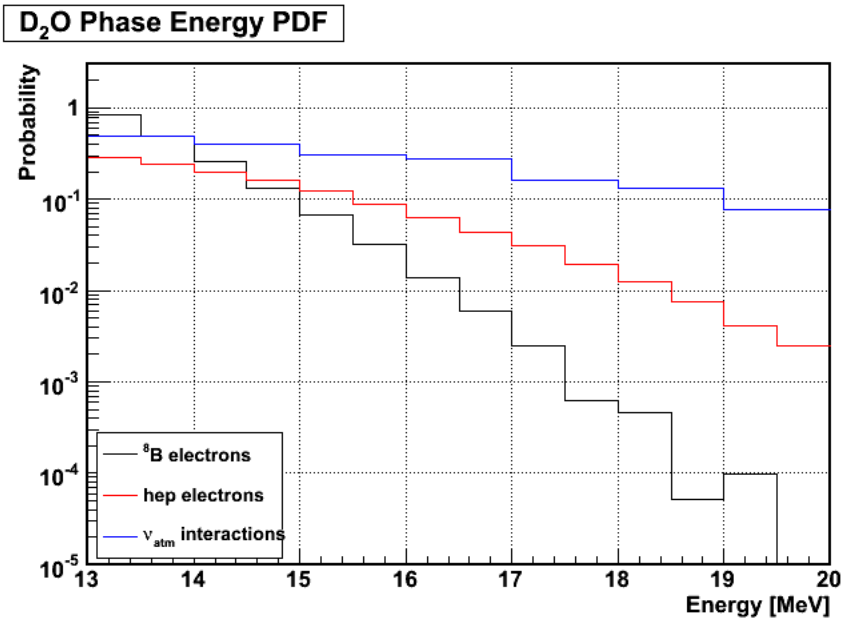

Figure 3.4: Energy probability distribution functions for the D_2O phase.

If we now scale each of the PDFs in Figures 3.4, 3.5, and 3.6 by the SSM as determined by SNOMAN we can see the predicted three phase energy spectrum. This is shown in Figure 3.7. We can again see that the dominant signal at lower energies is the ^{8}B signal, and *hep* at higher energies, while the ν_{atm} signal is nearly flat.

3.2.2 $\cos\theta_{Sun}$

The $\cos\theta_{Sun}$ observable is defined as the angle between the direction of the Čerenkov light cone and the direction to the Sun. This can be useful to distinguish events that have a dependence on the direction to the Sun from those that do not. In SNO this is used to aid in separating the NC signal, which has no dependence on the direction to the Sun, from the CC and ES signals, which have a mild and strong dependence, respectively. It also helps with background radiation, which also has no solar direction dependence.

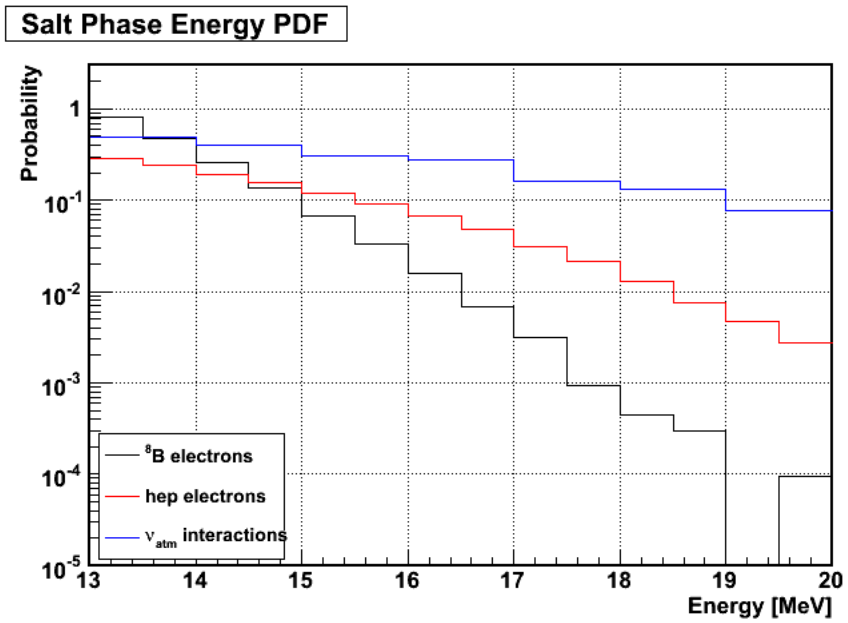

Figure 3.5: Energy probability distribution functions for the salt phase.

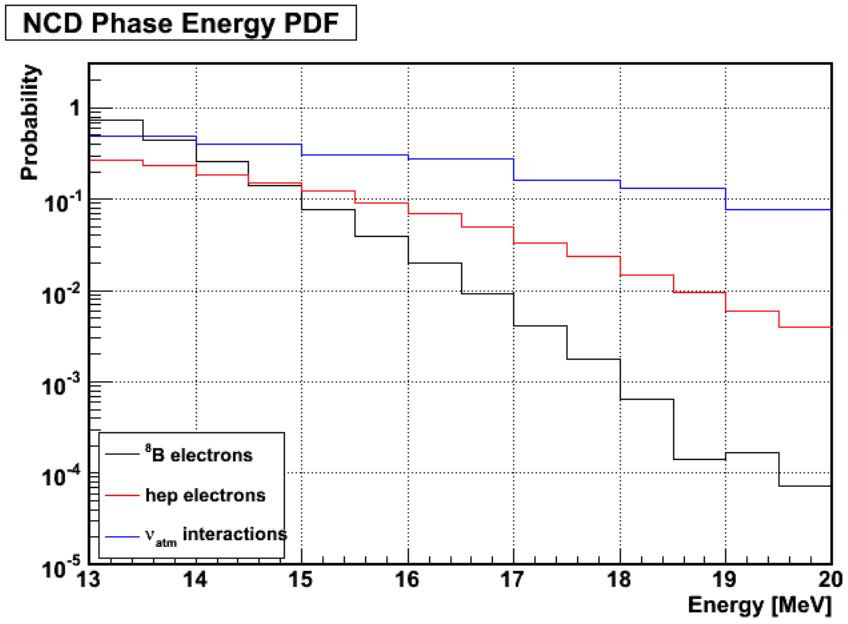

Figure 3.6: Energy probability distribution functions for the NCD phase.

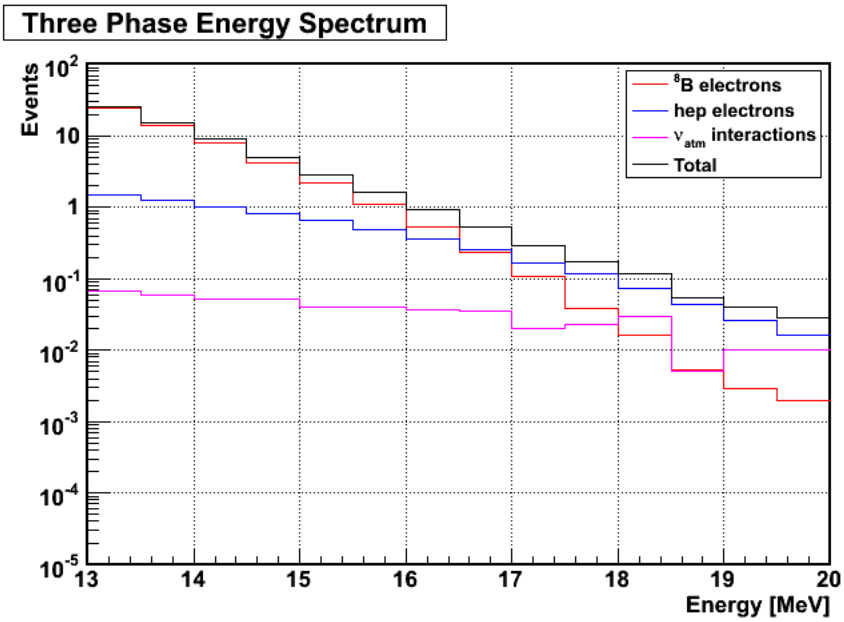

Figure 3.7: Predicted three phase energy spectrum.

Because we are observing the CC+ES reaction for both ^{8}B and *hep* signals, the $\cos\theta_{Sun}$ distribution is very similar, and thus has very little influence in separating the signals. However, the atmospheric neutrino flux has some $\cos\theta_{Sun}$ dependence, but with so few events expected this will have little impact.

Because the effect of this observable is so small the PDF is coarsely binned. The $\cos\theta_{Sun}$ PDFs for the three phases are shown in Figures 3.8 to 3.10.

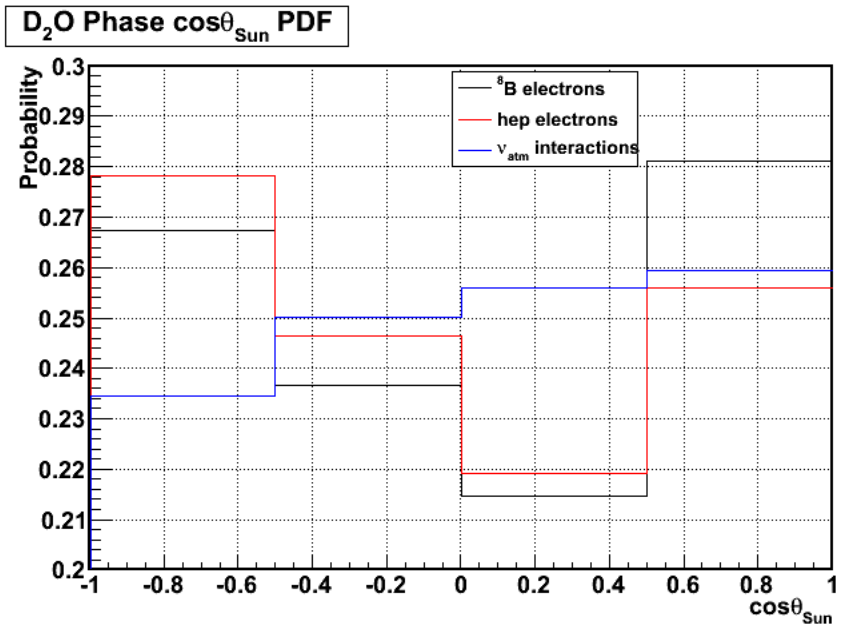

Figure 3.8: $\cos\theta_{Sun}$ probability distribution functions for the D_2O phase.

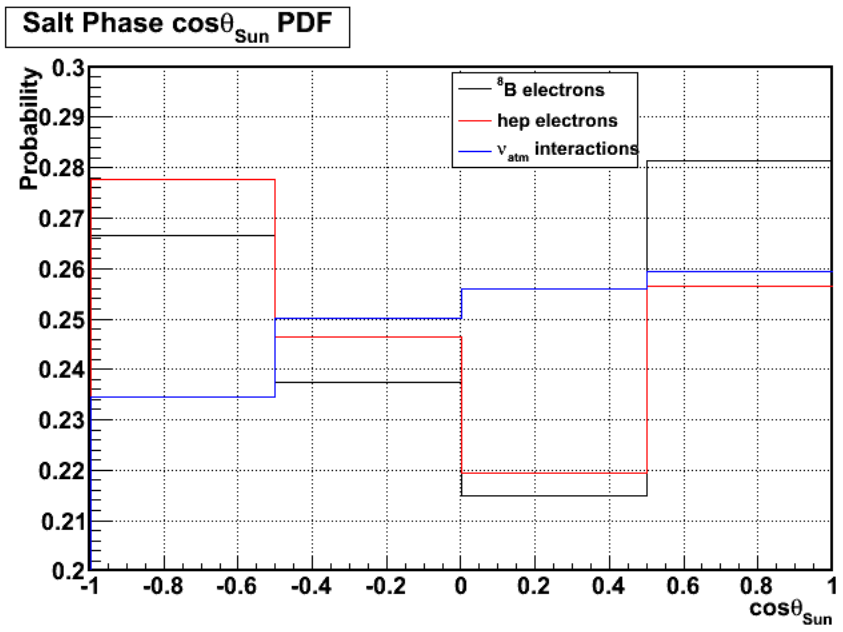

Figure 3.9: $\cos\theta_{Sun}$ probability distribution functions for the salt phase.

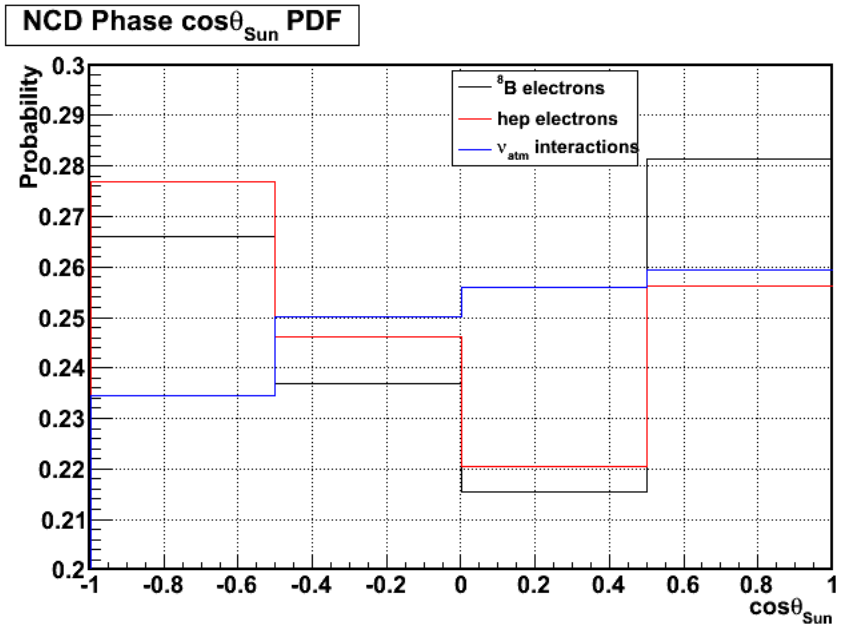

Figure 3.10: $\cos\theta_{Sun}$ probability distribution functions for the NCD phase.

Chapter 4

Signal Extraction

The goal of the signal extraction is to precisely determine the contributions of each signal type to our data set based on our knowledge of the signal characteristics. The signals that make up the data set have been outlined in Chapter 3. The signal extraction method should also allow for constraints to be placed on any signal due to previous measurements and calibrations. The signal extraction method will have the ability to *float systematic uncertainties* as will be discussed in section 6.1. This section will also outline the statistical method used to separate the data into its component signals. It will describe how the method was verified for accuracy. This section will finally detail how the 90% confidence limit will be determined in the case no statistically significant signal can be measured.

4.1 Likelihood Function

The likelihood function provides a way to statistically separate a data set into the various signals that compose the data. We start with a function, $f(x;\theta)$, which is the probability distribution function; i.e. $\int f(x;\theta)\,dx = 1$, and the value of $f(x;\theta)dx$ gives the probability, for a given θ, that x lies within a bin of size dx. We then measure the observable, x, multiple times, $(x_1, ..., x_n)$. For the set of multiple events, i, the likelihood is defined as the function:

$$L(\theta) = \prod_{i=1}^{data} f(x_i; \theta), \tag{4.1}$$

where $L(\theta)$ is the likelihood for a given parameter θ. In our case, the x_i are our observables and θ is the normalization of the signal flux, which we will now call Φ. The function $f(x_i; \Phi)$ is the joint probability distribution function of x_i.

If we were to repeat this same measurement k times we would expect the measured parameter Φ to be Poisson distributed around its true value. The likelihood is then multiplied by a Poisson probability, resulting in the *extended likelihood function*:

$$L(\Phi) = \frac{\Phi^n}{n!} e^{-\Phi} \prod_{i=1}^{n} f(x_i; \Phi) = \frac{e^{-\Phi}}{n!} \prod_{i=1}^{n} \Phi f(x_i; \Phi). \tag{4.2}$$

where n is the number of events in the data set.

For computational simplicity it is usually written in the form of log likelihood function

$$\log L(\Phi) = -\Phi + \sum_{i=1}^{n} \log \left(\Phi f(x_i; \Phi) \right), \tag{4.3}$$

where all the constant terms have been dropped. We can drop these terms because the constant terms do not contribute to the overall normalization of the parameters. The PDF $f(x_i; \theta)$ is made up of several signals, so we can write it as

$$f(x; \Phi) = \sum_{j=1} \Phi_j f_j(x_i), \tag{4.4}$$

and

$$\Phi = \sum_{j=1}^{m} \Phi_j \tag{4.5}$$

giving,

$$\log L(\Phi_j) = -\sum_{j=1}^{m} \Phi_j + \sum_{i=1}^{n} \log \left(\sum_{j=1}^{m} \Phi_j f_j(x_i) \right). \tag{4.6}$$

where m is the number of signals.

The maximum value of the log likelihood function results in the best set of parameters, Φ.

4.1.1 Extension to Three Phases

To extend the signal extraction over three phases, we must acknowledge that the PDF of each distribution, $f_j(x_i)$, will be different in each phase. We must then sum the log likelihood from each phase.

The modified log likelihood will look like:

$$\begin{aligned} \log L = & \sum_{i=0}^{data} \log \left(\sum_{j=0}^{signal} \Phi_j f_j\left(x_i\right) \right) - \sum_{j=0}^{signal} \Phi_j \\ & + \sum_{i'=0}^{data'} \log \left(\sum_{j=0}^{signal} \Phi'_j f'_j\left(x'_i\right) \right) - \sum_{j=0}^{signal} \Phi'_j \\ & + \sum_{i''=0}^{data''} \log \left(\sum_{j=0}^{signal} \Phi''_j f''_j\left(x''_i\right) \right) - \sum_{j=0}^{signal} \Phi''_j, \end{aligned} \tag{4.7}$$

where the prime and double prime notation denotes the addition of the salt and NCD data sets to the D_2O data set, respectively.

4.1.2 Adding a Constraint to the Likelihood

In cases where extra information is known, presumably from some calibration or measurement, it is possible to influence the likelihood in a way that reflects this knowledge. We can modify our likelihood by prior knowledge about some parameter α, by applying a Gaussian constraint $G(\alpha)$, to our likelihood by

$$L' = L\ G(\alpha) \tag{4.8}$$

where L' is the modified likelihood. In the case of a log likelihood, this becomes:

$$\log L' = \log L + \log G(\alpha), \tag{4.9}$$

This means that Equation (4.6) would be modified to be:

$$\log L = \sum_{i=0}^{data} \log \left(\sum_{j=0}^{signal} \Phi_j f_j \left(x_i \right) \right) - \sum_{j=0}^{signal} \Phi_j + \log(G(\alpha)) \tag{4.10}$$

Explicitly, $G(\alpha)$ is:

$$\log(g(\alpha)) = -\frac{1}{2} \left(\frac{\alpha - \mu_\alpha}{\sigma_\alpha} \right)^2 \tag{4.11}$$

where μ_α is the mean measured value of α and σ_α is the standard deviation of the measured value of α. Since Equation (4.11) constrains the likelihood function, we call it a *penalty term.*

4.1.3 Bayesian Prior

This analysis uses a Bayesian approach. That is, we start with Bayes' Theorem:

$$P(\theta|x_i) = \frac{P(x_i|\theta)P(\theta)}{P(x_i)} \tag{4.12}$$

where $P(\theta)$ is the prior probability of θ; this is assumed to be known before starting. $P(\theta|x_i)$ is the posterior probability, which depends on x_i, $P(x_i|\theta)$ is the likelihood of x_i given θ, and our likelihood function, $P(x_i)$ is the probability of x_i which is just a normalization constant [50].

Because our parameter θ is the number of events, we will consider that we have a uniform prior above zero, as shown in Equation (4.13). This means that the resulting posterior distribution is the likelihood distribution. The prior must be defined this way because we cannot allow negative (unphysical) values for the number of events in any signal.

$$P(\theta) = \begin{cases} 1 & \text{if } \theta \geq 0 \\ 0 & \text{if } \theta < 0 \end{cases} \tag{4.13}$$

4.2 Markov Chain Monte Carlo

To find the best set of parameters that maximizes the likelihood in the previous sections, we use a Markov Chain Monte Carlo (MCMC). The MCMC uses the

Metropolis-Hastings algorithm [51] to efficiently map out the likelihood space. The Metropolis-Hastings algorithm is a simple technique to build a Markov Chain. The algorithm is:

1. Select an initial set of parameters, θ.

2. Smear this set of parameters. Use a symmetric function, such as a gaussian centered on the parameter, with a width of $\frac{1}{3}$ the statistical uncertainty (θ').

3. Calculate the likelihood with these initial parameters (P(θ')).

4. Smear the previous set of parameters. Again, use a symmetric function centered on the parameter, with a width of $\frac{1}{3}$ the statistical uncertainty (θ).

5. Calculate the likelihood with these new parameters (P(θ)).

6. Take the ratio, R, between the new likelihood and the old. Take this number or one, whichever is lower.

$$R = \text{Min}\left(\frac{P(\theta)}{P(\theta')}, 1\right) \tag{4.14}$$

7. Randomly select a number between zero and one (α).

8. If the ratio, R, is greater than the random number, α, record the new set of parameters. If not, record the previous set.

$$\alpha = \begin{cases} \leq R & \text{then } P(\theta) \text{ Accepted} \\ > R & \text{then } P(\theta) \text{ Rejected} \end{cases} \tag{4.15}$$

9. Repeat steps 4–8.

Continue this procedure until the likelihood distribution is smooth. For each run, the initial steps of the Markov chain must be removed as *burn-in*. The initial steps are used to allow the chain to converge to the most probable points in the likelihood space. Section 7.1 will show the tests of convergence used.

This algorithm starts from an initial set of parameters and *walks* to the best-fit position in the likelihood space. This walking can be seen in Figure 4.1 and is removed with the burn-in cut. A simple check will verify that the algorithm has converged on the right location in likelihood space. This check tracks the value of the likelihood for each step in the Markov chain (See Figure 4.2). We can see that after 800 steps the Markov Chain has converged around the maximum of the likelihood. The Markov Chain will sample the space around the maximum likelihood for a statistically significant number of points, which can be seen by the slight variation around the converged likelihood in Figure 4.2. It has been proven that the Metropolis-Hastings algorithm will result in a sampling of parameter space in proportion to the likelihood.

The MCMC method has a significant advantage over minimization routines to find the best fit parameters of the likelihood function. Minimization routines, such as *minuit* [52], have trouble with *choppy* likelihood shapes and have problems with large numbers of parameters. The MCMC method produces the likelihood shape and allows the interpretation to be shown more directly. For example, we can see that in Figure 4.3 the signal extraction has produced a smooth posterior distribution (likelihood space when normalized). A simple way to represent the best fit value and its uncertainty is to use the mean and RMS of the posterior distribution. This choice is independent of the MCMC method giving much more freedom in the analysis. There is no limit to the number of parameters that can be used in the MCMC, but a large number of parameters can lead to longer convergence times. The MCMC method also

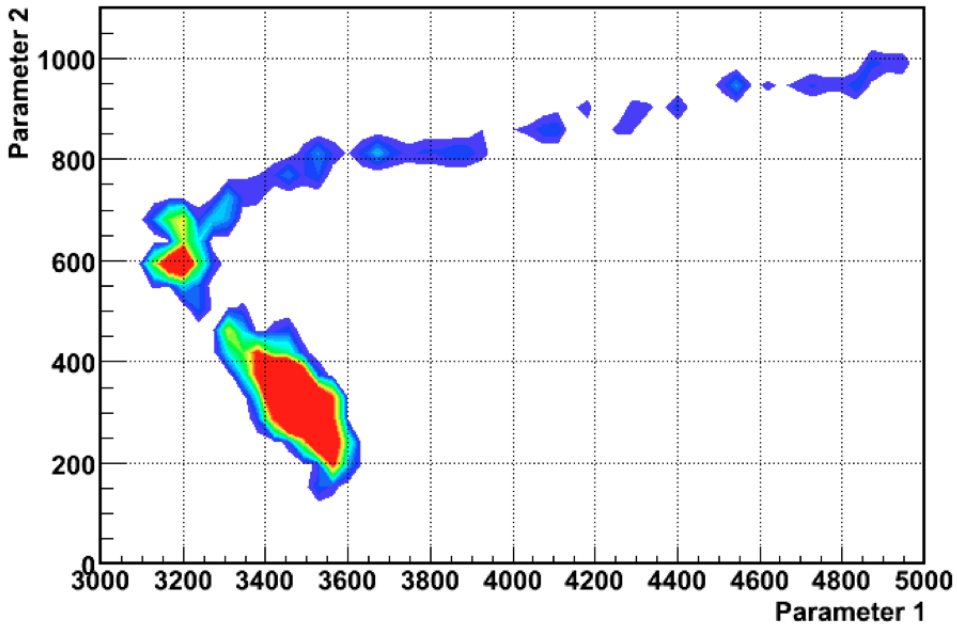

Figure 4.1: A simple example of a Markov Chain using two parameters chosen based on the Metropolis-Hastings algorithm. After the burn-in period, the algorithm ensure that the parameters are selected in proportion to their likelihood. During the burn-in period, which starts at arbitrary parameters, regions with lower likelihood are sampled. The colour scale represents the number of MCMC steps, the blue points are the lowest on the scale, and the red region is the largest.

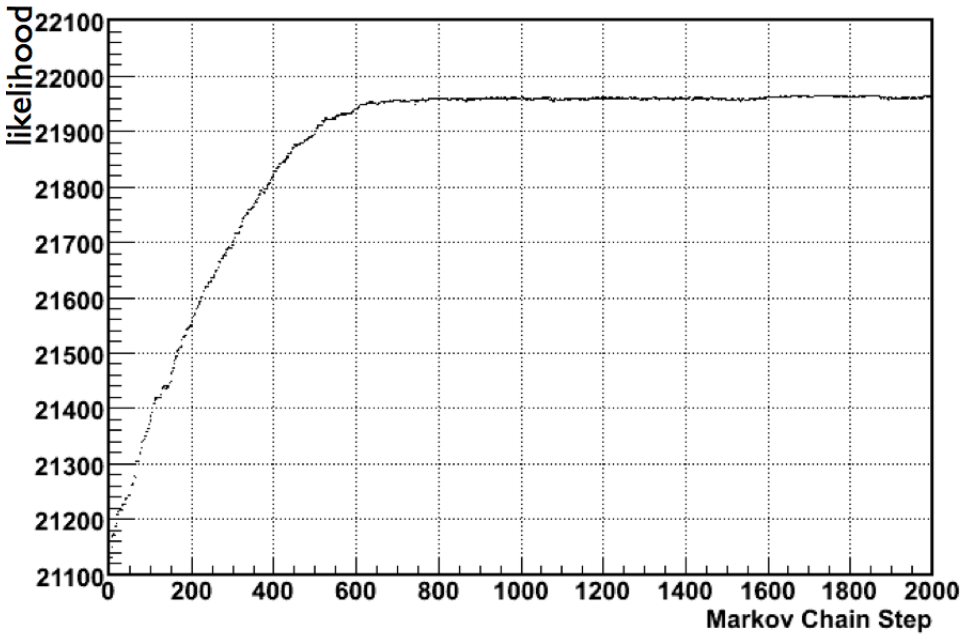

Figure 4.2: A check of the convergence of the Metropolis-Hastings algorithm. We see convergence after 800 Markov chain steps.

does not require that the likelihood space is a Gaussian distribution. Since it just samples the likelihood space, even non-Gaussian likelihoods can be sampled accurately.

4.3 Ensemble Test

Ensemble tests are used to verify that the signal extraction technique is correct. The procedure is to create sets of data, with a known number of events for each class, such as data drawn from the MC. Section 4.3.1 will explain the creation of the data sets. Each data set in the ensemble then has the signal extraction fit preformed individually and the result recorded. Each result should then be in statistical agreement with the known answer. The method to determine agreement is explained in Section 4.3.2.

Simply put, we use the likelihood function to generate the ensemble data,

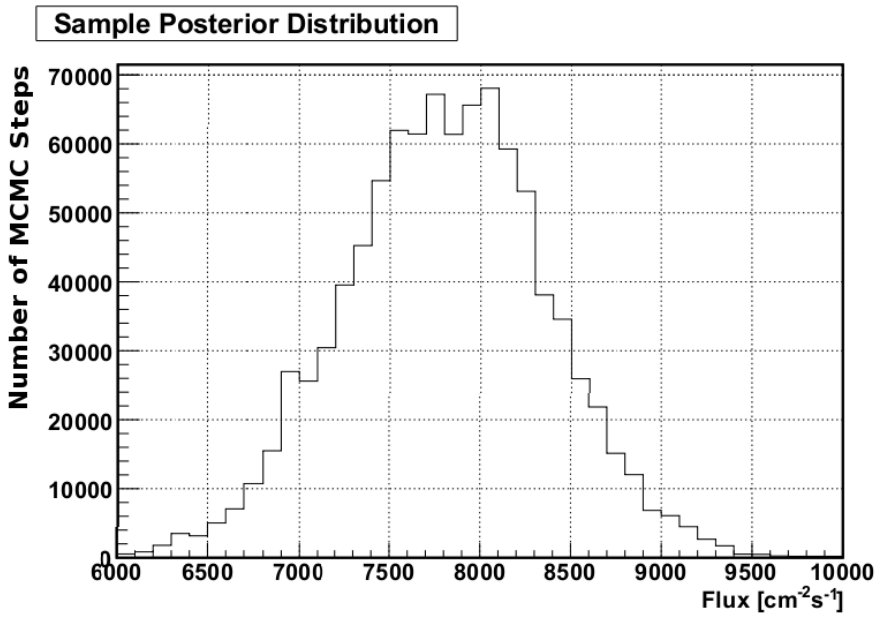

Figure 4.3: Sample posterior distribution, using data drawn randomly for the MC as a sample data set for the signal extraction.

we then use the likelihood function to fit the data, and measure our agreement. This gives us confidence that we understand the accuracy and precision as well as limitations of our signal extraction technique.

4.3.1 Fake Data Set

Each fake data set is drawn from the SNOMAN simulation. Each data set contains events from each expected signal type. In the case of the *hep* analysis, these signals are *hep*, ^{8}B, and atmospheric neutrinos. In each case only the CC and ES events are used because the NC events are below our energy threshold. Each data set contains data from all three phases, generated with phase-specific detector parameters.

Normally the number of events of each signal are chosen by using the expected results. This was not done for our test because the expected number of *hep* and atmospheric neutrino events is so low that the mean of the posterior distribution may not be the same as the number of events in the fake data

set. For example, the *hep* signal may be low enough that the resolution of the posterior distribution can be statistically consistent with zero. Our prior (Equation (4.13)) restricts us to a non-negative mean which necessarily leads to a biased mean.

When there are many events in the data set, the posterior distribution is a Gaussian distribution and the mean is a good representation of the expected result. In testing this signal extraction, the number of events of each signal is increased until the signal is large enough that the mean of each posterior distribution is the same as the mean as determined by a Gaussian fit.

When creating the fake data sets, each data set has a fixed number of events, which is selected according to a Poisson distribution about the expected number of events. The mean of the Poisson distribution is the expected number of events. Table 4.1 shows the mean of the Poisson distribution used for each signal.

Signal	Mean
hep (Phase I)	207
^{8}B (Phase I)	238
ν_{atm} (Phase I)	10
hep (Phase II)	207
^{8}B (Phase II)	238
ν_{atm} (Phase II)	10
hep (Phase III)	207
^{8}B (Phase III)	238
ν_{atm} (Phase III)	10

Table 4.1: Mean number of events used to create ensemble data sets. The number of events in each signal are arbitrarily increased from expected to allow the signal extract fit values that are not consistent with zero. This was only done as a check for errors in the routine, and no information about the sensitivity about the *hep* signal extraction should be inferred from this study.

The reason we use a Poisson distribution is because that is the statistical probability distribution that naturally defines a process with a fixed rate of independent events. If for example we were to run SNO multiple times, we

would not expect to measure the same flux in each experiment. The flux measurement in each experiment would be distributed as a Poisson distribution around the *true* flux.

This distribution has also been taken into consideration in the definition of the extended maximum log likelihood function used for the signal extraction, as shown in Equation (4.6).

4.3.2 Bias and Pull

Once ensemble data sets are created as described in Section 4.3.1 then we must quantify how well the signal extraction fits each signal. There are two sides to this question. 1) The bias test which measures how well we determine the mean; 2) The pull test which measures how well we determine the uncertainty.

To calculate the pull and bias for each parameter in an individual data set we:

1. Start with the parameter posterior distribution.
2. Remove the burn-in MCMC steps, and use the resulting posterior distribution.
3. Fit a symmetric Gaussian to the peak using the range ± 2 standard deviations. The size of the bins have little impact on this, as long as there are more than 10 bins within 1 standard deviation.
4. Use the mean of the Gaussian as μ_{fit}, and the width of the Gaussian as σ_{Gauss}.
5. Use these to find the bias and pull, Equations (4.16) and 4.18 respectively, for the given parameter in the data set.

The bias for each signal for an individual data set is calculated by:

$$\text{bias} = \frac{\mu_{\text{fit}} - \mu_{\text{input}}}{\mu_{\text{input}}}, \tag{4.16}$$

where μ_{fit} is the mean of the fit and μ_{input} is the mean of the Poisson distribution for the signal being calculated.

Once the biases for each data set (and signal) are calculated we can then look at the bias distributions. Figure 4.4 shows the bias distribution for each signal. Each parameter distribution has one entry from each of the ensemble data sets. These are Gaussian distributed. The mean is said to be the bias for the signal, and the error on the bias is given by the standard error:

$$\sigma_{\text{bias}} = \frac{\sigma_{\text{Gauss}}}{\sqrt{n}}, \tag{4.17}$$

where σ_{Gauss} is one standard deviation of the Gaussian fit. This is equal to the RMS of the distribution, if the distribution is truly Gaussian, and n is the number of ensemble fake data sets. Also, if the fit is Gaussian, σ_{bias} is the same as the error on the mean, μ_{fit}.

The bias mean and uncertainties for each signal are summarized in Figure 4.5. The bias for each signal is less than 3%, and less than 2 standard deviations from zero. This is small enough to have negligible impact on the signal extraction.

Like the bias, the pull calculation is done for each ensemble fake data set (and signal). The pull is calculated using:

$$\text{pull} = \frac{\mu_{\text{fit}} - \mu_{\text{input}}}{\sigma_{\text{fit}}}, \tag{4.18}$$

where μ_{fit} is the mean of the fit, μ_{input} is mean of the Poisson distribution for the given signal and σ_{fit} is the uncertainty of the fit for the given signal.

Once the pulls have been calculated for each fake data set the pull distributions for each signal can be made, as shown in Figure 4.6. Again, like the bias distributions, there is one entry in each pull distribution for each ensemble data set. These distributions are also Gaussian in nature. The mean and one standard deviation of this distribution is used to measure the pull and uncertainty.

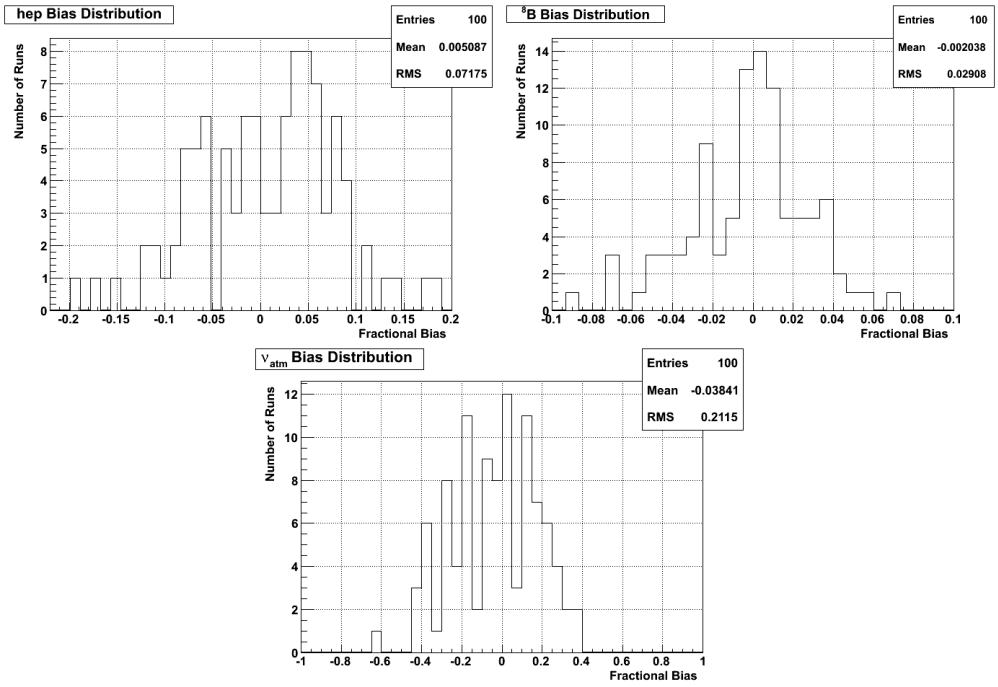

Figure 4.4: Bias distribution for each signal.

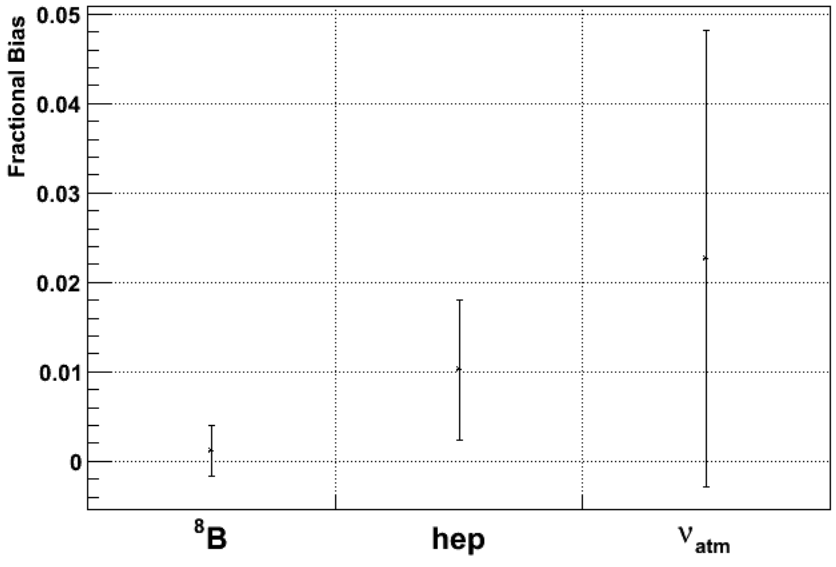

Figure 4.5: Summary of the bias distributions.

If the signal extraction is working properly, we expect the uncertainty of the pull to be one. This is because the μ_{fit} term in Equation (4.18) is drawn from a Poisson distribution. So if μ_{fit} is distributed around μ_{input} then the width of the pull should be one. We can see from Figure 4.7 that the width of the pull distribution is roughly one for all the signals as expected. We see a larger pull for the atmospheric neutrino flux. This may mean the the uncertainty in the atmospheric neutrino rate is less than the constraint.

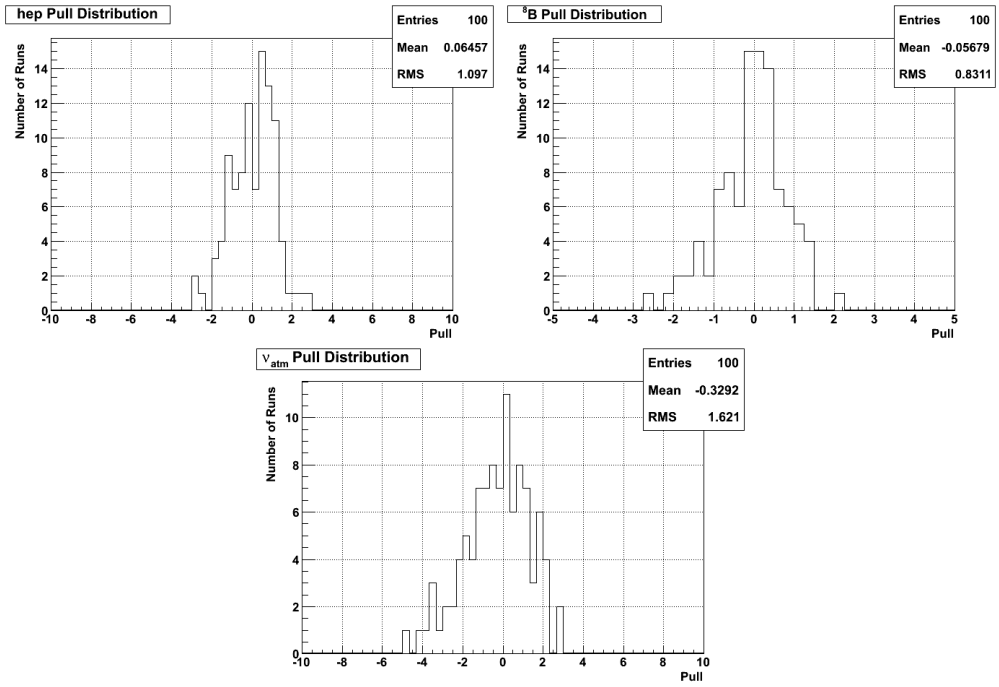

Figure 4.6: Pull distribution for each signal.

4.4 Setting an Upper Flux Limit

In searching for rare processes, the measurement of the parameter can be close to (consistent with) zero, as may be the situation in our analysis. In this case an upper limit is given to the parameter. In the case of the *hep* flux, we use the posterior (likelihood) distribution to make this measurement. In the case of an upper limit measurement, we will use a 90% confidence limit ($1 - \alpha = 0.90$).

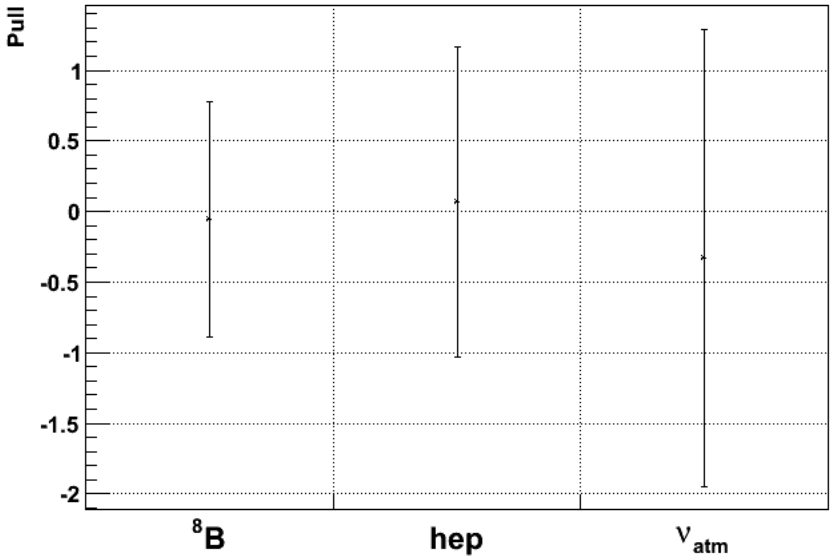

Figure 4.7: Summary of the pull distributions.

This upper limit is set by integrating the likelihood distribution, $P(\theta)$:

$$1 - \alpha = \int_0^{\theta_{upper}} P(\theta) d(\theta) \tag{4.19}$$

where θ is the parameter (*hep* flux in our case). Solving Equation (4.19) for θ_{upper} gives us the limit. Figure 4.8 shows an example of a likelihood distribution that is consistant with zero. The shaded (blue) region represents the allowed 90% confidence level. We would quote the highest value of the number of events in this area as a 90% c.l. upper limit.

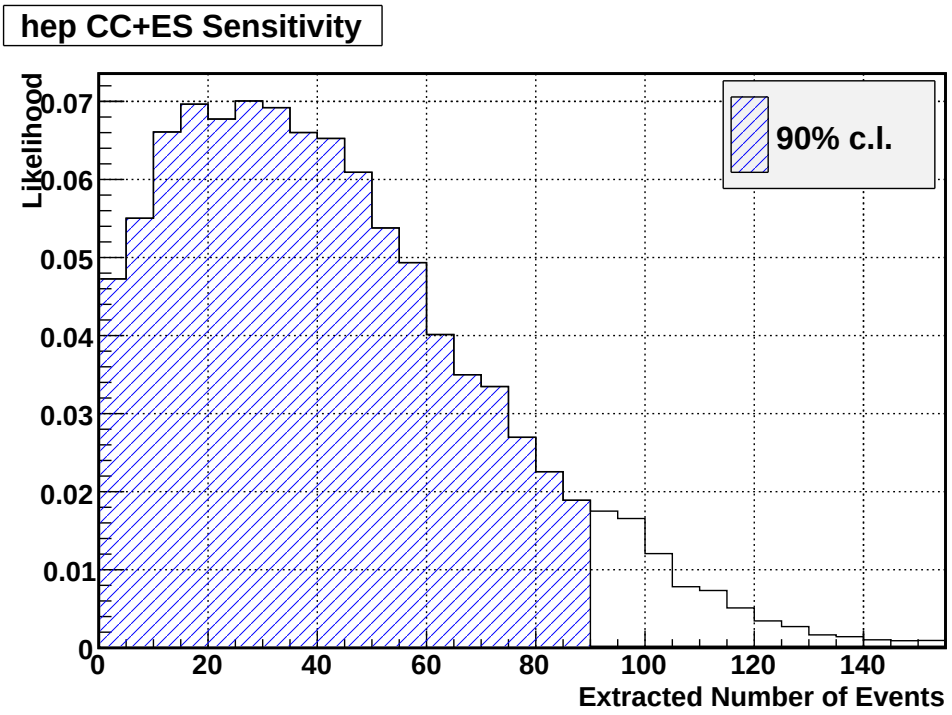

Figure 4.8: Sensitivity to *hep* events using MC as input data for a sample signal extraction.

Chapter 5

hep Search Methods

There are two methods that can be used to look for hep neutrinos. 1) Define a region of interest in the energy spectrum, estimate the number of expected background and signal events, and then count the number of events in this region. This is known as a *box* analysis. 2) Use the information in the shape of the spectrum to statistically distinguish the signals from the backgrounds. This is known as a spectral analysis.

This section will outline the box analysis as previously done by SNO [32]. It will then discuss the motivation to move to a spectral analysis. Finally I will compare the box analysis to the spectral analysis.

5.1 Box Analysis

A box analysis is conducted by defining a region of interest in the energy spectrum (a *box*) and counting the number of events observed in this box (energy range). The box limits are chosen to maximize the sensitivity to the desired signal. When trying to distinguish *hep* neutrinos from ^{8}B neutrinos only the CC and ES signals can be used. If a *hep* neutrino interacts via the NC reaction it would be impossible to distinguish from any other neutrino interacting via the NC reaction. The only observable difference that the *hep* neutrino has compared to other neutrinos is its energy distribution.

For the published SNO *hep* search [32], SNOMAN was used to determine the sensitivity to the *hep* signal for various energy thresholds. Figure 5.1 shows the expected number of signal and background events for a given low energy threshold, T_{thres}, while always using 20 MeV as the upper energy threshold. It also shows the corresponding sensitivity to the signal as a function of the energy threshold. The sensitivity is mainly determined by the signal to background ratio. We want the highest signal-to-noise ratio; that is, the highest number of *hep* events with the lowest number of background events. Based on the previous MC study the optimum energy window for the *hep* box was $14.3 < E < 20$ MeV.

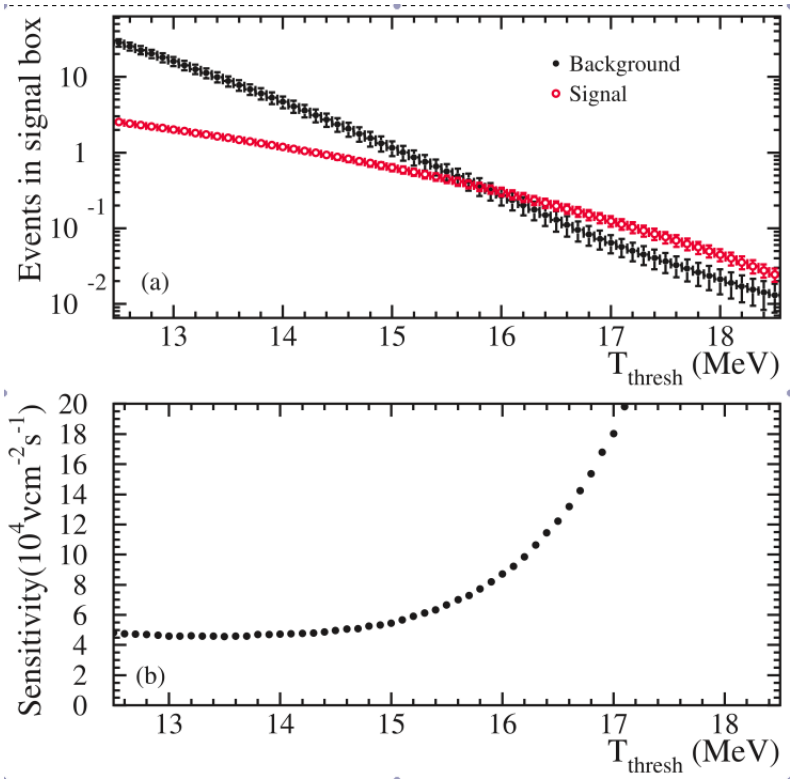

Figure 5.1: The sensitivity of the box analysis as a function of a lower energy threshold [32].

The Monte Carlo was used to simulate 10,000 *experiments* to predict the number of expected background and signal events in the box. Each experiment

was carried out by varying all of the input parameters within their uncertainties. The results from these experiments are shown in Figure 5.2. At the 68% c.l., 3.13 ± 0.60 background events and 0.99 ± 0.09 signal events were expected within the box.

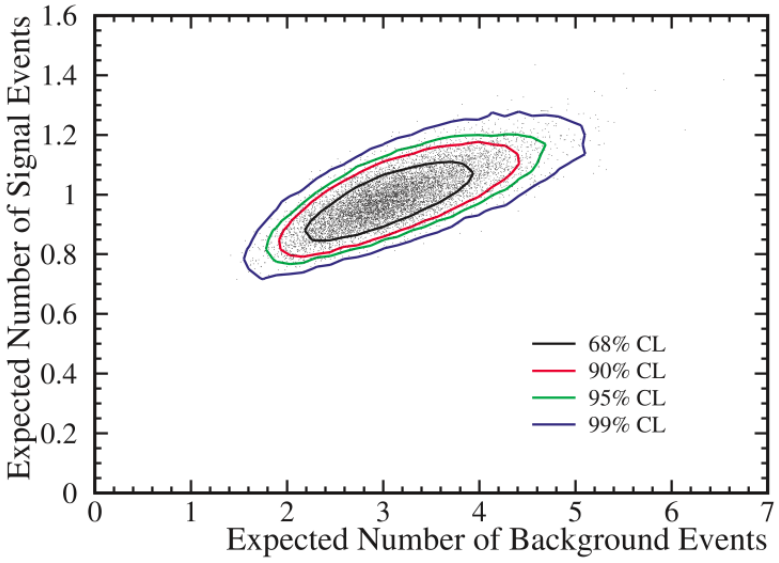

Figure 5.2: Monte Carlo simulations to determine the number of expected background and signal events [32].

5.2 Motivation for a Spectral Analysis

The main advantage of a box analysis is its simplicity. The main effort is optimizing the size of the box, as described in Section 5.1. This method only uses the number of events and neglects the energy distribution in the box where the events occur. This lost information can be harnessed by incorporating the spectral shape into the analysis.

Including the spectral information in the *hep* analysis allows us to distinguish the signal from the background. Figure 5.3 shows the PDFs from the D_2O phase. The main feature of interest for the *hep* analysis in Figure 5.3 is that the *hep* CC+ES PDF reaches out to a higher energy than the ^{8}B CC+ES

PDF. We can also see that the shape of the atmospheric neutrino PDF is significantly different from the ^{8}B and *hep* CC+ES PDFs. This means that by observing the energy shape of the data we can more precisely place the data into the most likely signal.

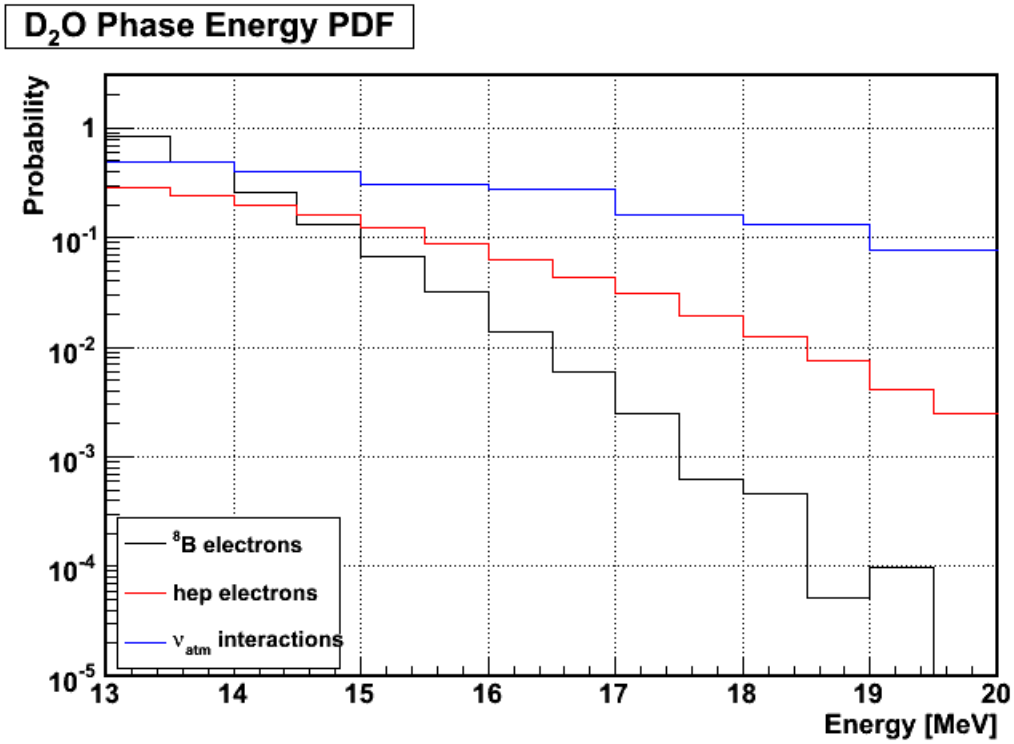

Figure 5.3: PDFs for the first phase of SNO using the SNO MC.

5.3 The Spectral Analysis in this Thesis

The goal of this thesis is to use the data from the three phases of SNO to conduct a search for *hep* neutrinos using a spectral analysis. The spectral signal extraction technique that will be used in this thesis has been outlined in Section 4. This section will highlight the improvements to the analysis over the previous box analysis.

The increase in sensitivity by including all three phases of data over just one can be estimated. The number of signal events, N_s, in one phase of running

with a live time, T, with a rate R_s would be

$$N_s = R_s \times T, \tag{5.1}$$

and number of background events, N_B, in one phase of running with a live time, T, with a rate R_B would be

$$N_B = R_B \times T. \tag{5.2}$$

This would result in a sensitivity, σ_N, in the number of events measured as

$$\sigma_N = \frac{N_s}{\sqrt{N_B}}. \tag{5.3}$$

If we then include all three phases for a live time $3T$ (assuming $T_1 = T_2 = T_3$), we expect the number of signal events to be $3N_s$, and the number of background events to be $3N_B$, assuming a constant rate throughout all the phases. This results in a sensitivity

$$\sigma_{3N} = \frac{3N_s}{\sqrt{3N_B}} = \sqrt{3}\sigma_N. \tag{5.4}$$

The addition of approximately three times the livetime to the analysis is expected to increase the sensitivity (decrease the uncertainty) by a factor of $\sqrt{3}$.

The MCMC signal extraction method has similarities to the box analysis. The box analysis used 10,000 MC experiments to define the probability of signal and background events, as shown in Figure 5.2. In each of these experiments the parameters were smeared within their uncertainties. The MCMC method does this more elegantly. It allows the parameters to be smeared for each step of the Markov chain sampling the likelihood space.

In this analysis the low energy threshold will be set to 13.5 MeV. At lower energies the *hep* energy spectrum is dominated by the ^{8}B energy spectrum, making the signal extraction difficult, and even less precise. Using a low energy

threshold does increase the number of *hep* events in our data set; however, the dominance and similar shapes of the energy spectra lead to a large correlation. This correlation increases the uncertainty (lowers the precision) of the *hep* measurement.

The 13.5 MeV energy threshold cuts out most of the ^{8}B events. Past SNO analyses have focused on measuring the ^{8}B flux. The *hep* analysis is going to use this past measurement as a constraint on the ^{8}B flux. Because so few of the ^{8}B events are in the region of interest for the *hep* analysis we are not concerned that these events could have significantly influenced past measurement, and thus we are not using these events to constrain themselves. That is, the past SNO ^{8}B flux measurement has been done with a significantly different sample than our data set. The constraint on the ^{8}B flux is applied as a Gaussian constraint on the likelihood, as described in Section 4.1.2.

5.4 Comparison of Projected Sensitivity of the Spectral Analysis to the Previous Phase I Box Analysis

This comparison was concluded before the 13.5 MeV low energy threshold was determined. This was done as a proof of concept for the spectral approach to the *hep* search. For this section only, the low energy threshold has been set to 6 MeV (consistent with the low energy threshold of previous analyses). This will not affect our conclusions, as our 13.5 MeV threshold should only improve our sensitivity.

To show the improvement that the spectral analysis can have over a box analysis, both methods were used on common data sets for comparison. We use Monte Carlo events as our simulated data. We set a known number of ^{8}B CC+ES, NC, and atmospheric neutrino type events and try to extract the number of *hep* CC+ES events.

A simple box analysis can be done using the same energy range as the previous SNO *hep* search, $14.3 < E < 20$ MeV. We first find the number of background, B, events (non-*hep* events, this can be done exactly since we are using MC). We then define an uncertainty on the number of background events, δB, by

$$\delta B = \sqrt{B}. \tag{5.5}$$

The number of events, N, in the energy window were counted. The number of signal events (*hep*), S, is then determined by

$$S = N - B. \tag{5.6}$$

The upper limit of the number of *hep* events is

$$S_{upper} = S + 1.282\ \delta B, \tag{5.7}$$

where 1.282σ is a 90% one-sided confidence limit [50].

The spectral (MCMC) analysis solves for the number of events in a different (much larger, $6 < \mathrm{E} < 60$ MeV) energy window, so to compare the box and spectral analysis they are converted to fluxes. The SSM *hep* flux, Φ_{SSM}, is $7.97 \times 10^3\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, so to determine the flux we use

$$\frac{\Phi_{upper}}{\Phi_{SSM}} = \frac{S_{upper}}{S_{SSM}}, \tag{5.8}$$

where

$$S_{SSM} = \begin{cases} = 1 & \text{for box analysis energy range} \\ = 19 & \text{for spectral analysis energy range} \end{cases} \tag{5.9}$$

S_{SSM} is determined from SNOMAN.

We run each analysis over 100 sets of data Poisson distributed at some *input* number of event of each signal type (See Section 4.3 for details on how the data sets are built). Ten different input values were chosen. The results for the box and spectral signal extractions are shown in Figure 5.4. The number of *hep* events are defined as the number of *hep* events in the spectral analysis

window. The results show that the spectral analysis is an improvement over the box analysis by approximately 30%.

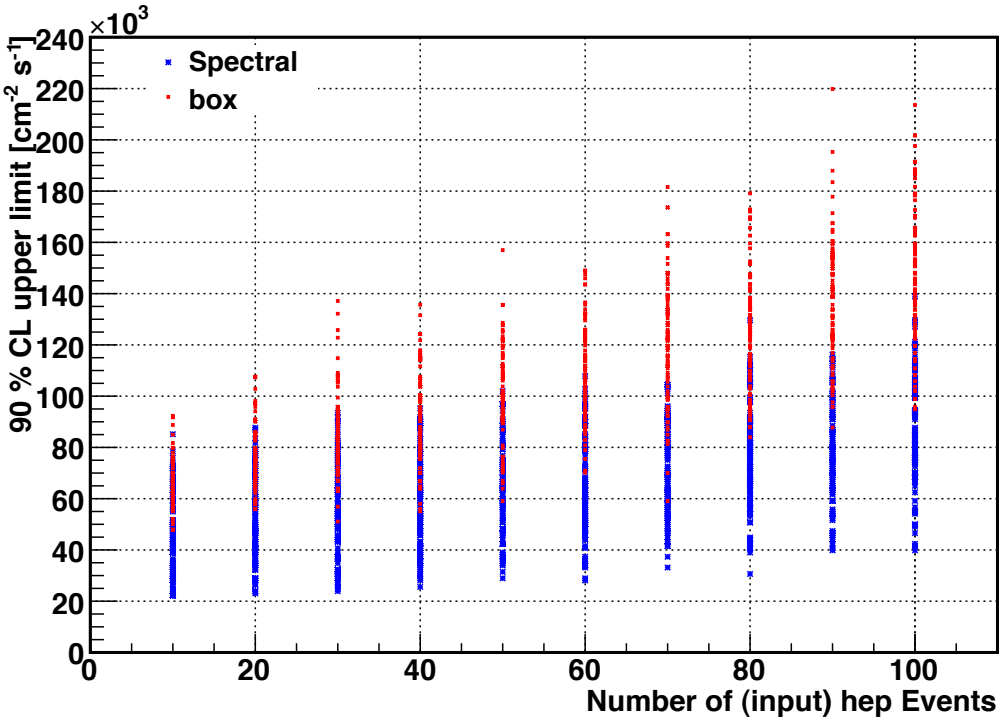

Figure 5.4: Comparison of a box analysis to a spectral analysis using the MC data sets. The inputted number of *hep* events are defined as the number of *hep* events in the spectral analysis window.

5.5 Comparison of Results using Only the Phase I Data Set for the Box and Spectral Analysis

Figure 5.5 shows the results from the first SNO *hep* search. Section 5.1 details the analysis used and states that there were 3.13 ± 0.60 expected background events and 0.99 ± 0.09 signal events expected. The box contained two events, consistent with background events. This allows an upper *hep* flux limit at a 90% confidence level of $2.3 \times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [32].

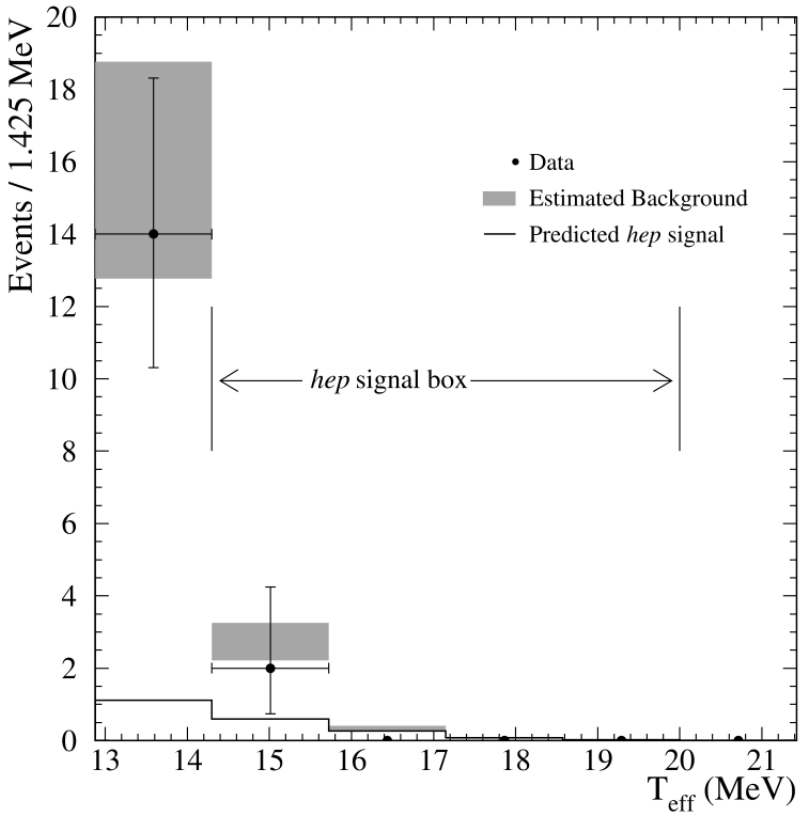

Figure 5.5: First SNO *hep* search results showing the *hep* signal box and observed data [32].

The details of the spectral analysis will be discussed in Chapter 6. This section will just outline the results of the box analysis and the spectral analysis using only the first phase of the SNO experiment. Figure 5.6 shows the 90% confidence level of the *hep* flux extracted using a spectral analysis. The *hep* flux at the 90% confidence level is 1.4×10^4 cm^{-2} s^{-1}. This limit is 1.6 times lower than the previous SNO *hep* search on the same data.

The SSM prediction for the *hep* flux is $(7.97 \pm 1.24) \times 10^3$ cm^{-2} s^{-1}, meaning that this result is still above the SSM, but it is closing in. The remaining Chapters will describe all of the parameters that go into the signal extraction, and conclude with the three phase *hep* search results.

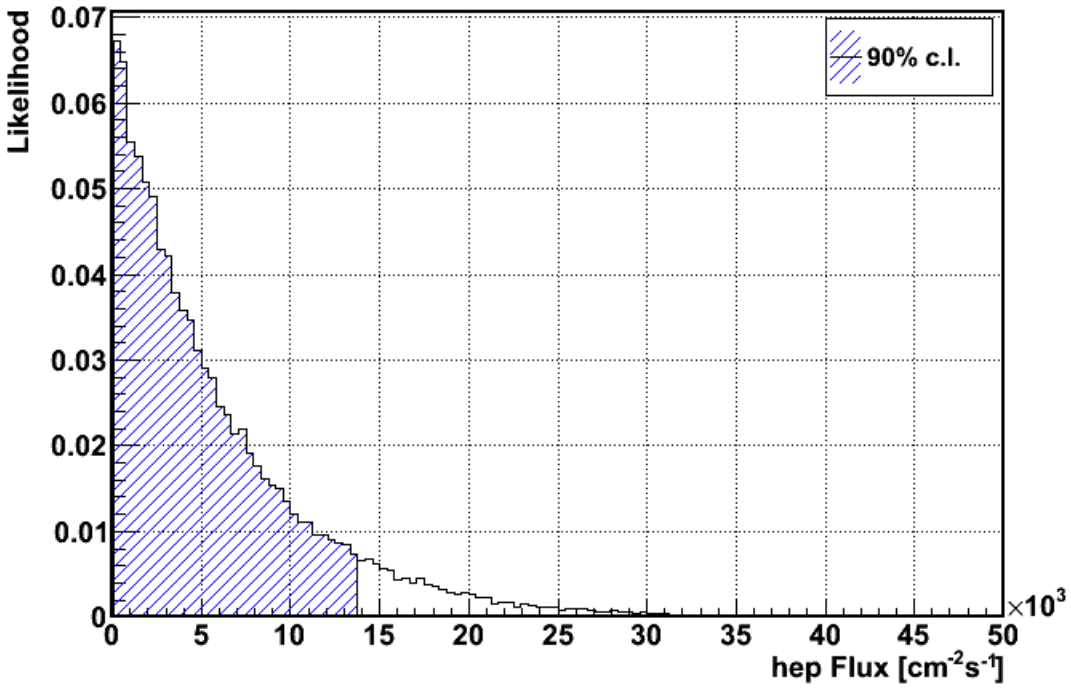

Figure 5.6: The likelihood distribution, integrated to a 90% confidence level, of the *hep* flux using the phase I data set.

Chapter 6

Systematic Uncertainties and Determination of the Low Energy Threshold

6.1 Systematic Uncertainties

Systematic uncertainties are defined in SNO as differences between the SNOMAN and the data as determined by calibrations. The section will first explain the difference in application of the uncertainties between an extended likelihood approach and a χ^2 approach. It will then define the terms that characterize the energy systematic uncertainties and summarize the values used for each uncertainty in the *hep* signal extraction.

Uncertainties in the χ^2 and Extended Likelihood Approaches

We now compare how we would deal with systematic uncertainties in a χ^2 analysis to a likelihood approach. Let us consider a flat data distribution, that is, a flat PDF with n data points, where the number in each bin i is N_i and depends on some parameter p. Lets also assume that there is some efficiency ϵ_i in each bin. The number of events in a given bin is then

$$N_i = \frac{\bar{N}(p)}{\epsilon_i}. \tag{6.1}$$

The uncertainty is determined by

$$\sigma = \sqrt{\sigma_{stat}^2 + \sigma_{sys}^2}, \tag{6.2}$$

where the σ_{stat} is the statistical uncertainty and σ_{sys} is the systematic uncertainty. The systematic uncertainty can be found by solving for $\frac{\bar{N}}{\epsilon_i}$ for a calibration source and then calculating the mean and RMS of $\left(\frac{\bar{N}}{\epsilon_i} - \frac{\bar{N}}{\epsilon_0}\right)$. The χ^2 is then given by:

$$\chi^2 = \sum_i^n \frac{\left(\frac{\bar{N}}{\epsilon_i} - \frac{\bar{N}}{\epsilon_0}\right)^2}{\sigma^2}. \tag{6.3}$$

We can see that we will get a good χ^2 for any shape of the PDF that is consistent with the uncertainty σ. For a good fit, we expect $\chi^2 = 1$.

This is different then the extended likelihood approach, where the systematic uncertainties are applied as constraints (penalty terms). In the case of the likelihood, we find the best fit parameters for the PDF, but there is no defined *good* value.

The likelihood approach is preferred because it can easily handle low statistics analysis. σ_{stat} becomes hard to define, and thus finding a best fit parameter in the χ^2 approach becomes difficult. Also, for large statistics and small σ_{sys} the two approaches converge to one another.

6.1.1 Defining the Systematic Uncertainties

The estimation of the energy scale and energy resolution uncertainty is defined as the difference between looking at the data from the ^{16}N calibration source (see Section 2.1.1) and comparing them to the SNOMAN simulation. This source was a well-defined and tagged gamma particle source, providing a precise and known energy in SNO, and therefore allowed for the most precise measurement of the energy response. This measurement was then verified by using the ^{8}Li calibration source (see Section 2.1.3). Constraints were able to

be applied from the pT source (see Section 2.1.2) in the first phase, as well as further verification by using the Michel elections as a constraint.

To evaluate the systematic uncertainty we compare the MC to data for given calibration sources. These sources are placed in finite locations $\vec{x}$ in the SNO volume at given times t. That is, we measure the discrete energy response of the detector $E_j(\vec{x}, t)$ for n calibration runs j. We then can take the average of our calibration,

$$E_0 = \frac{\sum_j^n E_j(\vec{x}, t)}{n}, \tag{6.4}$$

where E_0 is the average energy for all calibrations. The uncertainty is then measured to be the spread in the energy calibration measurements, σ_{E_0}. In a perfectly calibrated detector we would uniformly sample the volume (because we expect the data to interact uniformly in SNO) as $n \to \infty$.

To apply this exactly we would use a systematic uncertainty parameter for each region in $\vec{x}$ and t. Although possible, this would be very difficult and impractical for large values of n. Instead this $E_j(\vec{x}, t)$ is averaged over, as shown in Equation (6.4), with an uncertainty, σ_{E_0}.

The calibration sources were not always deployed in a spatially uniform way, as the calibration system was restricted to two planes in the volume. There were other variations in the location of deployment; e.g., we calibrated more often in lower regions than higher regions.

To account for all the variations in calibration deployment locations, times, and averaging, the measured uncertainty σ_{E_0} is widened. That is, we added an additional uncertainty δ to the width of the distribution σ_{E_0} to get

$$\sigma_{a,r} = \sigma_{E_0} + \delta. \tag{6.5}$$

This approximation is exactly right if all $E_j(\vec{x}, t)$ terms are the same, that

is if $E_j(\vec{x},t) = E_0$. However, using $E_0 \pm \sigma_{a,r}$ is a large enough distribution to ensure that we are not underestimating our systematic uncertainty. The resulting measurement is then what is used as the systematic uncertainties, the a and r terms in the following discussion, specifically in Equations (6.6) to 6.9.

To apply the systematic uncertainties in the signal extraction we must apply the systematic uncertainties to the PDFs. The energy of a simulated event T is modified to a new energy T_{new} by a series of equations:

$$T_{new} = (1 + a_0 + a_x)T, \tag{6.6}$$

$$T_{new} = \left(1 + a_1 \frac{T - 5.05}{19.0 - 5.05}\right) T, \tag{6.7}$$

$$T_{new} = TG(0, r_y)), \tag{6.8}$$

and

$$T_{new} = T + r_{\mathrm{III}}(T - T_g), \tag{6.9}$$

where a_0 is the linear energy scaling component and correlated between all three phases, a_x is the linear energy scaling component that is uncorrelated between the phases (x =I, II, or III), a_1 is the non-linear (quadratic) energy scaling component, r_y, and r_{III} are the linear energy resolution components (y =I, or II), T_g is the energy as generated by the Monte Carlo, and $G(= c, d)$ is a Gaussian with a centre located at c, and a width of d.

The energy scale is applied in this way, $(a_0 + a_x)$, to account for differences that we believe feed into our energy scale systematic uncertainty. There are many effects that can cause energy scale uncertainty but are characteristics of the detector and therefore do not change over time. These include

the uncertainty in the ^{16}N source, which we use to set the scale. There are other time-dependent causes, such as PMT breakdown and PMT concentrator degradation, that make us consider an additional phase dependent energy scale uncertainty.

Equations (6.6) to 6.9 can be thought of in two parts. First, the $(1 + a_0 + a_x)T$ and $\left(1 + a_1 \frac{T-5.05}{19.0-5.05}\right) T$ terms are a fractional shift in the energy, this just scales the energy up or down. Equation (6.7) is specifically designed to add energy-dependent scaling effects into the signal extraction. This shift is set to zero at the ^{16}N energy (5.05 MeV electron kinetic energy) where the scale was best measured, and to the scaling at the pT energy (19.0 MeV electron kinetic energy). We then linearly interpolate between.

The final terms are a widening or narrowing of the energy distribution, applied slightly different. The $T_{new} = TG(0, r_y)$ term, used in phases I and II, convolves the energy from the energy fitter with an extra Gaussian with a width, r_y and the $T + r_{\mathrm{III}}(T - T_g)$ term, used in phase III, multiplies the difference between the true energy and the energy from the energy fitter by an extra scaling term. These are essentially the same, however the formalisms are different to maintain consistency with past analyses. Combining all these terms ensures that we can accurately account for any differences in the simulation through out the phases.

When estimating the energy scale and non-linearity systematic uncertainties, the estimation was done for each phase, then the most conservative value was taken and that value was applied in each phase. Table 6.1 shows the value used for each energy systematic uncertainty, for each phase. Table 7.2 shows the summary of the effect of each systematic uncertainty.

To apply this to the code, the energy scaling is written in the parameterization:

$$T' = (1 + s_0 + s_1 T_0) T_0. \tag{6.10}$$

	Phase I	Phase II	Phase III
a_0	0 ± 0.0041	0 ± 0.0041	0 ± 0.0041
a_x	$0^{+0.0039}_{-0.0047}$	$0^{+0.0034}_{-0.0032}$	0 ± 0.0081
a_1	0 ± 0.0069	0 ± 0.0069	0 ± 0.0069
r_x	$0.155^{+0.041}_{-0.080}$	$0.168^{+0.041}_{-0.080}$	0.0161 ± 0.0141

Table 6.1: Summary of energy systematics for all phases. x refers to the given phase, I, II, or III. Note here that $r_{\rm I}$ and $r_{\rm II}$ are the values that correspond to Δ in Section 6.1.2.

This formalism was used for reason historical to the first SNO *hep* analysis, but as shown in Section 6.1.2 the (a_0+a_x) and a_1 terms can easily be re-written to be equivalent.

6.1.2 Re-parameterizing Energy Systematic Uncertainties from ^{8}B Three Phase Analysis to *hep* Analysis

Energy Scale Systematic Uncertainty

The *hep* energy scale systematic uncertainty has been implemented into the the signal extraction code in the form of:

$$T' = (1 + s_0 + s_1 T_0)T_0, \tag{6.11}$$

where T' is the re-mapped energy, T_0 is the energy from the SNO energy fitter, s_0 is the linear energy scale, and s_1 is the quadratic energy scale parameter.

This parameterization of the energy scale systematic uncertainty was used for historical reasons. The first SNO *hep* search used this form, and this analysis was designed using that search as a starting point. However, to be consistent with the three phase ^{8}B analysis the application of the energy scale systematics must be the same.

In the three phase analysis the energy scale is done in two steps, first the energy scale term:

$$T' = (1 + a_0^E)T_0, \tag{6.12}$$

where a_0^E is the three phase energy scale parameter ($a_0 + a_x$ in Section 6.1.1). The second step is the energy non-linearity:

$$T'' = \left(1 + a_1 \frac{T' - 5.05}{19.0 - 5.05}\right) T', \tag{6.13}$$

where T'' is the re-mapped energy, and c_0^E is the three phase energy non-linearity parameter.

Putting Equation (6.12) into Equation (6.12) and solving for the resulting re-mapped energy, T'', we get:

$$T'' = \left(1 + a_0^E - \frac{5.05}{19.0 - 5.05} a_1 - \frac{5.05}{19.0 - 5.05} a_0^E a_1 + \frac{(1 + a_0^E)^2}{19.0 - 5.05} a_1 T_0\right) T_0, \tag{6.14}$$

now comparing Equation (6.14) to Equation (6.11) we get the parameterization:

$$s_0 = a_0^E - \frac{5.05}{19.0 - 5.05} a_1 (1 + a_0^E), \tag{6.15}$$

and

$$s_1 = \frac{(1 + a_0^E)^2}{19.0 - 5.05} a_1. \tag{6.16}$$

In the signal extraction we vary the a_0^E and a_1 terms and use the same constraints as the three phase ^{8}B analysis, but they are applied using the s_0 and s_1 parameters internally to the *hep* signal extraction code.

Energy Resolution Systematic Uncertainty

Similar to the energy scale systematic uncertainty situation, the energy resolution systematic uncertainty needs to be re-parameterized from the three phase ^{8}B analysis to the *hep* implementation for consistency. The *hep* signal extraction applies the energy resolution systematic uncertainty in the form:

$$T' = T_0 + (T_0 - T_g) r_x, \tag{6.17}$$

where T' is the re-mapped energy, T_0 is the energy as returned by the SNO energy fitter, T_g is the true energy (MC generated energy), and r_x is the *hep* energy resolution parameter in phase x, where $x =$ I, II, or III.

Equation (6.17) is the same form as that used in the NCD phase of the three phase ^{8}B analysis. The D_2O and salt phases use the same parameterization as the LETA analysis.

The LETA analysis convolve each fitted energy with a Gaussian, $G(1, 0, \Delta)$, normalized to one, a mean of zero, and with a width (one standard deviation) equal to the energy resolution, Δ. That is,

$$T' = T_0 \; G(1, 0, \Delta). \tag{6.18}$$

Figure 6.1 shows this convolution. The black curve represents the energy resolution, with a width σ, and the red curve represents the Gaussian, $G(1, 0, \Delta)$, which is an additional energy resolution, which we will smear the energy by.

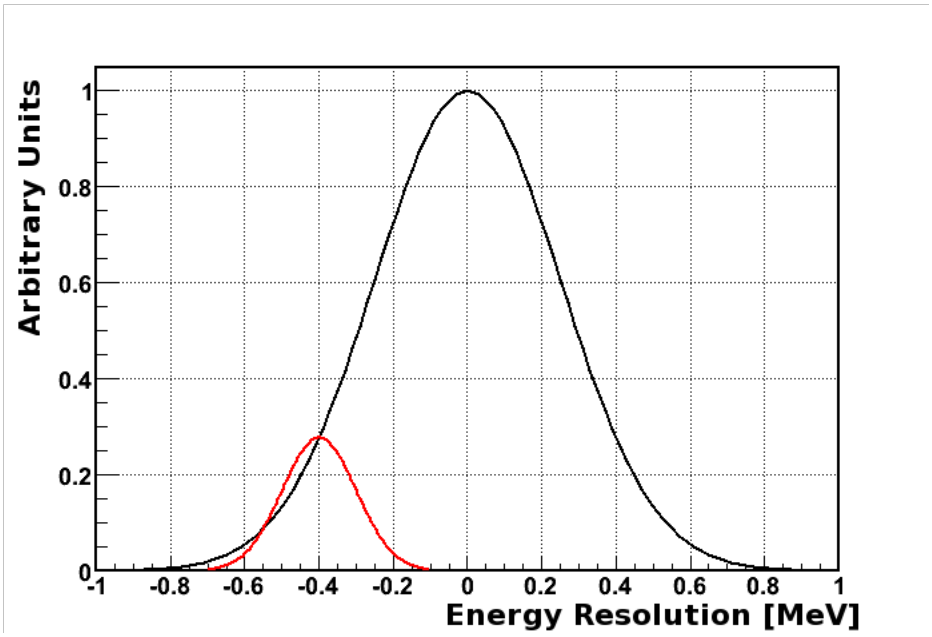

Figure 6.1: The convolution of the black curve by the red curve. The black curve has a width σ and the red curve has a width Δ.

In Equation (6.17) the width, σ' is given by:

$$\sigma' = r_x \sigma, \tag{6.19}$$

where σ is the true resolution between the fitted energy and the true MC generated energy.

In Equation (6.18) the width, σ' is given by:

$$\sigma' = \sqrt{\Delta^2 + \sigma^2}. \tag{6.20}$$

Combining Equations (6.19) and 6.20 we find:

$$r_x = \sqrt{1 + \frac{\Delta^2}{\sigma^2}} - 1, \tag{6.21}$$

where -1 comes from the T_0 term in Equation (6.17), which is the scaling part of the expression. The σ term has been evaluated previously in Reference [37] and is given by:

$$\sigma = -0.185 + 0.413\sqrt{T_0} + 0.0254 T_0. \tag{6.22}$$

The resolution, Δ, will be varied in the code using the same constraints as the three phase ^{8}B analysis. This will then be re-parameterized by Equation (6.21) which will then be used directly in the *hep* signal extraction code.

6.1.3 Floating Systematic Uncertainties

There are two way systematic uncertainties can be considered by a signal extraction; 1) To be *shifted and re-fit* and 2) *floated*. The first is a very simple and conservative approach where several signal extractions are performed, each with a different likelihood function and PDFs defined with different fixed parameters. The sensitivity of the fit to a change in parameters is determined by calculating the derivative of the extracted values from the fit with respect to the variation of the systematic parameters.

The second approach is to modify the (log) likelihood function in the signal extraction (Equation (4.6)) by a penalty term as mentioned in Section 4.1.2,

and to explicitly incorporate the dependence on a systematic parameter into the likelihood function. In this approach the uncertainties are assumed to be Gaussian. In this case we add a Gaussian to the log likelihood function giving:

$$\log L = \sum_{i=0}^{data} \log \left(\sum_{j=0}^{signal} \Phi_j f_j \left(x_i \right) \right) - \sum_{j=0}^{signal} \Phi_j - \left(\frac{\alpha - \mu_\alpha}{2\sigma_\alpha} \right), \tag{6.23}$$

where the first two term are Equation (4.6) and the last term is the Gaussian penalty term, constraining the systematic uncertainty, α, to the mean, μ_α, with a width, σ_α. This is analogous to adding another signal with an analytic Gaussian PDF. As many of these penalty terms can be added as necessary. In this analysis we float the three energy systematic uncertainties (one in each phase), we float a correlated energy scale systematic uncertainty that is correlated through all the phases, we float the non-linearity of the energy response systematic uncertanty, and we float the three energy resolution systematic uncertainties (one in each phase).

It should be noted here that when floating the systematic uncertainties, one needs to be aware of the extracted results. The systematic uncertainty parameter results, in general, should not be more precise (narrower posterior distribution) then the constraint, or penalty term. If this occurs, the set of parameter must be examined. That is, we do not expect the signal extraction to give better results then the calibration data to the systematic uncertainty. There are cases where this may look like the case. These cases may included the situation where two systematic uncertainties are constraining the same physical parameter, and thus may act to constrain each other, resulting in a narrower posterior distribution. With this all being said, we still prefer the floating systematics method, because of its ability to measure the correlations between the various parameters.

6.1.4 Normalization of PDFs After Application of Systematic Uncertainties

The usual way in which SNO has added systematic uncertainties to the signal extraction has been to modify the PDFs (defined in Section 3) by rebuilding (or *smearing*) them for various statistically allowed sets of systematic uncertainty values. This method is relatively easy, however each of the PDFs can have millions of entries making the rebuilding part extremely computationally long. As the rebuilding of the PDFs must be done for every choice in the sampling of the parameters. In the case of the Markov chain, that is a rebuilding for each MCMC step

The other option is to modify the mapping from data to MC space, keeping the PDFs the same shape. This is what we have chosen to use for the *hep* analysis. This section will detail how to maintain normalized PDFs with this approach.

The definition of a PDF requires that it is always the case that:

$$N(S)\int_{x_1}^{x_2} f(x,S)dx = 1, \tag{6.24}$$

where $f(x)$ is a probability distribution, $f(x)dx$ is the probability of the PDF at x in data space, and $N(S)$ is the normalization constant and S is the systematic uncertainty parameters..

We now define the analogous variable in MC as y, and the mapping from data to MC space as:

$$y = g(x,S), \tag{6.25}$$

and

$$dy = g'(x,S)dx, \tag{6.26}$$

putting Equations (6.25) and 6.26 into Equation (6.24) we get:

$$\int_{y_1=g(x_1,S)}^{y_2=g(x_2,S)} N(S)\frac{f(g^{-1}(y,S))}{g'(g^{-1}(y,S))}dy = 1. \tag{6.27}$$

We now define the value that is taken directly from the MC for a given y as:

$$h(y,S) = \frac{f(g^{-1}(y,S))}{g'(g^{-1}(y,S))}, \tag{6.28}$$

and the normalization of $h(y)$ for a given set of systematic uncertainty values, $N(S)$, is given by as:

$$N(S) = \frac{1}{\int_{y_1}^{y^2} h(y,S)dy}. \tag{6.29}$$

This results in the final normalized expression:

$$\int_{x_1}^{x_2} Nh(y,S)g'(x)dx = 1, \tag{6.30}$$

which re-normalizes the MC, allowing the use of the MC as they are, while ensuring the condition of Equation (6.24). Equation (6.30) states that multiplying the value from the MC distribution, $h(y)$, by the normalization for a given set of systematics, N, and the derivative of the mapping, $g'(x)$, we get the probability of an event with the observable, x.

6.1.5 Normalization of PDFs in the Case of Interpolation

To determine the effect of certain systematic uncertainties on the PDFs, it is often complicated to use an explicit analytic form as is outlined in the method of Section6.1.4. This is true in the case of the energy resolution systematic uncertainty.

Instead a table of PDFs was initially calculated for known values of the resolution and used for interpolation durning the signal extraction. This section will show the way in which we interpolated between two PDFs and normalized the result.

Starting with two normalized PDFs:

$$\int_{x_1}^{x_2} f_1(x)dx = 1, \tag{6.31}$$

at some value α_1, and

$$\int_{x_1}^{x_2} f_2(x)dx = 1, \tag{6.32}$$

at α_2. To interpolate to a point α between α_1 and α_2 we use:

$$f_a(x) = \frac{\alpha_2 - \alpha}{\alpha_2 - \alpha_1} f_1(x) + \frac{\alpha - \alpha_1}{\alpha_2 - \alpha_1} f_2(x), \tag{6.33}$$

where $f_\alpha(x)$ is the value of the PDF at α between $f_1(x)$ and $f_2(x)$. To normalize the $f_\alpha(x)$ PDF we require:

$$N_\alpha \int_{x_1}^{x_2} \left(\frac{\alpha_2 - \alpha}{\alpha_2 - \alpha_1} f_1(x) + \frac{\alpha - \alpha_1}{\alpha_2 - \alpha_1} f_2(x) \right) dx = 1, \tag{6.34}$$

where N_α is the normalization constant at α More specifically, as will be explained later in Section 6.1.1 we know that a takes the form:

$$\alpha = r_0, \tag{6.35}$$

where r_0 is a systematic uncertainty. Continuing with the normalization we get:

$$N_\alpha \int_{x_1}^{x_2} \left(\frac{\alpha_2 - r_0}{\alpha_2 - \alpha_1} f_1(x) + \frac{r_0 - \alpha_1}{\alpha_2 - \alpha_1} f_2(x) \right) dx = 1, \tag{6.36}$$

expanding:

$$\frac{1}{N_\alpha} = \int_{x_1}^{x_2} \left(\frac{\alpha_2 - r_0}{\alpha_2 - \alpha_1} f_1(x) + \frac{r_0 - \alpha_1}{\alpha_2 - \alpha_1} f_2(x) \right) dx, \tag{6.37}$$

because $f_1(x)$ and $f_2(x)$ integrate to one, the sum also integrates to one, leaving the normalization constant, N_a being:

$$N_\alpha = 1. \tag{6.38}$$

This means that for any selection of the energy resolution and interpolating between the calculated PDFs we maintain normalization.

6.2 Constraining the Signals

Previous measurements have been done that can aid in extracting the best possible *hep* measurement [6]. Given that the *hep* flux is totally dominated at low energies by the ^{8}B flux, we have chosen a low energy threshold that is high enough so that the residual ^{8}B flux above the threshold is comparable to the expected *hep* flux. With so few events we have limited statistics with which to limit the ^{8}B flux; however, we are able to use the previous SNO solar analyses to constrain the ^{8}B flux.

This previous ^{8}B flux measurement can be used to constrain the number of ^{8}B events in this *hep* analysis, assuming that we understand the energy spectrum of the ^{8}B and *hep* signals. Also, the atmospheric background was studied in much detail for the first SNO *hep* search [32] and discussed in Section 2.4.1. Each of these measured signals can be used to constrain the likelihood function, as is shown in Section 4.1.2.

This section will summarize the measurement of each signal that is used as the constraint. The ^{8}B flux constraint is shown in Table 6.2 and the atmospheric neutrino background is shown in Table 6.3.

^{8}B	mean	uncertainty
All phases	5.05	$^{+0.159}_{-0.152}$

Table 6.2: Constraint on the ^{8}B flux. In units of $\times 10^6\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

ν_{atm}	mean	uncertainty
Phase I	0.122	0.087
Phase II	0.156	0.111
Phase III	0.153	0.109
Rate ($\times 10^{-4}\nu_{atm}$/day)	3.89	2.84

Table 6.3: Number of expected ν_{atm} background events for each phase, and the constraint on the rate of ν_{atm} events in the SNO detector.

6.3 Flux to Event Conversion

The flux of each signal is assumed to be the same in all three phases; however, the number of events in each phase is not. The number of events varies for many reasons such as different live times between phases, different energy resolutions, different thresholds on the electronics, and different amount of absorption in the media. These efficiencies can be seen when we compare the ^{8}B or *hep* spectra from each phase to each other. They are not identical. The number of events in a phase, i, for a signal is given by:

$$N_i = \Phi_{Sun}\ \epsilon_i\ P_{ee}\ T_{\mathrm{live}_i}, \tag{6.39}$$

where Φ_{Sun} is the solar flux, ϵ_i is the capture efficiency (including neutrino-deuteron cross-sections, and detector efficiencies) for phase i, P_{ee} is the probability that an electron neutrino created in the Sun is still an electron neutrino what it reaches the detector, and T_{live_i} is the detector neutrino live time for phase i.

The SNOMAN software has all of the known components of ϵ_i and T_{live_i} included and we have evaluated systematic uncertainties that we can add to the P_{ee} distortion to our PDFs before we do the signal extraction, making this conversion simpler. To calculated the number of events we can use the MC from SNOMAN to do the calculation. This is done by:

$$\frac{N_{\mathrm{data}}}{\Phi_{\mathrm{data}}} = \frac{N_{\mathrm{MC}}}{\Phi_{\mathrm{MC}}}, \tag{6.40}$$

giving:

$$N_{\mathrm{data}} = \Phi_{\mathrm{data}} \frac{N_{\mathrm{MC}}}{\Phi_{\mathrm{MC}}}, \tag{6.41}$$

where N_{data} is the number of expected events in the data, Φ_{data} is the flux that is expected in the data, N_{MC} is the number of events in the Monte Carlo, and Φ_{MC} is the flux used to generate the Monte Carlo. In the signal extraction the

flux, Φ_{data} is the same over all three phases for one step of the Markov chain and the $\frac{N_{\text{MC}}}{\Phi_{\text{MC}}}$ ratio is unique to each phase, resulting in N_{data} being unique to each phase. The Φ_{data} term is then varied for each step of the Markov chain converging on the best fit value.

6.4 Determination of the Lower Energy Threshold

The sensitivity (the upper flux limit at 90% c.l.) of the *hep* analysis a function of the low energy threshold. There are two forces playing off against one another in the determination of the threshold. The first is that, as we lower the threshold we increase the number of signal (*hep*) events. The second is as we raise the threshold we reduce the number of background events (^{8}B).

The placement of the low energy threshold was determined by calculating the sensitivity to the *hep* signal using different low energy thresholds. To do this test, 100 data sets were created by drawing the expected number of signal events from the Monte Carlo for the given low energy threshold. A signal extraction was then done over each of these 100 data sets for a given low energy threshold. The 90% upper confidence limit on the *hep* flux was calculated using the method described in Section 4.4. These upper limits for a given low energy threshold were then recorded.

Low energy thresholds between 12–15 MeV were scanned in 0.5 MeV increments. Figure 6.2 shows the results for each low energy threshold. From Figure 6.2 we can see that the best sensitivity for the *hep* is achieved with a low energy threshold of 13.5 MeV. We want the low energy threshold to have as few ^{8}B neutrinos as possible because they are a background to the *hep* measurement, and to maintain independence between the *hep* measurement and the data set used in the LETA analysis that provides us with our ^{8}B constraint. At 14 MeV we can see that we are harming our sensitivity, so

we place the low energy threshold at 13.5 MeV.

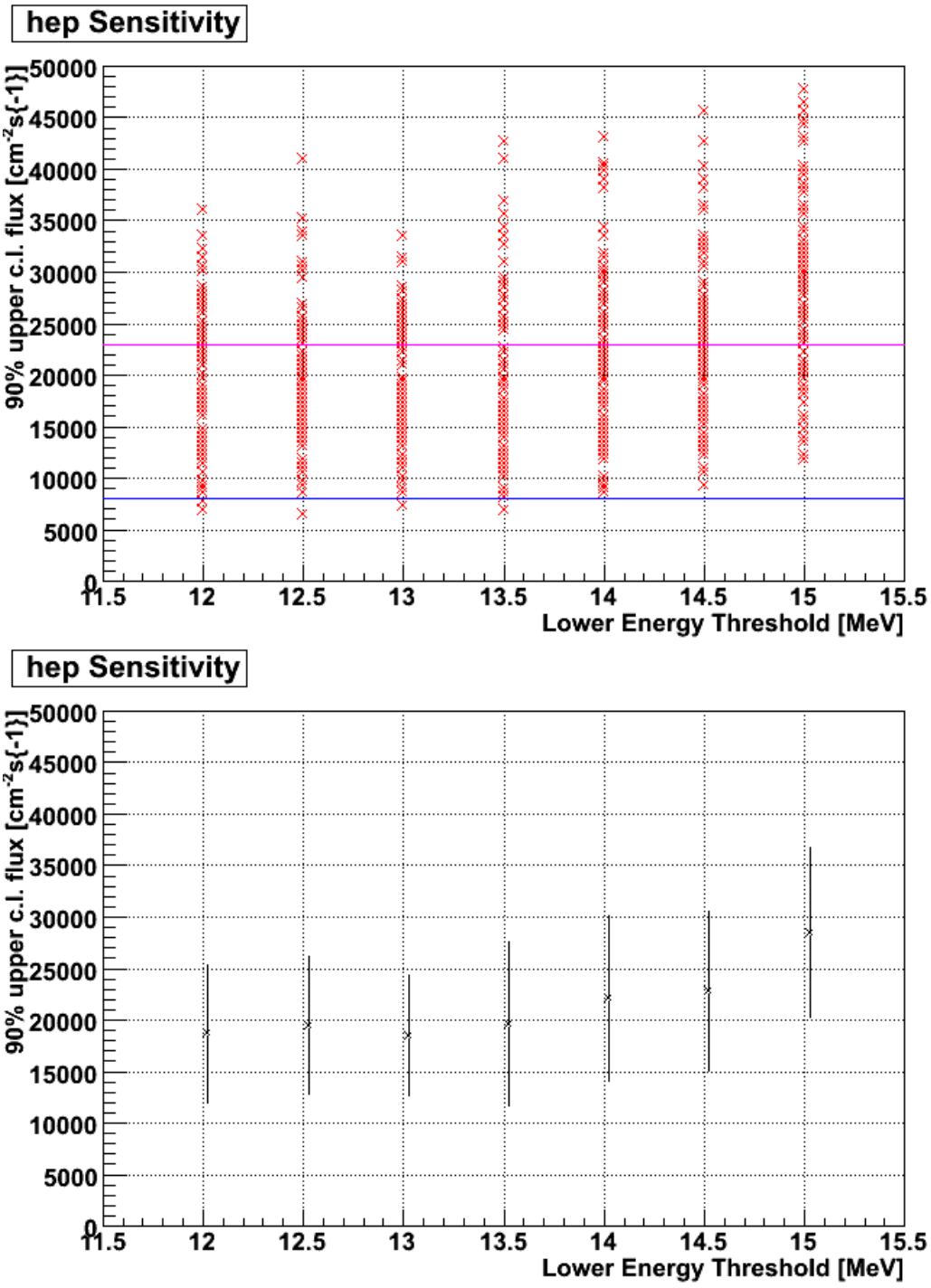

Figure 6.2: The sensitivity to various lower energy thresholds to determine the optimal placement for the three phase *hep* analysis. The upper panel shows the 90% upper c.l. flux from each of the 100 data sets used to probe sensitivity. The bottom panel shows the mean and RMS distribution from the points in the upper panel.

Chapter 7

Flux Results

Up to now the method of signal extraction has been presented, the systematic uncertainties have been described and we have decided on the optimal range in which to analyze to get the most sensitive result. The final step is to run the analysis. Once the signal extraction has been completed, the MCMC must be verified that it has reached convergence. Sections 7.1 and 7.2 will present the tests and confirmation that all signals have sampled the likelihood space, converged to a most likely value, and have no dependence on the initial parameters.

The Data has been presented in Section 2.3, and this section will present the results. This section will include the all correlations between parameters in the signal extraction as well as the effect of each systematic uncertainty on the *hep* result.

At the time of writing, the SNO internal analysis review committee has not approved this analysis to fully examine the data. What will be presented, is a 1/3 random selection of the actual data set. This thesis will present a comparison of a MC simulation of a three phase data set, to the 1/3 data set, and state conclusions using this comparison. All sections will contain results from the 1/3 data set unless otherwise stated.

7.1 Convergence

The MCMC method samples the likelihood space for various choices of the parameters. This method inherently means that some of the initial samples are more correlated with the initial guess of the parameter than with the actual shape of the likelihood space. The steps at the start of the Markov chain that we remove we call burn-in.

The number of steps required in the burn-in period is determined by the number of the parameters, the correlations between the parameters, and the size of the Markov step. It takes many steps of the Markov chain before any meaning came be taking from the results. That is, we have to wait for the Markov chain to converge.

One way to test convergence is to remove the burn-in steps then split the posterior into two halves. If the posterior distribution for the two halves are different, the chain has not converged. Figure 7.1 shows that the chain is converged for the signals, and Figure 7.2 shows that the chain has converged for the systematics uncertainties.

A better test to verify convergence is to check that the likelihood value does not change as a function of MCMC step, this is shown in Figure 7.3. This check is also used for the determination the burn in period, as we can see how long the likelihood value takes to converge. Figure 7.3 shows that after 100 steps the likelihood value no longer changes by large amounts.

Extending this procedure we can look at Figures 7.4 and 7.5 to see how each parameter converges as a function Markov chain step. This also provides another important check. These distributions explicitly show how the MCMC samples each parameter. This allows us to see if the parameter changes its mean value as a function of MCMC step, this would be a sign of non-convergence and is a symptom of a poor choice for MCMC step size for that

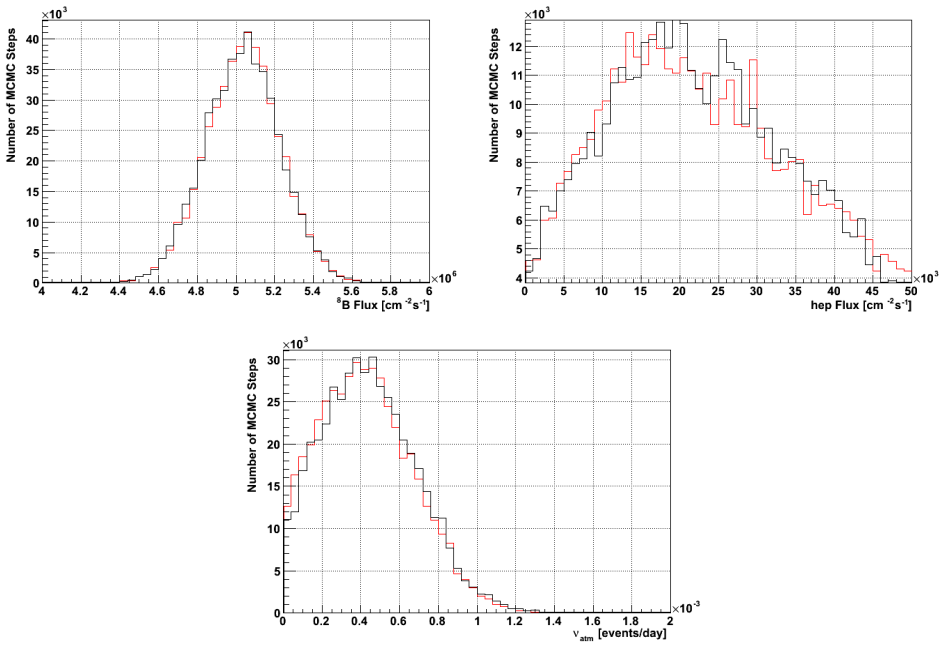

Figure 7.1: A comparison between the first half and the second half of the posterior distribution for the signals.

parameter.

7.2 Auto-Correlation Function

The auto-correlation function (ACF) is a way of measuring the *lag* that the Markov chain has before it *forgets* the value of a step. This is important to measure because the likelihood space can have no dependence on the initial guess giving to the Markov chain.

A side effect of this lag is that if the posterior is fit to a Gaussian the agreement in most cases looks good by eye but will have a poor χ^2 value. This is because of the Markov chain's inherent correlation to the previous value in the chain. It should be noted that this feature is what defines a Markov chain.

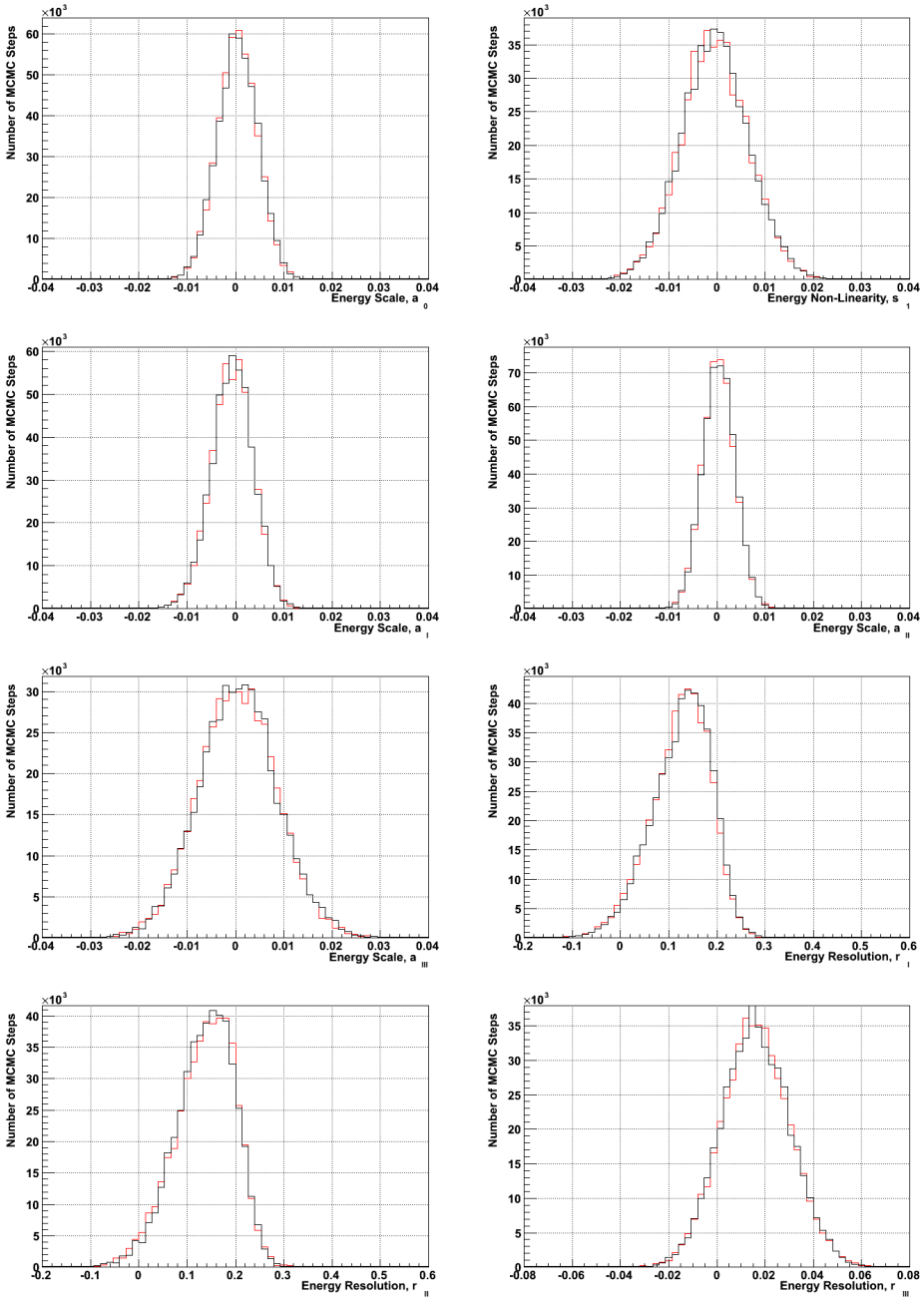

Figure 7.2: A comparison between the first half and the second half of the posterior distribution for the systematic uncertainties.

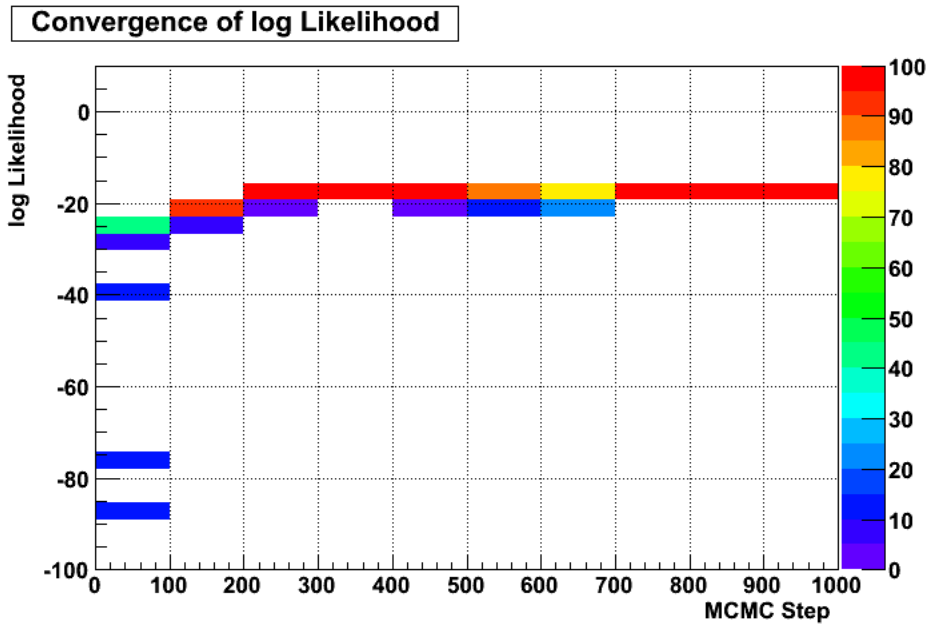

Figure 7.3: The convergence of the log likelihood as a function of the Markov chain step.

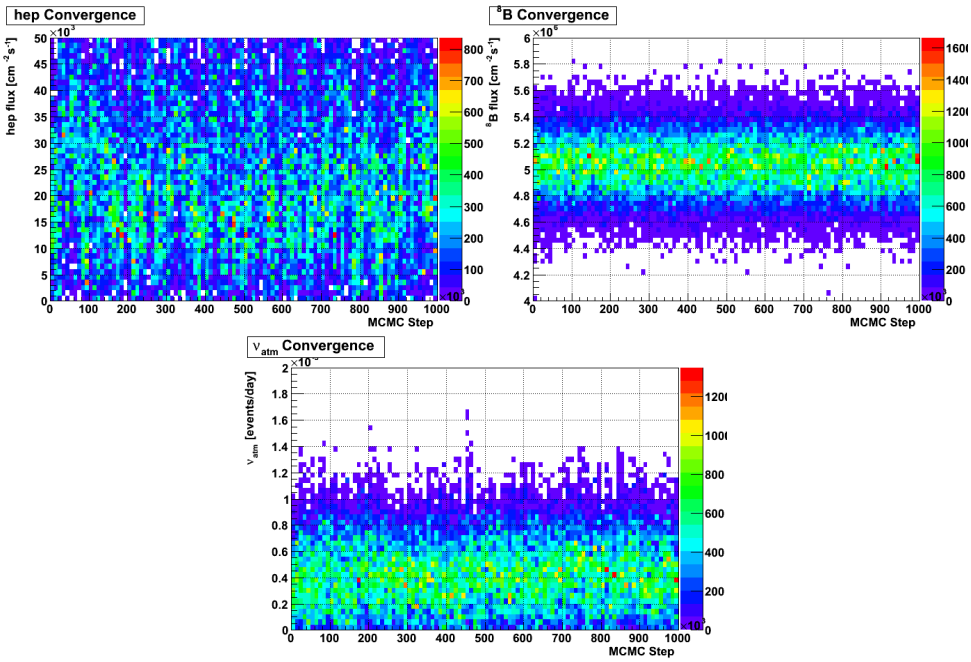

Figure 7.4: The sampling of the signals.

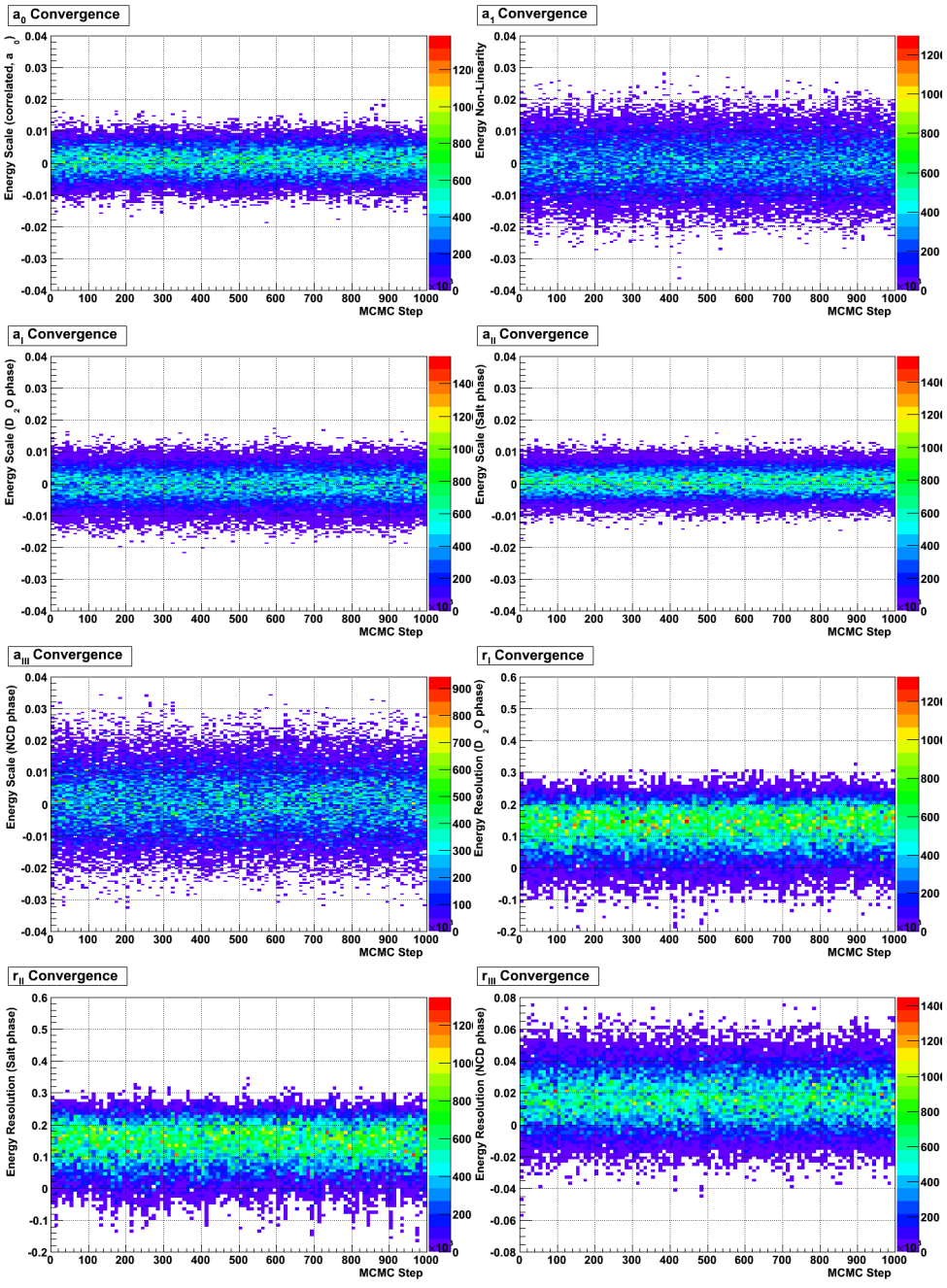

Figure 7.5: The sampling of the energy systematic uncertainties.

The auto-correlation function is defined as:

$$\mathrm{ACF}(h) = \frac{\sum_{i}^{i-h}(\theta_i - \bar{\theta})(\theta_{i+h} - \bar{\theta})}{\sqrt{\sum_{i}^{i-h}(\theta_i - \bar{\theta})^2 \times \sum_{i}^{i-h}(\theta_{i+h} - \bar{\theta})^2}}. \tag{7.1}$$

The ACF for all parameters in the signal extraction is shown in Figure 7.6. We can see that the MCMC step quickly become uncorrelated to the initial parameters. This gives us confidence that we are sampling the likelihood properly. To compare, the burn-in period selected for this analysis is 60,000 MCMC steps.

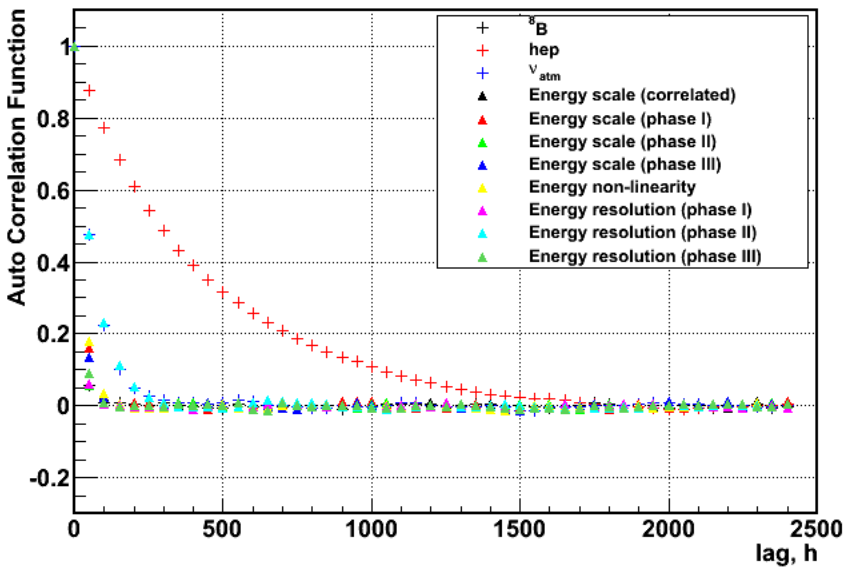

Figure 7.6: The auto-correlation function for the three phase *hep* analysis.

7.3 Correlations Between Parameters

The correlation matrix is given in Table 7.1. The correlation matrix is the representation of the correlation coefficients, ρ_{xy}, between the parameters x

and y. The correlation coefficient is defined as:

$$\rho_{xy} = \frac{V_{xy}}{\sigma_x \sigma_y}, \tag{7.2}$$

where V_{xy} is the covariance between parameters x and y and σ_x is the variance in x. Further details can be found in Reference [50].

Figure 7.7 shows the correlations between the signals. The distributions of the correlations between the all of the systematic uncertainty parameters can be found in Appendix A.

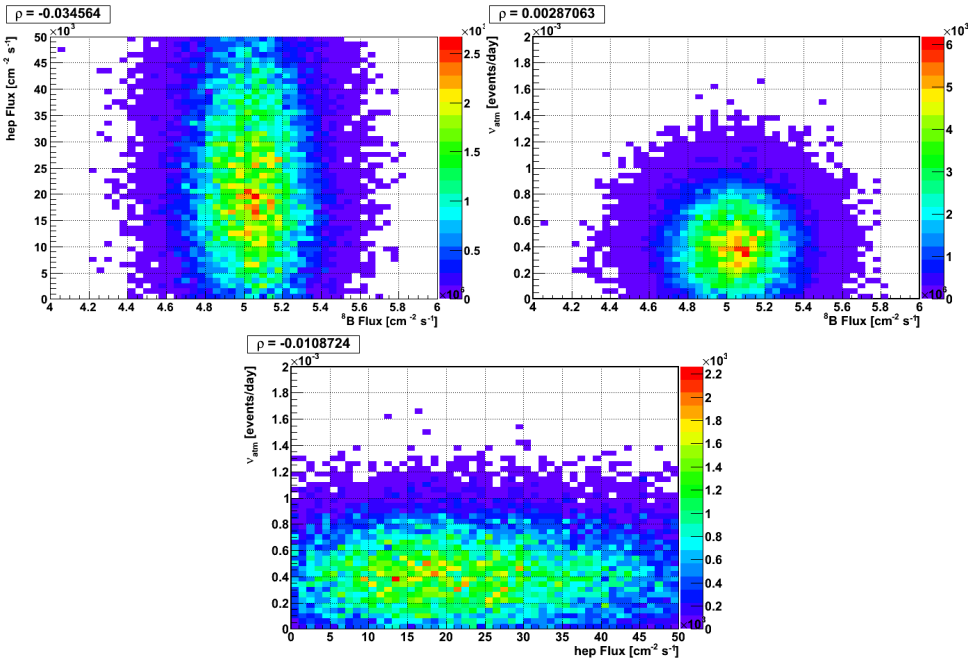

Figure 7.7: Correlation between the signals.

7.4 ^{8}B Result

Figure 7.8 shows the likelihood distribution as determined by MCMC signal extraction for the ^{8}B flux. This distribution shows that the best fit value for the ^{8}B flux is $(5.0 \pm 0.2) \times 10^6\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. This value was constrained by $\left(5.046^{+0.192}_{-0.196}\right) \times 10^6\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which came from the previous LETA analysis [6].

	^{8}B	*hep*	ν_{atm}	a_0	a_I	a_{II}	a_{III}	a_1	r_I	r_{II}	r_{III}
^{8}B	1.00000	-0.03456	0.00287	0.00265	-0.00723	-0.00362	0.00263	-0.01307	0.00818	-0.01244	-0.00372
hep	-0.03456	1.00000	-0.01087	-0.02805	0.00118	-0.00185	-0.04467	0.03761	-0.00807	-0.00474	0.00486
ν_{atm}	0.00287	-0.01087	1.00000	-0.00926	0.01483	-0.00182	-0.01186	-0.00368	0.01297	0.01302	0.01108
a_0	0.00265	-0.02805	-0.00926	1.00000	0.00463	0.01963	-0.01060	-0.01871	0.00331	-0.01060	0.00048
a_I	-0.00723	0.00118	0.01483	0.00463	1.00000	0.00811	0.00594	0.00146	-0.01722	0.00235	0.00653
a_{II}	-0.00362	-0.00185	-0.00182	0.01963	0.00811	1.00000	0.00359	0.00336	-0.00474	0.00315	-0.00069
a_{III}	0.00263	-0.04467	-0.01186	0.01358	0.00594	0.00359	1.00000	-0.00299	0.00691	-0.00009	-0.00229
a_1	-0.01307	0.03761	-0.00368	-0.01871	0.00146	0.00336	-0.00299	1.00000	-0.00800	0.00857	-0.00207
r_I	0.00818	-0.00807	0.01297	0.00331	-0.01722	-0.00474	0.00691	-0.00800	1.00000	-0.01793	-0.02041
r_{II}	-0.01244	-0.00474	0.01302	-0.01060	0.00235	0.00315	-0.00009	0.00857	-0.01793	1.00000	-0.00258
r_{III}	-0.00372	0.00486	0.01108	0.00048	0.00653	-0.00069	-0.00229	-0.00207	-0.02041	-0.00258	1.00000

Table 7.1: Three phase correlation matrix. The correlation between all of the signals and systematic uncertainties included in the signal extraction.

We can see that the extracted result compares well with the previous analysis, as expected.

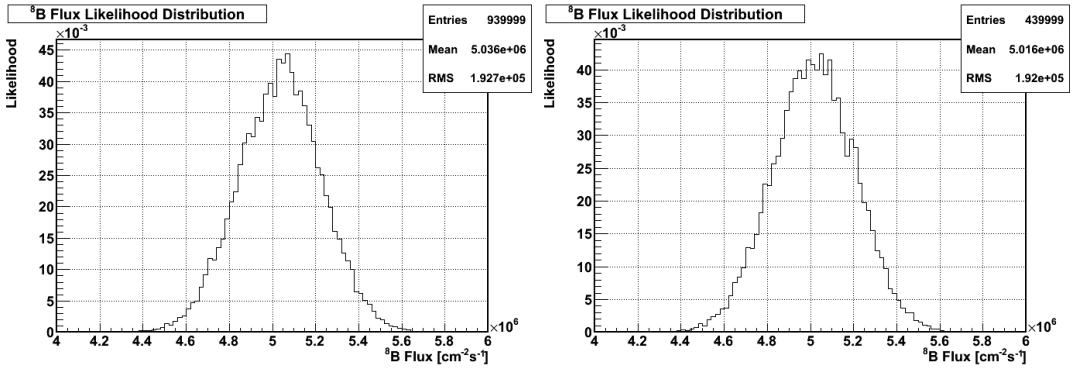

Figure 7.8: The extracted three phase ^{8}B flux. The left panel shows the result using the data set. The right panel is the result from the MC data set.

7.5 Atmospheric Neutrino Result

The atmospheric neutrino flux had been previously estimated by prior analysis, and was constrained by these measurements. The constraint, as listed in Table 6.3, is $(3.98 \pm 2.84) \times 10^{-4}$ ν_{atm}/day. Figure 7.9 shows the extraction of the atmospheric neutrino background, which is in good agreement with the measured constraints.

The rate of atmospheric neutrinos in the SNO detector extracted from the data set was $(4.36 \pm 2.43) \times 10^{-4}$ ν_{atm}/day. As mentioned in Section 2.4.1, this is the rate of atmospheric neutrinos that pass all of the *hep* analysis cuts, and which fall into this energy range, and which do not have any associated secondary events.

Figure 7.9 shows good agreement between the posterior distribution and the constraint.

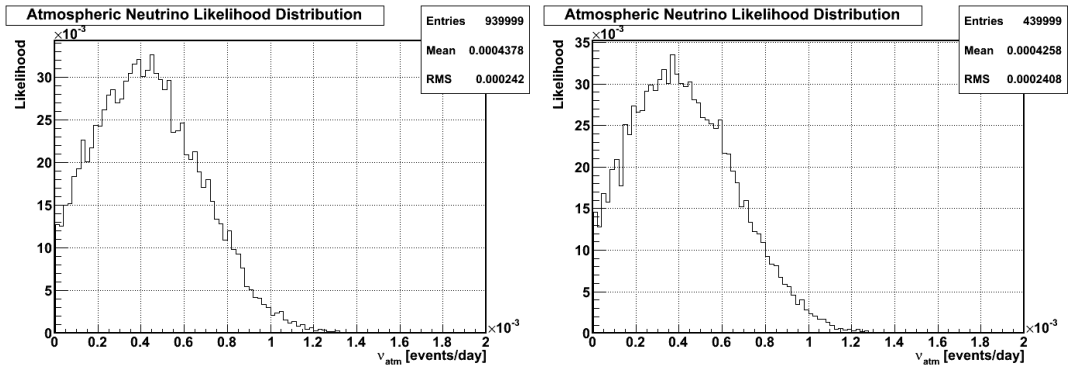

Figure 7.9: The extracted three phase atmospheric neutrino signal. The left panel shows the result using the data set. The right panel is the result from the MC data set.

7.6 *hep* Result

Figure 7.10 is the resulting likelihood distribution for the *hep* flux as determined my the MCMC signal extraction. This extraction was calculated simultaneously with a ^{8}B and atmospheric neutrino background, which were constrained to previous measurements. The resulting *hep* flux measurement is constant with a null measurement.

Using the method outlined in Section 4.4, we obtain a 90% upper confidence limit on the *hep* flux of $2.0 \times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ using the MC data (Figure 7.10) and a we are able to place a limit of $4.2 \times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. This limit is larger the previous best upper limit of $2.3 \times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The new limit can be compared to the SSM, which predicts a flux of $(7.97 \pm 1.24) \times 10^3\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

We notice in Figure 7.10 that the most probable value is not zero. This alludes to a non-zero *hep* flux, as predicted by the SSM. As stated in Reference [50] placing limits can be difficult when near a boundary. The suggested strategy is to present the method chosen so that the result can be used by others.

The simplest approach is to use the mean and RMS of the distribution. Using this approach we get the measured *hep* flux to be $(2.40 \pm 1.25) \times$

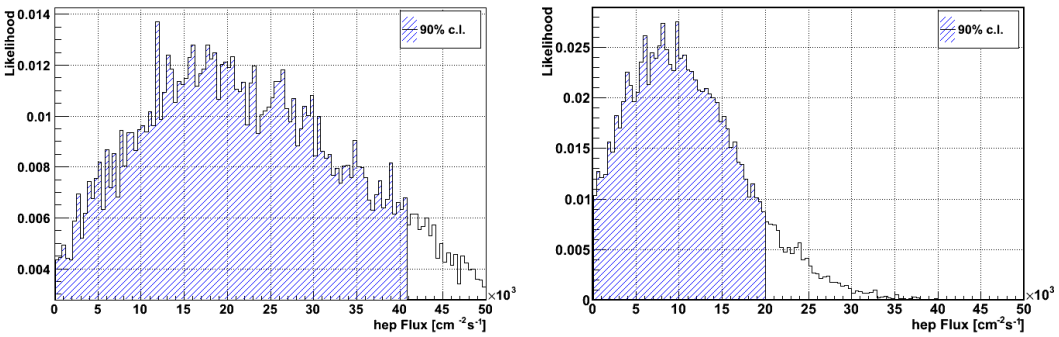

Figure 7.10: The extracted three phase *hep* flux, using a 90% c.l. The left panel shows the result using the data set. The right panel is the result from the MC data set.

$10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

In general, the Bayesian interval can be determine by integrating the probability in the range $[\theta_{lo}, \theta_{up}]$ to a confidence level $1-\alpha$. This is done with:

$$1-\alpha = \int_{\theta_{lo}}^{\theta_{up}} p(\theta|x)d\theta, \tag{7.3}$$

where $p(\theta|x)$ is the posterior distribution.

Here I will show the results using the application of Equation (7.3). To make this consistent with traditional analyses we chose our most probably value to be the peak of the posterior distribution, and chose our 1 σ error bands to include 68.27%, and to so that $p(\theta_{lo}) = p(\theta_{up})$. This is presented in Figure 7.11.

From Figure 7.11 we obtain a range for the MC data set to be $(1.10^{+0.44}_{-1.07}) \times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, and a range $(2.40^{+1.19}_{-1.60}) \times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ for the data set. We can see that these results are consistent, meaning a good agreement between our MC prediction and the data set.

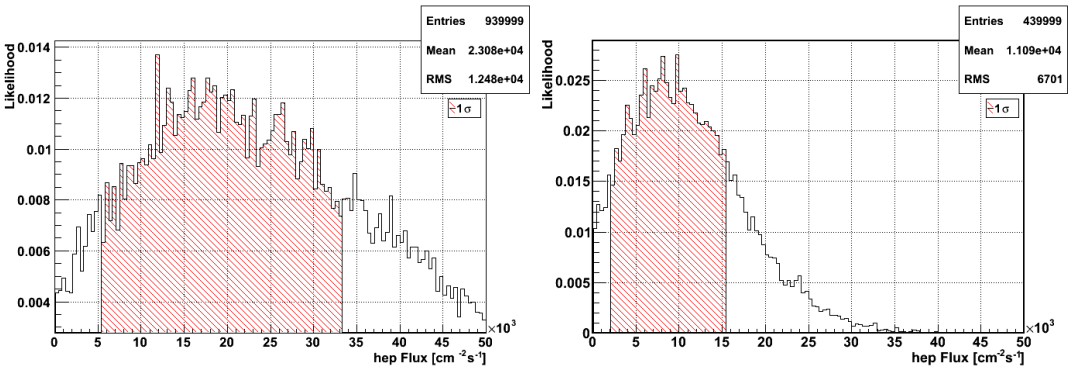

Figure 7.11: The 68.27% or 1σ confidence levels for the *hep* flux. The left panel shows the result using the data set. The right panel is the result from the MC data set.

7.7 Comparison of the Signal Extraction to the Data

To show how well the model in this thesis represents the data, we compare the most probable results from each parameter to the data. That is, we use the most probable value from Figures 7.8 7.9, and 7.10 to scale their PDF. Figure 7.12 shows good agreement between the signal extraction and the data for both the data set and MC data set. The error bars are the combined (in quadrature) statistical error and the error produced by the 1 σ range in flux from the signal extraction (the systematic error).

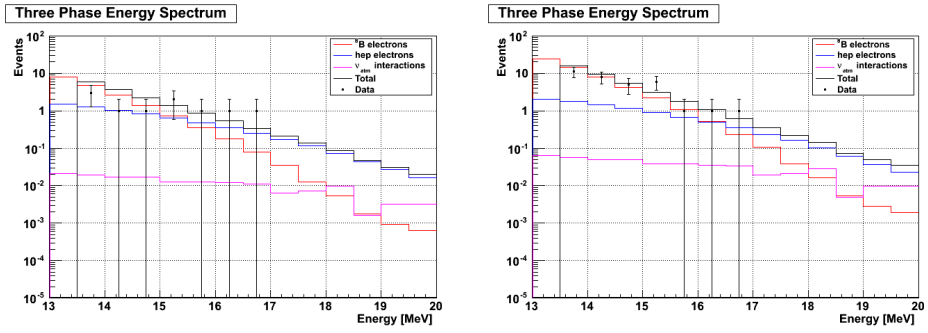

Figure 7.12: The result of the MCMC signal extraction, using the the most probably result for each parameter, compared to the data. The left panel is the data set, the right panel is the MC data set.

7.8 The Floated Systematic Parameters Result

As discussed the systematic parameters associated with the uncertainties were allowed to float in the signal extraction. Figures 7.13 and 7.14 show the final posterior distributions of these systematic uncertainties. These were constrained using the values in Table 6.1, we can see that the values from the fit are in good agreement.

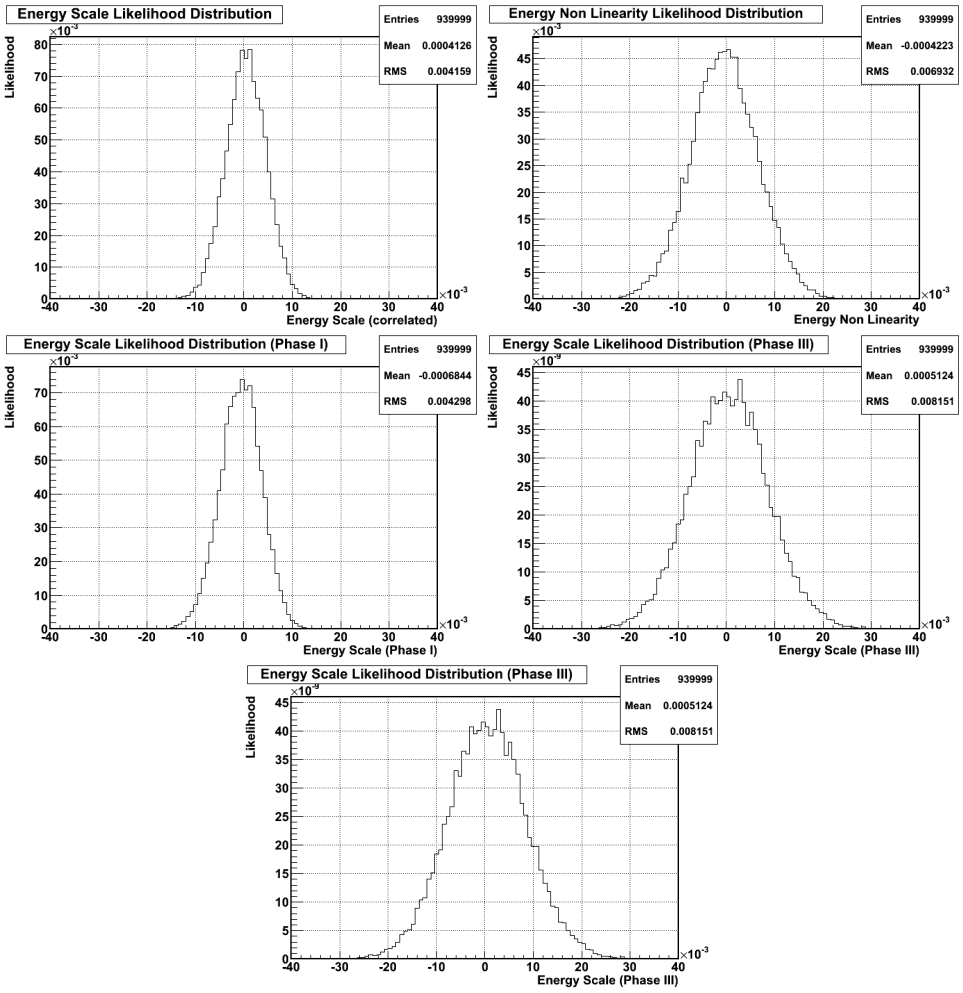

Figure 7.13: The likelihood distributions of the floated energy scale systematic uncertainties.

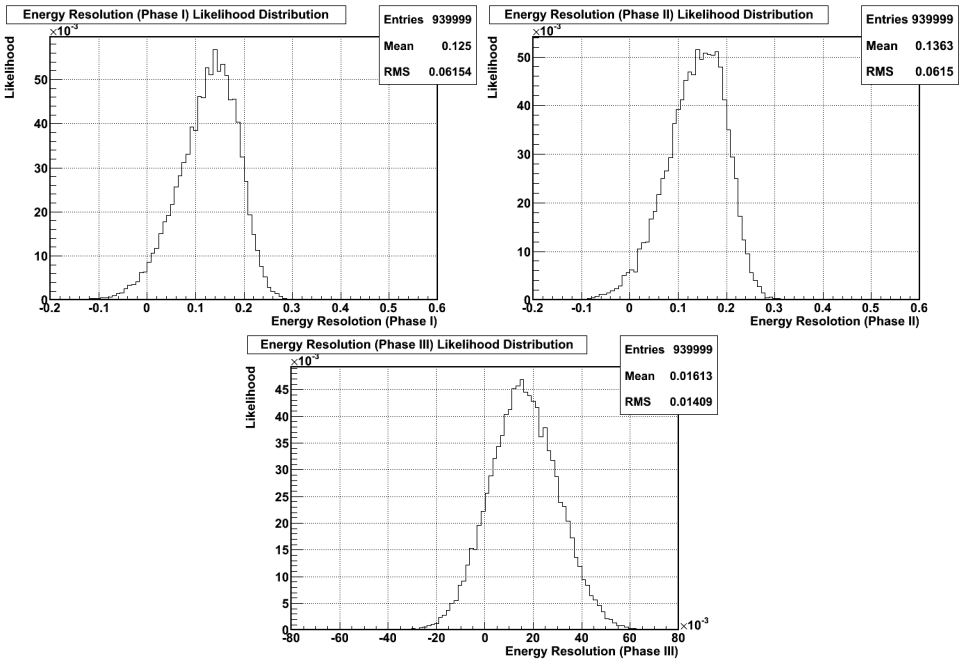

Figure 7.14: The likelihood distributions of the floated energy resolution systematic uncertainties.

7.9 Effect of Each Systematic Uncertainty on the *hep* Flux

To measure the size of each systematic uncertainty on the *hep* flux a test was done where all of the systematic uncertainties were fixed to their nominal values. Each parameter was then individually fixed to $+1\sigma$ and then -1σ, while still holding all others at their nominal value. The fit was re-calculated for each of these situations. In the case of P_{ee}, both neutrino oscillation parameters, $\tan^2\theta_{12}$ and Δm^2_{12}, were shifted at the same time in the direction specified by the $\pm 1\sigma$. The effect of changing each of these systematic uncertainties in isolation is recorded in Table 7.2.

We can see that the size of the effect on the results largest on the phase dependent energy scale systematic uncertainties, and that the energy resolution systematic uncertainties have very little impact on the results. In any case, none of the effects on the results due to the systematics are very large. This

is consistent with the correlations, where we see no strong correlation between the systematics and the *hep* flux.

Systematic	Phase	*hep* Flux Limit (90% c.l.) ($\times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$)	Effect on *hep* Flux Limit (%)	*hep* Flux Most Probable Value ($\times 10^4\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$)	Effect on *hep* Flux Most Probable Value (%)
Nominal	N/A	4.19	N/A	$2.44^{+1.15}_{-1.65}$	N/A
a_0 (+)	I, II, III	4.19	0	$2.40^{+1.15}_{-1.65}$	$1.7^{+1.1}_{-5.3}$
a_0 (−)	I, II, III	4.23	1.9	$2.44^{+1.23}_{-1.56}$	$0^{+2.2}_{-11.4}$
a_{I} (+)	I	4.19	0	$2.44^{+1.15}_{-1.65}$	0±0
a_{I} (−)	I	4.19	0	$2.44^{+1.15}_{-1.65}$	0±0
a_{II} (+)	II	4.23	1.9	$2.44^{+1.23}_{-1.60}$	$0^{+2.2}_{-6.3}$
a_{II} (−)	II	4.23	1.9	$2.44^{+1.19}_{-1.60}$	$0^{+2.2}_{-6.3}$
a_{III} (+)	III	4.15	0.9	$2.35^{+1.10}_{-1.69}$	$3.8^{+4}_{-19.7}$
a_{III} (−)	III	4.23	1.9	$2.48^{+1.23}_{-1.56}$	$1.6^{+3.3}_{-16.5}$
a_1 (+)	I, II, III	4.23	1.9	$2.44^{+1.23}_{-1.60}$	$0^{+2.2}_{-6.3}$
a_1 (−)	I, II, III	4.15	0.9	$2.35^{+1.19}_{-1.60}$	$3.8^{+1.4}_{-5.3}$
r_{I} (+)	I	4.19	0	$2.44^{+1.11}_{-1.65}$	$0^{+1.1}_{-0}$
r_{I} (−)	I	4.19	0	$2.44^{+1.11}_{-1.65}$	$0^{+1.1}_{-0}$
r_{II} (+)	II	4.19	0	$2.44^{+1.15}_{-1.65}$	0±0
r_{II} (−)	II	4.19	0	$2.44^{+1.15}_{-1.65}$	0±0
r_{III} (+)	III	4.19	0	$2.44^{+1.15}_{-1.65}$	0±0
r_{III} (−)	III	4.19	0	$2.44^{+1.15}_{-1.65}$	0±0
P_{ee} (+)	I,II,III	4.19	0	$2.44^{+1.19}_{-1.60}$	$0^{+2.2}_{-6.3}$
P_{ee} (−)	I,II,III	4.19	0	$2.44^{+1.19}_{-1.60}$	$0^{+2.2}_{-6.3}$

Table 7.2: The 1σ effect of each systematic uncertainty on the *hep* flux. The nominal value is the signal extraction with all of the systematic uncertainties fixed to their nominal values. The (+) and (−) labels refer to the application of the $+1\sigma$ and -1σ of the systematic uncertainty, while holding the others fixed to their nominal values.

Chapter 8

Conclusion

Physicists have been building neutrino detectors with energy resolution for 20 years, but have been unable to measure the elusive *hep* neutrino. The *hep* neutrino is hidden by neutrinos produced by the ^{8}B solar reaction for most of its energy range, and thus limits *hep* searches to a very localized energy range. Even now, energy resolution is just good enough to resolve the end point energies between the two signals. Super-Kamiokande, though large, only uses the elastic scattering reaction to measure neutrinos. This reaction gives quite poor energy resolution, especially in comparison to the charged current reaction.

SNOs ability to measure neutrino energies that are produced through the charged current reaction gives it a huge advantage, even at a much smaller active volume. This thesis describes SNOs *hep* search, using all the neutrino available over the three phases. That is over 1,000 days of neutrino data! Only a very few *hep* events occur at the very tail of the energy spectrum, so the longer you look, the more precisely you can measure these events.

This thesis has defined and verified a method to search for the *hep* neutrinos. This method is a Bayesian approach, using a Markov Chain Monte Carlo. This MCMC is applied using a Metropolis-Hastings algorithm, which correctly and efficiently maps out the likelihood space for the signals.

The signals that are considered are the ^{8}B and *hep* neutrino fluxes, and the atmospheric neutrino background. An energy range was chosen so that the ^{8}B and *hep* energy distributions would look different, and break the correlation that plagues their separation at lower energies. This was successful as shown in Table 7.1. The effect of the ^{8}B flux was also reduced by using prior SNO analyses to add a constraint to ^{8}B flux, this was shown in Section 6.2.

Also, events were chosen to minimize any contaminants from all background events, from electronic to natural sources. This cleaning was discussed in Section 2.3. This cleaning led to an estimation of the atmospheric background of an order of magnitude lower than the expected *hep* flux as shown in Sections 2.4.1 and 6.2.

All of this work to optimize the accuracy and precision of the signal extraction, combined with limiting the various backgrounds, has led to the best possible result of the *hep* flux as determined by the SNO experiment.

An upper 90% c.l. of $4.2 \times 10^4 \mathrm{cm}^{-2}\mathrm{s}^{-1}$ is placed on the *hep* flux using in this analysis, using 1/3 of the data. This seems worse when compared to the previous Super-Kamiokande result of $7.3 \times 10^4 \mathrm{cm}^{-2}\mathrm{s}^{-1}$, and to the previous SNO measurement of $2.3 \times 10^4 \mathrm{cm}^{-2}\mathrm{s}^{-1}$. However, this is the first time that a measurement, albeit less than 2σ from null, is observed. This measured value for the *hep* flux is $(2.40^{+1.19}_{-1.60}) \times 10^4 \mathrm{cm}^{-2}\mathrm{s}^{-1}$.

This is not a discovery however, this is still very exciting. We are now at a sensitivity to show some evidence that there is in fact a solar *hep* neutrino flux. At this time, we are only using one-third of the available data, as the analysis is under review by an internal SNO analysis committee. When the full data set is observed, this value may improve. When comparing our result to the Monte Carlo prediction for the full data set we expect a 90% upper c.l. of $2.0 \times 10^4 \mathrm{cm}^{-2}\mathrm{s}^{-1}$ and a *hep* flux value of $(1.10^{+0.44}_{-1.07}) \times 10^4 \mathrm{cm}^{-2}\mathrm{s}^{-1}$, which is consistent with the one-third data *hep* flux measurement, and close to the

SSM prediction of $(7.970 \pm 1.236) \times 10^3 \mathrm{cm}^{-2}\mathrm{s}^{-1}$.

Bibliography

[1] J. N. Bahcall, M. H. Pinsonneault, and S. Basu, *Solar models: Current epoch and time dependences, neutrinos, and helioseismological properties*, *The Astrophysical Journal* **555** (2001), no. 2 990.

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

[3] L. Miramonti, *Solar neutrino detection*, *AIP Conf. Proc.* **1123** (2009) 166–173, [0901.3443].

[4] **SNO** Collaboration, Q. R. Ahmad *et al.*, *Measurement of the charged current interactions produced by B-8 solar neutrinos at the Sudbury Neutrino Observatory*, *Phys. Rev. Lett.* **87** (2001) 071301, [nucl-ex/0106015].

[5] K. S. Hirata *et al.*, *Observation of ^8B solar neutrinos in the kamiokande-II detector*, *Phys. Rev. Lett.* **63** (Jul, 1989) 16–19.

[6] **SNO** Collaboration, B. Aharmim *et al.*, *Low Energy Threshold Analysis of the Phase I and Phase II Data Sets of the Sudbury Neutrino Observatory*, *Phys. Rev. C.* (2010).

[7] **SNO** Collaboration, B. Aharmim *et al.*, *An Independent Measurement of the Total Active 8B Solar Neutrino Flux Using an Array of 3He Proportional Counters at the Sudbury Neutrino Observatory*, *Phys. Rev. Lett.* **101** (2008) 111301, [0806.0989].

[8] J. P. Cravens, K. Abe, T. Iida, K. Ishihara, J. Kameda, Y. Koshio, A. Minamino, C. Mitsuda, M. Miura, S. Moriyama, M. Nakahata, S. Nakayama, Y. Obayashi, H. Ogawa, H. Sekiya, M. Shiozawa, Y. Suzuki, A. Takeda, Y. Takeuchi, K. Ueshima, H. Watanabe, S. Yamada, I. Higuchi, C. Ishihara, M. Ishitsuka, T. Kajita, and K. Kaneyuki, *Solar neutrino measurements in super-kamiokande-ii*, *Phys. Rev. D* **78** (Aug, 2008) 032002.

[9] R. Davis, *Solar neutrinos: II. experimental*, *Phys. Rev. Lett.* **12** (March, 1964).

[10] V. Gribov and B. Pontecorvo, *Neutrino astronomy and lepton charge*, *Physics Letters B* **28** (1969), no. 7 493 – 496.

[11] Z. Maki, M. Nakagawa, Y. Ohnuki, and S. Sakata, *A Unified Model for Elementary Particles*, *Progress of Theoretical Physics* **23** (June, 1960) 1174–1180.

[12] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the unified model of elementary particles*, *Progress of Theoretical Physics* **28** (1962), no. 5 870–880.

[13] J. N. Bahcall, *Neutrino Astrophysics*. Cambridge University Press, 1989.

[14] L. Wolfenstein, *Neutrino oscillations in matter*, *Phys. Rev. D* **17** (May, 1978) 2369–2374.

[15] S. P. Mikheev and A. Y. Smirnov, *Resonance enhancement of oscillations in matter and solar neutrino spectroscopy*, *Sov. J. Nucl. Phys.* **42** (1985) 913–917.

[16] J. N. Bahcall, A. M. Serenelli, and S. Basu, *10,000 Standard Solar Models: a Monte Carlo Simulation*, *Astrophys. J. Suppl.* **165** (2006) 400–431, [astro-ph/0511337].

[17] Y. Koshio, *Recent Super-Kamiokande Results on Solar Neutrinos*, in *Les Rencontres de Physique de la Vallee d'Aoste La Thuile*, (Aosta Valley), 2000.

[18] D. Kielczewska, *Experimental Results on Neutrino Oscillations using Atmospheric, Solar and Accelerator Beams*, in *Acta Physica Polonica, vol. B31, 1181*, (Cracow, Poland), 2000.

[19] Y. Koshio, *Solar neutrinos results and oscillation analysis from Super-Kamiokande*, in *26th Inter. Cosmic Ray Conf*, 1999.

[20] Y. Suzuki, *Solar Neutrinos*, in *Lepton-Photon Symposium 99*, 1999.

[21] L. Marcucci, R. Schiavilla, M. Viviani, A. Kievsky, and S. Rosati, *Realistic calculation of the He-3+p (hep) astrophysical factor*, *Phys Rev Lett.* **84** (JUN 26, 2000) 5959–5962.

[22] L. Marcucci, R. Schiavilla, M. Viviani, A. Kievsky, S. Rosati, and J. Beacom, *Weak proton capture on He-3*, *Phys. Rev. C* **63** (JAN, 2001).

[23] T.-S. Park, K. Kubodera, D.-P. Min, and M. Rho, *The solar hep processes in effective field theory*, *PiN Newslett.* **16** (2002) 334–336, [nucl-th/0110084].

[24] T. S. Park *et al.*, *Parameter-free effective field theory calculation for the solar proton fusion and hep processes*, *Phys. Rev.* **C67** (2003) 055206, [nucl-th/0208055].

[25] W. Alberico, S. Bilenky, and W. Grimus, *On a possibility to determine the S-factor of the hep process in experiments with thermal (cold) neutrons*, *Astroparticel Physics* **15** (APR, 2001) 211–215.

[26] G. Fiorentini, V. Berezinsky, S. Degl'Innocenti, and B. Ricci, *Bounds on hep neutrinos*, *Phys. Lett. B* **444** (DEC 24, 1998) 387–390.

[27] C. Horowitz, *High energy solar neutrinos and p-wave contributions to He-3(p,ve(+))He-4*, *Phys. Rev. C* **60** (AUG, 1999).

[28] M. Coraddu, M. Lissia, G. Mezzorani, and P. Quarati, *Super-Kamiokande hep neutrino best fit: a possible signal of non-Maxwellian solar plasma*, *Physica A-Statistical Mechanics and its Applications* **326** (AUG 15, 2003) 473–481.

[29] K. Kubodera and T. Park, *The solar hep process*, *Annual Review of Nuclear and Particle Scince* **54** (2004) 19–37.

[30] S. Fukuda, Y. Fukuda, M. Ishitsuka, Y. Itow, T. Kajita, J. Kameda, K. Kaneyuki, K. Kobayashi, Y. Koshio, M. Miura, S. Moriyama, M. Nakahata, S. Nakayama, A. Okada, N. Sakurai, M. Shiozawa, Y. Suzuki, H. Takeuchi, Y. Takeuchi, T. Toshito, Y. Totsuka, S. Yamada, S. Desai, M. Earl, E. Kearns, M. D. Messier, and K. Scholberg, *Solar b8 and hep neutrino measurements from 1258 days of super-kamiokande data*, *Phys. Rev. Lett.* **86** (Jun, 2001) 5651–5655.

[31] **Super-Kamkiokande** Collaboration, J. Hosaka *et al.*, *Solar neutrino measurements in Super-Kamiokande-I*, *Phys. Rev.* **D73** (2006) 112001, [hep-ex/0508053].

[32] **SNO** Collaboration, B. Aharmim *et al.*, *A search for neutrinos from the solar hep reaction and the diffuse supernova neutrino background with the sudbury neutrino observatory*, *Astrophys. J.* **653** (2006) 1545–1551, [hep-ex/0607010].

[33] J. Boger *et al.*, *The sudbury neutrino observatory*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **449** (2000), no. 1-2 172 – 207.

[34] **GALLEX** Collaboration, P. Anselmann *et al.*, *Solar neutrinos observed by GALLEX at Gran Sasso.*, *Phys. Lett.* **B285** (1992) 376–389.

[35] A. I. Abazov *et al.*, *Search for neutrinos from the sun using the reaction* $^{71}Ga(\nu_e,e^-)^{71}Ge$, *Phys. Rev. Lett.* **67** (1991), no. 24 3332–3335.

[36] B. A. Moffat, R. J. Ford, F. A. Duncan, K. Graham, A. L. Hallin, C. A. W. Hearns, J. Maneira, P. Skensved, and D. R. Grant, *Optical calibration hardware for the sudbury neutrino observatory*, *Nucl. Instrum. Methods Phys. Res., A* **554** (Jul, 2005) 255–265. 17 p.

[37] R. F. MacLellan, *The energy calibration for the solar neutrino analysis of all three phases of the Sudbury Neutrino Observatory.* PhD thesis, Queen's University, 2009.

[38] M. R. Dragowsky, A. Hamer, Y. D. Chan, R. Deal, E. D. Earle, W. Frati, E. Gaudette, A. Hallin, C. Hearns, J. Hewett, G. Jonkmans, Y. Kajiyama, A. B. McDonald, B. A. Moffat, E. B. Norman, B. Sur, and N. Tagg, *The 16n calibration source for the sudbury neutrino observatory*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **481** (2002), no. 1-3 284 – 296.

[39] **The LETA working group** Collaboration, "The low energy threshold analysis (leta) unidoc v. 0.1." internal document, February, 2009.

[40] A. W. P. Poon, R. J. Komar, C. E. Waltham, M. C. Browne, R. G. H. Robertson, N. P. Kherani, and H. B. Mak, *A compact 3h(p,gamma)4he 19.8-mev gamma-ray source for energy calibration at the sudbury neutrino observatory*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **452** (21 September 2000) 115–129(15).

[41] **The hep analysis group** Collaboration, "The hep analysis: Unified document version 1.0." internal document, April, 2006.

[42] N. J. Tagg, A. S. Hamer, B. Sur, E. D. Earle, R. L. Helmer, G. Jonkmans, B. A. Moffat, and J. J. Simpson, *The 8li calibration source for the sudbury neutrino observatory*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **489** (2002), no. 1-3 178 – 188.

[43] F. Zhang, *A study on the energy nonlineraty of the NCD phase*, 2008. Internal Document.

[44] D. Casper, *The nuance neutrino simulation, and the future*, *Nuclear Physics B - Proceedings Supplements* **112** (2002) 161.

[45] G. D. Barr, T. K. Gaisser, P. Lipari, S. Robbins, and T. Stanev, *A three-dimensional calculation of atmospheric neutrinos*, *Physical Review D* **70** (2004) 023006.

[46] G. Battistoni, A. Ferrari, T. Montaruli, and P. R. Sala, *The atmospheric neutrino fluxes below 100 mev: the fluka results*, *Astropart. Phys.* **23** (2005) 526–534.

[47] "MCNP4A, a monte carlo n-particle transport code system." Radiation Shielding Information Center, Los Alamos National Laboratory, 1993.

[48] *in the Proceedings of the International Conference on Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications (MC 2000) Conference*, 2000.

[49] H. H. W. R. Nelson and D. W. O. Rogers, "The EGS4 code system." SLAC report 265, 1985.

[50] G. Cowan, *Statistcal Data Analysis*. Oxford Science Publications, 1998.

[51] W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, *Biometrika* (1970).

[52] "Minuit - function minimization and error analysis." CERN Program Library entry D506, 1998.

Appendix A

The Correlation between Parameters in the Signal Extraction

This appendix shows the remaining correlation distributions, as described in Section 7.3. These figures show the correlation, ρ, between each of the parameters in the signal extraction.

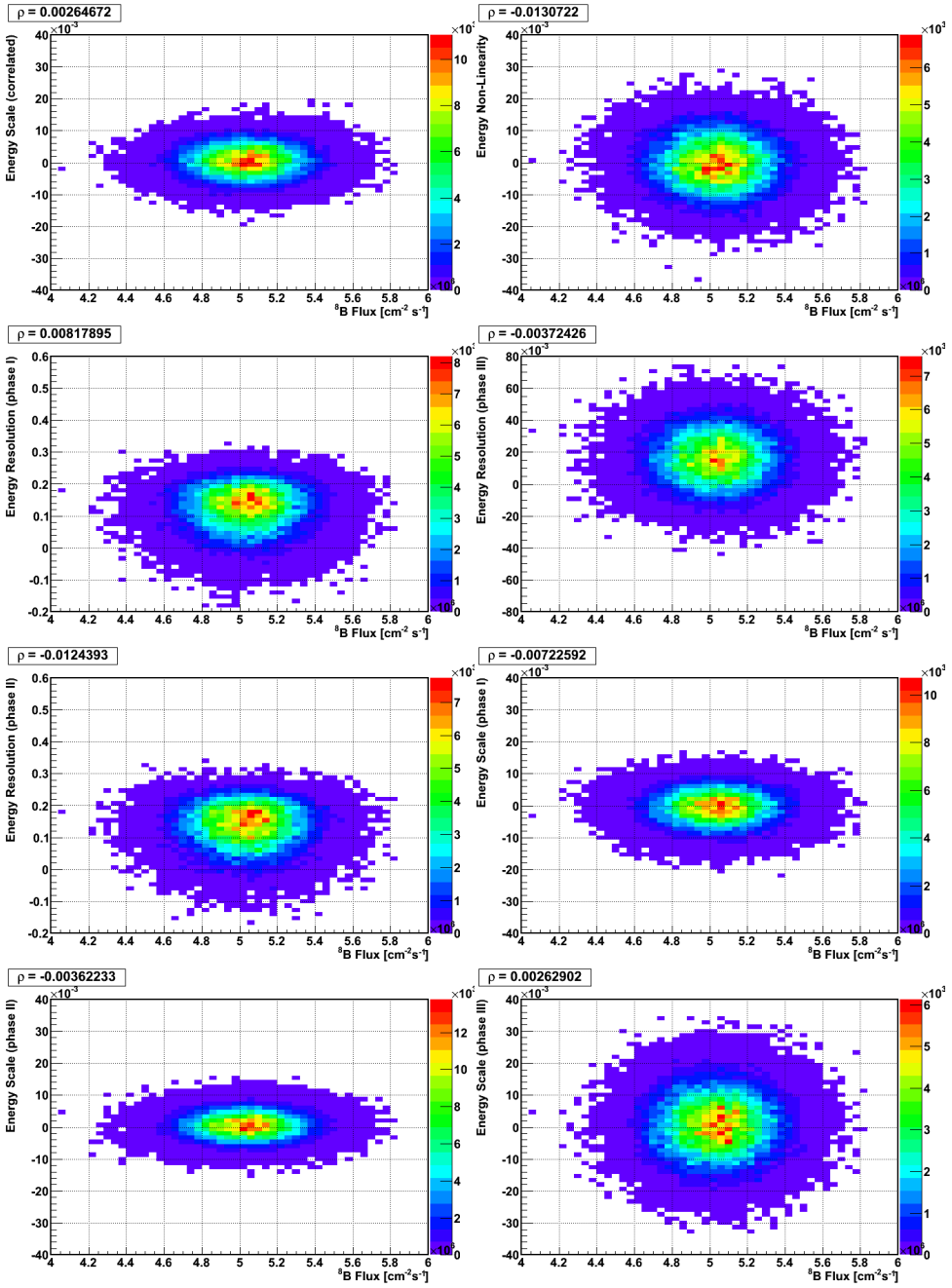

Figure A.1: Correlation between the ^{8}B flux and the energy systematic uncertainties.

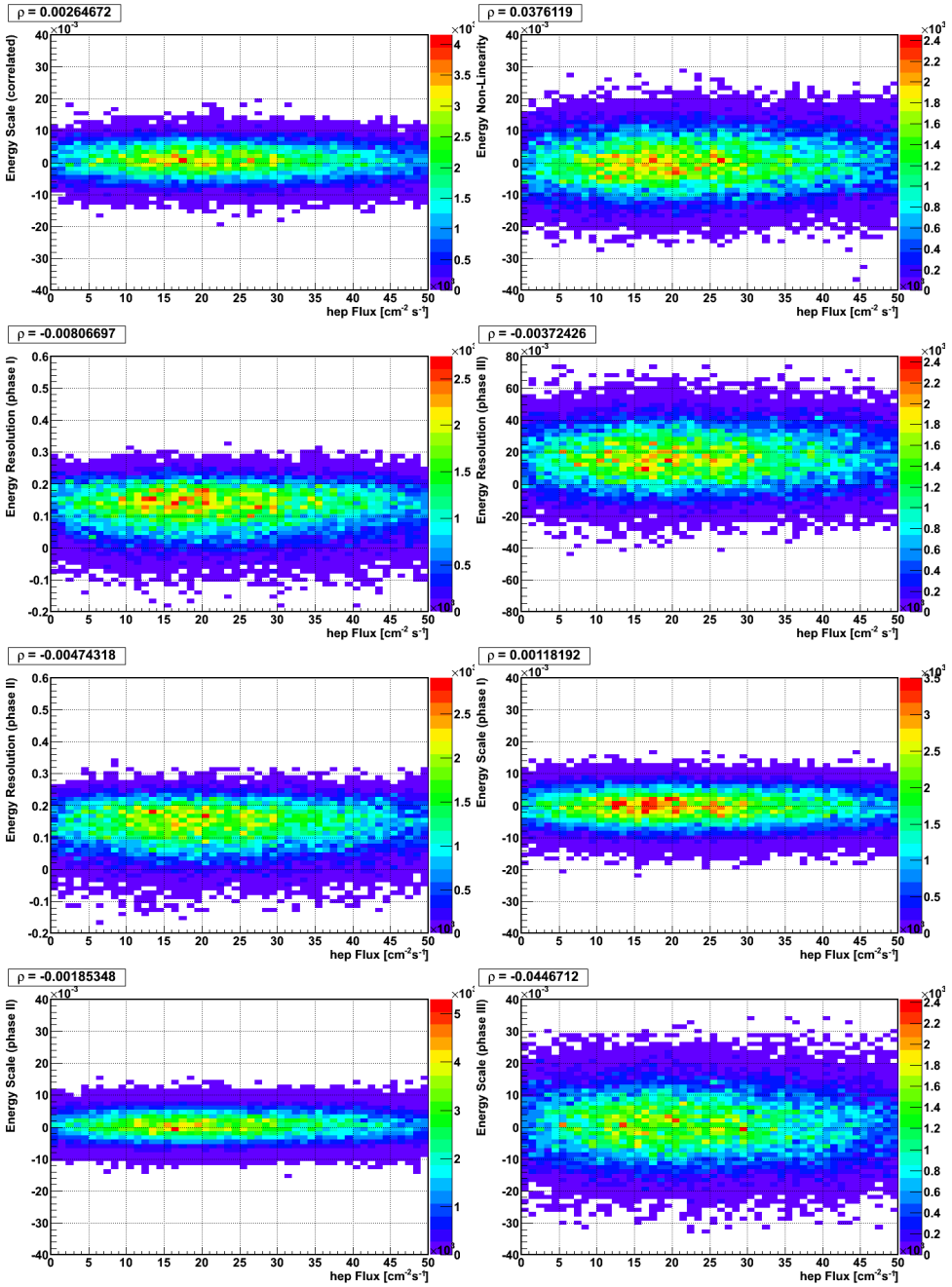

Figure A.2: Correlation between the *hep* flux and the energy scale systematic uncertainty.

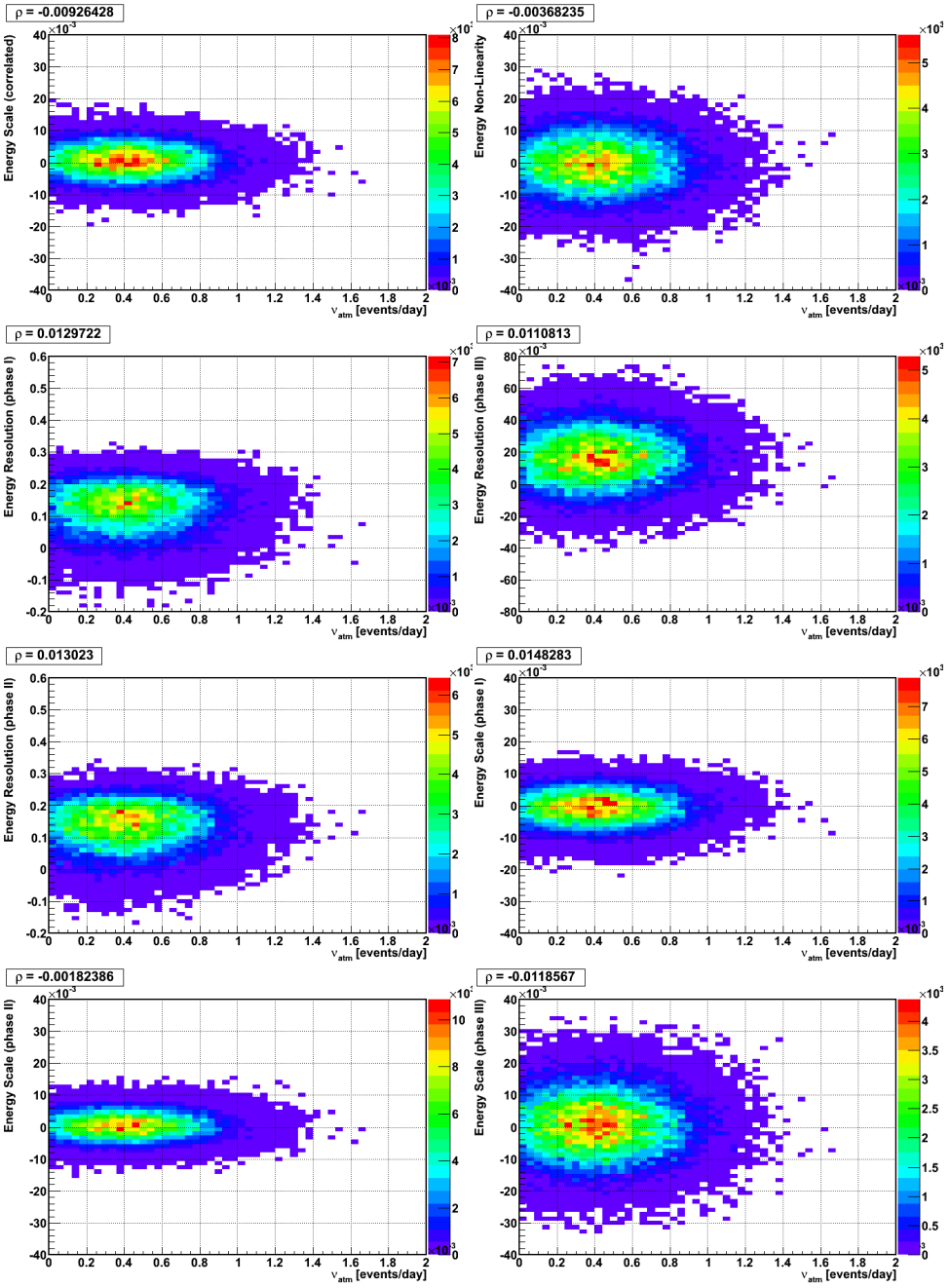

Figure A.3: Correlation between the ν_{atm} and the energy systematic uncertainties.

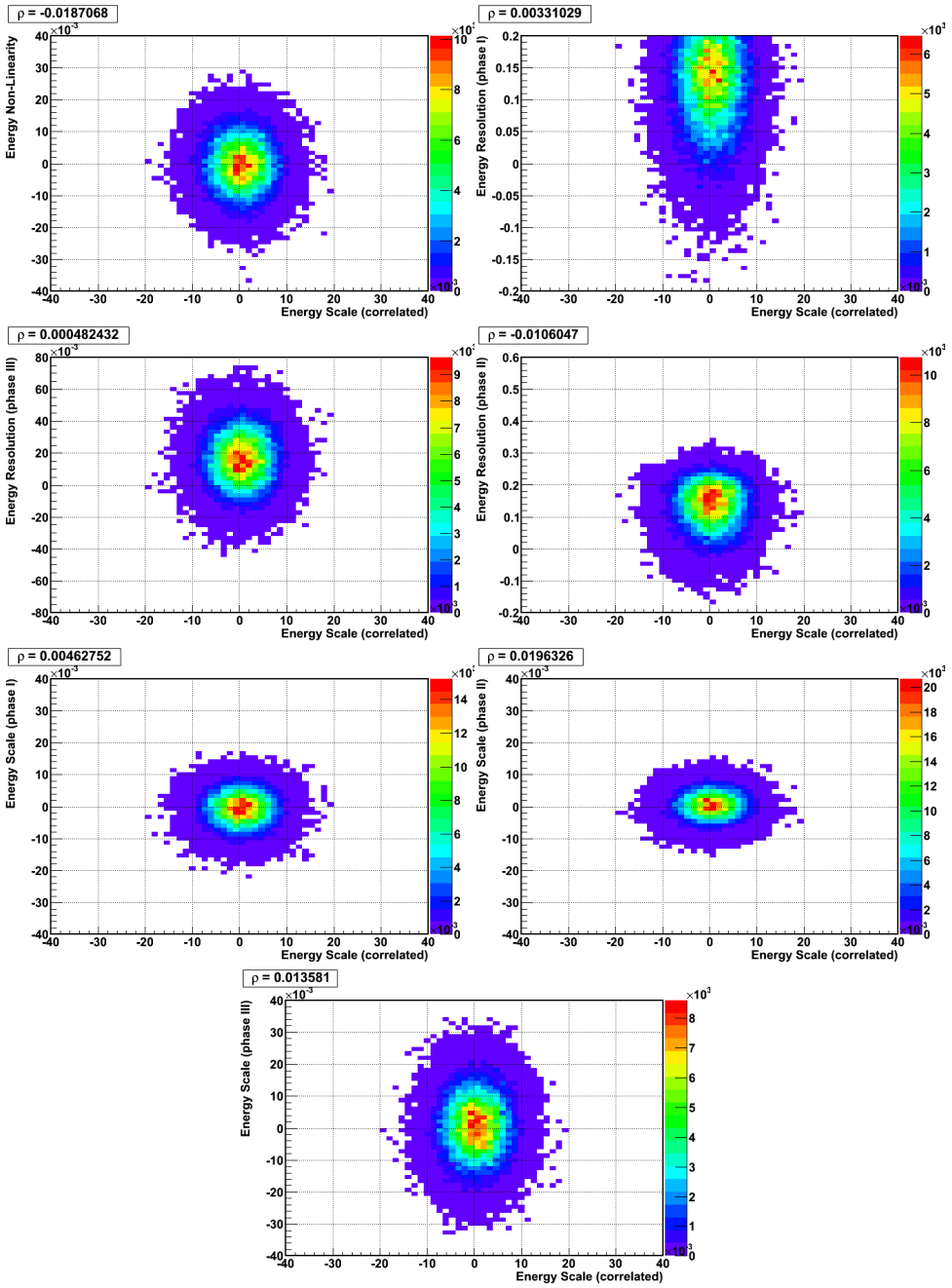

Figure A.4: Correlations between the energy scale, energy non-linearity, and energy resolution systematic uncertainties.

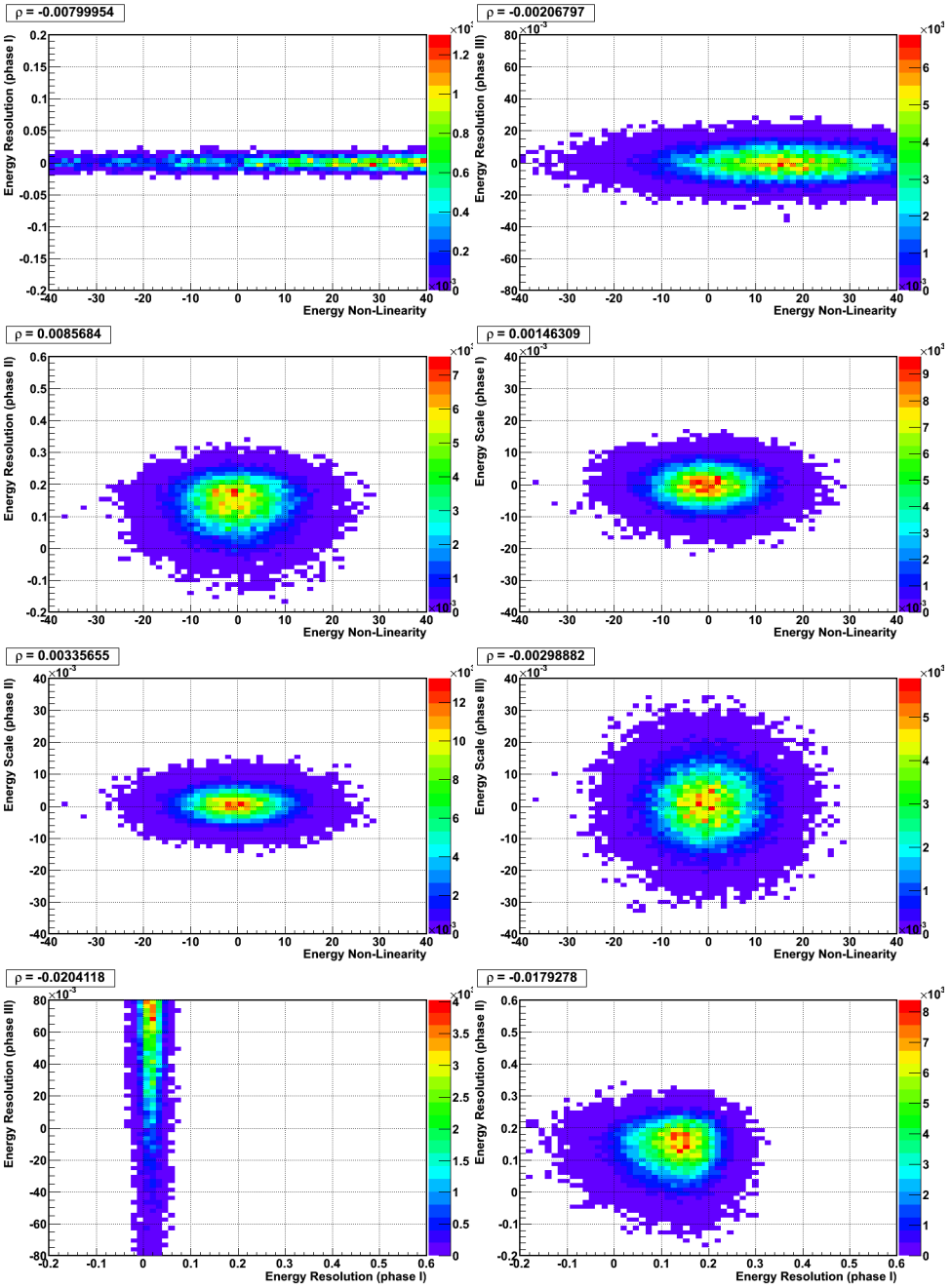

Figure A.5: Correlations between the energy scale, energy non-linearity, and energy resolution systematic uncertainties.

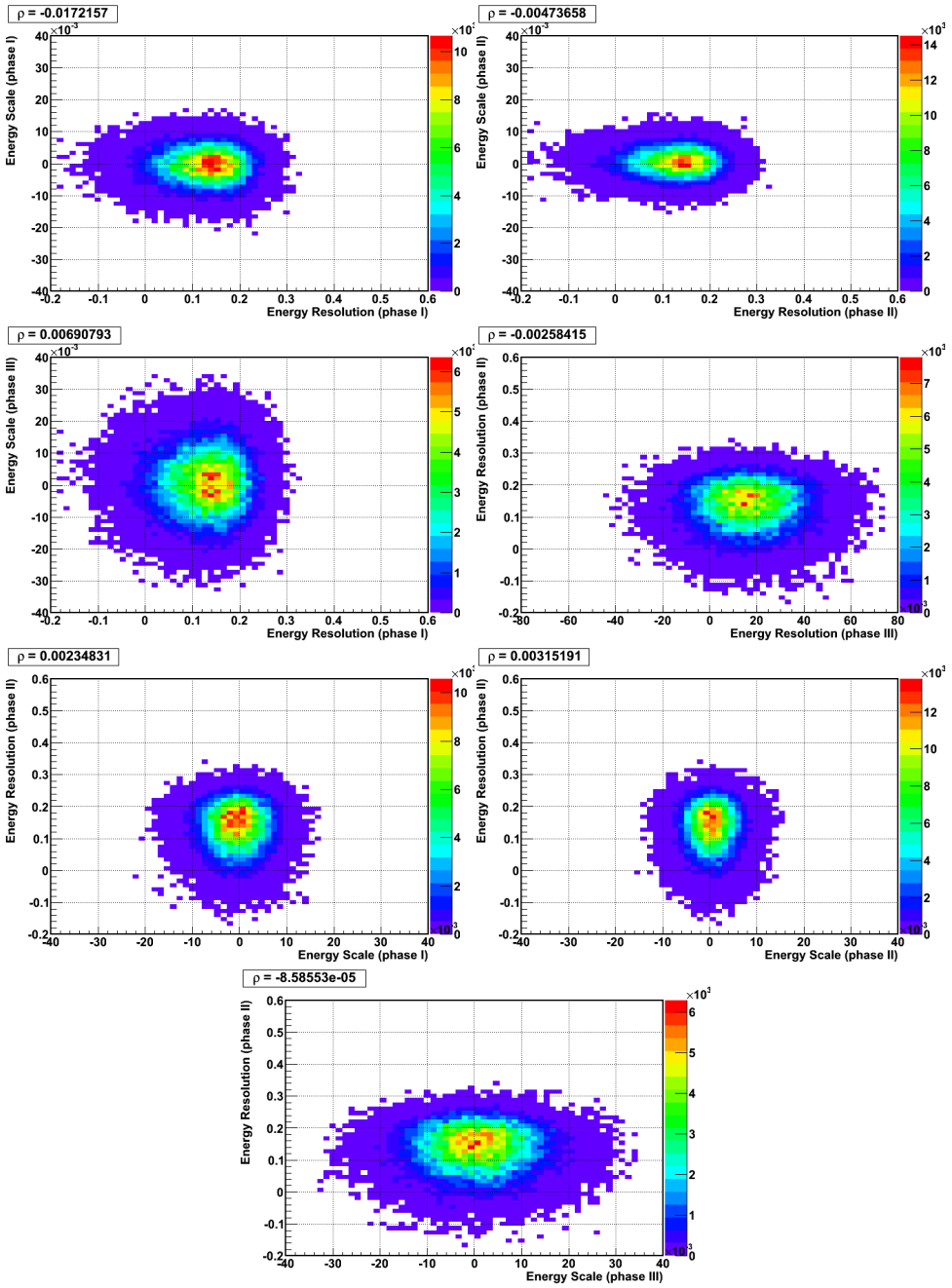

Figure A.6: Correlations between the energy scale, energy non-linearity, and energy resolution systematic uncertainties.

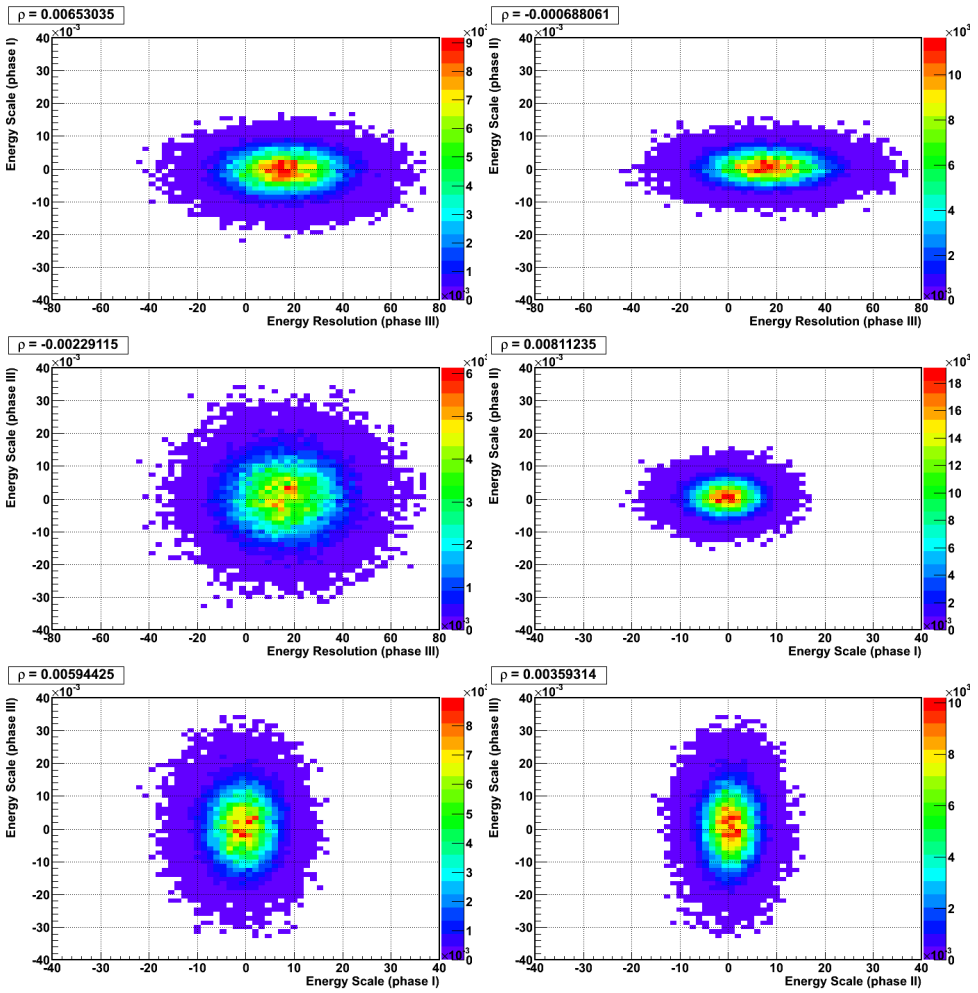

Figure A.7: Correlations between the energy scale, energy non-linearity, and energy resolution systematic uncertainties.