University of Alberta

A Combined Three-Phase Signal Extraction of the Sudbury Neutrino Observatory Data Using Markov Chain Monte Carlo Technique

by

Shahnoor Habib

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 2011
Edmonton, Alberta

Abstract

Neutrino physics has entered an era of precision, after proving that the Standard Solar Model is a viable theory and going beyond the current Standard Model of particle physics by proving that neutrinos possess nonzero masses. The Sudbury Neutrino Observatory (SNO) experiment, along with other experiments, has restricted neutrino mixing angle (θ_{12}) and the mass square difference (Δm^2_{21}) to lie within the large mixing solution area.

SNO, located 2 km underground in Sudbury, Canada, was an ultraclean heavy-water (D_2O) imaging detector for observing neutrinos produced by fusion reactions in the Sun. Neutrino interactions with heavy water resulted in flashes of light called Čerenkov radiation which was detected by an array of photomultiplier tubes. SNO took data from November 1999 to November 2006, totalling 1082 days of data taking.

This work describes an improved measurement of the mixing parameters from a combined fit of all the data. For the signal extraction fit on the data consisting of 4

observable of an event – radial position, recoil electron energy, direction relative to the Sun and event isotropy – Markov Chain Monte Carlo (MCMC) method based on Metropolis algorithm was employed. The nuisance parameters (systematics), weighted by external constraints, were allowed to vary in the fit. The goal of the thesis was to extract the survival probabilities of electron neutrinos and determine the total flux of active-flavour neutrinos from ^{8}B decay in the Sun measured through the neutral current interactions of neutrinos on deuterium. The ^{8}B flux from the fit is $(5.24 \pm 0.02) \times 10^6$ cm^{-2} s^{-1}; uncertainty from statistics and systematics is 3.56%. Along with ^{8}B flux, the fit extracted energy spectra of charged current interactions of neutrinos on deuterium and elastic scattering interactions of neutrinos on electrons. The fit described the energy-dependent day survival probability of solar neutrinos as a quadratic equation and asymmetry on the day survival probability as a linear equation. Four polynomial coefficients of the survival probability were extracted from the fit: constant coefficient as 0.3206 ± 0.0197, linear coefficient as 0.005 ± 0.008 and quadratic coefficient as -0.0014 ± 0.0033. There are two coefficients on the day-night asymmetry: constant coefficient as 0.0496 ± 0.0347 and the linear coefficient as -0.018 ± 0.028. The day-night asymmetry (0.0496) observed is 1.4σ away from zero. Using these findings, the oscillation space in terms of Δm^2_{21} and θ_{12} will be further constrained. Compared to the previous published SNO results, the uncertainty on ^{8}B went down from 3.83% to 3.56% and average ^{8}B ν_e survival probability (p_0) went down from 6.57% to 6.14%. If the data were analysed with the same assumptions, the decrease in uncertainties would have been approximately twice as big; however, more conservative systematic uncertainties were assigned in some cases.

Acknowledgements

I would like to thank my advisor, Dr. Aksel Hallin, who taught me a thing or two about testing the consistency of a code and write a simulation code consisting of just couple of lines to test a hypothesis (why assume when you can test it).

Next, I would like to thank Blair Jamieson, a post-doctoral fellow from the University of British Columbia, whose code I inherited and later expanded to include survival probability of electron neutrinos and sundry other features to the code.

From Berta Beltran, I learnt how to stand on my own two feet and dig up information without bothering her to look it up for me. From Chris Howard, I learned how not to take your work personally. He worked diligently but was quite comfortable in presenting his result, at any stage (complete, incomplete, with/without a bug in the code) for a discussion so he can get feedback which speeds up the work. Alex Wright, whom I met when I was a student at Queen's University in Kingston, got me started on my current research.

It would be remiss of me if I did not thank Pierre-Luc Drouin. Thanks for being extremely organized and helping me debug my codes on multiple occasions.

Many thanks to my friend, Raji (Rajalakshmi Sundaramoorthy), for reading my thesis and making various useful suggestions to improve it. Besides that she is a fabulous friend when you need one.

Table of Contents

List of Tables

List of Figures

Glossary	
AECL	Atomic Energy of Canada Limited
AV	Acrylic Vessel
CC	Charged Current Interaction
CP	Charge Parity violation
dof	degrees of freedom
ES	Elastic Scattering Interaction
GT	Global Trigger
GTID	Global Trigger Identification
ITR	In Time Ratio
KamLAND	Kamioka Liquid scintillator Anti Neutrino Detector
LETA	Low Energy Threshold Analysis
LMA	Large Mixing Angel
MC	Monte Carlo simulation code
MCMC	Markov Chain Monte Carlo
MSW	Mikheyev-Smirnov-Wolfenstein
NC	Neutral Current Interaction
NCD	Neutral Current Detector
NLL	Negative Log Likelihood
NHiT	Number of Photomultiplier tubes hit
OWL	Outward Looking PMT
PDF	Probability Density Function
PMT	Photomultiplier tubes
PSA	Pulse Shape Analysis
PSUP	PMT SUpport structure
RMS	Root Mean Square
SK	Super Kamiokande
SM	Standard Model
SNO	Sudbury Neutrino Observatory
SNOMAN	SNO Monte Carlo and ANalysis
SNP	Solar Neutrino Problem
SSM	Standard Solar Model

Table 1: List of acronyms.

Chapter 1

Introduction

1.1 Goal of the thesis

The goal of the thesis was to produce the most complete ^{8}B analysis of the solar neutrino data from Sudbury Neutrino Observatory (SNO). Signal extraction was carried out by Markov Chain Monte Carlo method based on Metropolis algorithm. The fit parameters consisted of ^{8}B flux from the measurement of the neutral current interactions in SNO and a set of polynomial parameters to describe the day time neutrino survival probability and a set of polynomial parameters to assign the asymmetry in the day and night neutrino survival probability.

1.2 Synopsis of Thesis

The first chapter is an introduction to solar neutrinos, Standard Solar Model, and solar neutrino experiments. The next chapter describes relevant features of the Sudbury Neutrino Observatory while the third chapter briefly describes the theory of neutrino oscillations. The third chapter also illustrates the importance of neutrino as a probe in the understanding of the mysteries of the Universe. The fourth chapter goes over the methodology of the signal extraction, describes the number of signals and backgrounds, the observable of the data, unique features of probability density functions that are used to distinguish various signals and backgrounds, the cuts applied on the data and the constraints applied on the fit. The fourth chapter also outlines the systematic uncertainties and the methods used for evaluating the goodness of fit. The Markov Chain Monte Carlo method exploited to extract the fit parameters of the models (Standard Solar Model and Neutrino Oscillation Model), is introduced in the fifth chapter. The sixth chapter focuses on a method to implement a constraint on the number of events when the width of the constraint is comparable to the constraint itself such that the Gaussian function traverses the negative region (non-physical). The seventh chapter goes over various cross-checks performed to make sure that the code is consistent. The eighth chapter presents the findings of running the MCMC code on a fit consisting of 4 signals and 1 background. This fit also floats various systematic uncertainties. The ninth chapter describes the result of the fit when a constraint from Pulse Shape Analysis (PSA) of the data from neutral current detectors is included in the fit. Five additional backgrounds were included in the fit. For this fit, nuisance parameters[1] were fixed. The tenth chapter goes over the result of an ensemble test when all the backgrounds are included. The next chapter has the results of ensemble tests for a fit on the third of the simulated datasets. After presenting the results, most of the chapter was devoted outlining the investigation carried out to discover the cause of bias in one of the main fitting parameter. Chapters 12 and 13 present the result of the fit on the third of the real data and the full real data respectively. The last chapter concludes the thesis.

[1] In a fit, there are parameters of interest and there are other (nuisance) parameters. The nuisance parameters, though not of interest, must be accounted for because of their effect on the parameters of interest. Examples are given in section 4.11.

1.3 Neutrinos in the Standard Model (SM)

1.3.1 The Standard Model in a Nut Shell

A scientific model is a description of nature, created by the human mind, to explain what happens in nature. For example, the matter we see around us is composed of elementary particles[2]. The Standard Model is a theory that describes properties of elementary particles and their interactions among themselves. Out of the four fundamental interactions – gravitation, electromagnetism, strong interaction and weak interaction – only gravity is not included in the SM. According to SM, there are 12 fundamental particles of spin $\frac{1}{2}\hbar$, known as fermions, which are classified according to their charges. The list includes 3 types (flavours) of charged leptons (electron, muon and tau), each with a corresponding neutrino, and six flavours of quarks. The six flavours of quarks are named: up (**u**), down (**d**), charm (**c**), strange (**s**), top (**t**) and bottom (**b**). Each particle has an associated antiparticle, with the same mass but opposite charge. For example, the antiparticle of the electron (e^-) is the positron (e^+). The SM particles are shown in figure 1.1. Each column of fermions is called a generation or a family. Electron and electron neutrino are part of one generation; tau and tau neutrino forms another generation and so on. In the quark sector, up and down form one generation and so on. The basic components of ordinary matter are electrons, protons and neutrons of which the later two are combinations of two types of quarks - **uud** and **udd** respectively. Protons and neutrons are called baryons and electrons and neutrinos are called leptons.

In the SM, neutrinos are electrically neutral fermions which interact with other particles via weak interactions only. Neutrinos come in three flavours [2]; each flavour is associated with a charged lepton: electrons with electron

[2] An elementary particle is not composed of any other particles, that is, it does not have an internal structure.

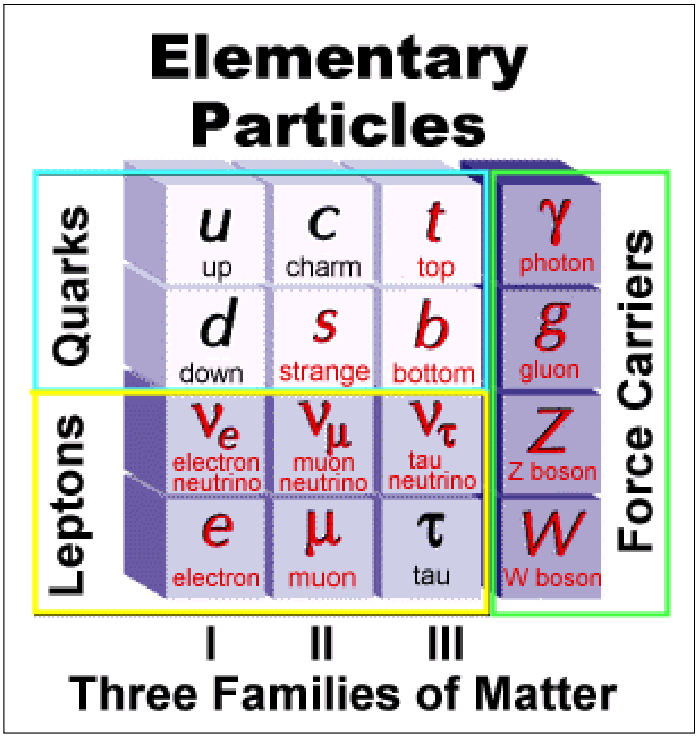


Figure 1.1: Fundamental fermions and bosons in the Standard Model. Figure from [1].

neutrinos (e, ν_e), muons with muon neutrinos (μ, ν_μ) and taus with tau neutrinos (τ, ν_τ). Each neutrino ν has an antineutrino which is represented by $\bar{\nu}$. For example, the antiparticle of ν_e is $\bar{\nu}_e$. Neutrinos are created in beta (β) decays which are described as:

$$n \longrightarrow p^+ + e^- + \bar{\nu}_e \tag{1.1}$$

$$p^+ \longrightarrow n + e^+ + \nu_e \tag{1.2}$$

In the β decay, described by equation (1.1) and shown in figure 1.2, at a fundamental level, a down quark is converted into an up quark via emission of a W^- boson which subsequently decays into an electron (e^-) and an electron antineutrino ($\bar{\nu}_e$).

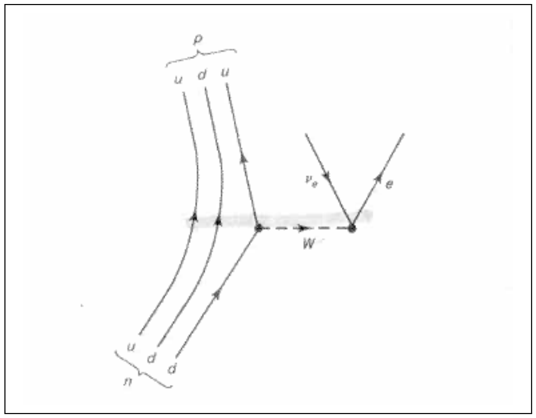


Figure 1.2: Decay of a neutron into a proton p^+, an electron e^-1 and an electron antineutrino ν_e mediated via a virtual W^- boson. Figure from [5].

1.4 The Role of Neutrinos in the Standard Solar Model (SSM)

There are various models to describe and predict the behaviour of the Sun. The Solar Standard Model refers to the model devised by John Bahcall ([6] and [7]) according to which stars have two mechanisms available to sustain their luminosity: pp cycle (listed in table 1.1) and Carbon-Nitrogen-Oxygen (CNO) cycle (listed in table 1.2). The primary source (98.5%) of solar energy is the pp chain. The rest is provided by the CNO cycle. The net pp chain reaction is:

$$4p \rightarrow {}^4\text{He} + 2e^+ + 2\nu_e + 26.73\,\text{MeV} \tag{1.3}$$

The pp chain reaction converts four protons into an α, two electron neutrinos ν_e, two positrons e^+ and energy which is released as gamma rays and kinetic energy of the particles. The average energy of the two neutrinos is $\langle E_\nu \rangle \approx 0.6$ MeV [8].

The pp chain burns hydrogen into helium in the core of the Sun. As hydrogen burns, the interior undergoes significant changes in size, luminosity and core temperature. The SSM fixes the initial elemental abundances according to the observed abundances in solar-system meteorites and the unmixed photosphere of the Sun. The model assumes that the Sun is in hydrostatic equilibrium; the outward pressure of photons and particle radiation is balanced by gravity. The model also assumes that initially Sun was of homogeneous composition and the change in the abundance of elements happens with time because of fusion within and not from diffusion from the outside. The model is evolved in time within certain constraints, for example, current photon luminosity, mass, radius and the age of the Sun as listed in table 1.4. When the model converges on the measured solar parameters, the model predicts the mass and temperature distribution in the Sun and the solar neutrino flux from the core. The energy spectrum of solar neutrinos due to the nuclear processes, listed in table 1.1, is shown in figure 1.3.

Reaction	ν_e **Energy (MeV)**
$p + p \rightarrow {}^2\mathrm{H} + e^+ + \nu_e$ (pp)	≤ 0.424
or	
$\mathrm{p} + e^- + \mathrm{p} \rightarrow {}^2\mathrm{H} + \nu_e$ (pep)	1.422
${}^2\mathrm{H} + \mathrm{p} \rightarrow {}^3\mathrm{He} + \gamma$	
${}^3\mathrm{He} + {}^3\mathrm{He} \rightarrow \alpha + 2\mathrm{p}$	
or	
${}^3\mathrm{He} + {}^4\mathrm{He} \rightarrow {}^7\mathrm{Be} + \gamma$	
${}^7\mathrm{Be} + e^- \rightarrow {}^7\mathrm{Li} + \nu_e$	(90%) 0.861
	(10%) 0.383
${}^7\mathrm{Li} + \mathrm{p} \rightarrow 2\ \alpha$	
or	
${}^7\mathrm{Be} + \mathrm{p} \rightarrow {}^8\mathrm{B} + \gamma$	
${}^8\mathrm{B} \rightarrow {}^8\mathrm{Be}^{\star} + e^+ + \nu_e$	< 15
${}^8\mathrm{Be}^{\star} \rightarrow 2\ \alpha$	
or	
${}^3\mathrm{He} + \mathrm{p} \rightarrow {}^4\mathrm{He} + e^+ + \nu_e$ (hep)	≤ 18.77

Table 1.1: Nuclear reactions in the proton-proton chain along with neutrino energy. Table from [8].

Reaction	ν_e **energy (MeV)**
^{12}C+p $\rightarrow$ ^{13}N+γ	
^{13}N $\rightarrow$ ^{13}C+e^++ν_e	≤ 1.199
^{13}C+p $\rightarrow$ ^{14}N+γ	
^{14}N+p $\rightarrow$ ^{15}O+γ	
^{15}O$\rightarrow$ ^{15}N+e^++ν_e	≤ 1.732
^{15}N +p $\rightarrow$ ^{12}C+ α	

Table 1.2: Nuclear reactions in the CNO chain along with neutrino energy. Table from [8].

Electron neutrinos are produced in four reactions in this chain. The CNO chain also produces neutrinos but is negligible in the Sun. Detection of CNO neutrinos will enable us to differentiate between various metallicity models of the Sun. The SSM makes a number of predictions which can be tested and one of these is the electron neutrino flux produced by each of the four reactions in the pp chain (table 1.3). The model also predicts where the different neutrino fluxes originate in the Sun, as shown in figure 1.4.

1.5 Solar Neutrino Problem

In the late 1960s, Ray Davis Homestake Experiment in a Gold Mine in South Dakota measured the flux of neutrinos from the Sun and detected a deficit [9]. According to SSM, Sun only produces electron neutrinos. The deficit of the solar electron neutrino flux (figure 1.5 and table 1.5) from the neutrino flux predicted by the SSM (table 1.3) is known as the Solar Neutrino Problem (SNP). The discrepancy, lasting from early seventies to about 2002, has since been resolved by introducing neutrino oscillations to the standard model.

There were several proposals to explain the deficit but any model requiring the change in the solar model has to overcome the success in the prediction of the total ^{8}B solar neutrino flux [10] listed in table 1.3 and the current observ-

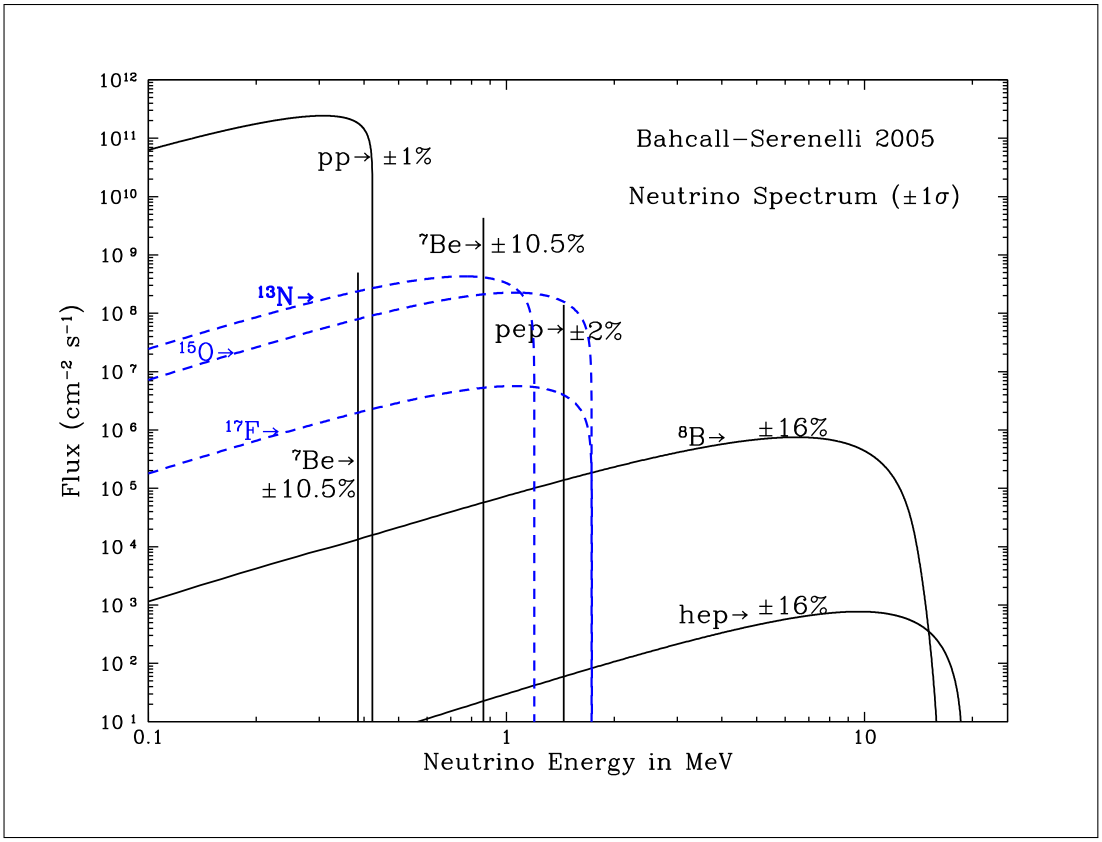


Figure 1.3: This figure, from [13], shows solar neutrino spectrum that is predicted by standard solar model from the CNO cycle (in blue dotted lines) and pp chain. Continuum source fluxes are given in units of neutrinos cm^{-2} s^{-1} MeV^{-1} while line source fluxes (^{7}Be and pep) are given in units of neutrinos cm^{-2} s^{-1}. The total theoretical errors for each source is also indicated in the figure.

able, for example, luminosity, radius and age of the Sun as listed in table 1.4. Additionally, the models must predict the core temperature because of its effects on the total numbers of neutrinos emitted. The core environment affects the neutrino flux from each of the ν reactions but not its energy spectrum. The model has to explain the energy-dependent distortion of the ν_e flux observed in experimental results. The solution to this problem, as demonstrated by SNO, is that solar neutrinos change flavour on the fly. The phenomena of ν oscillation – whereby neutrinos oscillate back and forth between different flavours – is physics beyond the Standard Model as SM assumes neutrinos to

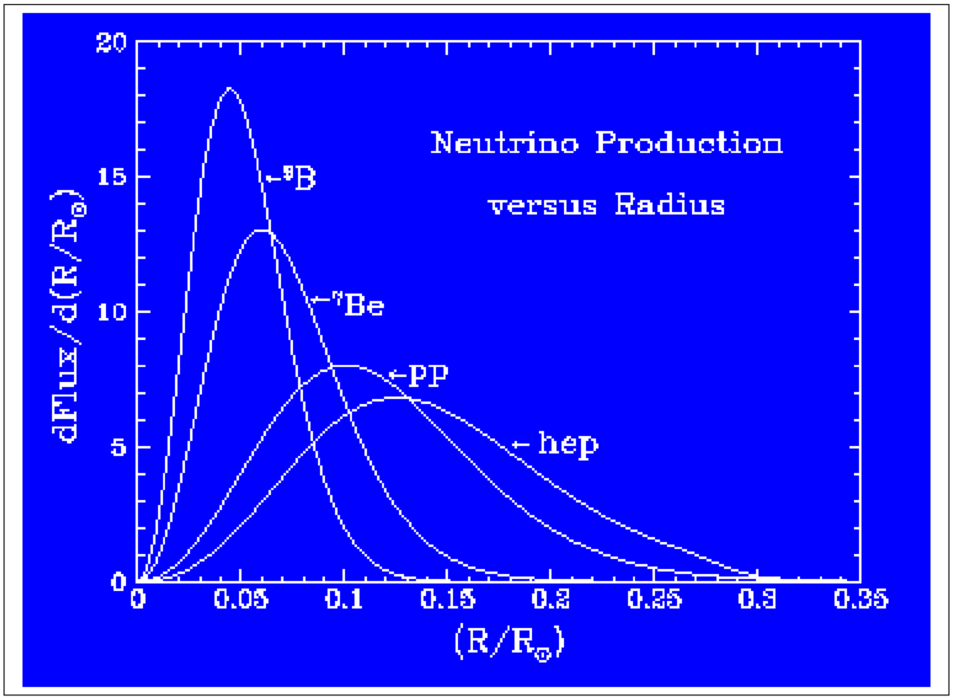


Figure 1.4: Neutrino production as a function of radial distance of the Sun. Figure from [8].

be massless, thereby, can not change flavours. In consequence of the oscillations, we know that neutrinos do have mass, this mass although very tiny[3], contributes as much to the universe as the combined mass of all the stars in the galaxies.

1.6 Experiments, Advantages and Constraints

Three experimental methods are employed to detect solar neutrinos: radiochemical, water Čerenkov and scintillator. Radiochemical experiments, only sensitive to electron neutrinos, do not measure the energy of the detected neutrinos but measure the rate of neutrino induced events above a fixed energy threshold. The advantage of radiochemical experiments is a low energy threshold. They can detect neutrinos with energy less than one MeV. Examples

[3]Cosmological constraints to the sum of ν mass $\Sigma = \Sigma m_\nu$ typically range below 1 eV [11]. The beta spectrum of tritium [15], limits the sum of active neutrinos to be between 0.05 and 8.4 eV.

Reaction	**Flux ($cm^{-2}s^{-1}$)**	**Maximum Energy (MeV)**
pp	$5.94(1.00^{+0.01}_{-0.01}) \times 10^{10}$	0.42
pep	$1.39(1.00^{+0.01}_{-0.01}) \times 10^{8}$	1.44
hep	$2.10{\times}10^{3}$	18.77
^{7}Be	$4.80(1.00^{+0.09}_{-0.09} \times 10^{9}$	0.86 (90%) 0.38 (10%)
^{8}B	$5.15(1.00^{+0.19}_{-0.14} \times 10^{6})$	≈ 15
^{13}N	$6.05(1.00^{+0.19}_{-0.13} \times 10^{8})$	1.20
^{15}O	$5.325.32(1.00^{+0.22}_{-0.15} \times 10^{8})$	1.73
^{17}F	$6.33(1.00^{+0.12}_{-0.11} \times 10^{6})$	1.74

Table 1.3: Predicted fluxes of neutrinos from the solar nuclear fusion reactions along with the maximum energy. The errors quoted, for the predictions from SSM BP98, are 1σ theoretical uncertainties. Reactions in column one are described in table 1.1. This table is from [12].

Parameter	**Value**
Mass $M_{\odot}$	$(1.9891 \pm 0.0004) \times 10^{23}$ g
Radius $R_{\odot}$	$(6.9599 \pm 0.0002) \times 10^{10}$ cm
Luminosity $L_{\odot}$	$(3.846 \pm 0.004) \times 10^{3}$ Joules/s
Neutrino Luminosity	$0.022\ L_{\odot}$
Age	$(4.52 \pm 0.04) \times 10^{9}$ yr

Table 1.4: Few observed solar parameters from [14].

of radiochemical detectors are Homestake, GALLEX (Gallium Experiment), GNO (Gallium Neutrino Observatory) and SAGE (Soviet-American Gallium

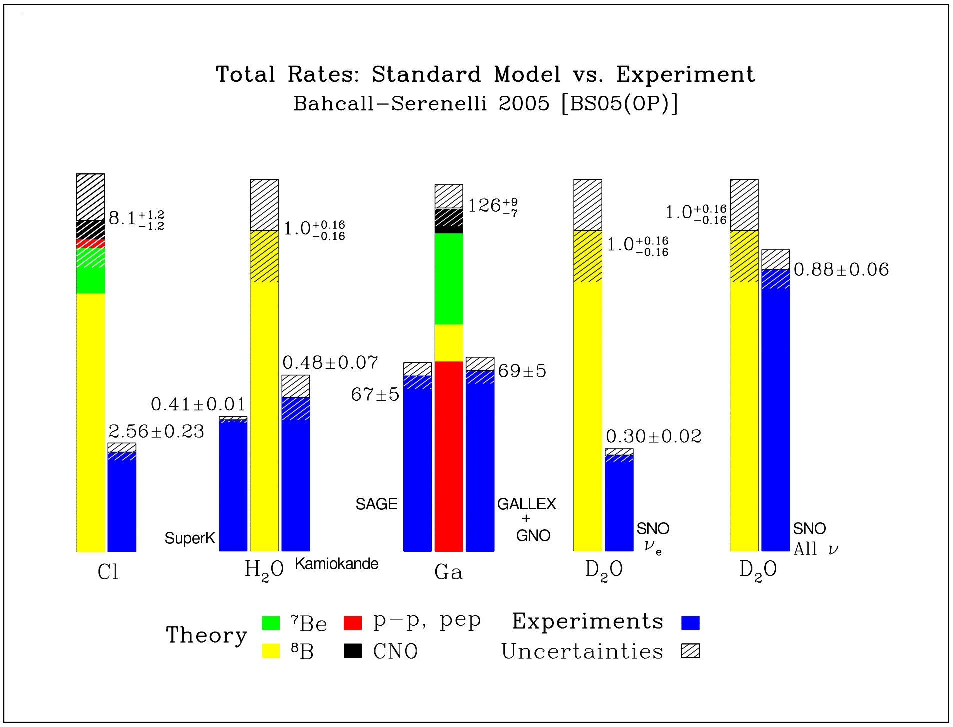


Figure 1.5: Comparison of solar neutrino flux observed in the experiments compared to the theory. Figure from [16]. The unit of gallium (SAGE and GALLEX+GNO) and chlorine (Homestake) experiments is Solar Neutrino Unit (SNU) which is 10^{-36} ν reaction per second per target atom. The unit of water Čerenkov experiments (Kamiokande, SuperKamiokande and SNO) is the flux obtained from the experiment divided by the flux predicted from the Standard Solar Model BS05.

Experiment). The targets consisted of either chlorine or gallium. Neutrinos were detected via the following reactions:

$$\nu_e +^{37} \text{Cl} \to e^- +^{37} \text{Ar} \tag{1.4}$$

$$\nu_e +^{71} \text{Ga} \to e^- +^{71} \text{Ge} \tag{1.5}$$

The neutrino flux was calculated by counting the occurrences of either argon or germanium by detecting their radioactive decays. The low-energy threshold (only 0.233 MeV) of the Gallium experiments (GALLEX, GNO and SAGE) enables them to observe neutrino captures from all pp chain neutrinos.

Examples of water Čerenkov experiments are Kamiokande, SuperKamiokande[4]

[4]Both located in Kamioka zinc mine in Japan.

(Super-K) and Sudbury Neutrino Observatory (SNO). The Kamiokande experiment, consisting of 680 tonnes of water, detected neutrinos via their elastic scattering[5] (ES) reaction: $\nu_x + e^- \rightarrow \nu_x + e^-$ where x is any flavour: e, μ or τ. The Čerenkov light, generated by recoil electrons, is detected by a set of photomultiplier tubes (PMTs) directed at the target volume. The advantage of water detectors is that they detect neutrinos in real-time, with directional, spectral and time information determined on an event-by-event basis. The disadvantage is that the threshold in energy is much higher than the radiochemical experiments, at the order of 5 MeV, which limits the detection sensitivity to neutrinos from ^{8}B and hep branches of the pp chain (figure 1.3). The cross section for interaction increases with neutrino energy, hence neutrinos from ^{8}B branch are easier to observe because of their higher energy range (0-14 MeV). Since the detected low energies are dominated by experimental backgrounds due to radioactivity, availability of the energy range 0 to 14 MeV enable us to put a detector threshold on the energy such that a large proportions of backgrounds are removed (figure 4.2) without incurring a comparable reduction in the statistics.

Experiment	**Detection Method**	**Flux (observed/predicted)**
Homestake	Radiochemical	0.34 ± 0.06
GALLEX & GNO[6]	Radiochemical	0.58 ± 0.07
SAGE[7]	Radiochemical	0.59 ± 0.07
Kamiokande	Water Čerenkov	0.55 ± 0.13
Super-K	Water Čerenkov	0.45 ± 0.08

Table 1.5: Pre-SNO results of the solar neutrino experiments in comparison to the prediction from the Bahcall-Pinsonneault BP2000 SSM. Table from [17].

.

Table 1.5 summarizes the results from the first generation of solar neutrino experiments. The discrepancy between the individual experimental results is

[5]The ES reaction, though sensitive to all flavours, has reduced sensitivity to ν_μ and ν_τ.

due to a difference in the energy threshold of the experiments which makes them sensitive to neutrinos from some or all four neutrino-producing reactions in the pp chain. Furthermore, the exact suppression is also dependent on the energy threshold of the experiment.

1.6.1 Super-Kamiokande

Super-Kamiokande, abbreviated as Super-K, is a large, underground, water Čerenkov neutrino detector located 1000 m underground in an active zinc mine in the Japanese Alps mountain ranges. It consists of a cylindrical stainless steel tank (41.4 m tall and 30.3 m in diameter) holding 50,000 tons of ultra-pure water. The tank volume is divided into two regions: a large inner region and a 2 metre wide outer region. The inner region is optically isolated from the outer region by a stainless steel superstructure. Mounted on the superstructure are 11,146 photomultiplier tubes (50.8 cm in diameter) that face the inner region and 1885 (20.3 cm in diameter) that face the outer region. Neutrinos are detected via $\nu_e - e$ elastic scattering. The Super-K detector, employs the same ES reaction as the Kamiokande detector, to monitor the neutrino flux but with a fiducial mass 33.1 times greater than the original experiment. The Čerenkov light, emitted by recoil electrons, is detected by the photomultiplier tubes. The interaction vertex, ring direction and flavour of the incoming neutrino is determined from the charge collected on the PMTs, the sharpness of the ring on the wall and the timing information recorded by each photomultiplier tube. Using elastic scattering interactions, Super-K provided the direct evidence that the Sun is a source of neutrinos, as shown in figure 1.6, and made critical contributions towards the resolution of the solar neutrino problem. In February 1987, Super-K detected neutrinos created by a supernova (SN 1987A, located in the Large Magellinic Cloud). Besides solar neutrinos, Super-K also detects interactions of ∼1 GeV neutrinos produced by interactions of cosmic rays with air molecules in the upper atmosphere. From these atmospheric neutrinos, Super-K glimpsed the first hint of neutrino oscillations. More ν_μ were detected coming from above than from below (figure 3.3 in chapter 3.). The data were in good agreement with two-flavour $\nu_\mu \Leftrightarrow \nu_\tau$ oscillations with $\sin^2 2\theta_{23} > 0.82$ and $5 \times 10^{-4} < \Delta m^2_{32} < 6 \times 10^{-3}$ eV^2 at 90% confidence level [18].

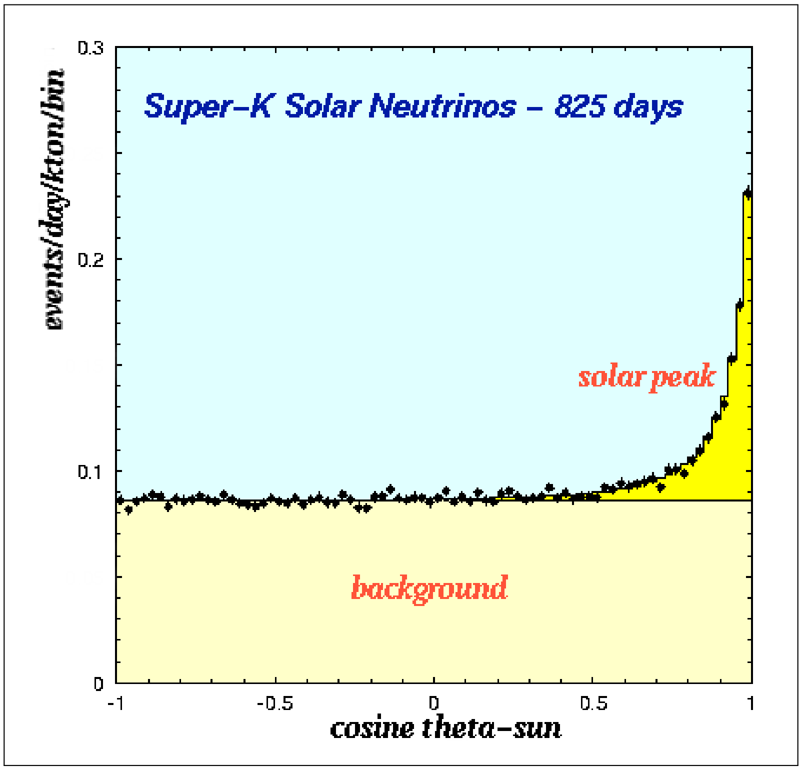


Figure 1.6: Plot of angular distribution of recoil electrons relative to the Sun. A peak at Cosine $\theta \approx 1$ points to solar neutrinos. After subtracting background due mostly from radon gas in the water, the area under the peak, is the measured number of solar neutrinos. Figure from [19].

1.6.2 KamLAND Result

The Kamioka Liquid scintillator Anti-neutrino Detector (KamLAND) is a reactor antineutrino experiment in Japan and detects antineutrinos ($\bar{\nu_e}$) from 53 nuclear reactors in the surrounding area [20]. The experiment probes θ_{12} and Δm^2_{21} neutrino mixing parameters without complications from the enhancement of neutrino oscillation in matter because the average distance between the reactors and the detector is roughly 180 km. It extracted neutrino oscillation parameters by observing two complete oscillation cycles in the $\bar{\nu}_e$ spectrum (figure 3.4 in chapter 3). KamLAND is located at the site of the former Kamiokande experiment. The heart of the experiment is a 18m diameter stainless steel sphere, containing liquid scintillator and, surrounded by 1879 50 cm diameter photomultiplier tubes (PMTs) for antineutrino detection. Electron antineutrinos are detected via inverse $\beta-$decay ($\bar{\nu}_e + p^+ \longrightarrow e^+ + n$) with a 1.8 MeV threshold. Assuming CPT[8] invariance, the result of global analysis (figure 1.7) of KamLAND, SNO and other solar ν experiments is $\Delta m^2_{21} = 7.59^{+0.21}_{-0.21} \times 10^{-5}\, eV^2$ and $\tan^2\theta_{12} = 0.47^{+0.06}_{-0.05}$ [20].

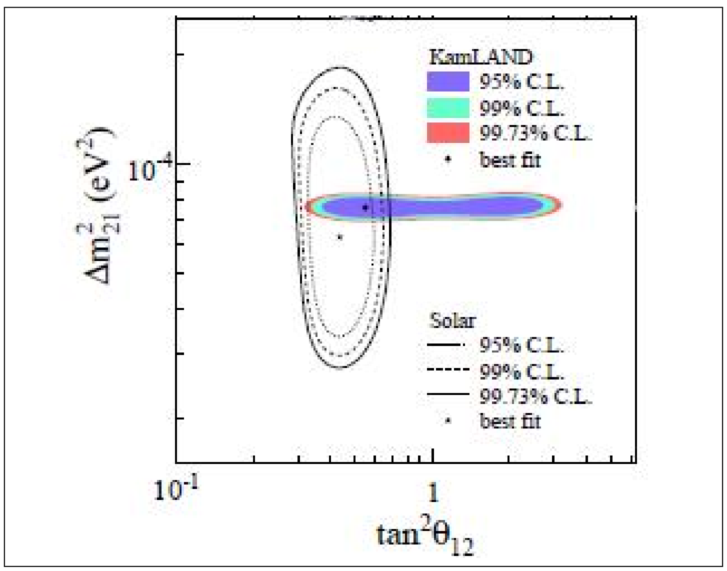


Figure 1.7: Global analysis of parameter space – Solar+KamLAND. Figure from [20].

1.7 The Sudbury Neutrino Observatory (SNO) detector

The Sudbury Neutrino Observatory will be described in detail in chapter 2.

[8]CPT invariance means that if a particle is replaced with its corresponding antiparticle (charge conjugation – C), and the space coordinates (parity – P) and time (T) are reversed, the physical laws are unchanged.

Chapter 2

Sudbury Neutrino Observatory

This chapter briefly describes the SNO detector. For a full technical report of the detector and all of its subsystems, refer to [21]. SNO, shown in figures 2.1 and 2.2, is an enormous optical instrument to detect short bursts of Čerenkov light associated with neutrino interactions. It is situated in Vale Inco's Creighton Mine in Sudbury, Ontario, Canada. The detector was proposed to: clarify the basic energy generation processes in the Sun, test the hypothesis of ν oscillation and determine the fundamental properties of neutrinos by studying ^{8}B neutrinos emitted from the core of the Sun. The ability of neutrinos to penetrate vast distances through dense matter without interacting makes them an excellent probe in investigating the processes that generate them. These include fusion reactions in the core of stars, supernovae explosions, radioactive decays in the Earth's core, mantel and crust. The ability to penetrate matter while weakly interacting with it makes neutrinos extremely hard to detect hence it is important to maximize the mass and sensitivity of the detector without increasing the backgrounds. To achieve this objective, one 10^6 kg of heavy water (D_2O) used as an active medium, was enclosed in a transparent acrylic vessel (12 m in diameter) to intercept about 10 neutrinos per day. Surrounding the acrylic vessel (AV) is a geodesic stainless-steel structure, 17.8 metres in diameter, for carrying 9438 inward-looking photomultiplier tubes (PMTs) (figure 2.2). The space between the AV and the PMT SUPport structure (PSUP) is filled with light water (H_2O). The barrel-shaped cavity, housing SNO target detector, is 22 metres in diameter at maximum and 34 metres in height. The space between the PSUP and the cavity walls is also filled with light water. The D_2O is unique because it offers equal sensitivity to all of the neutrino types (ν_e, ν_μ and ν_τ). A separate reaction (Charged current is described in equation (2.1)) has sensitivity to electron neutrinos (ν_e) only. The light water, surrounding the heavy water, provides both buoyancy for the vessel and radioactive shielding against external neutrons, radioactivity in the PMTs and radiation emanating from the rocks in the cavity. The location of SNO, 2 km underground, protects it from cosmic rays, especially cosmic ray induced muons.

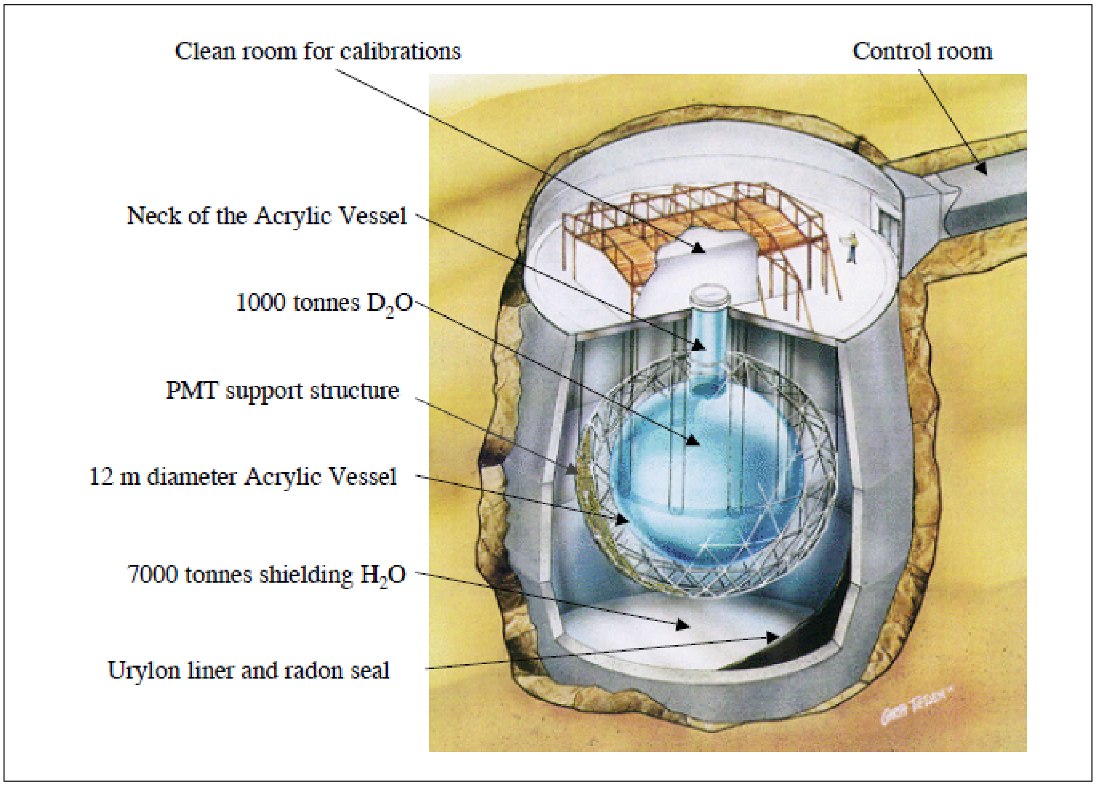


Figure 2.1: Artist's rendering of SNO detector, showing the acrylic vessel (AV), the PMT SUPport structure (PSUP), the control room, and the clean room above the neck of the AV.

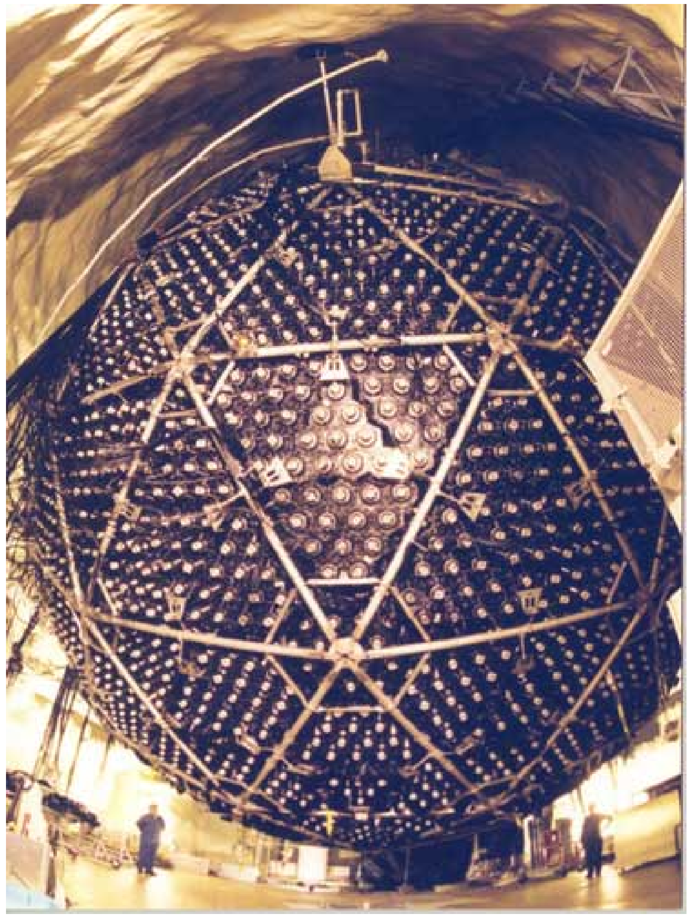


Figure 2.2: View of the PMT support structure (PSUP) in SNO.

The aim of Sudbury Neutrino Observatory (SNO), proposed by Herb Chen in 1984 [30], was to look into the solar neutrino problem [23] and test the neutrino oscillation model using solar neutrinos. To accomplish this goal, SNO detector was constructed in a large cavity, 2,000 metres below ground, in an active nickel mine near Sudbury, Ontario. Two factors are influential in detecting neutrinos: reduced backgrounds and increased detection volume. The rock overburden reduced the background rate of muons from cosmic radiation to roughly 70 per day. The shielding provided by the rock overburden is equivalent to 6010 metres of water. The increased detection volume consisted of 10^6 kg of heavy water (D_2O) which was borrowed from the Atomic Energy of Canada Limited (AECL). As shown in figure 2.1, the walls of the

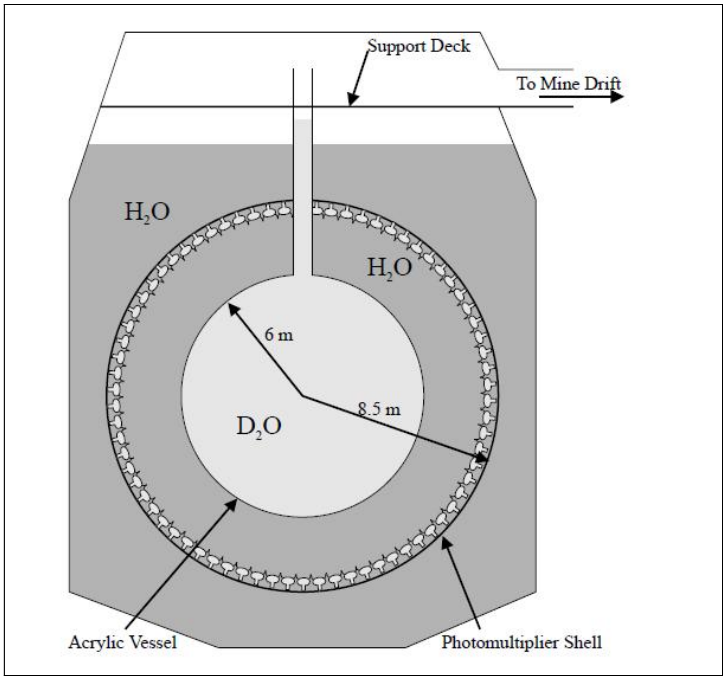


Figure 2.3: A schematic view of SNO.

cavity are covered with an urylon plastic liner to prevent any material from the surrounding rock leaking in the pure water.

The acrylic sphere, constructed by bonding 122 panels of ultraviolet transmitting acrylic together, has a thickness of 5.5 cm in most places. An opening, resembling a chimney or neck, at the top of the acrylic vessel is 1.5 metres in diameter and 6.8 metres in height. The AV is suspended by 10 ropes to the support deck which is shown in figure 2.3. The ropes are connected to 11.4 cm thick rope groove panels at the belly of the sphere. Figure 2.4 shows the acrylic sphere, the suspension ropes, rope groove panels and the neck.

Events within the detector were observed by watching Čerenkov light using 9438 inward-facing photomultiplier tubes (PMTs) while the 91 outward looking tubes (OWL) tag cosmic muon events and instrumental background. As seen in figure 2.3, these PMTs were mounted on a spherical PMT Support structure (PSUP) concentric with the AV. Twenty three PMTs are suspended in a rectangular frame facing inwards in the outer H_2O region. These PMTs, along with the 8 PMTs installed in the neck region of the AV, were used to reject instrumental backgrounds. Since neutrino interaction is a relatively rare low-energy process, SNO was designed to be ultraclean of radioactive backgrounds. The limits were set such that the total neutron background from photo-disintegrations is less than 10^{th} of the solar-neutrino rate (5,000 per year) for ν interactions in the fiducial volume of D_2O (R$\leq$550.0 cm). This leads to limits of 3×10^{-15} g/g of Th and 4.5×10^{-14} g/g of U in the D_2O [22].

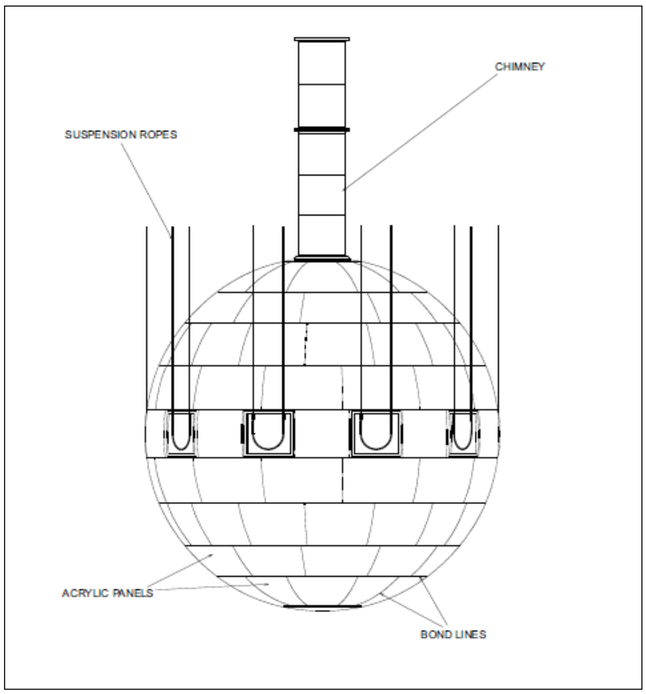


Figure 2.4: The diagram shows acrylic tiles, acrylic belly plates and grooves, ropes, and a chimney on the acrylic vessel (AV).

2.0.1 The Three Interactions

Neutrinos interact with matter via the exchange of $W^{\pm}$ or Z^0 bosons, as shown in the Feynman diagrams of figure 3.2. SNO measured the flux of all neutrinos - F_{ν_x} (where x is e, μ, τ) and the flux of electron neutrinos - F_{ν_e}. The difference between them $(F_{\nu_x} - F_{\nu_e})$ gives the flux of non-electron neutrinos. These fluxes were measured via different ways in which neutrinos interact with the heavy water. When a neutrino interacts with deuterium, electrons can be created which emit a flash of light called Čerenkov radiation which is picked by the PMTs and converted into electronic signals for analysis. An example of a reconstruction of a neutrino interaction is shown in figure 2.5.

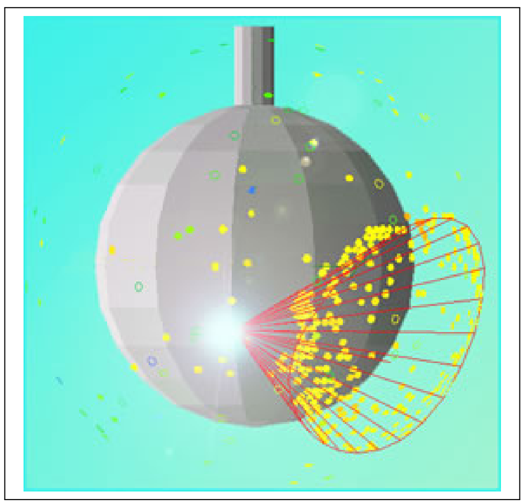


Figure 2.5: A reconstruction of a neutrino interaction, as captured by photomultiplier tubes, is shown here. Figure from [24].

SNO observed ^{8}B ($^8\text{B} \rightarrow {}^8\text{Be}^* + e^+ + \nu_e$) and hep ($^3\text{He} + p \rightarrow ^4 \text{He} + e^+ + \nu_e$) solar neutrinos via these reactions:

Charged Current (CC)

The (CC) reaction, as shown in figure 2.6, is specific to electron neutrinos only. The Q value of CC interaction is -1.4 MeV. In this interaction, the electron carries off most of energy and hence the energy of the electron is strongly correlated with the neutrino energy. A measurement of an energy spectrum of the CC reaction provides a very good sensitivity to spectral distortions produced by neutrino oscillations in the dense matter of the Sun.

In CC interactions, as described in equation (2.1) and shown in figure 3.2a, an electron neutrino interchanges a W boson with a deuterium nucleus thereby converting the neutron in the deuterium into a proton and transmuting itself into an electron[1]. According to the SSM, about 30 CC interactions per day are predicted for SNO in the absence of ν oscillations.

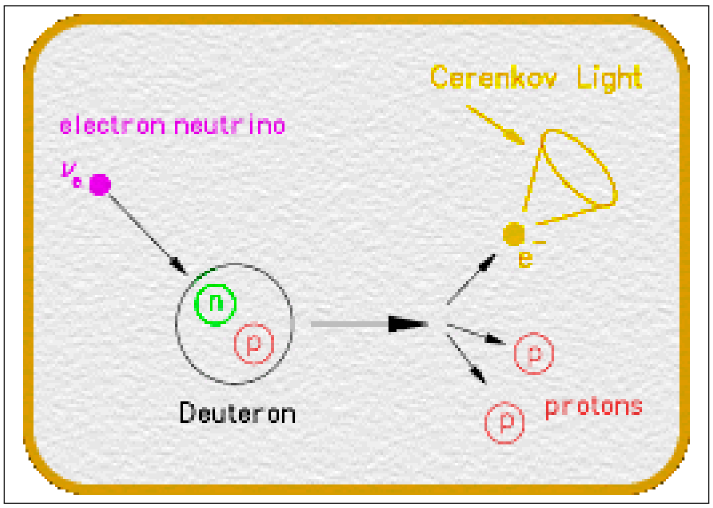


Figure 2.6: Charged current interaction in action. Figure from [25].

$$\nu_e + d \rightarrow p + p + e^- \tag{2.1}$$

Elastic Scattering (ES)

The ES interaction, as shown in figure 2.7 and described in equation (2.2), is sensitive to all flavours of neutrinos but with enhanced sensitivity to ν_e because of the availability of an additional channel to ν_e that of W boson, as shown in figure 3.2. According to SSM about 3 ES events per day are predicted

[1]Solar ν_μ and ν_τ are not energetic enough to interact with a deuterium nucleus producing two protons and a corresponding μ or τ because μ and τ are heavier than an electron and require ν_μ or ν_τ to be more energetic than an ν_e to initiate a CC reaction. Hence CC is only sensitive to ν_e.

for SNO for no ν oscillation.

$$\nu_x + e \rightarrow \nu_x^{'} + e^{'} \tag{2.2}$$

where ν_x refers to any active flavour of ν, e refers to electron and primes on the outgoing particles indicate that energy and momentum has changed by the scattering interaction. In the ES interaction, the electron recoils in roughly

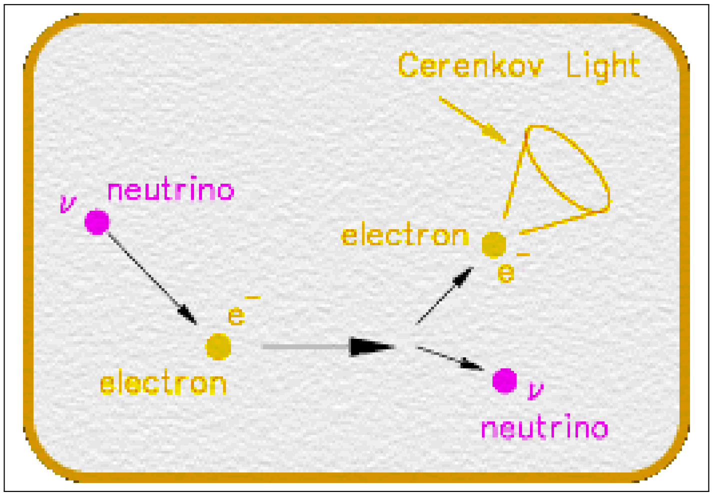


Figure 2.7: Elastic scattering interaction in action. Figure from [25].

the same direction that the ν was travelling therefore the electron "points back" to the Sun. Both SNO and Kamiokande/Super-Kamiokande make use of the electrons in water for the measurements of the solar flux of ^{8}B neutrinos. Therefore, the ES interactions allow a cross-check with the Super-Kamiokande results. The energy and direction of recoil electrons are measured by observing their Čerenkov light with photomultiplier tubes.

Neutral Current (NC)

The NC interaction, mediated by Z^0 boson, allows measurement of the total flux of ^{8}B neutrinos because it is equally sensitive to all three ν flavours described by the standard electroweak model. As seen in figure 2.8 and described in equation (2.3), an incident neutrino breaks up the deuterium (d) into a proton (p) and a free neutron (n). The liberated neutron is then thermalized as it scatters around in the heavy water. Gamma rays are emitted when the neutron is absorbed by another nucleus. The gamma rays scatter electrons with sufficient energy to produce Čerenkov radiation to be detected by the PMTs.

$$\nu_x + d \rightarrow n + p + \nu_x \tag{2.3}$$

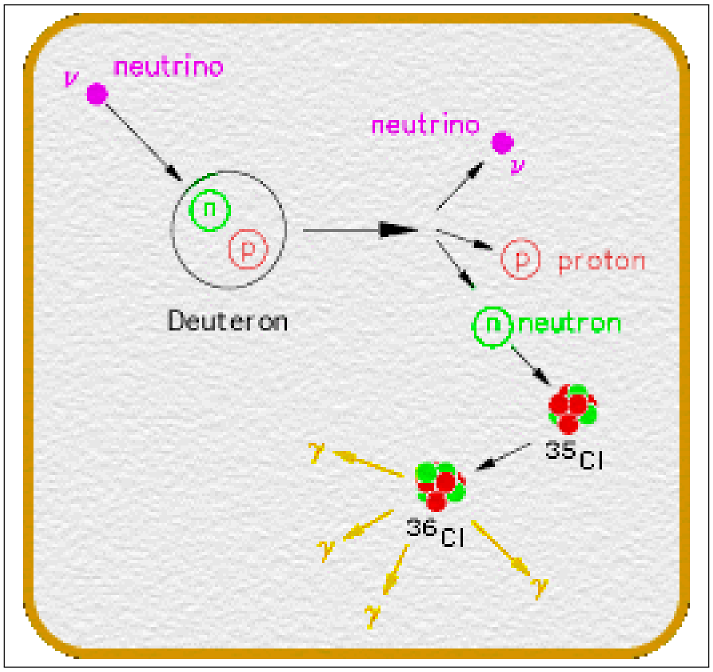


Figure 2.8: Neutral current interaction in action in the salt phase of SNO. The chlorine nucleus (Cl) of NaCl absorbs the neutron and emits a cascade of γ rays. Figure from [25].

The NC signal provides no information about either the energy or direction of the incident ν. The measurement of ^{8}B flux depends on the capture efficiency of neutrons in SNO and the resulting gamma ray cascade. The SSM predicts about 13.2 NC events per day for SNO. Neutral current, an inelastic scattering of neutrinos with deuterium, has a threshold of 2.2 MeV which is the binding energy of the deuterium. It involves liberation and recapturing of a neutron. Three different recapture mechanisms, employed to detect the neutrons, made SNO a three-phase-experiment.

The NC reaction is measured by observing the γ-rays from the subsequent capture of the free neutron in the first two phases, and by direct detection in the third phase. The NC provides the total neutrino flux to explore the solar models, irrespective of neutrino oscillations, since the reaction is equally sensitive to all non-sterile ν types.

SNO is also sensitive to charged current, elastic scattering and neutral current interactions from hep neutrinos from ^{3}He+p $\rightarrow$ ^{4}He + e^+ interaction in the Sun. The interactions induced by hep neutrinos are described in this thesis as hep CC, hep ES and hep NC respectively.

2.0.2 The Three Phases

In the first phase (D_2O phase) of SNO, the detected neutrons captured predominately on a deuterium (cross-section 0.5 millibarn[2]) in the D_2O with a release of a 6.25 MeV photon which imparted enough energy to electrons via Compton scattering or pair production (e^-e^+) to produce Čerenkov light for PMT arrays to detect. Using distributions of the reconstructed energy, position and orientation of the events, the NC was statistically separated from the CC and ES signals. The first phase ran from November 2, 1999 to May 31, 2001. Results, published in [26], proved that neutrinos undergo oscillations in flavour as they journey from the core of the Sun to the Earth. The number of NC events above 5.5 MeV was about a third of the measured number of CC events because the neutron capture efficiency with D_2O alone was only 14%.

In the second phase (Salt phase) of SNO, 2,000 kg of purified NaCl were added to the D_2O. While the salt concentration was only 0.2% by weight, salt enhanced the probability of neutron capture because the 44 barn thermal capture cross-section on ^{35}Cl is 88,000 times larger than the capture cross-section on deuterium resulting in an increase in the sensitivity by a factor of three to detect NC interactions. Another benefit of adding salt is a better separation of NC from CC and ES because absorption of a neutron on ^{35}Cl

[2] A barn is a unit of area equal to 10^{-24} cm^2, used to measure cross sections in physics.

produces a cascade of photons with energy totalling 8.6 MeV as compared to a single γ of energy 6.25 MeV produced when a neutrino interacts with a deuterium. The outcome from the second phase, published in [27], was precise measurements of the parameters that govern neutrino oscillations.

In the third phase of SNO (Neutral Current Detection (NCD) phase), an array of proportional counters called NCDs, was deployed in the heavy water to detect neutrons independent of the PMTs. The NCDs were filled with a mixture consisting of 85% of ^{3}He and 15% CF^4 by pressure. A total of forty strings, laser-welded assemblies of individual counters, were attached to anchor points on a 1 m^2 grid. Out of the forty strings, four contained ^{4}He instead of ^{3}He for assessing the backgrounds.

The NCDs only blocked 9% of the Čerenkov photons. The advantage of NCDs is that over 60% of the detected NC events were recorded separately on an event-by-event basis from the CC and ES signals. The separate readout reduced contribution of NC signal in the Čerenkov data which made possible reduction of the correlation between NC and CC from about -0.5 to better than -0.02. Furthermore, the CC signal in the NCD phase has substantially reduced contamination from neutron capture hence measurement of the neutrino energy spectrum via the CC reaction is made with increased precision [21]. The result of the third phase is described in [29].

2.1 Čerenkov Radiation

Neutrino interactions in SNO were observed by detecting the Čerenkov light emitted by relativistic electrons. For CC and ES, the electrons were the recoil electrons (equations (2.1) and (2.2)) and for NC, the electrons were Compton-scattered by the gamma rays released by the capture of neutron in a nucleus (equation (2.3)).

If a source emits energy, via waves, in all directions then the wave fronts will be spherical as shown in figure 2.9 from [31]. In a non-dispersive medium, the velocity of waves is described as: $v = \lambda\nu$ where λ and ν are wavelength and frequency of the wave respectively. If the source itself is moving such that it nearly keeps pace with its wave fronts ($v_s \approx v = \lambda\nu$) then the wave fronts look different as shown in figure 2.10a. If the source moves ($v_s = \beta c$) faster than the waves ($v = \frac{c}{n}$ where n is the refractive index of a medium) it generates, then all spherical wave fronts bunch along the surface of a cone (as shown in figure 2.10b) which signifies a shock wave and the cone is then referred to as the Mach cone. The surface of the cone is tangent to all the wave fronts with a half angle (from figure 2.9) described as:

$$v_s = \beta c \tag{2.4}$$

$$\cos\theta = \frac{vt}{v_s t} = \frac{c}{n v_s} \tag{2.5}$$

Thus, by simply measuring the Čerenkov cone opening angle, the velocity of the particles may be determined. An electrically charged particle emits electromagnetic waves due to its charge and motion. When a charged particle is travelling through a medium, its electromagnetic field disrupts the local electromagnetic field (EM) by displacing the electrons in the atoms of the medium. Photons are emitted as electrons relax back to the ground state. In normal circumstances, these photons destructively interfere with each other resulting in no radiation. However, when a charged particle travels through a medium at a pace (v) that exceeds the speed of light in the medium (v_t) then it outruns the electromagnetic waves that it emits, thereby creating a shock front which makes it possible for the photons to interfere constructively and intensify the observed radiation. The shock wave is analogous to sonic boom produced by an aircraft travelling faster than the speed of sound in air.

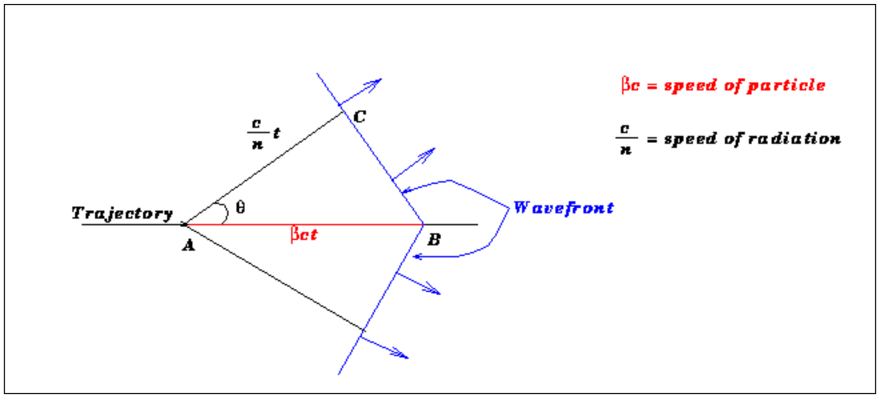


Figure 2.9: Spherical wave fronts surrounding a stationary source. Figure from [31].

Condition for a shock front to occur is:

$$v > v_t = \frac{c}{n} \tag{2.6}$$

$$T_T = (1/2)m{v_t}^2 \tag{2.7}$$

where **n** is a refractive index of the medium, **c** is velocity of light in vacuum, **m** is mass of the charged particle, v_t and T_T are the threshold velocity and threshold kinetic energy respectively. No Čerenkov radiation will be emitted if the kinetic energy of the charged particle drops down to below T_T.

The Čerenkov radiation is utilized in the Čerenkov detectors for detecting fast particles and determining their speeds or making a distinction between particles of different speeds. In D_2O, the angle of the opening cone is 41° for relativistic electrons. The electron direction is not constant due to scattering[3] hence the detected light does not correspond to a single ring pattern as the angle depends on the velocity of the charged particle (equation (2.5)). The electrons will emit Čerenkov light until their kinetic energy drops below the Čerenkov threshold of 0.262 MeV. Table 2.1 lists the index of refraction for various media along with the corresponding kinetic energy threshold for the Čerenkov radiation.

[3]Scattering causes reduction in velocity.

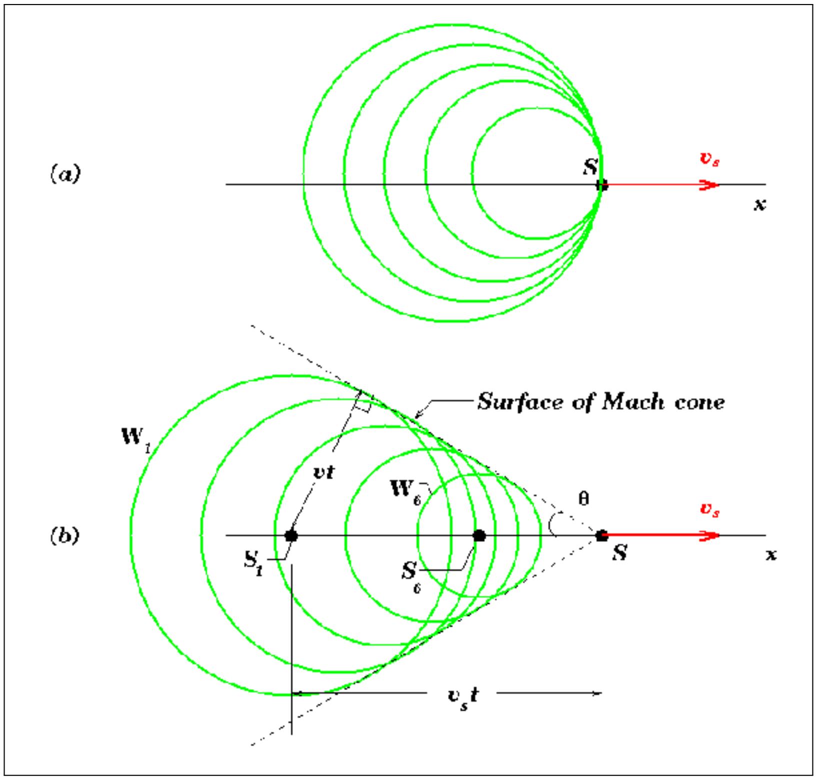


Figure 2.10: This diagram shows wave fronts when the source is moving at a speed comparable to the speed of the waves. (a) $v_s \approx v$ (b) $v_s > v$. Figure from [31].

Medium	**Refractive index n**	T_T **(MeV)**
D_2O	1.333	0.262
H_2O	1.338	0.258
acrylic	1.461	0.190
pyrex	1.474	0.185

Table 2.1: For the relevant materials in SNO, the table lists the refractive index (n) and the corresponding kinetic energy threshold (T_T) for Čerenkov radiation. Table from [32].

The second order differential Čerenkov spectrum is given by equation (2.8).

$$\frac{d^2N_C}{dxd\lambda} = \frac{2\pi\alpha z^2}{\lambda^2}\left(1 - \frac{1}{n^2(\lambda)\beta^2}\right) \tag{2.8}$$

where d^2N_C is the number of photons emitted in a track length equal to dx over a spectral range of $d\lambda$, **z** is a charge of a moving particle in units of electron charge **e**, $\beta \equiv v/c$, $n(\lambda)$ is the refractive index as a function of wavelength

(λ), and $\alpha \approx 1/137$ is the fine structure constant. The number of photons N_C emitted by an electron is approximately proportional to the electron's track length and hence its energy. It is ≈358 photons per cm [33] in the spectral range of 300 to 650 nm to which the PMTs are sensitive. The electron track length in D_2O is approximately 0.45 cm per MeV for electrons of kinetic energy between 5 and 15 MeV. Thus about 1140 photons are produced by a 7 MeV electron [34].

2.2 PMTs

Photomultiplier tubes, shown in figure 2.11, are extremely sensitive detectors of light. These detectors enable individual photons to be detected by multiplying the current produced by the photons by as much as 100 million times in multiple dynode stages. Three characteristics of PMTs are the transit time, the rise time and the transit time spread. The transit time is a time interval between arrival of a photon at the cathode and arrival of an amplified pulse at the anode. The rise time is a time required for a PMT anode signal to rise from 10% to 90% of the final charge collected. The transit time spread is due to different paths that the electron can take from the photocathode to the anode [37].

In SNO, the PMTs are eyes of the detector and upon sensing a single photon produce an electrical pulse that travels to the data acquisition electronics. The radioactivity levels of PMTs have to adhere to strict specifications of allowed maximum radioactivity levels. The measured concentration of uranium was less than 120 nanograms per gram, the thorium concentration was 90.0 nanograms per gram and the potassium concentration was 0.2 milligram per gram in the glass. Besides low-radioactivity levels, the constraints on the PMTs are low failure rate (since they can not be replaced), a high photon detection efficiency, a low noise rate and a narrow spread in the transit time.

The energy and position resolution largely depends on the spread in the transit time, the photon detection efficiency and the noise rate. The photocathode coverage with the PMTs alone is 31% hence each PMT (20 cm in diameter) is surrounded by a 27 cm diameter light concentrator (wavelength shifter) which increases the overall detector light collection to 54% of 4π. This configuration is shown in figure 2.13.

The sensitivity of a photocathode in a PMT is expressed as quantum efficiency (QE) which is simply defined as:

$$QE = \frac{\text{number of photoelectrons emitted}}{\text{number of incident photons}} \tag{2.9}$$

Quantum efficiency is a function of the wavelength or quantum energy of the incident photon. In terms of quantum efficiency, as shown in figure 2.12, PMTs in SNO exhibited a peak quantum efficiency of $\approx 21\%$ at 450 nm.

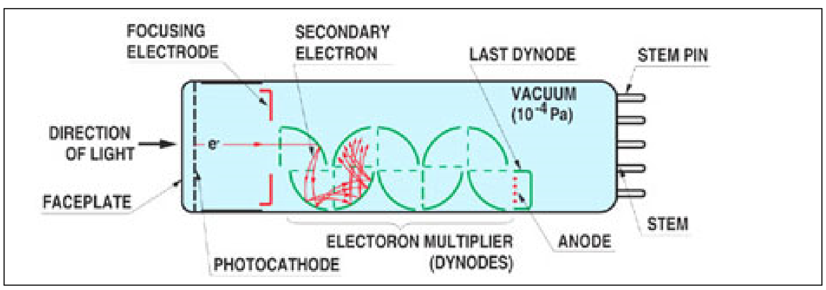


Figure 2.11: A diagram of a Photomultiplier Tube (Not SNO's). Figure from [35].

The PMTs in SNO detect the Čerenkov light emitted by relativistic electrons produced directly or indirectly in neutrino interaction. Because the light is emitted in a cone shape (figure 2.5), a characteristic ring-like pattern of activity is seen on the array of PMTs. The ring pattern is useful to infer direction, energy, and flavour information of the incident neutrino. A ring pattern with fuzzy and blurry edges, due to multiple Coulomb scattering, is characteristics of electron while a ring pattern with sharp edges indicates a muon.

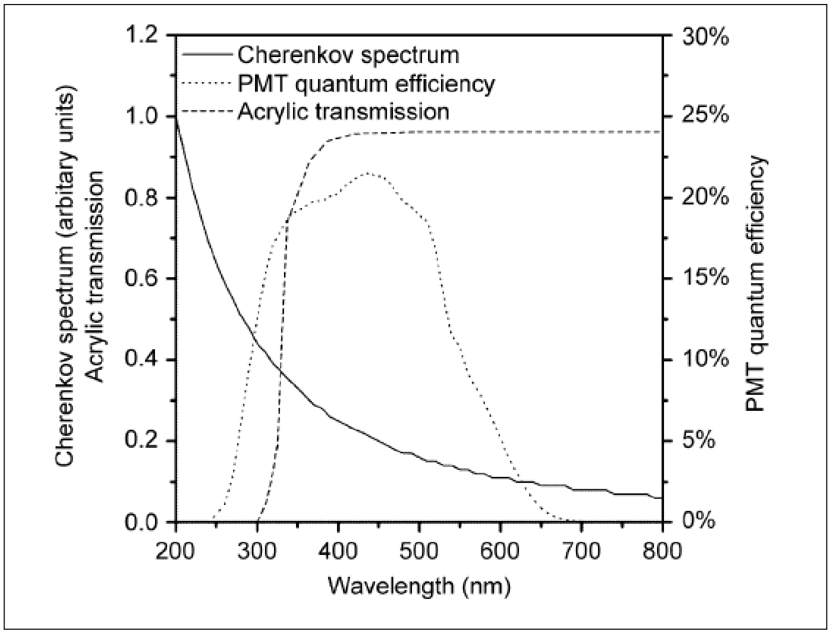


Figure 2.12: Figure shows the transmission of the SNO acrylic vessel and PMT quantum efficiency as a function of wavelength superimposed on the Čerenkov spectrum (in arbitrary units). Figure from [36].

2.3 The NCD Phase

This section describes the Neutral Current Detectors (NCDs) of the NCD phase. An NCD is a proportional counter to detect neutrons via a ^{3}He(n,p)^{3}H interaction which has a Q value of 0.764 MeV. The absorption of a thermal neutron in ^{3}He makes it unstable which then decays into a proton and a ^{3}H (tritium). To conserve energy and momentum, the end products are always emitted back to back; with proton carrying 0.573 MeV of kinetic energy and the tritium having 0.191 MeV [37]. The charged proton and ^{3}H ionize the gas inside the NCD, creating around 20,000 electron-ion pairs. The electrons are accelerated by a high voltage at the anode of the proportional counter. The accelerated electrons produce secondary ionization with sufficient energy to produce an avalanche. The movement of the electron-ion pairs in the NCD induces an electrical signal on the anode which is directly proportional to the energy of the original ionizing particles.

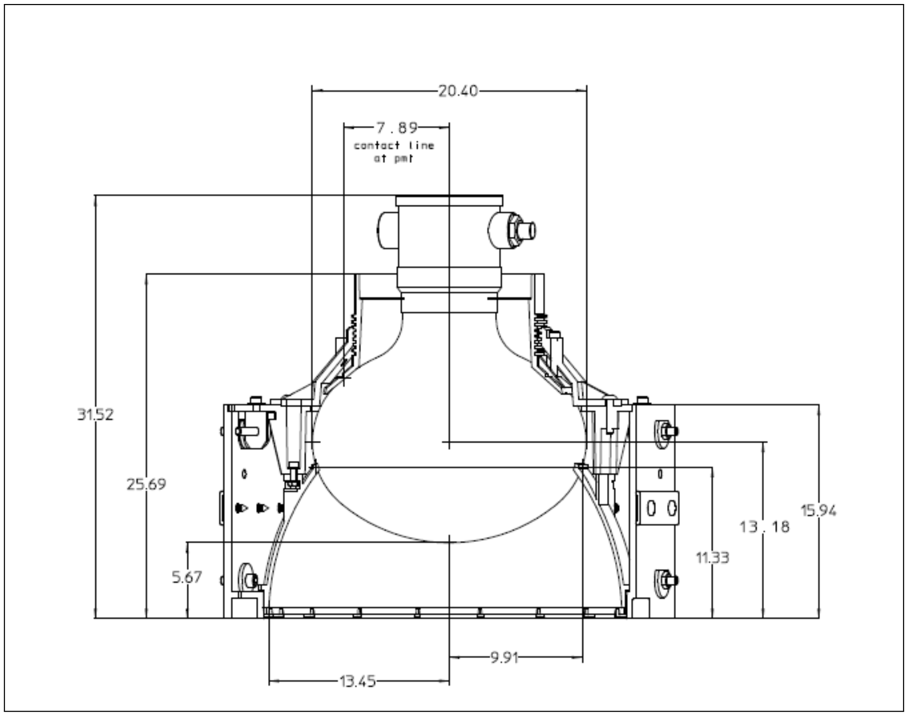


Figure 2.13: A schematic of the Hamamatsu R1408 Photomultiplier Tube along with a reflector assembly used in SNO.

The capture cross-section for thermal neutrons on ^{3}He is 5330 barns, about seven orders of magnitude larger than the capture cross section on deuterium (^{2}H). Compared to thermal neutrons, ^{3}He has negligible sensitivity to gamma rays which makes it an effective neutron detector material. The 36 NCD strings provide a neutron capture efficiency of 26%, giving a neutral current signal of about 3.3 events per day in the NCDs.

The neutron capture signature in the NCD array is read out through a separate data acquisition system from the Čerenkov light signal observed with the PMTs. Since the majority of NC events in the NCD phase are measured separately in the NCDs, the statistical correlation between the NC signal to the light collected in the PMTs is reduced compared to previous phases. Additionally, the NC flux measured in the NCDs is used to calibrate the NC contribution to the PMT signal which is then subtracted from the CC and ES signals, thus improving the CC/NC ratio which constrains the solar neutrino mixing angle.

2.4 Software

SNO Monte Carlo and ANalysis (SNOMAN) software, a package of FORTRAN routines, was used for data analysis and Monte Carlo simulation of the detector. The SNOMAN code models all significant detector geometries, such as the acrylic vessel, the acrylic tiles, the acrylic belly plates and grooves, the Kevlar ropes, the neck, the PSUP and the source container when a source was deployed. The simulation of a large Čerenkov detector is a very complex process. Simulation starts with an electron or neutron of a given momentum at a given location in the detector. The electron is tracked as it slows down. The tracking of the electron and the Čerenkov light produced is calculated using the proven electromagnetic shower code ESG4 [38]. The Čerenkov photons are then transported through the D_2O, acrylic vessel and H_2O, to the PMT sphere (figure 2.3 for a layout of SNO detector). The code takes into account scattering and absorption of photons in each medium while moving towards the PMTs. Including the calibration of PMTs, the simulation constructs an event in which each fired PMT (a PMT hit) is identified, and the time and charge of the PMT is recorded. The simulation also takes into account the PMT noise.

The purpose of reconstruction is to determine the event observables from the pattern of hits and the timings of the PMTs fired. The basic principle is to search for a point source from the hit pattern which would give the correct time of arrival of the photons at different PMT locations. This is complicated by the existence of PMT noise and delay of photons due to scattering en route to the PMTs. The large size of the PMTs and their timing response set a limit to the accuracy of the reconstruction carried out. Once the vertex of an event is known, the initial electron direction is estimated by computing the vector from the vertex to the centroid of the PMT hit locations. The accuracy of the estimation is limited mainly by multiple scattering of electrons.

2.4.1 Response to γ rays

Through Compton scattering and pair production, γ rays generate a shower of charged particles, which in turn, produce Čerenkov light. Compared to a single electron having the same energy, the total light produced by a γ ray is less because each charged particle ceases to produce light as soon as it drops below the Čerenkov threshold energy. In SNOMAN, the γ ray induced showers are calculated using the program EGS4 and the resulting Čerenkov light is computed as in the single electron case.

2.5 Generating an Event Trigger

The number of PMTs hit for a given event, defined as NHITs, is a function of the energy of the event. Low NHIT events, corresponding to low-energy events, are dominated by low-energy background events. To reduce the background events in the data, a simple hardware trigger is employed according to which only events having 13 or more hits are recorded.

For the analysis, the software trigger is set to 20 hits for an event which corresponds to $\approx$ 3 MeV. The diameter of the PSUP is 18 metres hence the time it takes photons, from a single event, to reach different PMTs can vary by as much as 66 nanoseconds (or more due to multiple reflections). The time window, for the primary trigger, was set to be 100 nanoseconds. A global trigger (GT) is initiated if 17 or more PMTs are hit (NHITs$\geq$ 17) within a 100 nanoseconds window. The hardware threshold can be adjusted, by the software, to be higher than 13 hits. The time of trigger is recorded by a 10 MHz and 50 MHz clock. For each GT, a global trigger identification number (GTID), the time the GT is generated, the identification number of each PMT fired and the digitized charge collected are stored. From all this information, the event position, the event energy, the direction of the event relative to the Sun *etc.*are extracted.

2.6 Calibration

Calibration of electronics was vital to maintain the accuracy of the data consisting of charge and timing information from the individual PMTs. Electronics calibration of SNO is covered in detail in [21]. In order to understand the response of the detector to different event types as a function of both energy and position (within the detector volume), detector calibrations were performed. The detector calibrations include: global light collection efficiency, the angular response of the PMTs, the optical attenuation lengths, the energy response of the detector (as a function of both energy and position within the detector) and acceptance of background events. A variety of optical and calibration sources were deployed in the detector. All calibration sources for the heavy water were deployed through the neck of the AV. Sources, intended for the light water between the AV and the PSUP, were deployed using guide tubes that are accessible from the deck above the detector. The calibration source manipulator system, designed and constructed at Queen's University, is a rope-and-pulley system that moves calibration sources throughout the 12 metres diameter detector with approximately 5 cm accuracy. A series of calibrations were performed to take the signals from the PMTs and transform them in terms of event energy, position and direction for further analysis to extract neutrino properties. The calibration sources used in SNO are: laser ball, ^{16}N, ^{252}Cf, ^{8}Li, AmBe, and last but not least sources constructed from ^{226}Ra and ^{232}U to model low-energy backgrounds from ^{238}U and ^{232}Th. Primary calibration sources, deployed in SNO, are outlined in figure 2.14. Calibration of SNO is described in detail in [21].

Calibration source	Details	Calibration
Pulsed nitrogen laser	337, 369, 385, 420, 505, 619 nm	Optical & timing calibration
^{16}N	6.13-MeV γ rays	Energy & reconstruction
^{8}Li	β spectrum	Energy & reconstruction
^{252}Cf	Neutrons	Neutron response
Am-Be	Neutrons	Neutron response
$^{3}H(p,\gamma)^{4}He$ ("pT")	19.8-MeV γ rays	Energy linearity
U, Th	$\beta - \gamma$	Backgrounds
^{88}Y	$\beta - \gamma$	Backgrounds
Dissolved Rn spike	$\beta - \gamma$	Backgrounds
In situ ^{24}Na activation	$\beta - \gamma$	Backgrounds

Figure 2.14: Primary calibration sources employed in SNO. Figure from [28].

2.7 Results from the Three Phases

The D_2O phase ran from November 1999 until May 2001 using only D_2O in the target volume. From 306 days of data, the measured CC, ES, and NC fluxes (in terms of $10^6\, cm^{-2}s^{-1}$) are given below:

$$\phi_{ES} = 2.39^{+0.24}_{-0.23}(stat.)^{+0.12}_{-0.12}(syst.) \tag{2.10}$$

$$\phi_{CC} = 1.76^{+0.06}_{-0.05}(stat.)^{+0.09}_{-0.09}(syst.) \tag{2.11}$$

$$\phi_{NC} = 5.09^{+0.44}_{-0.43}(stat.)^{+0.46}_{-0.43}(syst.) \tag{2.12}$$

The result is from [10]. While reporting the result, the first error (stat.) is due to statistics and the next one (syst.) is from the systematic uncertainties. The energy threshold for the first phase was 5.0 MeV. The ES flux is consistent with the precision measurement made by Super-K [40], and the NC flux was consistent with the prediction for ^8B flux in the Standard Solar Model. The fact that the CC flux is less than the NC flux proved the phenomena of ν oscillation, and the additional fact that the CC flux is also less than the ES flux because the ES[4] flux has contributions from μ or τ neutrinos provided a test of consistency. The correspondence between the fluxes and flavours is

[4] The additional interaction due to charged current available only to electron neutrinos, as shown in figure 3.2, makes ES predominately sensitive to ν_e. Equation(2.14) splits ES into its constituents.

listed below:

$$CC = \nu_e \tag{2.13}$$

$$ES = \frac{5}{6}\nu_e + \frac{1}{6}(\nu_\mu + \nu_\tau) \tag{2.14}$$

$$NC = \nu_e + \nu_\mu + \nu_\tau \tag{2.15}$$

Hence, under the assumption of unitarity which relates the NC, CC and ES rates directly and no oscillation between active and sterile neutrinos, a simple change of variables gives:

$$\phi_e = 1.76^{+0.05}_{-0.05}(stat)^{+0.09}_{-0.09}(syst) \tag{2.16}$$

$$\phi_{\mu\tau} = 3.41^{+0.45}_{-0.45}(stat)^{+0.48}_{-0.45}(syst) \tag{2.17}$$

Combining the statistical and systematic uncertainties in quadrature, $\phi_{\mu\tau}$ is $3.41^{+0.66}_{-0.64}$ which means that the null hypothesis is excluded at 5.3σ. The conclusion of the result ($CC < ES < NC =$ ^{8}B flux) published from the first phase is that neutrinos undergo transformation en route from the Sun to the Earth.

In the second phase, 2,000 kg of NaCl was added to the D_2O volume to increase the capture efficiency of neutrons, released in the NC interaction, by a factor of three greater than for pure D_2O. Furthermore, neutron capture in Chlorine resulted in multiple gammas totalling 8.6 MeV of energy while neutron capture in D_2O produces a single gamma with an energy of 6.25 MeV. The diffuse pattern of Čerenkov light from multiple gammas allows a better separation between CC and NC events. The salt phase ran from July 2001 to August 2002, collecting 391 days of data. The energy threshold for this phase was 5.5 MeV. Assuming undistorted CC and ES energy spectra, the measured fluxes (in terms of $10^6\, cm^{-2}\ s^{-1}$) are listed below:

$$\phi_{ES} = 2.34^{+0.23}_{-0.23}(stat.)^{+0.15}_{-0.14}(syst.) \tag{2.18}$$

$$\phi_{CC} = 1.72^{+0.05}_{-0.05}(stat.)^{+0.11}_{-0.11}(syst.) \tag{2.19}$$

$$\phi_{NC} = 4.81^{+0.19}_{-0.19}(stat.)^{+0.28}_{-0.27}(syst.) \tag{2.20}$$

The results from the salt phase [39] are consistent with the results obtained in the D_2O phase.

The final phase of SNO, the NCD phase, had 385.17 live days. The addition of the proportional counters, called Neutral Current Detectors, allows for a measurement of neutron capture that is systematically different from the mechanisms used in the previous two phases. The outcome (in terms of $10^6\, cm^{-2} s^{-1}$) is:

$$\phi_{ES} = 1.77^{+0.24}_{-0.21}(stat.)^{+0.09}_{-0.10}(syst.) \tag{2.21}$$

$$\phi_{CC} = 1.67^{+0.05}_{-0.04}(stat.)^{+0.07}_{-0.08}(syst.) \tag{2.22}$$

$$\phi_{NC} = 5.54^{+0.33}_{-0.31}(stat.)^{+0.36}_{-0.34}(syst.) \tag{2.23}$$

The result is from [29]. The ratio of the number of CC events to the number of NC events is:

$$\frac{\phi_{CC}}{\phi_{NC}} = 0.301 \pm 0.033 \tag{2.24}$$

Chapter 3

Neutrino Oscillation Theory

3.1 Introduction

Neutrinos started as a curiosity of physics but graduated to being a practical tool to unveil some of the hidden mysteries of the universe. This chapter illustrates the usefulness of neutrinos as a probe, for instance, the example of physics beyond the Standard Model is led by neutrino oscillations because the experimental evidence for massive neutrinos in neutrino oscillations is the first clear signal of physics beyond the Standard Model of elementary particles. This chapter relates the phenomenon of neutrino oscillation and goes over the experimental evidence gathered so far to characterize it.

3.2 Neutrinos as a Window to the Universe

Neutrinos are excellent tools to map the universe because they rarely interact with other particles and are not affected by magnetic fields, hence they travel in straight lines. These features are useful to know the source from which the detected neutrinos originated and the processes which produced the neutrinos in the first place. For example, the direction of neutrinos detected at SNO is highly correlated with Sun's direction in the sky and the energy distribution suggest that the neutrinos are from the decay of ^{8}B in the Sun.

As shown in figure 3.1, for massless neutrinos the spin is always opposite the linear momentum (left-handed) whereas the antineutrinos are always right-handed because the spin and linear momentum always point in the same direction[1]. Hence in the Standard Model (SM) one of the intrinsic properties of a neutrino is its negative helicity (left-handedness) therefore the right-handed field is zero. The antineutrino, antimatter partner of the neutrino, has positive helicity (right-handedness), consequently has no left-handed field. Since strong experimental evidences point to neutrinos having non-zero mass and in order to gain mass, a neutrino has to couple with Higgs' field which in turn require both left and right-handed fields so the question arises – why are only left-handed neutrinos detected in experiments?. In SM, a neutrino is distinct from its antimatter partner but now that a neutrino has mass and both left and right-handed field, is neutrino distinct from an antineutrino?. If there is no distinction between neutrinos and antineutrinos, then the conservation of lepton number[2] is not a fundamental law. The observation of neutrinoless double beta decay ($n \rightarrow p + e^- + \bar{\nu}_e$ and then $p + \nu_e \rightarrow n + e^+$) would clearly show that $\nu_i = \bar{\nu}_i$ and that $\mathbf{L}$, the lepton number to distinguish ν from an $\bar{\nu}$, is not conserved. If a neutrino and an antineutrino are same then neutrinos are Majorana particles; a name to honour Ettore Majorana who first proposed the possibility.

The objective of neutrino experiments is to answer these interesting ques-

[1]Place your right hand in front of your face and curl the fingers in the direction of orbital motion of $\bar{\nu}$, shown in the figure 3.1, then the thumb will point to the direction of momentum, as well as, the direction of spin. According to the right-hand rule, if fingers of the right hand describe the sense of rotation then the thumb points in the direction of spin. For a ν, if fingers of the left hand describe a sense of rotation and thumb points to the orientation of motion then using the right-hand rule, the spin direction comes up to be opposite to the direction of motion (thumb). Hence neutrinos are left-handed and antineutrinos are right-handed.

[2]ν and $\bar{\nu}$ are distinguished by the lepton number $\mathbf{L}$ which is +1 for ν and -1 for $\bar{\nu}$. According to SM, the lepton number is conserved in weak interaction. Since ν has no charge, the lepton number is the only indicator to differentiate a ν from an $\bar{\nu}$. The speed of a massive ν would always be less than the speed of light in vacuum, therefore in theory an observer can overtake a left-handed ν and sees a right-handed ν, thereby changing the lepton number from +1 to -1.

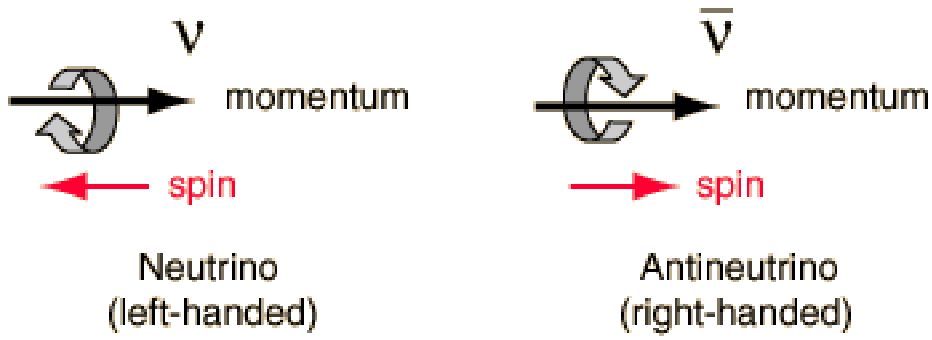


Figure 3.1: The handedness of Neutrinos in a pictorial form. Figure from [41].

tions by studying properties of neutrinos, for example, mass of each flavour, mixing parameters, and magnetic moment. Since neutrinos interact very weakly ($\sigma \approx 2.5 \times 10^{-18}$ barns for E_{ν_e}=10 MeV in ^{2}H [42]) with matter, they are exploited to study the interactions that produced them. For example, due to small interaction cross-sections, neutrinos emerge from the solar core in 2 seconds while it takes heat $\simeq$ 10,000 to 170,000 years to percolate up to the surface from the center [43]. Thus solar neutrinos are windows to an understanding of the inner workings of the Sun. Similarly geo-neutrinos are windows to the core of the Earth. Various models of supernova explosion were tested using neutrinos from Supernova 1987a [45]. If observed, relic neutrinos created during the Big Bang, can open up a new tool to observe the early universe. Neutrinos are also a tool to investigate the weak interaction and the Charge-Parity violation in the weak interaction. There are three generations of neutrinos. If there is a fourth generation of leptons and quarks then neutrinos being of the lightest mass might be the first ones to be discovered. Besides neutrinos, supernovae are predicted to emit gravitational waves (GW) [48] when the core of the star collapses due to gravity. Einstein predicted gravitational waves in his theory of General Relativity (GR) hence the goal of experiments like Laser Interferometer Gravitational Wave (LIGO) and Laser Interferometer Space Antenna (LISA) is to verify that the waves follow the model described in the General Relativity. The SN1987A neutrino data (total 11 events in three experiments), although limited, was enough to confirm the baseline model of gravitational collapse as well as put limits on neutrino mass. A detailed analysis of SN1987A is available at [44].

The direction of an electron scattered by a neutrino ($\nu_\chi + e^- \rightarrow \nu_\chi + e^-$ where χ is e$^-$, μ or τ.) is measured from the Čerenkov light cone, which is then reconstructed to provide direction to the supernova. Combining the neutrino data with the data from the gravitational waves detectors will help eliminate the backgrounds in the gravitational waves experiments.

Neutrinos are studied extensively because they might lead to physics beyond the Standard Model. Observation of oscillation among neutrino species is to-date the only concrete experimental evidence of a new physics.

3.3 Weak Interaction

Neutrino interactions are dominated by the weak force. Gravity is the only other known force to interact with neutrinos but its effects are insignificant in comparison to the weak interaction. Besides interacting with neutrinos, weak force is also the only force capable of changing flavour of one quark into another. It also has the distinction of being the only interaction[3] known to be mediated by massive gauge bosons ($W^\pm$ and Z^0). Since weak interaction acts on left-handed particles (right-handed anti-particles) it violates parity symmetry maximally. The Charge-Parity (CP) violation by the weak interaction is a not as strong effect. The examples of vertices of weak interaction are shown in figure 3.2. Two of the vertices involved a charged boson $W^\pm$, hence they are called Charged-Current (CC) interaction. The third one is called a Neutral Current (NC) because it involves a Z^0 boson.

[3] Mass of gluons is assumed to be zero and [47] analyse upper limits on a possible gluon mass.

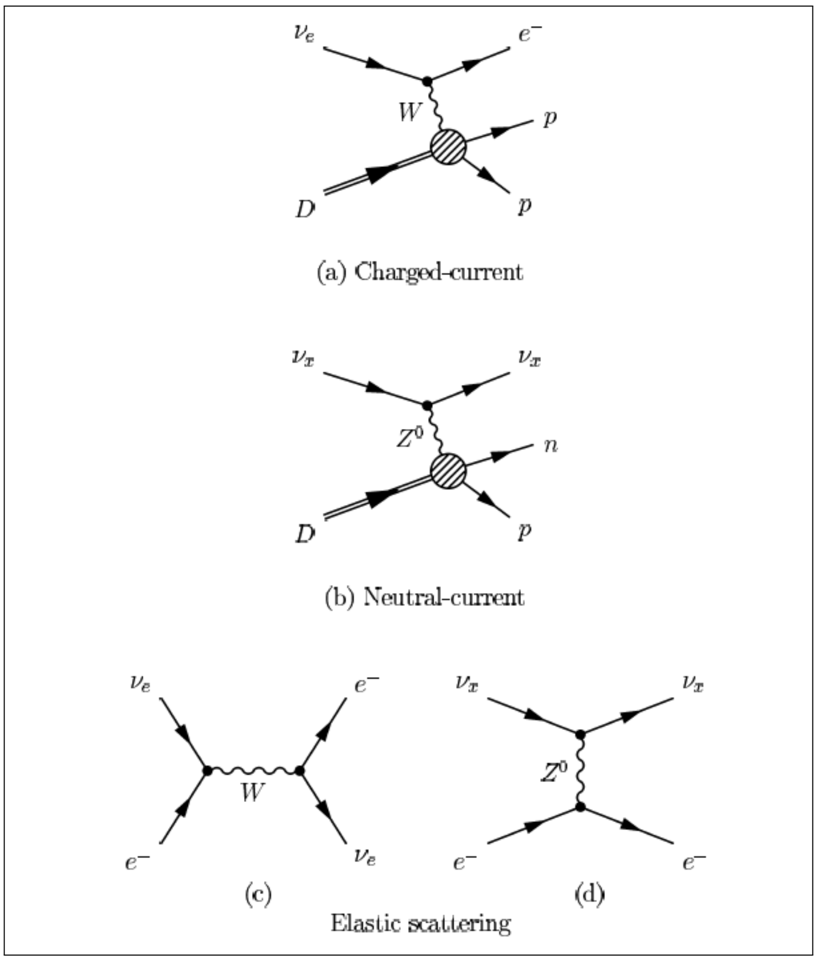


Figure 3.2: Charged current, neutral current and elastic scattering interactions. For solar neutrinos, only ν_e interact with electrons via $W^{\pm}$.

3.4 Neutrino Oscillations

Neutrinos undergo flavour change along their journey from the core of the Sun to the Earth, transforming from one flavour to another (ν_e to ν_μ or ν_τ), thus evading detection by instruments designed to detect only ν_e. The idea of ν oscillation was introduced by physicist Bruno Pontecorvo [57] in 1957 and developed further to include oscillations in matter by Wolfenstein [58] and then by Mikheyev and Smirnov [51]. Since Homestake, several experiments using atmospheric, solar, accelerator and reactor neutrinos have confirmed observing a deficit/disappearance in the neutrino flux but in 2001 Solar Neutrino Observatory conclusively proved that neutrino oscillation is the cause of the deficit by providing a clear evidence of neutrino flavour change [26]. For an early history on neutrino oscillations, refer to [49].

Neutrinos are created and observed through weak interaction as flavour eigenstates (ν_e, ν_μ, ν_τ) but propagate as a linear mixture of mass eigenstates (ν_1, ν_2, ν_3). While propagating, a neutrino is in a superposition of three flavour eigenstates,

$$|\nu\rangle = A_e|\nu_e\rangle + A_\mu|\nu_\nu\rangle + A_\tau|\nu_\tau\rangle \tag{3.1}$$

The flavour eigenstates and mass eigenstates are related by a mixing matrix $U_{\alpha i}$, known as PMNS (PontecorvoŰMakiŰNakagawaŰSakata matrix) matrix, where the index α denotes the flavour state and i the mass state. The mixing angles θ_{ij} gives the relationship between the flavour state i and the mass state j.

$$|\nu_\alpha\rangle = \sum U_{\alpha i}|\nu_i\rangle \tag{3.2}$$

The mixing matrix is parametrized as:

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}\,e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}\,e^{i\delta} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

where θ_{ij} is the mixing angle, $c_{ij} = \cos\theta_{ij}$, $s_{ij} = \sin\theta_{ij}$ and δ is the Charge Parity (CP) violating phase. The three mixing angles (θ_{ij} - ij = 12, 13, 23), and one complex phase angle δ are four parameters determining the amount of mixing. Since θ_{13} is less than 13° [59], the central matrix is reduced to the identity matrix and and $\Delta m^2_{23} \gg m^2_{12}$ makes it possible for mixing in the remaining sectors, θ_{23} and θ_{12}, to be approximated by a two flavour oscillation. This approximation reduced the complication of neutrino mixing from three sectors (ij = 12, 13, 23) to one sector (ij=12). For the two neutrino case, the transformation matrix U is expressed as:

$$U = \begin{pmatrix} \cos\,\theta & \sin\,\theta \\ -\sin\,\theta & \cos\,\theta \end{pmatrix} \tag{3.3}$$

$$\nu_e = \cos\theta\,\nu_1 + \sin\theta\,\nu_2 \tag{3.4}$$

$$\nu_\chi = -\sin\theta\,\nu_1 + \cos\theta\,\nu_2 \tag{3.5}$$

where ν_χ is an admixture of ν_μ and ν_τ flavour states – $\nu_\chi = \frac{1}{\sqrt{2}}(\nu_\mu - \nu_\tau)$, ν_1 and ν_2 are mass eigenstates with masses m_1 and m_2 and θ is the mixing angle. The change with time (assuming two state system) for the electron neutrino is:

$$|\nu_e(t)\rangle = |\nu_1\rangle e^{-iE_1t}\cos\,\theta + |\nu_2\rangle e^{-iE_2t}\sin\,\theta \neq \nu_e \tag{3.6}$$

Due to this change, there is a probability that at time $t{\neq}0$ an electron neutrino will not be detected as an electron neutrino because it is no longer an electron neutrino. The two states ν_1 and ν_2 propagate independently at different speeds owing to their different masses, m_1 and m_2. The difference in relative phase over time causes a periodic modification of the interference between the two states resulting in a finite possibility that a neutrino created as an ν_e will be observed as ν_μ. Using natural units ($\bar{h} = c = 1$), the survival probability of an ν_e, with energy E, to be detected as an ν_e or ν_χ after travelling a finite distance L is calculated from the above matrix to be:

$$P_{\nu_e \to \nu_e} = 1 - \sin^2(2\theta)\sin^2\left(\frac{\Delta m^2\, L}{4E}\right) \tag{3.7}$$

$$P_{\nu_e \to \nu_\chi} = \sin^2(2\theta)\sin^2\left(\frac{\Delta m^2\, L}{4E}\right) \tag{3.8}$$

$$\Delta m^2 = m_2^2 - m_1^2 \tag{3.9}$$

where Δm^2 is the mass splitting between the first and second neutrino mass eigenstates. Expressing Δm^2 in eV^2, L in km and E in GeV, equations (3.7) and (3.8) are expressed as:

$$P_{\nu_e \to \nu_e} = 1 - \sin^2(2\theta)\sin^2\left(1.27\frac{\Delta m^2(eV^2)\, L(km)}{E(GeV)}\right)$$

$$P_{\nu_e \to \nu_\chi} = \sin^2(2\theta)\sin^2\left(1.27\frac{\Delta m^2(eV^2)\, L(km)}{E(GeV)}\right)$$

The conversion probability(survival probability) that an ν_e will appear as ν_e or ν_μ/ν_τ is dependent on energy as well as oscillation parameters Δm^2 and $\sin^2(2\theta)$. The frequency of the oscillation is controlled by Δm^2 while the amplitude, determined by measuring the difference between the total solar neutrino flux (^{8}B flux) and CC, is controlled by $\sin^2(2\theta)$. No flavour transformation takes place if all neutrino flavours have equal mass ($m_1 = m_2$) or zero mass ($m_1 = 0, m_2 = 0$). In order for oscillation to occur, at least one neutrino flavour has to possess non-zero mass. The vacuum oscillation (L_v) is a distance over which an ν_e after oscillation is detected as an ν_e. Hence, L_v, from the Equation (3.7) is

$$\sin^2\left(\frac{\Delta m^2 L_v}{4E}\right) = 0 \tag{3.10}$$

$$\left(\frac{\Delta m^2 L_v}{4E}\right) = \pi \tag{3.11}$$

$$L_v = \frac{4\pi E}{\Delta m^2} \tag{3.12}$$

Redefining Equation (3.7) in terms of L_v,

$$P_{\nu_e \to \nu_e} = 1 - \sin^2(2\theta)\sin^2(\frac{\pi L}{L_v}) \tag{3.13}$$

$$P_{\nu_e \to \nu_\chi} = \sin^2(2\theta)\sin^2(\frac{\pi L}{L_v}) \tag{3.14}$$

Putting back $\hbar$ and $\mathbf{c}$ in the (3.7) results in:

$$P_{\nu_e \to \nu_e} = 1 - \sin^2(2\theta)\sin^2\left(\frac{\Delta m^2 c^4 L}{4E\hbar c}\right)$$

3.5 Mikheyev-Smirnov-Wolfenstein (MSW) Effect

The presence of matter complicates the simple formalism of neutrinos passing through the vacuum. Neutrinos not only oscillate while propagating in space but also oscillate while interacting with matter which changes its survival probability, a phenomenon known as matter or "MSW" (Mikheyev-Smirnov-Wolfenstein) effect. The neutrinos acquire effective masses from coherent scattering processes in the matter. The coherent scattering of ($\nu_e e^- \to \nu_e e^-$) via $W^\pm$ boson differentiates the electron neutrinos from the other neutrinos, as shown in the elastic scattering diagram in figure 3.2. The MSW propagation equation is:

$$i\frac{d}{dt}\begin{pmatrix}\nu_e \\ \nu_\mu\end{pmatrix} = \frac{1}{2}\begin{pmatrix}\frac{-\Delta m^2}{2E}\cos 2\theta + \sqrt{2}G_F N_e & \frac{\Delta m^2}{2E}\sin^2\theta \\ \frac{\Delta m^2}{2E}\sin^2\theta & \frac{\Delta m^2}{2E}\cos 2\theta - \sqrt{2}G_F N_e\end{pmatrix}\begin{pmatrix}\nu_e \\ \nu_\mu\end{pmatrix} \tag{3.15}$$

where G_F is the Fermi constant and N_e is the density of electrons in the media through which the ν travel. The factor of $\sqrt{2}$ was incorrectly omitted in the original paper by Wolfenstein.

The additional term ($\sqrt{2}G_F N_e$) in the diagonal favours the electron neutrino since the only charged leptons in normal matter are electrons. The additional interaction (term), shown in figure 3.2c, results in a different forward scattering amplitude for electron neutrino relative to the other neutrino types which changes its relative propagation and thereby its flavour superposition. The difference between the potential experienced by the electron ν_e and other flavours χ of neutrino is:

$$\Delta V = V_e - V_\chi = \sqrt{2} G_F N_e \tag{3.16}$$

For neutrinos travelling through matter, Equations (3.4) and (3.5) are rewritten to show the relationship of mass eigenstates (ν_{1m}, ν_{2m} where m stands for matter) in terms of flavour eigenstates (ν_e, ν_χ), and the angles for vacuum and matter oscillations are denoted by θ_ν and θ_m respectively.

$$\nu_{1m} = \cos\theta_m \, \nu_e + \sin\theta_m \, \nu_\chi \tag{3.17}$$

$$\nu_{2m} = -\sin\theta_m \, \nu_e + \cos\theta_m \, \nu_\chi \tag{3.18}$$

where θ_m is the matter mixing angle, defined by:

$$\tan\, 2\theta_m = \frac{\sin 2\theta_\nu}{\cos\, 2\theta_\nu - \sqrt{2} G_F N_e \frac{E}{\Delta m^2}} \tag{3.19}$$

where θ_ν is the mixing angle for the vacuum oscillation. The resonance occurs when

$$\cos\, 2\theta_\nu - \sqrt{2} G_F N_e \frac{E}{\Delta m^2} = 0 \tag{3.20}$$

and $\theta_m = 45°$. At resonance the matrix (Equation (3.15)) becomes degenerate since the diagonal terms are equal. Resonance can occur as neutrinos move from high-density core to lower density further from the core.

Hence, when the resonance conditions are met maximal mixing can occur even for small values of the vacuum mixing angle. The strength of the matter oscillation depends on energy of neutrinos as well as density of the matter through which the neutrinos are travelling. Therefore, neutrinos of different energies can have different degrees of matter oscillation. Wolfenstein's original paper looked at a case of neutrinos passing through a slab of matter of constant density. Mikheyev and Smirnov realized that with varying density the diagonal elements of the Equation (3.15) can become degenerate for certain values of N_e and neutrino energy (various combinations of $N_e E$) instead of one fixed ν energy **E**. Since the diagonal terms have an energy dependence, the suppression of ν_e is a function of energy.

3.5.1 Variable Electron Density

The Sun has a variable electron density, hence N_e is changing as neutrinos propagate through variable density medium. The electron density at the core of the Sun, where neutrinos are created, is at the highest. Hence $\theta_m \simeq \pi/2$ from Equation (3.19); substituting in Equation (3.18) results in electron neutrinos to be in ν_{2m} mass state. As the neutrinos propagate outwards, the density is decreasing. At the surface, $N_e \approx 0$ hence $\sin^2 2\theta_m \simeq \sin^2 2\theta_\nu$ from Equation (3.19), therefore, ν_{2m} mass state consist of both ν_e and ν_χ flavours. At the core ν_{2m} consist of ν_e only and at the surface ν_{2m} is the admixture of ν_e and ν_χ. If the density is changing adiabatically, neutrinos will encounter a layer where the resonance happens resulting in the flavour transformation. Resonant conversion means that the oscillation probability reaches its maximum amplitude and does not depend on the vacuum mixing angle θ_ν.

3.6 Predictions from MSW

The MSW effect leaves three distinct signatures on the observed neutrino spectrum.

1. **Matter Enhanced Oscillation**

   The ratio of the total number of detected neutrinos (NC in SNO) to the electron neutrinos deduced from the CC interaction provides nearly unequivocal evidence for neutrino oscillation. For the null hypothesis (no oscillation) the ratio should be one; from the vacuum oscillation, the minimum ratio should be 0.5 (Refer to Section 3.4). From the final SNO NCD analysis [29], the ratio is:

$$\frac{\phi_{CC}}{\phi_{NC}} = 0.301 \pm 0.033(\text{total}). \tag{3.21}$$

So although the data from the solar experiments indicate oscillation enhanced by matter, MSW theory also makes two additional experimentally testable predictions, neither of which have been observed.

2. **Spectral Distortion**

   Since the survival probability is a function of neutrino energy and the oscillation parameters (Δm^2 and θ), the spectral distortion is a complicated function. The resonance condition is never met for the low energy neutrinos; high energy neutrinos encounter a resonance phenomenon that suppressed electron neutrino flux for a certain values of the vacuum oscillation parameters, hence the ^{8}B neutrino spectral distribution is distorted.

3. **Day-Night Effect**

   The day-night asymmetry (A_{DN}) in neutrino flux is expressed as:

$$A_{DN} \equiv \frac{2(\phi_N - \phi_D)}{(\phi_D + \phi_N)} \tag{3.22}$$

   where ϕ_D and ϕ_N are the day and night ^{8}B flux respectively. The difference in the number of neutrinos detected from above (ϕ_D - neutrinos pass through very little of the Earth's matter) to the number detected from below (ϕ_N - neutrinos travel through large amounts of the Earth) results in the day-night asymmetry. This is caused by $\nu - e$ interactions in the Earth that regenerate the ν_e from ν_μ or ν_τ. The solar oscillation parameters predict fairly small A_{DN} (about 1.5% [52]) hence the prerequisite of measuring the day-night asymmetry is an excellent control of systematic effects to reduce systematic uncertainty and a large sample of data to reduce statistical uncertainty.

3.7 Experimental Evidence for Neutrino Oscillation

The weakness of the weak interaction makes it necessary for a neutrino experiment to have an intense source of neutrinos or a huge detector to detect neutrinos [54]. The source employed dictates the sensitivity of an experiment to the three sectors of neutrino oscillation.

3.7.1 Atmospheric Neutrinos

The collision of cosmic rays with nuclei in the upper atmosphere creates a shower of hadrons, mostly pions which decay to atmospheric neutrinos as shown in the following equations.

$$\pi^- \to \mu^- + \overline{\nu}_\mu \tag{3.23}$$

$$\mu^- \to e^- + \nu_\mu + \overline{\nu}_e \tag{3.24}$$

$$\pi^+ \to \mu^+ + \nu_\mu \tag{3.25}$$

$$\mu^+ \to e^+ + \nu_e + \overline{\nu}_\mu \tag{3.26}$$

At low energies (E_ν<1 GeV) there are approximately two $\nu_\mu + \overline{\nu}_\mu$ produced for each $\nu_e + \overline{\nu}_e$ as a consequence of the above decay sequence. The flavour ratio,

$$\mathbf{R} \equiv \frac{\nu_\mu + \overline{\nu}_\mu}{\nu_e + \overline{\nu}_e} \tag{3.27}$$

is a function of energy. As the energy of neutrinos increases above one GeV, relativistic muons can reach the ground before they decay [53], therefore, the ratio increases. Super-Kamiokande (SuperK) is a water Čerenkov detector which detects electron neutrino and muon neutrino by their interaction with the nuclei of hydrogen and oxygen in the 22.5 kiloton central fiducial mass of water [46]. The neutrino flavour is tagged by detecting and identifying the resulting charged lepton. The direction of Čerenkov ring corresponds to the direction of the outgoing lepton which is correlated to the incoming neutrino. Similarly the amount of Čerenkov light corresponds to the kinetic energy of lepton (electrons or stopped muons[4]) which is correlated with the neutrino energy. The distance travelled by the incoming neutrino is determined by the arrival angle of the incoming neutrino with respect to the overhead point for an observer (zenith). The range of flight distances from 15 km to 13 000 km (15 km for vertically downward-going neutrinos and 13 000 km for vertically-upward neutrinos) and the broad neutrino energy spectrum from sub-GeV to multi-GeV makes the atmospheric neutrinos excellent probes of θ_{23} and Δm^2_{32} sector of neutrino oscillation. Also experiments detecting atmospheric neutrinos are sensitive to oscillations with Δm^2 down to 10^{-4} eV^2 [18]. The result of the search is:

$$\chi \equiv \frac{\nu_\mu}{\nu_e} \tag{3.28}$$

$$\frac{\chi_D}{\chi_P} = 0.63 \pm 0.03(stat) \pm 0.05(syst) \tag{3.29}$$

where χ_D is the ratio of the number of muon neutrino to the number of electron neutrino from the data and χ_P is the ratio from the prediction. Considering the simple kinematics of pion decay, the ratio (**R** from equation (3.27)) is well predicted which entails that the ratio $\frac{\chi_D}{\chi_P}$ is expected to be 1.0. Cosmic rays are randomized by interstellar magnetic field, therefore they arrive at the Earth isotropically. Furthermore, cosmic rays, producing neutrinos with energy above 10 GeV, are not deflected by the Earth's magnetic field;

[4]Stopped muons are "fully contained" events and muons deposit all kinetic energy in the detector whereas "partially contained" event is where a particle exits the fiducial volume depositing only partial energy.

consequently the $\cos\,\theta$ distribution should be symmetric and the number of atmospheric neutrino should be equal for equal bins with $+\cos\ \theta$ and $-\cos\ \theta$. However the result of the experiment was not consistent with this scenario. Besides the ratio ($\mathbf{R}$), the up/down asymmetry (A$\equiv \frac{up-down}{up+down}$) predicted to be zero is greater than 6 standard deviations from the expected [78]. The mixing

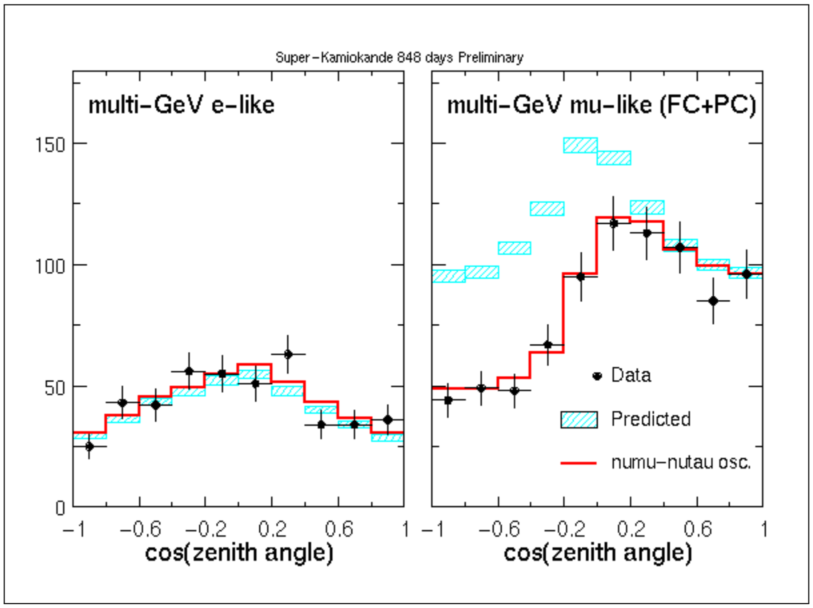


Figure 3.3: Plot shows the $\cos\theta$ distributions for electron neutrinos (ν_e) and muon neutrinos (ν_μ). The $\cos\theta$ distribution for ν_e is symmetric and fits the prediction but the distribution for the ν_μ is not symmetric. The $\cos\,\theta$ distribution for the muon neutrinos does not fit the expected but neutrino oscillation model fits the data. Figure from [77].

angle for atmospheric neutrinos is at its maximum which points to a complete mixing of flavours.

3.8 Solar Neutrinos

In 2001 Sudbury Neutrino Observatory in Canada provided the first direct evidence of solar neutrino oscillation. The result from the extensive statistical analysis was that 35% of the arriving neutrinos are electron neutrinos, the remaining consist of muon neutrinos or tau neutrinos. The total number of neutrinos agreed well with the SSM predictions based on the fusion reactions inside the Sun. The result from SNO also confirmed the interpretation of the anomaly of atmospheric neutrinos in terms of neutrino oscillations. A positive evidence for neutrino oscillations is a prof of non-zero rest mass of the neutrino.

3.8.1 Accelerator Neutrinos

The goal of experiments at accelerators is to perform precise measurement of oscillation parameters (Δm^2_{23} and $\sin^2_{23}$), determine the pattern of neutrino masses and investigate charge-parity (CP) violation in the neutrino sector. The probability of conversion of ν_μ to ν_τ is:

$$P_{\nu_\mu \to \nu_\tau} = \sin^2 2\theta \, \sin^2 \big(\frac{\Delta m^2 L}{4E_\nu}\big) \tag{3.30}$$

For determining the oscillation parameters:

$$\frac{\Delta m^2 c^4 L}{4c\hbar E_\nu} > 1 \tag{3.31}$$

For smaller L/E_ν:

$$\theta \simeq 0$$

$$\sin\theta \simeq \theta$$

$$P_{\nu_\mu \to \nu_\tau} \propto \sin^2 2\theta \, (\Delta m^2)^2 (\frac{L}{E_\nu})^2 \tag{3.32}$$

Table 3.1 lists the neutrino path lengths of various experiments.

Experiment	**L (km)**
K2K (Kek to Kamioka beam)	250
Fermilab to MINOS	730
CHGS (Cern to Gran Sasso)	730
JHF (Japan Hadron Facility)	290

Table 3.1: Path lengths **L** of various experiments. Table from [60].

3.8.2 Reactor Neutrinos

The first neutrinos to be detected by Federick Reines and Clyde Cowan were antineutrinos from a reactor adjacent to the Savannah River. For the detection, the detector uses the inverse beta decay ($\bar{\nu} + p \rightarrow n + e^+$); the prompt photons, emitted when e^+ annihilates with an electron of matter, are followed by a delayed photon when the neutron is absorbed by matter. The coincidence window within which the prompt and the delayed photons are emitted allowed the neutrino interactions to be separated from the backgrounds due to radioactivity and cosmic rays. The most precise measurement of antineutrinos from reactors was achieved by KamLAND. From the KamLAND high precision results, the oscillation pattern in the L/E is shown in figure 3.4.

258 $\bar{\nu}_e \rightarrow \bar{\nu}_e$ events observed.
365.2 ± 23.7 were the expected events without oscillations.
17.8 ± 7.3 expected background events.
disappearance confirmed at 99.998% C.L.
energy spectrum shows distortions with 99.6% C.L.
best-fit of KamLAND: $\Delta m^2 = 7.9^{+0.6}_{-0.5} \times 10^{-5}\, eV^2$
best-fit of KamLAND + solar data: $\Delta m^2 = 7.9^{+0.6}_{-0.5} \times 10^{-5}\, eV^2$, $\tan^2\theta = 0.40^{+0.16}_{-0.07}$

Table 3.2: KamLAND result on 1 November 2004. Data from [20]

3.8.3 Oscillation Parameters - θ_{12} and Δm^2_{21}

All of the solar neutrino results were combined to obtain the best estimate for the solar neutrino mixing parameters. The allowed regions in Δm^2 and $\tan^2\theta$ from χ^2 fit to data from all three phases of SNO is shown in figure 3.5a. From the global analysis of all solar neutrino data and the 2881 ton-year KamLAND reactor antineutrino results, the allowed regions are shown in figures 3.5b and 3.5c. The best-fit point to the Solar global plus KamLAND data

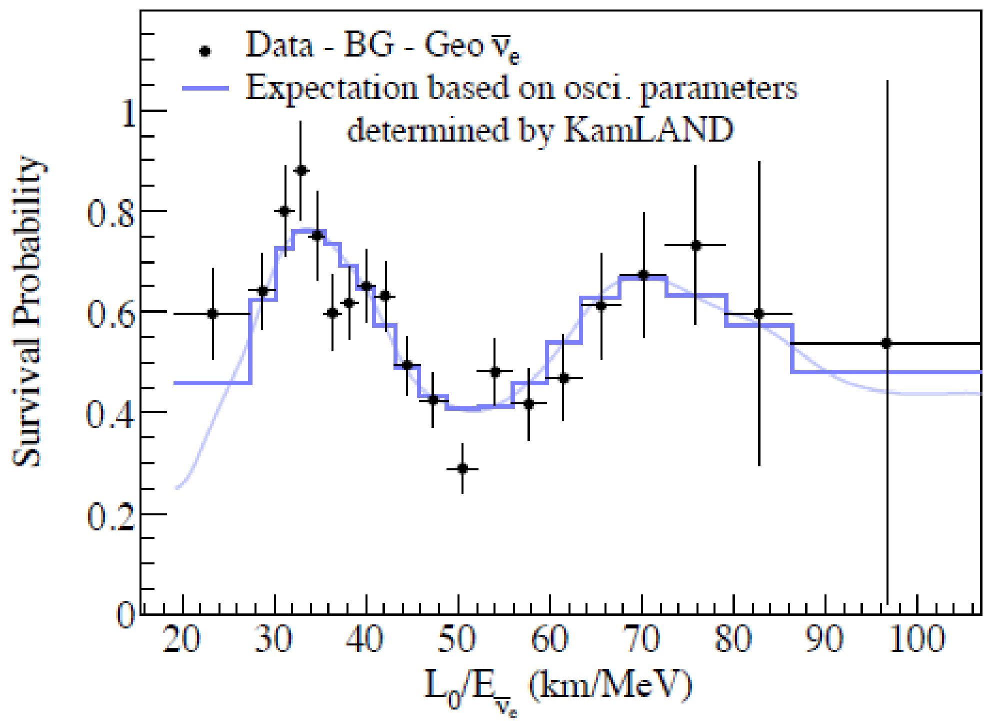


Figure 3.4: KamLAND obtained oscillation parameters from two cycles of L/E. Figure from [20].

yields $\Delta\, m^2 = 7.59^{+0.19}_{-0.24} \times 10^{-5}\, eV^2$ and $\theta = 34.4^{+1.3}_{-1.2}$ degrees, where the errors reflect marginalized 1-σ range. The mixing parameter space strongly favours the large mixing angle (LMA) region[5] and the maximal mixing is ruled out with very high significance – 5.3σ.

[5] The mass square difference Δm^2 ranges between the mass eigenstates from about 3 to 9 $\times 10^{-5}\, eV^2$ while the mixing angle θ is in the range of $\tan^2\theta \approx 0.25 - 0.65$.

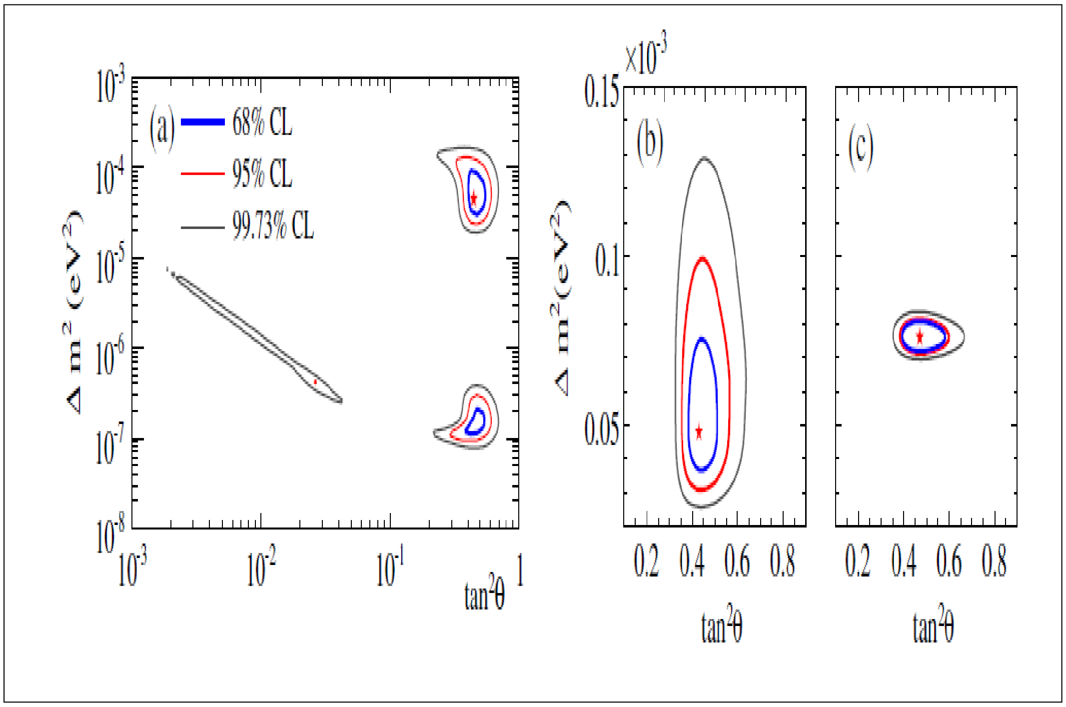


Figure 3.5: Neutrino-oscillation contours. (a) all three phases of SNO. The best-fit point is $\Delta m^2 = 4.57 \times 10^{-5}\,eV^2$, $\tan^2\,\theta = 0.447$, $f_B = 0.900$, with $\chi^2/d.o.f = 73.77/72$. (b) Solar Global: SNO, SK, Cl, Ga, Borexino. The best-fit point is $\Delta m^2 = 4.90 \times 10^{-5}\,eV^2$, $\tan^2\,\theta = 0.437$, $f_B = 0.916$. (c) Solar Global+KamLAND. The best-fit point is $\Delta\,m^2 = 7.59 \times 10^{-5}\,eV^2$, $\tan^2\theta = 0.468$, $f_B = 0.864$. KamLAND constrains Δm^2 and Solar Global constraints $\tan^2\theta$. Figure from [29].

Chapter 4

Signal Extraction Techniques

4.1 Introduction

In SNO data, it is impossible to distinguish CC, ES and NC event-by-event because the signal consist of the Čerenkov light from a recoil or Compton scattered electron. Hence the number of events, belonging to each event type, is estimated using a Markov Chain Monte Carlo fit. MCMC relies on the extended likelihood function to randomly draw samples of the posterior distributions for analysis. This section describes the likelihood function for the combined three phase analysis of SNO data and all the components that go in to it. The likelihood function begins with a list of all possible event types:

1. Charged current electrons – **CC**

2. Elastic scattering electrons – **ES**

3. Neutral current neutrons – **NC**

4. Background from (α, n) reactions on the surface of the acrylic vessel + NCD cables – **EX**

5. Background from atmospheric neutrinos – **Atmos**

6. Background from the internal radioactivity in the D_2O – **$\mathbf{d_2opd}$**

7. Backgrounds from increased radioactivity "hot spots" on NCD string K2 – **k2pd**

8. Background from increased radioactivity "hot spots" on NCD string K5 – **k5pd**

9. Backgrounds from NCDs – **ncdpd**

10. hep CC – CC initiated by hep ν

11. hep ES – ES initiated by hep ν

12. hep NC – NC initiated by hep ν

The observables used to distinguish the signals in the NCD phase, analysed in this thesis, are the event's reconstructed energy (E), reconstructed direction with respect to the Sun (cos $\theta_{\odot}$), and the events's reconstructed position (x,y,z). A radial parameter, ρ, is calculated as $\rho \equiv (\sqrt{x^2+y^2+z^2}/R_{AV})^3$ where $R_{AV} = 600.5$ cm is the radius of the acrylic vessel. The reconstruction technique that generate these observables for each event are described in section 4.8.

4.2 Generation of Probability Distribution Function

The likelihood method requires a Probability Distribution Function (PDF) for each event type in the dataset. The PDFs should be constructed to maximize the use of the available information while minimizing any bias in the fit. The correlations between observables were taken into account by building 3D PDFs for all signals and backgrounds in the fit. For the NCD phase, the three observables were volume-weighted variable ρ, the cosine of scattering angle with respect to the Sun-Earth direction cos θ_{Sun} and effective recoil electron energy (T_{eff}). The 3D PDFs were built, according to the table 4.1, from Monte Carlo generated events. The fiducial volume for the SNO analysis is 550 cm, hence $\rho_{max} \simeq 0.77$. Careful choice of a bin width has an impact on the bias of the signal extraction. The bins should be narrow enough to fully define the shape of the distribution so that no information is lost. However, statistical fluctuation due to too narrow bin can distort the PDF shape and produce a noticeable effect on the fit result.

Observable	**Number of Bins**	**Range**	**Bin Width**
ρ	10	0 to 0.77025	0.0770
$\cos\theta_{Sun}$	25	-1.0 to +1.0	0.08
Energy T_{eff}	13	6 to 20 MeV	0.5 MeV from 6-12 MeV bin 13 has events from 12-20 MeV

Table 4.1: Ranges and binning used for each observable in the 3D PDFs.

4.3 Signals and Backgrounds

The backgrounds in SNO consist of cosmic rays, muons, instrumental backgrounds and natural radioactivity from ^{238}U and ^{232}Th decay chains. The design of SNO, with signals expected in a few tens of events per day, was dictated largely by shielding and radioactivity considerations. All materials used in the construction of SNO were carefully selected to ensure that the neutrino signal was not overwhelmed by the radioactive backgrounds. Going from the outer to the inner regions (Various regions of SNO are shown in figure 2.3), the levels of uranium and thorium are on the order of parts per million for the rock, parts per billion for the PSUP, parts per trillion for the AV and parts per 10^{15} for the D_2O. The first three items (cosmic rays, muons and instrumental backgrounds) can be removed by the low level cuts[1] applied to the data. The radioactive backgrounds, still remaining in the data after the low-level cuts, are classified as external and internal backgrounds: internal backgrounds are those events occurring within the D_2O volume itself and the external events are events occurring outside the D_2O volume, that is, on the AV, in the H_2O and/or the glass of the PMTs. The external backgrounds consist of $\beta - \gamma$ from the PMTs, ^{214}Bi and ^{208}Tl in AV and H_2O. The internal backgrounds are due to ^{214}Bi and ^{208}Tl in the D_2O. Thorium and uranium in the NCD nickel (bulk) and the presence of two areas of increased activity [strings K2 and K5], referred to as hot spots, increased the amount of radioactivity in the detector [65].

The CC, ES, and NC event types, from the list in Section 4.1, are initiated by neutrinos but the EX, Atmospheric and $d_2$opd are due to neutron capture on ^{2}H. The intrinsic radioactivity backgrounds in the detector and from the surrounding rocks in the cavity are due to ^{232}U and ^{232}Th decay chains, described in tables A.2 and A.1. Besides the direct radiations, produced by these radionuclides, alphas and neutrons (byproducts of the radiations) produce indirect radiations by interacting mostly through reactions (α,pγ) and (α,n) with various light elements, in particular ^{29}Si, ^{30}Si, ^{27}Al, ^{26}Mg, and ^{23}Na. The neutrons released in these reactions or those from the spontaneous fission can be captured to produce γ-rays with energies extending up to nearly 10 MeV. These γ-rays can penetrate the detector and are the main source of background emanating from the rock walls of the detector cavity. Backgrounds set a limit on the low energy threshold for the analysis since the finite energy and spatial

[1] Low level cuts are: remove bursts of light which last for microseconds or longer, remove events when a single PMT or small set of adjacent PMTs record a very high charge, remove events when the charge integrate to zero (electronic noise), remove PMT hits without a global trigger (orphans) or where a PMT is hit multiple times (burst event), remove events which also include hits on outward looking PMTs, or on special PMTs installed in the neck, and eliminate events which occur less than 20 seconds after a muon, or 250 milliseconds after a likely atmospheric neutrino.

resolution of the detector will allow a fraction of these events to be inseparable from the solar neutrino events.

4.4 Neutral Current Backgrounds

The neutral current reaction ($\nu_x + d \rightarrow \nu_x + p + n$) with the release of a detectable neutron is unique to SNO and has its corresponding unique backgrounds. Any event mimicking the neutrino disintegration of ^{2}H is a background to NC. The most important source of these background neutrons are photo-disintegration neutrons. A gamma ray (γ-ray) with energy greater or equal to 2.2 MeV can split a deuterium into a proton and a free neutron through a process called photo-disintegration. The background neutrons are not distinguishable from the neutrons released when neutrinos with at least 2.2 MeV split deuterium nuclei into protons and neutrons. There are only two $\beta - \gamma$ decays in the ^{238}U and ^{232}Th chains that can produce sufficiently energetic gammas to contribute to photo-disintegration – ^{214}Bi and ^{208}Tl. The ^{208}Tl nuclei came largely from decays of intrinsic ^{232}Th though the most likely source of ^{214}Bi is from decays of ^{222}Rn entering the detector from mine air and the remaining ^{214}Bi nuclei originated from decays of intrinsic ^{238}U. The decay schemes of ^{214}Bi to ^{214}Po and ^{208}Tl to ^{208}Pb are outlined in figure 4.1.

One source of external neutrons is the (α,n) reaction. During the construction, Radon ^{222}Rn (from the ^{232}U chain) in the mine air came in contact with the acrylic surface. The decay product of radon that decayed within the interior volume of the vessel are carried by electric fields to the surface, where they are deposited and subsequently decay to ^{210}Pb which covers the surface of the detector. This is a problem since ^{210}Pb, with a half-life ($T_{1/2}$) of 22.3 years, is a long-lived source of ^{210}Po. When ^{210}Po decays to ^{206}Pb, it emits a 5.4 MeV α which can interact primarily with light elements (^{2}H,^{13}C,^{17}O and ^{18}O in H_2O, D_2O and acrylic ($C_5H_8O_2$)) emitting a free neutron. The largest

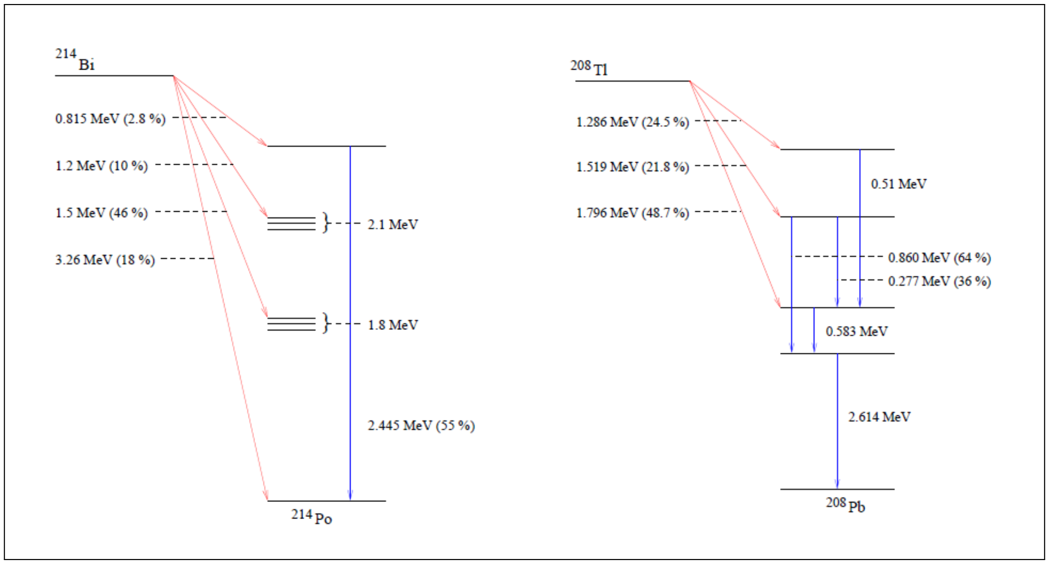


Figure 4.1: Simplified decay schemes for ^{214}Bi and ^{208}Tl. Blue lines (going straight down) represent γ transitions and red lines (slant in direction) represent β transitions. A γ with minimum 2.2 MeV can photo-disintegrate a ^{2}H. Figure from [65].

source of (α,n) reactions was the acrylic. The deployment of NCDs in the D_2O also contributed to (α,n) reactions. Even though the NCDs were electropolished, a small amount of ^{210}Po remained on the surfaces of NCDs which contributed to the total number of neutrons produced by (α,n) reactions. Using an external α counter on a representative sample of the NCDs, the rate of neutron production was determined to be $1.38 \pm 0.28 \times 10^{-2}$ neutrons per day [66]. Another source of external neutrons is the photo-disintegration of deuterium from γ rays that originate outside the D_2O volume.

Neutrons from internal photo-disintegration are not discernible from NC neutrons so the intrinsic radioactivity was measured independently and its neutron production was subtracted from the final NC result. External neutrons, on the other hand, have a distinct radial profile – as shown in blue in figure 4.3 – which enables MCMC to statistically tell them apart from signals and other backgrounds.

4.5 Low-Energy $\beta - \gamma$ decays

Although the energy threshold of 6.0 MeV is above the energies of β particles and γ rays produced in natural radioactive decay, there are three ways in which these events may be reconstructed to resemble a neutrino event in the fiducial volume (i) the combined signals from coincident β particles and γ rays or the random coincidences of two or more decays may exceed the trigger threshold and hence contribute to the data (ii) a neutron, released by a low-energy (at least 2.2 MeV) γ ray induced photo-disintegration of ^{2}H, mimicking an NC event (iii) even though the radioactivity decay rate in $\approx$ 10,000 PMTs is 10^8 to 10^9 decay per day[2], the events rarely reconstruct in the D_2O volume because the PMTs are located far away from the D_2O and the presence of H_2O between PMTs and D_2O shields the volume of interest from the $\beta - \gamma$ decays of the PMTs. However multiple events in the PMTs or from the PMTs and other materials happening in coincidence, within the trigger time window, will occasionally be reconstructed into the D_2O volume. Single β activity in a particular PMT glass is easier to reject because a large signal is confined to that particular PMT.

4.6 Atmospheric Neutrinos, Muons and Muon Followers

The only particles which can penetrate the 2,039 metres of rock overburden and enter the sensitive volume of SNO are neutrinos and high energy muons. Atmospheric neutrinos constitute as background to the solar neutrinos but 1 event per 1 Gigagram per year per 10 MeV bin width is not significant for SNO [67]. Most of the atmospheric ν interactions deposit a large ($\geq$ 20 MeV)

[2]Statistically, with 10^9 decays/day there are events with a high energy but they are rejected because of their large reconstruction error.

amount of energy and multiple charge particles at the interaction vertex[3] but a small number of interactions release single neutrons (the elastic scattering of $\nu_\mu + n \rightarrow \nu_\mu + n$ and resonance production of pions via $\nu_\mu + p \rightarrow \Delta^+ \rightarrow \nu_\mu + \pi^+ n$) or low energy photons without a detectable tag. Furthermore, there is a possibility of a ν interaction creating an excited state of $^{16}O^*$ that de-excites to give photons in 6 MeV range which is a background for the charged current analysis.

Muons, at a rate of approximately 3 per hour in SNO, are also not a significant background since the energy deposited in the detector is much greater than the energy deposited by neutrino induced events. The problems with muons are the spallation[4] products they generate when they interact with hydrogen and oxygen nuclei in the detector. Muon-induced ^{16}O spallation products, such as, ^8B, ^{12}B and ^{12}N are high-energy β emitters (13 to 16 MeV energies) with half-lives between 10 to 800 milliseconds. These spallation events are identified by their characteristic time signature, a high energy muon signal followed by a β decay signal. The spallation products are mostly neutrons which lead to a signal that is indistinguishable from the NC neutrino signal. The leakage of the spallation neutron events into the dataset is less than 0.014 neutrons per day for the D_2O phase compared to $\approx$ 1.3 neutrons/day [68] from the radioactivity in the detector. We include it in the list of other backgrounds, which are small, but which are tallied, along with an uncertainty, and subtracted from the NC signal. When 150 PMTs are hit or/and five outward looking PMTs are also hit, the events are tagged as muon events. Events occurring within 20 seconds of an event, tagged as muon event, are also removed to prevent spallation neutrons (muon followers) from entering the dataset. An additional neutron background cut imposed a 250×10^{-3} dead

[3]These events are cut by a burst cut and the energy cut ($6 \geq E < 20$ MeV)

[4]spallation is a high-energy nuclear reaction in which several nucleons are released from the nucleus of a target atom.

time (in software) following every event in which the total number of PMTs which registered a hit exceeded 60 [10].

4.7 Backgrounds to Charged Current and Elastic Scattering

Energetic electrons due to β decay and Compton-Scattered γ rays ($\gamma + e^- \rightarrow \gamma + e^-$) which reconstruct inside the fiducial volume constitute backgrounds to CC ($\nu_e + d \rightarrow p + p + e^-$) and ES ($\nu_x + e^- \rightarrow \nu_x + e^-$). The sources of β decays, shown in tables A.2 and A.1 and the main sources of γ rays with energy exceeding 2.2 MeV are 2.445 MeV γ-ray from the decay of ^{214}Bi and 2.615 MeV γ-ray from the decay of ^{208}Tl shown in figure 4.1.

Additionally neutrons from NC, from photo-disintegration of ^{2}H and from outside the D_2O vessel can be captured by ^{2}H producing a 6.25 MeV γ-ray. In the salt phase, a neutron capture in chlorine resulted in a cascade of γ rays with a total energy of 8.65 MeV. The gamma rays, from neutron capture, imparts energy via Compton scattering to electrons beyond Čerenkov threshold, thus making them similar to CC or ES electrons. In the NCD phase, a neutron captured in the NCDs did not constitute as a background to the CC and ES interactions resulting in reduced uncertainty on the number of CC events in the NCD phase.

4.8 Observables

Observables are the reconstructed attributes of an event derived from the hit patterns recorded by the PMTs of the detector. For SNO, the observables for statistically separating the signals and backgrounds in the NCD phase are energy (T_{eff}) radial position (ρ) and direction of an event relative to the Sun ($\cos\ \theta_{\odot}$). Distributions shown in figures 4.2 to 4.5 were used to create probability distribution functions (PDFs) for performing an extended maximum likelihood fit to the data using MCMC Metropolis algorithm. These PDFs were generated from Monte Carlo simulations assuming no flavour transformation and the shape of the standard ^{8}B spectrum. The Monte Carlo simulation included a detailed model of the physics of neutrino interactions and radioactive decays within the detector and a meticulous description of the detector geometry.

4.8.1 Energy

The reconstructed kinetic energy of an event, called T_{eff}, is the most probable energy of a single electron that produced the hit pattern of PMTs observed. The CC, ES and $ES_{\mu\tau}$ spectra depend on the shape of incident neutrino spectrum. The observable energy is very important for separating β decays or γ rays from neutron capture from the higher energy neutrino-events which also produce electrons. Figure 4.2 shows the energy distribution of NC, CC, EX, K5PD and NCDPD along with a line at the 6.0 MeV energy threshold.

4.8.2 Vertex of Event

One of the observable is a vertex (x,y,z) where the event occurred. For convenience, when making the PDFs, instead of using the vertex, the normalized cubic radius ($\rho = (\frac{R}{R_{AV}})^3, R_{AV} = 600$ cm) is used because of the spherical shape of the detector (figure 2.1). The radial distance (R) is calculated as:

$$R = \sqrt{(x^2 + y^2 + z^2)} \tag{4.1}$$

In terms of ρ, an event on the surface of AV will have ρ =1 and an event at the center will have ρ =0 , and events distributed uniformly throughout the detector will have a flat distribution in ρ. Figure 4.3 illustrate radial distributions of NC, CC, EX and NCDPD along with fiducial volume cut at $\rho < 0.77025$.

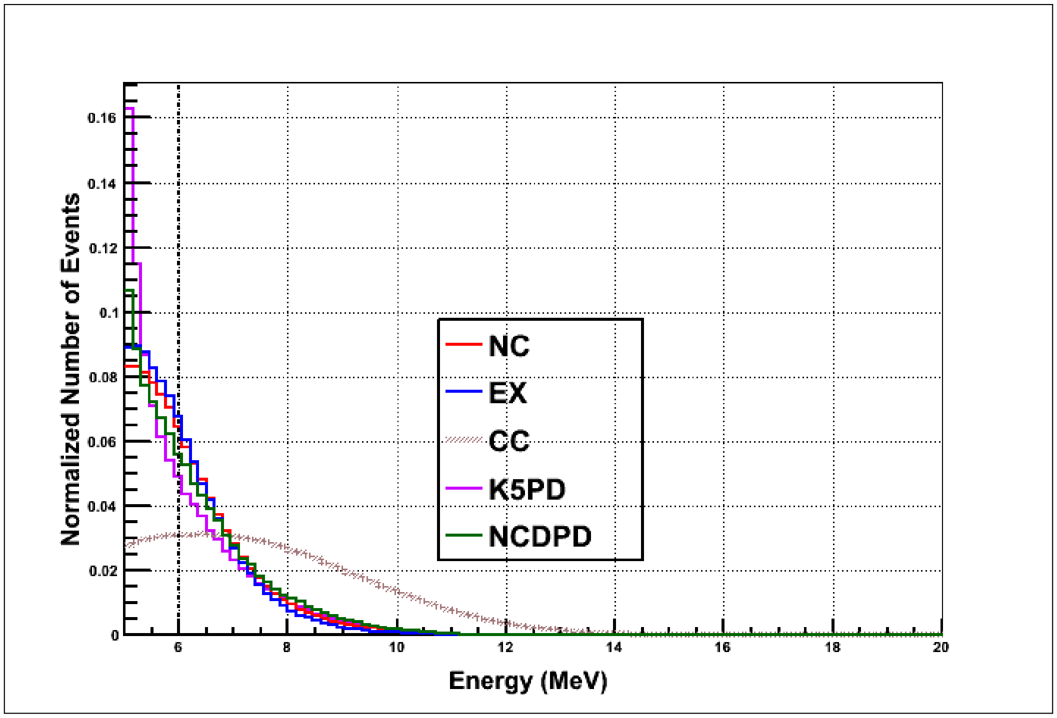


Figure 4.2: Distribution of energy for for NC, CC, EX, K5PD and NCDPD in the energy window of 5-20 MeV. For this plot fiducial volume cut is applied at $\rho < 0.77025$.

Radial position is important in separating external events from the internal events; compare the radial distribution of external neutrons in blue to NCDPD neutrons in green in figure 4.3. This shows that ρ is a very useful tool in statistically separating external from the internal backgrounds, for instance NCDPD. As a neutron thermalizes, it wanders off and since hydrogen in the acrylic ($C_5H_8O_2$) and light water (H_2O) is an efficient neutron sink, the radial distribution of NC (in red) is not as flat as the radial distribution of CC (in cyan) in figure 4.3. Hence radial distribution also provides a weak handle on NC events. Although the D_2O volume extends to R=600 cm, background events from the rest of the detector leak into this volume. By defining a fiducial volume of 550 cm, a large number of these events were rejected.

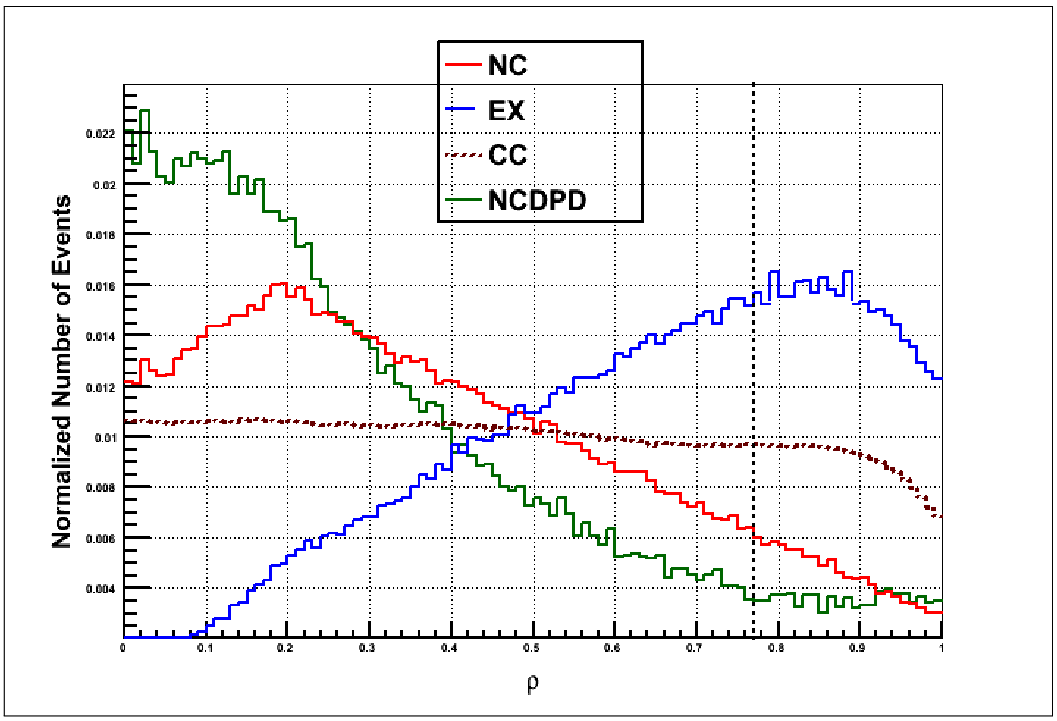


Figure 4.3: The normalized radial distribution for the NC, CC, and EX in the energy window of 6-20 MeV. The power of ρ to separate the external events from the internal events is very clear as external backgrounds have a steep distribution in ρ. Energy range for this plot is 6 to 20 MeV.

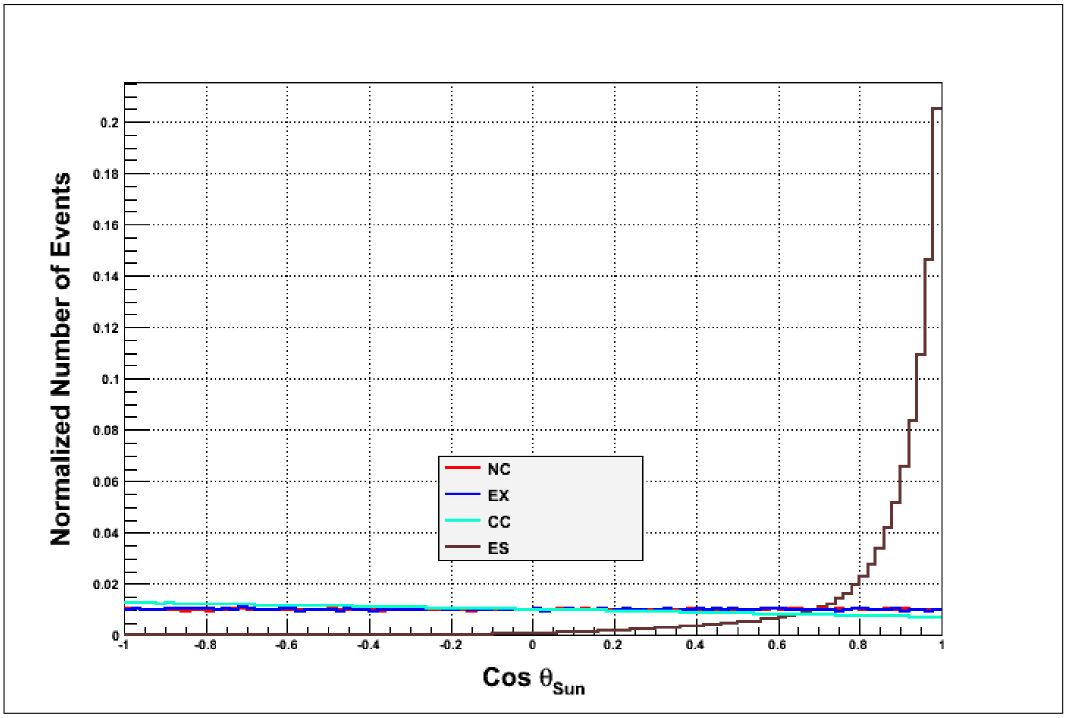


Figure 4.4: Distribution of Cos $\theta_\odot$ for EX, CC, ES and NC. Both energy ($6.0 \leq E < 20$) and fiducial volume ($\rho < 0.77025$) cuts are applied to plot the distributions. ES peak at cos $\theta \simeq 1$ pointing away from the Sun proves that Sun is the origin of neutrinos.

4.8.3 Cosine $\theta_\odot$

Cosine θ is a reconstructed direction of an event relative to the direction of a ν arriving from the Sun. Cosine $\theta = 1$ means forward scattered electron and Cosine θ=-1 means electron scattered in the backward direction. Recoil electrons from ES have a strong forward scattering peak (figure 4.4.). This confirms that the Sun is the source of neutrinos. The angular correlation is key to separate ES events from other events in SNO.

The recoil electrons from CC also have a weak angular dependence as shown in figure 4.5. The distribution of angles between the incident neutrino and the recoil electron is described as: $1 - 0.340 \cos \theta_\odot$ [69]. This feature was used to help separate CC from other signals and backgrounds in SNO.

The NC distribution is flat because a γ from a neutron capture has no

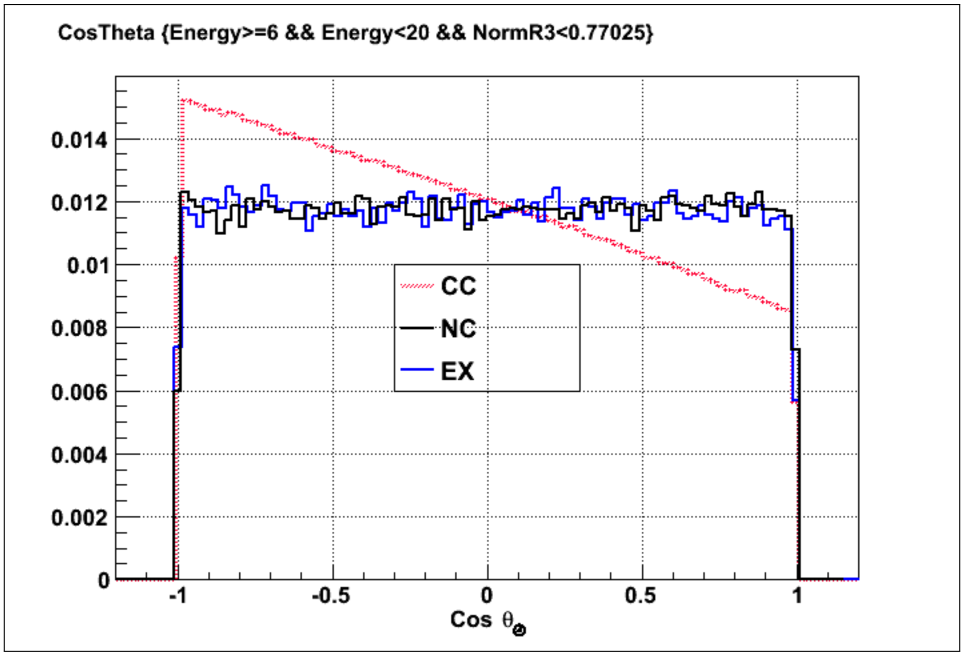


Figure 4.5: Distribution of Cos $\theta_{\odot}$ for CC (dotted line in pink). Both energy ($6.0 \leq E < 20$) and fiducial volume ($\rho < 0.77025$) cuts are applied to plot the distributions. For comparison the distributions of Cos $\theta_{\odot}$ for NC and EX is also included.

memory of the incident neutrino direction and since backgrounds had no correlation with the Sun's position in the sky, they also exhibited a completely flat distribution.

4.9 Data Selection Cuts

One example of instrumental backgrounds is static discharges inside the PMTs, which generate flashes of light. Such events are called "flashers" for any given PMT. Although flashers are rare, they contribute significantly when integrated over the entire array. The events belonging to instrumental backgrounds are distinguishable from Čerenkov light and so can be identified and removed based on an analysis of charge and timing distribution of the triggered PMTs, the spatial distribution of PMT hits and/or the firing of specific photomultiplier veto tubes. A number of cuts were applied in the analysis to remove the instrumental backgrounds; these are fully described in reference [70]. Each cut returns a simple binary decision which was stored in a Data Analysis Mask Number (DAMN) bank. A DAMN mask is applied to the dataset in a bitwise manner to select events passing specific cuts for a specific analysis. Another level of cuts, applied both to the Monte Carlo and the real data, are called high level cuts. These are applied to the reconstructed quantities, for example, event energy, event position and direction and event time. Backgrounds which originate outside of the fiducial volume but reconstructed far from the their true location will generally have very unusual hit patterns. The purpose of high-level cuts are to remove these events. Examples of high-level cuts include:

- In-Time Ratio (ITR): ITR is a ratio of the number of PMTs hit in the prompt time window of ±10 nanoseconds compared to the number of PMTs hit outside the window. The ITR cut is ITR>0.55. A Čerenkov event was fairly instantaneous, therefore had a high ITR≈ 0.75; misreconstructed and non-Čerenkov events, such as electronic noise, produced light which was spread over a longer period of time and consequently had smaller values of ITR.

- The energy estimator, for the D_2O and Salt phases, returns a most probable energy, as well as an uncertainty on the energy. Since misreconstructed events tend to have very large uncertainties, they can be thus removed.

The high-level cuts are: $\rho <0.77025$, -1$\leq$ Cos $\theta_\odot <1.0$ and 6 MeV $\leq$ energy$<$20 MeV.

4.10 The Likelihood Function

The likelihood of an event is a probability of observing that event given the measured values ($\vec{x}_d$) of the observables (Energy, cos $\theta_\odot$, ρ) and models of event classes (PDFs) to which that event might belong. The likelihood function is the product of the probabilities for each event in the dataset:

$$\mathcal{L} = \prod_{d=1}^{N_D} F(\vec{x}_d, \vec{P}) \tag{4.2}$$

where N_D is a number of events in the dataset, $\vec{x}_d$ are the values measured for observables for an event **d** and F$(\vec{x}_d, \vec{P})$ is the probability density function returning the probability of observing the event with observables $\vec{x}_d$ and the current values of the fit parameters $\vec{P}$. The SNO dataset consists of different signal classes. In this case, the probability of measuring a specific set of observable values $(\vec{x}_d)$ becomes a linear combination of the probability of measuring those values for an event of each signal type:

$$F(\vec{x}_d, \vec{P}) = \sum_{i=1}^{N_s} c_i F_i(\vec{x}_d, \vec{P}) \tag{4.3}$$

$$= \sum_{i=1}^{N_s} \frac{n_i}{N_D} F_i(\vec{x}_d, \vec{P}) \tag{4.4}$$

where N$_s$ is the total number of signals in the fit, c_i is a component of $\vec{P}$ representing the probability of observing an event of signal type i which is the fraction of the events of that type in the dataset $(\frac{n_i}{N_D})$, n$_i$ is the number of events belonging to signal type i in the dataset, N$_D$ is the number of events in the dataset and F$_i(\vec{x}_d, \vec{P})$ is the probability of event type i having observables values $(\vec{x}_d)$ and the current values of the fit parameters $\vec{P}$. The number of events observed in the dataset is actually Poisson distributed around the true mean value of the model, μ. The number of events for a signal type i (n$_i$) represents Poisson fluctuation about the value μ_i. In order to fit for the true value μ_i for each signal, the Poisson fluctuation of n$_i$ is taken into consideration in the likelihood function. The *extended* likelihood function is:

$$\mathcal{L} = \mu^{N_D} \frac{e^{-\mu}}{N_D!} \prod_{d=1}^{N_D} F(\vec{x}_d, \vec{P})$$

$$= \frac{e^{-\mu}}{N_D!} \prod_{d=1}^{N_D} \mu F(\vec{x}_d, \vec{P}) \tag{4.5}$$

where $\mu = \sum_{i=1}^{N_s} \mu_i$ such that $\mu_i = \mu(n_i/N_D)$. Taking the log of the likelihood function:

$$\log \mathcal{L} = -\mu - \log N_D! + \sum_{d=1}^{N_D} \log \big(\mu F(\vec{x}_d, \vec{P})\big) \tag{4.6}$$

Ignoring $\log N_D!$ because it is a constant number and has no influence on the posterior distribution and substituting equation (4.4) in the above equation results in the following equation:

$$\log \mathcal{L} = -\sum_{i=1}^{N_s} \mu_i + \sum_{d=1}^{N_D} \log \big(\mu \sum_{i=1}^{N_s} (n_i/N_D) F_i(\vec{x}_d, \vec{P})\big) \tag{4.7}$$

Substituting $\mu_i = \mu(n_i/N_D)$ results in:

$$\log \mathcal{L} = -\sum_{i=1}^{N_s} \mu_i + \sum_{d=1}^{N_D} \log \big(\sum_{i=1}^{N_s} \mu_i F_i(\vec{x}_d, \vec{P})\big) \tag{4.8}$$

where $\vec{P}$ (which include μ_i) are the parameters fitted in the MCMC fit.

Taking into account the day-night asymmetry due to MSW effect and the detector response, the likelihood equation changes to:

$$\log \mathcal{L} = -\sum_{i=1}^{2N_s} \mu_i + \sum_{d=1}^{N_D} \log \big(\sum_{i=1}^{2N_s} \mu_i F_i(\vec{x}_d, \vec{P})\big) \tag{4.9}$$

where each event class was split into two cases: day and night.

This is a simple likelihood function without any constraints or floating systematic uncertainties. From this point on, I will explain different components that go into the calculation of a number of events; application of different systematic uncertainties (Subsection 4.10.3), constraints applied on the systematic uncertainties (Subsection 4.10.4) and the role the systematic uncertainties play in the calculation of the number of events (Subsection 4.10.5), floating PMT NC detection efficiency (Subsection 4.10.1), constraints applied from external measurements on the number of events (Subsection 4.10.6), the role of survival probability of ν in computing the number of events of CC, ES and $\mathrm{ES}_{\mu\tau}$ (Subsection 4.10.7), constraints applied from different analysis (Subsections 4.10.8 and 4.10.9) to reduce the systematic uncertainties of the final result from the MCMC fit. Subsection 4.11 is a synopsis on the computation of the number of events of different event classes and at the end in the summary 4.13 all the components of the log likelihood will be combined in the final log likelihood function (4.80). Section 4.12 describes determining the goodness of a fit by performing pull and bias study of the fit.

4.10.1 Flux to Event Conversion Factor (F2EF) and Fiducial Volume Correction – $\mathbf{S}_i$

One of the parameters, neutron capture efficiency, is floated via flux-to-event conversion factor which includes livetime, neutron detection efficiency and correction factors. F2EF is used to convert the number of events, that passed the high-level cuts, into a flux; F2EF for NC is calculated as:

$$N_{nc}^{model} = R_{nc}\,\mathrm{LT}\,\epsilon_{nc}\,\epsilon_{\mathrm{corr}} \tag{4.10}$$

$$\mathrm{F2EF} = f_{nc} = \frac{N_{nc}^{\mathrm{model}}}{\phi_{\mathrm{SNOMAN}}} \tag{4.11}$$

where LT=392.89 days is the livetime of the NCD phase, ϕ_{SNOMAN} is NC flux in Monte Carlo ($5.14\times \nu s\, 10^6 cm^{-2} sec^{-1}$), $\epsilon_{nc} = 0.0485$ is the neutron capture efficiency for the PMTs in the NCD phase and $R_{nc} = 13.27$ is the rate of neutrons per day expected in SNO after high level cuts. Since the simulation does not mirror the data perfectly, corrections ϵ_{corr} are applied to the number of the predicted events.

The fiducial volume correction is calculated each time the parameters are changed which entails rebuilding the PDFs for the changed values of the parameters; S_i is a ratio of the number of events (N_k^{mc} for step k in MCMC) inside the ρ, Energy and Cos $\theta_\odot$ cuts for the current values of the systematic parameters to the number of events ($N_{\mathrm{default}}^{mc}$) passing the same cuts for the default systematic parameters used to build PDFs from the SNOMAN Monte Carlo. Mathematically expressed as:

$$S_i = \frac{N_k^{mc}}{N_{\text{default}}^{mc}} \tag{4.12}$$

4.10.2 Calculation of the Number of Events for the Neutral Current and the Backgrounds

This section describes the methodology for calculating the expected number of day (equation (4.20)) and night (equation (4.21)) events for the backgrounds: EX, $d_2$opd, atmospheric neutrinos, k2pd, k5pd and ncdpd. To take into account the possibility that a systematic uncertainty or number of events for a background might have different values for a day and night data, a day-night analysis was carried out [92]. One option is to split the uncertainty into two uncertainties β_d and β_n where β_d is used when calculating fluxes or building/normalizing the PDFs for the day while β_n is used for the night data. A better option is to float an average on β and an effective day-night asymmetry on β. Mathematically expressed as:

$$\beta = \frac{(\beta_n + \beta_n)}{2} \tag{4.13}$$

$$A_\beta = \frac{2(\beta_n - \beta_d)}{(\beta_n + \beta_d)} \tag{4.14}$$

Inverting these two equations gives β_n and β_d in terms of β and A_β.

$$\beta_n = (1 + A_\beta/2)\beta \tag{4.15}$$

$$\beta_d = (1 - A_\beta/2)\beta \tag{4.16}$$

$$A = 2\left(\frac{r_n - r_d}{r_n + r_d}\right) \tag{4.17}$$

$$r \equiv \frac{r_d + r_n}{2} \tag{4.18}$$

$$\alpha = \frac{r}{r_{nom}} \tag{4.19}$$

$$N_{Day} = \alpha\,\frac{n_{nom}}{N(\vec{\eta})}(1.0 - 0.5\,A)N_d(\vec{\eta}) \tag{4.20}$$

$$N_{Night} = \alpha\,\frac{r_{nom}}{N(\vec{\eta})}\,(1.0 + 0.5\,A)N_n(\vec{\eta}) \tag{4.21}$$

$$N = N_{Day} + N_{Night} \tag{4.22}$$

where

- α is average background rate r in the data in terms of a most likely value from an external measurement r_{nom}. This is the default rate used in the MC.
- N – the number of background events.
- A – the day-night background asymmetry for this event type.
- r_d and r_n – rates in the data for a day and night respectively.
- r_{nom} - the nominal background data rate corresponding to the most likely value from an external measurement.
- $N(\vec{\eta})$ – the number of background events in the Monte Carlo which satisfy the cuts after the application of current systematic parameters $\vec{\eta}$.
- n_{nom} – the number of background events corresponding to the external measurement.
- $N_d(\vec{\eta})$ and $N_n(\vec{\eta})$ – the number of day and night Monte Carlo generated events that satisfy the cuts after the application of current systematic parameters $\vec{\eta}$.

Except $d_2$opd and EX neutrons background, the day-night asymmetry A is zero for all other backgrounds. In the MCMC fit, α is floated – along with A (where applicable) – rather than the number of events.

The number of events for NC were calculated as:

$$N_{Day} = f_{8B}\, f_{NC}^{PMT} \big(\frac{N_d(\vec{\eta})}{N_d(\vec{\eta}) + N_n(\vec{\eta})}\big) \tag{4.23}$$

$$N_{Night} = f_{8B}\, f_{NC}^{PMT} \big(\frac{N_n(\vec{\eta})}{N_d(\vec{\eta}) + N_n(\vec{\eta})}\big) \tag{4.24}$$

$$N = N_{Day} + N_{Night} \tag{4.25}$$

where f_{8B} and f_{NC}^{PMT} are current MCMC values of ^{8}B flux and flux-to-event ratio for NC from the PMTs.

4.10.3 Systematic Uncertainties

The position, time and direction of an event were reconstructed by simultaneously fitting them using the hit times and locations of the hit PMTs. After the vertex and direction reconstruction from the maximum log likelihood method, they were used in the estimation of event energy. Hence the dominant sources of systematic uncertainty in the signal extraction fit are concerned with the reconstruction accuracy and are listed in tables 4.2 to 4.5. Data from calibration sources deployed within the detector were compared to Monte Carlo predictions (from the vertex and direction reconstruction algorithm) and the full size of the difference was taken as the magnitude of the uncertainty. The differences between the calibration data and the reconstructed data from Monte Carlo were parametrized as four types:

(i) vertex offset is a constant offset between an event's true and reconstructed position.

(ii) vertex scale is a position dependent bias in the reconstructed position that is proportional to the difference between the reconstructed event and the actual location of the source in the calibration data.

(iii) vertex resolution is a width of a distribution of the reconstructed event positions.

(iv) angular resolution is the width of the distribution of reconstructed event directions relative to the initial electron direction of the source.

Since these uncertainties can alter the predictions for the number of events reconstructed inside the fiducial volume and can additionally alter the shape of the PDFs used in the signal extraction, we incorporate systematic uncertainties into the analysis as parameters in the likelihood function which are allowed to "float" meaning vary within $\pm 10\sigma$ where σ is width of constraint from external measurement or found from trial and error to get good acceptance in the Markov Chain and autocorrelation coefficient to drop to zero within the first 10,000 steps and remain stable (figure 5.2). The purpose of floating the systematics is to properly calculate the correlation between systematic effects and to allow the data to tell us how the scales and resolution differ between data and Monte Carlo within the constraints from the calibration data. By modelling the differences as a possible remapping of the observables for Monte Carlo generated events, the PDFs are rebuilt using the scaled and smeared values of the observables on each evaluation of the log likelihood calculation. The parameters of the remapping are fitted to determine the extent of remapping allowed to the MC observables while still matching the ρ, Cosine $\theta_{\odot}$ and E_m of the data. Day-night events are selected by using a day-night tag in the Monte Carlo; for the NCD phase day-night tag is 30 for day events and 31 for night events. Tables 4.2 and 4.3 list the parameters involved in the determination of the energy of an event, tables 4.4 and 4.5 list the systematic parameters involved in the determination of a direction and a location of an event. One of the prediction of a "matter enhanced" oscillation is an asymmetry in a day and night fluxes but the asymmetry can also arise because of the variations in detector response over a 24-hour time scale, for instance, diurnal variations in a laboratory's temperature. SNO's 2 km underground location isolate it from diurnal effects but limits must be placed on their size.

Directional systematics arise because SNO is not completely spherically symmetric, and because the directions of CC and ES events are correlated with the time of the day. ES events, in particular, preferentially illuminate the upper half during the night and lower half during the day. Any differences in the up-down response of the detector or variations in detector response with the direction of the event will manifest these directional differences as a day-night asymmetry. The γ-rays emitted by neutron capture have random directions, hence asymmetry in detector response produce no day-night effects on NC events.

Energy Systematics

Energy Scale – a_0^E

Parameter	**Description**	**Central Value**	**Constraint**
a_1	Energy scale (correlated)	0	± 0.0041
a_2	Energy scale	0	± 0.0081
a_3	Energy scale Diurnal asymmetry	0	± 0.0038
a_4	Energy scale Directional asymmetry for *es* only	0	± 0.0099
c_0	Energy non-linearity	0	± 0.0069

Table 4.2: Various parameters for the uncertainty in the energy scale.

For the day events,

$$a_0^E = (1.0 + a_1 + a_2)\,(1 - 0.5\,a_3 - 0.5\,a_4) \tag{4.26}$$

For the night events,

$$a_0^E = (1.0 + a_1 + a_2)\,(1 + 0.5\,a_3 + 0.5\,a_4) \tag{4.27}$$

where a_4 is zero for all classes except ES and $ES_{\mu\tau}$. The energy non-linearity systematic uncertainty – applied to CC, ES, hep CC and hep ES – accounts for possible changes in the energy scale away from the ^{16}N source used to calibrate the energy scale. This uncertainty is correlated between all three phases.

Energy Resolution – b_0^E

Parameter	**Central Value**	**Constraint**
Energy resolution b_0 for a neutron	0.0119	± 0.0104
Energy resolution for an electron	0.0162	± 0.0141
Directional asymmetry in resolution for ES only – b_1	0	± 0.012

Table 4.3: Various parameters for the uncertainty in the energy resolution. Energy resolution for a neutron and an electron is 100% correlated.

For the day events,

$$b_0^E = b_0\,(1 - 0.5\,b_1) \tag{4.28}$$

For the night events,

$$b_0^E = b_0\,(1 + 0.5\,b_1) \tag{4.29}$$

where b_1 is zero for all classes except ES and $ES_{\mu\tau}$.

Equations for remapping energy

Equations (4.30) is applied for CC, ES, $ES_{\mu\tau}$, hep CC and hep ES. Equation (4.31) is for NC and neutron backgrounds.

$$T_{remap} = a_o^E\,T_0 + 1.3613\,b_0^E\,(T_0 - T_g) + c^0\,T_0(T_0 - 5.05\,\mathrm{MeV})/(19.0 - 5.05)\,\mathrm{MeV} \tag{4.30}$$

$$T_{remap} = a_o^E\,T_0 + b_0^E\,(T_0 - 5.65\,\mathrm{MeV}) + c^0\,T_0(T_0 - 5.05\,\mathrm{MeV})/(19.0 - 5.05)\,\mathrm{MeV} \tag{4.31}$$

where T_0 and T_g are the reconstructed energy and the Monte Carlo energy respectively and T_{remap} is the remapped energy which can be directly compared to the data. To calculate the energy resolution of neutral current (nc) and 9 backgrounds, the number 5.65 MeV is used as a mean T_g for every event in the MC in equation (4.31). Equation (4.30) is employed to remap the energy of electron classes, for example, cc, es and $es_{\nu\mu}$. Neutron classes are neutral current and all the neutron backgrounds (ex, atmos, $d_2$opd, k2pd, k5pd, ncdpd, hep NC) and electron classes are CC, ES, $ES_{\mu\tau}$, hep CC and hep ES.

Angular Resolution (Cosine $\theta_\odot$) for ES only

Parameter	**Description**	**Central Value**	**Constraint**
b_0^θ	Resolution	0.0	± 0.12
b_1^θ	Directional asymmetry	0.0	± 0.069

Table 4.4: Uncertainty in the angular resolution.

$$\text{For day events } Cos\,\theta_{remap} = 1. + (1 + b_0^\theta)(1 - b_1^\theta)(Cos\,\theta - 1) \tag{4.32}$$

$$\text{For night events } Cos\,\theta_{remap} = 1. + (1 + b_0^\theta)(1 + b_1^\theta)(Cos\,\theta - 1) \tag{4.33}$$

where θ is the angle of the Monte Carlo event relative to the direction of the Sun and $Cos\,\theta_{remap}$ is the remapped observable to build the 3D PDF (ρ, Cos $\theta_\odot$, T_0). Events that are pushed passed cos $\theta_\odot = \pm 1.0$ are randomly assigned a Cos θ value in the interval [-1.0,1.0].

Vertex (x,y,z) Systematics

Monte Carlo events are generated at vertex (x_g,y_g,z_g)[5] and energy T_g while vertex (x,y,z) with the energy T_0 are the reconstructed vertex and energy of the Monte Carlo events.

Parameter	**Description**	**Central Value**	**Width of Constraint**
a_1^{xyz}	x,y,z coordinate scale (100% correlated)	0.0	$+$**0.0029-0.0077**
a_2^{xyz}	Diurnal asymmetry	0.0	$\pm$0.0015
a_3^{xyz}	Directional asymmetry for ES only	0.0	$\pm$0.0018
a_0^x	x offset	0.0	$\pm$4.0
a_0^y	y offset	0.0	$\pm$4.0
a_0^z	z offset	5.0	$\pm$4.0
b_0^{xy}	x,y resolution constant	0.06546	$\pm$0.000818124
b_1^{xy}	x,y resolution linear	-5.501$\times$ 10^{-5}	$\pm 3.66098\times10^{-9}$
b_2^{xy}	x,y resolution quadrature -	3.9×10^{-7}	$\pm 3.92118\times10^{-14}$
b_s^z	z scale	0.0	$+$**0.0015-0.0012**
b_0^z	z resolution constant	0.07096	0.00078696
b_1^z	z resolution linear	1.155×10^{-4}	$\pm 6.80761\times10^{-9}$
c_0^{xyz}	energy dependence in the fiducial volume uncertainty	0.0	$+$**0.0087-0.0067**

Table 4.5: Uncertainties in the reconstruction of a vertex.

$$a_0^{xyz} = (1.0 \pm\ 0.5\, a_2^{xyx}\ \pm\ 0.5\, a_3^{xyz}) \tag{4.34}$$

[5]For analysis purposes Cartesian coordinates are defined such that the center of the acrylic vessel is at (x,y,z)=(0,0,0) and the neck of the acrylic vessel is located symmetrically about the positive z axis.

where + is for a night event and - is for a day event. We find the corresponding remapped MC variables:

$$x_{remap} = a_0^x + (1.0 + a_1^{xyz}) a_0^{xyz}\, x + (b_0^{xy} + b_1^{xy}\, z + b_2^{xy}\, z^2)\,(x - x_g) \tag{4.35}$$

$$y_{remap} = a_0^y + (1.0 + a_1^{xyz}) a_0^{xyz}\, y + (b_0^{xy} + b_1^{xy}\, z + b_2^{xy}\, z^2)\,(y - y_g) \tag{4.36}$$

$$z_{remap} = a_0^z + (1.0 + a_1^{xyz} + b_s^z)\, z + (b_0^z + b_1^z\, z)\,(z - z_g) \tag{4.37}$$

$$R = \sqrt{(x_{remap}^2 + y_{remap}^2 + z_{remap}^2} \tag{4.38}$$

$$\rho = (R/600.0)^3 \tag{4.39}$$

4.10.4 Application of Constraints on the Systematic Uncertainties

A penalty is applied on the fit if an external measurement is used to directly constrain a fit parameter. Since all systematic uncertainties were measured from comparing the calibration data to the Monte Carlo simulation of the calibration data, a constraint term for each systematic uncertainty is added to the log likelihood function. Values for the mean and sigma of these constraints terms comes from calibration measurements or Monte Carlo simulation. For the Čerenkov data, constraints were largely obtained from ^{16}N calibration data. The energy uncertainties are described in [29] and the reconstruction uncertainties are described in [90]. When the systematic uncertainties are not correlated among themselves, the constraint term is calculated as:

$$-\log \mathcal{L} = \frac{1}{2} \sum_i \left(\frac{p_i - \bar{p_i}}{\sigma_{p_i}}\right)^2 \tag{4.40}$$

$$\tag{4.41}$$

where p_i, $\bar{p_i}$ and σ_{p_i} represent the current value of the systematic uncertainty i in the MCMC fit, its mean and constraint width respectively. The three parameters (b_0^{xy}, b_1^{xy} and b_2^{xy}) – common to x and y coordinates – are correlated, and the constrained term is calculated using a covariance matrix:

$$-\log \mathcal{L} = \frac{1}{2}\sum_{i=0}^{2}\sum_{j=0}^{2}(b_i^{xy} - b_i^{\bar{xy}})(b_j^{xy} - b_j^{\bar{xy}})(V_{b^{xy}}^{-1})_{ij} \tag{4.42}$$

Similarly, the constraint term for two parameters, for a z-dependent vertex resolution, is calculated as:

$$-\log \mathcal{L} = \frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}(b_i^{z} - b_i^{\bar{z}})(b_j^{z} - b_j^{\bar{z}})(V_{b^{z}}^{-1})_{ij} \tag{4.43}$$

where the covariance matrices $V_{b^{xy}}$ and V_{b^z} [90] are:

$$\begin{pmatrix} 0.000818124 & -2.24984 \times 10^{-7} & -4.19131 \times 10^{-9} \\ -2.24984 \times 10^{-7} & 3.66098 \times 10^{-9} & 3.71423 \times 10^{-12} \\ -4.19131 \times 10^{-9} & 3.71423 \times 10^{-12} & 3.92118 \times 10^{-14} \end{pmatrix}$$

$$\begin{pmatrix} 0.00078696 & 3.47188 \times 10^{-7} \\ 3.47188 \times 10^{-7} & 6.80761 \times 10^{-9} \end{pmatrix}$$

If the uncertainty has asymmetric errors, the constraint is applied to the likelihood as:

$$\mathcal{L} = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_- + \sigma_+}e^{-0.5\left[(x-\mu)/\sigma\right]^2} \tag{4.44}$$

$$\sigma = \begin{cases} \sigma_- & \text{if } x < \mu \\ \sigma_+ & \text{if } x \geq \mu \end{cases} \tag{4.45}$$

where σ_- and σ_+ allow for asymmetric uncertainties, **x** is the current value of the uncertainty in the MCMC fit, μ and σ are the mean and width of the uncertainty respectively. The log likelihood term, corresponding to these constraints, is simplified, after dropping the constants, to:

$$\log \mathcal{L} = -0.5(\frac{x-\mu}{\sigma})^2 \tag{4.46}$$

4.10.5 Application of energy dependent fiducial volume and uncertainty in the shape of ^{8}B neutrino energy spectrum

Fiducial volume in SNO is fixed at 550.0 cm. The fiducial volume uncertainty is applied by assigning an uncertainty to every event's vertex (x,y,z) in the Monte

Carlo. The energy dependent fiducial volume and uncertainty in the shape of the ^{8}B ν energy spectrum are coupled together because they are applied as weight to some PDFs. The uncertainty to the shape of ^{8}B ν energy spectrum is applied in accordance with the limits from the paper published by Winter *et al.*in [95]. Henceforth this uncertainty will be called Winter uncertainty in this thesis.

Each MC event gets weighted by a factor W_{ij} (ij: current step i for class j) when building the PDF:

$$W_{ij} = 1.0 + c_0^{xyz}(a_0^E T_{0ij} - 5.05\,\mathrm{MeV}) \tag{4.47}$$

W_{ij} in equation (4.47) is an energy-dependent fiducial volume factor applied around the midpoint of the ^{16}N energy (5.05 MeV), where $a_0^E T_{ij}$ is the scaled reconstructed effective electron kinetic energy. The Winter ^{8}B spectral shape uncertainty is propagated in MCMC by reweighing the CC, ES and NC events using the function:

$$W'_{ij} = W_{ij}\big(1.0 + (\frac{w}{3})(0.018 - 0.001999 \times E_{ij} - 0.000088769 \times E_{ij}^2)\big) \tag{4.48}$$

$$W_j = \sum_{i=1}^{N_j} W_{ij}/N_j \text{ for backgrounds} \tag{4.49}$$

$$W_j = \sum_{i=1}^{N_j} W'_{ij}/N_j \text{ for signals} \tag{4.50}$$

$$W_{nc} = \sum_{i=1}^{N_j} \big(1.0 + (\frac{w}{3})(0.018 - 0.001999 \times E_{ij} - 0.000088769 \times E_{ij}^2)\big)/N_j \tag{4.51}$$

where W_{ij} is used for the backgrounds and W'_{ij} is used for the signals. These are applied as a weight to the MC event i used to define the PDF belonging to the class j, W_j (equation (4.49) for the backgrounds or (4.50) for the signals) is the factor by which the number of expected events is modified, c_0^{xyz} takes into account dependence of the reconstruction of a vertex of an event on the energy of the event, w is a systematic parameter which has a normal distribution of N(0,1), E_{ij} is a neutrino energy and N_j is the number of events that pass the cuts for class j. The expected number of neutral current events on the NCD-side of the NCD phase is modified by W_{nc}. The calculated number of day and night events on the Čerenkov-side, n_j, is modified by:

$$N_j = n_j * W_j \tag{4.52}$$

where n_j is the number of events belonging to the class j, and N_j is the modified number of events for the class j after the application of the above mentioned uncertainties. The application of N_j in the calculation of the log likelihood is described in equation (4.80).

4.10.6 Constraints from the Backgrounds added to the Likelihood Function

Ex-situ/in-situ radioassays were performed to measure the concentrations of ^{214}Bi and ^{208}Tl in the detector. These concentrations are converted to the expected number of events in the analysis window which were used to constrain the number of events for each background. The constraints were assumed to be Gaussian and applied in the likelihood function as:

$$\log \mathcal{L} = \sum_{i=1}^{m} \log\left(C_i(\mu_i)\right) \tag{4.53}$$

where C_i is the Gaussian probability of obtaining μ_i background events of type i and m is the number of backgrounds. Substituting the Gaussian function $e^{\frac{-(N_i-\mu_i)^2}{(2\sigma_i^2)}}$ in equation (4.53) and taking the log will result in:

$$\log \mathcal{L} = -\sum_{i=1}^{m} \frac{(N_i-\mu_i)^2}{(2\sigma_i^2)} \tag{4.54}$$

where N_i is the number of background events of type i in the current step. If σ_i is comparable to μ_i then the Gaussian distribution will not be symmetric because the number of events can not be negative. This issue is discussed in detail in chapter 6.

4.10.7 P_{ee} Survival Probability with Day Night Asymmetry

This section describes how the PDFs are distorted and number of events are calculated for CC, ES and $ES_{\mu\tau}$ using the survival probability equations and ^{8}B flux. The basic assumption is that the flux of electron neutrinos is modified by a factor:

$$P_{ee_D}(E_\nu) = p_0 + p_1\,(E_\nu - 10\,\mathrm{MeV}) + p_2\,(E_\nu - 10\,\mathrm{MeV})^2 \tag{4.55}$$

$$A_{ee} \equiv \frac{2(\phi_N - \phi_D)}{(\phi_N + \phi_D)} = a_0 + a_1\,(E_\nu - 10\,\mathrm{MeV}) \tag{4.56}$$

$$P_{ee_N}(E_\nu) = \left(p_0 + p_1\,(E_\nu - 10\,\mathrm{MeV}) + p_2\,(E_\nu - 10\,\mathrm{MeV})^2\right)\frac{2 + A_{ee}}{2 - A_{ee}} \tag{4.57}$$

where E_ν is neutrino energy and A is energy-dependent day-night asymmetry on the survival probability. The survival probabilities were parametrized in this way to reduce correlations between p_0 and the higher order terms by expanding all functions around the peak (10.0 MeV) of ^{8}B energy spectrum. An advantage of the asymmetry ratio is that most systematics cancel out except those that scale day and night fluxes differently. Equations (4.55) and (4.57), quadratic equations to represent the survival probability of an electron neutrino, are used to distort the day **D** and night **N** PDFs of CC and ES.

For $ES_{\nu\tau}$ following equations are used:

$$P_{ee_D}(E_\nu) = 1 - \left(p_0 + p_1\,(E_\nu - 10\,\mathrm{MeV}) + p_2\,(E_\nu - 10\,\mathrm{MeV})^2\right) \tag{4.58}$$

$$P_{ee_N}(E_\nu) = \left(1 - (p_0 + p_1\,(E_\nu - 10\,\mathrm{MeV}) + p_2\,(E_\nu - 10\,\mathrm{MeV})^2)\right)\frac{2 + A_{ee}}{2 - A_{ee}} \tag{4.59}$$

The reason for the difference is that Elastic scattering interaction has contributions from electron, μ and τ neutrinos, hence if P_{ee} is the survival probability for an electron neutrino then the survival probability for μ and τ neutrino will be $1 - P_{ee}$ hence the non-electron components of the day and night ES flux, respectively, are scaled by 1-$P_{ee_D}(E_\nu)$ and 1-$P_{ee_N}(E_\nu)$.

The 3D PDF ($f_D(\vec{x})$ where $\vec{x} = \rho, \cos\theta_\odot, E_m)$) for the day instance **D** is distorted using the 2D PDF $h_D(E_\nu, E_m)$ in neutrino energy (E_ν) and measured energy (E_m) via:

$$f_D(\vec{x}, P_{ee}) = \int f_D(\vec{x}) \frac{h_D(E_\nu, E_m) P_{ee_D}(E_\nu)}{h_D(E_\nu, E_m)} \, dE_\nu \tag{4.60}$$

The 3D PDF, $f_N(\vec{x})$, for the night instance **N** is distorted using the 2D PDF $h_N(E_\nu, E_m)$ via:

$$f_N(\vec{x}, P_{ee}) = \int f_N(\vec{x}) \frac{h_N(E_\nu, E_m) P_{ee_N}(E_\nu)}{h_N(E_\nu, E_m)} \, dE_\nu \tag{4.61}$$

Figures 4.6 to 4.9 show $h_D(E_\nu, E_m)$ and $h_N(E_\nu, E_m)$ and their projections on the X axis. Besides distorting the 3D PDFs, these histograms are also used in the calculations of the number of events for the CC and ES event classes.

Using the histograms shown in figures 4.6 and 4.7, the number of CC events is calculated:

$$CC = CC_D + CC_N \tag{4.62}$$

$$CC_D = f_{nc} \, S_{cc} \, \phi_{nc} \, (\frac{\sigma_{cc} \, \epsilon_{cc}}{\sigma_{nc} \, \epsilon_{nc}}) p_{ee_d} \, R_D \tag{4.63}$$

$$CC_N = f_{nc} \, S_{cc} \, \phi_{nc} \, (\frac{\sigma_{cc} \, \epsilon_{cc}}{\sigma_{nc} \, \epsilon_{nc}}) p_{ee_n} \, (1 - R_D) \tag{4.64}$$

where S_{cc} is a fiducial volume correction for CC, R_D is a ratio of the number day events to the total number of events in the Monte Carlo, ϕ_{nc} is the ^{8}B flux, f_{nc} is flux to event conversion factor and the variable p_{ee_d} (a ratio of the number of events with given values of P_{ee} from equations (4.55) and (4.57) to the number of events with P_{ee} equal to 1.0) is calculated as:

$$p_{ee_d} = \frac{\iint h_D(E_\nu, E_m) \, P_{ee}(E_\nu, day) dE_\nu \, dE_m}{\iint h(E_\nu, E_m) dE_\nu \, dE_m} \tag{4.65}$$

$$p_{ee_n} = \frac{\iint h_N(E_\nu, E_m) \, P_{ee}(E_\nu, night) dE_\nu \, dE_m}{\iint h(E_\nu, E_m) dE_\nu \, dE_m} \tag{4.66}$$

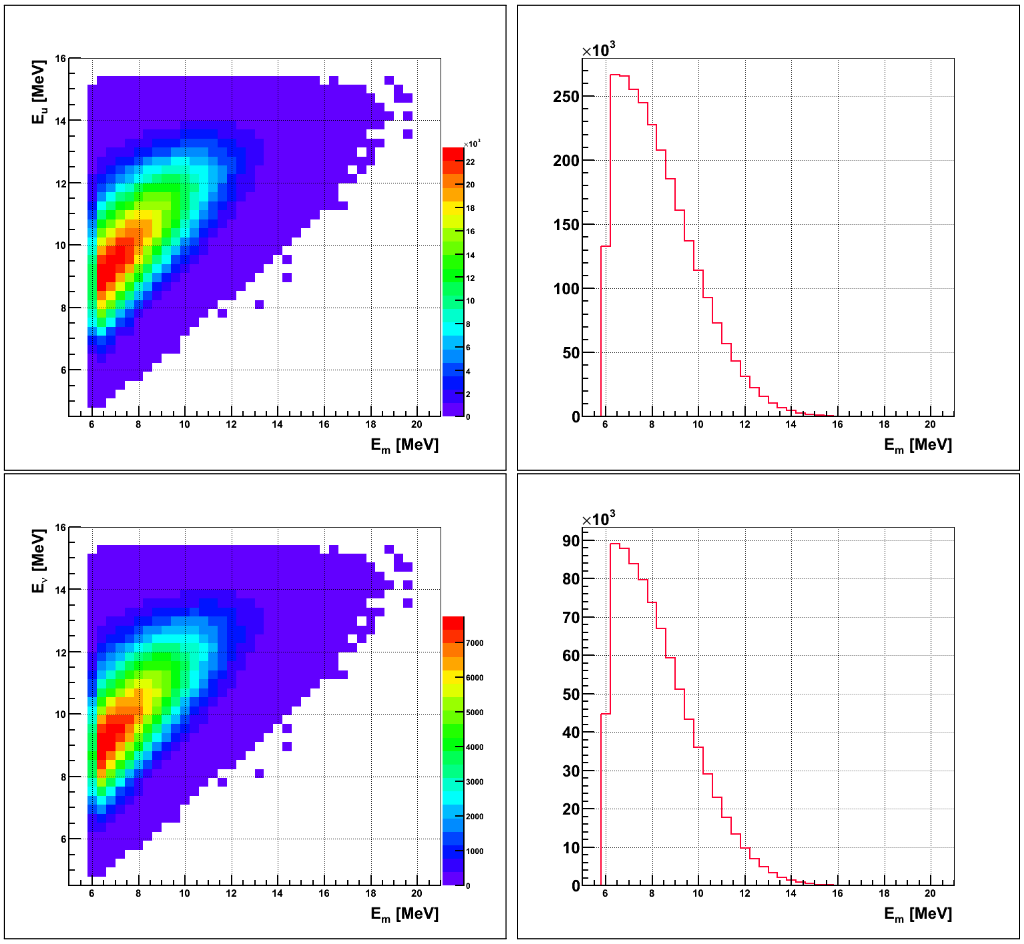


Figure 4.6: Top left is a 2D distribution of neutrino energy versus measured energy ($h_D(E_\nu, E_m)$) and the top right is a projection of the 2D distribution on the measured energy ($\int h_D(E_\nu, E_m)\, dE_\nu$). Bottom left is a 2D distribution of neutrino energy versus measured energy weighted by $P_{ee_D}(E_\nu)$ ($h_D(E_\nu, E_m)P_{ee_D}(E_\nu)$) and the right plot shows $\int h_D(E_\nu, E_m)P_{ee_D}(E_\nu)dE_\nu$. The right 1D histograms are projections on E_ν of the 2D histograms on the left. The correlation of E_m with E_ν, as seen in this plot is taken into consideration when distorting the 3D PDFs. The reduction in the number of events – comparing the top plots to the bottom plots – is due to ν oscillations which is applied as distortions. In the 2D histograms, the number of events is shown in the color pallet.

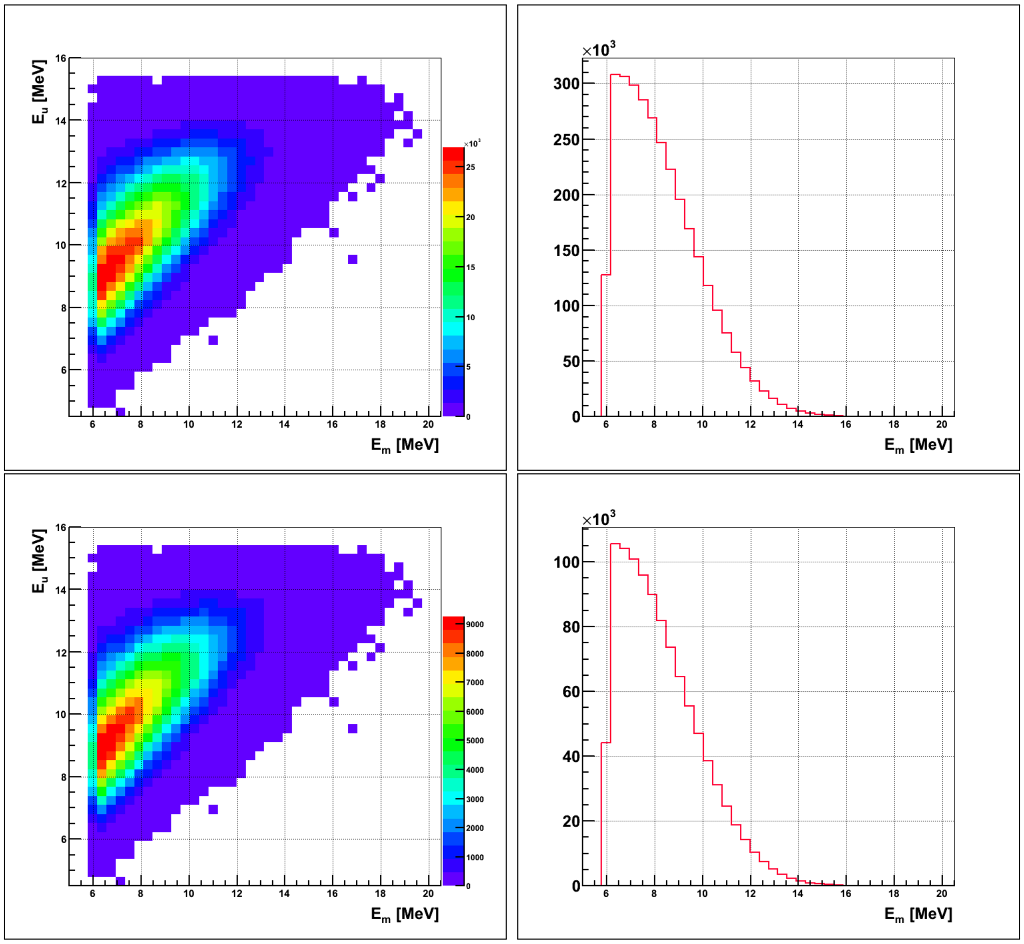


Figure 4.7: Top left is a 2D distribution of neutrino energy versus measured energy ($h_N(E_\nu, E_m)$) and the top right is $\int h_N(E_\nu, E_m)\, dE_\nu$. Bottom left is $h_N(E_\nu, E_m)$ weighted by $P_{ee_N}(E_\nu)$ ($h_N(E_\nu, E_m)P_{ee_N}(E_\nu)$) and the right plot shows $\int h_N(E_\nu, E_m)P_{ee_N}(E_\nu)dE_\nu$. The right 1D histograms are projections on E_ν of the 2D histograms on the left. The correlation of E_m with E_ν, as seen in this plot is taken into consideration when distorting the 3D PDFs. The reduction in the number of events – comparing the top plots to the bottom plots – is due to ν oscillations which is applied as distortions. In the 2D histograms, the number of events is shown in the color pallet.

Tabulated values used for $\frac{\sigma_\chi\,\epsilon_\chi}{\sigma_{nc}\,\epsilon_{nc}}$ (χ is CC, ES and $\mathrm{ES}_{\mu\tau}$) are listed in table 4.6.

Similar equations are used to calculate the number of events for ES and $\mathrm{ES}_{\mu\tau}$.

Class	**Cross Section Ratio**
cc	5603.875/240.569
es	481.596/240.569
$\mathrm{es}_{\mu\tau}$	74.833/240.569

Table 4.6: Values of $\frac{\sigma_\chi\,\epsilon_\chi}{\sigma_{nc}\,\epsilon_{nc}}$ (χ is CC, ES and $\mathrm{ES}_{\mu\tau}$) used in the MCMC fit. Values are from [90].

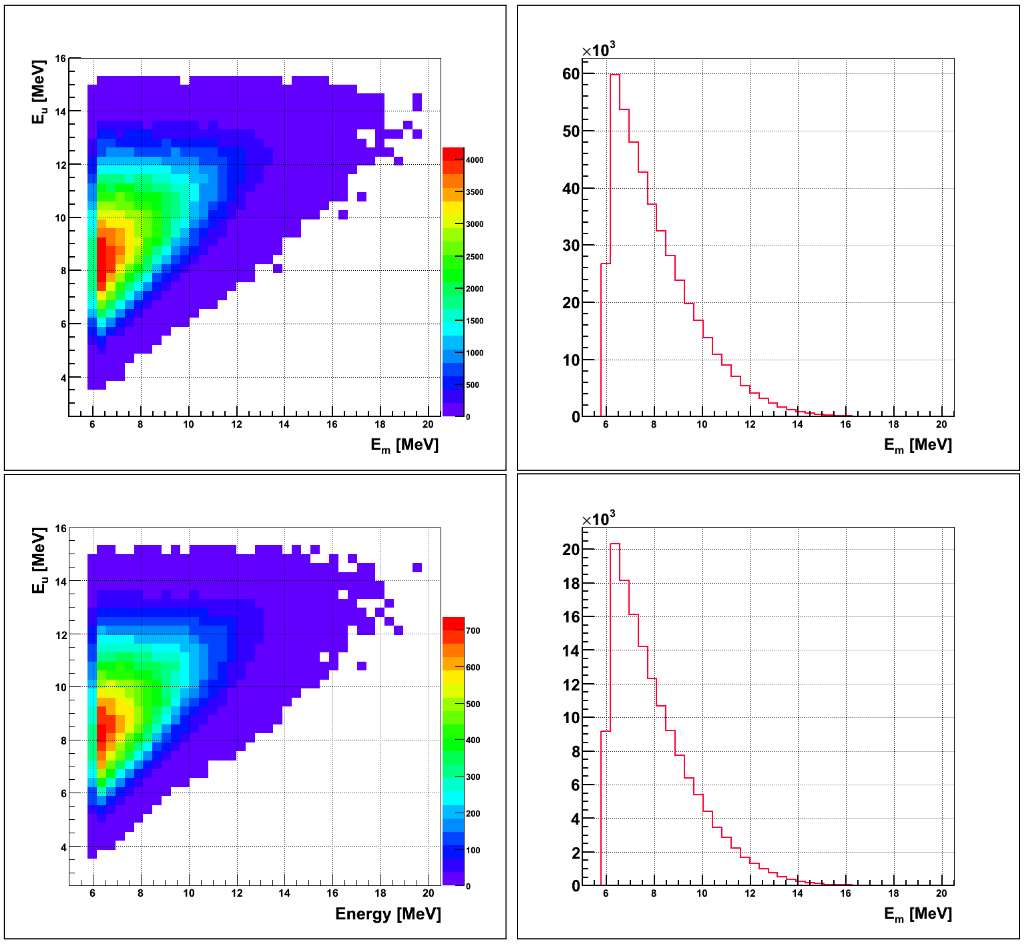


Figure 4.8: Top left is a histogram $h_D(E_\nu, E_m)$ of neutrino energy versus measured energy and the top right is $\int h_D(E_\nu, E_m)\, dE_\nu$. Bottom left is $h_D(E_\nu, E_m) P_{ee_D}(E_\nu)$ neutrino energy versus measured energy weighted by $P_{ee_D}(E_\nu)$ and the right shows the plot of $\int h_D(E_\nu, E_m) P_{ee_D}(E_\nu) dE_\nu$. The correlation of E$_m$ with E$_\nu$, as seen in this plot is taken into consideration when distorting the 3D PDFs. The reduction in the number of events – comparing the top plots to the bottom plots – is due to ν oscillations which is applied as distortions. In the 2D histograms, the number of events is shown in the color pallet.

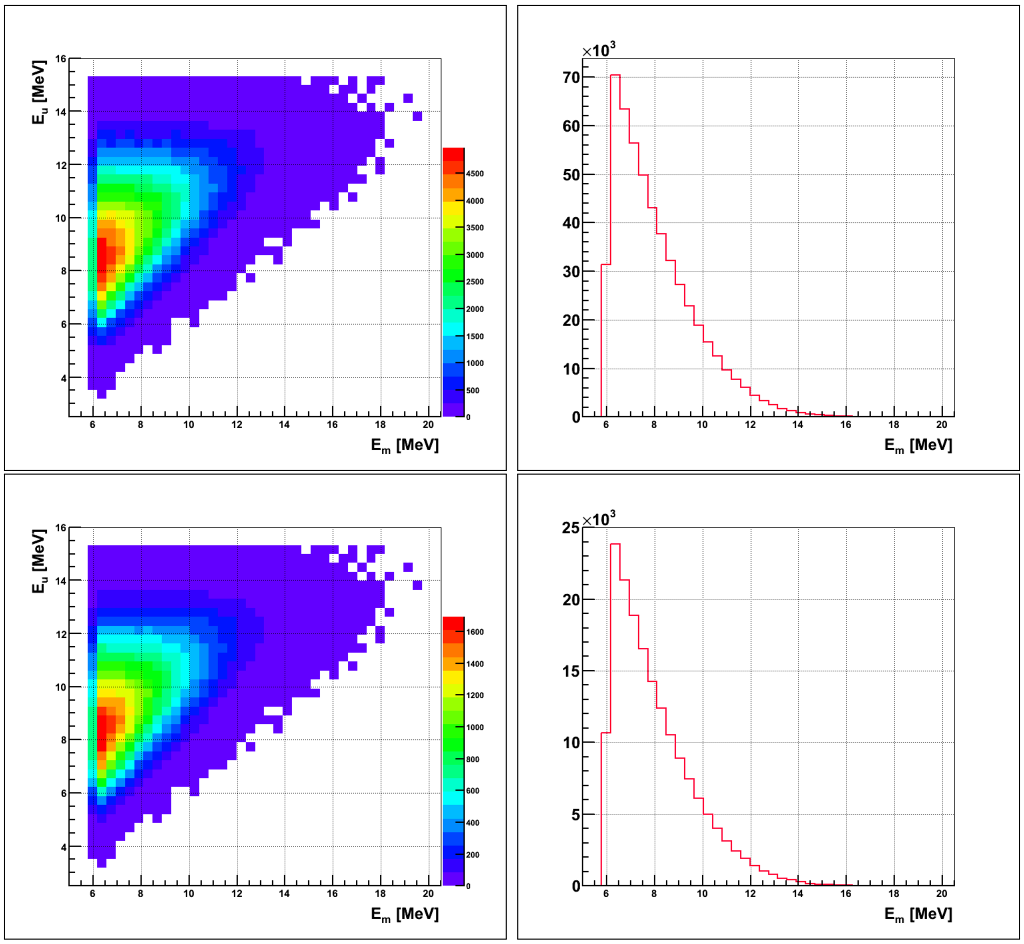


Figure 4.9: Top left is a histogram $h_N(E_\nu, E_m)$ of neutrino energy versus measured energy and the top right is $\int h_N(E_\nu, E_m)\, dE_\nu$. Bottom left is $h_N(E_\nu, E_m)P_{ee_N}(E_\nu)$ neutrino energy versus measured energy weighted by $P_{ee_N}(E_\nu)$ and the right shows the plot of $\int h_N(E_\nu, E_m)P_{ee_N}(E_\nu)dE_\nu$. The correlation of E$_m$ with E$_\nu$, as seen in this plot is taken into consideration when distorting the 3D PDFs. The reduction in the number of events – comparing the top plots to the bottom plots – is due to ν oscillations which is applied as distortions. In the 2D histograms, the number of events is shown in the color pallet.

4.10.8 Constraint from Low Energy Threshold Analysis of the D_2O and Salt Phases

The Low Energy Threshold Analysis (LETA) analysis is the simultaneous fit of two independent datasets from the first two phases of the SNO experiment. Since the flux of neutrinos from the Sun is assumed to be constant over the SNO operation time, the rate of neutrino events in the two phases can not vary independently. Full details of the LETA fit is found in the following references [39], [61], [73] and [74]. The constraint from LETA is applied as:

$$-\log \mathcal{L}_{\mathrm{LETA}} = ((\vec{\phi} - \vec{\phi}_{bf})^T/\vec{\sigma}^T)V^{-1}((\vec{\phi} - \vec{\phi}_{bf})/\vec{\sigma}) \tag{4.67}$$

where $\vec{\phi_{bf}}$ and $\vec{\sigma}$ are the 6×6 matrices containing the best-fit results and their uncertainties from a LETA fit and $\vec{\sigma}^T$ is the transpose[6] of the matrix $\vec{\sigma}$. The parameters constrained are ^{8}B scale, P_{ee} parameters (p_0, p_1, p_2), and day-night parameters (a_0 and a_1). The values of the above mentioned parameters from the MCMC fit, for the current step, is enclosed in the $\vec{\phi}$ matrix. The correlation among the parameters is represented in the LETA constraint equation (4.67) as an inverse of the correlation matrix $\vec{V}^{-1}$.

4.10.9 PSA Constraint

The pulse shape analysis (PSA References [71] and [72]) provides us with an independent measure of the number of neutrons derived from the NCD detector signals. This enables us the constrain the number of neutrons from the NCDs. The PSA constraint is applied as:

$$-\log \mathcal{L} = \frac{\left(\mathrm{PSA} - \sum_{i=1}^{m} N_i \kappa_i\right)^2}{(2\sigma^2)} \tag{4.68}$$

[6]The transpose of a $m \times n$ matrix $\mathbf{A}$ is another matrix , with n rows and m columns, designated as A^T such that $[A^T]_{ij}$=$[A]_{ji}$; the rows and columns of $\vec{A}$ matrix are switched in the $\vec{A}^T$ matrix.

where PSA is the central value of the PSA constraint and σ is its width, m is the number of neutron signals, N_i is the number of events for the neutron event type i for the current step in the MCMC fit and κ_i is the ratio that converts (N_i) the number of events belonging to the PMTs to the number of events belonging to the NCDs. Table 4.7 lists the values of κ_i. Except for neutral current where flux-to-event from the NCDs (f_{NC}^{NCD}) is a fit parameter, all other κ values are constant.

Neutron Signal Type	κ_i
Neutral current	^{8}B flux $\times$ f_{NC}^{NCD}
External neutrons	40.9/20.754
k2pd	32.8/9.402
k5pd	31.6/8.378
ncdpd	35.6/5.938
$d_2$opd	31.0/8.305
Atmos	13.6/24.681
hepNC	4.363

Table 4.7: Values of κ listed for various backgrounds [29]. Information used to apply PSA constraint on the number of neutron events from the NCDs.

4.11 Synopsis of the Calculation of the Number of Events

This section brings it all together, that is, the equations to calculate the number of events for each class. Each time an event passes the high level cuts, after applying the current systematic uncertainties, the number $N(\vec{\chi})$ is incremented by one. If it is a day event then $N_d(\vec{\chi})$ is incremented by one or else $N_n(\vec{\chi})$ is incremented by one. $N(\vec{\mathrm{def}})$ is the number of events that pass the default systematic uncertainties.

Each MC event, belonging to the backgrounds, gets weighted by a factor – W_i – which is calculated as:

$$W_i = 1.0 + c_0^{xyz}(a_0^E T_{0i} - 5.05\,\mathrm{MeV})$$

For CC, ES, ES$_{\mu\tau}$ and NC, the weight calculated in equation (4.69) is multiplied by another weight due to uncertainty in ^{8}B ν energy spectrum.

$$W_i' = W_i\big(1.0 + \frac{w}{3}(0.018 - 0.001999 \times E_i - 0.000088769 \times E_i^2)\big)$$

$$W_{nc} = \sum_{i=1}^{N_{nc}} \big(1.0 + \frac{w}{3}(0.018 - 0.001999 \times E_i - 0.000088769 \times E_i^2)\big)/N_{nc}$$

where W_i' is the weight applied to the PDFs belonging to CC, ES, ES$_{\mu\tau}$ and NC, $a_0^E T_i$ is the scaled reconstructed energy for the current event, c_0^{xyz} (energy dependent fiducial volume uncertainty) takes into account energy dependence of the reconstruction of a vertex of an event, w is the Winter uncertainty applied to CC, ES and NC, E_i is a neutrino energy for the current event i and N$_{nc}$ is the number of NC events that pass the high-level cuts after applying the current systematics. The expected number of neutral current events on the NCD-side of the NCD phase is modified by the Winter uncertainty using W_{nc}.

1. Equation to calculate the number of events for the backgrounds other than external neutrons and d$_2$opd.

$$N = \alpha N_{nom} \frac{N(\vec{\chi})}{N(\vec{\mathrm{def}})} \Big(\frac{N_d(\vec{\chi})}{N(\vec{\chi})} \frac{\sum_{i=1}^{N_d} W_i}{N_d(\vec{\chi})} + \frac{N_n(\vec{\chi})}{N(\vec{\chi})} \frac{\sum_{i=1}^{N_n} W_i}{N_n(\vec{\chi})} \Big) \tag{4.69}$$

2. Equation to calculate the number of events for the external neutrons and

$d_2$opd.

$$N = \alpha N_{nom} \frac{N(\vec{\chi})}{N(\vec{\text{def}})} \Big([1 - 0.5A] \frac{N_d(\vec{\chi})}{N(\vec{\chi})} \frac{\sum_{i=1}^{N_d} W_i}{N_d(\vec{\chi})} + [1 + 0.5A] \frac{N_n(\vec{\chi})}{N(\vec{\chi})} \frac{\sum_{i=1}^{N_n} W_i}{N_n(\vec{\chi})} \Big) \tag{4.70}$$

3. Equation to calculate the number of events for the neutral current:

$$N_{nc} =^8 B\, f2e \frac{N(\vec{\chi})}{N(\vec{\text{def}})} \Big(\frac{N_d(\vec{\chi})}{N(\vec{\chi})} \frac{\sum_{i=1}^{N_d} W_i'}{N_d(\vec{\chi})} + \frac{N_n(\vec{\chi})}{N(\vec{\chi})} \frac{\sum_{i=1}^{N_n} W_i'}{N_n(\vec{\chi})} \Big) \tag{4.71}$$

where f2e is a flux-to-event ratio for the ^{8}B flux detected by the PMTs. The number of NC events, for the PSA constraint, is calculated as:

$$N_{nc}^{\text{PSA}} = {}^8B\, f2e_{ncd}\, W_{nc} \tag{4.72}$$

where f2e$_{ncd}$ is a flux-to-event ratio for the ^{8}B flux detected by the NCDs and W$_{nc}$ is from equation (4.69). The effects of the application of systematics, to events from Monte Carlo detected by the PMTs, are not applied in the calculation of the number of events detected by the NCDs because the systematic uncertainties belonging to the PMTs and NCDs are uncorrelated.

Looking at the equations (4.69) to (4.71), it is clear that these equations can be simplified by cancelling out the terms that can be cancelled but the code was developed to calculate the ratios separately.

4. Following equations are used to distort the day and night PDFs of CC and ES.

$$\begin{aligned} P_{ee_D}(E_\nu) &= p_0 + p_1\,(E_\nu - 10\,\text{MeV}) + p_2\,(E_\nu - 10\,\text{MeV})^2 \\ A_{ee} \equiv \frac{2(\phi_N - \phi_D)}{(\phi_N + \phi_D)} &= a_0 + a_1\,(E_\nu - 10\,\text{MeV}) \\ P_{ee_N}(E_\nu) = \big(p_0 + p_1\,(E_\nu - 10\,\text{MeV}) &+ p_2\,(E_\nu - 10\,\text{MeV})^2\big)\,\frac{2 + A_{ee}}{2 - A_{ee}} \end{aligned}$$

where D stands for a day event and N stands for a night event. For $\mathrm{ES}_{\nu\tau}$ following equations are used:

$$P_{ee_D}(E_\nu) = 1 - \left(p_0 + p_1\,(E_\nu - 10\,\mathrm{MeV}) + p_2\,(E_\nu - 10\,\mathrm{MeV})^2\right)$$
$$P_{ee_N}(E_\nu) = \left(1 - (p_0 + p_1\,(E_\nu - 10\,\mathrm{MeV}) + p_2\,(E_\nu - 10\,\mathrm{MeV})^2)\right)\frac{2 + A_{ee}}{2 - A_{ee}}$$

Equation to calculate the number of charged current (CC) events using the ^{8}B flux, f2e and the parameters of the survival probability equation is:

$$N_{cc} = {}^8B\,f2e\,\frac{\sigma_{cc}\epsilon_{cc}}{\sigma_{nc}\epsilon_{nc}}\frac{N(\vec{\chi})}{N(\vec{\mathrm{def}})}\left(p_{ee_d}\,\frac{N_d(\vec{\chi})}{N(\vec{\chi})}\frac{\sum_{i=1}^{N_d} W_i'}{N_d(\vec{\chi})} + p_{ee_n}\,\frac{N_n(\vec{\chi})}{N(\vec{\chi})}\frac{\sum_{i=1}^{N_n} W_i'}{N_n(\vec{\chi})}\right) \tag{4.73}$$

where p_{ee_d} and p_{ee_n} are calculated as:

$$p_{ee_d} = \frac{\iint h_D(E_\nu, E_m)\,P_{ee}(E_\nu, day) dE_\nu\, dE_m}{\iint h(E_\nu, E_m) dE_\nu\, dE_m}$$
$$p_{ee_n} = \frac{\iint h_N(E_\nu, E_m)\,P_{ee}(E_\nu, night) dE_\nu\, dE_m}{\iint h(E_\nu, E_m) dE_\nu\, dE_m}$$

Similar equations are used for ES and $\mathrm{ES}_{\mu\tau}$.

5. For the PSA constraint, the number of events is calculated as:

$$N = \alpha N_{nom}\left[(1 - 0.5A)\frac{N_d(\vec{\chi})}{N(\vec{\chi})} + (1 + 0.5A)\frac{N_n(\vec{\chi})}{N(\vec{\chi})}\right] \tag{4.74}$$

The day-night asymmetry A is zero for all the backgrounds except the external neutrons and the $d_2$opd.

4.12 Evaluating Fit Biases

The purpose of the ensemble (multiple fake datasets) test is to undertake a pull and bias study. Each fake dataset corresponds to a data from an experiment. When an experiment is repeated multiple times (multiple fake datasets), there is statistical uncertainty in data. For example, if there is a constraint from an external measurement in each experiment then the constraint will have a normal distribution too. If this fact is taken into consideration when fitting N datasets then the pull distribution should tend toward a normal distribution, as shown in figure 4.10, with a mean of 0 and width of 1.0. A correct pull distribution demonstrates that the fit is unbiased and the error-estimation procedure is accurate.

This section describes how the pull and bias of a fit is evaluated. Several fits were performed on multiple fake datasets in the process of development of the MCMC code and for each fit, pull and bias study was carried out.

$$\text{Bias}(x) = \frac{N(x) - E(x)}{E(x)} \tag{4.75}$$

$$\text{Pull}(x) = \frac{N(x) - E(x)}{\sigma(x)} \tag{4.76}$$

where $N(x)$ is the fitted value, $E(x)$ is the expected value and $\sigma(x)$ is the uncertainty of the fitted value. From the bias equation, the bias is the fractional shift in the fit result from the expected value. From the computation of bias we examine whether the fit result agrees with the Monte Carlo input and the pull of the fit is computed to examine whether the error generated by the fit agrees with the spread of the fit result. For the unbiased fit, both the pull and bias should be distributed around zero and the pull should have RMS of one.

The bias and pull of a parameter x is computed from each fake data file from the ensemble and the resulting distributions are fitted to a Gaussian function described in equation (4.77). The mean (μ) and variance from the Gaussian function is used to calculate the average pull and bias of the parameter x according to equations (4.78) and (4.79). The drawn error bars in the bias plots indicate the uncertainty on the average bias which corresponds to the sample standard deviation of the test parameter $\mathbf{x}$ divided by the square root of the number of samples $\sigma/\sqrt{N}$. The drawn error bars on the pull plots show the sample standard deviation σ of the test parameter $\mathbf{x}$ and not the

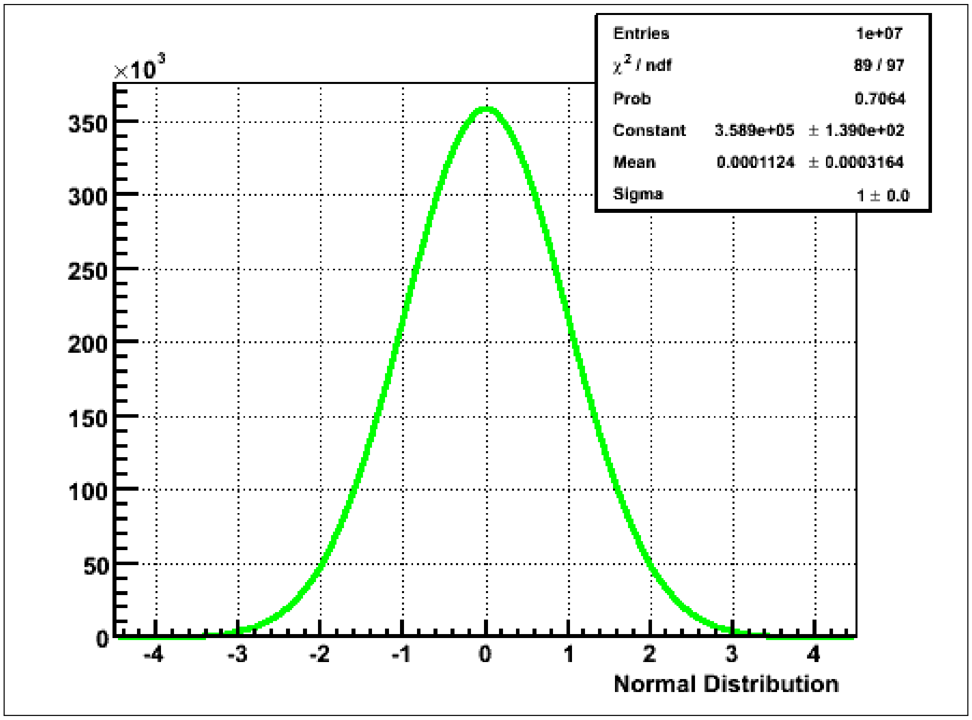


Figure 4.10: Figure shows the normal distribution ($f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$) with a mean ($\mu$) of zero and a width ($\sigma$) of one.

uncertainty on the average pull.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \tag{4.77}$$

$$Bias(x) = \mu \pm \frac{\sigma}{\sqrt{N}} \tag{4.78}$$

$$Pull(x) = \mu \pm \sigma \tag{4.79}$$

4.13 Summary

The MCMC fit perform the fit on Čerenkov data of the NCD phase of SNO and add data from NCDs of the NCD phase as a constraint from pulse shape analysis and incorporate data from the first two phases of SNO as constraints from the low-energy threshold analysis. Signal extraction extracts a ^{8}B flux and a set of polynomial parameters to describe the day time survival probability (P_{ee_D}) and a set of polynomial parameters to attribute the asymmetry A_{ee}

in the day and night survival probability. The negative log likelihood (NLL) function, after adding all the constraints, is:

$$
\begin{aligned}
-\log\mathcal{L} = \sum_{i=1}^{2m} N_i - \sum_{d=1}^{N} \log\big(\sum_{i=1}^{2m}(N_i)F_i(\vec{x}_d,\vec{P})\big) \\
+\frac{(\mathrm{PSA} - B\,\epsilon_1 - n_1\kappa_1 - n_2\kappa_2 - n_3\kappa_3 + n_4\kappa_4 + n_5\kappa_5 + n_6\kappa_6 - 4.363)^2}{2\,(\sigma)^2} \\
+\frac{(\overline{\alpha}_1-\alpha_1)^2}{2\sigma_1^2}+\frac{(\overline{\alpha}_2-\alpha_2)^2}{2\sigma_2^2}+\frac{(\overline{\alpha}_3-\alpha_3)^2}{2\sigma_3^2}+\frac{(\overline{\alpha}_4-\alpha_4)^2}{2\sigma_4^2} \\
+\frac{(\overline{\alpha}_5-\alpha_5)^2}{2\sigma_5^2}+\frac{(\overline{\alpha}_6-\alpha_6)^2}{2\sigma_6^2}+\frac{(\overline{\xi}_0-\xi_0)^2}{2\sigma_{\xi_0}^2}+\frac{(\overline{\xi}_1-\xi_1)^2}{2\sigma_{\xi_1}^2} \\
+\frac{(\overline{\epsilon}_1-\epsilon_1)^2}{2\sigma_{\epsilon_1}^2}+\frac{(\overline{\epsilon}_0-\epsilon_0)^2}{2\sigma_{\epsilon_0}^2} \\
+\frac{1}{2}\sum_i(\frac{p_i-\bar{p}_i}{\sigma_{p_i}})^2+\frac{1}{2}\sum_{i=0}^{2}\sum_{j=0}^{2}(b_i^{xy}-\bar{b_i^{xy}})(b_j^{xy}-\bar{b_j^{xy}})(V_{b^{xy}}^{-1})_{ij} \\
+\frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}(b_i^{z}-\bar{b_i^{z}})(b_j^{z}-\bar{b_j^{z}})(V_{b^{z}}^{-1})_{ij} \\
+((\vec{\phi}-\vec{\phi}_{bf})^T/\vec{\sigma}^T)V^{-1}((\vec{\phi}-\vec{\phi}_{bf})/\vec{\sigma})
\end{aligned} \quad (4.80)
$$

where $\mathrm{F}_i(\vec{x}_d,\vec{P})$ is the probability density function, for the class i, giving the probability of observing the event d with observables $\vec{x}_d$ and the current values of the fit parameters $\vec{P}$, $N_1, N_2, \ldots, N_m$ are the number of events for **m=13** event classes and $n_1, n_2, \ldots, n_6$ are the number of events computed using equation (4.74), for the calculation of PSA constraint. The ^{8}B flux is designated by **B** and **PSA** is the PSA constraint for the current fake dataset and σ is the width of the constraint. The values of κ for various backgrounds is listed in table 4.7. The average number 4.363 is the number of NC interactions from the hep neutrinos (^{3}He+p $\rightarrow$ ^{4}He + e^+ + ν_e) expected to be detected in the NCDs. Number of interactions in SNO from hep neutrinos are fixed in the MCMC fit. In the constraint terms, α_1 to α_6 are the values of α (equation (4.19)) in the current MCMC step for EX, $d_2$opd , atmospheric neutrinos, k2pd, k5pd and ncdpd respectively. The $\overline{\alpha}_1$, $\overline{\alpha}_2$, $\overline{\alpha}_3$, $\overline{\alpha}_4$, $\overline{\alpha}_5$ and

$\overline{\alpha}_6$ are the constraints for EX, $d_2$opd , atmospheric neutrinos, k2pd, k5pd and ncdpd respectively for a given fake dataset. The day-night asymmetries for the external neutrons and D2OPD are represented by ξ_0 and ξ_1 respectively. The flux-to-event for the NCDs and PMTs are represented by ϵ_1 and ϵ_0 respectively. The PMT and NCD neutron capture efficiencies in the NCD phase appear as a component to the flux-to-event conversion factors. In the likelihood equation, p_i, $\bar{p}_i$ and σ_{p_i} represent the current value of the PMT systematic parameter i in the MCMC fit, its mean and constraint width respectively. The next two terms are calculation of the constraint for the systematic uncertainties that are correlated. For more details, refer to 4.10.4. The $\vec{\phi}_{bf}$ and $\vec{\sigma}$ are the matrices containing the best-fit results and their uncertainties from a LETA fit, $\vec{\sigma}^T$ is the transpose of matrix $\vec{\sigma}$ and values of the parameters, constrained by LETA, for the current step in the MCMC fit is enclosed in the $\vec{\phi}$ matrix. The matrix $\vec{V}$ is the correlation of the parameters provided by the LETA fit.

The fit for datasets generated using the full Monte Carlo is described in chapters 8, 9 and 10 and the fit for datasets generated using the third of the Monte Carlo is described in chapter 11.

The code, used in the extraction of fit parameters from the data, was developed in a series of steps and for each step, the code was tested on an ensemble of fake datasets. The MCMC code was originally created by Juergen Wendland and later expanded by Blair Jamieson from whom I inherited the code. I expanded the code to include survival probability equations, day-night asymmetries to account for matter effect on neutrino oscillations and the day-night asymmetries in the backgrounds (EX and $d_2$opd) to account for the possible variations in the detector response with time of the day and day-night asymmetries of various background sources during the day versus during the night. Besides the diurnal asymmetries, directional asymmetries were also added to account for possible up-down asymmetries in the detector.

Various new penalties were added along the line to reduce the uncertainties from the MCMC outcome, which included using the constraints from the Low Energy Threshold Analysis (LETA) of the data from the D_2O and Salt phases of SNO and Pulse Shape Analysis (PSA) of the signals from the NCDs of the NCD phase. During various stages of development, several tests were created to cross-check the validity of the code and to flush out the bugs, if any, lurking in the code. The progress was not like climbing a stairway to code heaven; more likely a snake and ladders game. A bug, implanted while in pursuit of another bug, can take weeks of investigation but nevertheless a fruitful effort to understand a complicated code. The thesis is not a history of the work done to reach the final goal but a description of various milestones.

Chapter 5

Markov Chain Monte Carlo Method

5.1 Introduction

This chapter covers the algorithm employed to extract the fit parameters from the data. Before delving into the details of the algorithm, a brief definition of the terms used in this chapter is in order. The definitions of terminology, in this chapter, are based on information gained from Wikipedia especially

[http://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo]

which should be referred to for more detail.

Monte Carlo Simulation: Calculations based on the application of random numbers are generally known as Monte Carlo (MC) simulations. The MC approach is not limited to the calculation of probabilities but can also be used to calculate the integral of complicated functions. In experimental particle physics, MC simulation is used for designing detectors, understanding their behaviour and comparing data to theory. The following list, although by no means complete, illustrate the diversity of the application of MC simulation: nuclear reactor design, radiation cancer therapy, traffic flow, stellar evolution, oil well exploration and Dow Jones forecasting.

Random Walk: Random walk consist of taking random steps where the probability of taking a step in any direction is equal and not influenced by the previous steps. Movement of pollen grains in a glass of water is an example of random motion known as Brownian motion. Other examples include diffusion of dye in an unstirred glass and fluctuation in price of a stock.

Random Process: A system undergoing a discrete random process means that the system will be at random states at different steps. The steps are often thought of as time, but formally the steps are just integers, and the random process is a mapping of steps to states. The change in the state from one step to the next is called a transition, and a probability associated with each transition in the state is called a transition probability.

Markov Chain: A Markov chain is named after Andrey Markov. The chain is generated by using the current sample values to randomly propose the next sample values [80]. Given its current state, the transition probability of the future step depends only on the current state of the system. Mathematically it is described as:

$$P(X_{n+1}|X_1, X_2 \ldots, X_n) = P(X_{n+1}|X_n) \tag{5.1}$$

Burn-in: The practise of throwing away some number of iterations at the start of an MCMC run when MCMC has not yet converged is known as burn-in. This practise is a necessity because the choice of initial values of the parameters is independent of likelihood. The chain can start in a region of very low likelihood and then walk to the region of highest likelihood. For a converged chain, the chain stays in the region around the highest likelihood.

Markov chain Monte Carlo simulation is a class of algorithms to simulate a random walk for the purpose of sampling through probability distributions. The random walk is undertaken long enough to ensure as complete tour through the likelihood distribution as possible considering limitations of computational resources. The purpose of taking a large number of random steps is to achieve an equilibrium distribution after removing the burn-in period. MCMC is one of the simulation technique to explore high-dimensional probability distributions by generating statistically consistent samples from the target distribution. One of the goal of MCMC is computation of the expected values of fit parameters. The expectation values are calculated with samples drawn more proportionally from higher likelihood regions.

The validity of MCMC depends critically on the rate of convergence to an equilibrium distribution. Constructing a Markov chain with the desired properties is not difficult. The challenge is determining the number of steps required to achieve an equilibrium distribution. For a complicated code, the number of steps is limited by the computational resources. A chain with a rapid mixing is a good chain because it will quickly attain an equilibrium distribution starting from any arbitrary state. The Metropolis-Hastings algorithm is a method to generate a Markov chain using a proposal density Q(x^{t+1}, x^t) which depends on the current state x^t, to generate a new proposal state x^{t+1}. This proposal is "accepted" as the next value ($x^{t+1} = x'$) if α (a random number between 0 and 1) satisfies:

$$\alpha < \min\Big(\frac{P(x')Q(x^t; x')}{P(x^t)Q(x'; x^t)}, 1\Big) \text{ [80]} \tag{5.2}$$

The current value of x is retained ($x^{t+1} = x^t$) if the proposal is rejected. Any probability distribution P(x) can be used by the Metropolis-Hastings algorithm to draw samples. For the proposal density, a Gaussian function centred on the current state x^t: Q$(x'; x^t) \sim N(x^t, \sigma^2 I)$, may be used. This proposal density generates samples centred around the current state x^t with variance σ^2. The Metropolis algorithm requires $Q(x; y) = Q(y; x)$ – a symmetric function. In that case, equation (5.2) is reduced to:

$$\alpha < \min\Big(\frac{P(x')}{P(x^t)}, 1\Big) \tag{5.3}$$

because

$$\frac{Q(x^t; x')}{Q(x'; x^t)} = 1 \tag{5.4}$$

In MCMC, we assume the proposed density to be symmetric. The challenges in MCMC fitting are:

1. **Determine whether a MCMC fit has converged.**

   A plot of the log likelihood versus steps is one of the methods to find out if the fit has converged. An example is shown in figure 5.1. To make sure that the MCMC fit has converged and is independent of the opening values of the parameters, each MCMC chain was started at different random initial values of the parameters. The range of the parameters was within $\pm 3\sigma$ of the mean (expected values from a theoretical model) where σ is the constraint on the mean from an independent measurement or uncertainty on the mean which was selected after several trial and errors to ensure good acceptance in the MCMC fit. A check was carried out to test the robustness of the code for which the initial values of the parameters were selected within $\pm 10\sigma$ instead of $\pm 3\sigma$. The check is described in section 10.3. Since the check demonstrated that convergence was achieved within 4000 steps, the initial values were selected within $\pm 3\sigma$ instead of $\pm 10\sigma$ because of computational limitations. The convergence of a Markov chain is different from the **Maximum Likelihood Estimation (MLE)** because the former does not convergence to an estimate like the latter but instead calculates the probability distribution of the value of a parameter in the volume of high likelihood.

   Following are the convergence issues: when has the chain moved from its starting values and started sampling from its stationary distribution? and how large a sample is required for obtaining estimates within acceptable accuracy? Autocorrelation, described in section 5.2, is an important

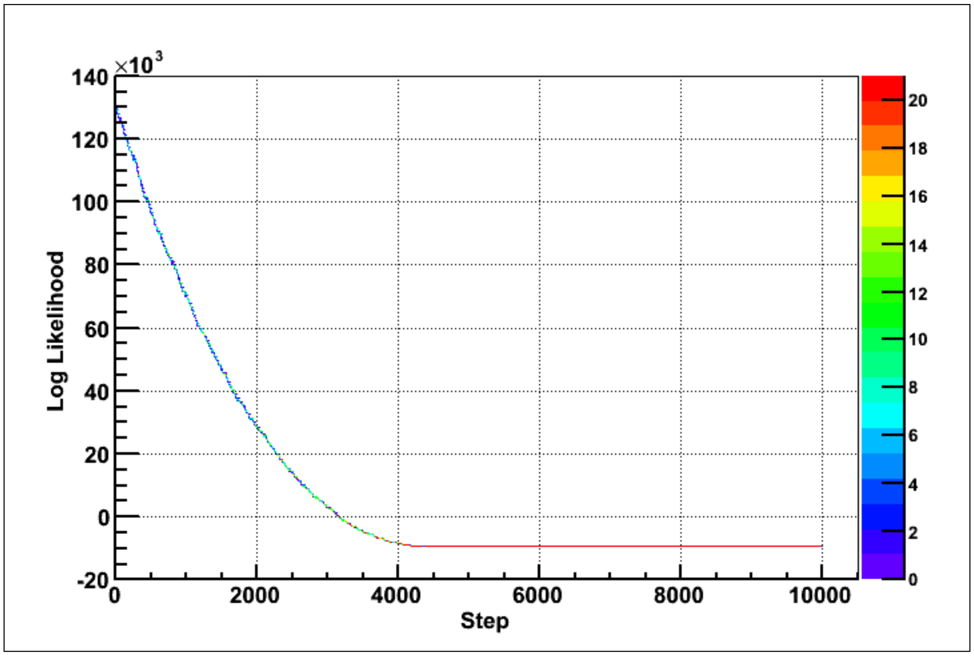


Figure 5.1: Log Likelihood (shown on the vertical axis) versus various time steps (shown on the horizontal axis). This plot shows that the MCMC fit has converged around 4000 steps.

measurement in the consideration of the chain length. A badly mixed chain will have a high autocorrelation and will need a longer run time to give an estimation of fit parameters to the required accuracy.

2. **Determine a burn-in period.**

   See section 5.2 on how to select the burn-in period.

3. **Step size of the MCMC fit**

   If the step size (σ^2 – of a Gaussian function for the proposal density function) is too narrow then MCMC will not sample enough of the parameter space to find the best fit values in a finite number of steps; if the step size is too wide, steps are rarely accepted resulting in a low acceptance rate because the proposals most likely populate regions of much lower probability density. The ideal acceptance rate for an N-dimensional Gaussian proposal density function is $\approx 23\%$ [81].

4. **Extracting values from the Posterior Distribution Functions** After a posterior distribution is determined, it typically goes through a "post processor". Post-processing is performed to determine the fit values and their uncertainties from the posterior distributions. Typical pull and bias studies are carried out to test the behaviour of the fit. Various techniques are available to pick the best fit:

    (a) Use the mean and Root Mean Square (RMS) of a posterior distribution as the estimation of the parameter and its uncertainty respectively.

    (b) Fit a Gaussian function to the posterior distribution and use the mean and sigma ($\mu \pm \sigma$) of the function as the estimation of the parameter and its uncertainty respectively.

    (c) The mode of a continuous posterior distribution is a value $\mathbf{x}$ at which the distribution is at its peak. The peak ($\sim$ mode) and RMS of the posterior distribution were considered as the estimation of the parameter and its uncertainty respectively. The mean, median and mode belong to a normal distribution all coincide.

For a history of MCMC and a list of useful references pertaining to MCMC, refer to [82]. MCMC methods were introduced in the 1950s to efficiently sample an unknown probability distribution. The time needed in MCMC, to sample a distribution, grows approximately linearly rather than scaling exponentially with the number of parameters varied. For this reason, MCMC methods are particularly useful to evaluate integrals in many dimensions. Examples of applications in physics include estimation of cosmological parameters [83] and for analysing the orbits of extrasolar planets [84].

5.2 Autocorrelation Function

Autocorrelation is one of the convergence diagnostics used to determine how many initial steps should be discarded from the output of the MCMC fit such that the remaining samples represent the target distribution of interest and how many steps are necessary in the chain.

Given measurements $Y_1, Y_2 \ldots Y_N$ at time $X_1, X_2 \ldots X_N$ the correlation between observables separated by k time steps is:

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2} \text{ from [85]} \tag{5.5}$$

where $\bar{Y} = \sum_{i=1}^{N} \frac{Y_i}{N}$. Autocorrelation is a correlation coefficient between two values of the same variable Y_i and Y_{i+k} at X_i and X_{i+k} where k is the lag because one of the pair of observations (Y_i) lags the other (Y_{i+k}) by **k** periods or samples. Autocorrelation is a tool to find the degree of relationship of a signal with itself at different times. An example is shown in figure 5.2. The value of r_k lies between -1 and 1, with $+1$ indicating perfect correlation and -1 indicating perfect anti-correlation. Positive autocorrelation is a sign of "persistence" which is a tendency of a system to frequently return to the same state. Autocorrelation implies that a time series is predictable, as future values are correlated with current and past values. This behaviour reduces the effective sample size. The quantity r_k is known as the autocorrelation coefficient at lag k.

If the posterior distribution is not random in time then the information in each observation is not totally independent from the information in other observations.

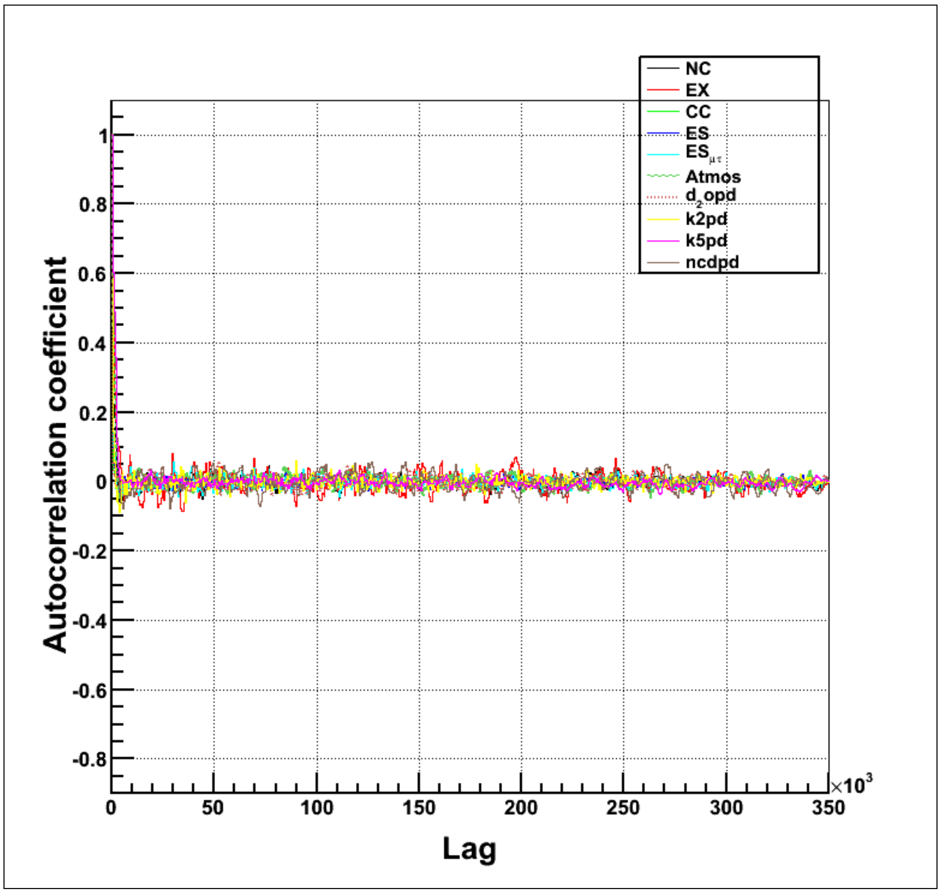


Figure 5.2: Plot showing autocorrelation coefficient versus lag. This is an example of autocorrelation function applied to a MCMC fit. The parameters of the MCMC fit, shown in the legend, are the parameters that we use for signal extraction in SNO, and are defined explicitly in chapter 4. Selecting the burn-in from autocorrelation plot is a judgement call based on experience.

5.3 SIGnal EXtractrion (SIGEX) with MCMC

MCMC fit is based on Bayesian analysis which require a joint probability P(D, θ) consisting of a prior P(θ) and a likelihood P(D$|\theta$):

$$P(D, \theta) = P(D|\theta)P(\theta) \tag{5.6}$$

where D denote the observed data and θ denote model parameters. Bayes' theorem [80] is used to determine the probability distribution of θ conditional on D.

$$P(\theta|D) = \frac{P(\theta)P(D|\theta)}{\int P(\theta)P(D|\theta)d\theta} \tag{5.7}$$

$$P(\theta|D) \propto P(\theta)P(D|\theta \tag{5.8}$$

P$(\theta|D)$, the object of all Bayesian inference, is a distribution of unknown θ given the known D. It is known as posterior because it is obtained after the data is observed. Frequentists employ the P$(D|\theta)$ distribution and Bayesian utilize the P$(\theta|D)$ distribution for signal extraction.

A MCMC, based on Metropolis algorithm [79], is utilized for estimating parameters and uncertainties on the parameters. Traditionally, SNO has minimized Negative Log Likelihood (**NLL**) function – via MINUIT [86] – with respect to all the parameters for the purpose of obtaining the best fit values of the parameters. The curvature of $-\log(\mathcal{L})$ at the minimum was used in the calculation of the uncertainties on the parameters. Minimizing the **NLL** is very challenging with so many parameters because the likelihood function is based on binned Monte Carlo and is choppy everywhere which makes the minimum ill-defined. The likelihood function, used in MCMC, is explained in detail in chapter 4. MCMC generates random samples, of possible values of all the fit parameters, drawn from the joint probability distribution described in equation (5.7). The parameters of interest are determined by integrating over all nuisance parameters.

The algorithm is to take a random walk through a parameter space, that is, propose the values of the fit parameters for the next (n+1) step $\vec{x}_{\text{prop}}$ using the last accepted values $\vec{x}_n$.

$$\vec{x}_{prop} = \vec{x}_n + \vec{\epsilon} \tag{5.9}$$

where $\vec{\epsilon} = N(0, \frac{\sigma}{3})$, a Gaussian of mean zero and width that is roughly 1/3 of the expected statistical uncertainty or the constraint uncertainty. The value 1/3 is chosen to make the acceptance rate $\approx$ 25%.

The probability of each step is calculated as: $\frac{\mathcal{L}_{prop}}{\mathcal{L}_{curr}}$ where $\mathcal{L}_{prop}$ and $\mathcal{L}_{curr}$ are the likelihoods of the proposed step and the current step respectively. The acceptance probability is $\min\big(1, \frac{\mathcal{L}_{prop}}{\mathcal{L}_{curr}}\big)$ which is compared to a random number (α) between 0 and 1. The step is accepted if α does not exceed the acceptance probability (equation (5.10)); if the step is accepted, the parameter values are updated (equation (5.11)) or else the current point in the chain is retained to make the next proposal.

$$\alpha \leq \min\big(1, \frac{\mathcal{L}_{prop}}{\mathcal{L}_{curr}}\big) \tag{5.10}$$

$$\vec{x}_{n+1} = \vec{x}_{prop} \text{ If accepted} \tag{5.11}$$

Using this methodology, resulting distribution of parameters from the chain will have a frequency distribution given by $\mathcal{L}$.

5.4 Prior in the MCMC fit

The prior P(θ) in MCMC fit are the constraints from low-energy threshold (LETA) analysis of combined first two phases of SNO, pulse shape analysis (PSA) of the data from NCDs of the NCD phase, constraints applied on backgrounds from external measurements and constraints applied on systematic uncertainties from calibration sources and last but least the number of events were constrained to be positive.

5.5 Comparison of the first half to the second half of the posterior distribution

Various measures exist to indicate the degree of convergence of the MCMC code. In addition to the autocorrelation function and the likelihood versus step, discussed in section 5.2, another method is to plot the first and second half of the posterior distribution, after removing the burn-in, on the same diagram. If the distributions are similar, then the MCMC fit has converged. Some of the first and second halves are displayed for the parameters of the MCMC fit. These parameters were explained in chapter 4. In figures 5.3 to 5.5, the distributions shown in red and blue are very similar which is an indication that the fit converged.

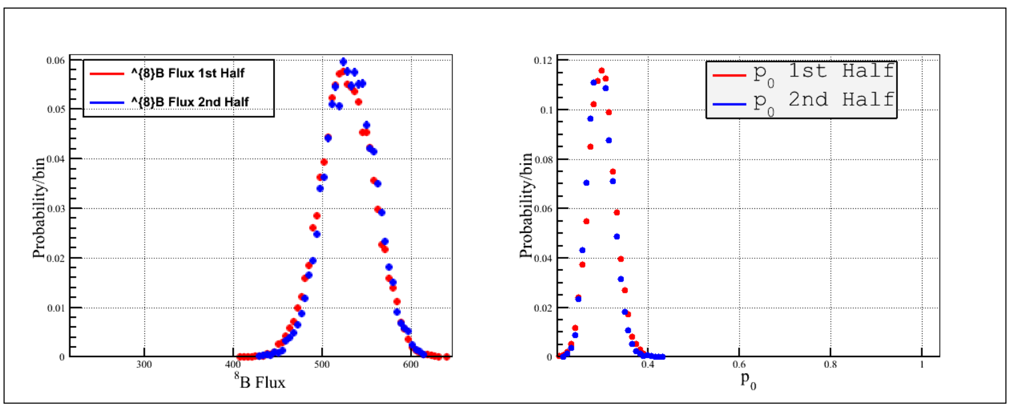


Figure 5.3: Comparing the distribution of the first half (shown in red) and the second half of the MCMC fit (shown in blue), after removing the burn-in period, for two parameters (labelled ^{8}B flux and p_0) in a MCMC fit.

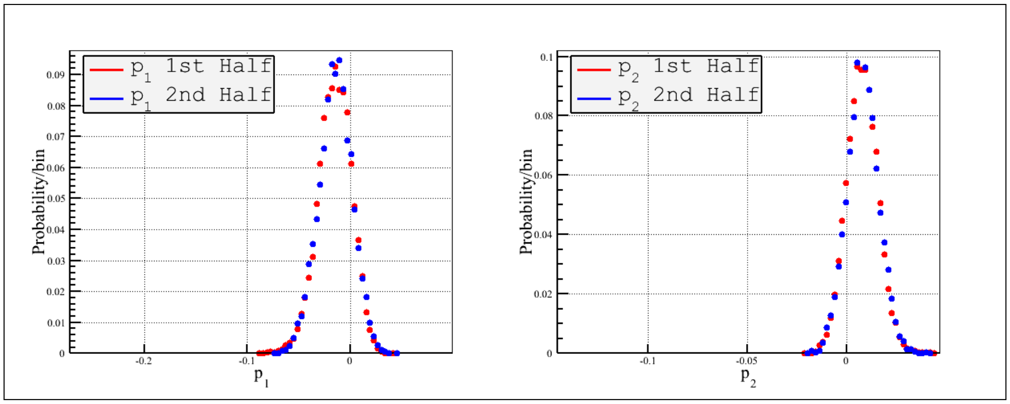


Figure 5.4: Comparing the distribution of the first half (shown in red) and the second half of the MCMC fit (shown in blue), after removing the burn-in period, for two parameters (labelled p_1, and p_2) in a MCMC fit.

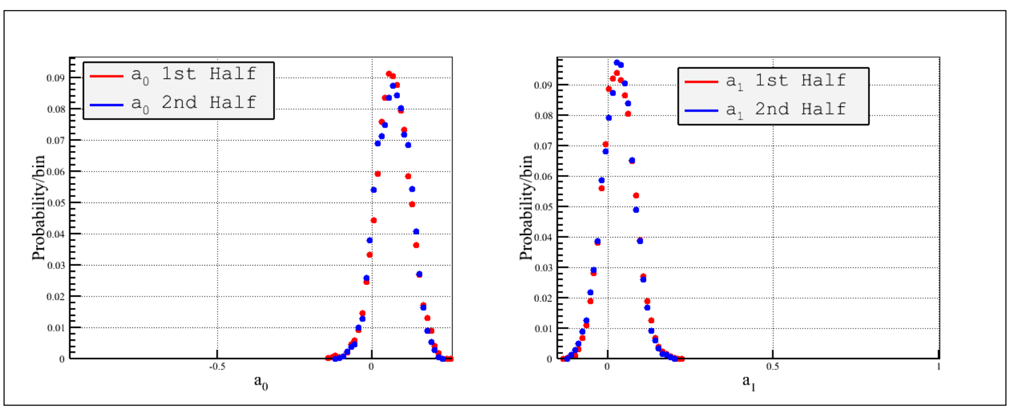


Figure 5.5: Comparing the distribution of the first half (shown in red) and the second half of the MCMC fit (shown in blue), after removing the burn-in, for day-night asymmetry parameters a_0 and a_1.

Chapter 6

Mean or Centroid: Does It Matter in the Fit?

6.1 Introduction to the Fit

The fitting procedure for SNO starts with an extended maximum likelihood fit to untangle **m** different signals; each with its normalized PDF ($F(\vec{x}_d)$) defined for some observable vector $\vec{x}_d$. The log likelihood function is defined as:

$$\mathcal{L} = \sum_{d=1}^{N_D} \log\big(\sum_{i=0}^{m} \mu_i F_i(\vec{x}_d)\big) - \sum_{i=0}^{m} \mu_i \tag{6.1}$$

where **m** is the number of signals in the data and N_D is number of observed events in the data. The parameters varied in the fit, μ_i, are the expected number of events for each signal type i. The PDF $F_i(\vec{x}_d)$ for signal type i is used to calculate the probability of measuring an event with observable values $\vec{x}_d$ to be of signal type i. Besides signal extraction fit to extract the number of events belonging to the backgrounds, independent measurements were carried out, such as low level radio assays of U and Th decay chain products in D_2O and H_2O, to measure the backgrounds in these regions. Any constraints from independent measurements are added as penalties to the likelihood function. Assuming a Gaussian distribution for an uncertainty on the external measurement, the penalty is imposed as:

$$\exp\big(\frac{-(N-\tilde{N})^2}{2\,\sigma^2}\big) \tag{6.2}$$

where $\tilde{N}$ is a central value and $\tilde{\sigma}$ is an uncertainty on the central value and N is the number of events in a current step of the MCMC fit. Taking the log of equation (6.2) and inserting in equation (6.1) results in:

$$\mathcal{L} = \sum_{d=1}^{N_D} \log\big(\sum_{i=0}^{m} \mu_i F_i(\vec{x_d})\big) - \sum_{i=0}^{m} \mu_i - \frac{(N-\tilde{N})^2}{2\,\sigma^2} \tag{6.3}$$

6.2 Mean or Centroid

The normal or Gaussian distribution is a continuous probability distribution function that describes data that cluster around the mean of the distribution. The function is described mathematically as:

$$f(x) = \frac{1}{\sqrt{2\,\pi\sigma^2}}\,\exp(-\frac{(x-\mu)^2}{2\,\sigma^2}) \tag{6.4}$$

Integrating the above equation results in:

$$\int f(x)dx = 1 \tag{6.5}$$

where μ is the centroid and σ^2 is the variance of the function. The associated graph is bell-shaped with a peak at the centroid. The mean of the function is defined as:

$$\langle x\rangle = \frac{\int x\,f(x)dx}{\int f(x)\,dx} \tag{6.6}$$

where the integrals are taken over the domain of the function $f(x)$. For the Gaussian function, the centroid and the mean are equivalent as shown:

$$\bar{x} = \frac{\int_{-\infty}^{+\infty} x\,\exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)dx}{\int_{-\infty}^{+\infty} \exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big)dx} \tag{6.7}$$

$$\bar{x} = \frac{\sigma\,\sqrt{(2\pi)}\,\mu}{\sigma\,\sqrt{(2\pi)}}$$

$$\bar{x} = \mu \tag{6.8}$$

The domain is from $+\infty$ to $-\infty$. In a complex experiment like SNO, several variables are measured independently of the main experiment. In SNO, the number of external neutrons was measured independently and fitted to a Gaussian function. The Gaussian function yielded 40.9 ± 20.6 as the number of external neutrons from the NCDs. The number of external neutrons measured from the PMTs of the NCD phase was calculated to be 20.6 ± 10.4. This information is applied as a constraint in a fit. Before fitting the actual data, the reliability of the fit is tested on a number of fake data sets. Each fake data set corresponds to an experiment along with its own set of independent measurements and each independent measurement is fitted to a Gaussian function which yields a mean and an uncertainty on the mean ($\mu \pm \delta\mu$). According to the Central Limit Theorem [89], the distribution of the means from each independent measurement should follow a Gaussian distribution. Hence for each fake data set, the constraint is selected as a random draw of the Gaussian function Gauss($\mu, \delta\mu$) (Reasons for the procedure is given in [74]). The distribution, using Gauss(20.6,10.4), is plotted in a figure 6.1. As the distribution covers the region less than zero and since the number of external neutrons can not be negative, the Gaussian function has to be truncated to positive regions. For the truncated Gaussian function, the mean and the centroid are not the same ($\bar{x} \neq \mu$). The constraint term in the likelihood function (equation (6.3)) requires $\tilde{N}$ to be the centroid but if instead of a centroid, one substitutes a mean into the equation, one necessarily gets a bias. Figure 6.2 shows the same Gaussian as in figure 6.1, but truncated to positive values. Although both figures have the same centroid (20.6), the means are different. The bias, due to the difference between the mean and the centroid, is defined as:

$$\text{Bias} = \frac{N - \tilde{N}}{N} \tag{6.9}$$

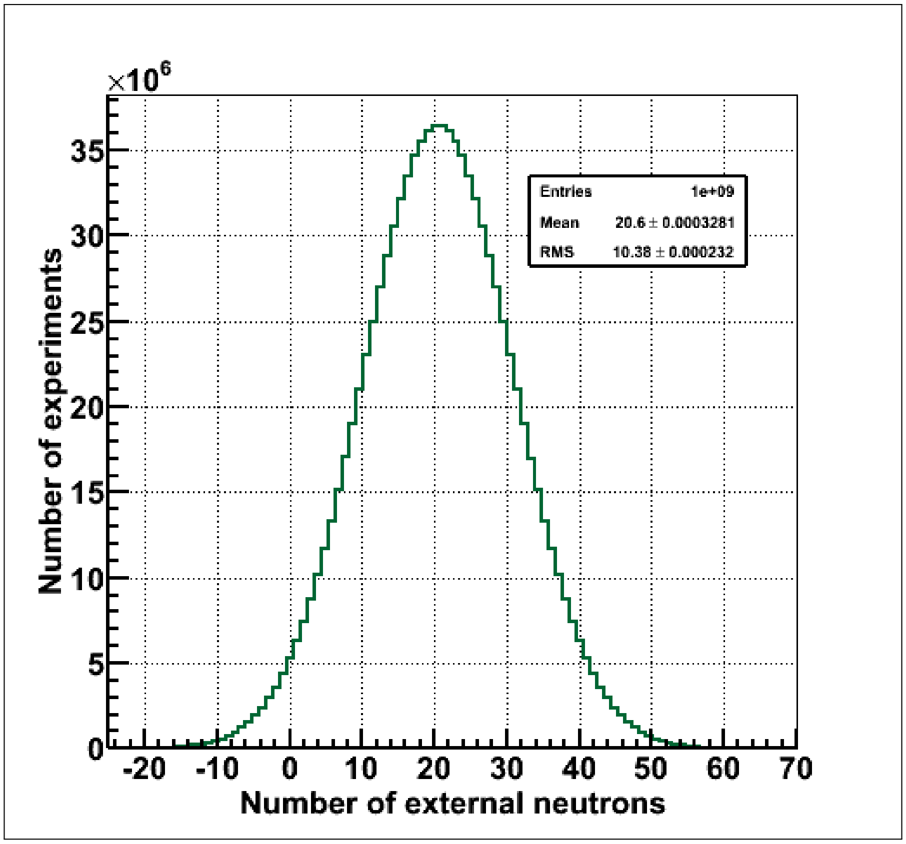


Figure 6.1: Distribution of the number of external neutrons using the Gaussian function (equation (6.4)) with a mean (μ) of 20.6 and σ of 10.4. A $\sigma = 10.4$ comparable to the mean $\mu = 20.6$ will result in a Gaussian function traversing the negative region.

where $\tilde{N}$=20.6±0.003 is the centroid from figure 6.1 and N=21.19±0.003 is the mean from figure 6.2. Using these values, the bias and the uncertainty on the bias are calculated as $0.029 \pm 5.89e-6$ (equations (6.10) and (6.11)). The bias is clearly much greater than its uncertainty.

$$\text{bias} = \frac{(21.19 - 20.6)}{20.6} = 0.029 \tag{6.10}$$

$$\delta\text{bias} = \text{bias}\sqrt{(0.003/20.6)^2 + (0.003/21.19)^2} = 5.89 \times 10^{-6} \tag{6.11}$$

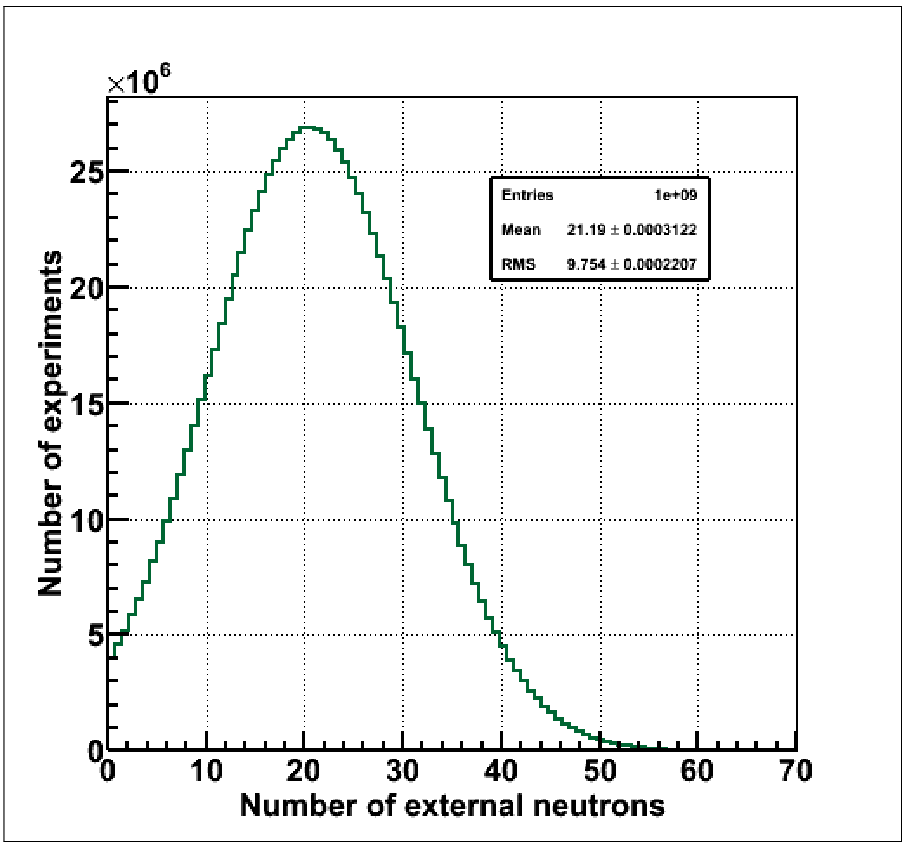


Figure 6.2: Effect of truncating the Gaussian function to the positive region only. The mean of 21.19 is different from the centroid of 20.6

6.3 Solution of the Problem

For a truncated Gaussian function, equation (6.7) is modified by changing the limits of integration. The centroid μ, corresponding to the mean $(\bar{x})$, is found by solving the following equation:

$$\bar{x} = \frac{\int_0^\infty x \exp\left(-\frac{(x-\mu)^2}{2\,\sigma^2}\right) dx}{\int_0^\infty \exp\left(-\frac{(x-\mu)^2}{2\,\sigma^2}\right) dx} \tag{6.12}$$

The integration is:

$$\bar{x} = \mu + \frac{\sqrt{\frac{2}{\pi}}\sigma\ \exp\left(\frac{-\mu^2}{2\,\sigma^2}\right)}{\mathrm{Erfc}\left(\frac{-\mu}{\sqrt{2}\sigma}\right)} \tag{6.13}$$

$$\mathrm{Erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2}\, dt \tag{6.14}$$

We want to invert equation (6.13) to solve for the Centroid (μ) in terms of the mean $\bar{x}$. Since solving equation (6.13) is difficult, a ROOT utility TF1 (defined as equation 6.13) was utilized to solve the problem. Then the TF1 routine GetX [87] was availed to obtain the Centroid (μ) corresponding to the given mean $(\bar{x})$. The Centroid was then applied in the calculation of the penalty term of the likelihood (equation (6.3)). We test the procedure by following this recipe.

1. Randomly draw a mean $(\bar{x})$ from Gauss(20,10).
2. Find the centroid (μ) corresponding to the mean $(\bar{x})$ using equation (6.13).
3. Plot a histogram using random draws from Gauss(μ,10).
4. If the solution is correct then the mean of the histogram should be $\bar{x}$.

Two cases, plotted in figures 6.3 and 6.4, proved that the solution is indeed correct.

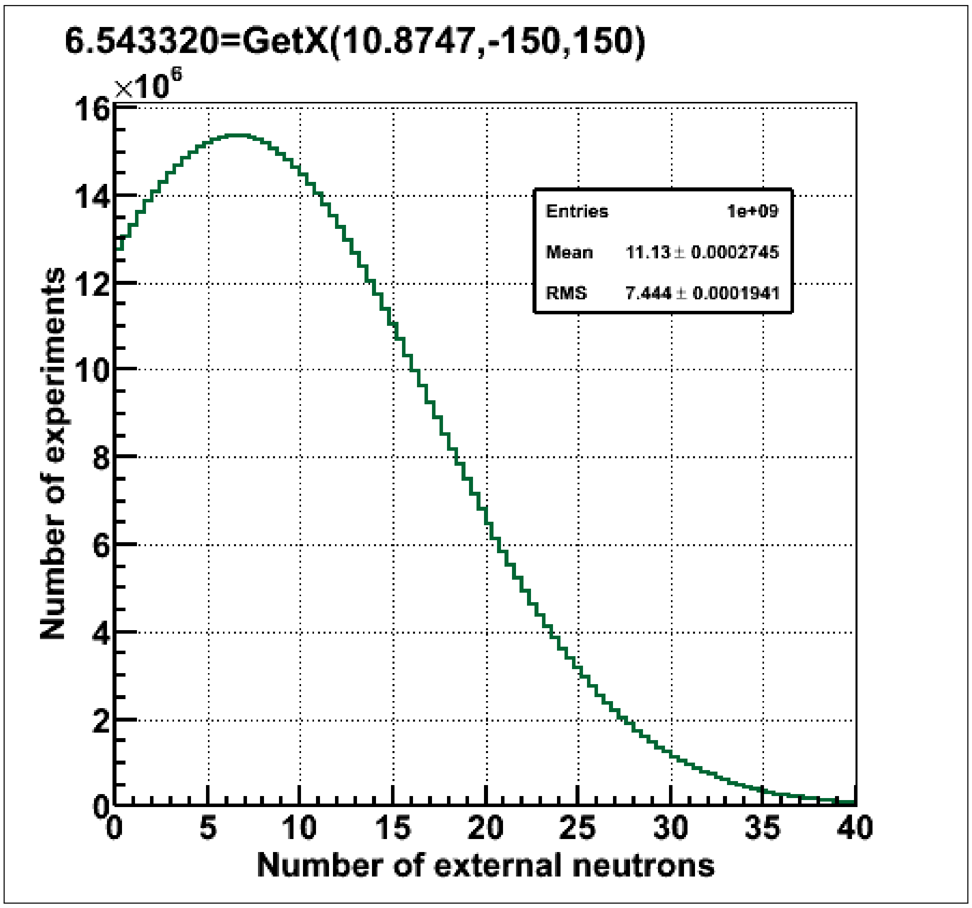


Figure 6.3: Centroid 6.54332, corresponding to the mean of 10.8747, resulted in a histogram with a mean $\pm$ RMS corresponding to 11 ± 7.

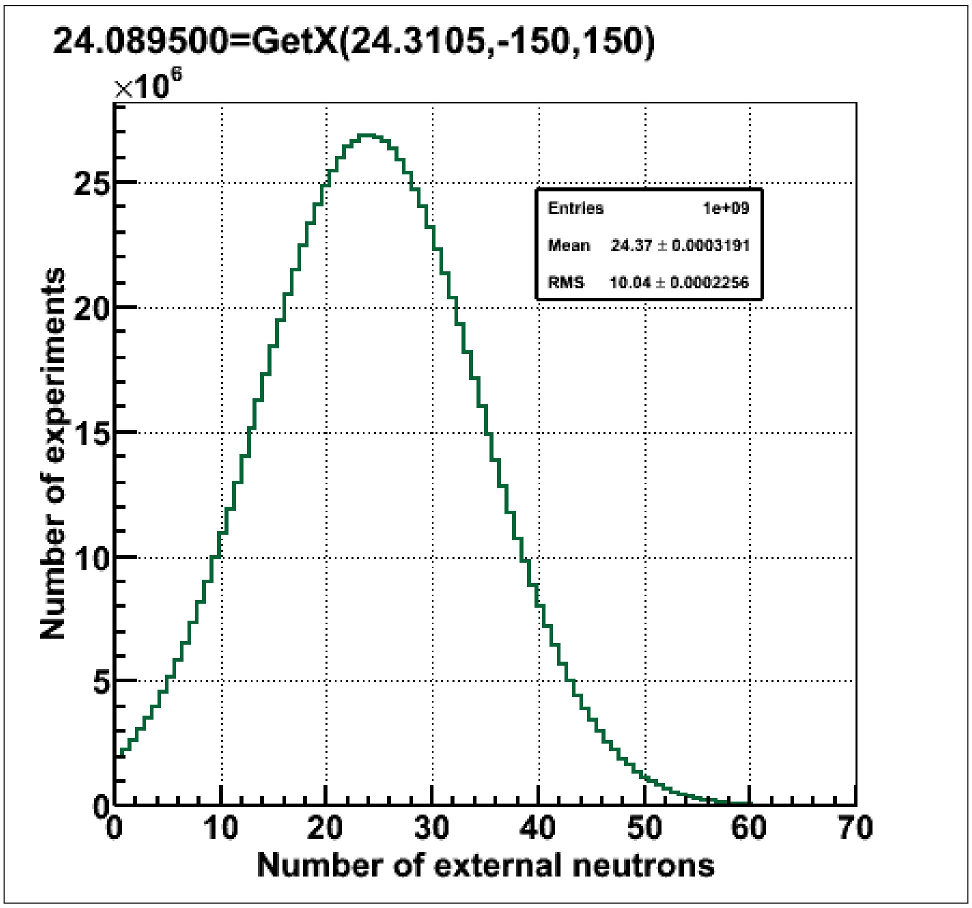


Figure 6.4: Centroid 24.0895, corresponding to the mean of 24.3105, resulted in a histogram with a mean $\pm$ RMS corresponding to 24.37 ± 10.04.

6.4 Introduction of the Toy Monte Carlo

A toy Monte Carlo (MC) was developed to test whether properly taking into account the dissimilarity between the mean and the centroid of the truncated Gaussian function results in a reduced bias compared to when the distinctness is not considered. To determine if the fit is unbiased and uncertainties are properly propagated, a pull and bias study is performed. The pull and bias are defined as follows:

$$\mathrm{Pull}_{ij} = \frac{N_{ij} - \tilde{N}_j}{\sigma_{ij}} \tag{6.15}$$

$$\mathrm{Bias}_{ij} = \frac{N_{ij} - \tilde{N}_j}{N_{ij}} \tag{6.16}$$

where N_{ij} and σ_{ij} are the number of events and the uncertainty on the number of events for the signal **j** obtained from fitting the fake data set number **i**, $\tilde{N}_j$ is the mean Poisson number for the signal **j** in the fake data sets.

Our toy Monte Carlo has two signal types – **A** and **B** – and the purpose of a fit is to determine the number of events N_A and N_B in our data set. To test the query posed in the title of this note, the PDFs for the signals **A** and **B** were formed using random draws from $\mathrm{Gauss}(6, 1)$ and $\mathrm{Gauss}(6.1, 1)$ respectively. The nearly identical PDFs, shown in figure 6.5, were selected so that a constraint ($\mathrm{Gauss}(20, 10)$) on the signal **B** will provide a separation of the signals. Likelihood function is described in equation (6.3) where m goes from 1 to 2, $\tilde{N}$ is the central value of the constraint and $\tilde{\sigma}$ is 10. The central value of the constraint is independently Gaussian-fluctuated according to the uncertainty $\sigma = 10$, hence the central value is changing from one fake data set to another fake data set. The reason for the above procedure is described in [74]. In the MCMC fit, the number of steps was 20,000 and the burn-in period was 5000. After removing the burn-in period, the Mean and RMS of the

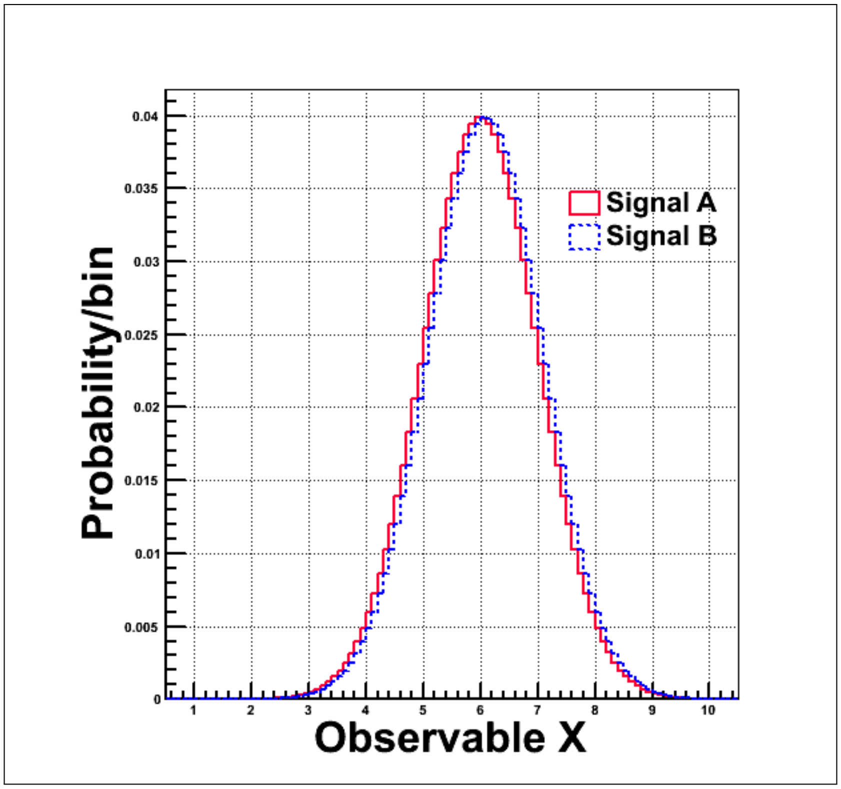


Figure 6.5: PDFs for the two signal types A and B, defined over a hypothetical observable $\mathbf{X}$.

posterior distribution were used as the number of events and the uncertainty on the number of events respectively.

6.4.1 Introduction to Different Cases

Three cases were considered.

- Case 1: The number of each event type is drawn from a Poisson distribution for each fake data set. Mean number of events in the Poisson distributions were 250 and 100 for the signals **A** and **B** respectively. Case 1 is a base case because the distribution of Gauss(100,10), as shown in figure 6.6, does not venture into negative regions. Hence the pull and bias of Case 2 and Case 3 will be compared to pull and bias of Case 1.

- Case 2: Mean number of events, in the Poisson distributions, were changed from 100 to 20 for the signal **B**. For a given fake data set the

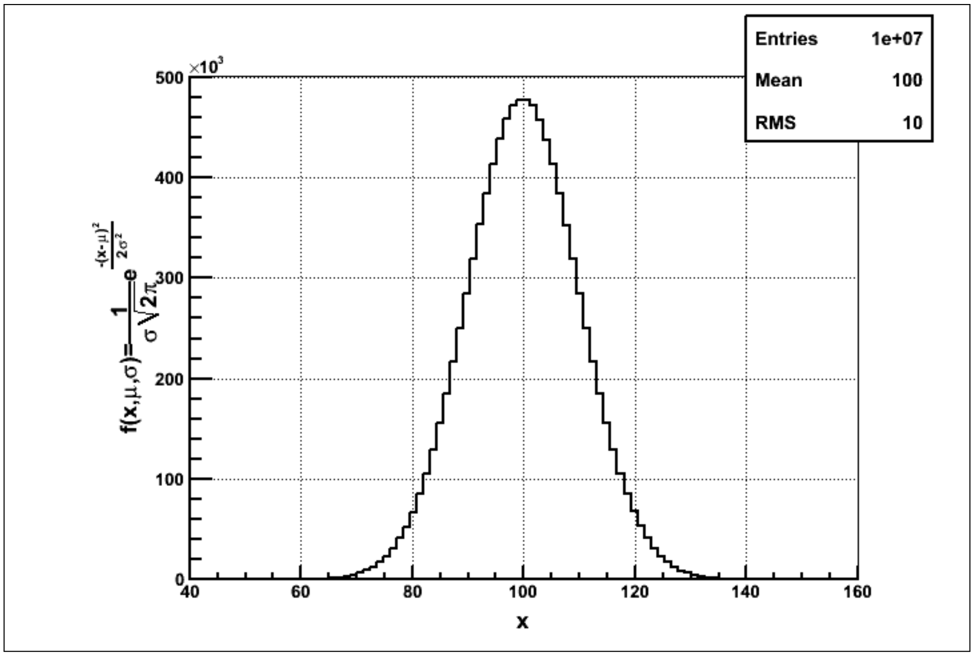


Figure 6.6: Distribution of Gauss(100,10) does not yield negative number of events for the signal **B**.

number of events for the signal **A** was always selected using the Poisson function but for the signal **B**, one of the two functions was used: Poisson or Gauss(20,10). For each fake data set, a random number was drawn from Gauss(20,10) as the mean number of events to be set in motion as $\tilde{N}$ in the penalty term of the likelihood equation (6.3). No correction was applied to account for the discrepancy between the centroid and the mean of the truncated Gaussian function Gauss(20,10).

- Case 3: Same as case 2 except that the difference between mean and centroid is properly accounted for, by finding a centroid corresponding to the mean and then using the centroid as $\tilde{N}$ in the penalty term of the likelihood equation (6.3).

	Poisson Mean	**Constraint**
Case 1	A 250 B 100	NO 100 ± 10
Case 2/3	A 250 B 20	NO 20 ± 10

Table 6.1: Quick overview of the Toy Model.

6.5 Result and Discussion

Two thousand fake datasets were used in a MCMC fit to extract the number of events of signals, N_A and N_B, from the data. Table 6.2 was utilized to plot figures 6.12 and 6.13. From the figure 6.12 and the table 6.2, the width of the pull distribution seems consistent with 1.0. From plots shown in the figure 6.13, it is conclusive that the disparity between the centroid (μ) and the mean ($\langle x \rangle$) of the Gaussian function matters in the fit. Case 1 is a base line case; the number of events for signals **A** and **B** are 250 and 100 respectively and the constraint on signal **B** is 100$\pm$10. Since the width of the constraint is 10% of the central value (100), the mean and the centroid of the Gaussian distribution are equivalent. For the Case 2 and 3, the number of events for the signal **A** remains same as 250 but the number of events for the signal **B** drops down from 100 to 20. Now the width of the constrain (10) is 50% of the central value (20), hence the mean and the centroid of the Gaussian distribution are no longer equivalent. The bias of fitting signal **B** for Case 2 and Case 3 will always be greater than the bias of fitting signal **B** for the Case 1 (shown in green in figure 6.13) but the goal of the current exercise is to show that the bias of signal **B** will be worse in Case 2 (shown in dotted red in figure 6.13) than in Case 3 (shown in blue in figure 6.13).

The application of equation (6.13) in the fit reduced the bias of the signal **B** from (0.086$\pm$0.009) to (0.028$\pm$0.009) when the Gaussian function was used to randomly draw the number of events for the Signal **B** and from (0.083 ± 0.009) to (-0.005 ± 0.009) when the Poisson distribution was employed instead of the Gaussian distribution to randomly draw the number of events for the Signal **B**.

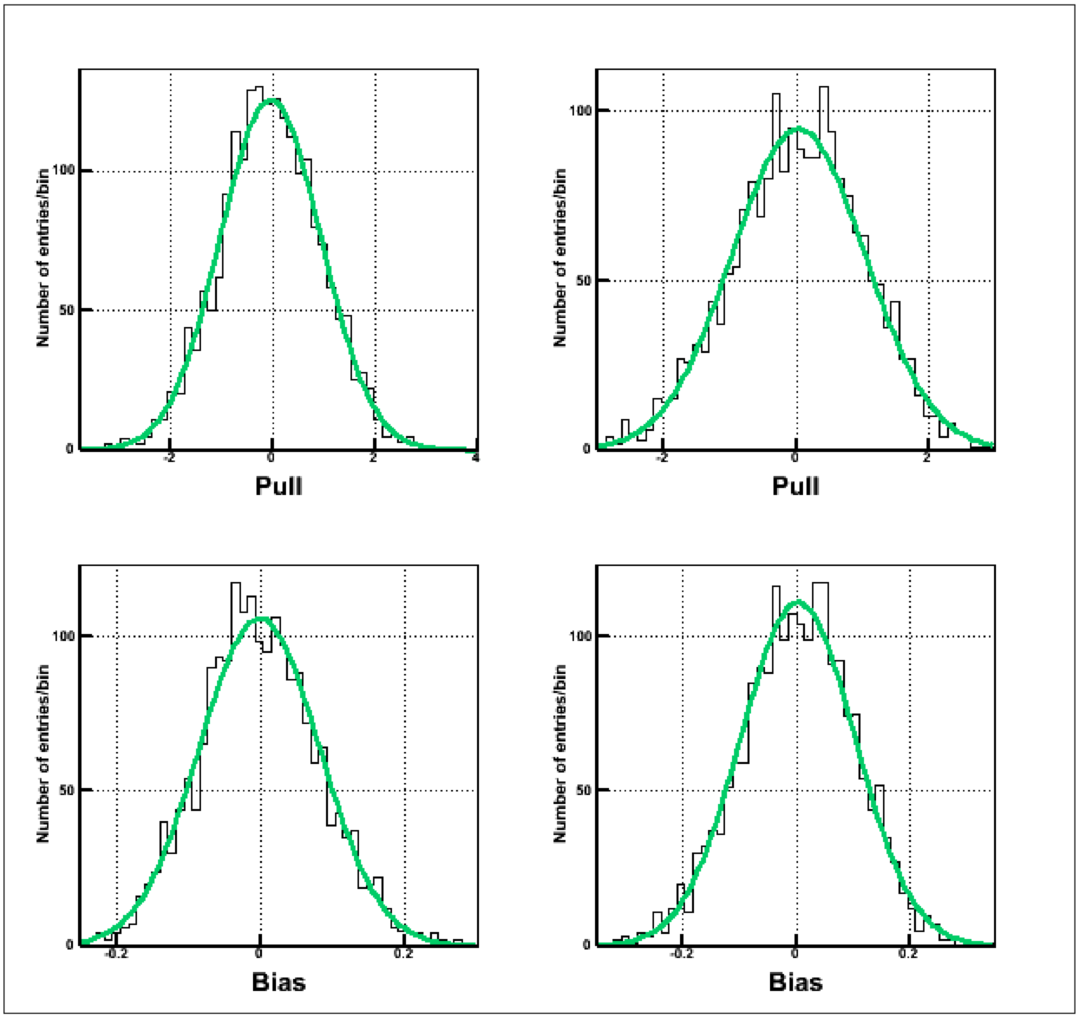


Figure 6.7: Pull and bias plots for the signals **A** and **B** for the Case 1 using Gaussian distribution to generate events for the signal **B**.

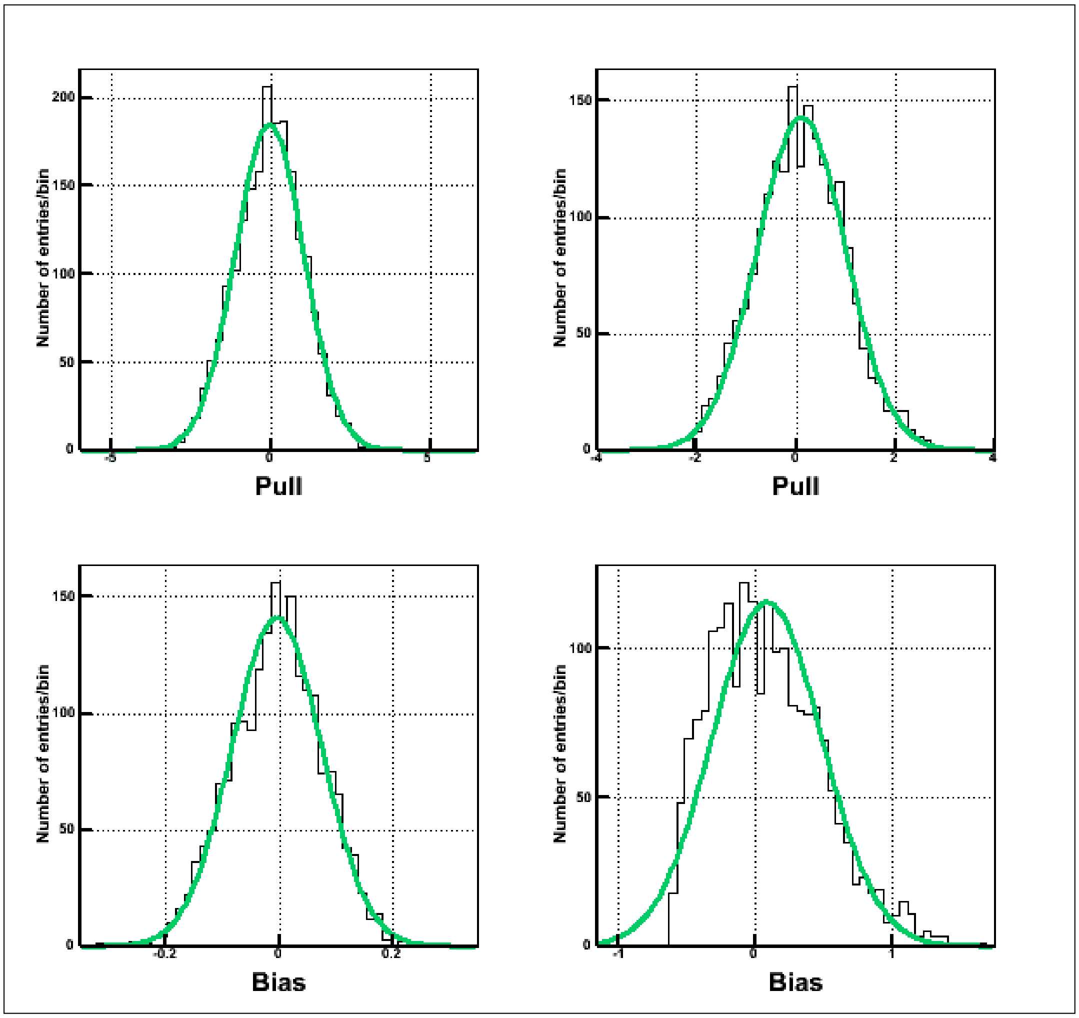


Figure 6.8: Pull and bias plots for the signals **A** and **B** for the Case 2 using Gaussian distribution to generate events for the signal **B**.

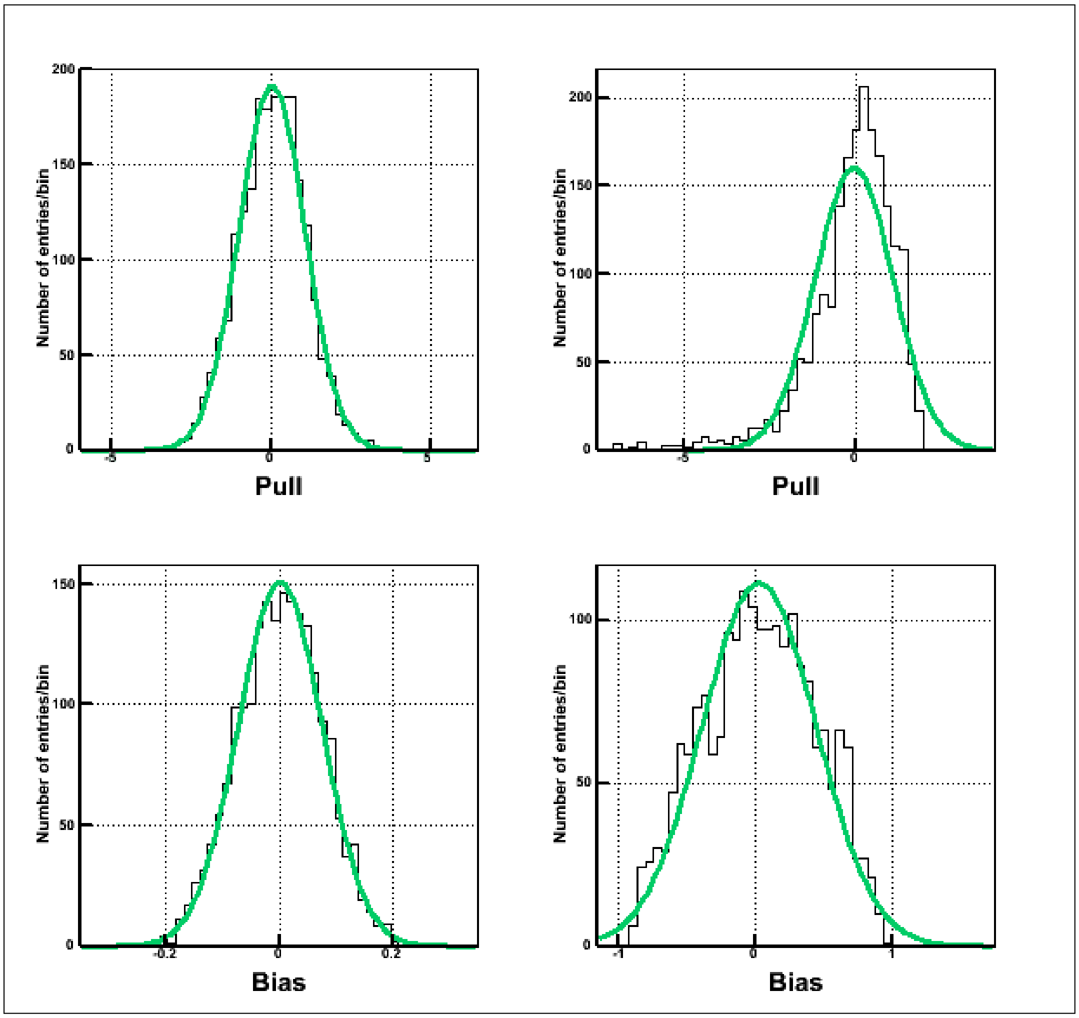


Figure 6.9: Pull and bias plots for the signals **A** and **B** for the Case 3 using Gaussian distribution to generate events for the signal **B**.

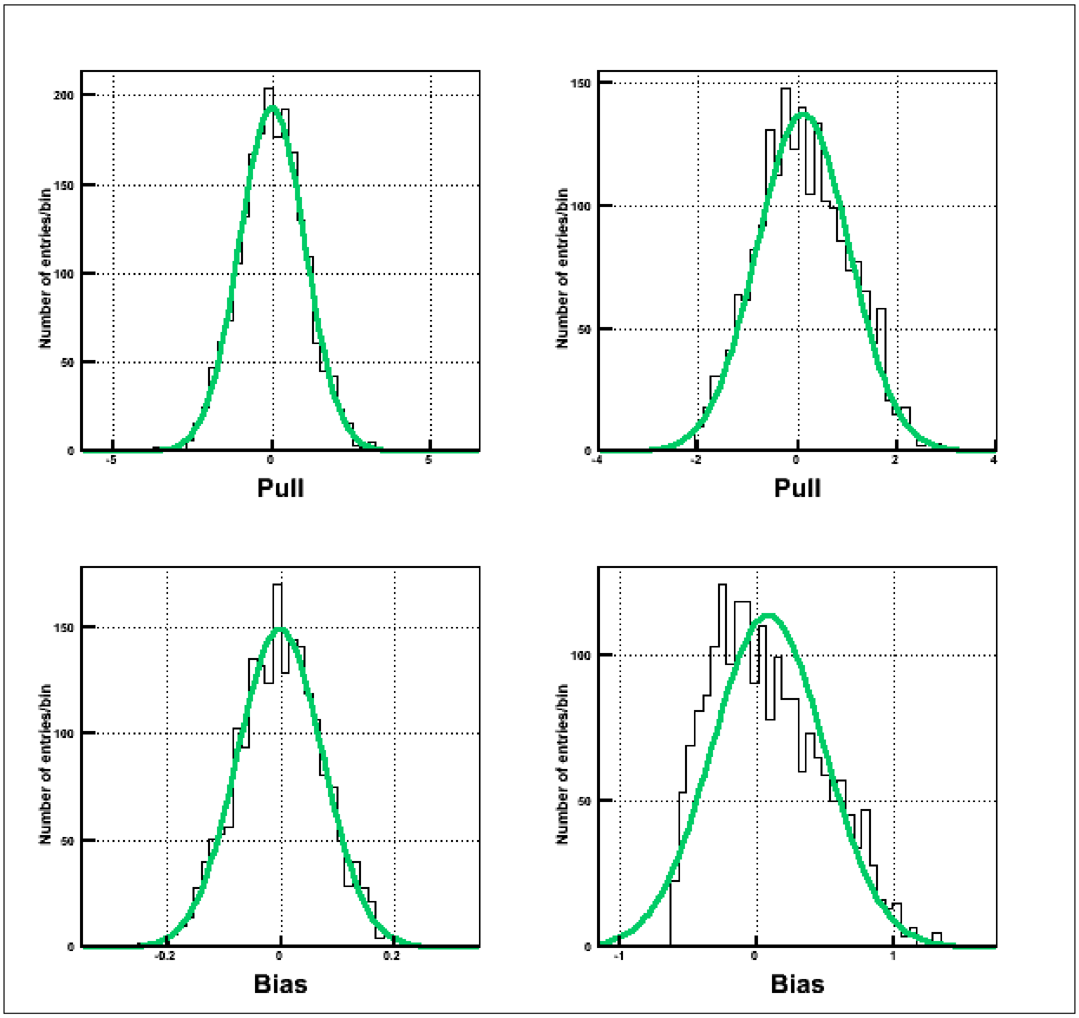


Figure 6.10: Pull and bias plots for the signals **A** and **B** for the Case 2 using Poisson distribution to generate events for the signal **B**.

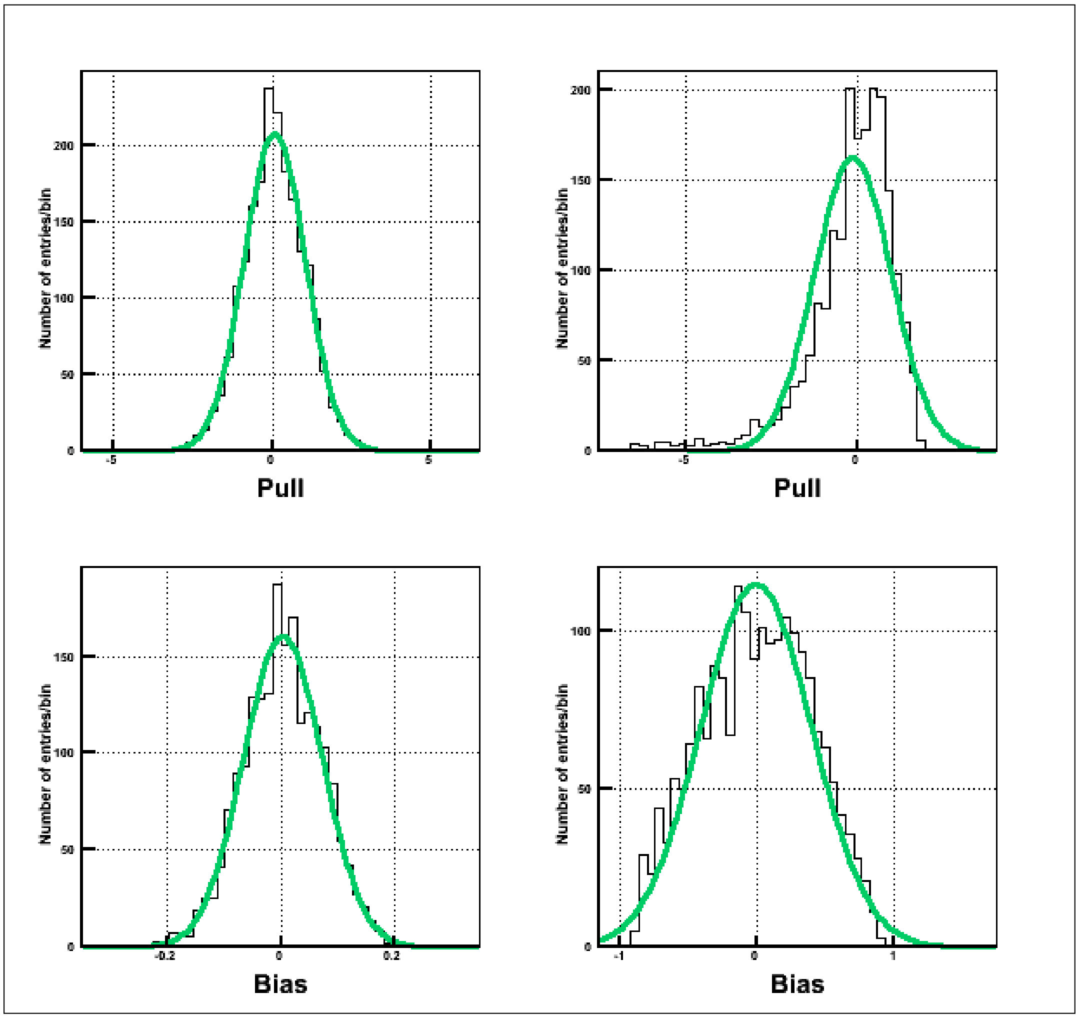


Figure 6.11: Pull and bias plots for the signals **A** and **B** for the Case 3 using Poisson distribution to generate events for the signal **B**.

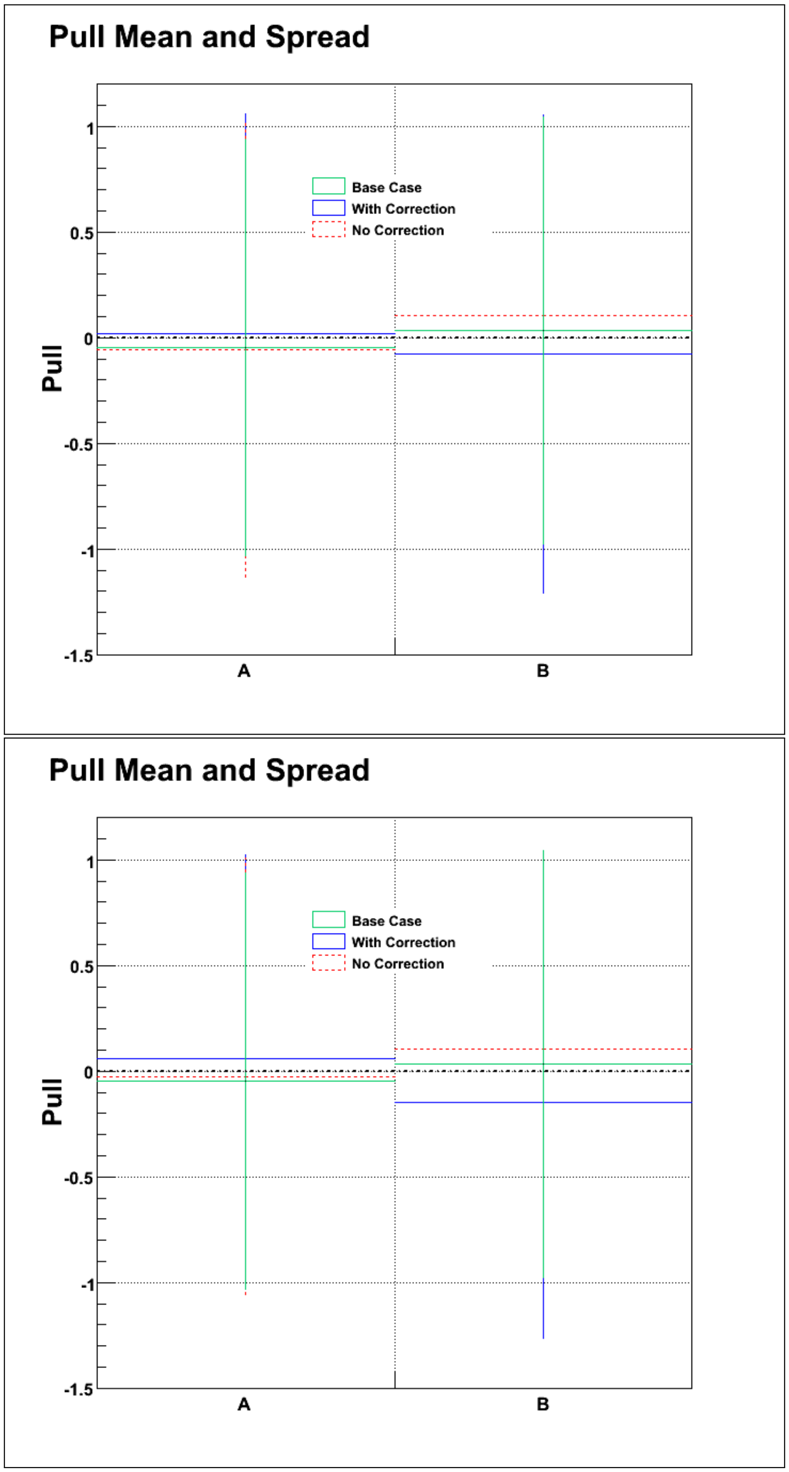


Figure 6.12: Comparing Pull for Case 1, Case 2 and Case 3. Top plot compares the pull spread when the number of events for the Signal **B** is drawn using the Gauss$(20, 10)$ distribution while the bottom compares it using the Poisson distribution instead.

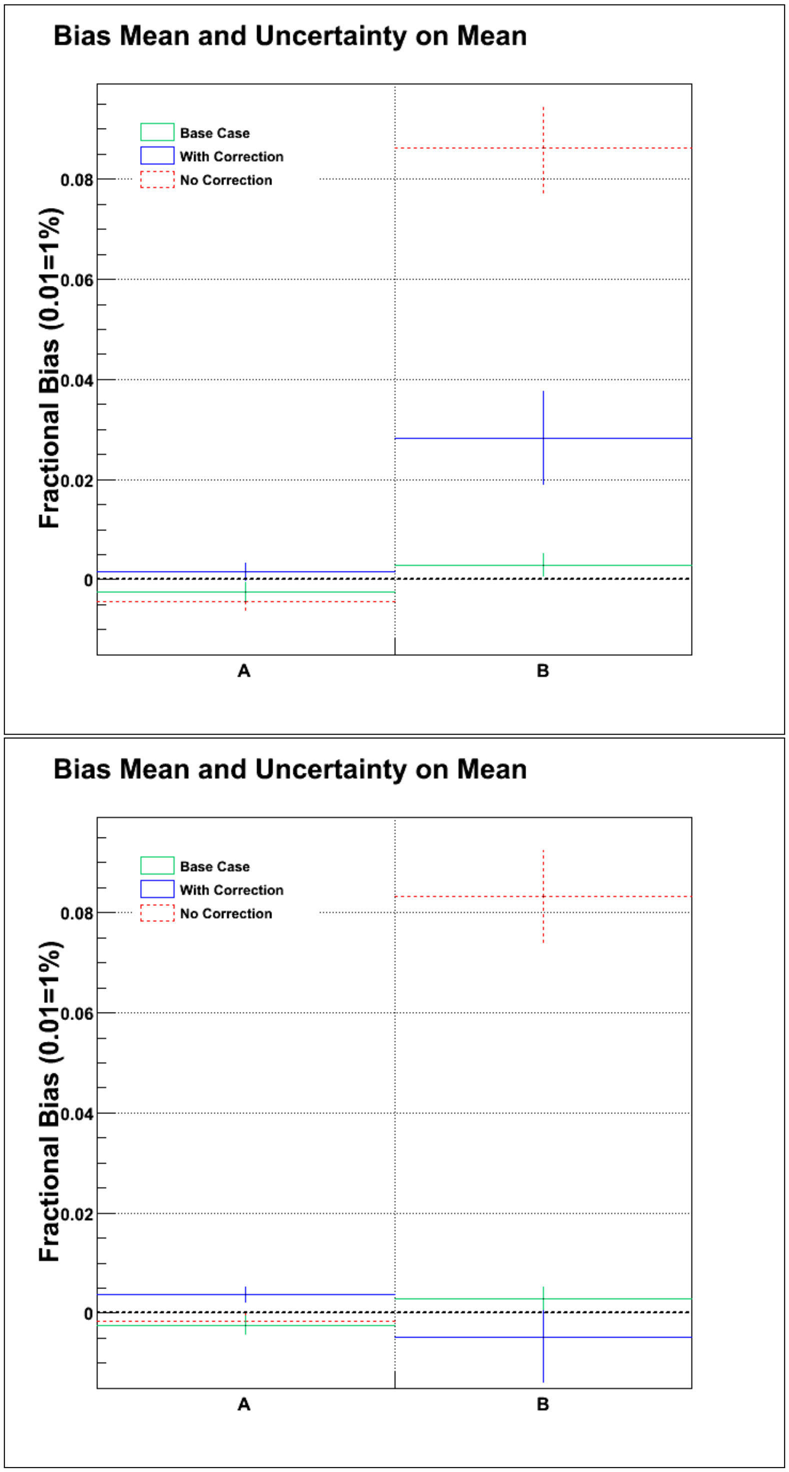


Figure 6.13: Comparing Bias for Case 1, Case 2 and Case 3. Top plot compares the bias spread when the number of events for the Signal **B** is drawn using the Gauss(20, 10) distribution while the bottom compares it using the Poisson distribution instead.

Case	Signal	Pull	Bias	Width of Pull	δBias
Case 1	A	-0.0451	-0.0024	0.9855	0.0019
	B	0.0336	0.0029	1.0091	0.0023
Case 2 (Poisson)	A	-0.0248	-0.0017	1.0328	0.0017
	B	0.1060	0.0833	0.9281	0.009
Case 3 (Poisson)	A	0.0596	0.0037	0.9627	0.0016
	B	-0.1491	-0.0048	1.1136	0.0091
Case 2 (Gaussian)	A	-0.0560	-0.0044	1.0788	0.0018
	B	0.1035	0.0862	0.8921	0.009
Case 3 (Gaussian)	A	0.0217	0.0016	1.0397	0.0017
	B	-0.0764	0.0283	1.1329	0.0093

Table 6.2: Pull and bias in a tabulated form from $\mu \pm \sigma$ of a Gaussian fit on the distributions shown in figures 6.7 to 6.11.

6.6 Conclusion

As demonstrated conclusively the distinction of the mean and the centroid of the Gaussian function is relevant to the bias of the fit. The conclusion is to continue applying equation (6.13) in the MCMC fit in situations where the Gaussian function is truncated due to the fact that the number of events can not be negative.

Chapter 7

Cross-Checks

Several cross-checks were carried out to make sure that the PDFs are normalized correctly and that the likelihood function, used in the MCMC fit, is correct.

1. Verification of the Normalization Method: It is important that the histograms, used in the likelihood function, be probability density function (PDF) with $\int f(x)dx = 1$, that is, the area under the curve is unity. Since the PDFs are model of the event types found in the data and the goal of the fit is to determine the number of events belonging to each event type, any error in the calculation of the normalization will result in the fit giving wrong number of events for the event classes in the data. Hence the calculation of the normalization factor is very important. Using the likelihood function as a method in Root TF3 function, the integral shown in equations (7.1) and (7.2) were verified to be unity.

$$\iiint h(\rho, \cos\,\theta, E)\, dE\, d\rho\, dcos\,\theta \tag{7.1}$$

   Similarly the integrals of the 4D PDFs (equation (7.2)) were verified to be unity by combining a 1-Dimensional and a 3-Dimensional ROOT functions (TF1 [87] and TF3 [88] respectively).

$$\iiiint h(\rho, \cos\,\theta, E|\cos\,\theta_{zenith})\,dE\,d\rho\,d\cos\,\theta\,d\cos\,\theta_{zenith} \tag{7.2}$$

2. Verifications of elements of Survival Probability (p_0, p_1, p_2, a_0 and a_1): This test was performed to make sure that the equations for survival probability ((4.55), (4.57), (4.58) and (4.59)) used to distort the 3 dimensional ($g(\rho, \cos\,\theta, E)$ and 2 dimensional ($h(E_\nu, E)$) histograms result in the same distortion of the energy **E**, that is, $\int h(E_\nu, E)\,dE_\nu = \iint g(\rho, \cos\,\theta, E)\,d\rho\,d\cos\,\theta$. The χ^2 (defined in equation (7.3)) test on $H = \int h(E_\nu, E)\,dE_\nu$ and $G = \iint g(\rho, \cos\,\theta, E)\,d\rho\,d\cos\,\theta$ resulted in χ^2 =0.0 with a probability of unity.

$$\chi^2 = \sum_{j=1}^{N} \frac{\left(Y_j - \Lambda_j\right)^2}{\sigma_{Y_j}^2} \tag{7.3}$$

where Y_j and Λ_j are the bin content of the *jth* bin of H and G respectively, σ_{Y_j} is the error of the *jth* bin belonging to H, and N is a number of bins of the histograms H and G.

The reason for zero χ^2 is that both histograms are exactly the same if the distortion was performed correctly. This test was performed for several MCMC steps with parameters of survival probability starting far away from the actual values used in the generation of the fake data sets. The reason was to make sure that selecting the parameters far away from the actual values will not cause $\int h(E_\nu, E)\,dE_\nu \neq \iint g(\rho, \cos\,\theta, E)\,d\rho\,d\cos\,\theta$. This test #2 was performed on several MCMC fits for several steps of the fit but none showed any problem.

3. Verification of the Likelihood Function: A routine was developed to test the probability density functions used in the MCMC fit against the likelihood functions to make sure that PDFs used in the fit are calculated correctly. The methodology to create a PDF **h**(x,y,z), to compare with

the PDF H(x,y,z) used in the MCMC fit, is called acceptance-rejection and it generates samples from the probability distribution function which in this case was the likelihood function. Following steps were undertaken to perform this check:

(a) In a loop of 50,000 steps, generate random observables x_j, y_j, z_j within the cuts:

$$x_{min} \leq x_j < x_{max} \tag{7.4}$$

$$y_{min} \leq y_j < y_{max} \tag{7.5}$$

$$z_{min} \leq z_j < z_{max} \tag{7.6}$$

(b) Calculate Likelihood **L** using the observables x_j, y_j, z_j.

(c) Determine the maximum of the Likelihood $\mathcal{L}_{max}$ in the loop.

(d) Next, in a loop of 19 millions steps, generate events x_j, y_j, z_j.

(e) Calculate a likelihood $\mathcal{L}$ using x_j, y_j, z_j.

(f) Generate a random number **R** between 0 and 1.0; an event (x_j, y_j, z_j) is accepted if

$$\mathcal{L} < R\mathcal{L}_{max} \tag{7.7}$$

(g) Populate histogram **h** with accepted values x_j, y_j, z_j

(h) After the loop has finished, compare $\mathbf{h}(x_j, y_j, z_j)$ to $H(x_j, y_j, z_j)$ in a χ^2 test. The comparison was done in three steps: 3D was compared against the 3D PDF, 2D projections (xz, yz and xy) were compared against 2D projections and last of all, all three 1D projections (x, y and z) were compared visually as well using a χ^2 test. The probability was calculated using TMath:Prob(χ^2, dof) where dof stands for degrees of freedom.

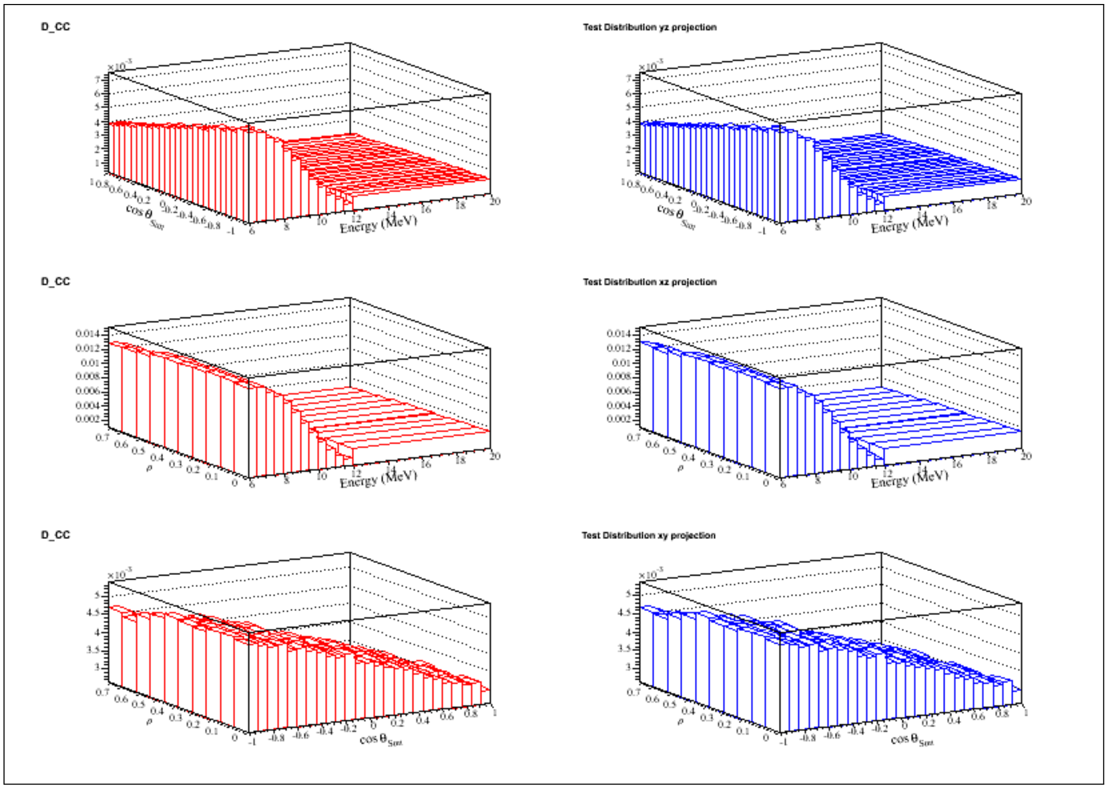


Figure 7.1: Comparing **day** 3D PDF for CC (left) to the one generated using the likelihood function (right). Top shows yz projection, middle shows xz projection and the bottom plot displays xy projection.

The purpose of carrying out this check is to ascertain that the PDFs generated using the Likelihood function (we will call them the tester PDFs) were similar in shape to the original PDFs. The χ^2 test comparing both PDFs indicates the goodness of the fit. This exercise also confirmed that the normalization for each PDF is correctly calculated and the cuts were correctly applied in various sections of the code.

Some example plots are shown in the following figures.

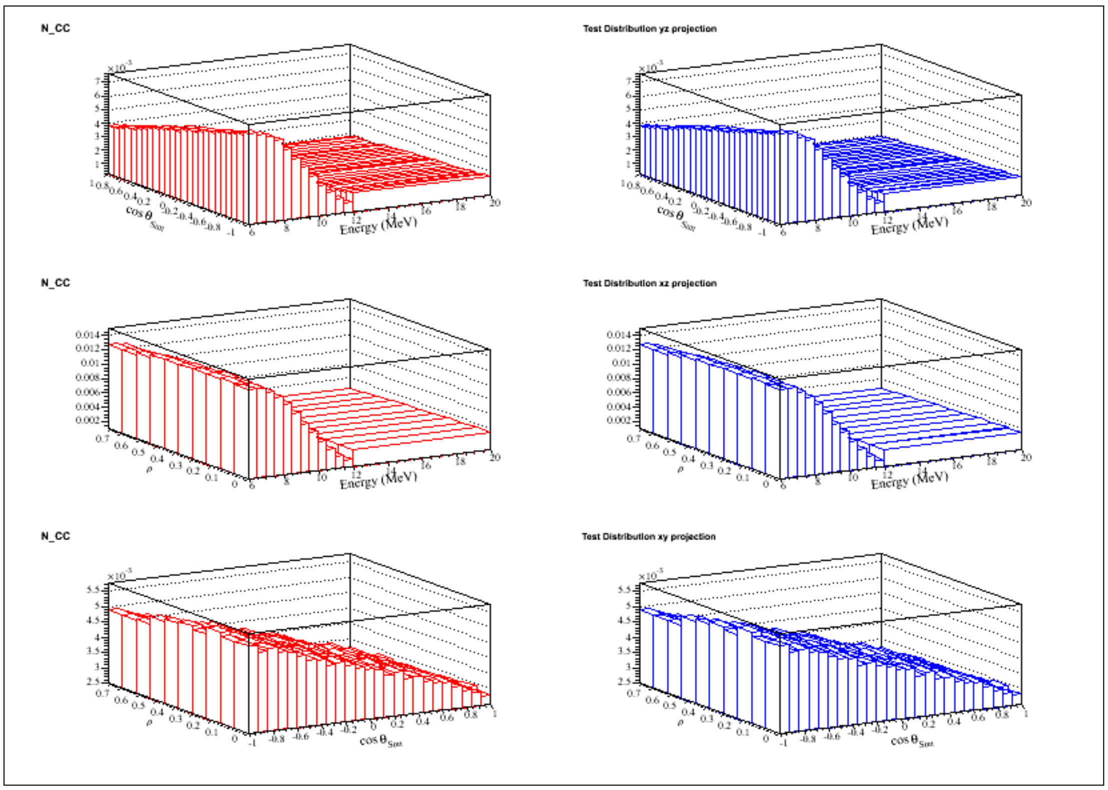


Figure 7.2: Comparing **night** 3D PDF for CC (left) to the one generated using the likelihood function (right). Top shows yz projection, middle shows xz projection and the bottom plot displays xy projection.

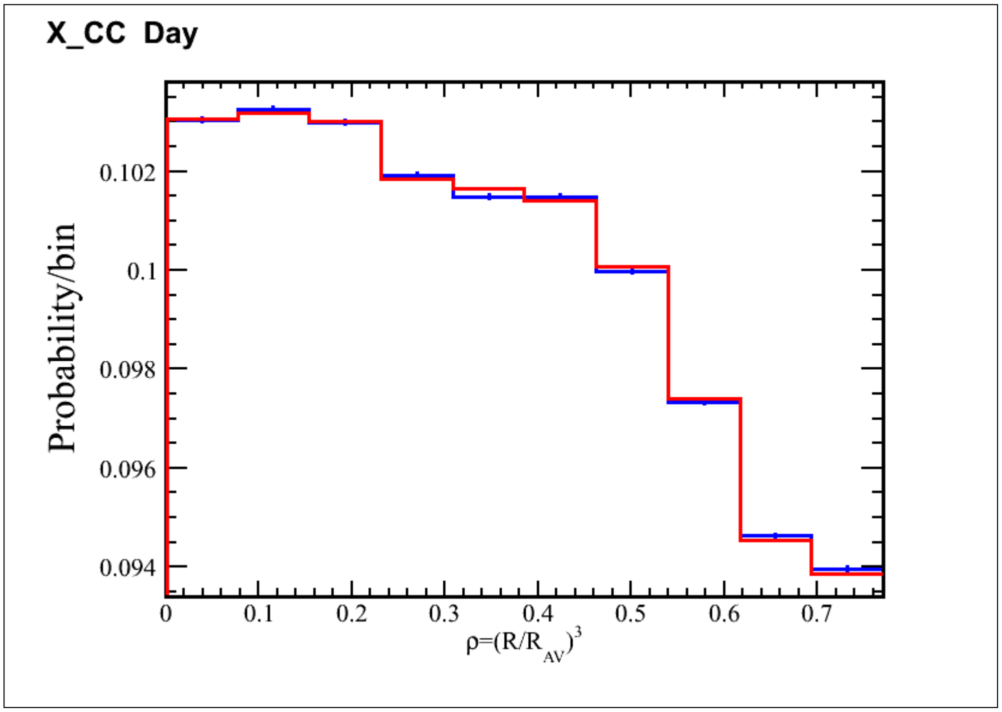


Figure 7.3: Comparing ρ projection of CC **day** – blue shows the x projection of the PDF used in the MCMC fit and red shows the x projection of the tester PDF.

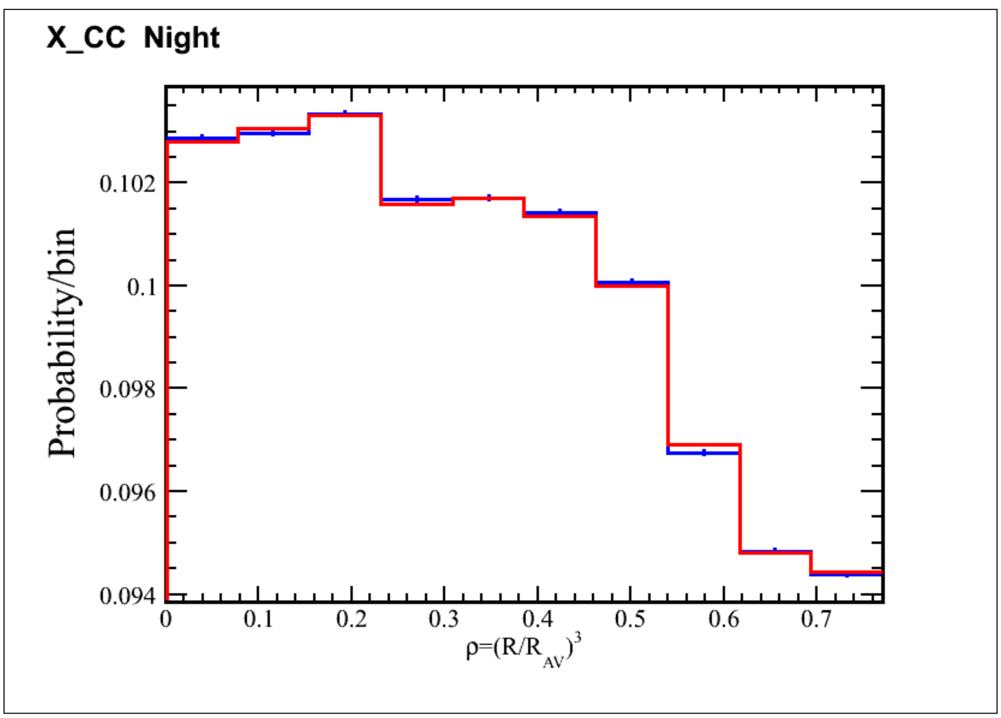


Figure 7.4: Comparing ρ projection of CC **night** – blue shows the x projection of the PDF used in the MCMC fit and red shows the x projection of the tester PDF.

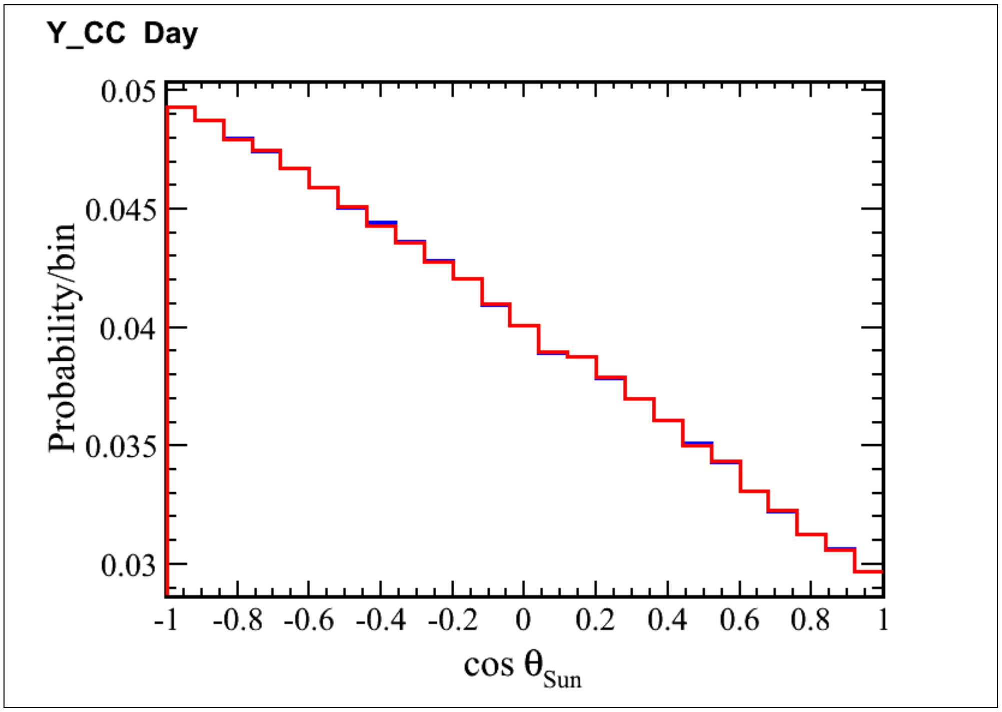


Figure 7.5: Comparing $\cos\theta_{Sun}$ projection of CC **day** – blue shows the y projection of the PDF used in the MCMC fit and red shows the y projection of the tester PDF.

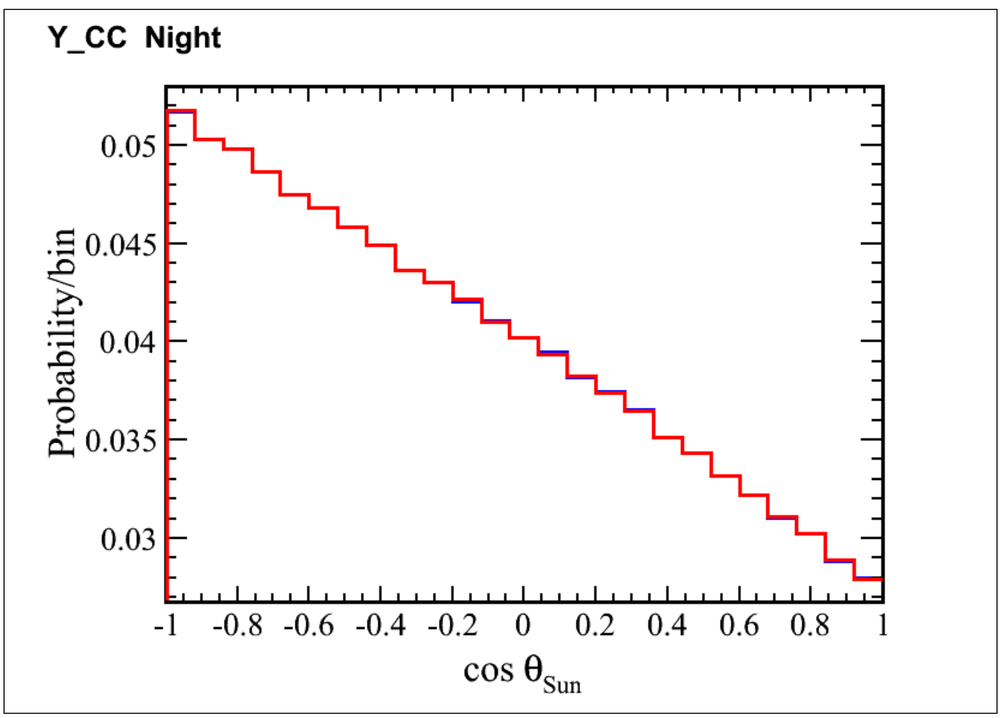


Figure 7.6: Comparing $\cos\theta_{Sun}$ projection of CC **night** – blue shows the y projection of the PDF used in the MCMC fit and red shows the y projection of the tester PDF.

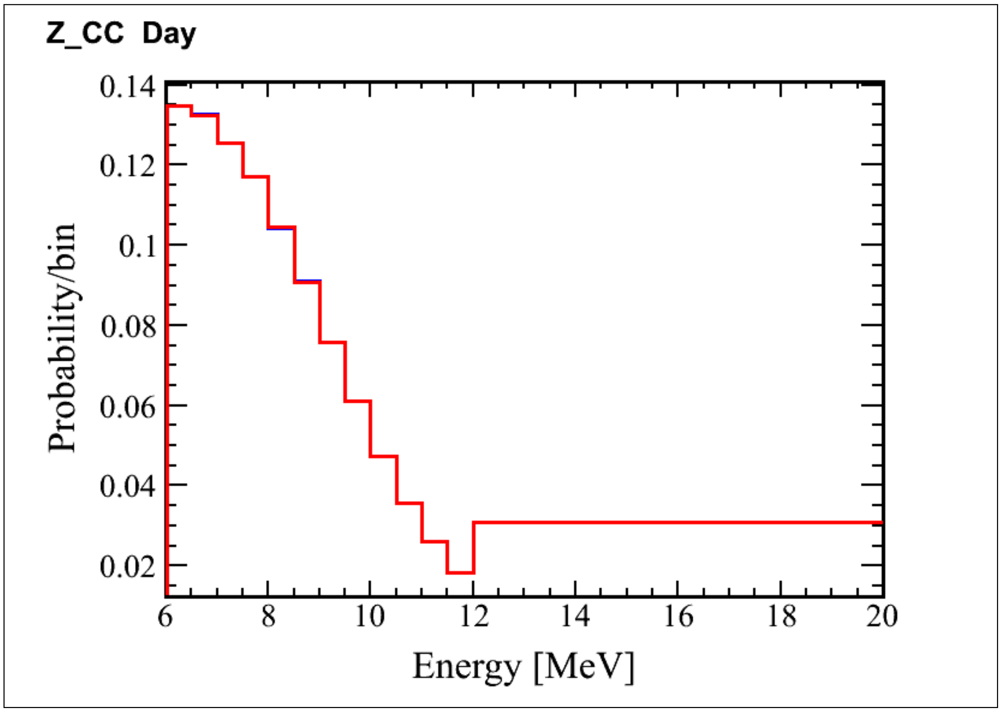


Figure 7.7: Comparing energy projection of CC **day** – blue shows the z projection of the PDF used in the MCMC fit and red shows the z projection of the tester PDF.

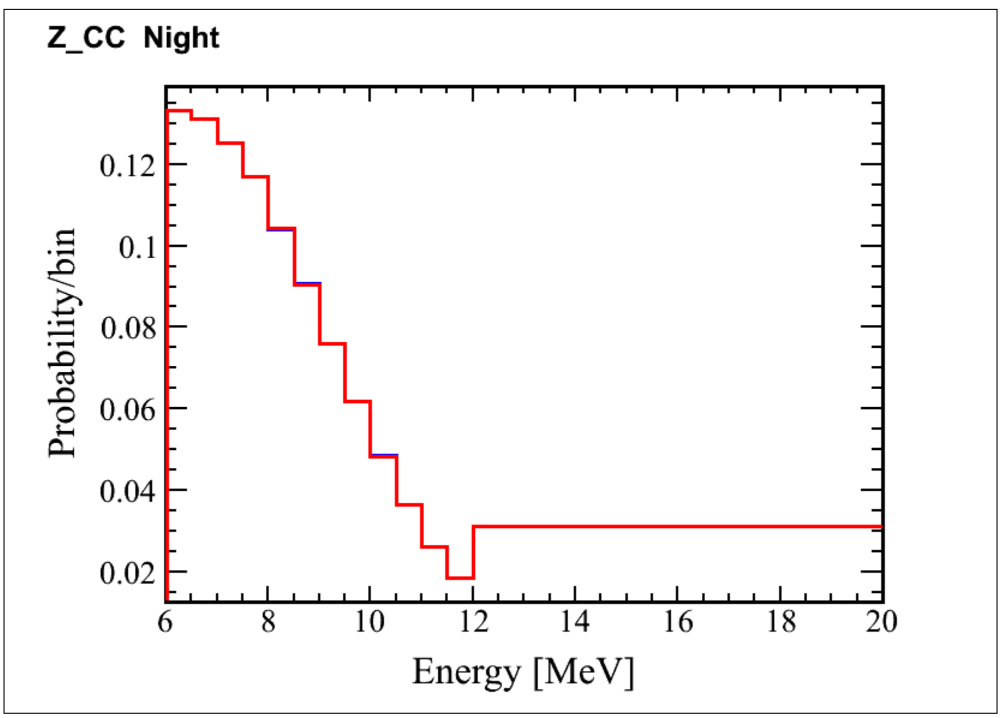


Figure 7.8: Comparing energy projection of CC **night** – blue shows the z projection of the PDF used in the MCMC fit and red shows the z projection of the tester PDF.

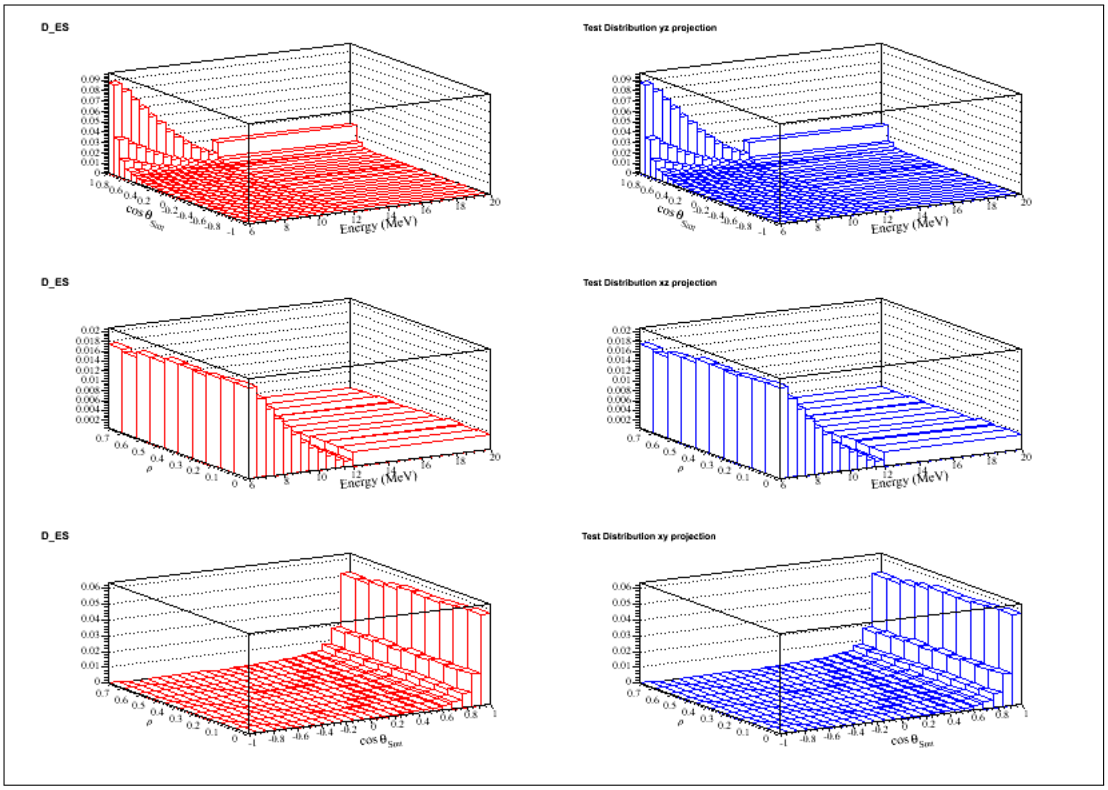


Figure 7.9: Comparing **day** 3D PDF for ES (left) to the one generated using the likelihood function (right). Top shows yz projection, middle shows xz projection and the bottom plot displays xy projection.

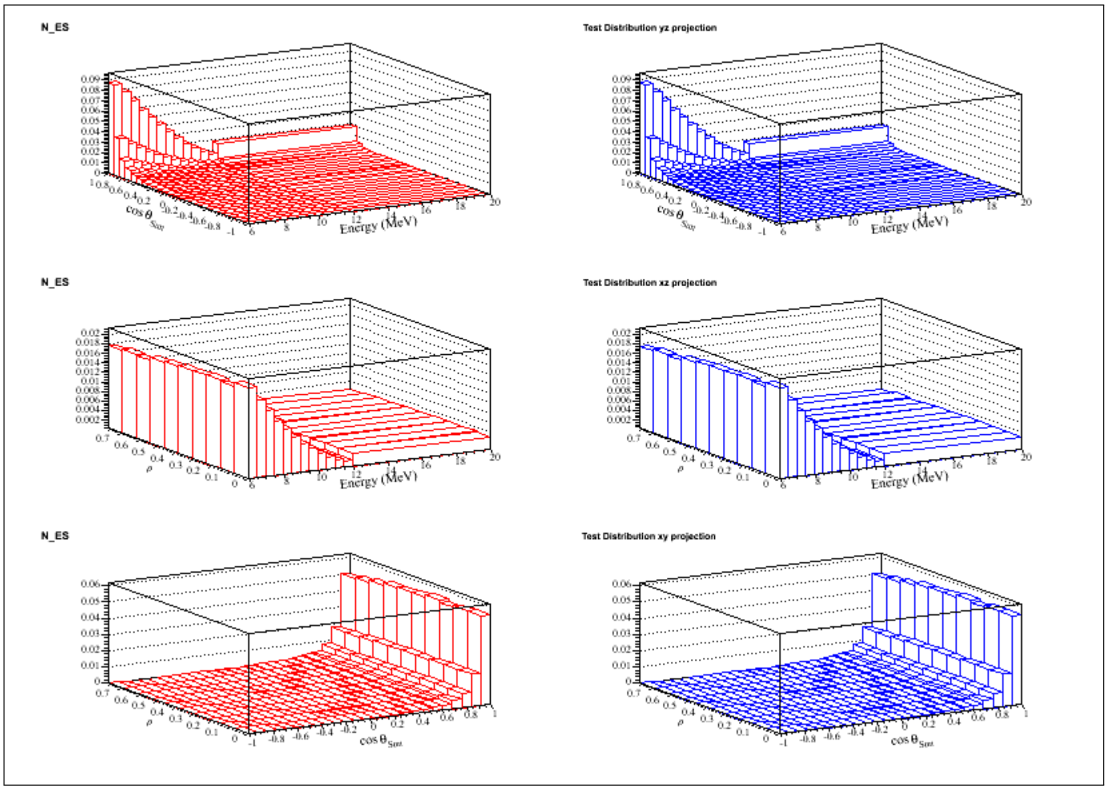


Figure 7.10: Comparing **night** 3D PDF for ES (left) to the one generated using the likelihood function (right). Top shows yz projection, middle shows xz projection and the bottom plot displays xy projection.

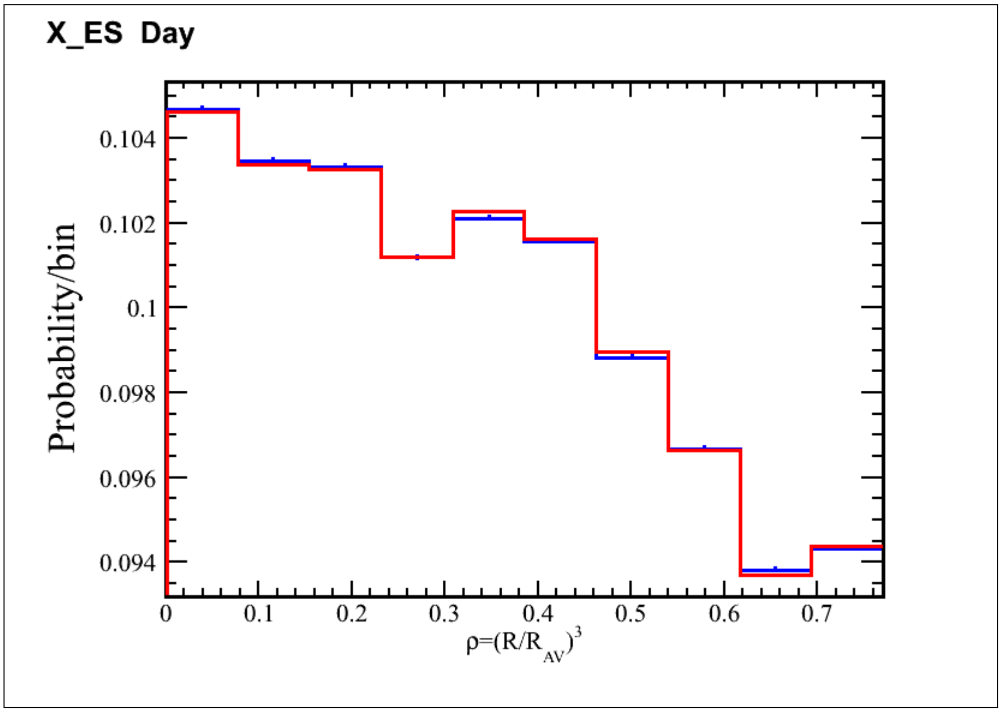


Figure 7.11: Comparing ρ projection of ES **day** – blue shows the x projection of the PDF used in the MCMC fit and red shows the x projection of the tester PDF.

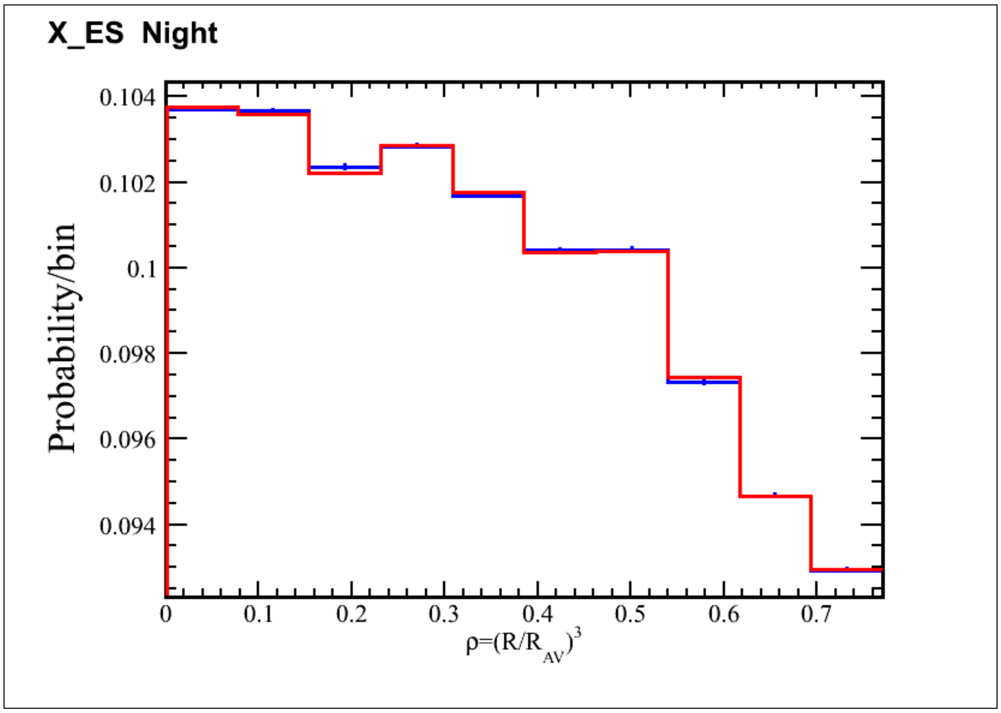


Figure 7.12: Comparing ρ projection of ES **night** – blue shows the x projection of the PDF used in the MCMC fit and red shows the x projection of the tester PDF.

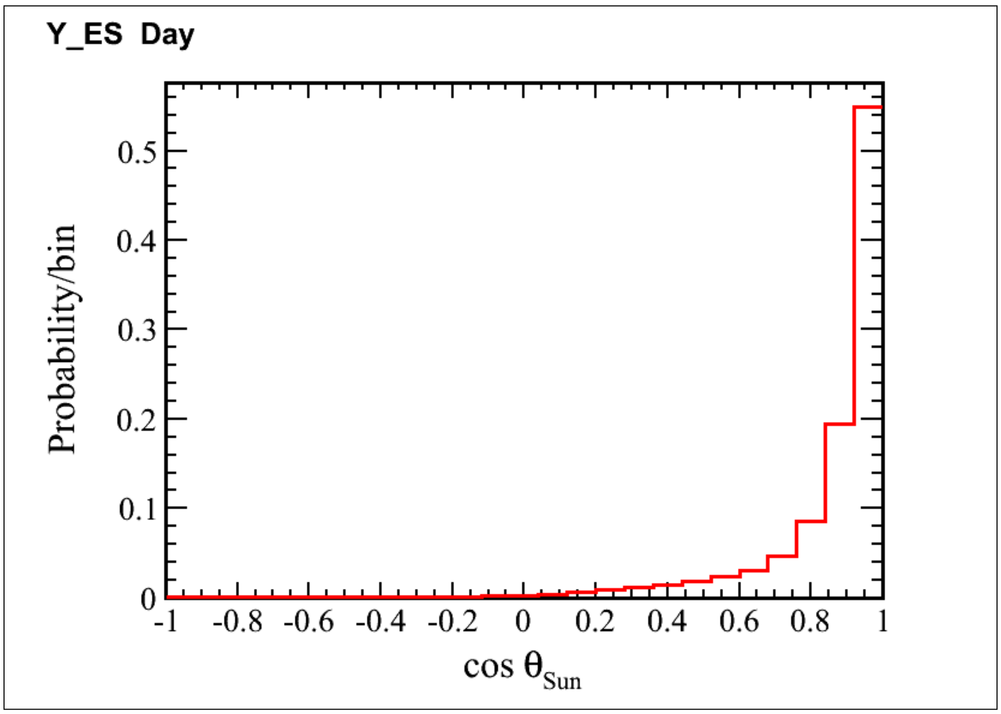


Figure 7.13: Comparing cos θ_{Sun} projection of ES **day** – blue shows the y projection of the PDF used in the MCMC fit and red shows the y projection of the tester PDF. Y projections of ES are very similar hence on this plot, blue is not visible because it is covered by red.

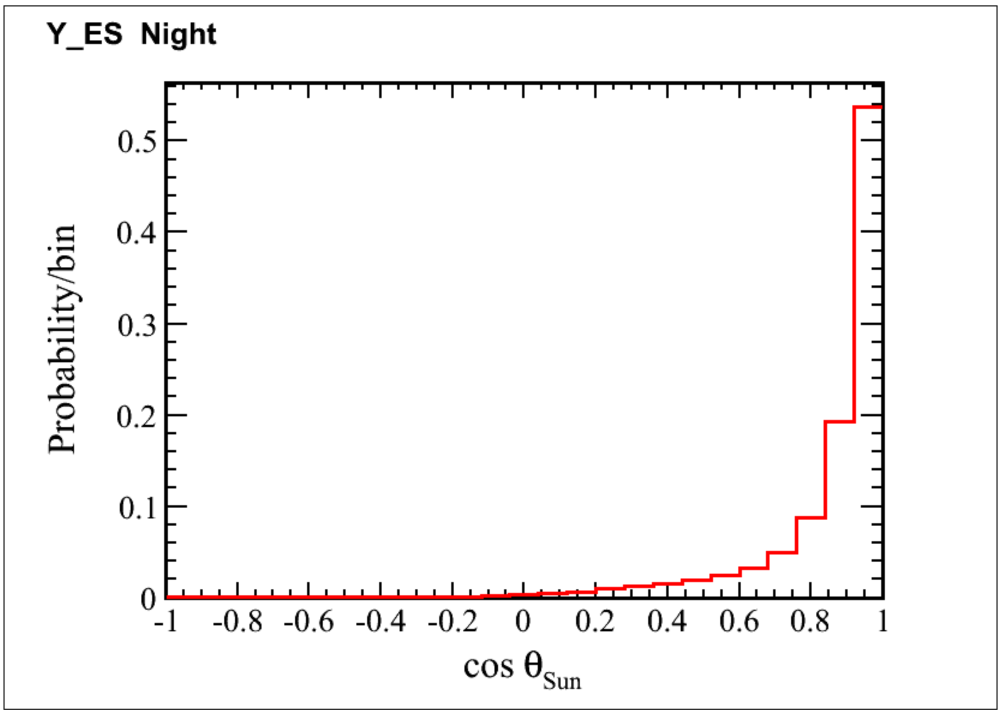


Figure 7.14: Comparing cos θ_{Sun} projection of ES **night** – blue represents the y projection of the PDF used in the MCMC fit and red represents the y projection of the tester PDF. Y projections of ES are very similar hence on this plot, blue is not visible because it is covered by red.

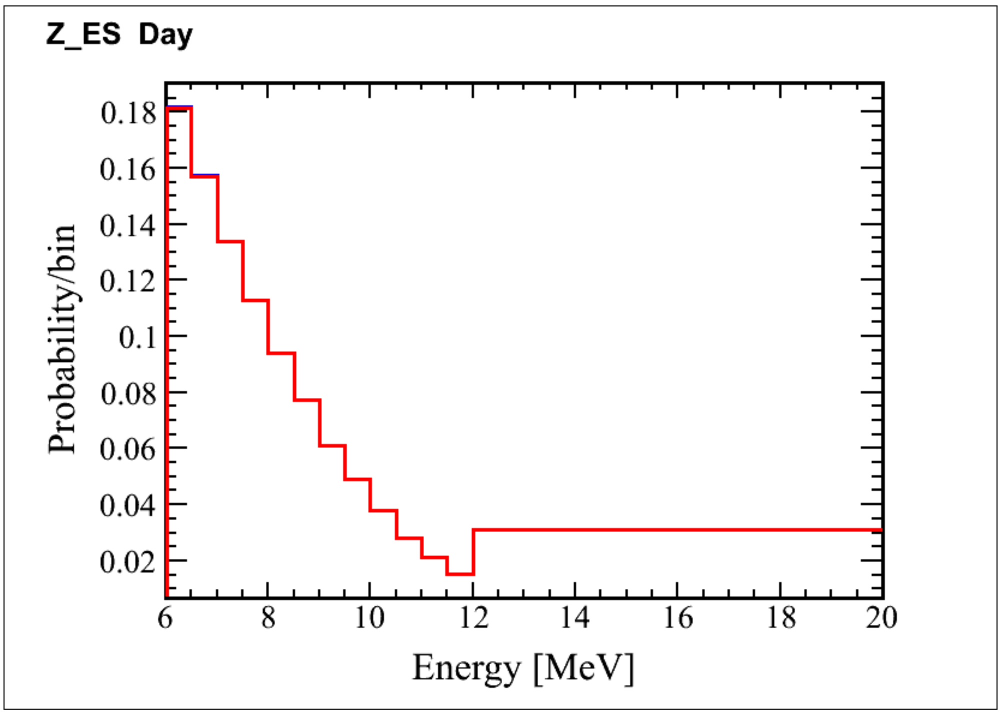


Figure 7.15: Comparing energy projection of ES **day** – blue shows the z projection of the PDF used in the MCMC fit and red shows the z projection of the tester PDF. Z projections of ES are very similar hence on this plot, blue is not visible because it is covered by red.

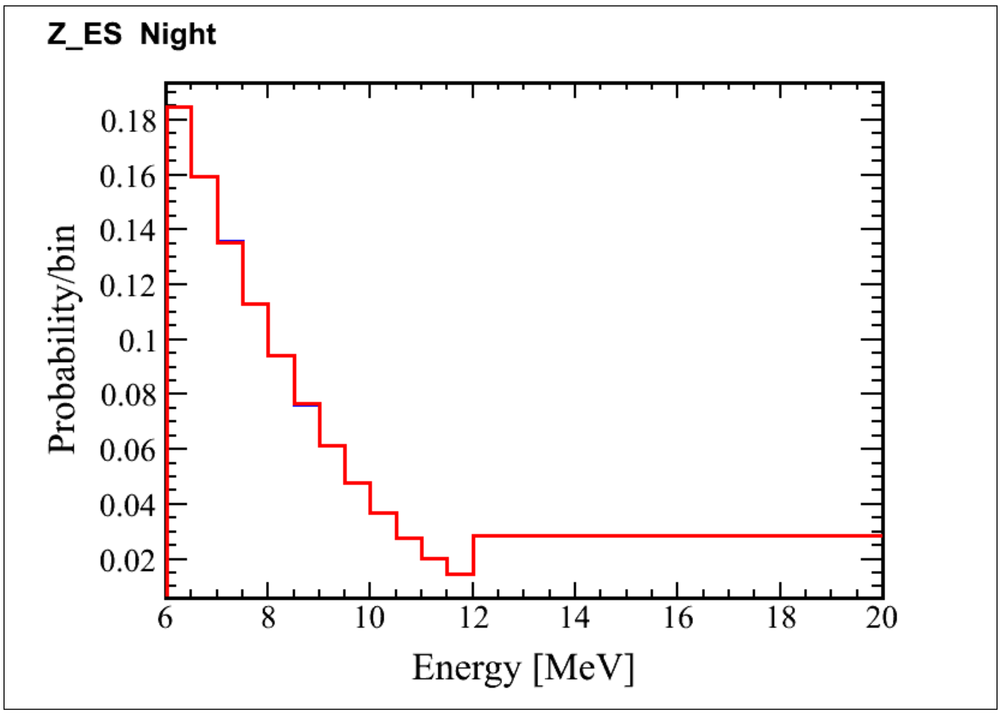


Figure 7.16: Comparing energy projection of ES **night** – blue shows the z projection of the PDF used in the MCMC fit and red shows the z projection of the tester PDF. Z projections of ES are very similar hence on this plot, blue is not visible because it is covered by red.

4. Final Test: For a fake data set, we know the number of events for each class so we generate the PDFs for each class using the TF3 function with Likelihood function of the MCMC fit as a method. The TF3 function generates $\rho, \cos\theta, E$ events for the 3D PDF, within the limits specified for each variable, and then uses the Likelihood method to draw the samples of the PDFs. Once all the PDFs were generated, the projections of the PDF for each class, corresponding to each axis, is normalized using the number of events for the class. For a given axis, the projections from all the classes were added up and fitted against the corresponding projection of the data. See examples in figures 7.17, 7.18 and 7.19.

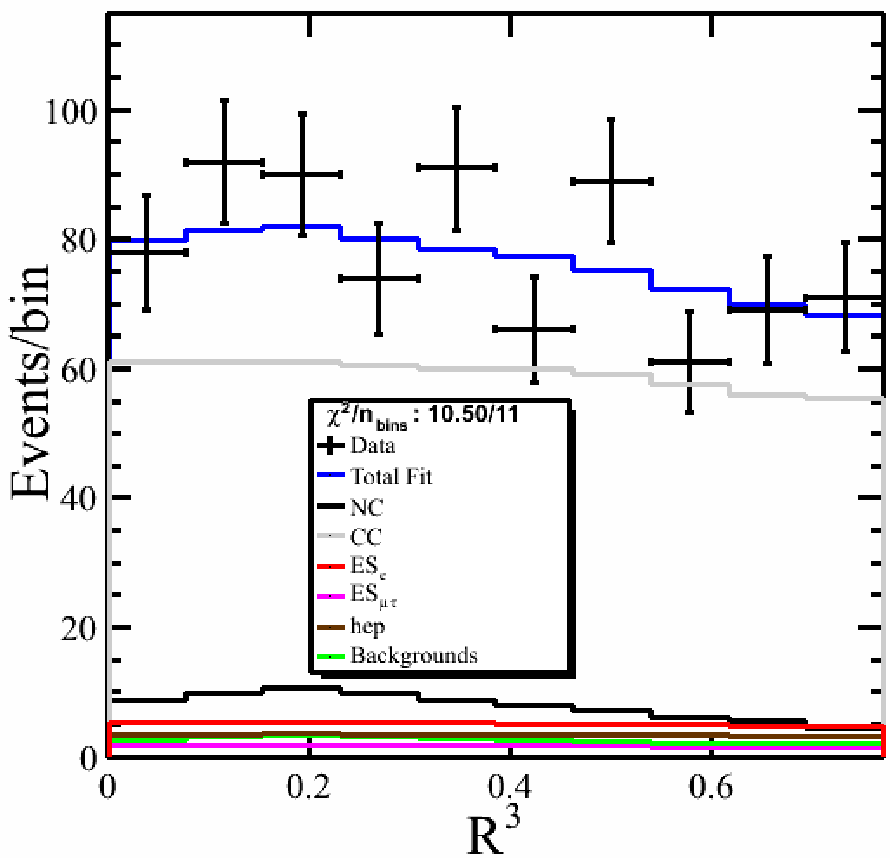


Figure 7.17: Projection of Čerenkov data for observable $\rho = (R/R_{AV})^3$ (captioned as R^3 on the x axis) overlaid with distributions from Monte Carlo simulation of the signals. The distribution, from the Monte Carlo simulation of the signal, is scaled by the number of events for the signal extracted from the fit. This test #4 was undertaken to test the likelihood function of the MCMC fit. Legend shows χ^2 with the number of bins in the histogram.

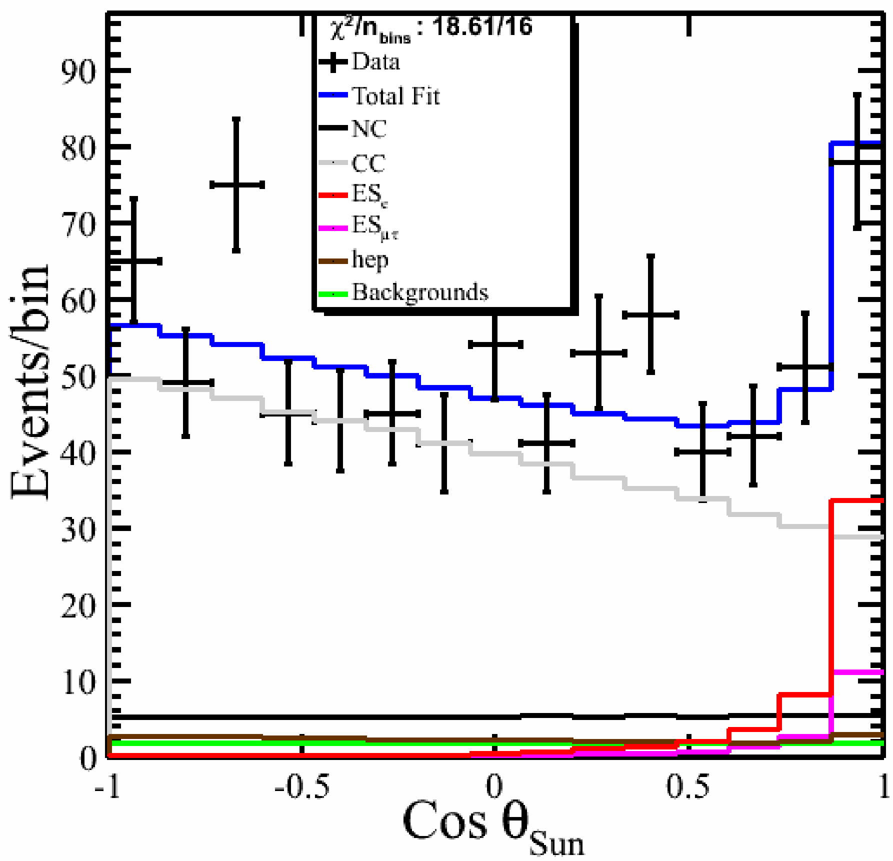


Figure 7.18: Projection of Čerenkov data for observable cos θ_{Sun} overlaid with known number of events for signals. This test #4 was undertaken to test the likelihood function of the MCMC fit. Legend shows χ^2 with the number of bins in the histogram.

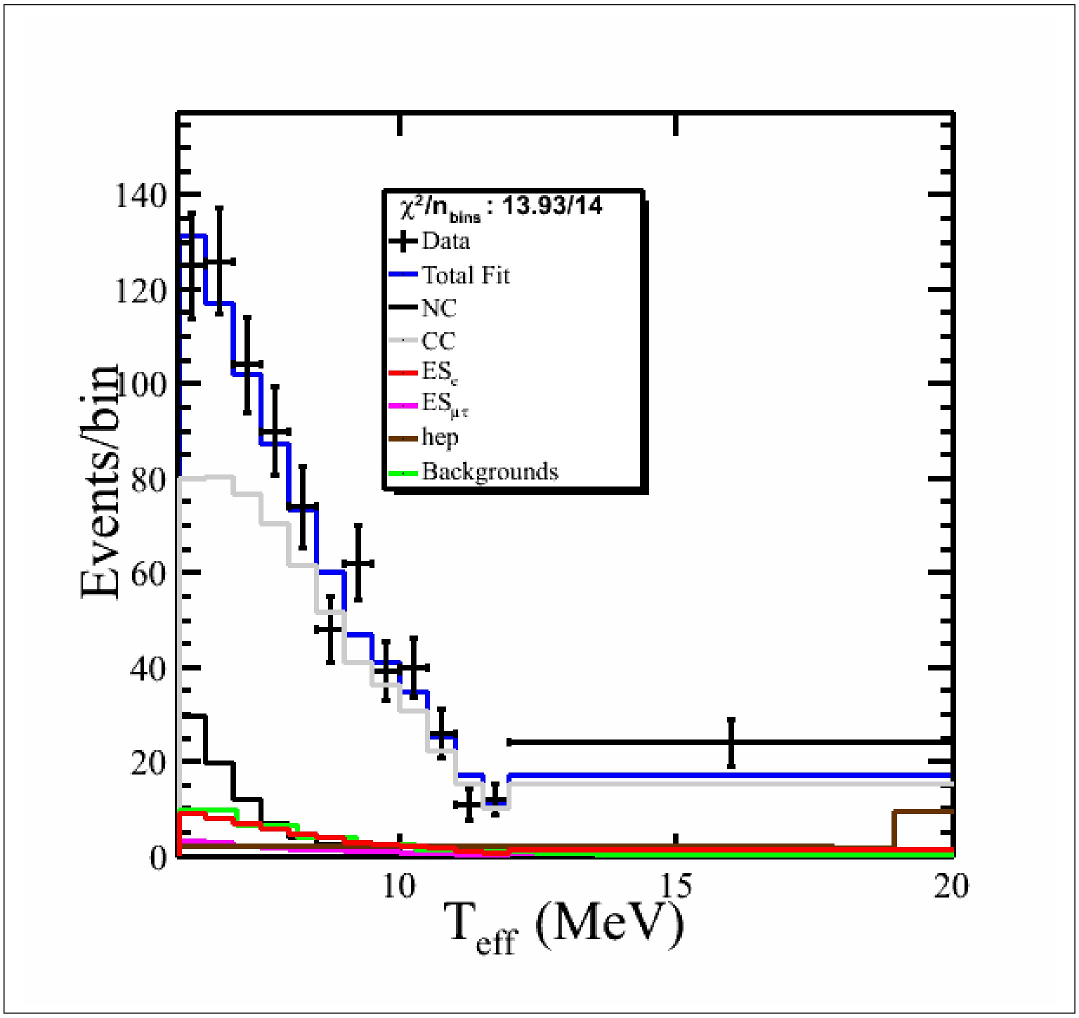


Figure 7.19: Energy spectrum $T_{\rm eff}$ in the Čerenkov data overlaid with energy spectra from NC, CC, ES, ES$_{\mu\tau}$ and all the backgrounds. This test #4 was undertaken to test the likelihood function of the MCMC fit. Legend shows χ^2 with the number of bins in the histogram.

7.1 Summary

This chapter outlined numerous tests undertaken to check the validity of the code. Getting an expected answer (for example 20/4=5 and 100/20=5) from the code is a necessary condition for the correctness of the code but it is not a sufficient condition; to ascertain that it is 20/4 and not 100/20, it is important to check each component that goes into the calculation of the expected answer (5) from the code and this chapter demonstrated that each component of the complicated program is doing exactly what it is supposed to be doing. Several improvements to the code came about while testing various components of the code so this was not an exercise in futility. The conclusion is that the MCMC fit can deliver the result. Various other tests were also performed which are described in section 11.5.1.

Chapter 8

MCMC Ensemble Test for a fit with 7 signal parameters, 19 systematic parameters and 3 constraints

This chapter describes the result of MCMC fit when the systematic parameters are allowed to float. The fit floats p_0, p_1, p_2, a_0, a_1, number of nc events, number of ex events, uncertainty in NC neutron detection efficiency in the Čerenkov data and 19 systematic parameters due to uncertainties in the shape of the PDFs. Three constraints are applied in the fit: total number of neutrons, number of external neutrons and flux-to-event ratio. There are five classes: neutral current (nc), acrylic vessel neutron photo-disintegrations (ex or av), charged-current (cc), elastic scattering for electron neutrinos (es_e), and elastic scattering for μ and τ neutrinos ($es_{\mu\tau}$). The data and PDFs are three dimensional in normalized radius-cubed (ρ), cosine of the event's direction from the Sun ($\cos\theta_\odot$), and the kinetic energy (E_m). The cuts applied in the signal extraction are: $0.0 \le \rho < 0.77025$, $1.0 \le \cos\theta_\odot < 1.0$ and 6 MeV $\le \mathrm{E}_m <$ 20 MeV. The SNOMAN Monte Carlo was used to build the PDFs for the signal extraction. The 3 dimensional (3D) PDF is built with the kinetic energy binning of 0.5 MeV between 6 MeV and 12 MeV and a single bin between 12 MeV and 20 MeV. Two systematic parameters applied on the energy scale and vertex scale take into account the fact that the systematics considered for the NCD analysis might have different values for the day and night data caused by possible time variations in the detector response. Additionally possible up-down asymmetries in the detector was considered for the **es** signal in terms of energy scale, energy resolution, vertex scale and the direction of the event – $\cos\theta_{\odot}$. For a detailed description of the systematics, see section 4.10.3.

8.0.1 Negative Log Likelihood (NLL) Equation

The NLL, used as a joint probability distribution in the MCMC fit, is described as:

$$\begin{aligned}\log\,\mathcal{L} = \sum_{i=1}^{2m} N_i - \sum_{d=1}^{N} \log\big(\sum_{i=1}^{2m}(N_i)F_i(\vec{x}_d,\vec{P})\big) \\ + \frac{(\overline{f}_{01} - N_0 - N_1)^2}{2\sigma_{01}^2} + \frac{(\overline{N}_1 - N_1)^2}{2\sigma_1^2} + \frac{(\overline{\epsilon} - \epsilon)^2}{2\sigma_\epsilon^2} \\ +\frac{1}{2}\sum_i(\frac{p_i - \bar{p}_i}{\sigma_{p_i}})^2 + \frac{1}{2}\sum_{i=0}^{2}\sum_{j=0}^{2}(b_i^{xy} - b_i^{\bar{x}\bar{y}})(b_j^{xy} - b_j^{\bar{x}\bar{y}})(V_{b^{xy}}^{-1})_{ij} \\ +\frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}(b_i^z - \bar{b_i^z})(b_j^z - \bar{b_j^z})(V_{b^z}^{-1})_{ij}\end{aligned} \tag{8.1}$$

where $\mathrm{F}_i(\vec{x}_d,\vec{P})$ is the probability density function, for the class i, giving the probability of observing the event d with observables $\vec{x}_d$ and with the current values of the fit parameters $\vec{P}$, $N_1, N_2, \ldots, N_m$ are the number of events for **m**=**5** event classes and **i** goes from 1 to **N** data entries. In the constraint terms, N_0 and N_1 are the number of neutral current events and external neutrons (EX) respectively for a current MCMC step and $\overline{N}_1$ is the constraint on the EX for a given simulated dataset. Similarly $\overline{f}_{01}$ is the constraint on the total number of neutrons and σ_{01} is its uncertainty. The current value of the flux-to-event ratio is ϵ, $\bar{\epsilon}$ is the constraint on ϵ and σ_ϵ is the uncertainty on the constraint. In the likelihood equation, p_i, $\bar{p}_i$ and σ_{p_i} represent the current value of the

PMT systematic parameter i in the MCMC fit, its mean and constraint width respectively. The next two terms are calculation of the constraint for the systematic uncertainties that are correlated. The sources of neutrons in this ensemble test are from neutral current and external neutrons. Table 8.3 gives a quick overview of the salient features of this ensemble test and section 8.2 describes the constraints in detail.

8.1 Description of Simulated Datasets

Before testing the MCMC fit on the real data, it was tested on a set of simulated data files. These sets were generated from events from the full SNOMAN Monte Carlo simulation to resemble the real data as closely as possible. Previous SNO results were used to estimate the number of events for each signal and *ex situ* measurements were used to estimate the number of background events expected in the analysis window. For a more detailed account of fake data set generation, consult reference [90].

Since the extended likelihood function implicitly assumes statistical fluctuations in the number of events of each signal type, these were randomly drawn from a Poisson distribution for each data set. Ensemble testing comprises running the full signal extraction on each of the simulated fake datasets to (I) fine tune the fitting algorithm, (II) to assure that the statistical and systematic uncertainties are properly propagated to the estimation of the number of events for each signal type in the fit, (III) to adjust the PDF configuration – if needed – and (IV) to make sure that the pull and bias distribution of the fitting parameters follow an expected pattern from fitting N number of simulated datasets. Table 8.1 lists the mean number of events used in a Poisson distribution for the generation of a number of events of each signal in the simulated datasets.

Event Class	Mean Number of events for the Poisson Distribution
cc	1845.276
es	161.607
$es_{\mu\tau}$	49.721
nc	240.569
ex	20.754

Table 8.1: Poisson parameter for each class used in the creation of the simulated datasets.

8.2 Constraints in the Fit

Three constraints were applied to the fit. Table 8.2 list the central values and the widths of the constraints. The central value of the total number of

Constraint On	Central Value	Width/Uncertainty of Central Value
f_{nc}	0.467578	0.00603
Number of External Neutrons	20.754	10.453115
Number of Neutrons	261.323	10.4533 (equation (8.3))

Table 8.2: List of constraints, their central values and the uncertainties on the central values.

neutrons ($\bar{N}$) in the fit is calculated as:

$$\bar{N} = N_{EX} + N_{NC} \tag{8.2}$$

where N_{EX} and N_{NC} are the mean number of external neutrons and the NC events used in the Poisson distribution for the generation of simulated datasets.

The width (σ) is calculated as:

$$\begin{aligned}\sigma_{EX} &= 10.453115 \\ \sigma_{NC} &= \frac{1}{\sqrt{N_{NC}}} \\ \sigma &= \sqrt{(\sigma_{EX}^2 + \sigma_{NC}^2)}\end{aligned} \tag{8.3}$$

8.3 Result

Two hundred and one independent fits were performed, each with a different random seed so that the fits are statistically independent.

Number of simulated datasets	201
Number of steps	55,000
Number of steps removed as burn-in	20,000
Number of constraints	3
Number of parameters	27
Number of event classes	5

Table 8.3: Quick overview of the ensemble test.

The bias and pull were calculated for each dataset and the distributions of these values, across the collection of 201 simulated datasets, were used to determine whether the extraction was bias free. The peak and RMS of the posterior distributions were used to get the best-fit value and its uncertainty. These values were used in equations (4.76) and (4.75) to calculate the pull and bias. The autocorrelation versus lag, shown in figures 8.1, 8.2 and 8.3, was plotted to get an estimate of the burn-in period. Autocorrelation function is covered in section 5.2. Pull distributions are shown in figures 8.4 and 8.8; except neutron detection efficiency ϵ_{nc}, the pull of all other parameters are approximately 1.0. The reason for pull width of ϵ_{nc} (Last row and 4^{th} column in table 8.4 – 0.7159) to be less than 1.0 is that ϵ_{nc} is also constrained by the total number of neutrons. Bias distribution is shown in figure 8.6 while figure 8.7 illustrates bias in terms of its uncertainty to ascertain that the biases of fit parameters are zero within their uncertainties, that is, all fit parameters agree with the Monte Carlo inputs.

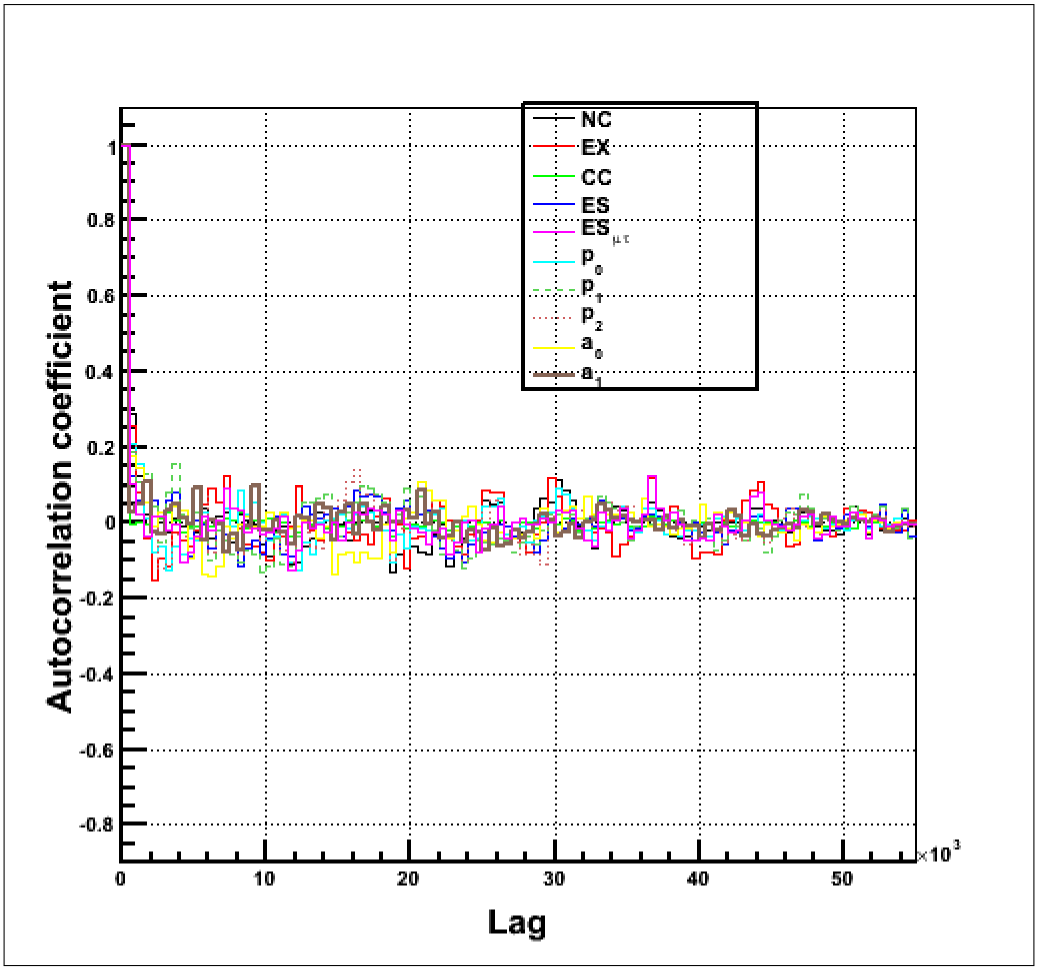


Figure 8.1: Autocorrelation coefficient versus lag for floating the systematics. This plot shows the autocorrelation of signals and the parameters of survival probability equation. From this plot, burn-in of 20,000 steps was selected.

The width of the pull of external neutrons is more than 1, as seen in figure 8.4, because the likelihood function (posterior distribution) was not symmetric for some of the simulated datasets resulting in the long tail in the pull distribution of the external neutrons which is shown in figure 8.5.

The arrangement of the Chapters in the dissertation gives a false sense

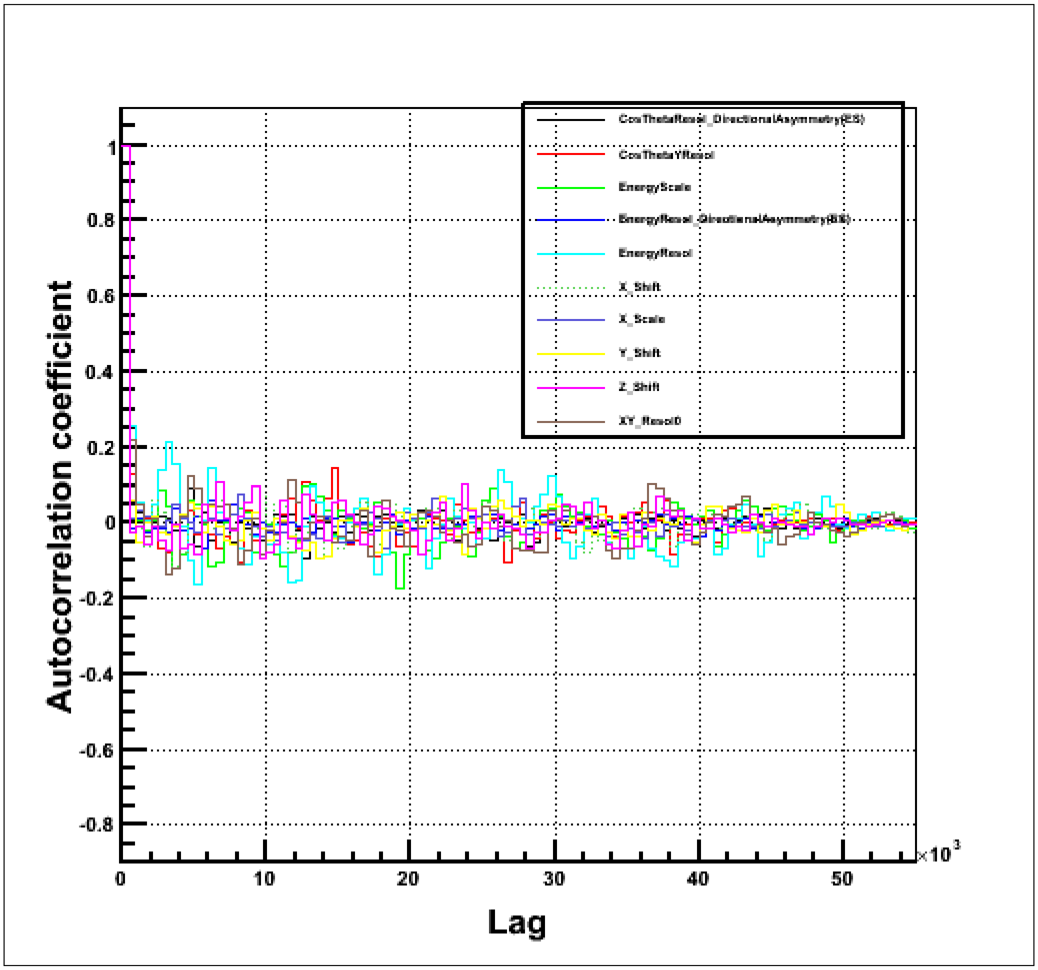


Figure 8.2: A plot showing autocorrelation coefficient versus lag of various systematic parameters. In total 19 systematics were floated in the fit; autocorrelation of ten is shown in this diagram.

of timing; the result, shown in this chapter, is from the earliest version of the code which was run at Western Canada Research Grid (WestGrid [120]). While the development was going on at the University of Alberta to include the PSA constraint and the day-night asymmetries for the external neutrons and D_2O photo-disintegrations ($d_2$opd), the code at West Grid was being developed to test the implementation of the systematic uncertainties listed in table 8.5. Once the MCMC fit with the PSA constraint was fully formed, testing of the code started at the University of Alberta; first running MCMC code with a limited number of backgrounds (result described in chapter 9) and next

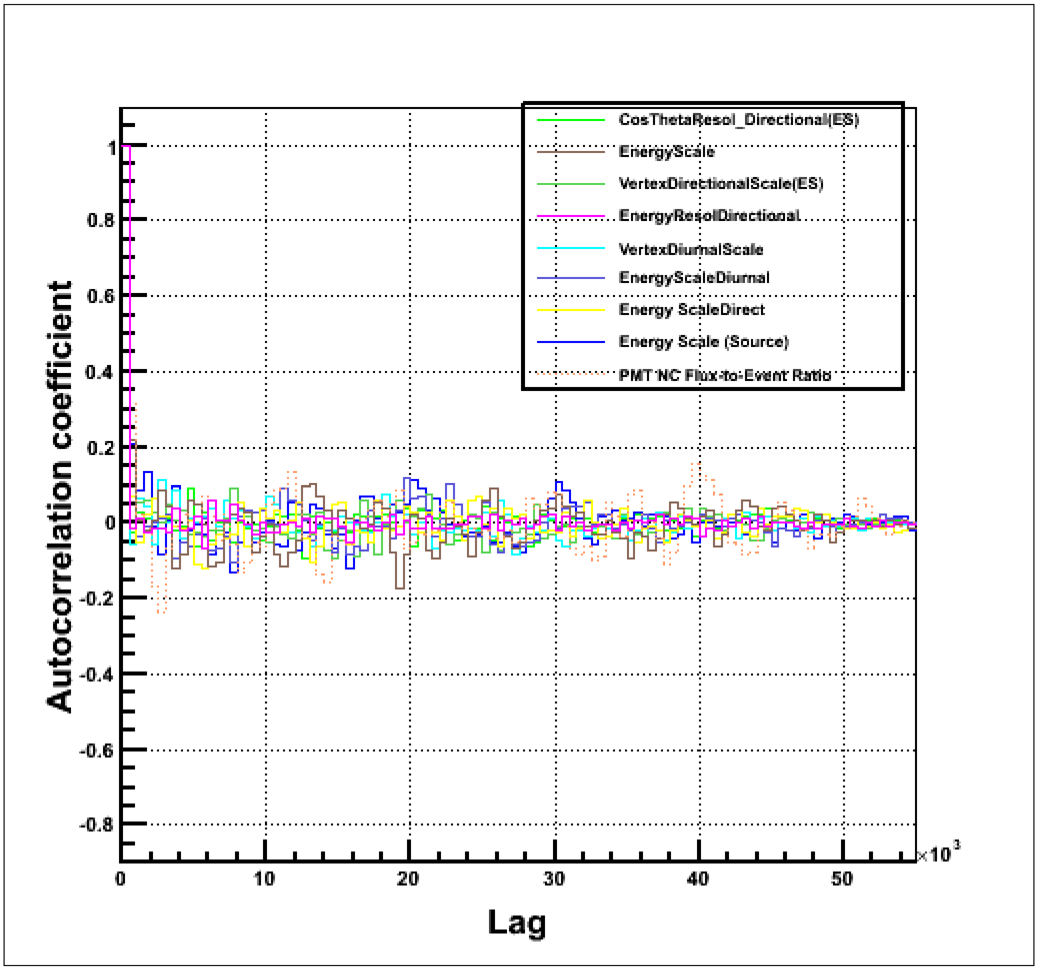


Figure 8.3: Autocorrelation coefficient versus lag of additional systematic parameters.

including all the backgrounds (results described in chapter 10). These tests did not include floating the systematic uncertainties except neutron detection efficiencies for the PMTs and NCDs and, as mentioned above, the day-night asymmetries for the external neutrons and D_2O photo-disintegration. These MCMC fits, with and without floating the systematic uncertainties, were running simultaneously on WestGrid and the University of Alberta. After the successful implementation of PSA constraint, constraint from LETA was added; the results are described in chapter 10.

As mentioned, the MCMC code went through several major developments and cross checks were performed for each new addition to the code. During

	Pull	**Bias**	**Pull Width**	**Bias Error**
nc	-0.148414	-0.0111318	1.03769	0.00694771
ex	-0.25303	0.0297015	1.73406	0.0352745
p_0	0.00691543	0.0131642	0.919249	0.00807544
p_1	0.0683312	-0.0654726	1.16261	0.166914
p_2	-0.0420322	-0.337115	0.92251	0.49617
a_0	-0.020796	-0.0304478	1.19134	0.162445
a_1	-0.00905473	-0.191791	1.08107	0.835545
f_{nc}	-0.0382465	-0.00125871	0.715907	0.00130726

Table 8.4: Pull and bias data, in tabular form, used to plot figures 8.4, 8.6 and 8.7.

the cross checks it was discovered later (while fitting third of the real data) that the equation to calculate the number of events for CC, ES and ES$_{\mu\tau}$ has an incorrect factor in it, that is, the number of events were multiplied by the fiducial volume correction of neutral current (nc) rather than their own. This bug escaped notice for two reasons (1) it is relevant only when the systematics are floated and most of the tests performed did not float the systematic uncertainties and (2) the effect is small for ES and ES$_{\mu\tau}$ and matters only for CC because the number of CC events dominates number of events from all other event classes (table 8.1). Following equations were used instead of the correct equations (which had S$_{cc}$ instead of S$_{nc}$) described for CC in equations (4.63) and (4.64).

$$\chi = \chi_D + \chi_N \tag{8.4}$$

$$\chi_D = f_{nc}\, S_{nc}\, \phi_{nc}\, (\frac{\sigma_\chi\, \epsilon_\chi}{\sigma_{nc}\, \epsilon_{nc}}) p_{ee_d}\, R_D \tag{8.5}$$

$$\chi_N = f_{nc}\, S_{nc}\, \phi_{nc}\, (\frac{\sigma_\chi\, \epsilon_\chi}{\sigma_{nc}\, \epsilon_{nc}}) p_{ee_n}\, (1 - R_D) \tag{8.6}$$

where χ =CC, ES and ES$_{\mu\tau}$, S$_{nc}$ is the fiducial volume correction for NC, R_D is the ratio of the number day events to the total number of events in the Monte Carlo, ϕ_{nc} is the ^{8}B flux, f_{nc} is a flux-to-event conversion factor and

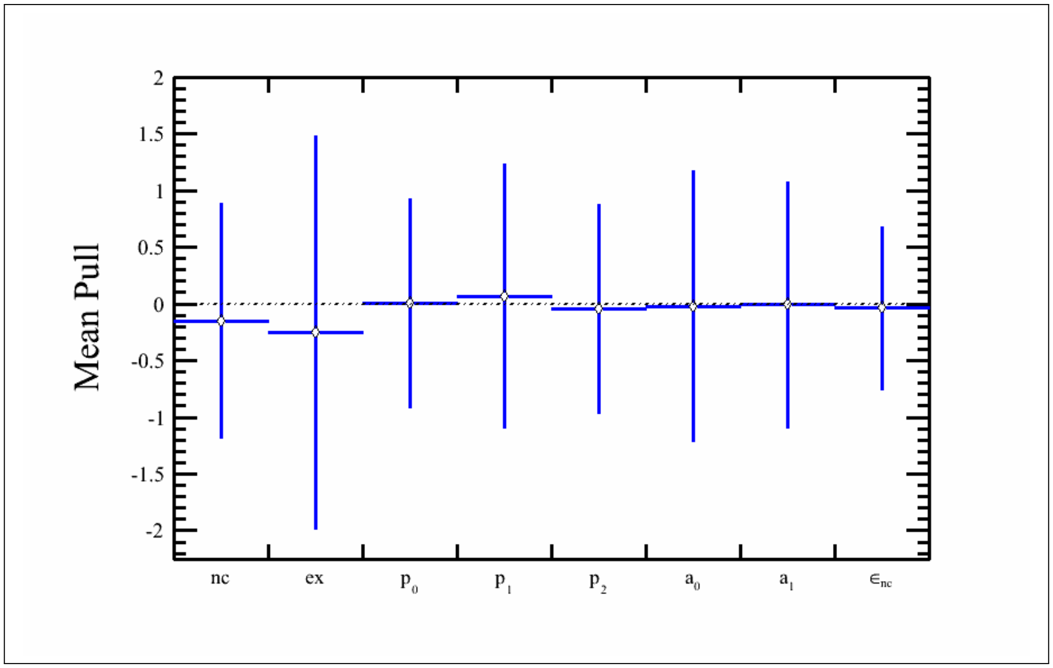


Figure 8.4: Pull distribution of the MCMC fit for the case where systematic parameters were allowed to float. Table 8.4 was used to plot this figure. The fit has five event classes, three constraints and 27 fit parameters. This plot shows the pull distribution for the 8 parameters of the MCMC fit. The peak and RMS of the posterior distributions were used to get the best-fit value and its uncertainty.

the variable p_{ee_d} (p_{ee_n}) is a ratio of the number of events for day (night) classes with given values of P_{ee} (from equations (4.55) and 4.57 or equations (4.58) and (4.59)) to the number of events with P_{ee} equal to 1.0.

The uncertainty in using the wrong value of fiducial volume correction is calculated as $(S_{cc} - S_{nc})/S_{cc}$ where S_{nc} and S_{cc} are from the first step of the Markov chain. From the 8 Markov chains from a fit on the full data, the maximum uncertainty is 6%. The purpose of this fit was to check the implementation of the systematic uncertainties while various fits, without floating the systematic uncertainties, were being tested at UofA. We wanted to make sure that when we start testing MCMC code that includes the systematics the code has already been tested in terms of step sizes and acceptance. So from that perspective, the code at WestGrid was a success.

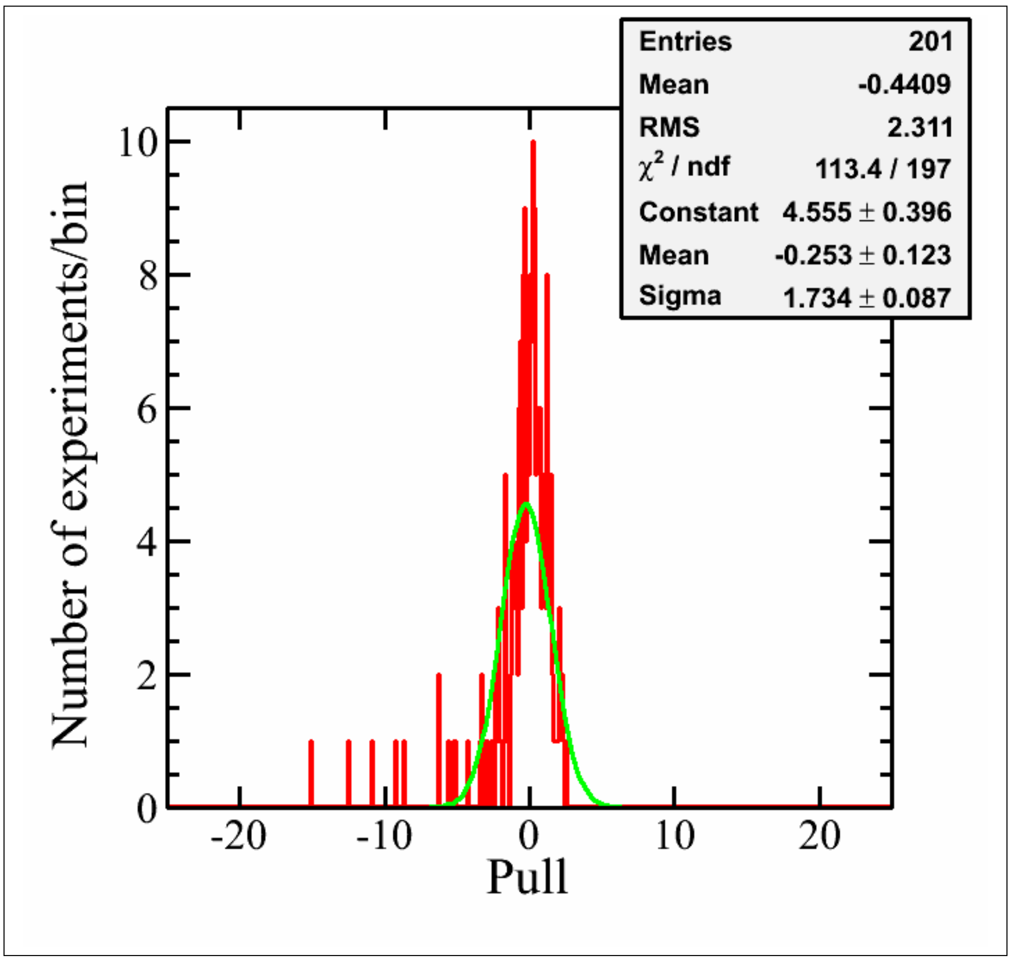


Figure 8.5: Long tail on the left is the reason for the wider pull width of the external neutrons shown in figure 8.4.

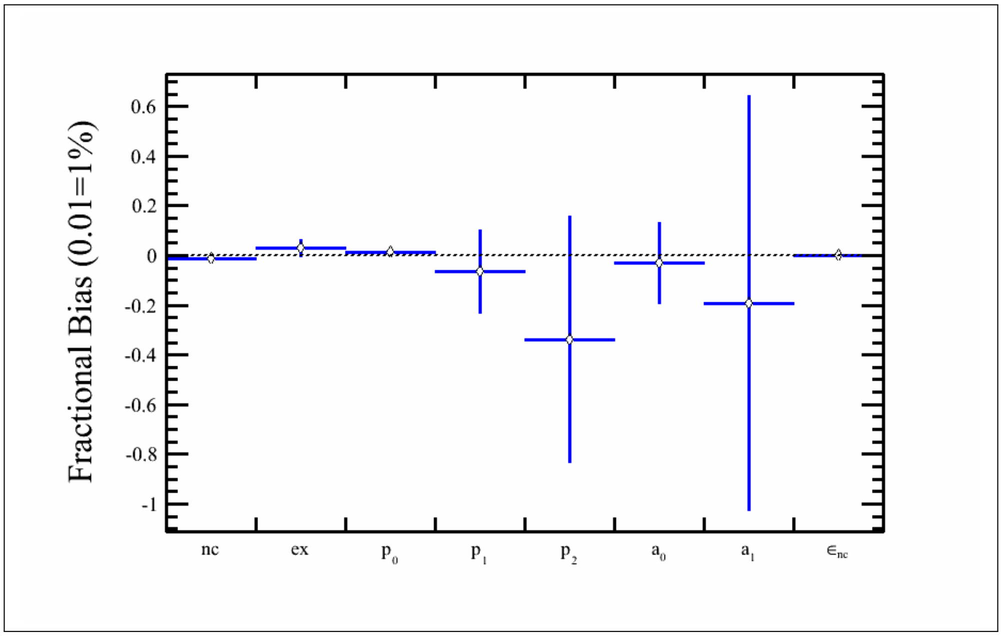


Figure 8.6: Bias distribution of the MCMC fit while floating the systematic parameters. Table 8.4 was used to plot this figure. The fit has five event classes, three constraints and 27 fit parameters. This plot shows the bias distribution for the 8 parameters of the MCMC fit. The peak and RMS of the posterior distributions were used to get the best-fit value and its uncertainty.

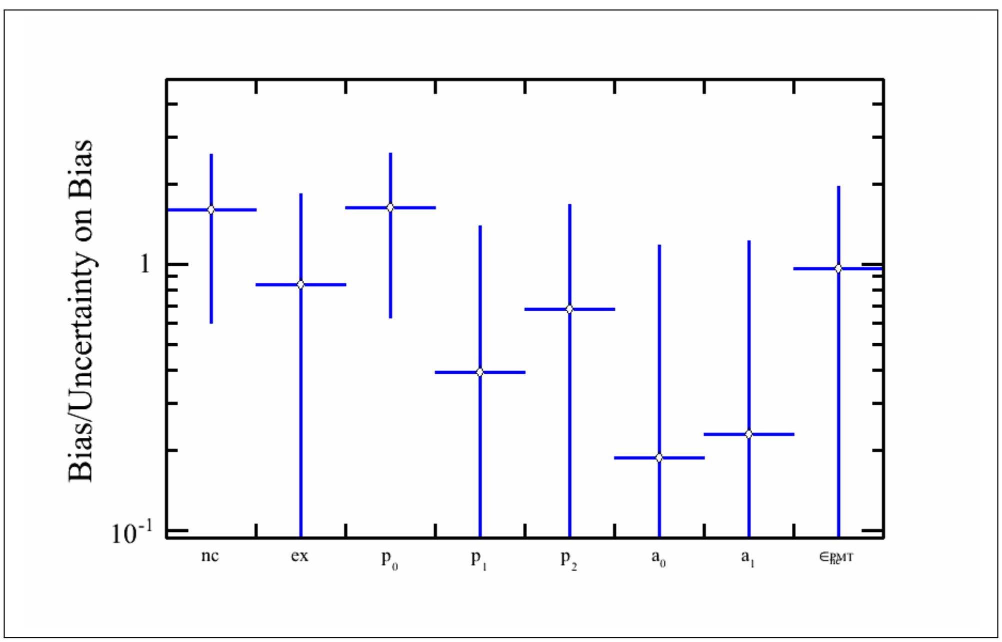


Figure 8.7: The plot shows bias divided by the uncertainty on the bias for the case of floating the systematic parameters. This plot shows that the bias for the fit parameters is consistent with zero for an ensemble test involving 201 simulated datasets.

Name	**Description**	**Pull**	**Pull Width**
Sys_1	Direction Asymmetry in $\cos\theta$ for ES only	0.0233341	1.03253
Sys_2	$\cos\theta$ Resolution	0.114876	0.943028
Sys_3	Energy Scale	-0.092193	1.04703
Sys_4	Directional Asymmetry in Energy Resolution for ES only	-0.102643	1.059
Sys_5	Energy Resolution	-0.0257215	0.950999
Sys_6	X Shift	-0.0593584	1.08257
Sys_7	Vertex Scale	-0.0748767	0.999712
Sys_8	Y Shift	0.0396525	1.01556
Sys_9	Z Shift	0.0262243	1.13231
Sys_{10}	XY Resolution – constant term	-0.0554756	1.05777
Sys_{11}	XY Resolution – linear term	-0.0847274	0.988964
Sys_{12}	XY Resolution – quadratic term	-0.0401506	1.03712
Sys_{13}	Z Resolution – constant term	0.155381	1.05093
Sys_{14}	Z Resolution – linear	0.110996	0.987656
Sys_{15}	Vertex Diurnal	0.0385573	0.941561
Sys_{16}	Vertex Direction	-0.0228358	0.876688
Sys_{17}	Energy Scale Diurnal	0.0338809	0.972683
Sys_{18}	Energy Scale Direction	-0.0114926	0.93307
Sys_{19}	Energy Scale (correlated)	0.0181612	1.09638

Table 8.5: Data, in a tabular form, to plot figure 8.8.

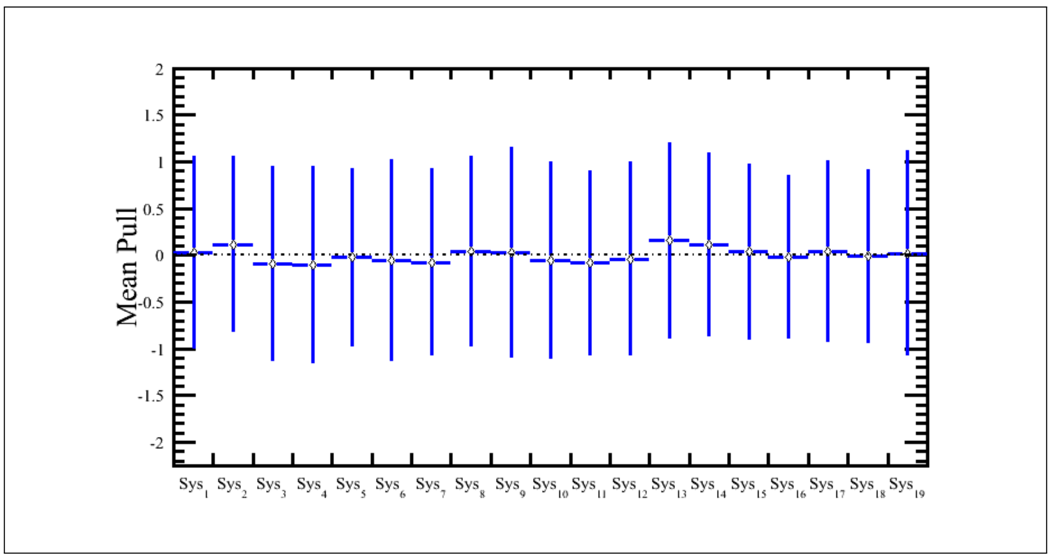


Figure 8.8: Pull spread of the systematics. This plot shows that the systematics are correctly treated in the MCMC fit because the pull is approximately zero and the width of pull is approximately one for all the systematics. The mean of the posterior distribution was used in the calculation of the pull.

8.4 Summary

This chapter described the result from the first MCMC ensemble test. This is a simple fit as it does not float the day-night asymmetries for the backgrounds from external neutrons and $d_2$opd. Also the constraint on the total number of neutrons is the mean Poisson number of neutrons in the simulated datasets. The pull (figures 8.4 and 8.8) and bias plots (figures 8.6 and 8.7) show that the code is ready to tackle the next challenge, that is, add PSA constraint, additional backgrounds besides the external neutrons and float the day-night asymmetries for two of the backgrounds.

Chapter 9

Adding the Constraint from Pulse Shape Analysis (PSA) and Backgrounds

This chapter describes Monte Carlo Markov Chain (MCMC) fit on 448 simulated datasets (version 4) using PSA method described in section 4.10.9. Details on simulated data generation are described in reference [90]. For this ensemble test, following are the parameters in the likelihood function of the NCD phase: Neutral Current (nc), D_2O neutron backgrounds ($d_2$opd), acrylic vessel neutron photo-disintegrations (ex), neutron backgrounds due to hot spots on K2 and K5 NCD strings, atmospheric neutrino backgrounds (atmos), NCD background neutrons (ncdpd), hepCC, hepES and hepNC. The last three are due to interactions of hep neutrinos in the heavy water. Since ncdpd, k2pd and k5pd have limited statistics, to test the PSA method, these were not added; chapter 10 describes the result when all the backgrounds were included. The number of events for hepCC, hepES ad hepNC are fixed in the MCMC fit and the number of events for Charged Current (cc), Elastic Scattering (es) due to electron neutrinos, Elastic Scattering due to ν_μ and ν_τ neutrinos ($es_{\mu\tau}$) are calculated using the ^{8}B flux and the survival probability equations as described in chapter 4. Besides floating the number of events for the signals and backgrounds, other floating parameters are three P_{ee} variables

(p_0, p_1 and p_2), two day-night asymmetries (a_0 and a_1) to take into account the matter effect caused by propagation of neutrinos through matter in the Sun and the Earth and the day-night asymmetries for the two backgrounds (external neutrons and $d_2$opd). The PSA method added a constraint from pulse shape analysis to the fit (equation (9.6)). For details on this analysis, consult reference [90]. A quick overview of the ensemble test is listed in a table 9.1.

The flux-to-event ratios, to convert flux to number of events, are calculated as:

$$f_{nc}^{\text{PMT}} = N_{nc}/\phi_{MC} \approx 0.46758 \tag{9.1}$$

$$f_{nc}^{\text{NCD}} = \frac{13.2744702 * 392.89 * 0.211 * 0.8250158}{514.5} = 1.76460 \tag{9.2}$$

In equation (9.1), N_{nc} =240.569 is the actual number of NC events/SSM set and $\phi_{MC} = 514.5 \times 10^4\ \nu\ \text{cm}^{-2}\ s^{-1}$ is the input SSM ^{8}B ν flux in the SNOMAN Monte Carlo. In equation (9.2), 13.27 is the number of neutrons per day expected to interact in the SNO detector, 392.89 days is the lifetime of the NCD phase, 0.211 is the NCD neutron capture efficiency and 0.8250158 is the correction applied to the efficiency to account for the differences between the data and the Monte Carlo simulation used in the fit (See reference [90] for details). From herein, flux-to-event ratios $\text{f}_{nc}^{\text{PMT}}$ and $\text{f}_{nc}^{\text{NCD}}$ will be used synonymously with ϵ_{nc}^{PMT} and ϵ_{nc}^{NCD}

9.0.1 Generation of the Simulated datasets

The simulated dataset consisted of 448 files; for each file, the number of events for the signals were randomly drawn from a Poisson distribution with means listed in column two of table 9.2. These means will hereafter be called Poisson means. Another set of simulated data was created using a different seed and this set is called an *alternate* dataset while the first set is called a *regular*

Number of MCMC steps	250000
Burn-in	30000
Number of parameters floating	13
Total number of classes	10
Number of fixed classes	3
Number of constraints	7

Table 9.1: Quick overview of the ensemble test using constraint from PSA.

dataset. The reason for creating the alternative dataset was to determine the role of statistics in the pull and bias of the fit. Columns 3 and 4 list the mean number of events for each class in the regular simulated dataset and the alternate dataset. The number of events for CC, ES and $\text{ES}_{\nu\mu}$ were calculated from ^{8}B flux, P_{ee} parameters (p_0, p_1 and p_2) and the day-night asymmetry (a_0 and a_1). These parameters (p_0, p_1, a_0, a_1) are listed in a table 9.3. Once the number of events were determined for CC, ES and $\text{ES}_{\mu\tau}$, these were used as the mean number of events in the Poisson distribution ("Poisson means"), which was utilized to randomly draw the number of events, for all the event classes, to create each simulated data file.

9.0.2 Constraints on the Fit

Seven constraints were applied to the fit. Table 9.4 lists central values of the constraints along with associated uncertainties on the constraints. To achieve a correct pull distribution, the constraint is randomly drawn from a Gaussian distribution for each simulated dataset. If this is done properly and the fit is correct then the pull plot, displayed in figure 10.2, should show that the width of the pull for all the floating variables is consistent with one. Since the width of the constraint on the external neutrons is 50.37% of the constraint, the mean and the centroid of the Gaussian function do not coincide because the constraint is restricted to be positive. Therefore equation (9.3) was utilized to

Event Class	Mean Poisson Parameter	Mean from Regular Simulated dataset	Mean from Alternate Simulated dataset
CC	1845.276	1845 ± 44	1843 ± 42.09
ES	161.607	161.85 ± 13.42	161.59 ± 12.29
$ES_{\mu\tau}$	49.721	49.88 ± 7.14	49.8 ± 7.1
NC	240.569	240.3 ± 15.3	240.9 ± 21.4
EX	20.754	21.05 ± 4.46	20.623 ± 4.613
$d_2$opd	8.305	8.2 ± 2.8	8.265 ± 2.812
Atmos	24.681	24.57 ± 4.99	24.9 ± 4.7
hepCC	12.844	12.692 ± 3.55	12.835 ± 3.459
hepES	1.068	0.70 ± 0.67	0.7511 ± 0.6850
hepNC	1.156	1.03 ± 0.98	1.0616 ± 1.018

Table 9.2: Properties of the regular and alternate simulated datasets; column 2 lists the mean number of events used in the Poisson distribution to generate the simulated datasets and column 3 and 4 has the actual mean number of events from the simulated datasets – regular and alternate respectively. These were determined by plotting the number of events, belonging to each event class, in a separate histogram; the mean and RMS of the histogram is reported as mean number of events and its uncertainty in columns 2 and 3.

find the Centroid (μ), corresponding to the mean ($\bar{x}$) randomly drawn from the Gaussian distribution, to be applied in the calculation of the likelihood function (9.6).

$$\bar{x} = \mu + \frac{\sqrt{\frac{2}{\pi}}\sigma \ \exp\left(\frac{-\mu^2}{2\,\sigma^2}\right)}{Erfc\left(\frac{-\mu}{\sqrt{2}\sigma}\right)} \tag{9.3}$$

$$\mathrm{Erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2}\, dt \tag{9.4}$$

where $\bar{x}$ and μ are the mean and the centroid of the Gaussian function respectively. For detail on this, consult Chapter 6.

Parameter	**Actual Value**
p_0	0.325
p_1	-0.0088
p_2	0.00122
a_0	0.028
a_1	0.00478

Table 9.3: Table lists values of P_{ee} parameters and day-night asymmetry used in the generation of the simulated datasets.

Constraint On	**Central Value**	**Width/Uncertainty on the Central Value**
$\mathrm{f}_{nc}^{\mathrm{NCD}}$	1.76460447	0.041815285
$\mathrm{f}_{nc}^{\mathrm{PMT}}$	0.46758	0.0129054351%
External neutrons (EX)	1.0	10.453115/24.754
Day night asymmetry of EX	0.0	0.0112
Atmospheric neutrinos	1	4.8999/24.681
D_2O background ($d_2$opd)	1.0	1.28594/8.305
Day night asymmetry for $d_2$opd	0.0	0.112
PSA	997.752	4.5%

Table 9.4: List of Constraints, their central values and the uncertainties on the central values.

9.0.3 Negative Log Likelihood Equation

The NLL, used as a joint probability distribution in the MCMC fit, is described as:

$$
\begin{aligned}
-\log\,\mathcal{L} = & \sum_{i=1}^{2m} N_i - \sum_{d=1}^{N} \log\Big(\sum_{i=1}^{2m} (N_i) F_i(\vec{x}_d, \vec{P})\Big) \qquad (9.5)\\
& + \frac{(\mathrm{PSA} - B\,\epsilon_1 - N_1\kappa_1 - N_2\kappa_2 - N_3\kappa_3 - hep_{nc})^2}{2\,(\sigma_P)^2}\\
& + \frac{(\overline{\alpha}_1 - \alpha_1)^2}{2\sigma_1^2} + \frac{(\overline{\alpha}_2 - \alpha_2)^2}{2\sigma_2^2} + \frac{(\overline{\alpha}_3 - \alpha_3)^2}{2\sigma_3^2}\\
& + \frac{(\overline{\xi}_0 - \xi_0)^2}{2\sigma_{\xi_0}^2} + \frac{(\overline{\xi}_1 - \xi_1)^2}{2\sigma_{\xi_1}^2} + \frac{(\overline{\epsilon}_1 - \epsilon_1)^2}{2\sigma_{\epsilon_1}^2} + \frac{(\overline{\epsilon}_0 - \epsilon_0)^2}{2\sigma_{\epsilon_0}^2}\\
& + \frac{1}{2}\sum_i \Big(\frac{p_i - \bar{p}_i}{\sigma_{p_i}}\Big)^2 + \frac{1}{2}\sum_{i=0}^{2}\sum_{j=0}^{2} (b_i^{xy} - b_i^{\bar{xy}})(b_j^{xy} - b_j^{\bar{xy}})(V_{b^{xy}}^{-1})_{ij}\\
& + \frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1} (b_i^z - b_i^{\bar{z}})(b_j^z - b_j^{\bar{z}})(V_{b^z}^{-1})_{ij} \qquad (9.6)
\end{aligned}
$$

where $\mathrm{F}_i(\vec{x}_d, \vec{P})$ is the probability density function, for the class i, giving the probability of observing the event d with observables $\vec{x}_d$ and with the current values of the fit parameters $\vec{P}$, hep_{nc} is the number of neutral current events initiated by hep neutrinos, $N_1, N_2, \ldots, N_m$ are the number of events for **m**=**10** event classes and **i** goes from 1 to **N** data entries. In the constraint terms, α_1, α_2 and α_3 are the ratios of the average rate to the nominal rate for the current MCMC step and the $\overline{\alpha}_1$, $\overline{\alpha}_2$ and $\overline{\alpha}_3$ are the constraints for EX, $d_2$opd (photo-disintegration) and atmospheric neutrinos respectively for a given simulated dataset. The day-night asymmetries for the external neutrons and $d_2$opd are represented by ξ_0 and ξ_1 respectively. The flux-to-event for the NCDs and PMTs are represented by ϵ_1 and ϵ_0 respectively. The ^{8}B flux is designated by **B** and **PSA** is the PSA constraint for the current simulated dataset and σ_P is the width of the PSA constraint. P_{ee} parameters (p_0, p_1 and p_2) and the day-night asymmetry (a_0 and a_1) were used to calculate the number of events for CC, ES and $\mathrm{ES}_{\mu\tau}$. In the likelihood equation, p_i, $\bar{p}_i$ and σ_{p_i} represent the current value of the PMT systematic parameter i in the MCMC fit, its mean and constraint width respectively. The last two terms are calculation of the constraint for the systematic uncertainties that are correlated.

9.0.4 Result of Pull and Bias Testing

Results of the pull and bias for the fit are shown in figures 9.2, 9.4 and 9.5. The top plot shows the result of fitting the regular data and the bottom plot shows the result of fitting the alternate data. For the pull and bias distributions, the fit value and its uncertainty are the peak and RMS of the posterior distribution respectively. Since ^{8}B flux (measured by neutral current interactions) is highly correlated with P_{ee} parameter p_0, as seen in figure 9.1, we decided to use 2 Dimensional marginal likelihood function [91] to determine the best-fit for nc instead of the 1D we used for all the other parameters. For a likelihood function consisting of 5 parameters, the 2D marginal likelihood is described as:

$$h(x_1, x_2) = \int_{x_3(min)}^{x_3(max)} \int_{x_4(min)}^{x_4(max)} \int_{x_5(min)}^{x_5(max)} \mathcal{L}(x_3, x_4, x_5) dx_3 dx_4 dx_5 \tag{9.7}$$

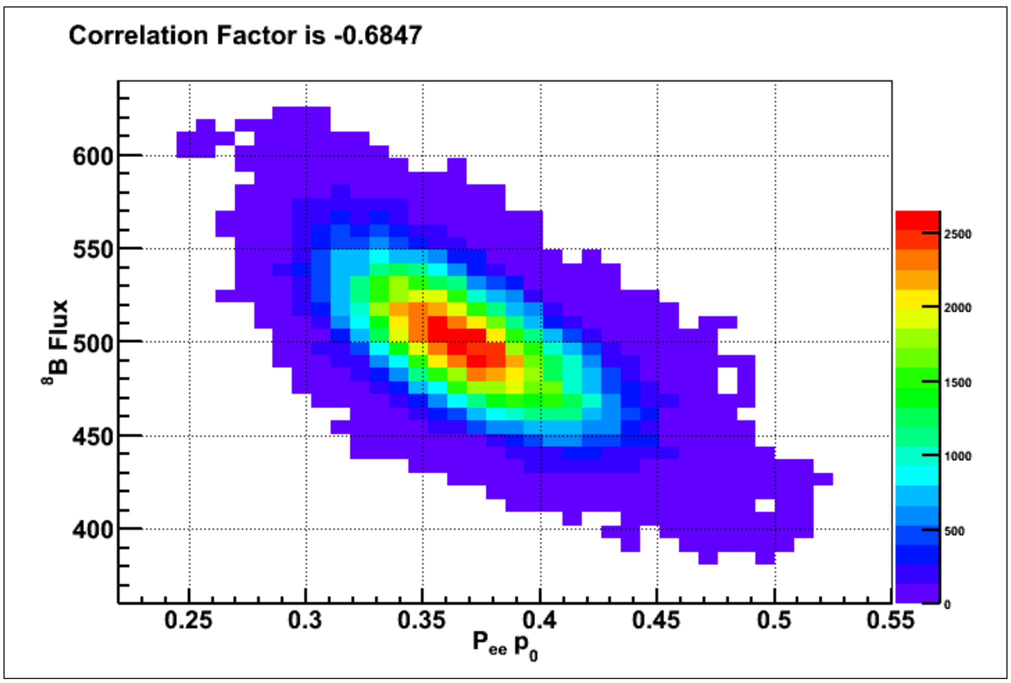


Figure 9.1: Correlation of ^{8}B flux with P_{ee} parameter p_0; the peak of the 2D histogram, in the ^{8}B flux dimension, is used as the best-fit in the calculation of the pull and bias shown in the second row of tables 9.5 and 9.6. Unit of ^{8}B flux is 10^4 cm^{-2} s^{-1}.

Parameter	**Pull**	**Bias**	**Width of Pull**	**Uncertainty on Bias**
nc^{1d}	-0.230942	-0.0125036	1.00191	0.0027206
nc^{2d}	-0.115403	-0.00580357	0.997218	0.00271135
ϵ_{nc}^{NCD}	0.0853571	0.00202282	0.968046	0.00107979
p_0	0.0397772	0.011134	0.982776	0.00437695
p_1	0.093125	-0.123214	1.02329	0.0916266
p_2	-0.00883929	-0.0585268	0.971979	0.3333
a_0	-0.0582154	-0.0979688	1.01031	0.0939415
a_1	-0.0213843	-0.248482	1.00527	0.520914
ϵ_{nc}^{PMT}	0.0250453	0.00048552	1.02351	0.00062449
ex	-0.130737	0.00133945	1.48868	0.0260593
ex_{Asym}	-0.0392864		1.00946	
d_2opd	-0.00272324	-0.00042858	0.979591	0.00717938
d_2opd_{Asym}	-0.077323		1.01282	
Atmos	0.0641083	0.0132866	1.00768	0.00942613

Table 9.5: Pull and Bias in tabulated form, for the regular dataset consisting of 448 files, to plot figures 9.2, 9.4 and 9.5. The pulls and biases were calculated using the peak and RMS of the posterior distributions. The actual value of day-night asymmetry in the simulated data is zero hence bias calculation for ex_{Asym} and d_2opd_{Asym} is not possible (equation (4.75)).

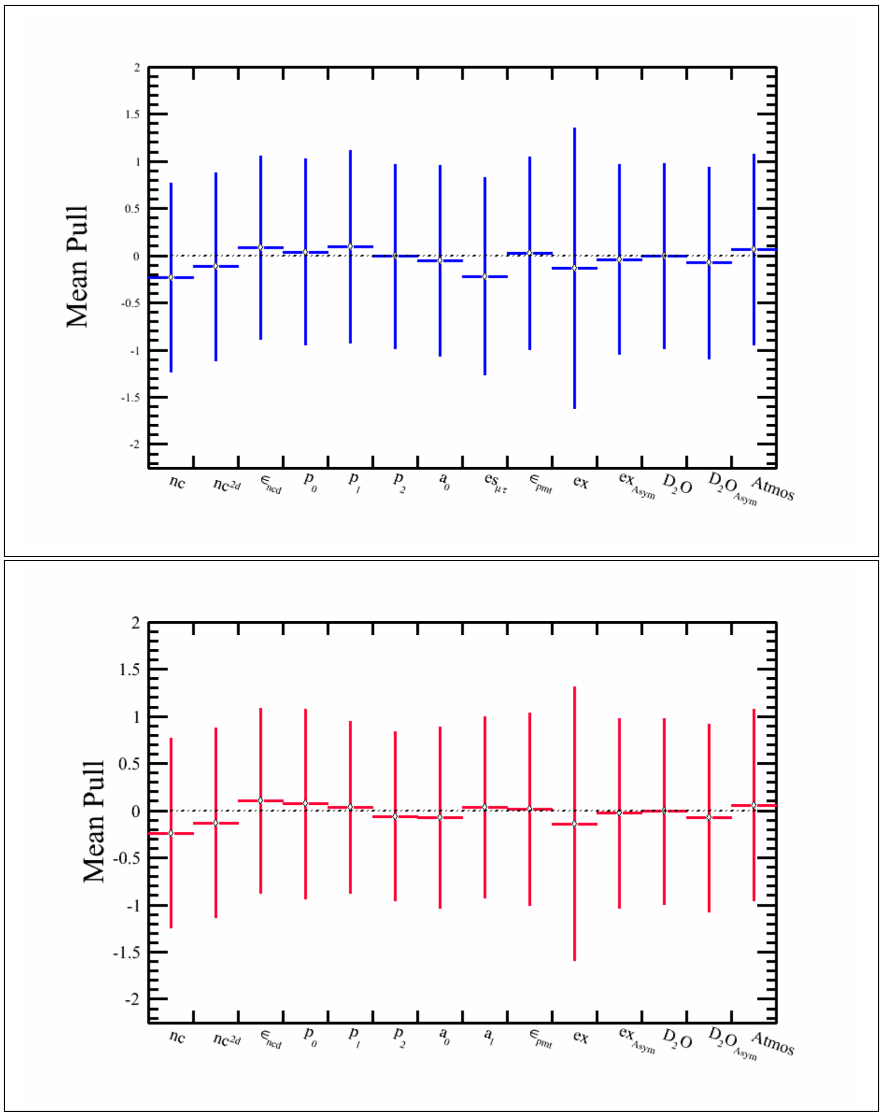


Figure 9.2: These are pull plots for the regular dataset blue, as well as, the alternate dataset in red.

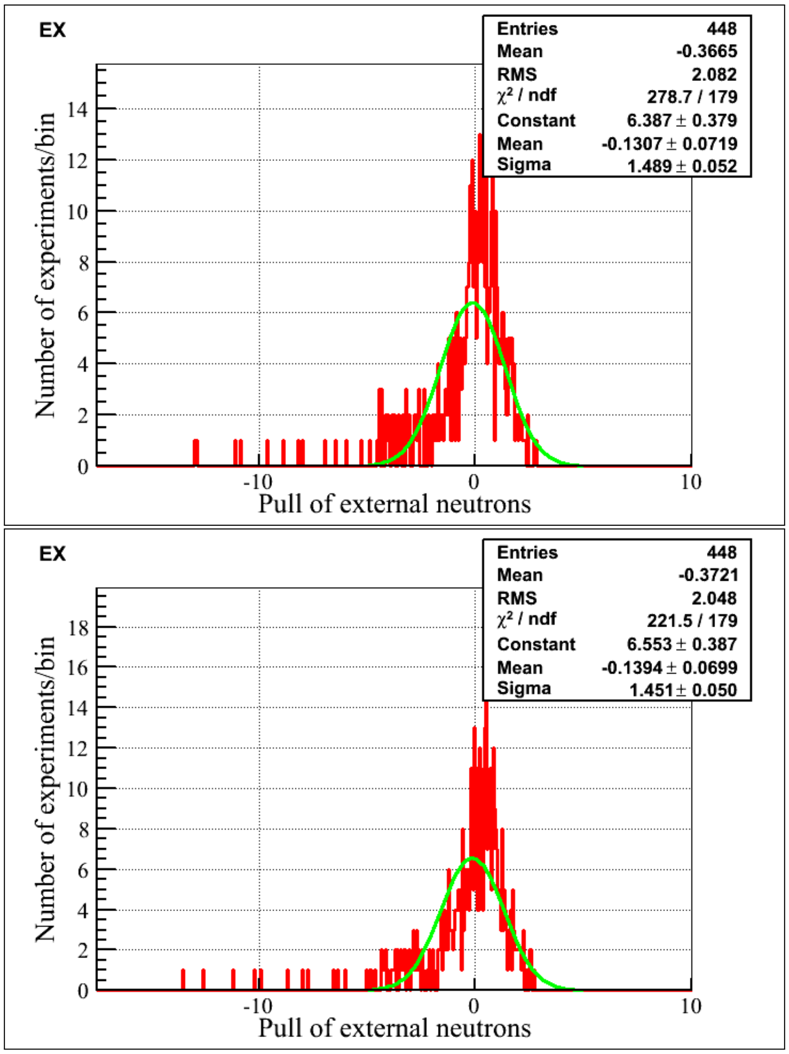


Figure 9.3: The pull distributions of external neutrons. The top plot is for the regular dataset and the bottom plot is for the alternative dataset. The tails cause the pull width to be greater than 1.0 as seen in figure 9.2.

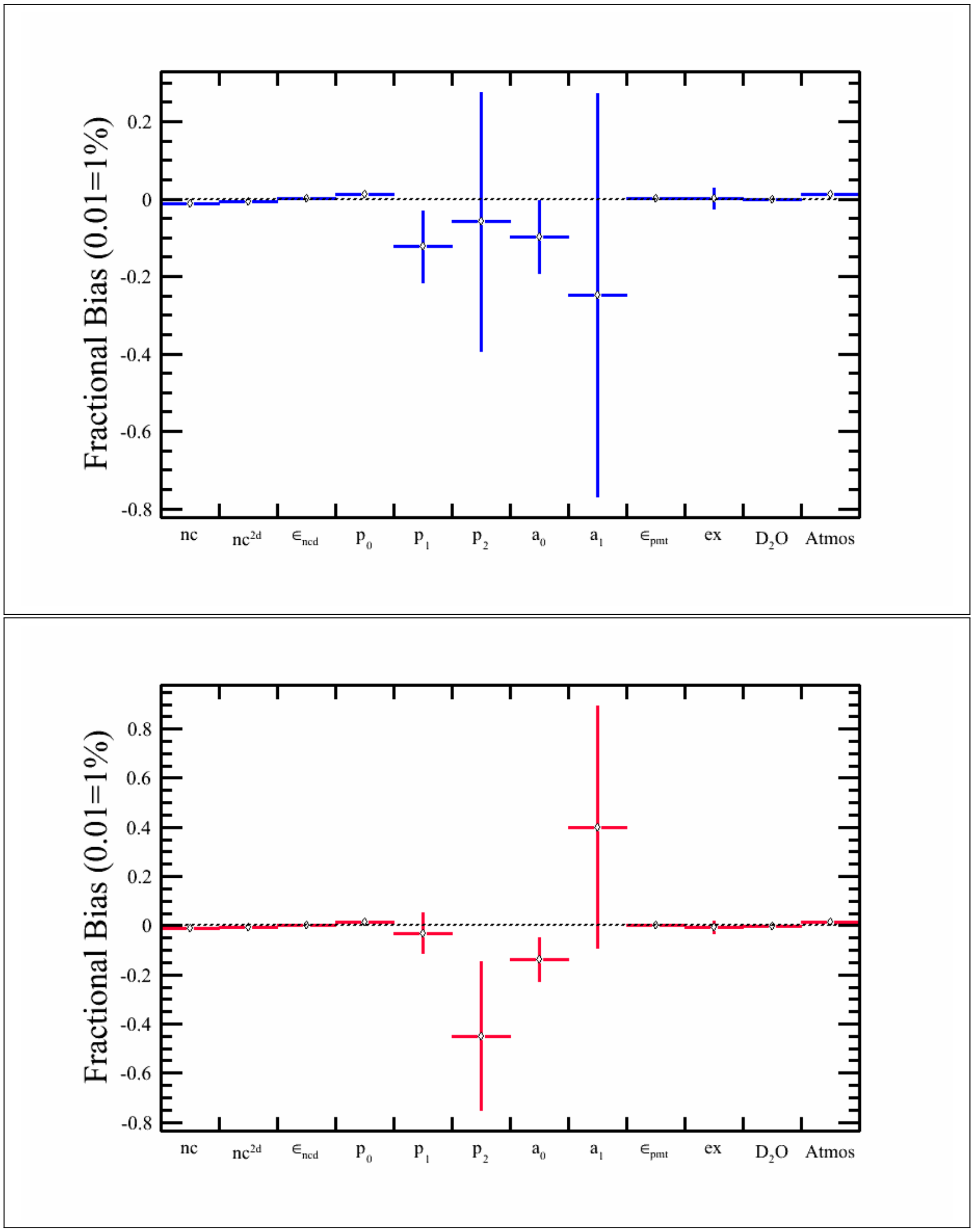


Figure 9.4: These are bias plots for the regular data in blue as well as for the alternate data in red. The role of statistics is evident in the bias of a_1 which flipped sign from the regular dataset to the alternate dataset.

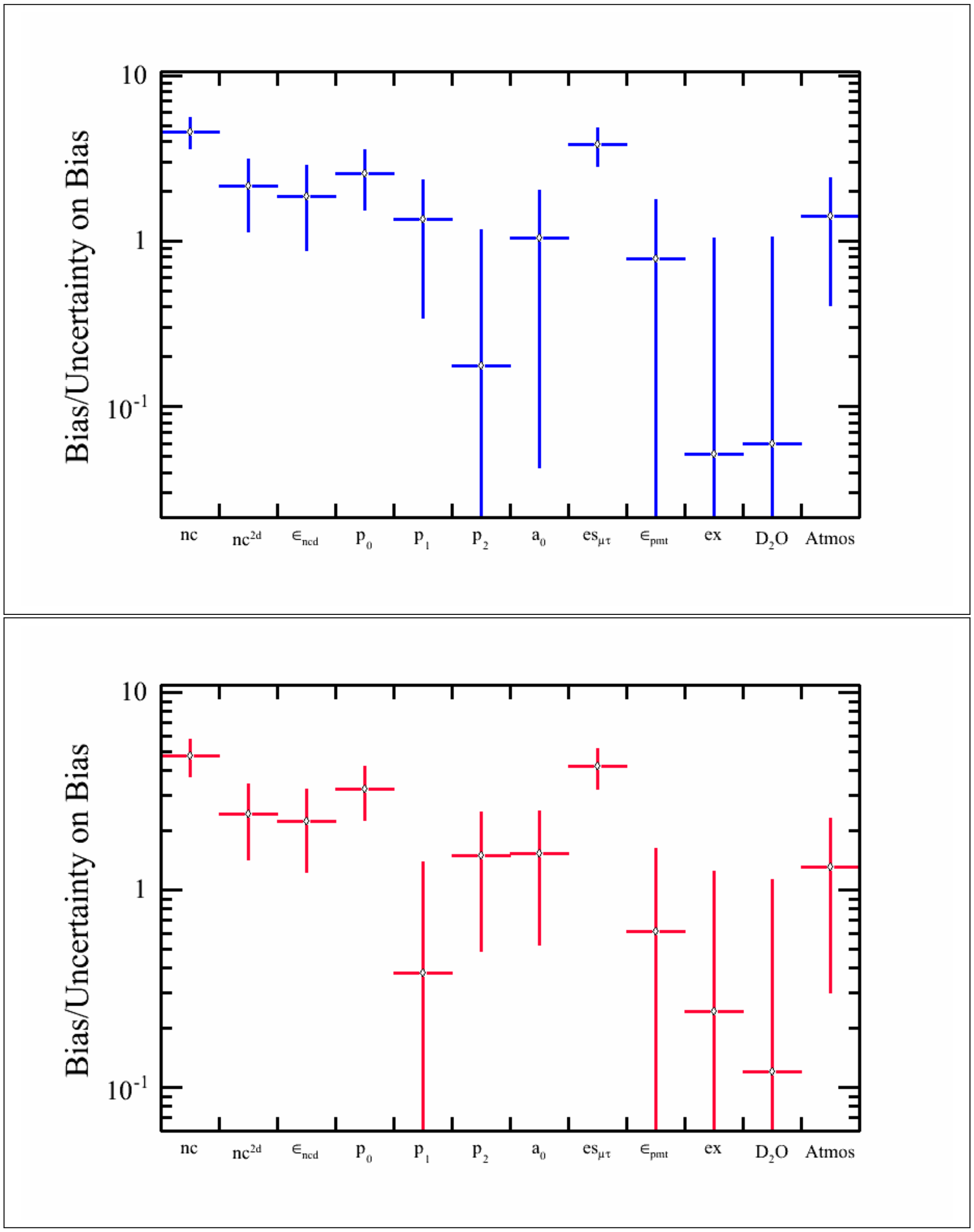


Figure 9.5: Spread of bias divided by the error in the bias for the regular data in blue and for the alternate data in red. The bias of NC and p_0 is not consistent with zero as hoped. The reason is explained in section 11.5.1.

Parameter	**Pull**	**Bias**	**Width of Pull**	**Uncertainty on Bias**
nc^{1d}	-0.238264	-0.0129786	1.00555	0.00272892
nc^{2d}	-0.128707	-0.00660357	1.00183	0.00272106
ϵ_{nc}^{NCD}	0.103482	0.00243735	0.980004	0.0010916
p_0	0.0716084	0.0145938	1.00723	0.0045035
p_1	0.0366964	-0.0308036	0.914152	0.0817548
p_2	-0.0601786	-0.449732	0.895417	0.302359
a_0	-0.0716522	-0.137188	0.962276	0.0901021
a_1	0.0365627	0.400714	0.96157	0.492561
ϵ_{nc}^{PMT}	0.0147325	0.000384841	1.0204	0.000624776
ex	-0.139419	-0.00625062	1.45068	0.025772
ex_{Asym}	-0.0271433		1.00452	
d_2opd	-0.00620543	-0.000870556	0.987645	0.00722651
d_2opd_{Asym}	-0.0781708		0.998703	
Atmos	0.0595101	0.0123081	1.01328	0.00946615

Table 9.6: Pull and bias in a tabulated form, for the alternate dataset consisting of 448 files, to plot figures 9.2, 9.4 and 9.5. The pulls and biases were calculated using the peak and RMS of the posterior distributions. The actual value of day-night asymmetry in the simulated data is zero hence bias calculation for ex_{Asym} and d_2opd_{Asym} is not possible (equation (4.75)).

9.1 Summary

This chapter described the MCMC fit result using the PSA constraint and adding backgrounds from D_2O photo-disintegration and atmospheric neutrinos. The biases are consistent with zero except for the bias in NC. Even though the bias in the neutral current (NC) is not consistent with zero using 1D or 2D marginal likelihood function, **(From 1D distribution -0.0125±0.0027 for the regular dataset and -0.013±0.0027 for the alternate dataset)**, this test was considered a success and we move on to the next ensemble test which included all the backgrounds in the fit. The reason for the bias in NC was investigated in detail and is covered in section 11.5.1.

Chapter 10

Ensemble Test with Signals, all the Backgrounds, and Fixed Systematic Parameters

10.0.1 Introduction

This chapter describes the MCMC fit using the PSA constraint (described in section 4.10.9). The parameters floating in the likelihood function are: neutral current (nc) , external neutrons (ex), day-night asymmetry for the external neutrons, $d_2$opd, day-night asymmetry for the $d_2$opd, atmospheric neutrinos, k2pd, k5pd, ncdpd, NCD NC detection efficiency, PMT NC detection efficiency, P_{ee} parameters (p_0, p_1 and p_2) and the day-night asymmetry (a_0 and a_1). Information from external measurements is used to apply constraints on the likelihood function. There are 11 or 12 constraints applied. Table 10.2 lists the parameters on which constraints are applied, the constraints applied and the widths/uncertainties of the constraints. Table 10.1 shows the Poisson means (Number of expected events, used as mean in the Poisson distribution, and henceforth will be called Poisson means.) for the generation of the simulated datasets and the means from the limited number of datasets. The number of steps in a MCMC fit is 350,0000 and 30,000 steps were removed as a burn-in. The number of simulated datasets in the ensemble test is 14 or 15. Three event classes (hepCC, hepES and hepNC) were not floated and the number of events for three event classes (CC, ES and $ES_{\mu\tau}$) were calculated from the ^{8}B flux and P_{ee} survival probability equation. Refer to chapter 4 for more details.

Parameter	**Poisson Means**	**Mean from Regular dataset**	**Mean from Alternate dataset**
nc	240.569	242.5 ± 11.84	237.5±19.45
ex	20.754	20.73 ± 2.768	21±7
k2pd	9.402	10.13 ± 2.918	8.267±2.977
k5pd	8.378	9.2 ± 2.4	8.267±1.611
$d_2$opd	8.305	7.067 ± 2.38	8.067±2.568
Atmos	24.681	23.13 ± 4.588	23.27±4.49
cc	1845.276	1843 ± 39.95	1836±29.81
es	161.607	162.7 ± 11.5	161.5±10.1
$es_{\mu\tau}$	49.721	50.13 ± 5.427	49.87±4.32
ncdpd	5.938	5.533 ± 1.996	6.667±3.218

Table 10.1: Expected number of events used as Poisson **means** in the generation of the simulated datasets and the mean number of events in the 15 simulated datasets.

Parameter	Constraint	Uncertainty on Constraint
ϵ_{nc}^{PMT}	0.46758	0.00603
ϵ_{nc}^{NCD}	1.764605	0.041815285
ex	1.0	10.453115/20.754
ex_{asym}	0.0	0.0112
k2pd	1.0	1.49056/9.402
k5pd	1.0	0.980968/8.378
$\mathrm{d_2opd}$	1.0	1.28594/8.305
$\mathrm{d_2opd}_{asym}$	0.0	0.112
Atmos	1.0	4.8999/24.681
ncdpd	1.0	2.0349/5.938
PSA	1097.752	49.39884

Table 10.2: Constraints and the uncertainties on the constraints applied on the parameters listed in column 1.

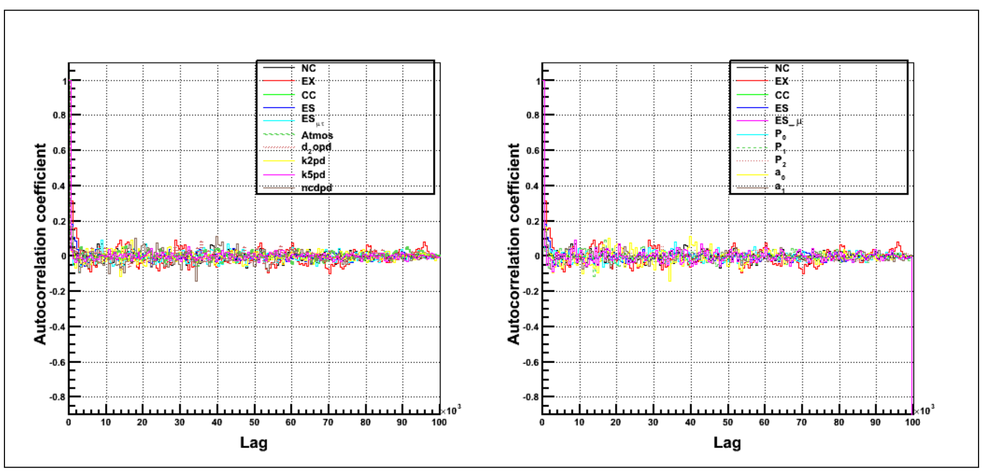


Figure 10.1: Autocorrelation coefficient versus lag for the parameters in the fit. Fifteen simulated datasets with 13 event classes were used in the ensemble test. The number of steps in the MCMC fit is 350,000 but only 100,000 are shown here for clarity. This fit does not float the systematic uncertainties.

10.0.2 Result

In the calculation of pull (equation (4.76)) and bias (equation (4.75)), for the best-fit of a fit parameter and its uncertainty, peak and RMS of the posterior distribution – belonging to the fit parameter – were used respectively. The peak is determined by the following algorithm:

1. Determine the RMS σ and the range (R) of the posterior distribution by projecting all the MCMC events along the single axis of the parameter.

2. 10^{th} of the σ is taken as a bin width of a histogram H and the number of bins are calculated as nbins = $(\mathrm{R}/(0.1\,\sigma))$.

3. The histogram H is populated with the parameter values from the MCMC chain.

4. Smoothing the histogram H and then searching for the bin with the highest value (maximum bin).

5. Peak is a bin center of the maximum bin.

For the regular data, the result is presented in a tabular form (table 10.3) as well as blue plots in figures 10.2 to 10.4; for the alternate datasets, the result in a tabular form is listed in a table 10.4 and pull and bias plots are shown in red in Figures 10.2 to 10.4.

Simulated datasets are created from a Monte Carlo simulation. Since random events are selected for inclusion in the simulated dataset, a different seed will result in different datasets. Hence statistics will play a role in the pull and bias spread. Increasing the number of datasets will reduce the role of statistics but with only 15 datasets, it is evident that statistics play a vital role. For example, looking at tables 10.3 and 10.4, the pull of k5pd is 0.602 for the alternate data but 0.115 for the regular data. Similarly the pull of p_2 is $+0.655$

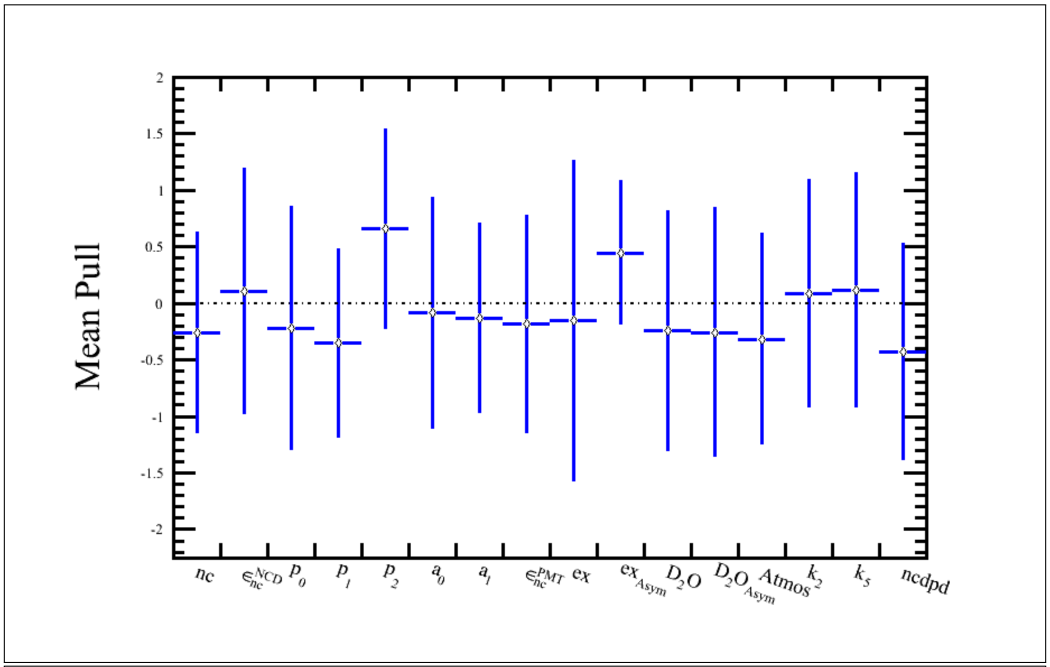
for the regular data but only -0.218 for the alternate data. The sign of pull flipped from the regular data to the alternate data for p_2. The pull of some of the parameters exceeded 0.5, for example p_2 and k5pd in the regular dataset and alternate dataset respectively, as shown in Figure 10.2. Section 10.2 covers a discussion of the statistics of pull in an ensemble consisting of 14 or 15 datasets. The error bars on the pull plots indicate the average spread of the parameter and not the uncertainty on the average pull. From statistics, the pull width as a function of the number of datasets **n**, is given as:

$$\sqrt{[(n-1)/n]}\left(1-\frac{1}{4(n-1)}\right). \tag{10.1}$$

Since n=14 or 15 for this analysis, the pull width should be 0.945 or 0.949.

Parameter	**Pull**	**Bias**	**Width of Pull**	**Uncertainty on Bias**
nc	-0.261713	-0.0155333	0.89048	0.0134721
ϵ_{nc}^{NCD}	0.106011	0.0003	1.08945	0.000792184
p_0	-0.222016	-0.0113	1.0788	0.0262884
p_1	-0.354	0.758667	0.834691	0.404804
p_2	0.655335	4.614	0.884342	1.65851
a_0	-0.0860025	-0.139411	1.02759	0.520458
a_1	-0.132667	-1.49877	0.842946	2.20641
ϵ_{nc}^{PMT}	-0.184667	-0.00230667	0.965865	0.00322137
ex	-0.157145	-0.0436701	1.42237	0.139222
ex_{Asym}	0.445728		0.636905	
$d_2$opd	-0.246017	-0.0378333	1.06408	0.0420035
d_2opd_{Asym}	-0.258038		1.10459	
Atmos	-0.319344	-0.0621667	0.937625	0.0475242
k2pd	0.0846684	0.0121667	1.01403	0.041383
k5pd	0.115336	0.0141667	1.03623	0.0312613
ncdpd	-0.431336	-0.117801	0.963052	0.0746731

Table 10.3: Pull and bias in tabulated form, used to plot distributions, shown in blue, in figures 10.2 to 10.4. The day-night asymmetry of the $d_2$opd and external neutrons in the regular datasets is zero, hence the bias is not applicable (equation (4.75)).



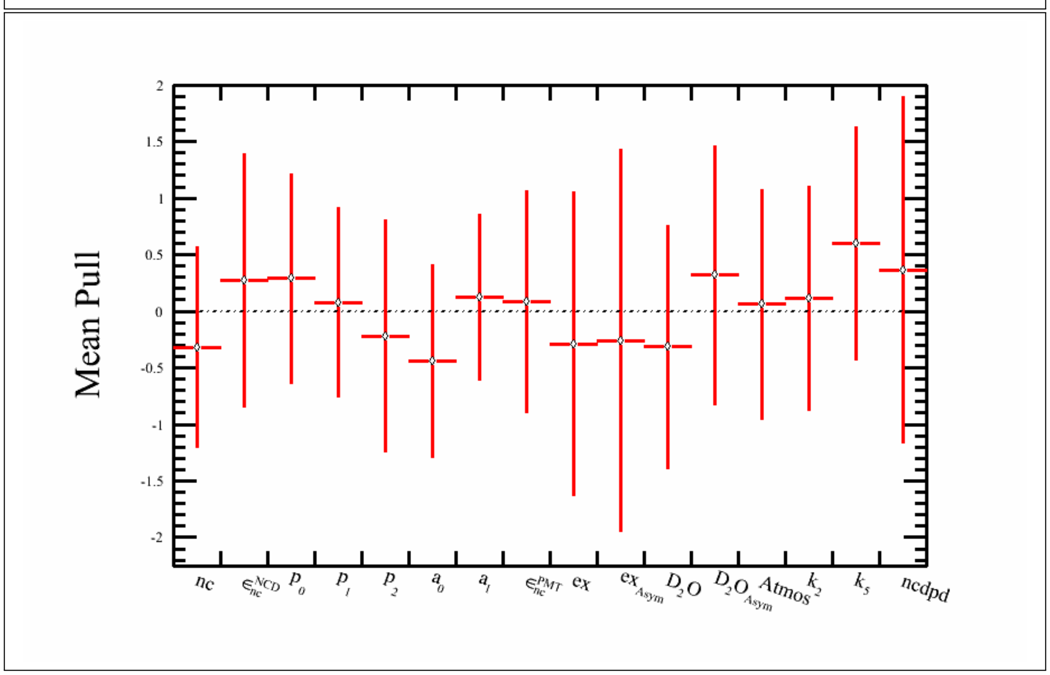


Figure 10.2: Pull spread for the 15 datasets. The peak and RMS of the posterior distribution were used as the best-fit and its uncertainty in the calculation of the pull of the fit. The top plot is the pull spread for the regular datasets and the bottom plot is for the alternate datasets. There are a number of sign flips between the data and the alternate data, for instance, p_2.

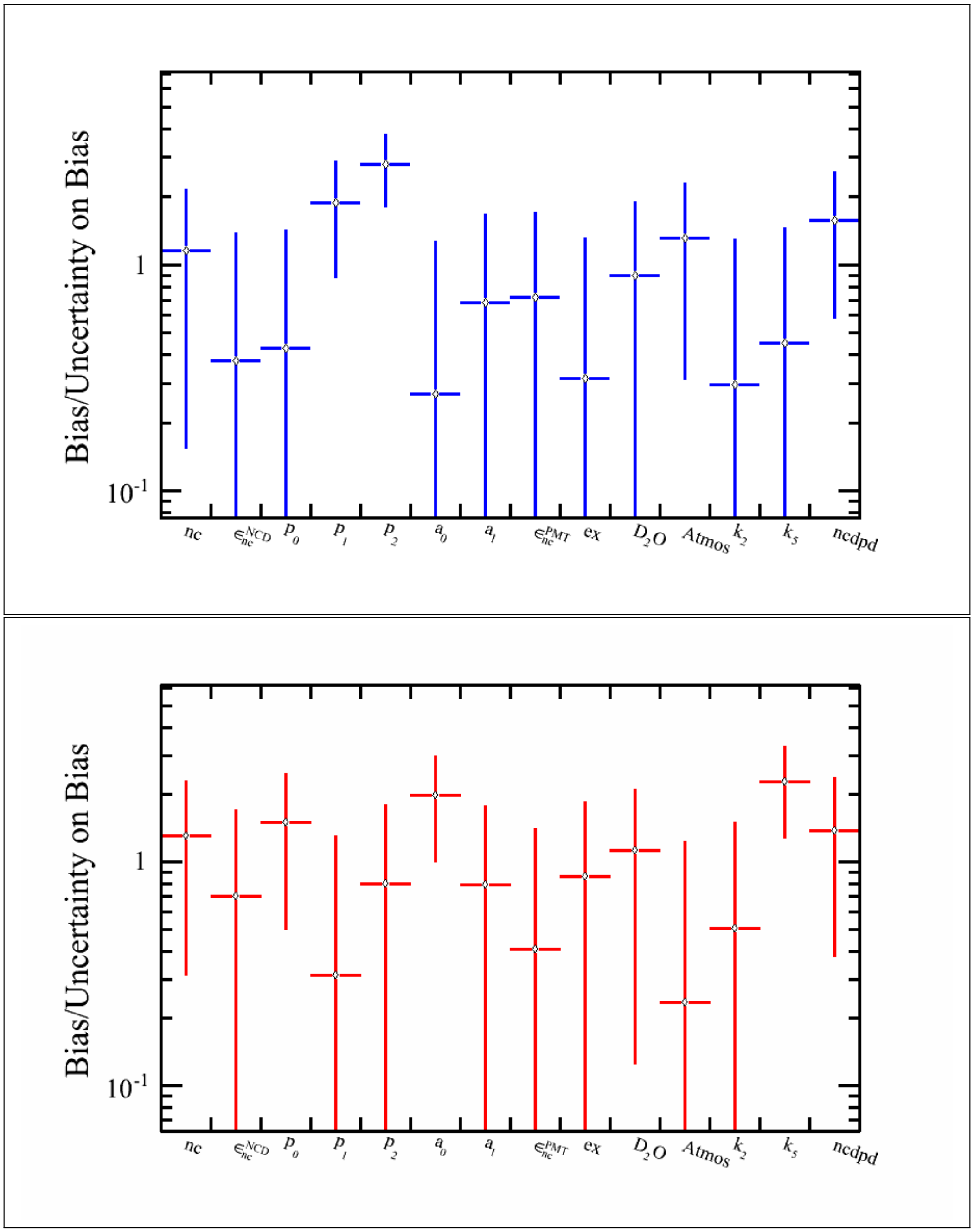


Figure 10.3: Spread of bias divided by the error in the bias. This ensemble test floats 16 parameters and does not include the systematics. The blue is for the regular datasets and the red is for the alternate datasets. The bias of p_2 is less than 1σ for the alternate datasets but more than 2σ for the data. The bias on a_0 is less than 1σ for the data but around 2σ for the alternate case.

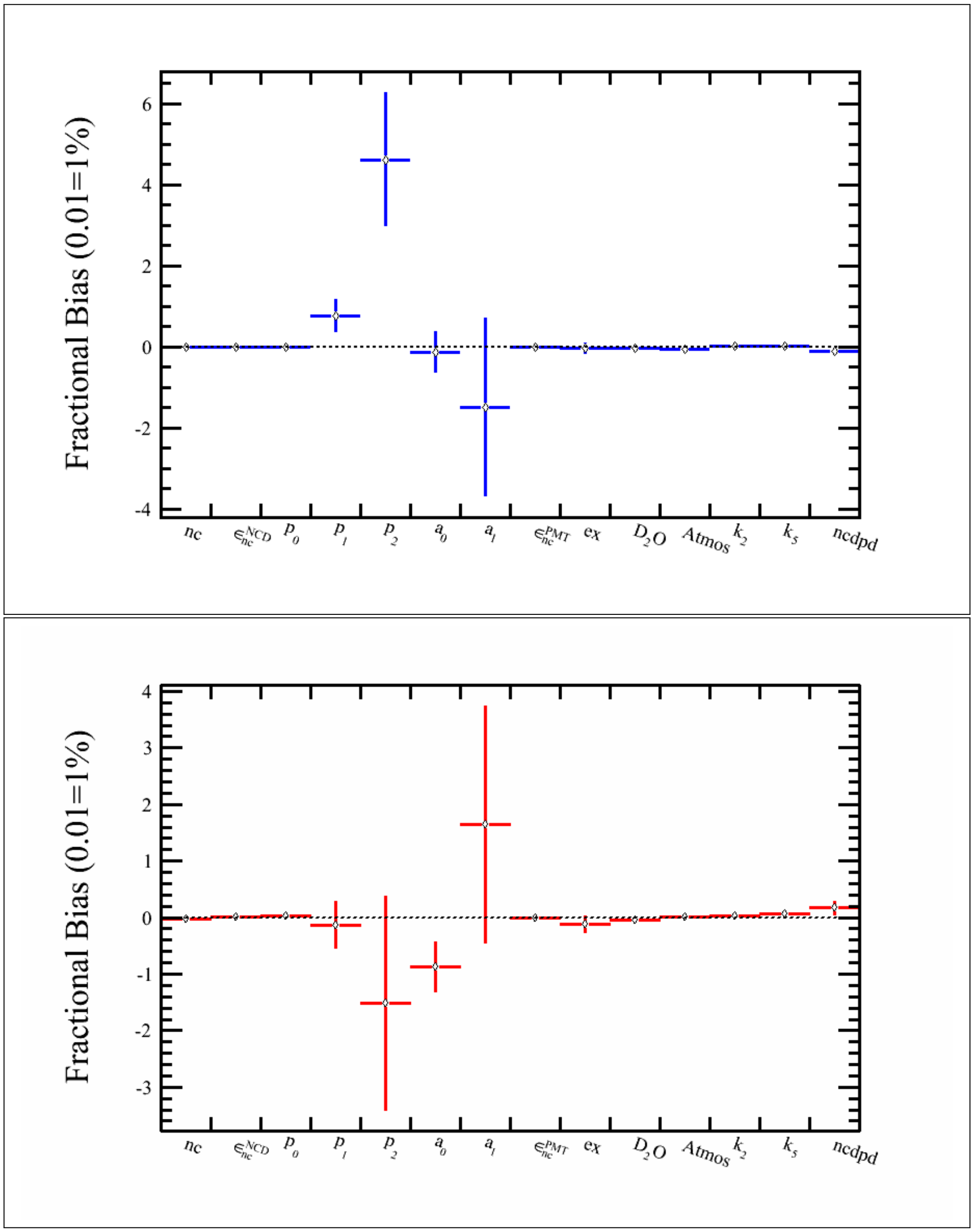


Figure 10.4: Spread of bias for the 15 regular datasets (blue) containing 13 event classes. The red is for the alternate datasets. There are a number of sign flips, for instance, p_2 and a_1.

Parameter	**Pull**	**Bias**	**Width of Pull**	**Uncertainty on Bias**
nc	-0.316215	-0.0184667	0.882103	0.0140898
ϵ_{nc}^{NCD}	0.272715	0.00207378	1.122	0.00293686
p_0	0.291334	0.0373	0.925968	0.0249086
p_1	0.078	-0.129333	0.837835	0.413784
p_2	-0.218004	-1.518	1.02159	1.89429
a_0	-0.444667	-0.87936	0.851902	0.440405
a_1	0.123333	1.64804	0.729518	2.09174
ϵ_{nc}^{PMT}	0.0860005	0.00134667	0.984209	0.00330405
ex	-0.289287	-0.123689	1.34256	0.144165
ex_{Asym}	-0.259242		1.68998	
$d_2$opd	-0.312695	-0.0485	1.07422	0.0430973
d_2opd_{Asym}	0.318098		1.14534	
Atmos	0.0620008	0.0121667	1.01325	0.0517788
k2pd	0.118001	0.0201667	0.991889	0.040105
k5pd	0.602032	0.0711667	1.03028	0.0312321
ncdpd	0.367007	0.16725	1.5316	0.121459

Table 10.4: Pull and bias in tabulated form, used to plot the distributions, shown in red, in figures 10.2 to 10.4. The day-night asymmetry of the $d_2$opd and external neutrons in the alternative datasets is zero, hence the bias is not applicable (equation (4.75)).

10.1 Including penalty from both the Low Energy Threshold Analysis (LETA) and the Pulse Shape Analysis (PSA)

For this fit, the constraints from LETA were added along with the constraint from the PSA in the likelihood function. PSA and LETA constraints are described in detail in Sections 4.10.9 and 4.10.8.

Fourteen datasets were used for this ensemble test. This fit also includes all 6 backgrounds along with CC, ES, NC and $ES_{\mu\tau}$. The parameters constrained by the LETA constraint are ^{8}B Scale and the 5 parameters of the survival probability equation (p_0, p_1, p_2, a_0 and a_1).

Parameter	Pull	Bias	Width of Pull	Uncertainty on Bias
nc	-0.33149	-0.00948571	0.883355	0.00665161
ϵ_{nc}^{NCD}	-0.317298	-0.00312461	1.17893	0.00323582
p_0	-0.18	-0.00771429	0.863663	0.0111242
p_1	-0.16	0.132571	0.692048	0.147339
p_2	0.509154	1.19314	1.20124	0.737018
a_0	0.202465	0.224572	0.981601	0.280461
a_1	-0.210286	-0.96	0.712436	0.948475
ϵ^{PMT}	-0.137993	-0.00146743	1.14366	0.00367554
ex	-0.0171429	0.012	0.865261	0.0965634
ex_{Asym}	-0.212742		0.974553	
d_2o	-0.406556	-0.0630932	1.02568	0.0420775
d_2o_{Asym}	-0.325714		1.00941	
Atmos	-0.362743	-0.0708	0.683359	0.0358792
k2pd	-0.0854673	-0.0127308	1.20072	0.0507381
k5pd	-0.373504	-0.0440585	1.04368	0.0325189
ncdpd	0.0377143	0.0302857	1.0265	0.0858535

Table 10.5: Pull and bias in tabular form, used to plot the distributions, shown in blue, in figures 10.5 to 10.7. The MCMC fit includes constraint from both LETA and PSA. The day-night asymmetry of the d_2opd and external neutrons in the regular datasets is zero, hence the bias is not applicable (equation (4.75)).

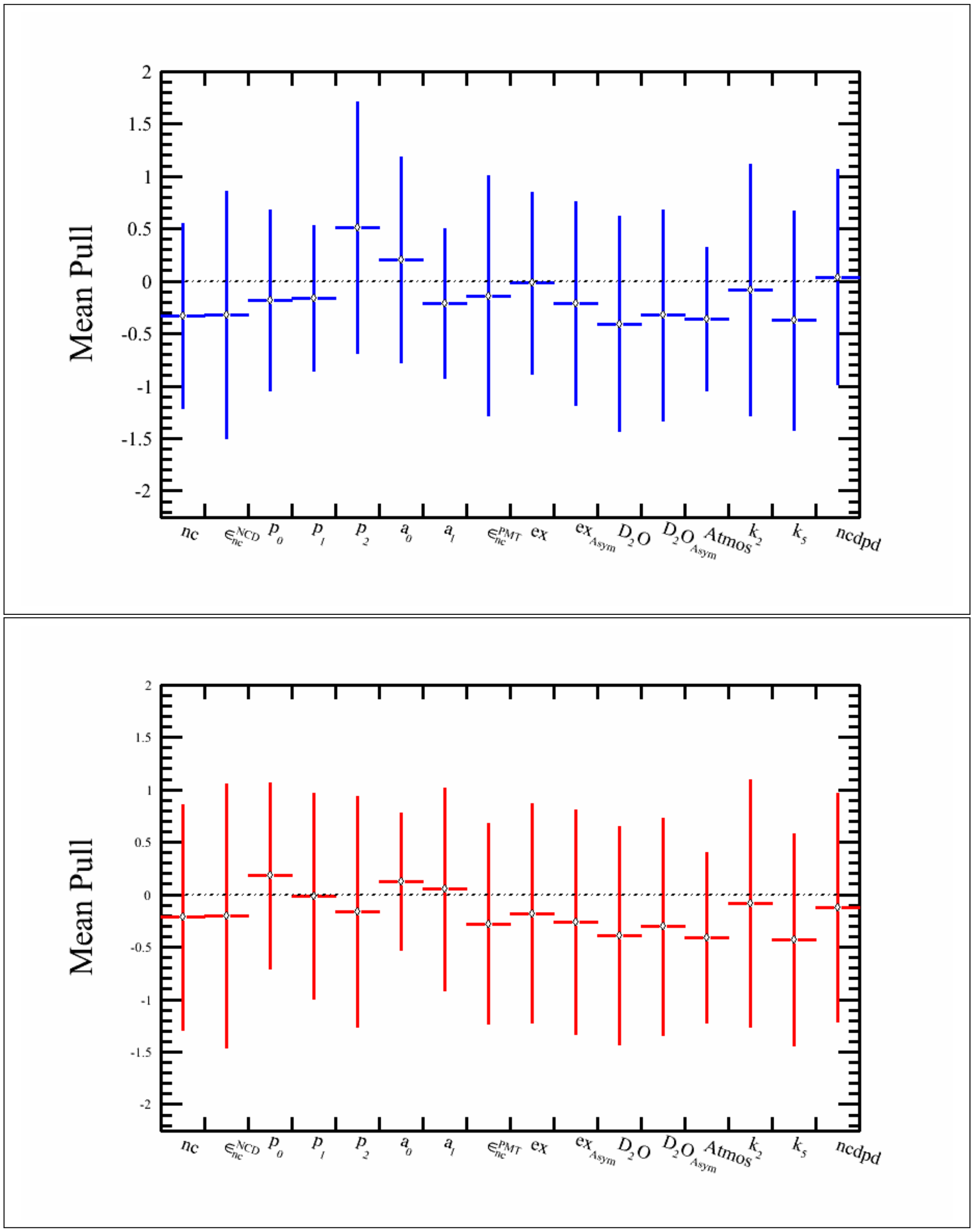


Figure 10.5: Pull spread with both PSA and LETA constraints. Ensemble test included fourteen datasets. Each MCMC run had 350,000 steps; 100,000 steps were removed as burn-in to assure the convergence of the remaining steps. Systematics were not floated in this ensemble test. The fit included 13 event classes and floated 16 parameters with the application of 12 constraints. The blue is for the data and red is for the alternate case. The parameters p_0, p_1, p_2 flipped signs between the regular data and the alternate data.

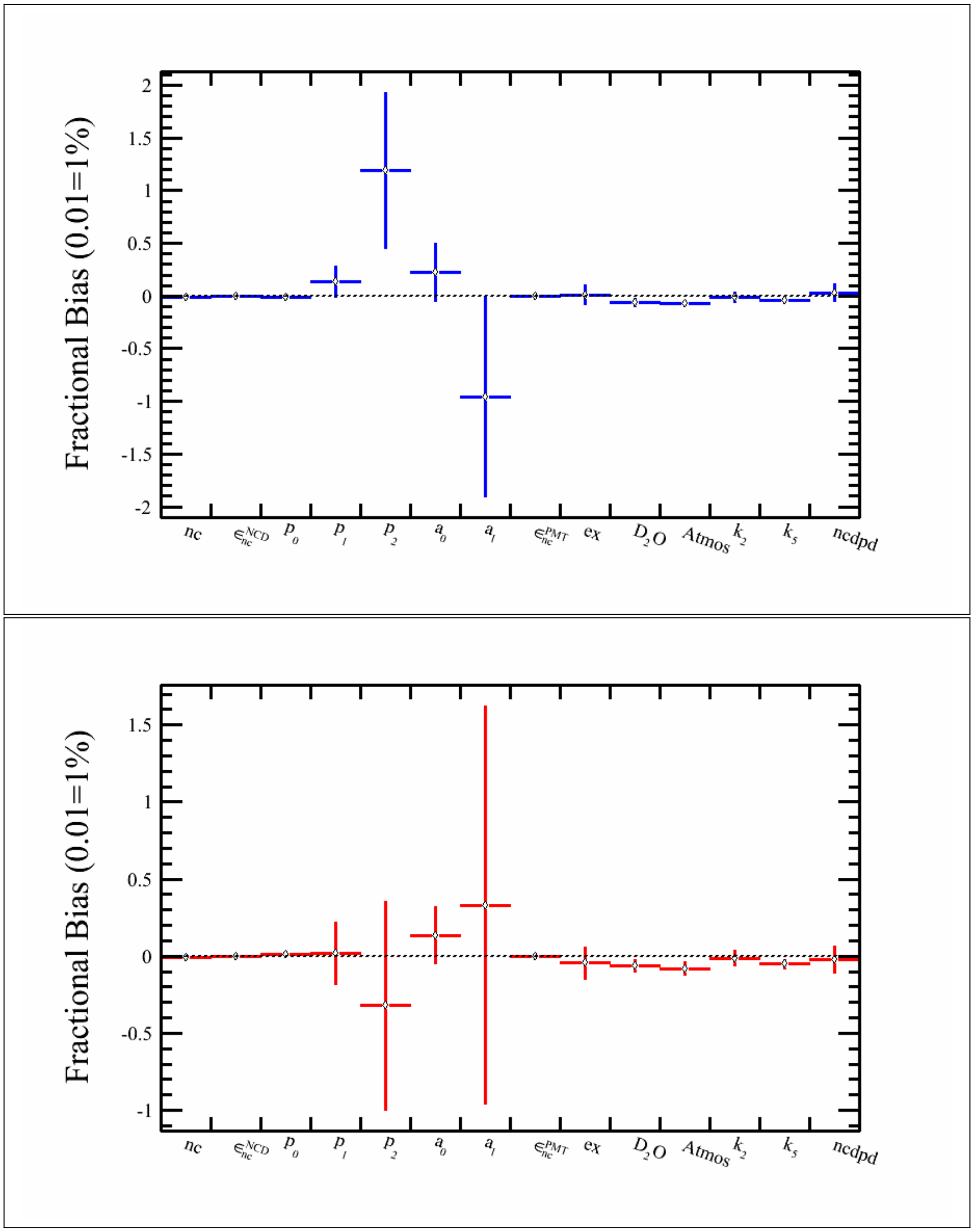


Figure 10.6: Spread of bias with both PSA and LETA constraints. Ensemble test included fourteen datasets. Each MCMC run had 350,000 steps; 100,000 steps were removed as burn-in to assure the convergence of the remaining steps. Systematics were not floated in this ensemble test. The fit included 13 event classes and floated 16 parameters with the application of 12 constraints. The blue is for the data and red is for the alternate case. The parameter p_2 flipped the sign between the regular dataset and the alternate dataset.

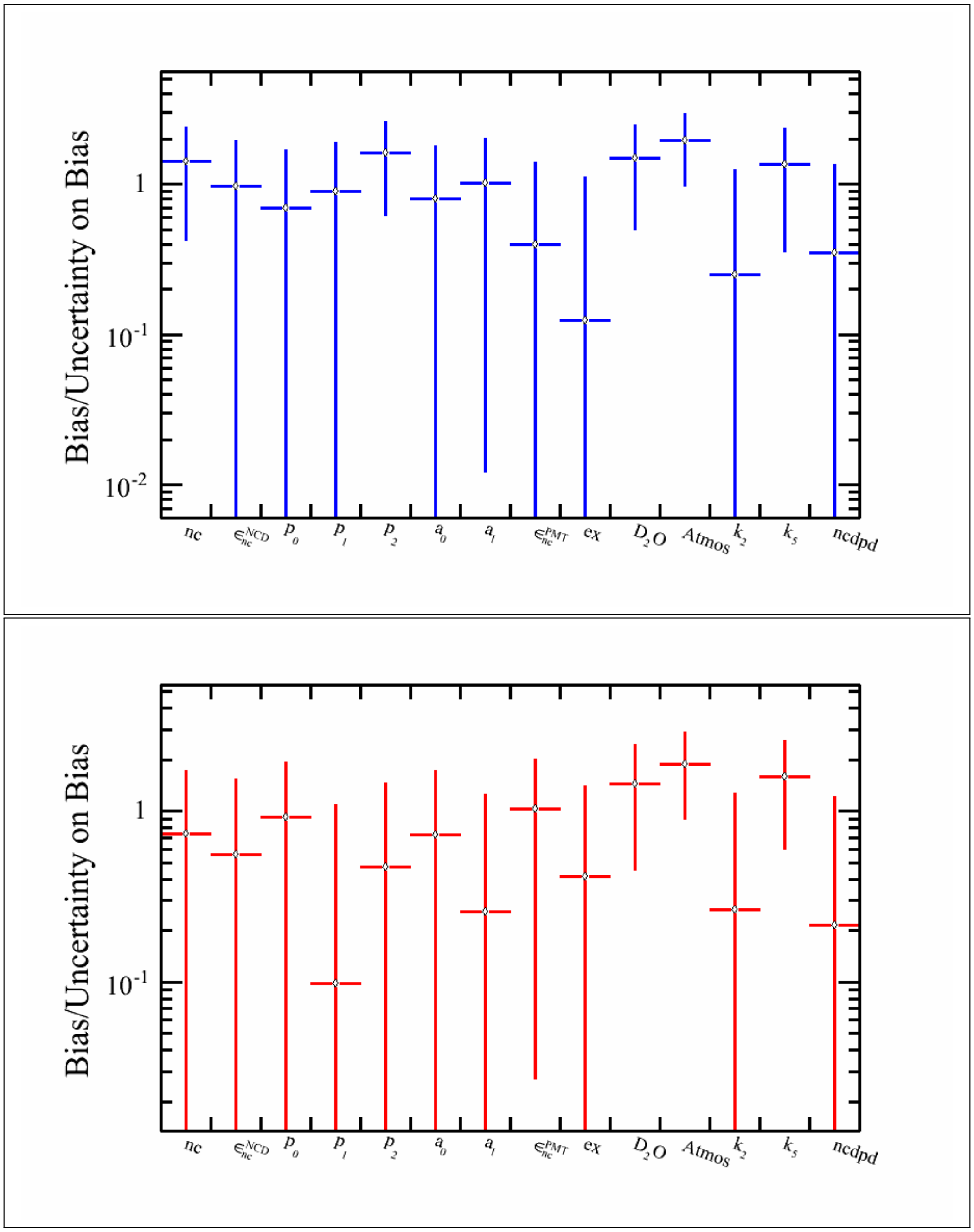


Figure 10.7: Spread of bias divided by the error in the Bias using both PSA and LETA constraints. Ensemble test included fourteen datasets. Each MCMC run had 350,000 steps; 100,000 steps were removed as burn-in to assure the convergence of the remaining steps. Systematics were not floated in this ensemble test. The fit included 13 event classes and floated 16 parameters with the application of 12 constraints. The blue is for the data and red is for the alternate case.

Parameter	**Pull**	**Bias**	**Width of Pull**	**Uncertainty on Bias**
nc	-0.216109	-0.006	1.07102	0.0081885
ϵ_{nc}^{NCD}	-0.203118	-0.00191767	1.2591	0.00345443
p_0	0.182857	0.0107143	0.885908	0.0115773
p_1	-0.0100001	0.02	0.980966	0.202757
p_2	-0.162878	-0.321429	1.09952	0.679468
a_0	0.125714	0.135	0.653438	0.185701
a_1	0.0528574	0.331429	0.965638	1.29104
ϵ_{nc}^{PMT}	-0.277144	-0.00305714	0.955609	0.00297688
ex	-0.178571	-0.0428571	1.04565	0.103571
ex_{Asym}	-0.261812		1.06897	
$d_2$opd	-0.391452	-0.0625	1.0424	0.0430739
d_2opd_{Asym}	-0.305726		1.03947	
Atmos	-0.408571	-0.08	0.811228	0.042172
k2pd	-0.0814626	-0.0132143	1.17777	0.0495151
k5pd	-0.430012	-0.0503571	1.01149	0.031553
ncdpd	-0.12287	-0.0188592	1.0902	0.0877943

Table 10.6: Pull and bias in tabular form, used to plot the distributions, shown in red, in figures 10.5 to 10.7. The MCMC fit includes constraint from both LETA and PSA. The day-night asymmetry of the $d_2$opd and external neutrons in the alternative datasets is zero, hence the bias is not applicable (equation (4.75)).

10.2 The Statistics of "Pulls"

Using a small toy Monte Carlo, it was shown that an ensemble test of 15 datasets will always result in broader pull distributions compared to say an ensemble of 50 or 100 datasets. Pull plots for the 15, 50 and 100 datasets are shown in figures 10.8 and table 10.7 lists pull, width of pull, χ^2 and degrees of freedom (**dof**) from fitting a Gaussian function to the pull spreads. The pull is 39.21 when using 15 files but improves to 0.399 and 0.133 with 50 and 100 files respectively. Non-zero number of entry at pull=-5 plays a bigger role in pull calculation when using 15 files compared to, for instance, 100 files. If the range of the fit is restricted from -2 to 1 in the top plot in Figure 10.8, the pull improves from 39 to 0.7 to with $\chi^2/dof = 6/17$.

Number of datasets	**Pull**	**Width of Pull**	χ^2	**Degrees of Freedom**
15	39	8	7	38
50	0.399	1.382	30	57
100	0.133	1.288	53	58

Table 10.7: Result of a toy Monte Carlo where different number of datasets were used in the analysis to quantify the effect of statistics on the pull of the fit.

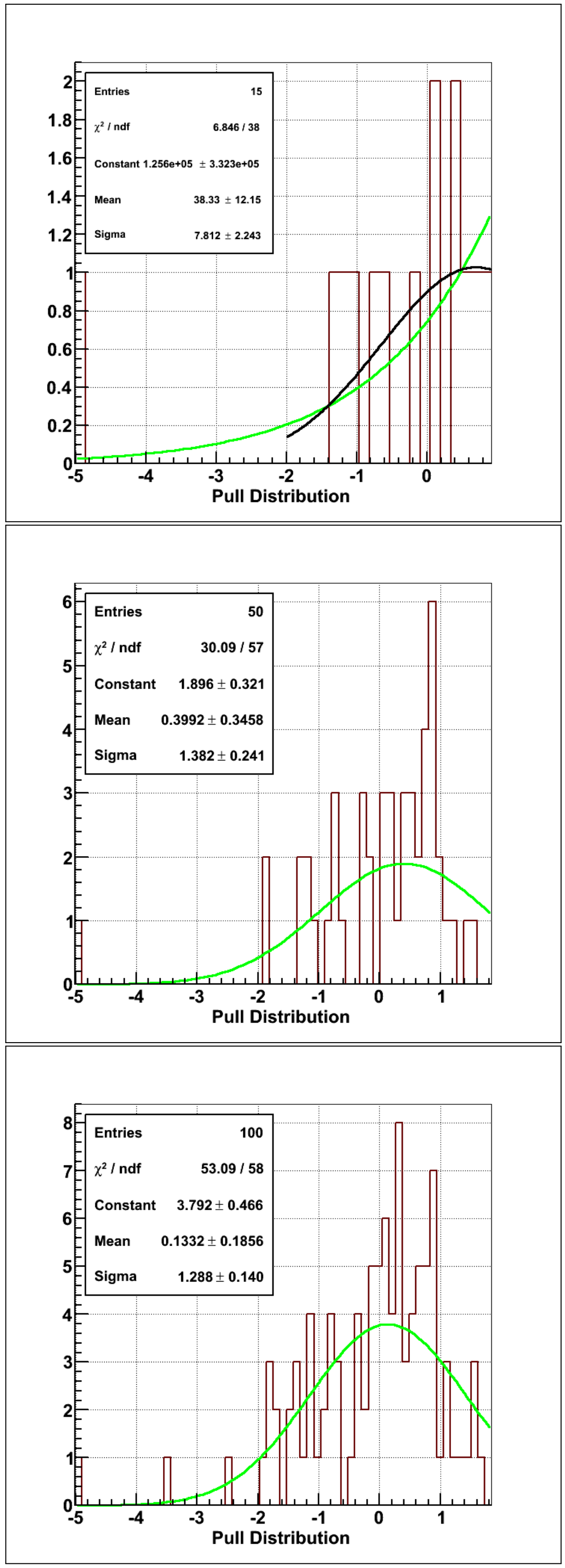


Figure 10.8: Pull spreads for 15, 50 and 100 sets. The green line shows a Gaussian fit of the distributions.

10.3 Convergence of the Markov Chain

One test of robustness of the MCMC code is the **convergence** fit where the starting points were picked from a flat distribution expressed as: Uniform$(\mu - 10\sigma, \mu + 10\sigma)$ where μ is the central value and σ is the uncertainty of the central value. MCMC is a robust code if the fit converged despite the fact that the initial values were far away ($\pm 10\sigma$) from the nominal values. This section describes the result of the **convergence** test. Plot 10.9 shows that the convergence of log likelihood was achieved around 4000 steps. Different parameters took different number of steps to converge, as seen in figures 10.11 to 10.15. To ensure the convergence, the burn-in period was selected to be 50,000 steps after consulting the autocorrelations plots, shown in figure 10.1. The number of steps in the burn-in period were rejected when estimating the parameter values. Table 10.8 gives a summary of the results of the **Convergence** fit; column one lists the constraints selected randomly for the dataset, column two lists the mean and RMS of the posterior distributions, after taking out the burn-in period, and the last column points to the figures that has the corresponding result.

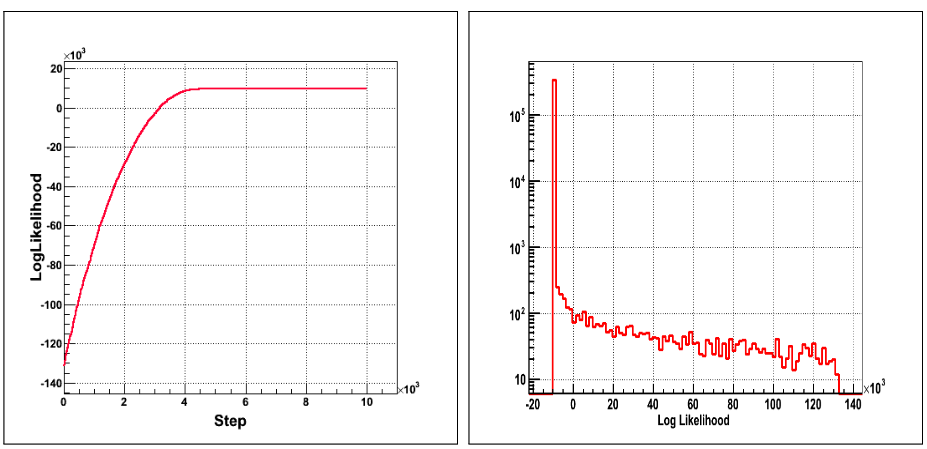


Figure 10.9: The left plot displays a histogram of log likelihood versus MCMC step. The plot on right is a histogram which shows the negative value of likelihood calculated for each step of the chain. The large negative likelihoods correspond to the steps where the chain has not yet converged.

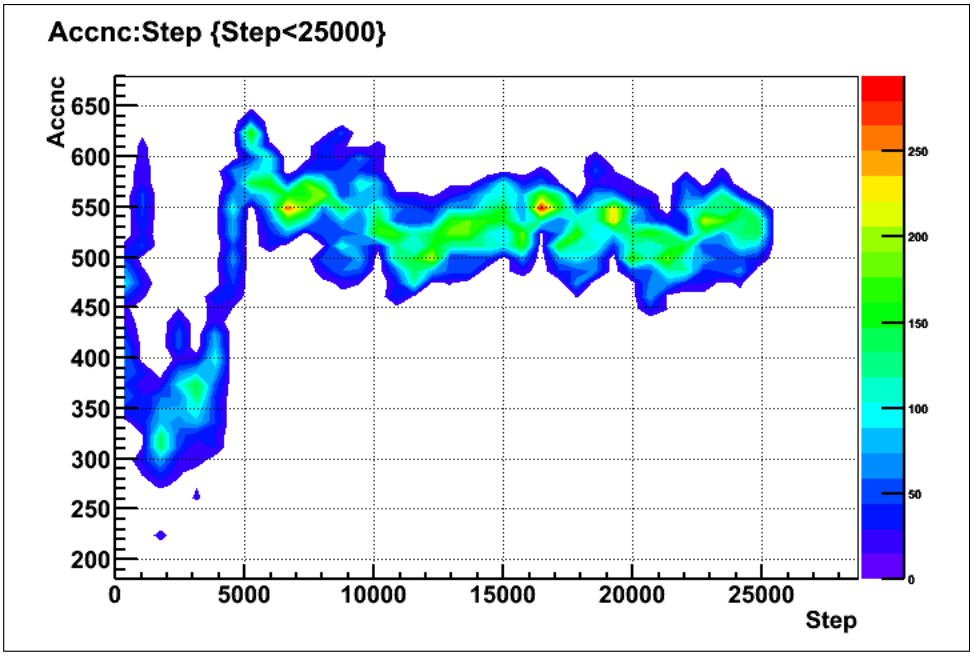


Figure 10.10: Convergence of neutral current (NC) flux. The yellow and red colours show the converged regions. For clarity not all MCMC steps (350,000) are shown in this plot.

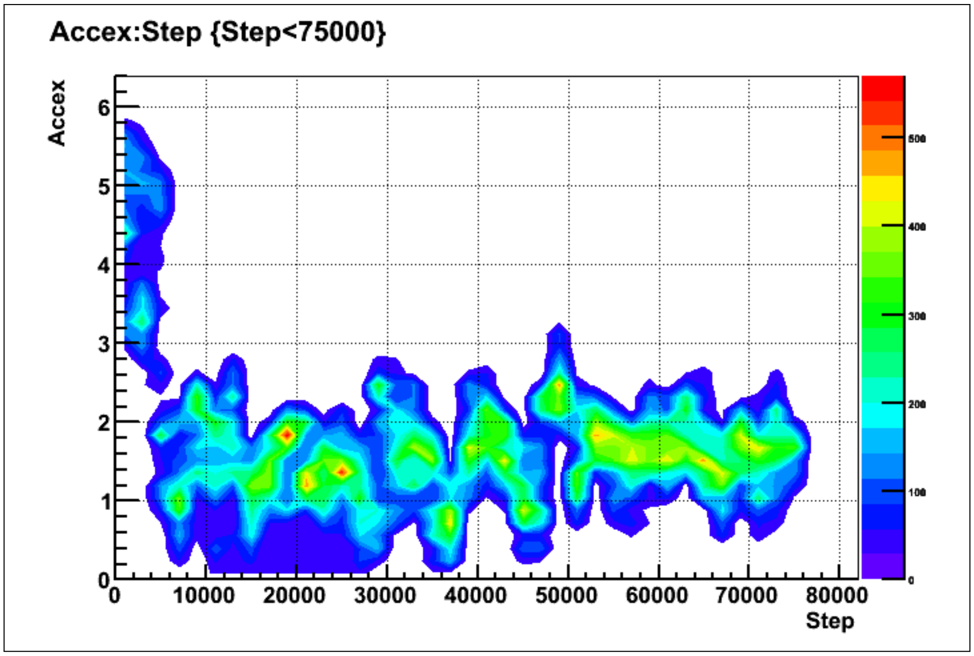


Figure 10.11: Convergence of the external neutrons. The yellow and red colours show the converged regions. For clarity not all MCMC steps (350,000) are shown in this plot.

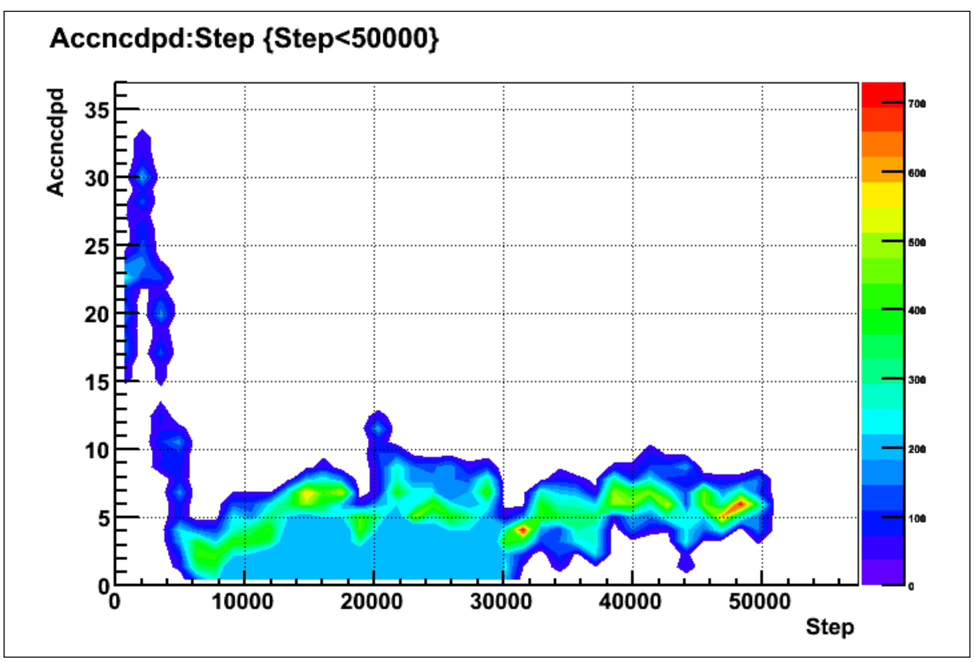


Figure 10.12: Convergence of the ncdpd background. The yellow and red colours show the converged regions. For clarity not all MCMC steps (350,000) are shown in this plot.

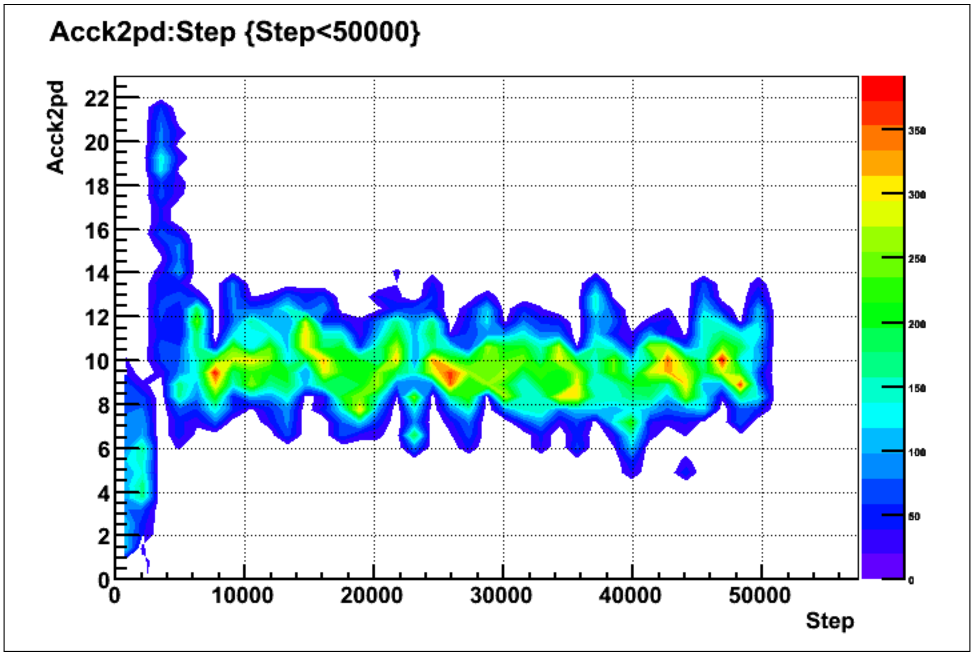


Figure 10.13: Convergence of the k2pd background. The yellow and red colours show the converged regions. For clarity not all MCMC steps (350,000) are shown in this plot.

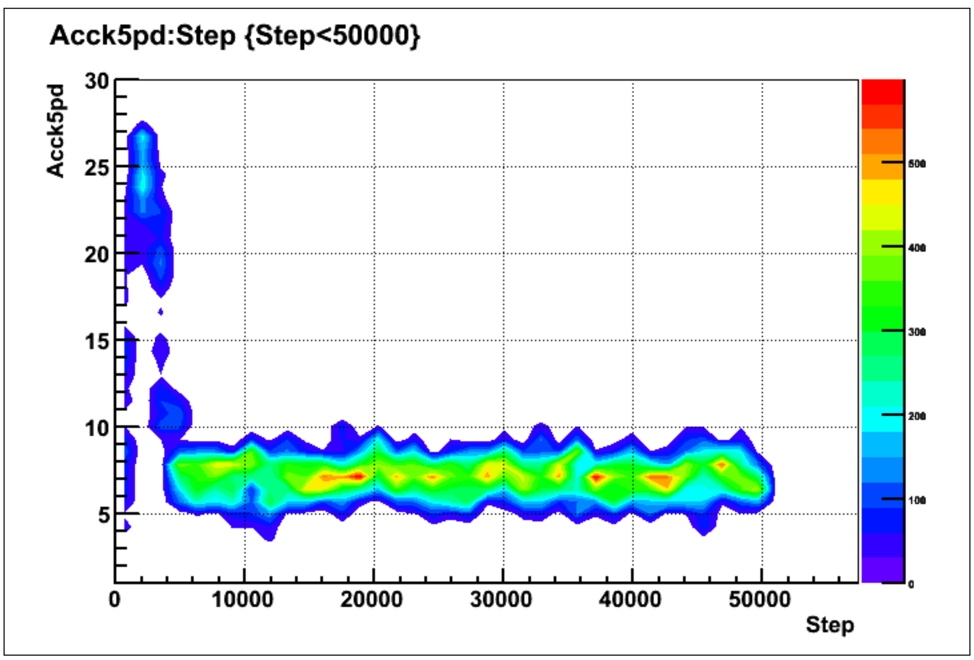


Figure 10.14: Convergence of the k5pd background. The yellow and red colours show the converged regions. For clarity not all MCMC steps (350,000) are shown in this plot.

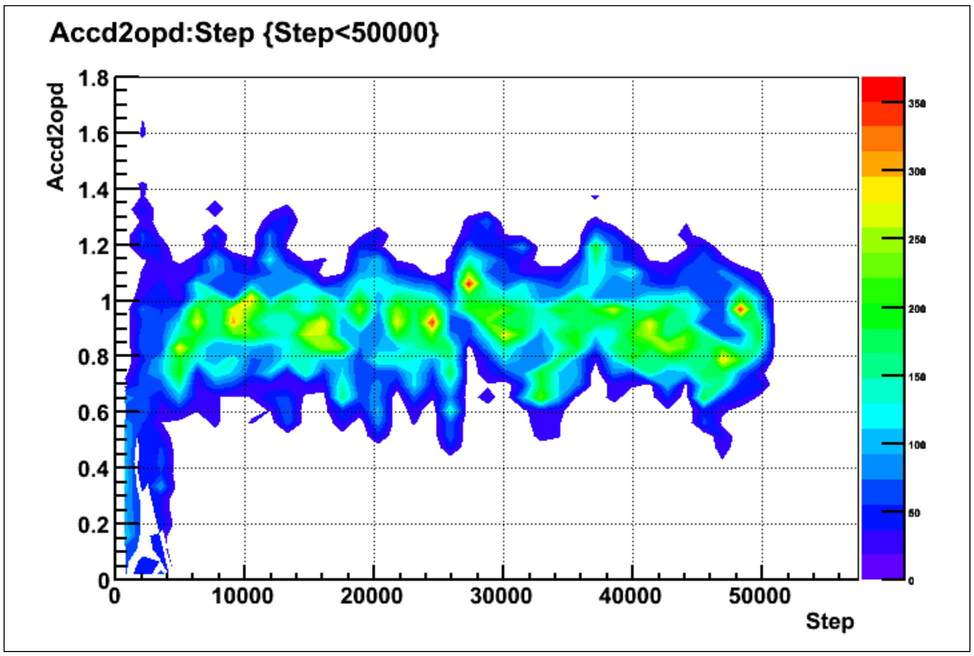


Figure 10.15: Convergence of $d_2$opd. The yellow and red colours show the converged regions. For clarity not all MCMC steps (350,000) are shown in this plot.

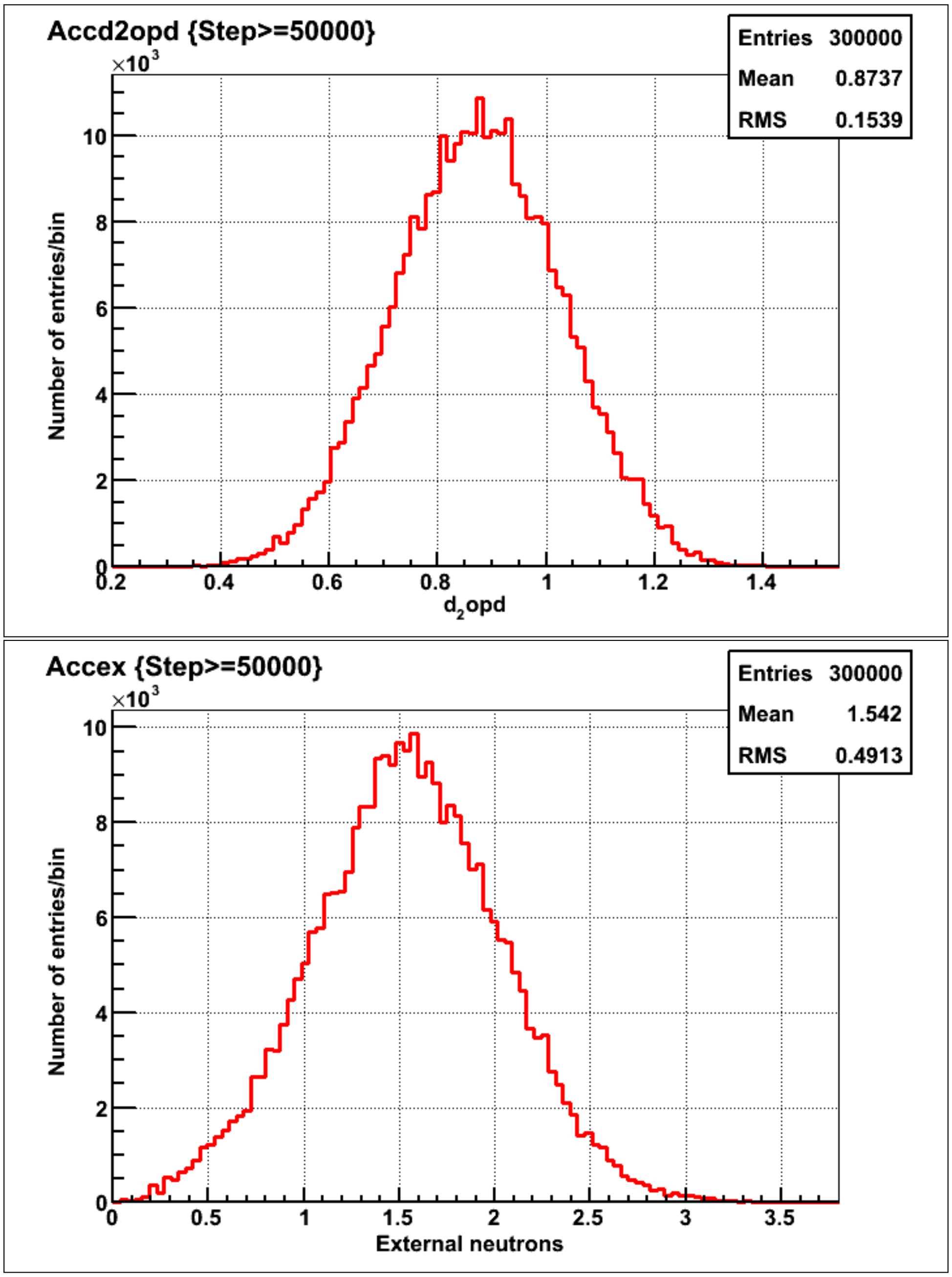


Figure 10.16: Posterior distributions of the $d_2$opd and external neutrons. The values are ratios of average rates to the nominal rates for the $d_2$opd and external neutrons. The fit value is taken as the mean of distribution and the uncertainty as the RMS.

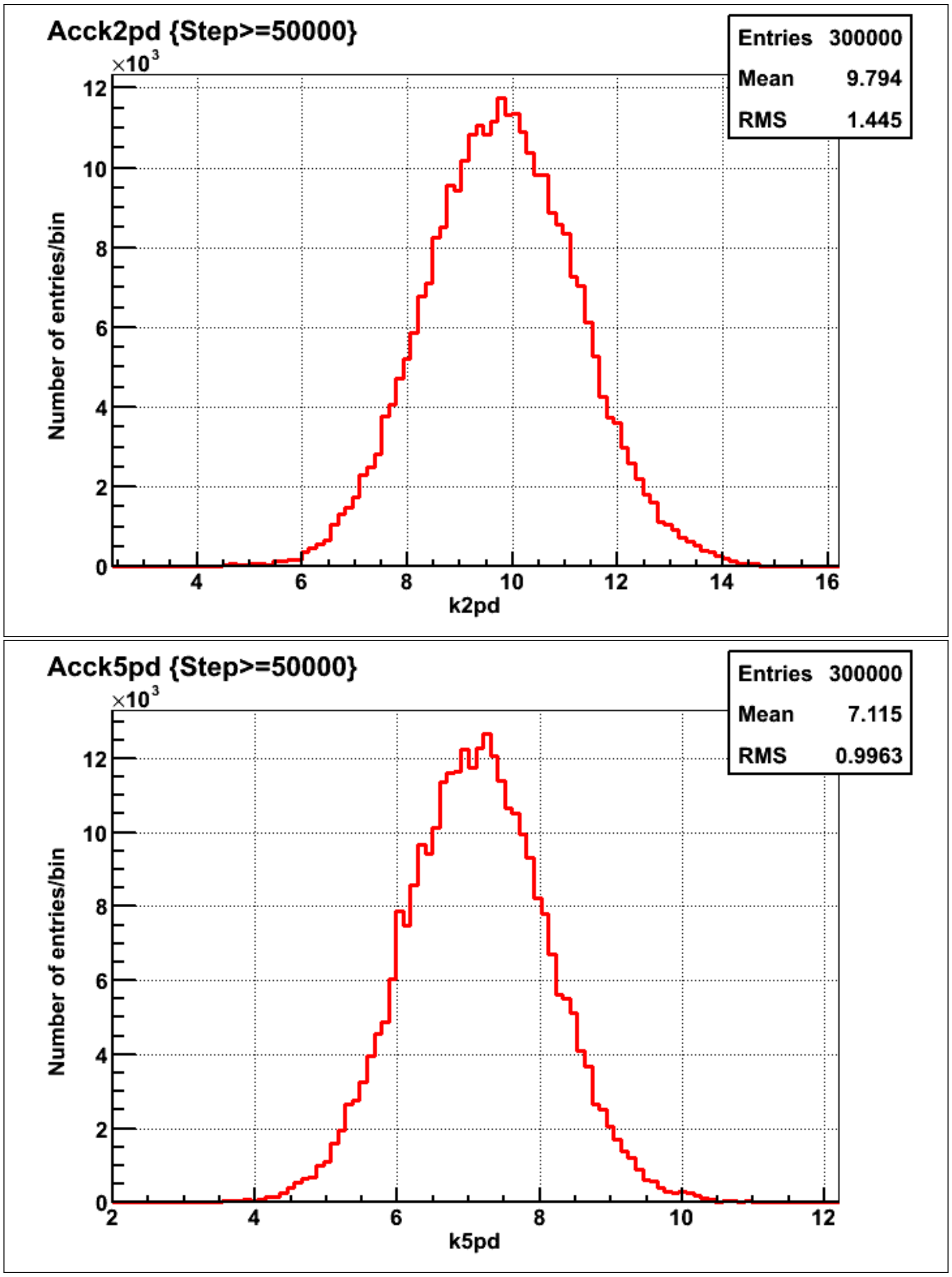


Figure 10.17: Posterior distributions showing the mean number of events of k2pd [top] and k5pd [bottom] neutrons.

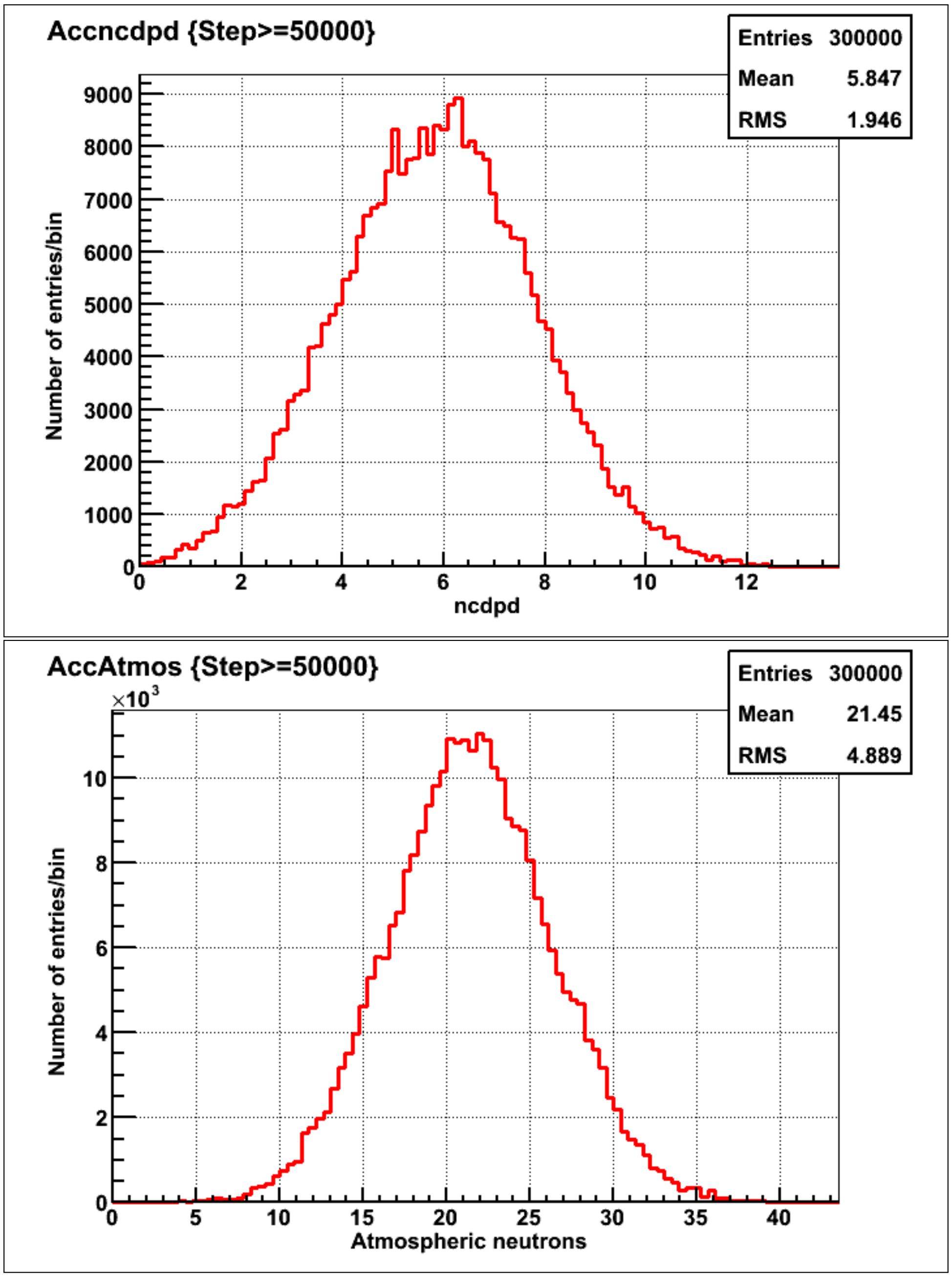


Figure 10.18: Posterior distributions showing the mean number of ncdpd neutrons and atmospheric neutrons.

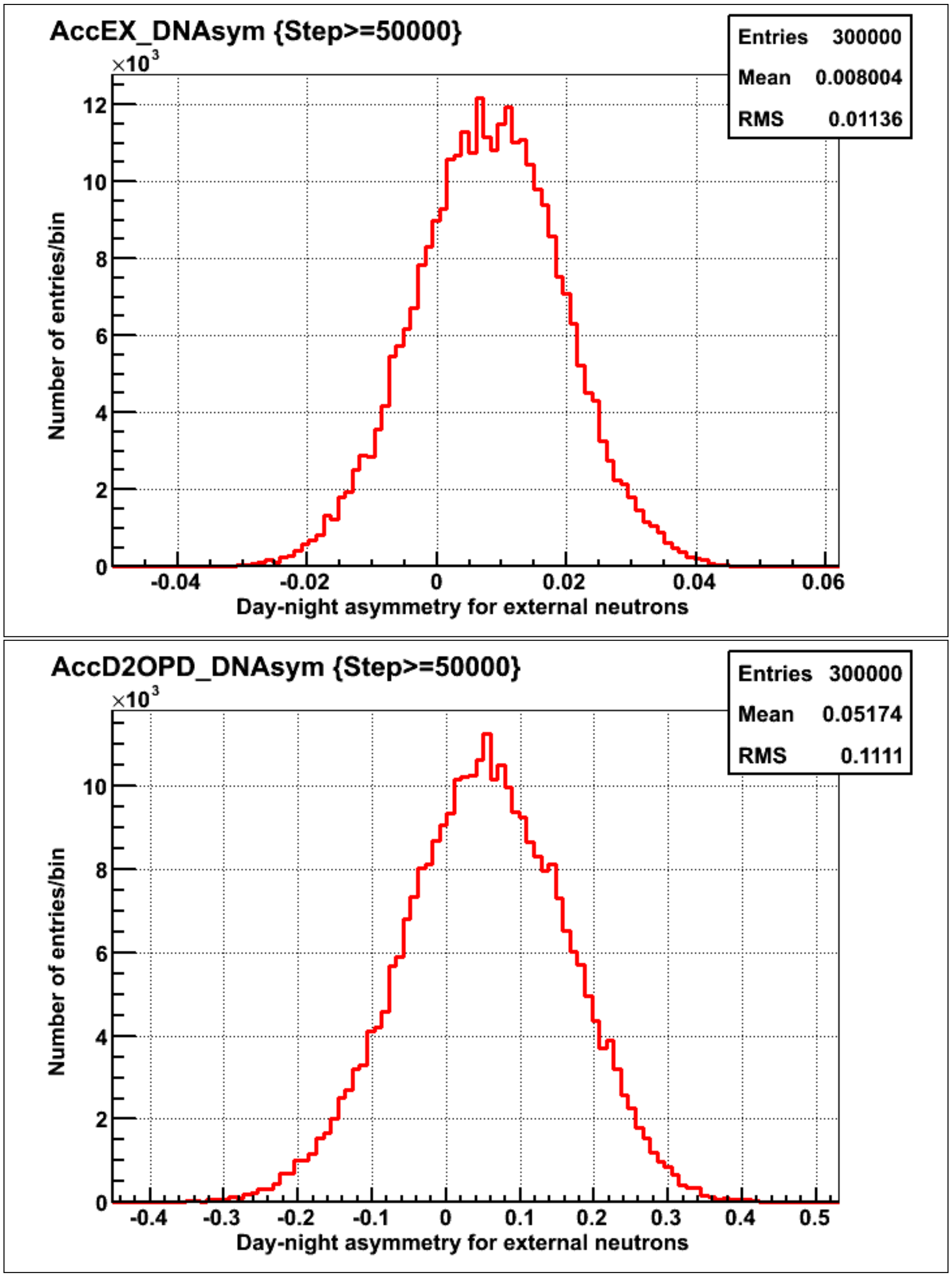


Figure 10.19: Posterior distributions of day-night asymmetry of external neutrons [top] and $d_2$opd [bottom] neutrons.

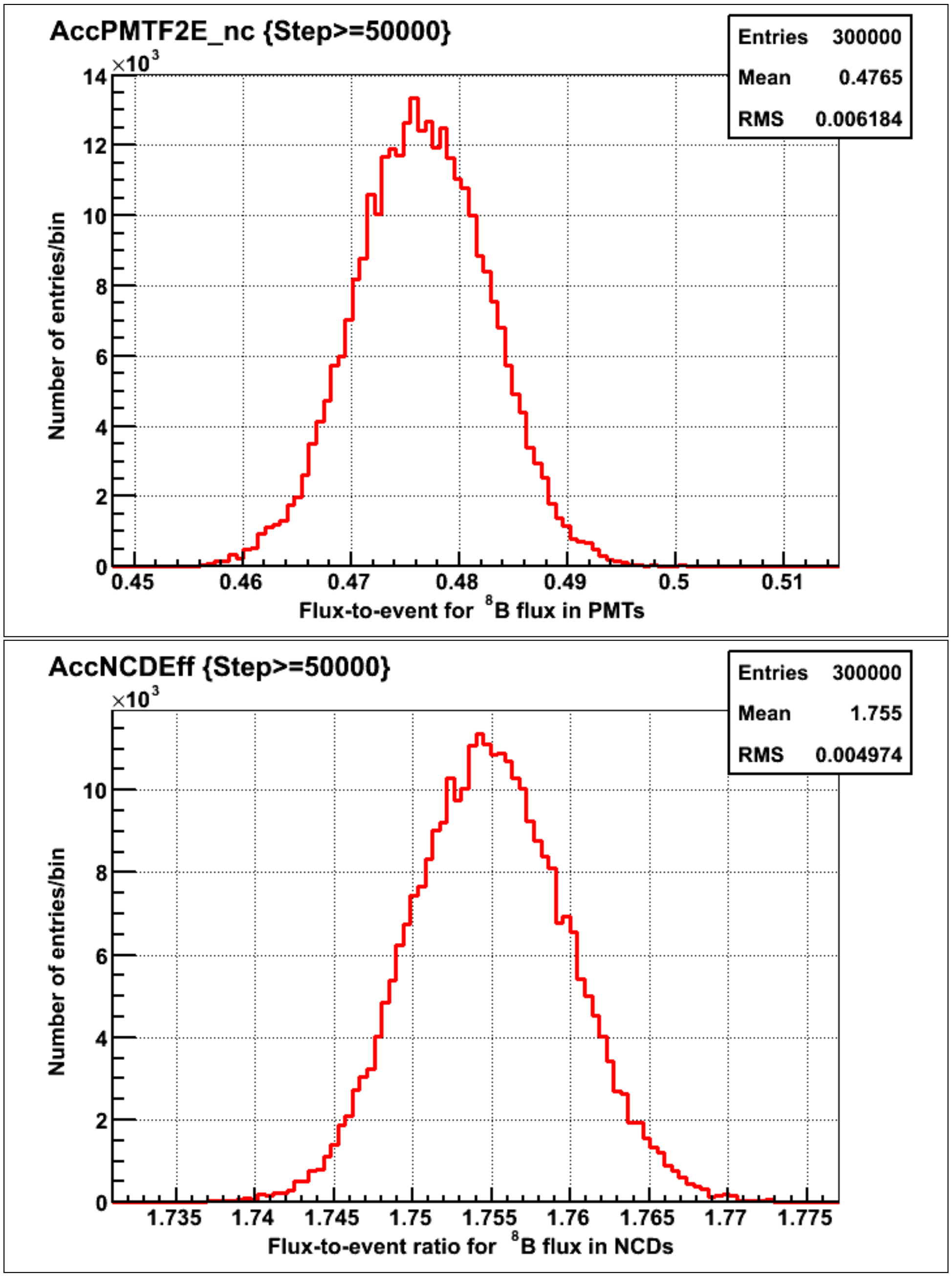


Figure 10.20: Plots for flux-to-event ratios for ^{8}B flux in PMTs [top] and NCDs [bottom].

Parameter	**Central Value of the Constraint**	**Fit Result**	**Figure Number**
$d_2$opd	0.876	0.8737 ± 0.1539	Top plot 10.16
ex	0.008	0.008 ± 0.011	Bottom plot 10.16
k2pd	9.79	9.79 ± 1.445	Top plot 10.17
k5pd	7.09	7.115 ± 0.9963	Bottom plot 10.17
ncdpd	5.685	5.847 ± 1.946	Top plot 10.18
Atmospheric	21.504	21.45 ± 4.89	Bottom plot 10.18
ex D/N	1.48915	1.542 ± 0.4913	Top plot 10.19
$d_2$opd D/N	0.0505	0.0517 ± 0.111	Bottom plot 10.19
PMT f2e	0.477	0.4765 ± 0.006	Top plot 10.20
NCD f2e	1.75465	1.755 ± 0.005	Bottom plot 10.20

Table 10.8: Comparing constraint to the result from the MCMC fit. Fit result, in column 3, consist of mean and RMS from the posterior distribution; D/N stands for day-night asymmetry and f2e is flux-to-event ratio.

10.3.1 Conclusion of the Convergence Test

This section shows that purposely selecting initial values far from away from the true values has no effect on the fit. Convergence was reached after 5,000 steps which is 1.4% of the total number of steps (350,000) that were used in the fit. For the **nominal** fits, starting values were randomly selected according to a Gaussian distribution ($\mu, \pm 3\sigma$) of width $\pm 3\sigma$ where μ is the central value of a constraint or a nominal value used in the generation of the datasets and σ is the width of the constraint or the uncertainty of the nominal value.

10.4 Comparing MCMC to QSigEx for the full Monte Carlo using the LETA constraint

The motivation for this ensemble test was to compare MCMC to QSigEx hence for this ensemble test, the constraints were not changed from one dataset to another dataset. This tests whether there are coding errors or algorithm problems in either of the two independent analyses. MCMC is the code that the author of this dissertation ran at University of Alberta and QSigEx is the code that Pierre-Luc Drouin from the University of Carleton ran. QSigEx is based on maximum likelihood method in TMinuit from ROOT [93]. MCMC is based on Markov Chain Monte Carlo method using Metropolis algorithm. Tables 10.9 and 10.10 illustrate the comparison between MCMC and QSigEx for the alternate case and regular datasets respectively. For both MCMC and QSigEx, the distribution of best-fit value, from each simulated dataset, was plotted for each parameter **x**. For MCMC, Root Mean Square (RMS) of the distribution was taken as an uncertainty on the best-fit value of the parameter **x** (See column 3 in tables 10.9 and 10.10). The mean shift, mean spread and the uncertainty on the mean shift was computed by plotting the difference, between the best-fit values from MCMC and QSigEx, divided by the average uncertainty (equation (10.2)):

$$\frac{2(q[x]-m[x])}{(\sigma_q[x]+\sigma_m[x])} \tag{10.2}$$

where $q[x]$ and $m[x]$ are the best-fit values for the parameter x, $\sigma_q[x]$ and $\sigma_m[x]$ are the uncertainties on the best-fits for the parameter x from QSigEx and MCMC respectively. The mean of the distribution is the mean shift and the RMS of the distribution is the mean spread, listed in column 6 and 7 of tables 10.9 and 10.10. The uncertainty on the mean shift is mean spread divided by the square root of the number of datasets which in this ensemble test was 14. The correlation between MCMC and QSigEx was obtained by plotting 2 Dimensional histogram (**Hist** $q[x] : m[x]$) and then using a function from ROOT Hist.GetCorrelationFactor(). See the last column in tables 10.9 and 10.10).

Comparing best-fit results (peaks of posterior distributions after removing the burn-in period) of a_0, a_1 and ^{8}B scale from MCMC to QSigEx in the left diagrams of the following figures 10.21, 10.22 and 10.23. The right plots compare computation of the equations (10.3) and (10.4) in red and blue respectively. The comparison is for each of the 14 fitted simulated datasets, shown in the X axis in the figures.

$$\frac{(q[x] - m[x])}{\sigma_q[x]} \tag{10.3}$$

$$\frac{(q[x] - m[x])}{\sigma_m[x]} \tag{10.4}$$

where $q[x]$ and $m[x]$ are the best-fit values for the parameter x, $\sigma_q[x]$ and $\sigma_m[x]$ are the uncertainties on the best-fits for the parameter x from QSigEx and MCMC respectively.

Comparing these plots with plots B.50, B.51 and B.52, it is evident that the restriction from LETA reduced the exploring region of the MCMC fit, hence the confidence levels have reduced too.

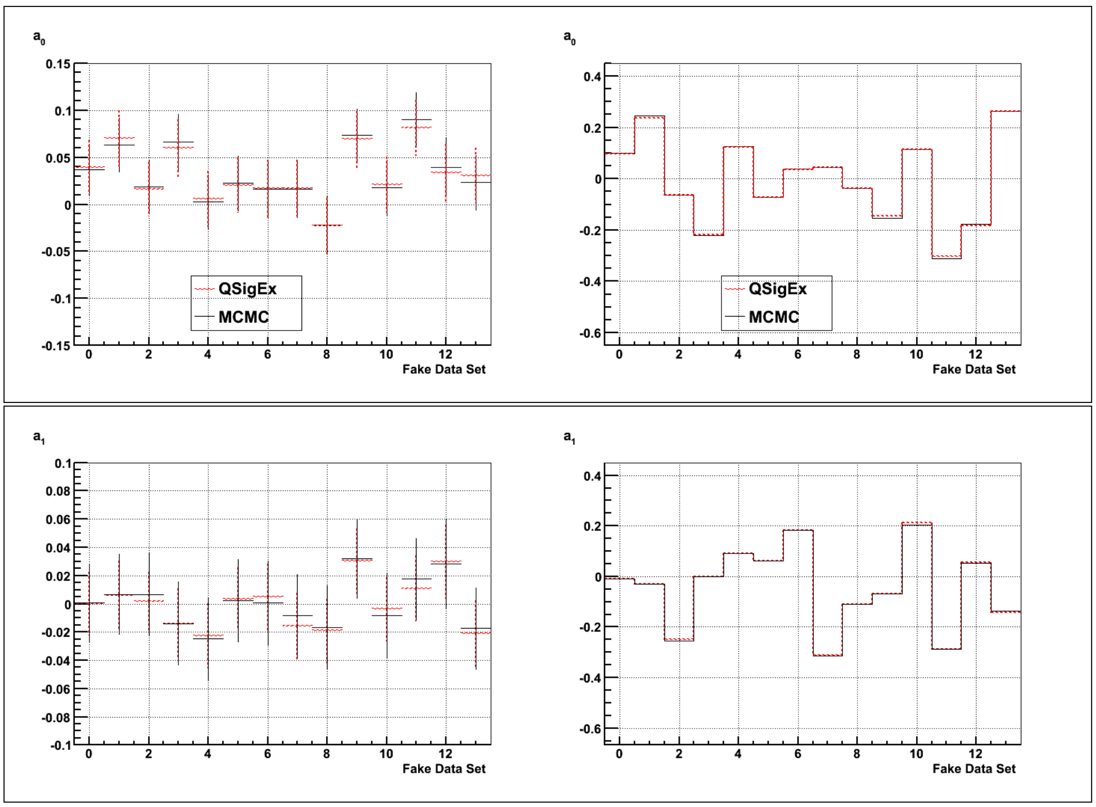


Figure 10.21: Comparing the best-fit of a_0 and a_1 along with their relative errors (from equations (10.3) and (10.4)), for each of the dataset, between MCMC and QSigEx. The ensemble test consist of 14 regular datasets.

Parameter	**Generated**	**QSigEx**	**MCMC**	**Mean Shift**	**Mean Spread**	**Correlation**
^{8}B	1.0	0.9937 ± 0.0226	0.993 ± 0.023	(-0.0295 ± 0.0624)	0.2336	0.955
p_0	0.325	0.3219 ± 0.0134	0.322 ± 0.013	(0.0135 ± 0.0614)	0.2297	0.965
p_1	-0.00888	-0.0099 ± 0.0053	(-0.0100 ± 0.0055)	(-0.0202 ± 0.0497)	0.1859	0.972
p_2	0.00122	(0.0028 ± 0.0034)	(0.0029 ± 0.0034)	(0.0408 ± 0.031)	0.1160	0.995
a_0	0.028	(0.0329 ± 0.0277)	(0.0331 ± 0.0294)	0.0077 ± 0.0437	0.1634	0.998
a_1	0.0048	(-0.00053 ± 0.00164)	(0.00042 ± 0.01640)	(0.0435 ± 0.04198)	0.1571	0.976

Table 10.9: Comparing result between QSigEx and MCMC for the regular datasets. Systematics were not floated.

Parameter	**Generated**	**QSigEx**	**MCMC**	**Mean Shift**	**Mean Spread**	**Correlation**
^{8}B Scale	1.0	0.9885 ± 0.0269	0.9939 ± 0.0302	-0.159 ± 0.053	0.1965	0.987
p_0	0.325	0.3275 ± 0.0099	0.3285 ± 0.0095	0.0657 ± 0.0348	0.1303	0.979
p_1	-0.00888	-0.0088 ± 0.0065	-0.0092 ± 0.0068	-0.0529 ± 0.0380	0.1424	0.990
p_2	0.00122	0.00081 ± 0.0031	0.00094 ± 0.00292	0.0498 ± 0.0373	0.1396	0.993
a_0	0.028	0.02844 ± 0.0205	0.0314 ± 0.0224	0.0994 ± 0.0619	0.2317	0.955
a_1	0.0048	0.0065 ± 0.0232	0.0067 ± 0.0257	0.0084 ± 0.0554	0.2074	0.985

Table 10.10: Comparing result between QSigEx and MCMC for the alternate datasets. Systematics were not floated.

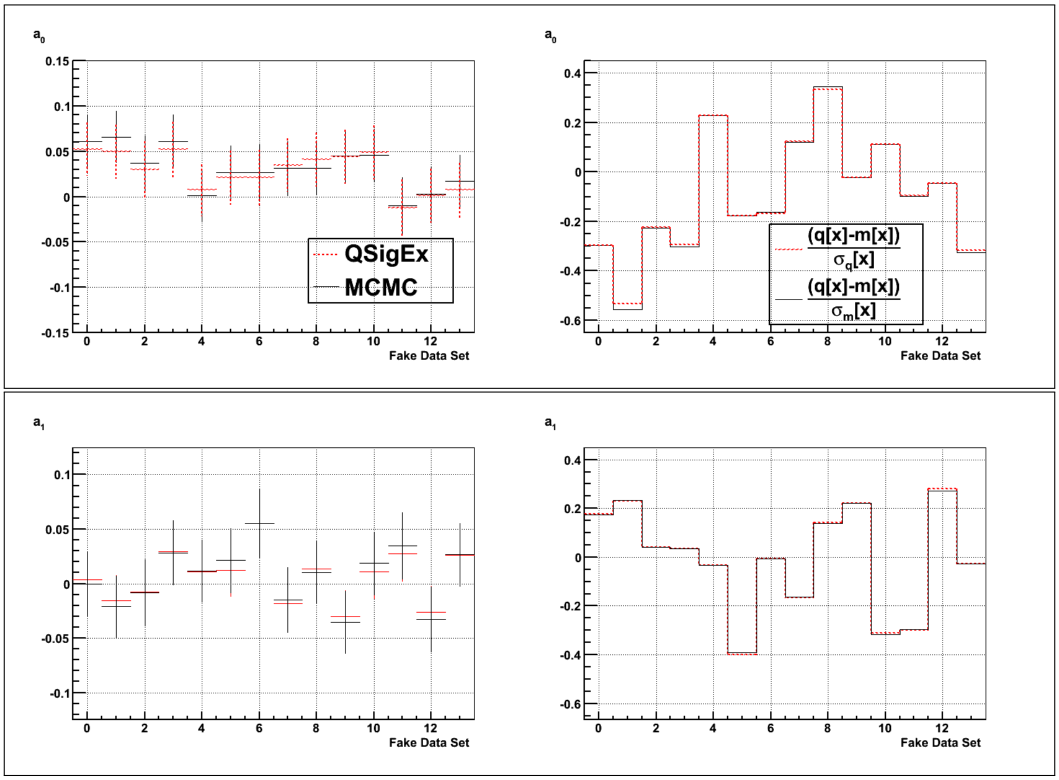


Figure 10.22: Comparing the best-fit results of a_0 and a_1 along with their relative errors (equations 10.3 and 10.4) between MCMC and QSigEx. The ensemble test consist of 14 alternate datasets.

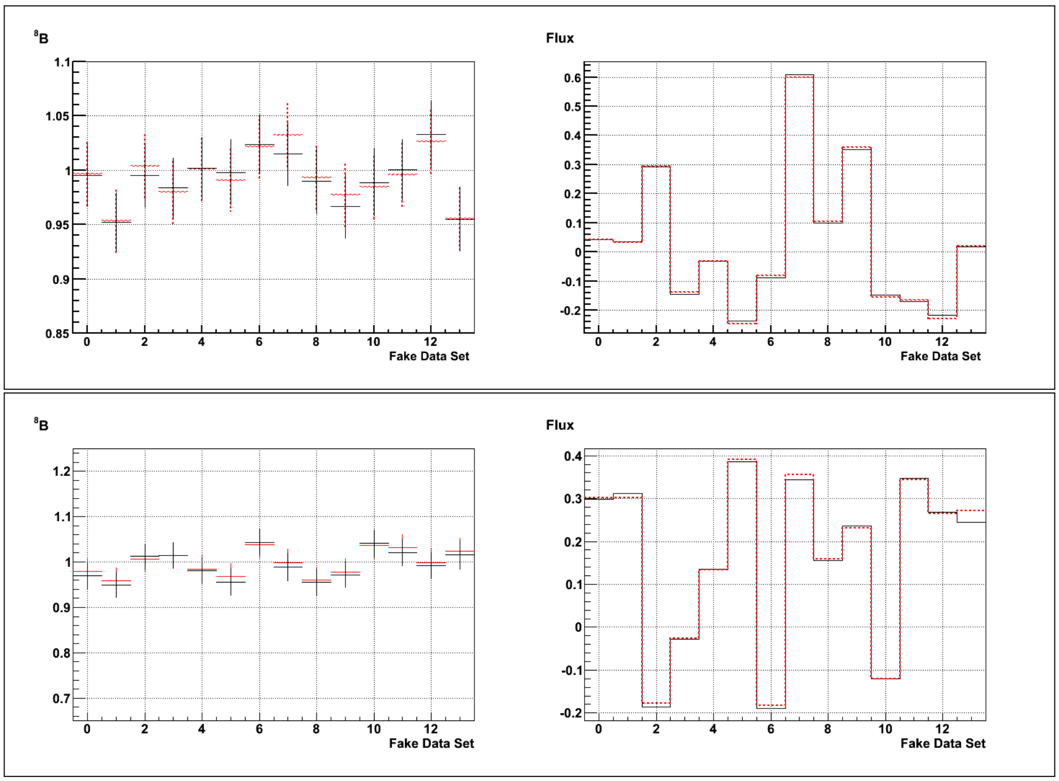


Figure 10.23: Comparing ^{8}B scale and the relative error in ^{8}B scale from MCMC to QSigEx for each of the 14 fitted datasets shown in the X axis. The top is for the regular dataset and the bottom plot is for the alternate dataset.

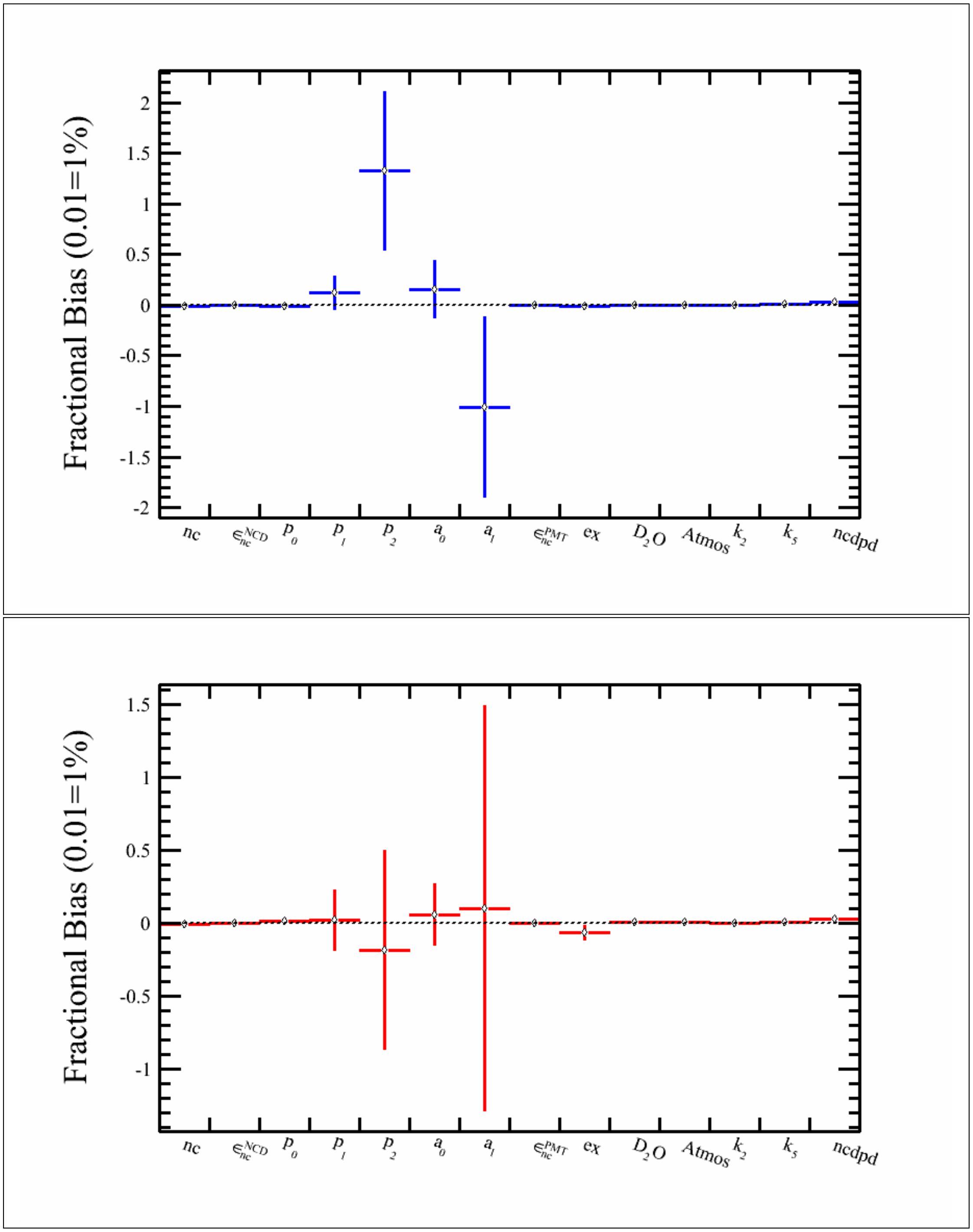
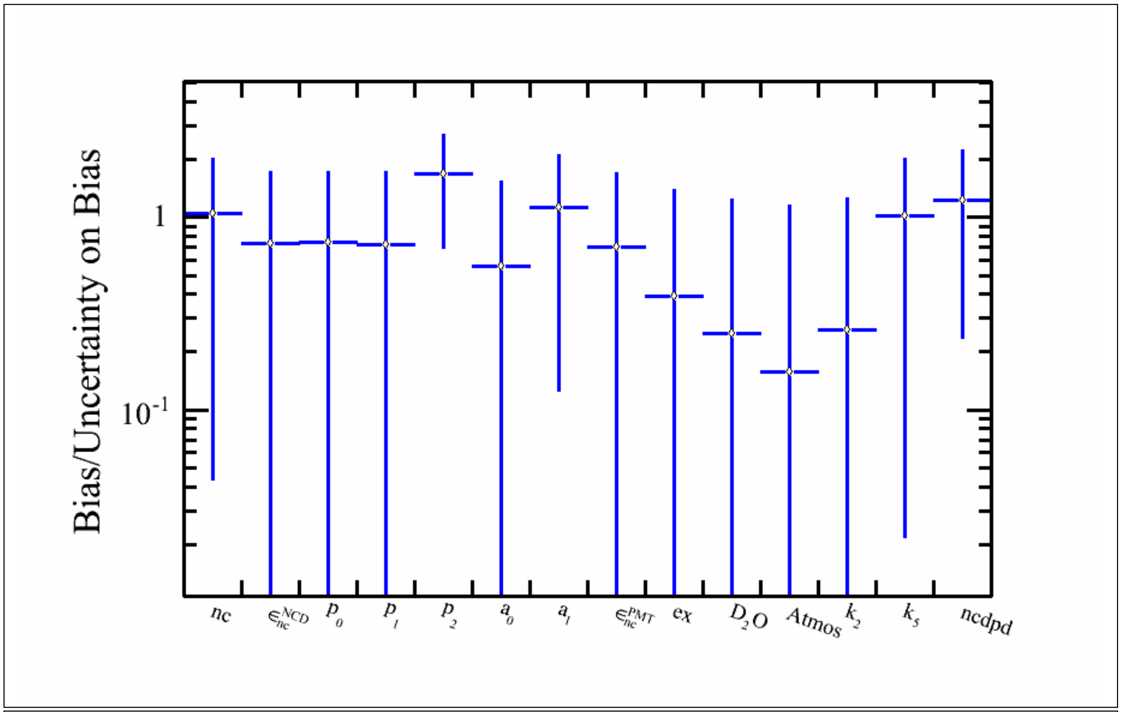


Figure 10.24: Comparing bias for the regular dataset in blue and for the alternate dataset in red.



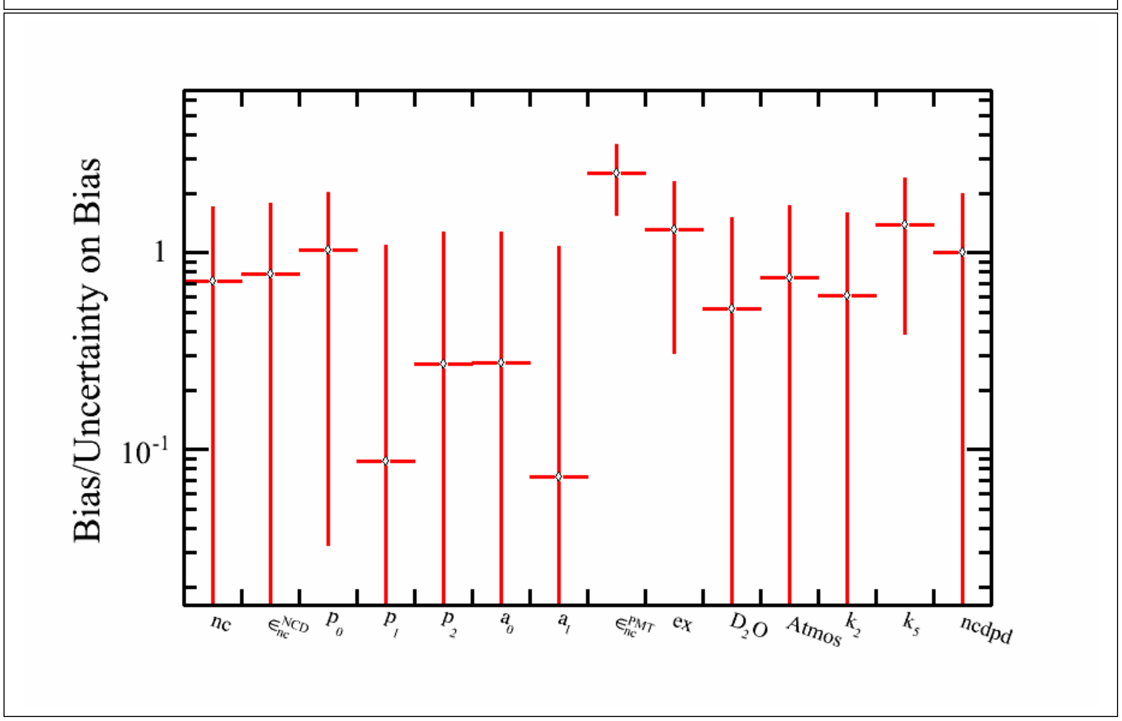
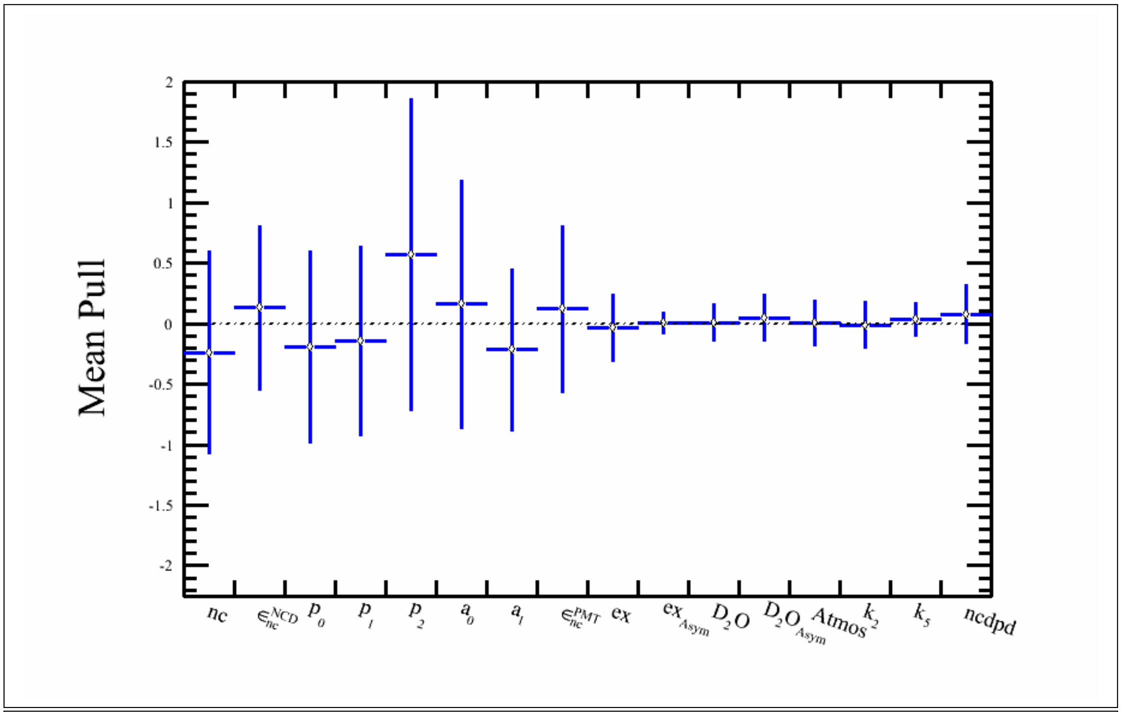


Figure 10.25: Comparing bias/uncertainty for the regular dataset in blue and for the alternate dataset in red.



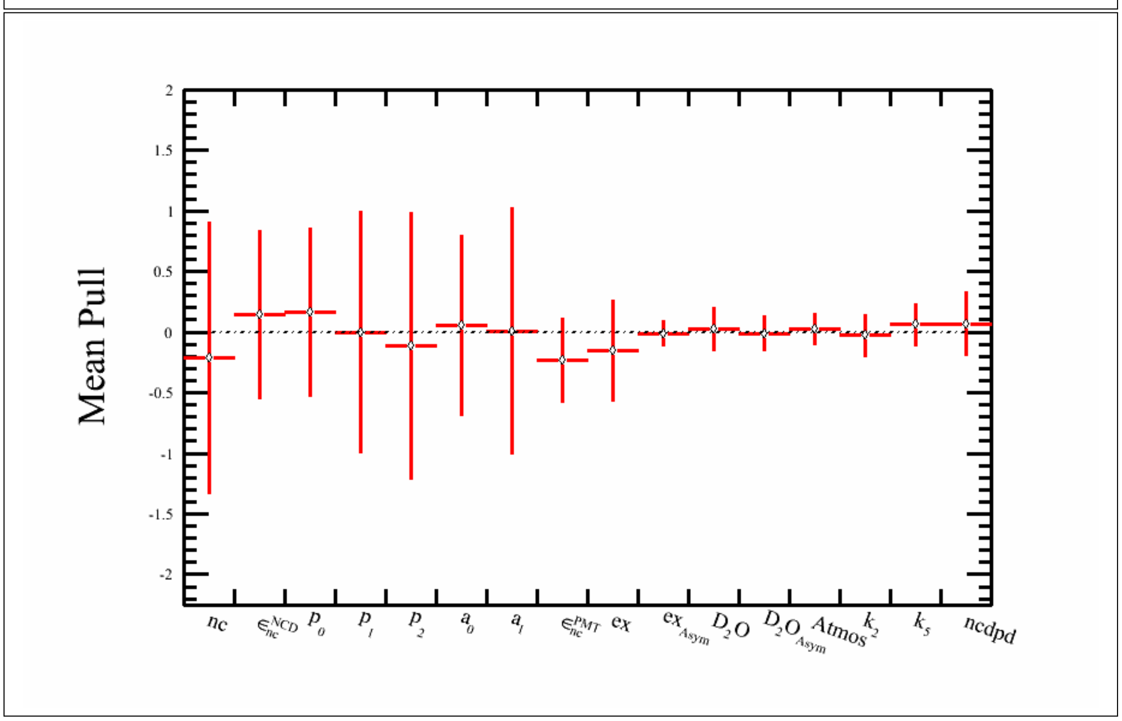


Figure 10.26: Comparing pull spread for the regular dataset in blue and for the alternate dataset in red. Since the constraints were not changed from one file to the next, the pull width is not 0.949.

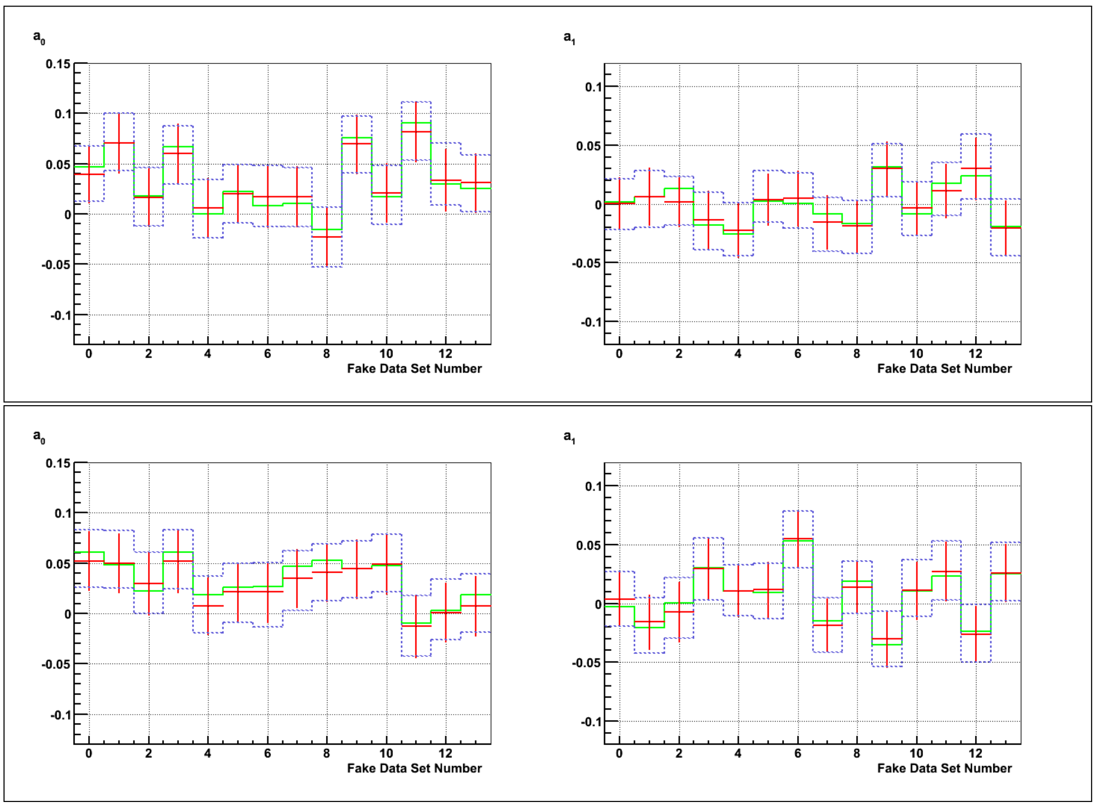


Figure 10.27: Showing best-fit result in green color for day-night asymmetries (a_0 and a_1) for each of the 14 fitted simulated datasets shown in the X axis. Systematics were not floated. The blue dotted lines show $\pm\sigma$ confidence levels from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red. The top is for the regular dataset and the bottom plot is for the alternate dataset.

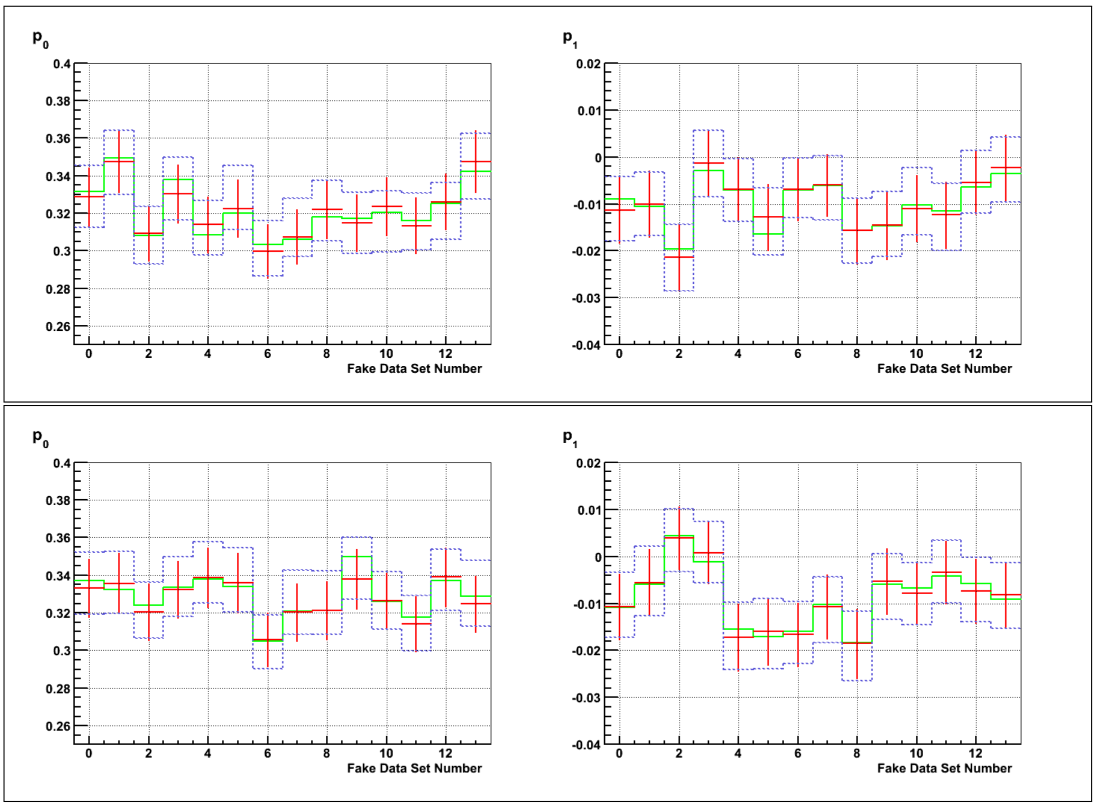


Figure 10.28: Showing best-fit result in green color for P_{ee} parameters (p_0 and p_1) for each of the 14 fitted simulated datasets shown in the X axis. Systematics were not floated. The blue dotted lines show $\pm\sigma$ confidence levels from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red. The top is for the regular dataset and the bottom plot is for the alternate dataset.

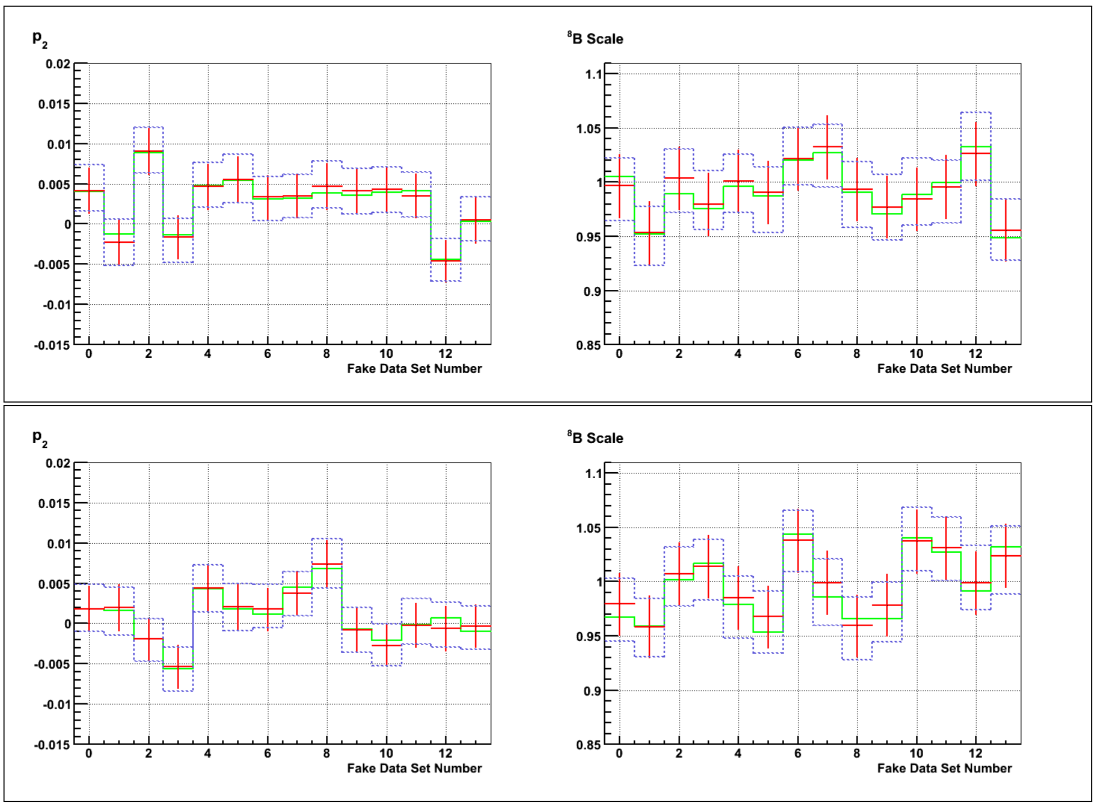


Figure 10.29: Showing best-fit result in green color for P_{ee} parameter (p_2) and ^{8}B scale for each of the 14 fitted simulated datasets shown in the X axis. Systematics were not floated. The blue dotted lines show $\pm\sigma$ confidence levels from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red. The top is for the regular dataset and the bottom plot is for the alternate dataset.

10.5 Summary

This chapter described two MCMC fit results (applying the PSA constraint and applying constraints from both PSA and LETA) which included all the backgrounds. The final section detailed the comparison between QSigEx and MCMC for the full PSA ensemble test including the LETA constraint. Even though the number of simulated datasets were limited (15 for PSA and 14 for PSA+LETA), the result showed that QSigEx and MCMC agree and converge to an unbiased result with correct uncertainties. Before running the code on the 1/3 data, it was tested on 1/3 simulated datasets; the result is described in chapter 11. This chapter also outlined the convergence test to check whether the likelihood function works if the values of the parameters fitted starts far from the actual values.

Chapter 11

Ensemble Test on 1/3 of the Simulated Dataset

11.1 Introduction

The difference between the full fit and the one third fit is that the expected number of events for each class corresponds to the expected number of events for the one-third of the real data. The ensemble test consisted of forty five simulated datasets. The motivation for the one third test is to compare the MCMC result to the result from the QSigex and for that reason the constraints were not changed from file to file; hence the width of the pull is not 0.983[1]. The fit included constraint from PSA on the total number of neutrons from the NCDs. The constraint from LETA was not included in this test. There are thirteen event classes including the 3 hep classes which are fixed in the MCMC fit. Besides floating the number of events for 7 event classes, the fit also floats:

1. NC capture efficiency uncertainty on the PMT side
2. NC capture efficiency on the NCD side

[1]The error bars on the pull plots indicate the average spread of the parameter and not the uncertainty on the average pull. From statistics, the pull width as a function of the number of datasets **n**, is given as $\sqrt{[(n-1)/n]}\big(1 - \frac{1}{4(n-1)}\big)$; if n=45 files, the width of the pull is 0.983.

3. Day-night diurnal asymmetry for the external neutron background

4. Day-night directional asymmetry uncertainty for the D_2O photo-disintegration background

Four cases were considered for the fit. Two for the regular-simulated dataset and two for the alternate-simulated dataset; (a) fit with fixed systematic uncertainties and (b) fit with additional 8 systematic uncertainties in the fit.

11.2 Fit with Fixed Systematics

The number of steps in the MCMC fit is 750,000; 100,000 of which is rejected as a burn-in period. Systematics were not floated for this ensemble test. The negative log likelihood (NLL) function, for the 1/3 fit is:

$$
\begin{aligned}
-\log\,\mathcal{L} = \sum_{i=1}^{2m} N_i - \sum_{d=1}^{N}\log\Big(\sum_{i=1}^{2m}(N_i)F_i(\vec{x}_d,\vec{P})\Big) \\
+\frac{(\mathrm{PSA} - B\,\epsilon_1 - N_1\kappa_1 - N_2\kappa_2 - N_3\kappa_3 - N_4\kappa_4 - N_5\kappa_5 - N_6\kappa_6 - 4.363/3)^2}{2\,(\sigma_P)^2} \\
+\frac{(\overline{\alpha}_1-\alpha_1)^2}{2\sigma_1^2}+\frac{(\overline{\alpha}_2-\alpha_2)^2}{2\sigma_2^2}+\frac{(\overline{\alpha}_3-\alpha_3)^2}{2\sigma_3^2}+\frac{(\overline{\alpha}_4-\alpha_4)^2}{2\sigma_4^2} \\
+\frac{(\overline{\alpha}_5-\alpha_5)^2}{2\sigma_5^2}+\frac{(\overline{\alpha}_6-\alpha_6)^2}{2\sigma_6^2}+\frac{(\overline{\xi}_0-\xi_0)^2}{2\sigma_{\xi_0}^2}+\frac{(\overline{\xi}_1-\xi_1)^2}{2\sigma_{\xi_1}^2} \\
+\frac{(\overline{\epsilon}_1-\epsilon_1)^2}{2\sigma_{\epsilon_1}^2}+\frac{(\overline{\epsilon}_0-\epsilon_0)^2}{2\sigma_{\epsilon_0}^2}+\sum_{k=1}^{8}\frac{(\overline{\beta}_k-\beta_k)^2}{2\sigma_k^2}
\end{aligned}
\tag{11.1}
$$

where $\mathrm{F}_i(\vec{x}_d,\vec{P})$ is the probability density function, for the class i, giving the probability of observing the event d with observables $\vec{x}_d$ and the current values of the fit parameters $\vec{P}$, $N_1, N_2, \ldots, N_m$ are the number of events for **m=13** event classes and **i** goes from 1 to **N** data entries. The ^{8}B flux is designated by **B** and **PSA** is the PSA constraint for the current dataset and σ_P is the width of the constraint which is 7.2% for the one third dataset. The values of

κ for various backgrounds is listed in the table 4.7. The number $4.363/3$ is the average number of NC interactions from the hep neutrinos (^{3}He+p $\rightarrow$ ^{4}He + $e^+ + \nu_e$) expected to be detected in the NCDs for the 1/3 fit. In the constraint terms, α_1, α_2, α_3, α_4, α_5 and α_6 are the values of α (equation (4.19)) in the current MCMC step for EX, $d_2$opd , atmospheric neutrons, k2pd, k5pd and ncdpd respectively. The $\overline{\alpha}_1$, $\overline{\alpha}_2$, $\overline{\alpha}_3$, $\overline{\alpha}_4$, $\overline{\alpha}_5$ and $\overline{\alpha}_6$ are the constraints for EX, $d_2$opd , Atmospheric neutrons, k2pd, k5pd and ncdpd respectively for a given dataset. The day-night asymmetries for the external neutrons and $d_2$opd are represented by ξ_0 and ξ_1 respectively. The flux-to-event for the NCDs and PMTs are represented by ϵ_1 and ϵ_0 respectively. The $\overline{\beta}_k$ is the central value of the kth systematic uncertainty and β_k is the current value of the systematic uncertainty in the MCMC fit. Eight systematic uncertainties were floated so k goes from 1 to 8. When the systematics are not floated then $k = 0$.

The pull and bias plots for the fit are shown in figures 11.1 to 11.3. Since the constraints were not changed from data file to data file, the width of the pull is not 0.983. From the bias plot, it appears that except NC all the biases are within zero of the uncertainties. Each fit result is derived from the peak of the posterior distributions. This is a modest attempt to make MCMC close to the maximum likelihood technique. The MCMC fit, for the regular dataset, resulted in -0.0177 ± 0.0036 bias in NC, 4.9σ away from zero, as shown in figure 11.4. For the alternate case, the bias is -0.0195 ± 0.0033 which is 5.9σ away from zero. Since the situation is different for the alternate dataset, it is evident that statistics played a role because the difference between the regular and alternate dataset is a seed used to generate them.

11.3 Floating 8 Systematics as parameters in the Fit

Following systematic parameters were floated are:

1. Phase-correlated energy scale
2. Energy scale (for the NCD phase)
3. Relative energy resolution shift
4. Radial vertex scale
5. $\cos\theta$ uncertainty for ES only
6. Day-night diurnal energy scale uncertainty
7. Day-night directional energy scale uncertainty for ES only
8. Day-night directional $\cos\theta$ uncertainty

The results are displayed in figures 11.5 and 11.6. For the regular dataset, 420,000 steps were taken in the fit for each dataset from which 50,000 steps were removed as burn-in; for the alternate dataset, 344,000 steps were taken in the fit for each dataset and burn-in was 80,000 steps.

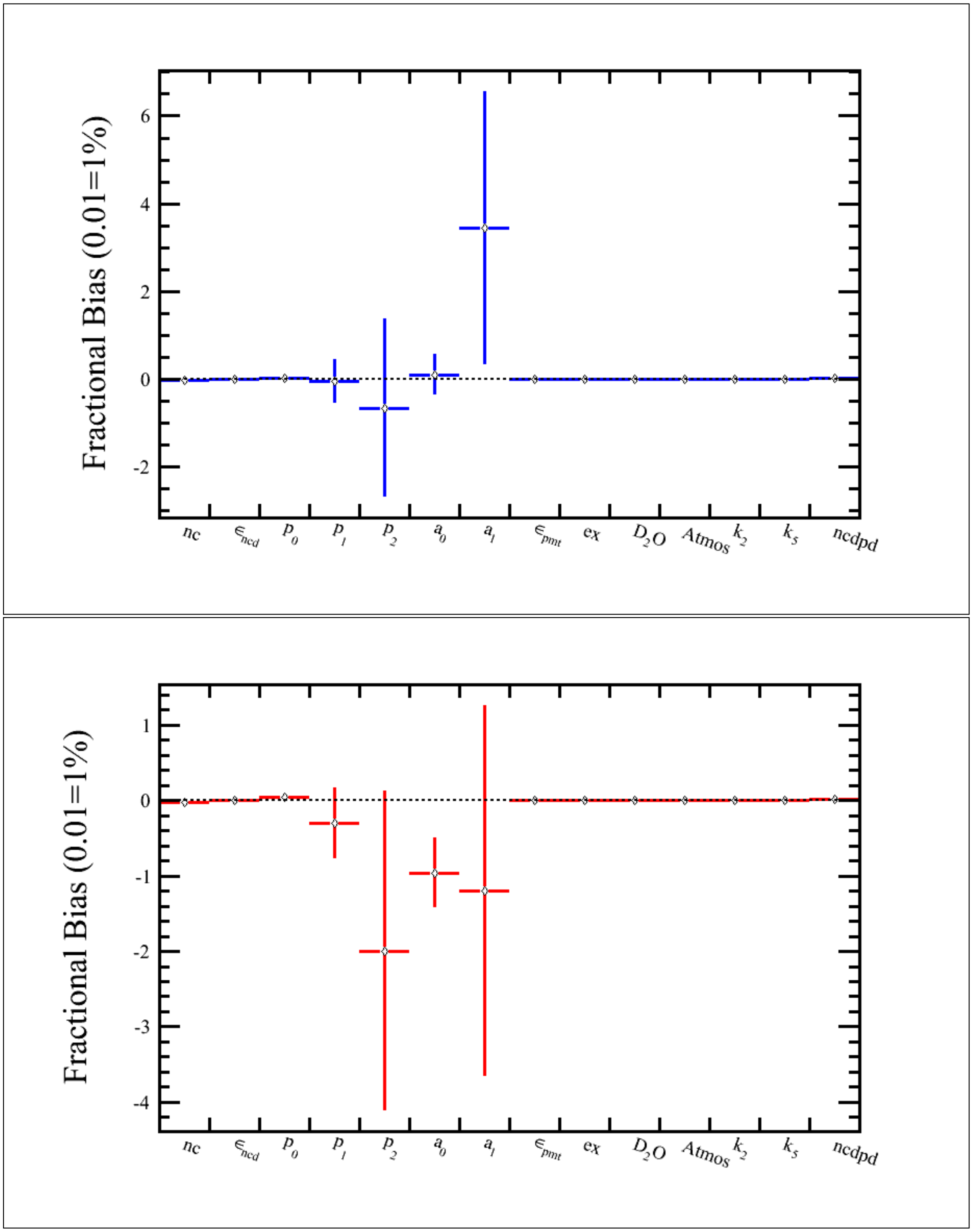


Figure 11.1: Result from the fit; the top plot is the bias spread for the regular dataset and the the bottom plot is the bias spread for the alternate dataset. The bias on a_1 changed sign from + to - between regular dataset and alternative dataset.

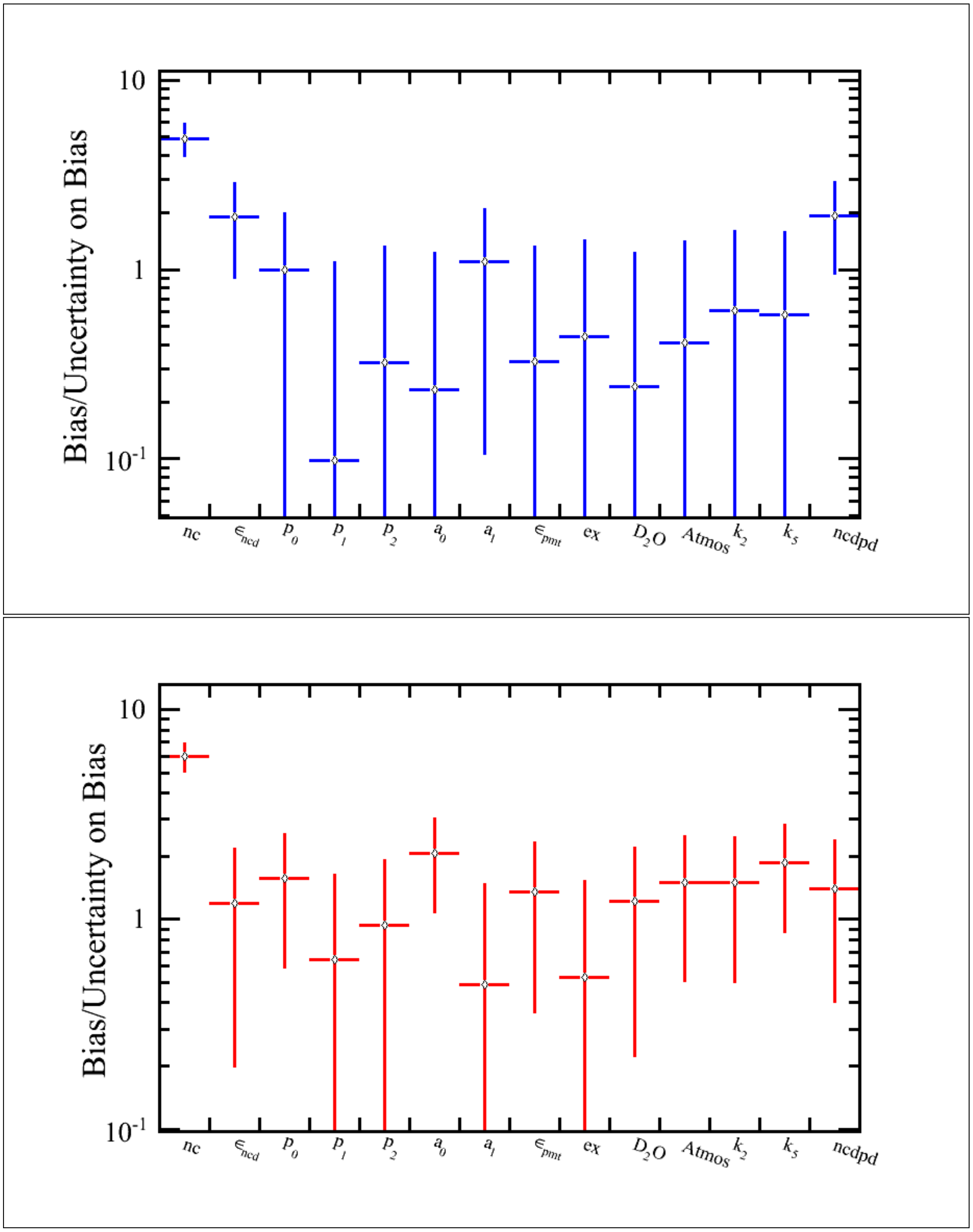


Figure 11.2: Bias divided by the uncertainty in the bias for the 1/3 fit; the top plot shows the result for the regular dataset and the bottom plot shows the result for the alternate dataset. No systematics were floated for this fit. The bias on NC detection efficiency in NCDs (second bin) and ncdpd (last bin) improved from regular dataset to alternate dataset.

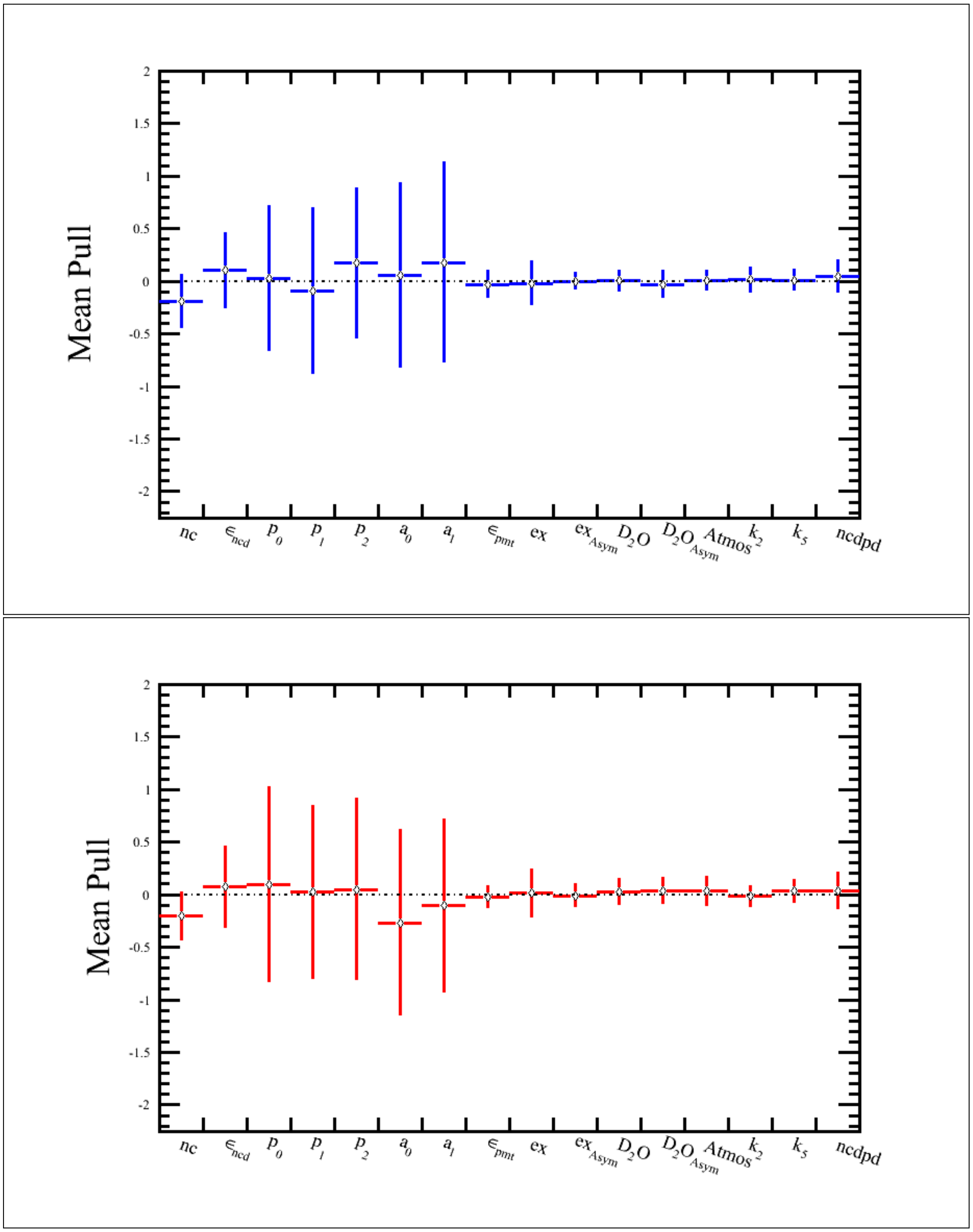


Figure 11.3: Pull spread for the regular dataset in blue and for the alternate dataset in red. Sine the constraints were not changed from one dataset to the next, the width of the pull is not 0.983. The pull of a_0 and a_1 flipped sign from + to - between regular to alternative datasets.

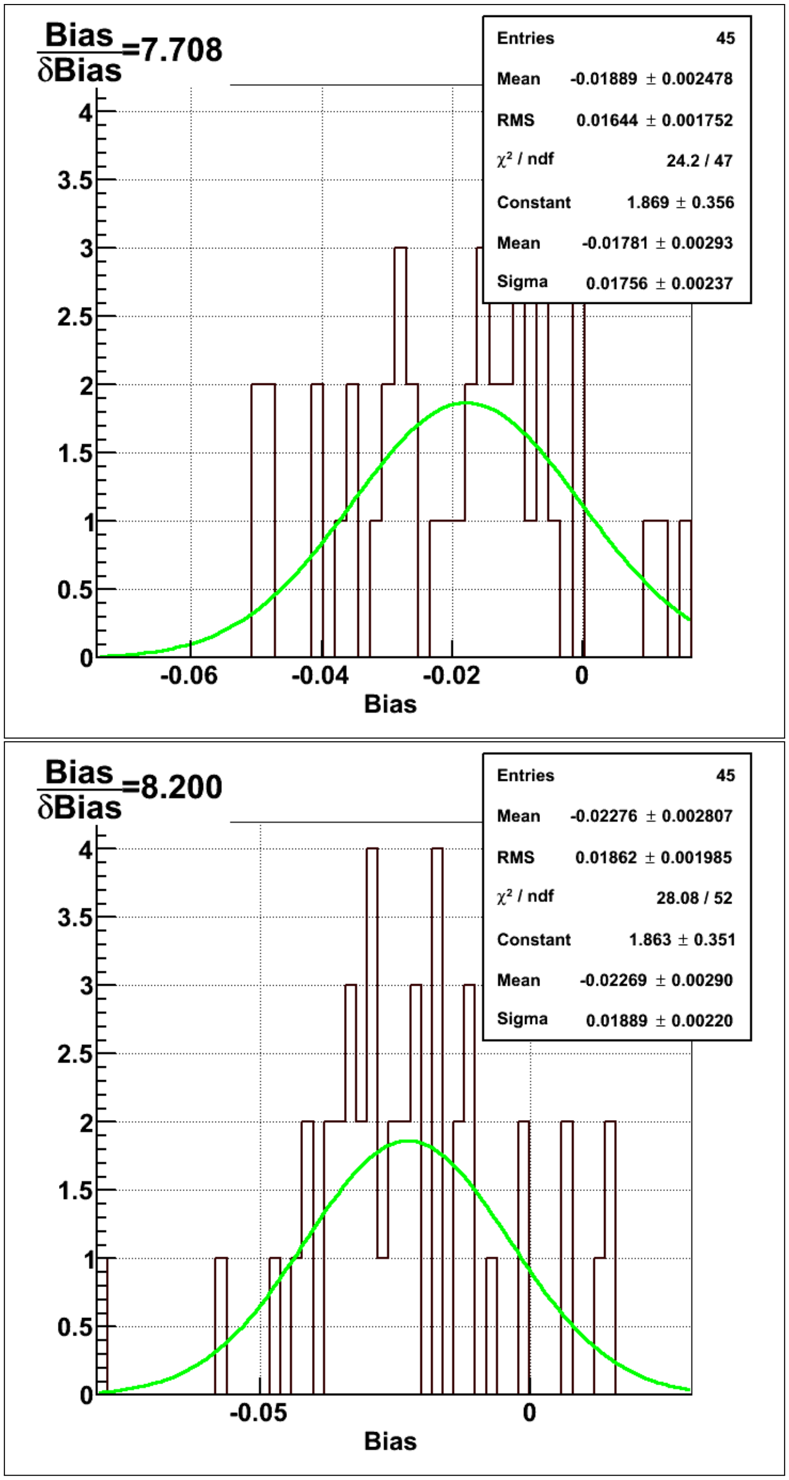


Figure 11.4: NC fit result of the third of datasets. The plot shows the Gaussian fit of the bias distribution of the ^{8}B flux for the regular dataset (top) and for the alternate dataset (bottom). The bias is better for the regular dataset than the alternate dataset.

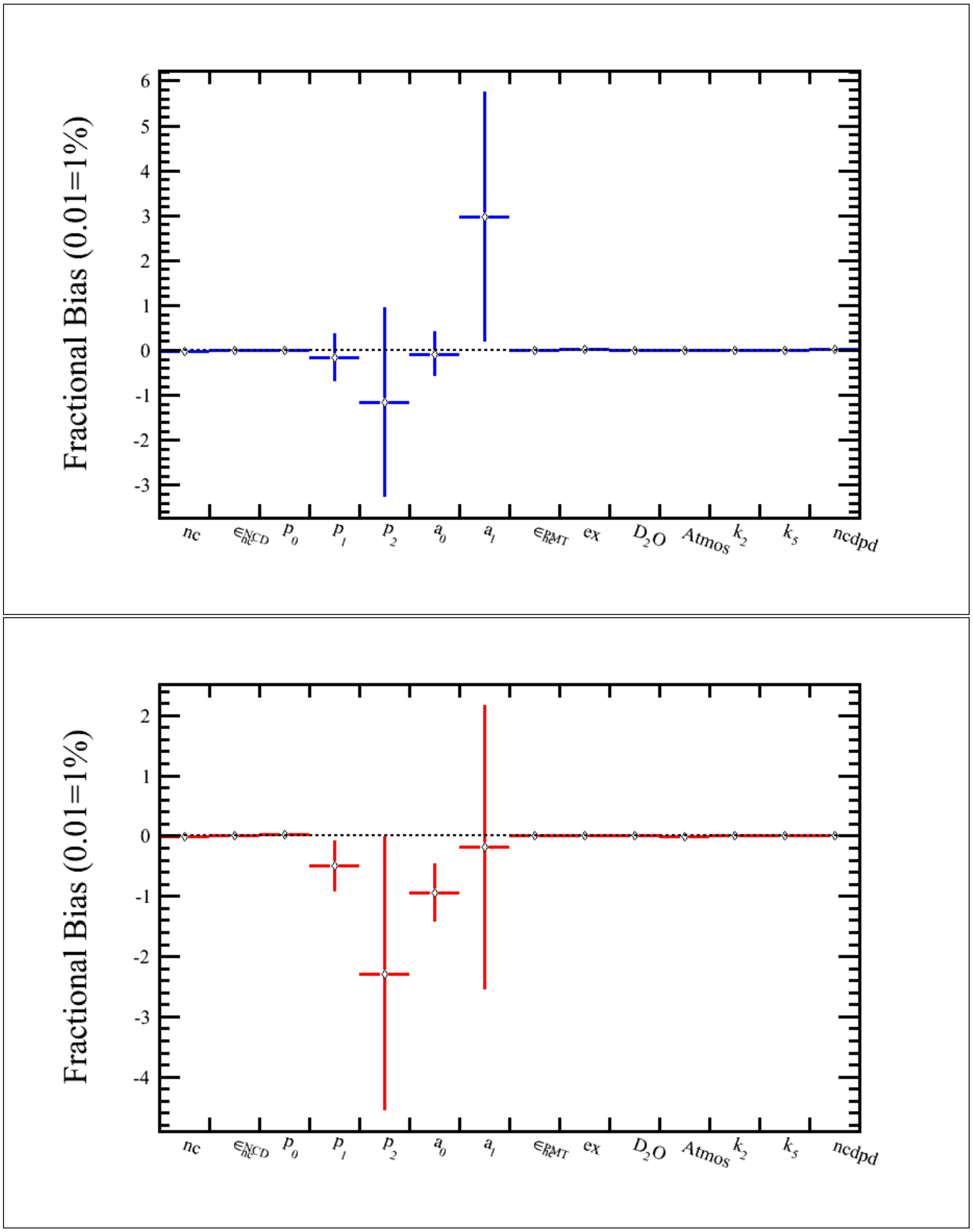


Figure 11.5: Result of 1/3 fit including 8 systematics. Top plot is the bias spread for the regular dataset and the bottom plot is the bias spread for the alternative dataset. The bias on a_1 flipped sign from + to - between the regular dataset and alternative dataset.

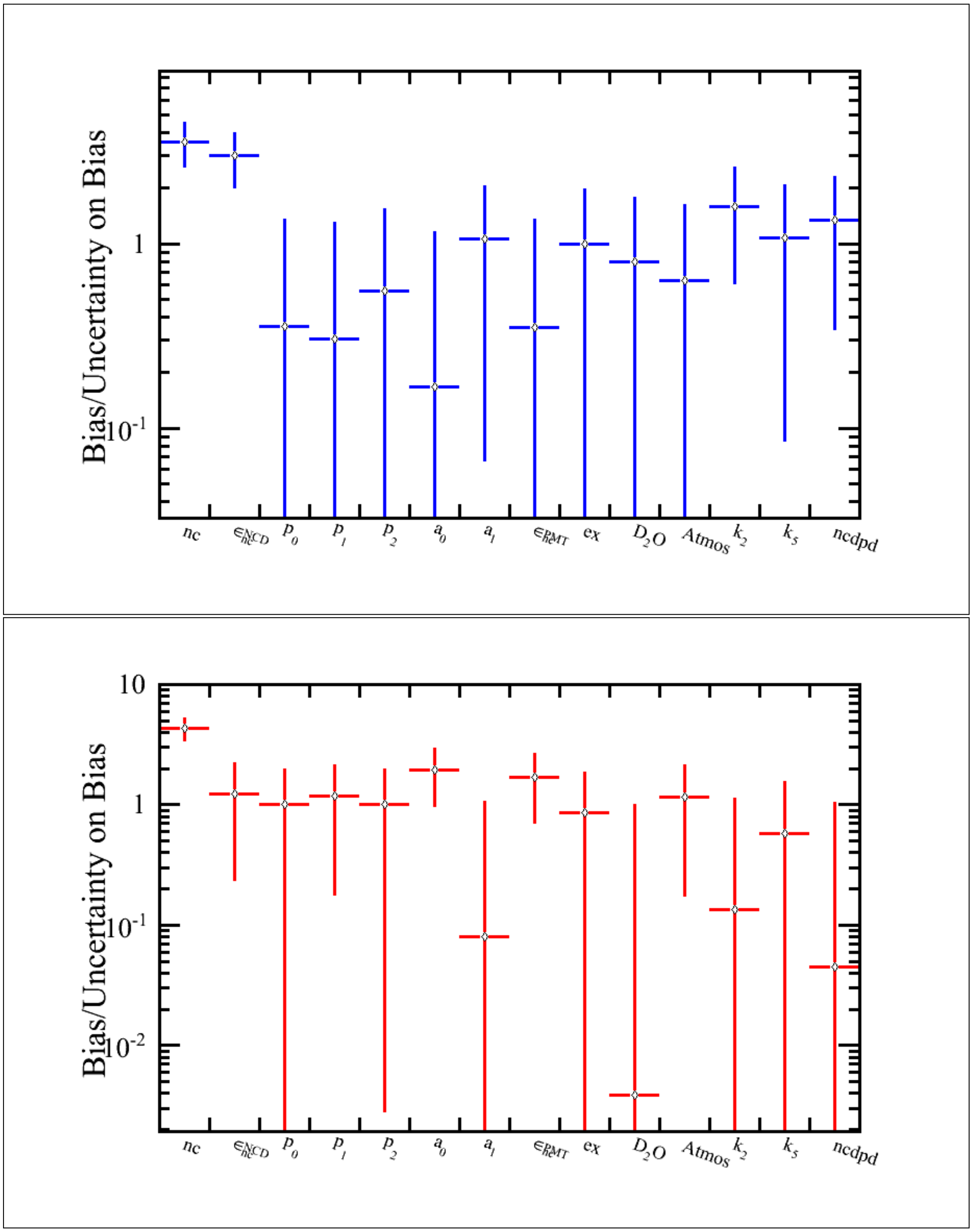


Figure 11.6: Bias divided by the uncertainty in the bias for the 1/3 fit. Top plot shows the result for the regular dataset and the bottom plot shows the result for the alternative dataset. Fit includes 8 systematics. Bias on NC detection efficiency in NCDs improved from regular dataset to alternative dataset hence it is an effect of statistics. Bias of all the parameters are zero within uncertainties except for NC which is more than 3σ away from zero.

This chapter describes the number of steps taken to figure out the cause of the bias. Table 11.1 lists the NC bias obtained for each step along with a reference to the corresponding figure that displays the NC bias distribution fitted to a Gaussian function. Plots are presented in the Appendix.

For the first nine "steps", as described below, all P_{ee} parameters were fixed to their nominal values.

1. Motivation: To find out if there is a configuration problem with the nominal number of events or a problem in the likelihood function.

   Information is passed to the MCMC program via a configuration file and if there is any misinformation in the configuration file then it will result in a biased fit. For this test, only three classes were included CC, ES and NC and NC neutron efficiencies were fixed. PSA constraint was not used. The result in NC bias is 0.005 ± 0.004, as shown in figure B.1.

2. Motivation: To test PMT side neutron efficiency in the likelihood function.

   Only three classes were included: CC, ES and NC. Floating PMT-side NC neutron efficiency. No PSA constraint was applied. The result, as shown in figure B.2 is 0.007 ± 0.004.

3. Motivation: To test PSA constraint.

   Fit contained only three classes: CC, ES and NC. NC neutron efficiencies were fixed. Fit included the PSA (with only NC neutrons) constraint. The result, as shown in figure B.3, is 0.003 ± 0.003.

4. Motivation: To test the PSA constraint with neutron efficiencies.

   Fit included only three classes: CC, ES and NC. Floating NC neutron efficiencies in the fit. Applied the PSA (with only NC neutrons) constraint. The result is shown in figure B.4.

5. Motivation: To test EX in the likelihood function.

   Fit had four event classes: CC, ES, NC and EX. NC neutron efficiencies were fixed. EX was fixed to its nominal value of 1.0. No PSA constraint was applied. The result, as shown in figure B.5, is 0.005 ± 0.003.

6. Motivation: To test the PSA constraint with NC and EX neutrons.

   Four event classed were included: CC, ES, NC and EX. NC neutron efficiencies were fixed. EX fixed to its nominal value of 1. Added PSA (with NC + EX neutrons) constraint. The result, as shown in figure B.6, is 0.003 ± 0.003.

7. Motivation: To test NC neutron detection efficiencies, for PMTs and NCDs, along with the PSA constraint.

   Four classes were included: CC, ES, NC and EX. Floating NC neutron efficiencies. EX fixed to its nominal value of 1. PSA (with NC + EX neutrons) constraint was also applied. The result is shown in figure B.7; NC bias came out to be 0.003 ± 0.003.

8. Motivation: To test EX and the PSA Constraint.

   Four classes were included: CC, ES, NC and EX. Neutron efficiencies were fixed. Floating EX with PSA constraint. The result, as shown in figure B.8, is 0.004 ± 0.003.

9. Motivation: To test the day-night asymmetry.

   Four classes were included: CC, ES, NC and EX. Neutron efficiencies were fixed. Floating the day-night asymmetry for EX. The result, as shown in figure B.9, is 0.003 ± 0.003.

None of the tests performed had a bias in NC. So the next step was performing a full MCMC fit floating the P_{ee} parameters, ^{8}B, *etc.*with one background removed at a time. Following steps were undertaken to flush out the cause of the problem.

Step 1: For the first test "Step", ncdpd background was removed, the other backgrounds were fixed to their nominal values. Result, as shown in figure B.10, is -0.017 ± 0.002. The bias in NC is 7σ away from zero.

Step 2: In the second step, k5pd background was removed while keeping the remaining backgrounds at the nominal level of 1.0. The bias in NC, as shown in figure B.11, is -0.015 ± 0.002 which is 7σ away from zero.

Step 3: In the third step, k2pd background was removed while keeping the remaining backgrounds at the nominal level of 1.0. The bias in NC, as shown in figure B.12, -0.016 ± 0.002.

Step 4: In the fourth step, d2opd background was removed while keeping the remaining backgrounds at the nominal level of 1.0. The bias in NC, as shown in figure B.13, is -0.016 ± 0.002.

Step 5: In the fifth step, hep background was removed while keeping the remaining backgrounds at the nominal level of 1.0. The bias in NC, as shown in figure B.14 , is -0.017 ± 0.002.

Step 6: In the sixth step, background of atmospheric neutrons was removed while keeping the remaining backgrounds at the nominal level of 1.0. The bias in NC, as shown in figure B.15 , is -0.017 ± 0.002.

Step 7: In the seventh step, background of external neutron was removed while keeping the remaining backgrounds at the nominal level of 1.0. The bias in NC, as shown in figure B.16 , is -0.016 ± 0.002.

Step 8: Since none of the backgrounds, by itself, reduced the NC bias, all of them were removed except the external neutrons. In the eight step, the fit had CC, ES, $ES_{\mu\tau}$, NC and EX. The NC bias, as shown in figure B.17, is -0.014 ± 0.002 which means that k2pd, k5pd, atmospheric neutrons, $d_2$opd and ncdpd were not causing the bias because they were not included in the fit.

Step 9: For the ninth step, except NC and P_{ee} parameters, all other parameters were fixed. The bias, as shown in figure B.18, did not disappear. It was -0.011 ± 0.003.

Step 10: For the tenth step, only NC was floated; everything else was fixed. The bias in NC disappeared at 0.0029 ± 0.0045, as shown in figure B.19.

The success of the operation lead to the next step which was to float everything except systematics and fixed p_1 and p_2. The promising result, shown in figure B.20, indicate that the bias might arise from distorting the PDFs using the P_{ee} parameters. Hence the distorted 3D PDF from MCMC was compared to the distorted 3D PDF from QSigex.

Figures B.19 and B.20 indicated that p_{ee} parameters might be causing the NC bias. To crack the problem, following steps were undertaken.

1. Calculated the number of CC, ES and $ES_{\mu\tau}$ using the nominal number of NC events and the nominal values for the p_{ee} parameters. The result, as displayed in a table 11.2, shows an excellent agreement between the values calculated by the fit and the Poisson means.

2. Determined the number of CC events in the 45 regular datasets (616 ± 24) and the 45 alternative datasets (610.18 ± 23.56). The biases calculated – $\frac{(616-615.09)}{24} \approx +0.0379$ for the regular dataset and $\frac{(610.18-615.09)}{23.56} \approx -0.208$ for the alternative datasets – indicate that limited statistics (only 45 files) is also a problem.

3. Determined which p_{ee} parameter is causing the problem by floating only one p_{ee} parameter at a time. Figures B.21 to B.24 indicate that p_0 and a_1 might be the cause of the bias in NC.

11.4 Checking the distortions of the Probability Distribution Functions

After determining that the bias arises from the P_{ee} parameters, we needed to ensure that we were calculating the PDFs properly. Two independent codes were developed to distort the 3D PDFs using the nominal values of p_{ee} parameters. The method of distortion was the one used in the MCMC fit. The purpose of the test was to verify that the distortion in the PDF from the MCMC fit was exactly the same as the distortion from QSigex because both MCMC and QSigEx used the same Monte Carlo, that is, the original PDFs are the same between MCMC and QSigEx. Figures B.25 to B.36 show an excellent agreement between the distortions from QSigEx and MCMC fits.

11.5 Checking distortion using the nominal values of p_{ee} to the distortion using values of p_{ee} from the fit

The next step was to determine the fit values of the p_{ee} parameters and compare the 3D PDF distorted by nominal values to the 3D PDF distorted by the fit result listed in table 11.3.

The result, shown in figure B.37, was that removing the last bin did not reduced the NC bias. For the reduced energy range, the Poisson mean number of NC events - for the bias calculation - was not taken as 240.569 but instead calculated as: $NC = 240.569 * \frac{122316}{122375}$ where 122316 is the number of MC events in the energy window of 6 to 12 MeV range and 122375 is the number of events in the full MCMC fit range of 6 to 20 MeV.

This set of tests systematically verified all parts of the MCMC code and showed that the likelihood function was accurately calculated and the MCMC method was working. However, there is still a bias in NC. The bias arises because MCMC implicitly integrates over the P_{ee} distributions when it calculates NC. Our conclusion is that this is fundamental to MCMC methods [94], therefore we do not expect the Markov chain (with Metropolis) to be unbiased– it does not find the point where the likelihood is maximized, but rather scans parameter space and generates steps of equal likelihood. So, for each parameter we get the probability distribution for that parameter, integrated over all the other parameters with appropriate weighting. We did not appreciate this at first, and spent a significant amount of time trying to track down the bias in NC. Ultimately, we found that the bias arises because the mean P_{ee} parameters are not the most likely P_{ee} parameters. That is one of the reason, the bias disappears when the they are constrained using the LETA result.

There exist schemes (See references [83] and [84]) where one can modify MCMC to converge to the point of maximum likelihood, and we considered implementing such a scheme but we decided not to pursue it because such modifications mean that the posterior distributions can no longer be used as projections of the likelihood function for setting the accurate confidence intervals. The price of the decision is that we end up with a biased "best value", but the confidence region should be right.

11.5.1 MCMC versus Maximum Likelihood Estimate (MLE)

The extensive set of tests, outlined in the previous section, demonstrated accuracy of the code and there was no evidence of differences in the likelihood function between the Monte Carlo, the MCMC fitting code, and the datasets.

However, there is a fundamental difference between the way in which one extracts parameter values from the MCMC fit and from a conventional likelihood fit. The usual point estimate in a Markov chain is the mean of the posterior probability distribution, and this generally will not coincide with the maximum likelihood estimate. The mean of the posterior distribution is a better estimator than the maximum likelihood estimator when the posterior is not symmetrical which is the case for various fit parameters in the MCMC fit.

For a given likelihood function $\mathcal{L}$, we find the total likelihood for an entire dataset by taking the product of the likelihoods of the observables of individual events: $\mathcal{L}(\vec{p}) = \prod_{i=1}^{N} \mathcal{L}(\vec{x}_i, \vec{p})$, where $\vec{p}$ refers to the parameters of the fit. The product is over all the events in the data, and $\vec{x}_i$ refers to the observables of an event i. For the standard maximum likelihood fit, a set of M parameters $\vec{p}$ are found that maximizes $\mathcal{L}$ – the full likelihood. With the Markov Chain, we typically obtain the fit values from single dimensional distributions f_j for the jth parameter that are obtained by integrating/marginalizing the likelihood distribution over the other parameters:

$$f_j(p_j) = \int dp_1 dp_2 \ldots dp_{j-1} dp_{j+1} \ldots dp_M \, \mathcal{L}(\vec{p}) \tag{11.2}$$

The value of the parameter can be extracted in several different ways:

1. The mean: $\langle p_j \rangle = \frac{\int dp_j p_j f_j(p_j)}{\int dp_j f_j(p_j)}$
2. The fit mean: fit a Gaussian function with mean $p_{j,fit}$ to f_j
3. The peak: find the value $p_{j,peak}$ which corresponds to the maximum of f_j.

None of these techniques correspond exactly to finding the point in M dimensional space that maximizes likelihood; since the maximum likelihood technique should be non-biased; this means, in general that the Markov Chain will be biased. However, it should be pointed out that, in general, the confidence intervals that are extracted from the Markov Chain should be exactly correct, though it is not straight-forward to find the confidence limits from the maximum likelihood fit. To test whether this is the explanation for our bias, we calculated the number of events for CC, ES and $ES_{\mu\tau}$ without/with floating any systematic uncertainties. For each case, the computation was performed twice; first using the p_{ee} parameters and ^{8}B scale from the **mean** of the posterior distributions and then from the **peak** of the posterior distributions. The set of values used in the computation of the derived parameters (first four rows) and the derived parameters (next four rows) are shown in tables 11.4 and 11.5.

This is a strong evidence that it is difference between the peak and the mean p_{ee} parameters that is causing the bias. When we find the number of CC events, we integrate over the P_{ee} parameters and therefore are using the biased, mean CC values rather than the unbiased peak CC value. The natural techniques involve integrating over them, which will essentially weigh the p_{ee} peak parameters with their likelihood; corresponding very closely to the way in which the mean, rather than the peak, is calculated. If instead of finding the peak from a 1D marginal likelihood function, we use a 2D marginal likelihood function which takes into account the correlation of ^{8}B flux with the P_{ee} parameter, as shown in figure 9.1, the bias on ^{8}B goes down from -0.01208$\pm$0.0034 to -0.0077$\pm$0.0031 for the regular dataset. For the alternate dataset, the bias reduced from -0.0169$\pm$0.0039 to -0.0056$\pm$0.0034. Reduction in bias when the maximum is from 2 Dimensional marginal likelihood (^{8}B and p_0) instead of 1 dimensional marginal likelihood (^{8}B flux) confirms that the hypothesis that the bias in ^{8}B is due to the marginalization over other parameters in the MCMC method.

Peak and mean of a posterior distribution (belonging to a parameter) from each of the 45 dataset are plotted into histograms and the histograms are fitted to a Gaussian function. The results of the fit (μ, σ) are reported in the first four rows of the tables 11.4 and 11.5 (Column 2/3 reports the fit result from the distribution of the means/peaks). Using either the mean or peak, the number of CC, ES and $ES_{\mu\tau}$ (next three rows in the tables) are computed from a procedure which is similar to the one used in the MCMC fit. The last row in the table 11.4 shows the calculation of CC bias from using the mean or the peak of the posterior distribution. The CC bias is around 50% smaller using the peak than using the mean even though the peak of a posterior distribution is not exactly equivalent to the peak from the maximum likelihood method. The consensus of the effort was that CC is unbiased using the p_{ee} parameters from **peak** but not when the p_{ee} parameters are from the **mean**.

11.6 Comparing Result from QSigEx and MCMC fit

The result of MCMC is compared to QSigEx in this section. MCMC is the signal extraction code run at the University of Alberta by the author of this thesis while QSigEx is the extraction code run by Pierre-Luc Drouin at the University of Carleton. There are four cases to compare; 1) data without floating systematic, 2) alternate data without floating systematic, 3) data with floating systematic and 4) alternate data with floating systematic. For each case 6 fit parameters are compared dataset by dataset ; ^{8}B flux, P_{ee} parameters (p_0, p_1 and p_2) and day-night parameters (a_0, a_1).

11.7 Calculating confidence intervals of MCMC

The upper and lower confidence limits are determined such that confidence limits at lower and upper limits are equal and the integral between them is 68%. The difference between the upper and lower limit is the range quoted for the $\pm\sigma$ errors on the fit.

This section shows the $\pm\sigma$ confidence intervals of MCMC result for all four cases mentioned in section 11.6. Figures B.50 to B.52 shows confidence intervals of various parameters for the regular datasets and figures B.53 to B.55 for the alternate datasets.

Figures B.56 to B.61 display the result when eight systematics were floated in the MCMC fit: first for the regular datasets and then for the alternate datasets.

11.8 Quantitative Comparison between MCMC and QSigEx

Tables 11.6 and 11.7 list a quantitative comparison between MCMC and QSigEx. How the quantities in the tables are computed is described in section 10.4.

Tables B.1 to B.6 show the comparison of the best-fit values of 6 parameters – extracted from fitting the regular data – between QSigEX and MCMC in tabular forms. Eight systematic uncertainties were floated for this fit. Column 3 (from equation (11.3)) has the mean of the 68% confidence intervals in MCMC along with its uncertainty (from equation (11.4)) and column 4 has the mean±RMS from the marginalized likelihood (ML) after taking out the burn-in period. The confidence intervals were calculated such that the area between upper level U and lower level L is 68% and bin content of the marginalized likelihood (after taking out the burn-in period) at L and U are equal.

$$\text{Best-fit} = (U + L)/2 \tag{11.3}$$

$$\delta\text{Best-fit} = (U - L)/2 \tag{11.4}$$

Test	**NC Bias**	**Figure Number**
Step 1	0.005 ± 0.004	B.1
Step 2	0.007 ± 0.004	B.2
Step 3	0.003 ± 0.003	B.3
Step 4	0.002 ± 0.003	B.4
Step 5	0.005 ± 0.003	B.5
Step 6	0.003 ± 0.003	B.6
Step 7	0.003 ± 0.003	B.7
Step 8	0.004 ± 0.003	B.8
Step 9	0.003 ± 0.003	B.9
Floating p_0	(-0.0037 ± 0.0024)	B.21
Floating p_1	(-0.0028 ± 0.0029)	B.22
Floating p_2	(0.000119 ± 0.003)	B.23
Floating a_0	(0.00026 ± 0.004)	Top B.24
Floating a_1	(-0.005 ± 0.003)	Bottom B.24
Removing NCDPD	-0.016 ± 0.002	B.10
Removing $k_5$pd	-0.015 ± 0.002	B.11
Removing $k_2$pd	-0.014 ± 0.003	B.12
Removing $d_2$opd	-0.015 ± 0.003	B.13
Removing hep	-0.016 ± 0.003	B.14
Removing Atmospheric neutrons	-0.017 ± 0.002	B.15
Removing External neutrons	-0.016 ± 0.002	B.16
Only CC, ES and NC floating	-0.014 ± 0.002	B.17
Only NC and P_{ee} parameters floating	-0.011 ± 0.003	B.18
Fixed p_{ee}	$+0.0029 \pm 0.0045$	B.19
Fixed p_1 and p_2	-0.0035 ± 0.0034	B.20

Table 11.1: List of the tests undertaken to resolve the source of NC bias along with the bias$\pm\delta$bias. For each test, the table also points to the figure where the result is illustrated.

Parameter	**Calculated**	**Poisson Mean**
CC	615.26	615.09
ES	53.87	53.87
$ES_{\mu\tau}$	16.57	16.57

Table 11.2: Comparison of the calculated values to the Poisson means. From the nominal values of ^8B flux and the parameters of survival probability, the number of CC, ES and $ES_{\mu\tau}$ were calculated. For example, the number of CC day and night events were computed using equations (4.63) and (4.64). The excellent fit between column 1 and column 2 indicates that the equations were correctly applied in the MCMC fit.

Parameter	**Fit Result**	**Nominal Value**
p_0	0.3343 ± 0.0412	0.325
p_1	-0.006 ± 0.04	-0.00888
p_2	-0.0008 ± 0.0192	0.00122
a_0	0.0298 ± 0.088	0.028
a_1	0.027 ± 0.080	0.00478

Table 11.3: Fit result of the p_{ee} parameters from the 45 datasets as compared to the nominal values.

Without Floating Systematic Uncertainties			
Parameter	**Fit Result Using Mean**	**Fit Result Using Peak**	**Nominal Value**
^{8}B Scale	0.9811 ± 0.0158	0.9815 ± 0.0221	1.0
p_0	0.3435 ± 0.0436	0.329 ± 0.042	0.325
p_1	-0.00259 ± 0.0478	-0.00839 ± 0.02888	-0.00888
p_2	-0.0049 ± 0.0283	0.00066 ± 0.01581	0.00122
CC	625.44	620.36	615.09
ES	53.42	54.21	53.87
$ES_{\nu\mu}$	16.61	16.55	16.57
CC Bias	$\frac{(625.44-615.09)}{615.09} \simeq 0.0168$	$\frac{(620.36-615.09}{615.09}) \simeq 8.57e-3$	

Table 11.4: Comparing mean versus the peak for the 45 datasets. Number of steps 750,000 with 400,000 steps removed as the burn-in period. The fit has 16 fit parameters and includes all the backgrounds (ex, d2opd, ncdpd, k2pd, k5pd and atmospheric neutrons). In this fit the day-night asymmetries for the external neutrons and d2opd backgrounds were also floated besides neutron detection efficiencies – both for the PMTs and the NCDs.

With Floating Systematic Uncertainties			
Parameter	**Fit Result Using Mean**	**Fit Result Using Peak**	**Nominal Value**
^{8}B Scale	0.9742 ± 0.0335	0.9743 ± 0.0494	1.0
p_0	0.3482 ± 0.056	0.3367 ± 0.059	0.325
p_1	-0.00737 ± 0.0347	-0.005 ± 0.049	-0.00888
p_2	-0.0035 ± 0.0262	-0.0039 ± 0.0327	0.00122
CC	636.9	615.07	615.09
ES	55.09	52.95	53.87
$ES_{\nu\mu}$	16.65	16.55	16.57
CC Bias	0.035	-3.25e-5	

Table 11.5: Comparing mean versus the peak for the 45 datasets. Number of steps were 30,000 and 10,000 steps were removed as burn-in. This fit floats eight systematics. The fit has 16+8 fit parameters and includes all the backgrounds (ex, d2opd, ncdpd, k2pd, k5pd and atmospheric neutrons). In this fit the day-night asymmetries for the external neutrons and d2opd backgrounds were also floated besides neutron detection efficiencies – both for the PMTs and the NCDs.

Parameter	**Generated**	**QSigEx**	**MCMC**	**Mean Shift**	**Mean Spread**	**Correlation**
^{8}B Scale	1.0	1.0020±0.0158	0.9793±0.0173	-0.2440±0.0131	0.0879	0.8822
p_0	0.325	0.3211±0.0353	0.3400±0.0451	0.3420±0.0371	0.2491	0.9476
p_1	-0.00888	-0.0132±0.0273	-0.0021±0.0465	0.2273±0.0644	0.4322	0.8843
p_2	0.00122	0.0022±0.0139	-0.0040±0.0277	-0.1941±0.0753	0.5048	0.8755
a_0	0.028	0.0336±0.0882	0.0261±0.0883	-0.0803±0.0295	0.1979	0.9736
a_1	0.0048	0.0300±0.0872	0.0225±0.0850	-0.0744±0.0296	0.1987	0.9706

Table 11.6: Comparing result between QSigEx and MCMC fit for the regular datasets. Eight systematics were floated. The best-fit and its uncertainty from MCMC is the mean and RMS of the posterior distribution (marginalized likelihood). Number of data files is 45 and each file has 420,000 steps with burn-in period as 50,000 steps.

Parameter	Generated	QSigEx	MCMC	Mean Shift	Mean Spread	Correlation
^{8}B Scale	1.0	1.0040±0.0219	0.9809±0.0204	-0.2776±0.0606	0.4069	0.5887
p_0	0.325	0.3267±0.0464	0.3467±0.0574	0.3534±0.0383	0.2569	0.9642
p_1	-0.00888	-0.0086±0.0278	-0.0012±0.0404	0.0995±0.0943	0.6329	0.7170
p_2	0.00122	0.0000±0.0166	-0.0065±0.0315	-0.1676±0.0834	0.5597	0.7880
a_0	0.028	0.0163±0.0909	0.0006±0.0908	-0.1797±0.0446	0.2994	0.9802
a_1	0.0048	-0.0005±0.0898	0.0056±0.0847	0.0793±0.0690	0.4630	0.9420

Table 11.7: Comparing result between QSigEx and MCMC fit for the alternate datasets. Eight systematics were floated. The best-fit and its uncertainty from MCMC is the mean and RMS of the posterior distribution (marginalized likelihood). Number of data files is 45 and each file has 344,000 steps with burn-in period as 80,000 steps.

11.9 Summary

After performing an exhaustive number of tests, during the fitting of simulated data ensemble (No LETA constraints applied), the conclusion is that the code, run under the following conditions (not floating systematics and floating 8 systematics) does have some small biases, which we have tracked down to those expected in a MCMC Metropolis-Hastings fit for this likelihood function and these data. However, the biases are small compared to a fit uncertainty, and the results, agree well with QSigEx. The ^{8}B flux shows a bias of -0.01208$\pm$0.0034 for the regular datasets and -0.0169 ± 0.0039 for the alternate datasets when 8 systematics are floated.

Chapter 12

MCMC Fit on 1/3 of the Real Dataset

Before performing a fit on the full data, the code was tested on the third of the data. This allows detailed comparison using real data while maintaining some blindness. The result from the fit, applied on the full data, will contribute to the final analysis published from SNO, for completeness sake, four new systematics were added. Since 4 new systematics were introduced, an additional test was performed to check the code. This chapter presents the result of the test along with the results of the fit. There are 45 parameters in the fit; 3 parameters – the number of events for hep CC, hep ES and hep NC – are fixed and the number of events for CC, ES and $ES_{\mu\tau}$ are calculated using ^{8}B flux and the application of the survival probability equation. Two parameters f2e^{pmt} and f2e^{ncd} are flux-to-event conversion factors for ^{8}B flux for the Čerenkov data and data from the NCDs respectively. There are 675,000 steps in the fit and the burn-in period is 75,000. PSA constraint is 375.1±28.7.

12.1 Checking Important Systematic Uncertainties

The four new systematic uncertainties are: (1) energy-dependent fiducial volume, (2) Z scale, (3) energy non-linearity and (4) uncertainty in the shape of

^{8}B neutrino energy spectrum (Winter uncertainty). Because these systematics cause very small effects, it is difficult to test the code so it was decided to perform the tests where all other parameters were fixed to ensure that the likelihood function follows the shape of the constraint applied on the systematic being tested. That is the reason that the log likelihood versus the uncertainty is a parabola with a centroid within one σ of the width of the constraint. For example, the top plots in figures 12.1 and 12.2 show that the centroids (0.098 and -0.00118) of Winter uncertainty and energy-dependent fiducial volume uncertainty are within the widths of their constraints: 0 ± 1 and $0^{+0.0087}_{-0.0067}$. Several other systematics were also tested, for instance, vertex scale, energy resolution and energy scale. These are plotted in figures 12.3, 12.6 and 12.7. Overall 12 systematics were checked, but only the important ones are described in this chapter. Table 12.1 lists name of the systematic uncertainty, the constraint applied, $\mu \pm \sigma$ from the Gaussian fit on the posterior distribution and the corresponding figure number showing the posterior distribution and the parabola of the constraint.

Comparing the bottom plot in figure 12.1 to the bottom plot in figure 12.4, we see there are bands in the former but not in the latter. The reason for a band structure is that the systematic uncertainty being tested (Winter uncertainty) modifies the number of events (section 4.11 describes the role of Winter uncertainty in calculation of number of events) which brings in additional constraints in the fit since the number of background events are also constrained (section 4.10.6). The application of additional constraints, besides the one on the Winter uncertainty that we were testing, causes formation of band structures. The log-likelihood plots of systematic uncertainties which do not affect the number of events, for instance figure 12.6 shows a clean and distinct parabola.

The log likelihood function to test the systematic uncertainty is:

$$
\begin{aligned}
-\log \mathcal{L} = \sum_{i=1}^{2m} N_i - \sum_{d=1}^{N} \log \big(\sum_{i=1}^{2m} (N_i) F_i(\vec{x}_d, \vec{P}) \big) \\
+\frac{1}{2}\sum_{i}(\frac{p_i - \bar{p}_i}{\sigma_{p_i}})^2 + \frac{1}{2}\sum_{i=0}^{2}\sum_{j=0}^{2}(b_i^{xy} - b_i^{\bar{x}y})(b_j^{xy} - b_j^{\bar{x}y})(V_{b^{xy}}^{-1})_{ij} \\
+\frac{1}{2}\sum_{i=0}^{1}\sum_{j=0}^{1}(b_i^{z} - \bar{b_i^{z}})(b_j^{z} - \bar{b_j^{z}})(V_{b^{z}}^{-1})_{ij}
\end{aligned}
\tag{12.1}
$$

where $\mathrm{F}_i(\vec{x}_d, \vec{P})$ is the probability density function, for the class i, giving the probability of observing an event d with observables $\vec{x}_d$ and the current values of the fit parameters $\vec{P}$, $N_1, N_2, \ldots, N_m$ are the number of events for **m=13** event classes and **i** goes from 1 to **N** data entries. In the likelihood equation (12.1), p_i, $\bar{p}_i$ and σ_{p_i} represent the current value of the PMT systematic parameter i in the MCMC fit, its mean and constraint width respectively. The next two terms are calculation of the constraint for the systematic uncertainties that are correlated and the correlation matrices between the correlated parameters is represented by V.

For asymmetric uncertainties $(\mu + |\sigma_+| + |\sigma_-|)$, σ_{p_i} applied in the log likelihood function 12.1 and for fitting the Gaussian function on posterior distributions (shown in figures 12.2 and 12.4) is determined as:

$$
\sigma_{p_i}(x) \equiv \begin{cases} \sigma_- & \text{if } x < \mu, \\ \sigma_+ & \text{if } x \geq \mu. \end{cases}
$$

The legends in the plots (figures 12.2 and 12.4) only shows one σ, for instance, the uncertainty on energy-dependent fiducial volume (figure 12.2) is $0^{+0.0087}_{-0.0067}$ but the legend shows σ as +0.0087.

12.2 Overview of the Result

Table 12.2 gives an overview of the result; the first column lists the name of parameter, the second column lists the best fit and its uncertainty, the

Uncertainty	Constraint	$\mu \pm \sigma$	Figure #
Winter	0 ± 1	0.098 ± 0.892	12.1
Energy-dependent fiducial volume	$0^{+0.0088}_{-0.0067}$	-0.001 ± 0.009	12.2
Vertex scale	$0^{+0.0029}_{-0.0077}$	0.001 ± 0.003	12.3
Z scale	$0^{+0.0015}_{-0.0012}$	-0.00003 ± 0.0015	12.4
Energy non-linearity	0 ± 0.0069	00.0018 ± 0.0063	12.5
Energy resolution	0.0119 ± 0.0104	0.013 ± 0.010	12.6
Energy scale	0 ± 0.0081	0.0002 ± 0.006	12.7

Table 12.1: Systematic uncertainty, the constraint applied, and $\mu \pm \sigma$ from fitting the posterior distribution to a Gaussian function. Last column points to the figure number corresponding to the systematic uncertainty.

third column shows the mean of the posterior distribution and the last column displays the difference between the mean and the peak in terms of the RMS of the posterior distribution. The good news is that for all the parameters, the difference between the mean and the peak in terms of the uncertainty is below 0.5σ (decided by the SNO collaboration)[1] which means that the posterior distribution from the MCMC is not very asymmetric.

12.2.1 Autocorrelation Plots

Figures 12.8 to 12.10 shows the application of autocorrelation function (equation (12.2)) to the MCMC fit of the 1/3 data. These plots indicate that step sizes selected for all the fit parameters were good otherwise for parameters whose step sizes are too narrow there is a lot of fluctuations going on till the end of the fit. Selecting step sizes is very challenging in a MCMC fit because

[1] The statistical uncertainty associated with the calculation of the peak, assuming that the posterior distribution has Gaussian distribution, is $1/\sqrt{N_{peak}}$ where N_{peak} is the number of entries in the bin where the peak is located. The uncertainty associated with the calculation of the mean is $1/\sqrt{N}$ where N is the number of entries in the posterior distribution. Since the uncertainty of the peak is higher than the uncertainty on the mean but both are extracted from fitting the same data, hence they should be within 0.5σ where σ is the RMS of the posterior distribution.

the consequence of very narrow step sizes is that the target distribution is not explored sufficiently although very broad step sizes result in a poor acceptance. Hence it takes a lot of trials and errors to finally select step sizes that balance the exploration of the target distribution as well give a good acceptance. There are 675,000 steps in the MCMC fit but to clearly show the drop of autocorrelation coefficient to zero within 10,000 steps, autocorrelation plots are displayed for only 375,000 steps.

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum\limits_{i=1}^{N}(Y_i - \bar{Y})^2} \text{ from [85]} \tag{12.2}$$

where N is the number of steps in the MCMC chain, $\bar{Y} = \sum\limits_{i=1}^{N} \frac{Y_i}{N}$, Y_i and Y_{i+k} are measurements at step i and $i+k$. Since Y_i lags behind Y_{i+k} by k steps, k is known as lag. In figures 12.8 to 12.10, lag is shown on the x axis and r_k – autocorrelation coefficient – is shown on the y axis.

12.3 Convergence Tests

To make sure that the fit has converged, MCMC result was randomly divided into two halves after taking out the burn-in period. Using a random number, 5,000 steps were added to one half or the other half. The difference in the mean of the posterior distribution from both halves (m_1 and m_2) were compared using the equations:

$$\frac{(m_2 - m_2) \times 2}{(\sigma_2 + \sigma_1)} \tag{12.3}$$

$$\sigma_0 = \sqrt{(m^2 + \sigma^2)} \tag{12.4}$$

where σ_1 and σ_2 are the standard deviations of the posterior distributions from both halves, m and σ are the mean and RMS around the mean (m) of the distributions shown in figure 12.11 and 12.12. The difference between the means (m_1 and m_2) in terms of the average uncertainty (equation (12.3)), calculated around zero (equation (12.4) instead of the average mean, should be less than $\frac{\sqrt{2}}{6}$. Table 12.3 shows that the convergence test, performed on the 6 parameters, show that the fit passed the convergence test.

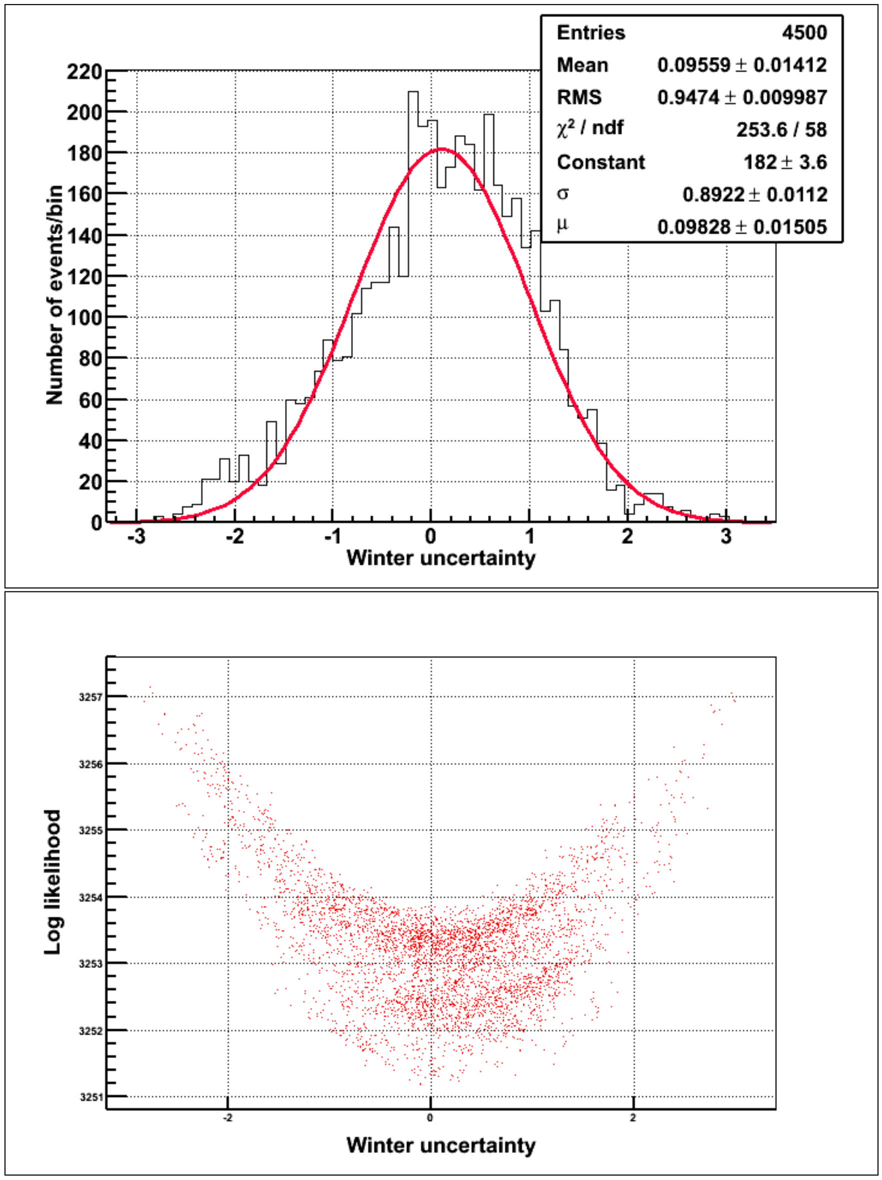


Figure 12.1: The top plot is the posterior distribution of the Winter uncertainty and the bottom plot is log likelihood versus the Winter uncertainty. The constraint applied was 0 ± 1. The constant term, from the Gaussian fit, only produces a constant offset in the negative log likelihood function and therefore has no impact on the best fit estimates from the posterior distributions.

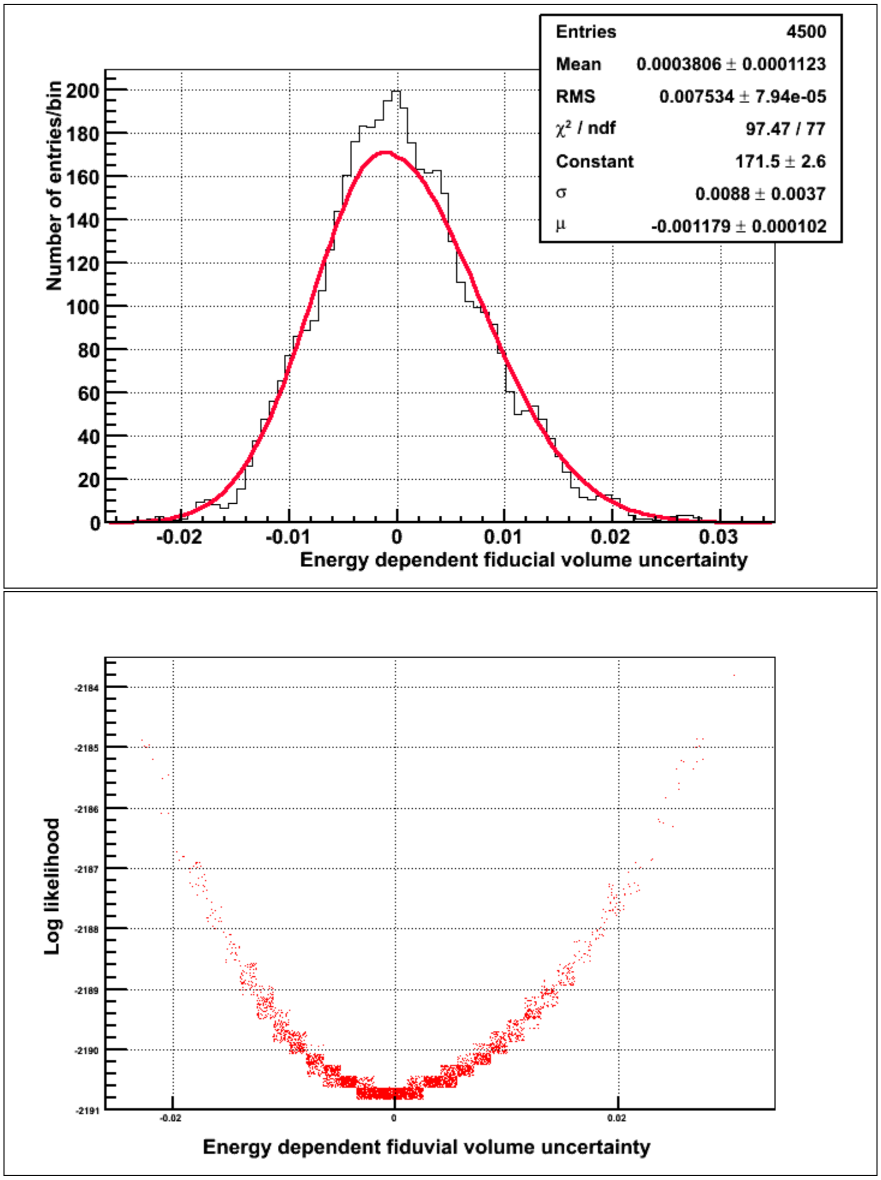


Figure 12.2: The top plot is the posterior distribution of energy-dependent fiducial volume and the bottom plot is log likelihood versus the energy-dependent fiducial volume. The constraint applied is $0^{+0.0087}_{-0.0067}$ and the centroid, -0.001179, is $0.1355\sigma_+$ and $0.1760\sigma_-$ away from zero.

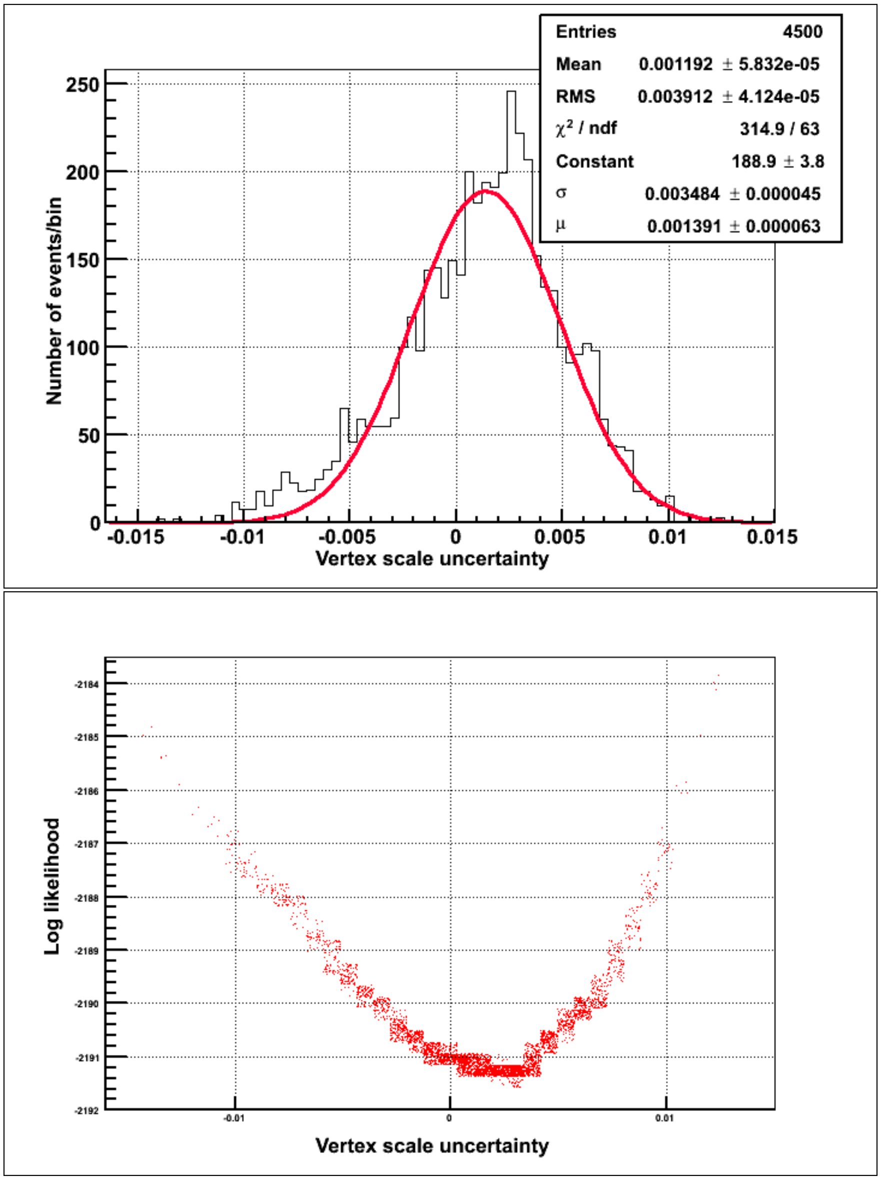


Figure 12.3: The top plot is the posterior distribution of vertex scale and the bottom plot is log likelihood versus the vertex scale. The constraint applied is $0^{+0.0029}_{-0.0077}$ hence the centroid, 0.001391, is $0.4797\sigma_+$ and $0.1806\sigma_-$ away from zero.

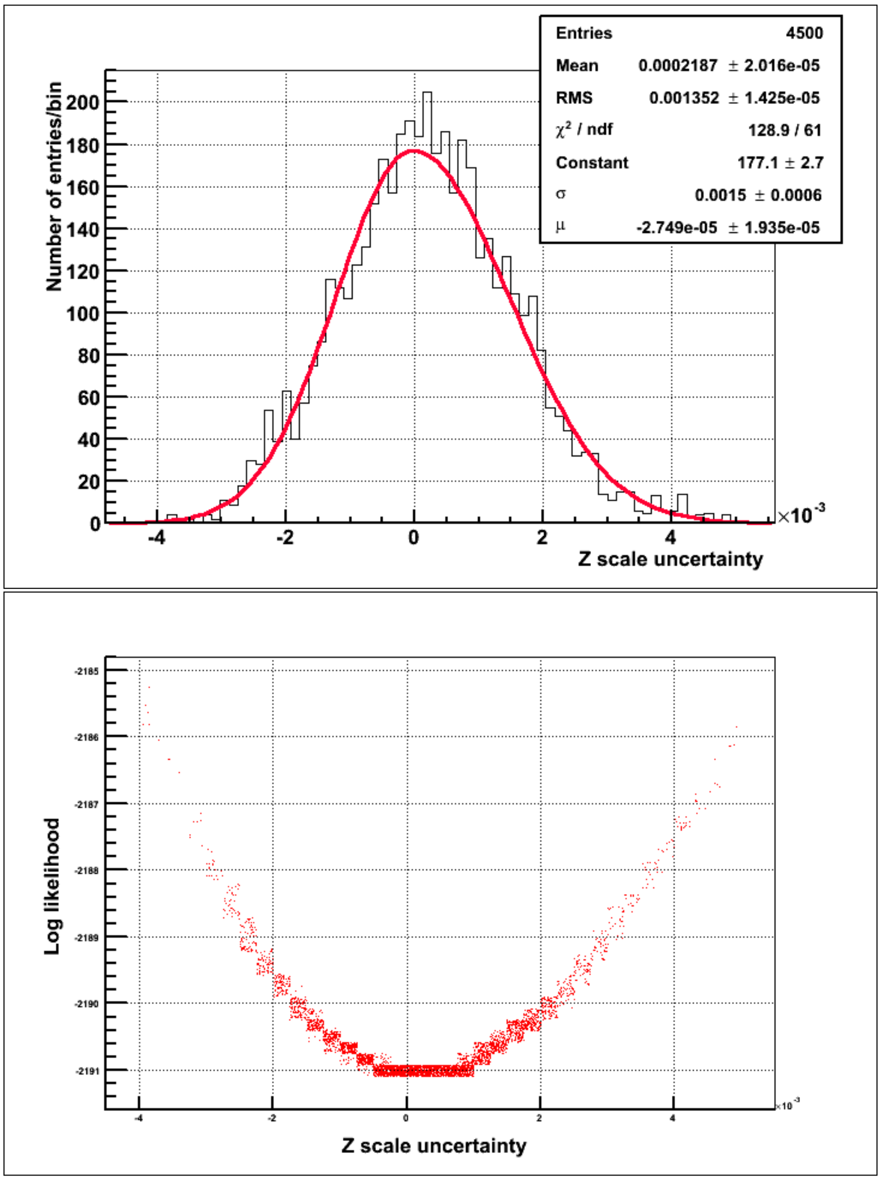


Figure 12.4: The top plot is the posterior distribution of Z scale and the bottom plot is log likelihood versus the Z scale. The constraint applied is $0^{+0.0015}_{-0.0012}$ hence the centroid, -2.749e-5, is $-0.018\sigma_+$ and $0.023\sigma_-$ away from zero.

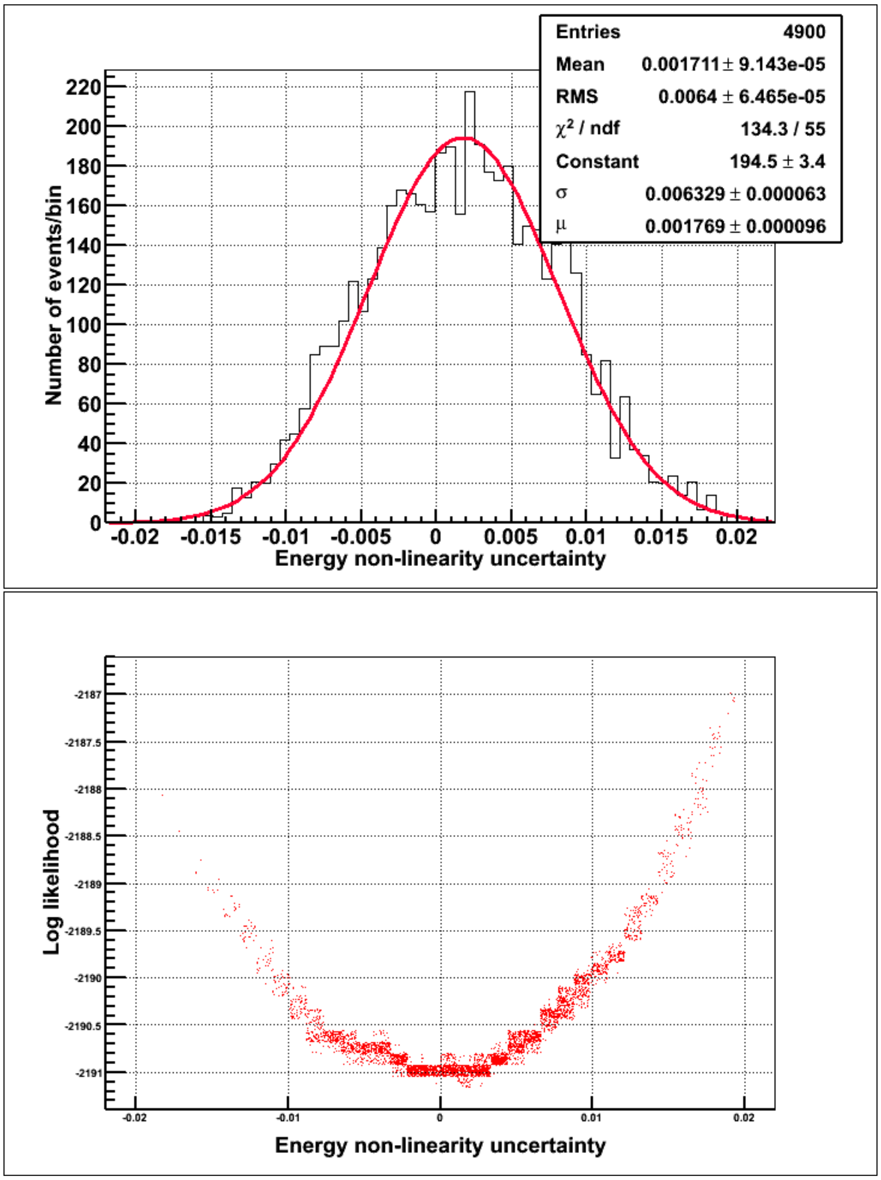


Figure 12.5: The top plot is the posterior distribution of energy non-linearity and the bottom plot is log likelihood versus the energy non-linearity. The constraint applied is 0 ± 0.0069 and the centroid from the Gaussian fit, $\mu = 0.001769$, is 0.256σ away from zero where $\sigma = 0.0069$ is width of the constraint.

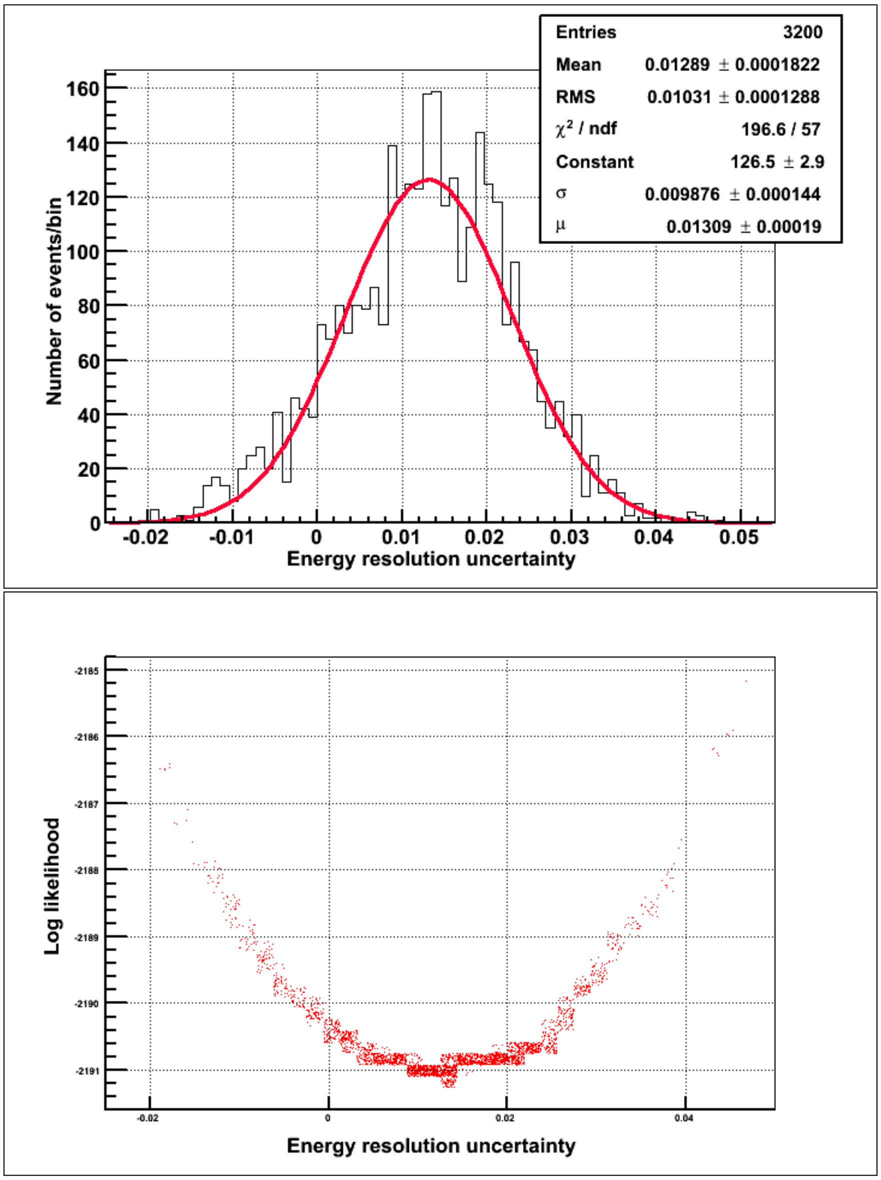


Figure 12.6: The top plot is the posterior distribution of energy resolution and the bottom plot is log likelihood versus the energy resolution. The constraint on energy resolution is 0.0119 ± 0.0104 hence the centroid $\mu = 0.0131$ is 0.1154σ away from 0.0119.

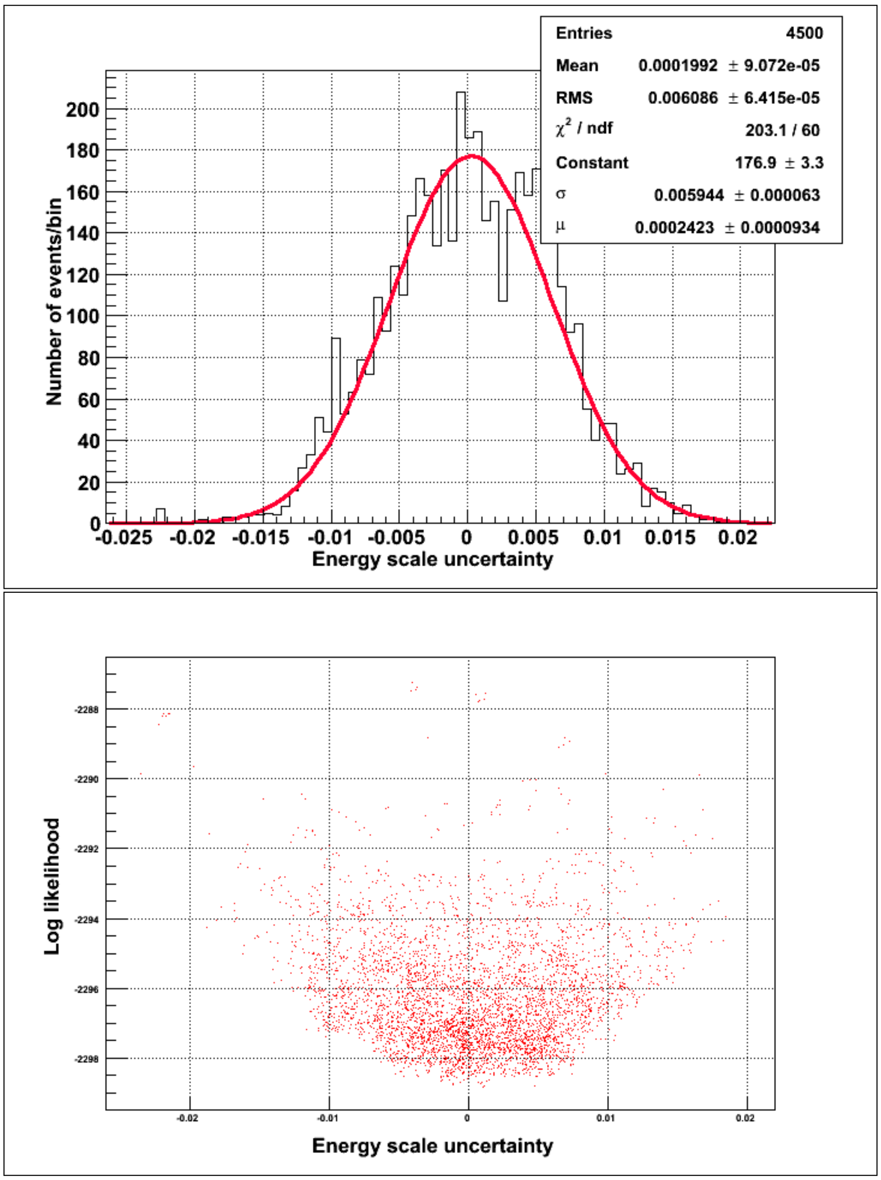


Figure 12.7: The top plot is the posterior distribution of energy scale and the bottom plot is log likelihood versus the energy scale. The constraint applied is 0 ± 0.0081.

Parameter	Peak±RMS	Mean	$\frac{\text{(Peak-Mean)}}{\text{RMS}}$
^{8}B scale	0.912727±0.0903166	0.914362	-0.0180996
p_0	0.335556±0.0709244	0.370174	-0.488102
p_1	-0.0136726±0.0448998	-0.00700323	-0.148538
p_2	-0.0198333±0.0432173	-0.0387461	0.437619
a_0	0.0788554±0.108772	0.0644824	0.132139
a_1	0.147683±0.147322	0.127117	0.139599
ϵ_{pmt}	0.466±0.0129844	0.467231	-0.0948358
CC	558.617±29.4819	569.665	-0.374767
Atmospheric neutrons	1.05±0.198532	0.99761	0.263886
K2PD	0.9925±0.164456	0.993089	-0.0035837
K5PD	1.00011±0.118572	1.00417	-0.0342191
NCDPD	1.04571±0.34213	1.00257	0.126108
ES	45.8297±4.71439	47.0332	-0.255299
EX	0.969595±0.472556	1.0862	-0.246749
EX day-night asymmetry	-0.0335729±0.0114606	-0.0342751	0.061268
D_2OPD	1.02484±0.148729	1.00733	0.117704
D_2O day-night asymmetry	-0.058427±0.116307	-0.0190298	-0.338735
$ES_{\mu\tau}$	18.7727±5.13597	20.0146	-0.24179
ϵ_{ncd}	1.77548±0.040568	1.76653	0.220554
Cos θ resolution direction	-0.00665±0.0692353	-0.00919638	0.0367786
Cos θ resolution	0.025±0.109935	0.0153701	0.0875961
Energy scale	0.00518±0.00852786	0.00418677	0.116469
Energy resolution direction	-0.00045±0.0120072	0.000777723	-0.102249
Energy resolution	0.01253±0.0105886	0.0133717	-0.0794951
X shift	-0.365±4.10277	-0.250553	-0.0278951
Vertex scale	-0.00146512±0.00557221	-0.00369108	0.399477
Y shift	-0.184524±3.68196	-0.40452	0.0597497
Z shift	-1.085±4.07116	-0.18085	-0.222086
XY resolution -constant term	0.076225±0.0285262	0.0695889	0.23263
XY resolution -linear term	-6.675e-05±6.1534e-05	-5.94605e-05	-0.118463
XY resolution -quadratic term	3.745e-07±1.98487e-07	3.6701e-07	0.0377378
Z resolution -constant term	0.06855±0.0286124	0.0707189	-0.0758041
Z resolution -quadratic	0.0001217±8.27124e-05	0.000112913	0.106233
Vertex scale diurnal	-8.1e-05±0.00150131	-4.93177e-05	-0.021103
Vertex scale direct	0.000505±0.00177122	-8.25761e-05	0.331735
Energy scale diurnal	-5.58511e-05±0.00387483	3.60054e-05	-0.0237059
Energy scale direct	-0.00203529±0.0101849	0.000329195	-0.232156
Energy scale correlated	0.00153963±0.00417378	0.00139816	0.0338959
Fiducial volume	0.00119474±0.00764633	0.00159754	-0.0526795
Energy non-linearity	0.00232±0.00715378	0.000914944	0.196407
Z scale	0.000447727±0.0014067	0.000247992	0.141988
Winter uncertainty	-0.25±1.0003	-0.0226684	-0.227264

Table 12.2: A listing of the peak (best fit) and its uncertainty (RMS of the posterior distribution) for the 42 parameters involved in the 1/3 data fit. Besides that, the table also lists the mean and the difference between the mean and the peak in units of the uncertainty.

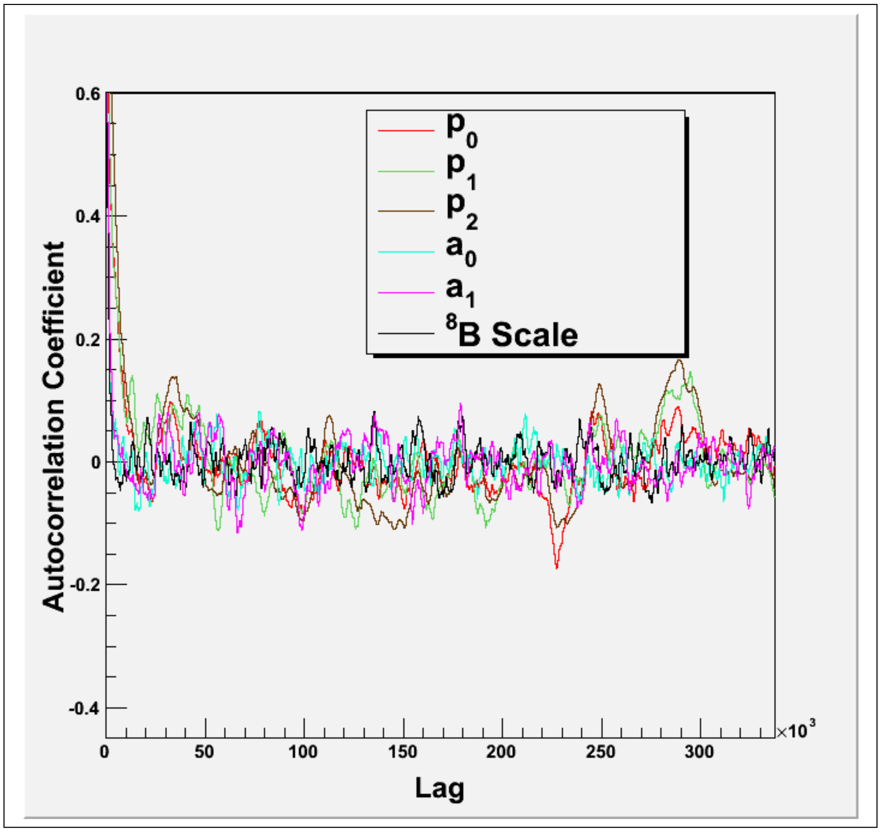


Figure 12.8: Autocorrelation plots showing the autocorrelation coefficient versus lag of ^{8}B scale and the P_{ee} parameters of the fit. There are 675,000 steps in the MCMC fit but to see the drop of autocorrelation coefficient to zero not all steps are shown in the figure.

Parameter	**σ_0 RMS around zero**
^{8}B scale	0.10129
p_0	0.14654
p_1	0.17582
p_2	0.18521
a_0	0.10859
a_1	0.12395

Table 12.3: Table lists RMS around zero (σ_0) for the 6 parameters.

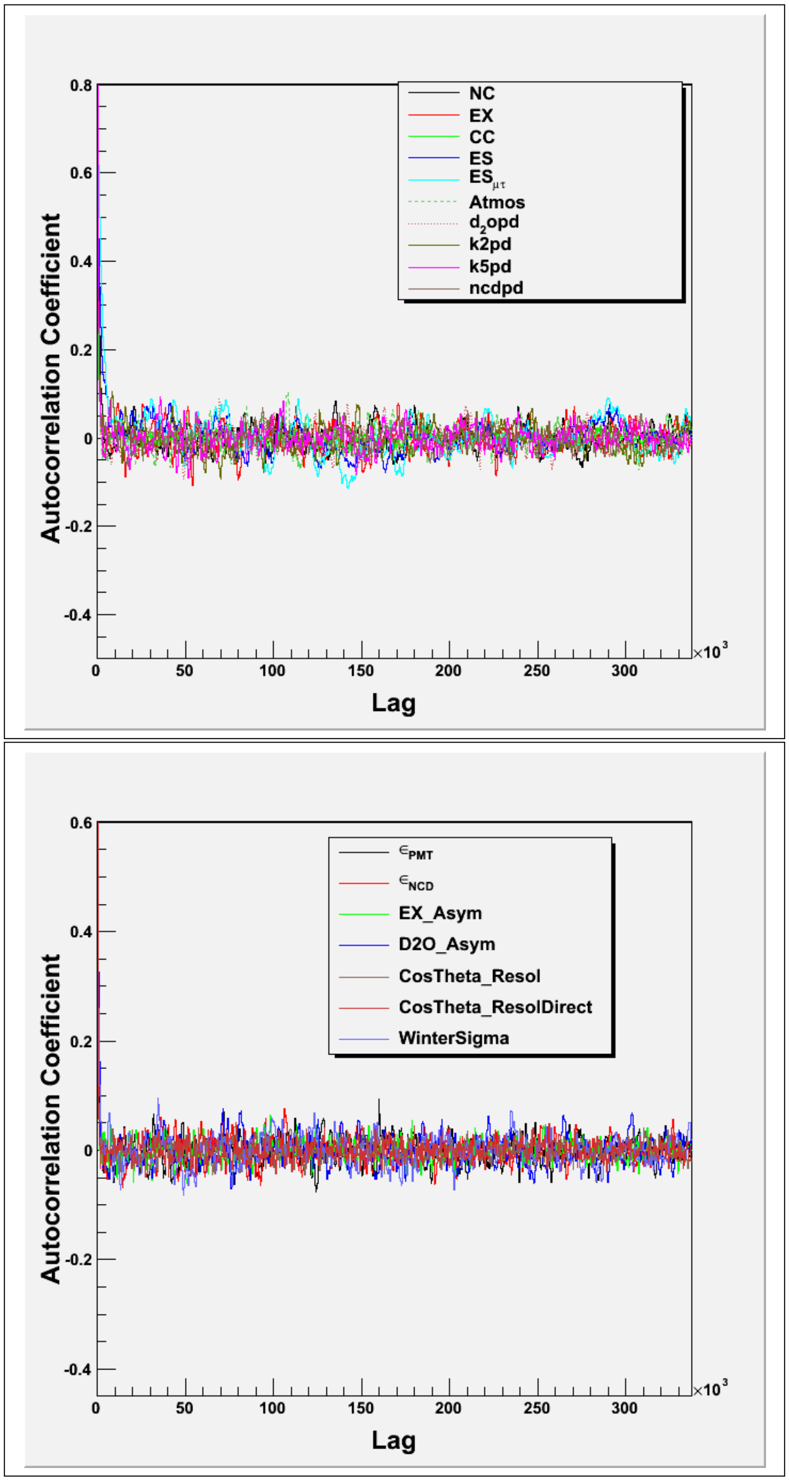


Figure 12.9: Autocorrelation plot showing autocorrelation coefficient versus lag: (top) for signals and backgrounds (bottom) miscellaneous parameters in the fit.

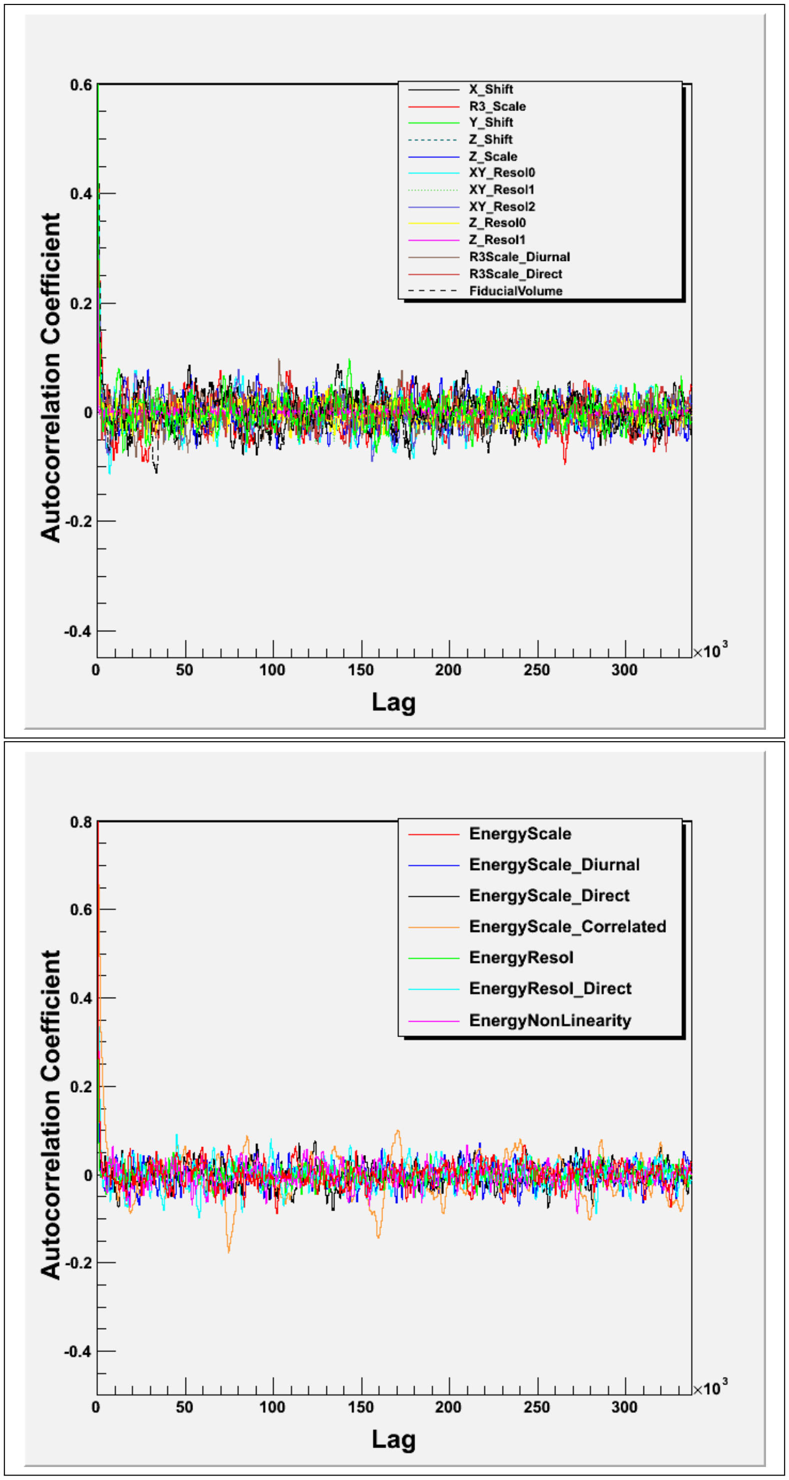


Figure 12.10: Autocorrelation coefficient versus lag for systematic uncertainties involved in the reconstruction of vertex and energy.

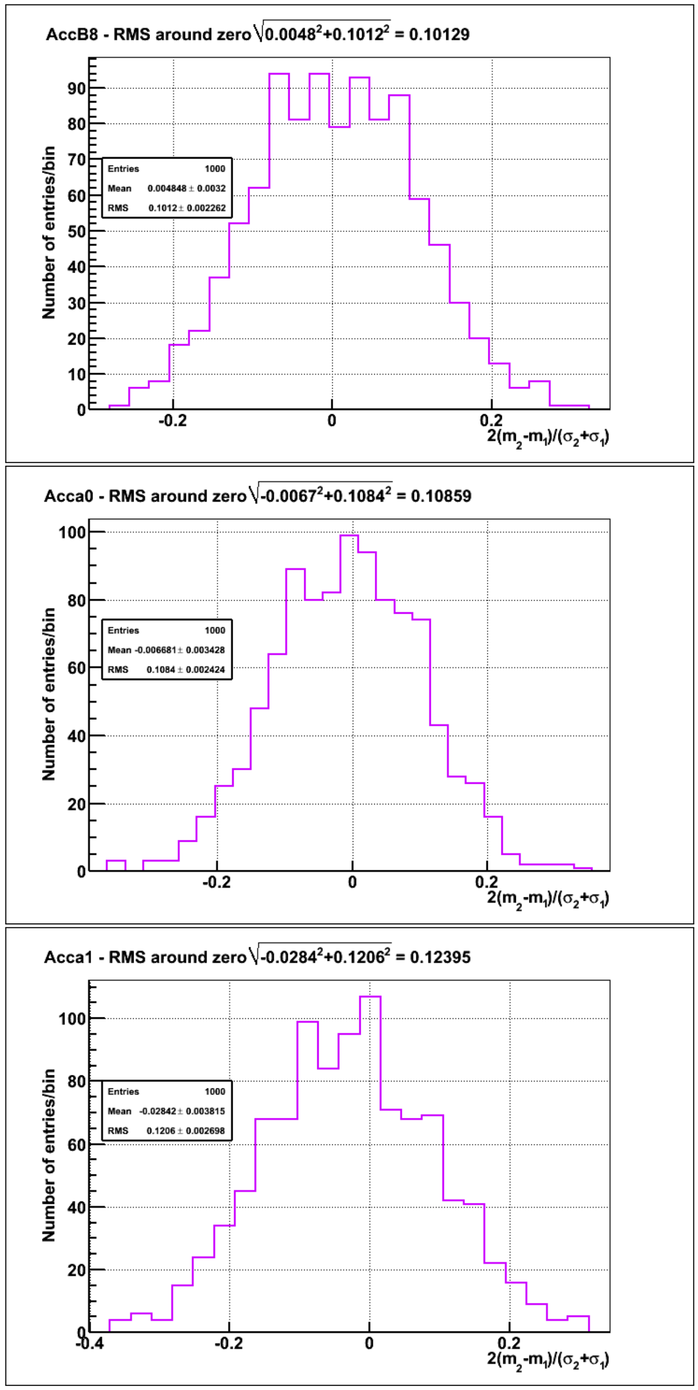


Figure 12.11: Convergence test for ^{8}B scale and day-night parameters a_0 and a_1.

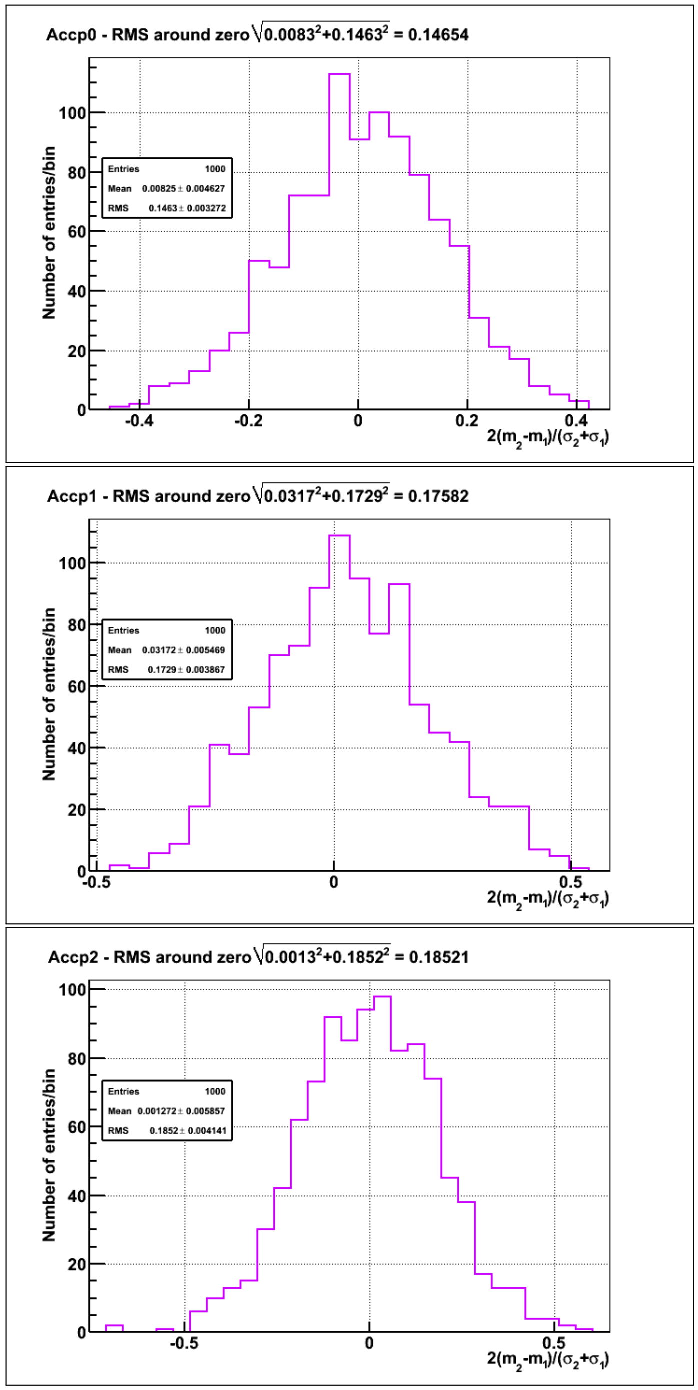


Figure 12.12: Convergence test for P_{ee} parameters p_0, p_1 and p_2.

12.4 1/3 Fit Using LETA Constraint

This section shows the result from the one-third fit using the constraint from LETA [90]. There are 275,000 steps out of which 50,000 steps were removed as burn-in. After removing the burn-in period, every 50^{th} step in the output was used in the analysis to break down the auto correlation in the MCMC output due to Markov property of the chain – the next step is proposed from the current step. Nine parameters were constrained from LETA fit; ^{8}B scale, p_0, p_1, p_3, a_0, a_1, energy scale correlated between PMTs and NCDs, energy non-linearity and the uncertainty in the shape of ^{8}B flux ([95]). The uncertainty of the first 6 parameters in column 2 in table 12.4 is narrower than in table 12.2 because of the LETA constraint.

Parameter	Peak±RMS	Mean	$\frac{\text{(Peak-Mean)}}{\text{RMS}}$
^{8}B scale	0.8975±0.0486087	0.902451	-0.101851
p_0	0.320667±0.0266209	0.312949	0.289908
p_1	-0.00350633±0.0114909	-0.00700466	0.304443
p_2	-0.003±0.00524458	-0.00234337	-0.125201
a_0	0.0779938±0.054771	0.0717547	0.113913
a_1	0.0289789±0.0425045	0.0152803	0.322287
f2e^{pmt}	0.465458±0.0121631	0.46675	-0.106266
CC	595.876±20.573	592.933	0.14308
Atmospheric neutrons	1.03065±0.195174	1.03672	-0.0310702
K2	1.01±0.160073	1.01455	-0.0284108
K5	1.01734±0.122808	1.01164	0.0463607
NCDPD	0.993056±0.303228	1.04974	-0.186952
ES	51.1872±2.23665	51.4787	-0.130314
EX	0.845339±0.477628	1.05253	-0.433792
EX day-night asymmetry	-0.0334444±0.0110389	-0.0330844	-0.0326189
D_2OPD	1.0375±0.148945	1.01028	0.18277
D_2O day-night asymmetry	0.0168293±0.11173	-0.00771537	0.219678
ES$_{\mu\tau}$	16.9307±1.79316	17.2637	-0.185707
f2e^{ncd}	1.77302±0.0419728	1.76544	0.180544
Cos θ resolution direction	-0.00225±0.0665388	-0.00811528	0.0881483
Cos θ resolution	-3.46945e-18±0.102177	0.00110065	-0.0107719
Energy scale	-0.002145±0.00753271	0.000330235	-0.328598
Energy resolution direction	-0.00208±0.0125062	0.000231762	-0.18485
Energy resolution	0.014085±0.0100358	0.0135538	0.0529283
X shift	-0.7075±3.92051	-0.218865	-0.124635
Vertex scale	-0.00173052±0.00548084	-0.00358492	0.338343
Y shift	-0.717262±3.87867	-0.111345	-0.156218
Z shift	-0.704545±4.0551	0.655036	-0.335277
XY resolution -constant term	0.0571382±0.0295222	0.0656156	-0.287154
XY resolution -linear term	-4.5225e-05±6.24351e-05	-4.46289e-05	-0.009548
XY resolution -quadratic term	4.2e-07±2.00981e-07	3.98836e-07	0.105303
Z resolution -constant term	0.075875±0.0288896	0.0698765	0.207634
Z resolution -linear term	0.0001249±8.40038e-05	0.000114327	0.125861
Vertex scale diurnal	-0.000175±0.00146668	-7.13101e-05	-0.0706969
Vertex scale direct	-2.9e-05±0.00175554	-1.39829e-05	-0.00855411
Energy scale diurnal	0.00185625±0.00386744	0.000932874	0.238757
Energy scale direct	0.00306081±0.0102824	0.000839598	0.21602
Energy scale correlated	0.00486458±0.00394943	0.00452603	0.0857234
Energy-dependent fiducial volume	-0.00239583±0.00706907	-0.00114739	-0.176606
Energy non-linearity	-0.0010275±0.00599211	0.000705498	-0.289213
Z scale	0.000102597±0.00131451	0.000393672	-0.221432
Winter uncertainty	-0.393571±0.945854	-0.232626	-0.170159

Table 12.4: A listing of the peak (best fit) and its uncertainty (RMS of the posterior distribution) for the 42 parameters involved in the 1/3 fit using LETA constraint. Besides that, the table also lists the mean and the difference between the mean and the peak in units of the uncertainty.

12.4.1 Asymmetric Systematic Uncertainties along with ^{8}B Winter Uncertainty With LETA Constraint

This is the first time that the code was run with Z scale, ^{8}B Winter uncertainty, and the energy-dependent fiducial volume with the LETA constraint. Following figures 12.13 to 12.16 show posterior distributions fitted with Gaussian functions for vertex scale, Z scale, ^{8}B Winter uncertainty and the energy-dependent fiducial volume. The burn-in is 10,000 and after burn-in every 50^{th} step was used to plot the posterior distributions.

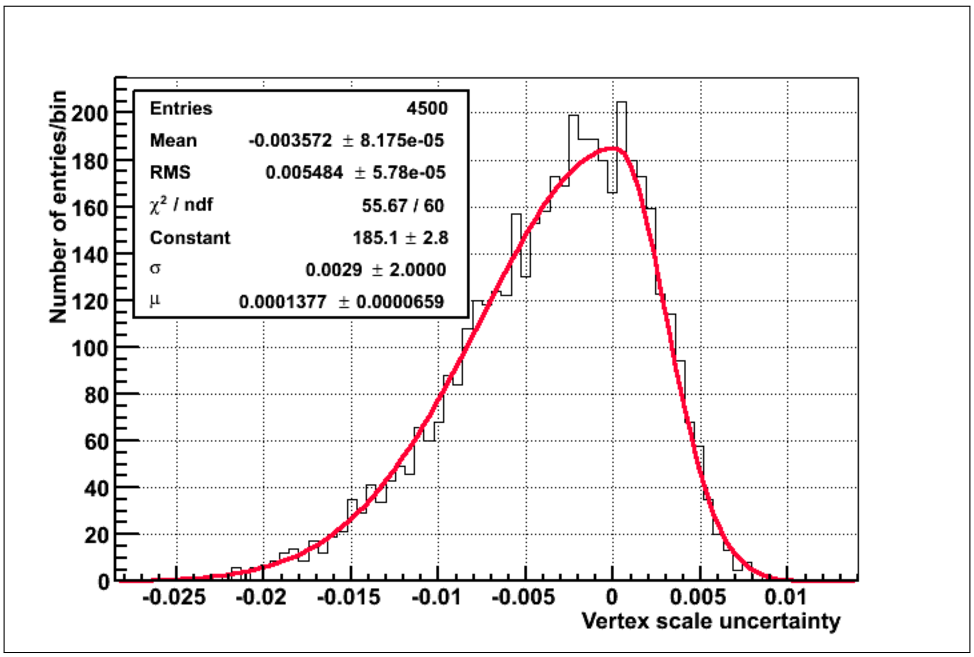


Figure 12.13: Posterior distribution of Vertex scale is shown in black and the Gaussian fit is shown in red. The constraint applied, in the MCMC fit, is $0^{+0.0029}_{-0.0077}$. Additional LETA constraint is used for this fit.

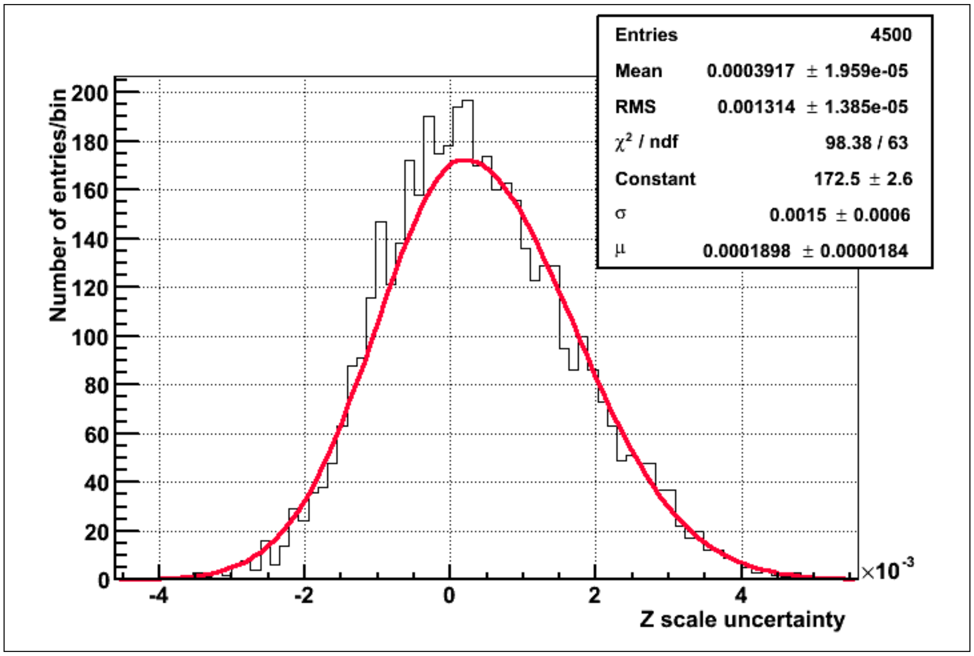


Figure 12.14: Posterior distribution of the Z scale is shown in black and the Gaussian fit is shown in red. The constraint applied, in the MCMC fit, is $0^{+0.0015}_{-0.0012}$. Additional LETA constraint is used for this fit.

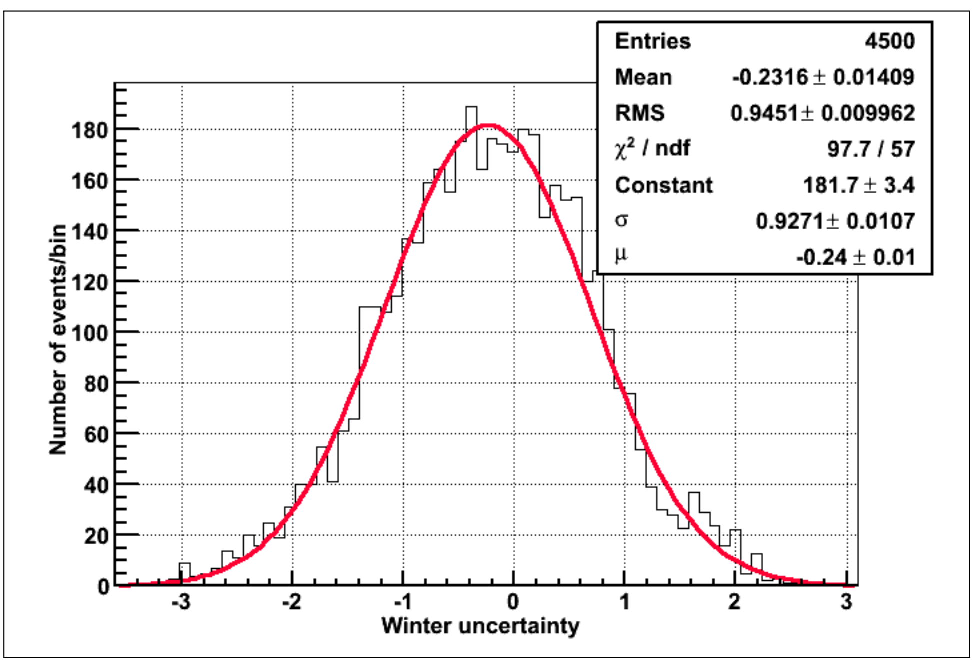


Figure 12.15: Posterior distribution of Winter uncertainty in shown in black and the Gaussian fit is shown in red. The constraint applied, in the MCMC fit, is 0 ± 1.0.

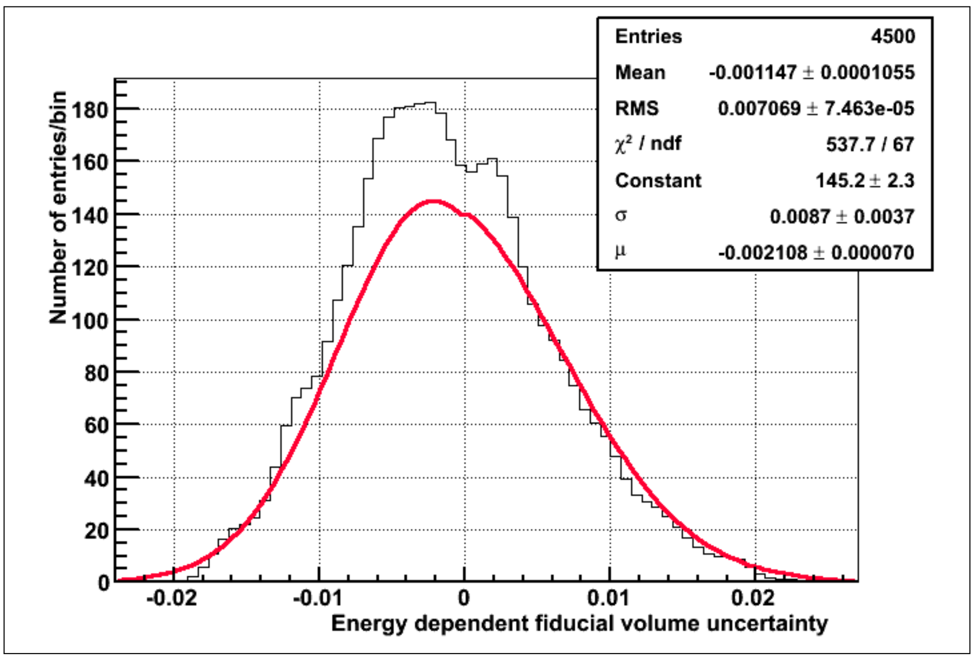


Figure 12.16: Posterior distribution of the energy-dependent fiducial volume uncertainty is shown in black and Gaussian fit of the posterior distribution is shown in red. The constraint applied, in the MCMC fit, is $0^{+0.0088}_{-0.0067}$. χ^2 of 538 with 67 degrees of freedom indicates that posterior probability distribution of energy-dependent fiducial volume uncertainty is not a Gaussian distribution as assumed in the calculation of the likelihood function. Additional LETA constraint is used for this fit.

12.5 Summary

This chapter described several cross-checks performed on the code: (1) to verify that application of the systematic uncertainties in the fit is correct, (2) to ensure convergence of the fit, (3) to ensure that autocorrelation coefficient drops down to zero within 10,000 steps and remain stable throughout the chain. The chapter also outlines the results from two fits: first the NCD-only fit and second the NCD fit with constraint from LETA. The results from both fits were compared between MCMC and QSigEx and agreement between the results propel us to the next step, that is, fitting the full data.

Chapter 13

Fit on the Full Data

13.1 Finally Fitting Full Data

The final fit is on the full dataset. The number of steps in the fit is 750,000 from which 25,000 steps are removed as a burn-in period. Figure 13.1 shows posterior density functions from the MCMC fit for the 6 parameters. The PSA constraint applied is 1115±79. The constraints from LETA are shown in table 13.1. The result of the fit is shown in table 13.2. The next table 13.3 shows the result when the systematic uncertainties are fixed at their nominal values. Comparing the two tables, it seems that the uncertainty from statistics dominate, for instance the total uncertainty on ^{8}B scale is 3.5% but from statistics alone the uncertainty is 3.4%. The correlation among the parameters of interest is shown in table 13.4. The projections of the fit on the three observables ($\cos\theta_{Sun}$, volume-weighted variable ρ and electron effective kinetic energy T_{eff}) are shown in figures 13.2 to 13.4. The number of background events in the Čerenkov data of the NCD phase is listed in table 13.6.

The χ^2, listed in table 13.5 for the one-dimensional projections of the fit in three observables, is evaluated using statistic and systematic uncertainties.

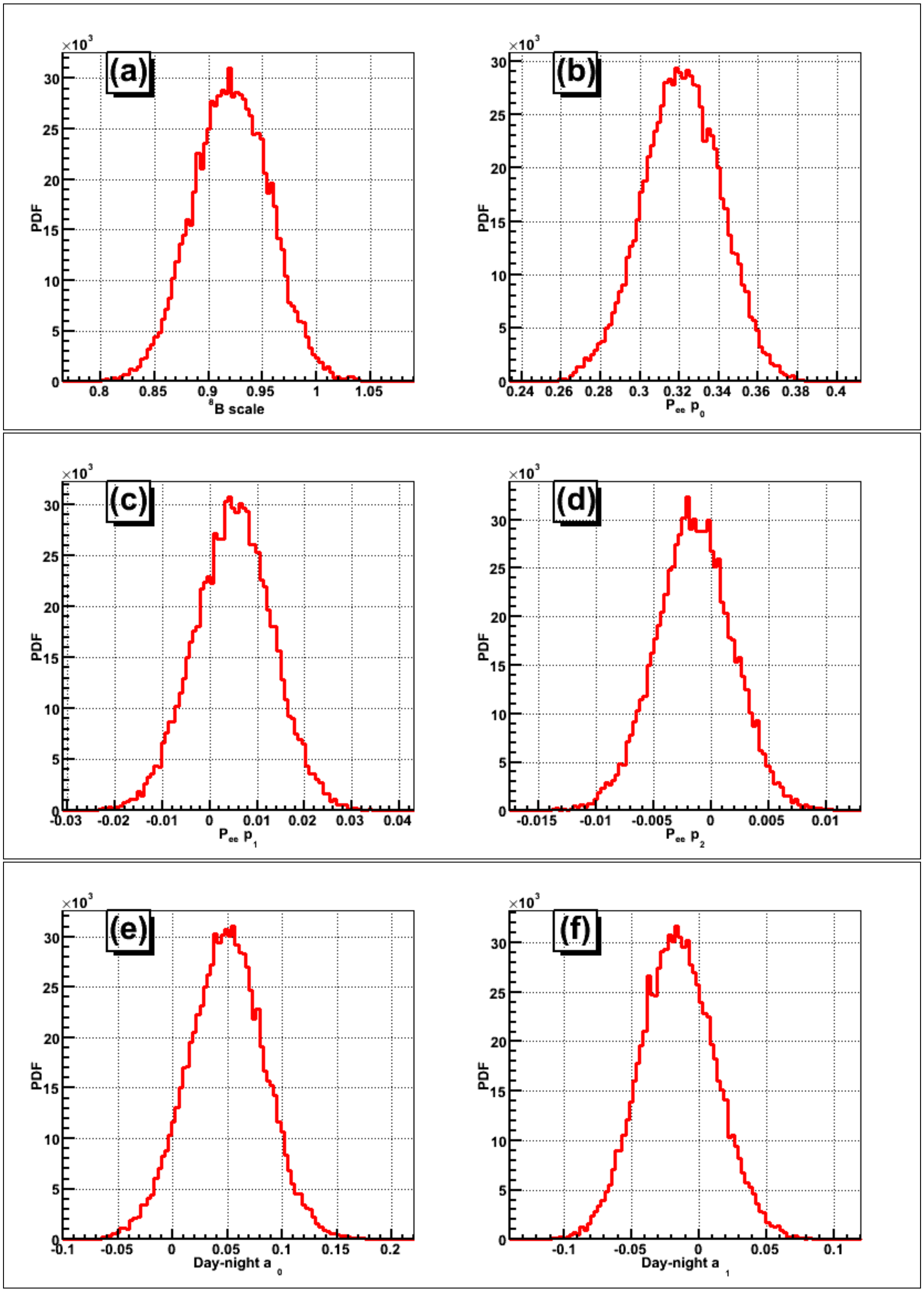


Figure 13.1: Posterior density functions (PDFs) from MCMC fit of 6 parameters; (a) ^{8}B scale, (b) constant term (p_0), (c) linear term (p_1) and (d) quadratic term (p_2) of the electron survival probability described in equation (13.1), (e) constant term (a_0) and (f) linear term (a_1) of the day-night asymmetry described in equation (13.2). These PDFs were used to determine the best-fits described in the first 6 rows in a table 13.2.

Parameter	Constraint	Width of the Constraint
^{8}B	0.9316508	0.0380755
p_0	0.3192673	0.0217642
p_1	0.0073795	0.0093338
p_2	-0.000617	0.0036361
a_0	0.0266399	0.0425075
a_1	-0.022025	0.0318915
Energy Scale Correlated	-0.001628	0.0030291
Energy Non Linearity	0	0.0069
Winter Uncertainty	0	1

Table 13.1: Constraints from the LETA fit [90].

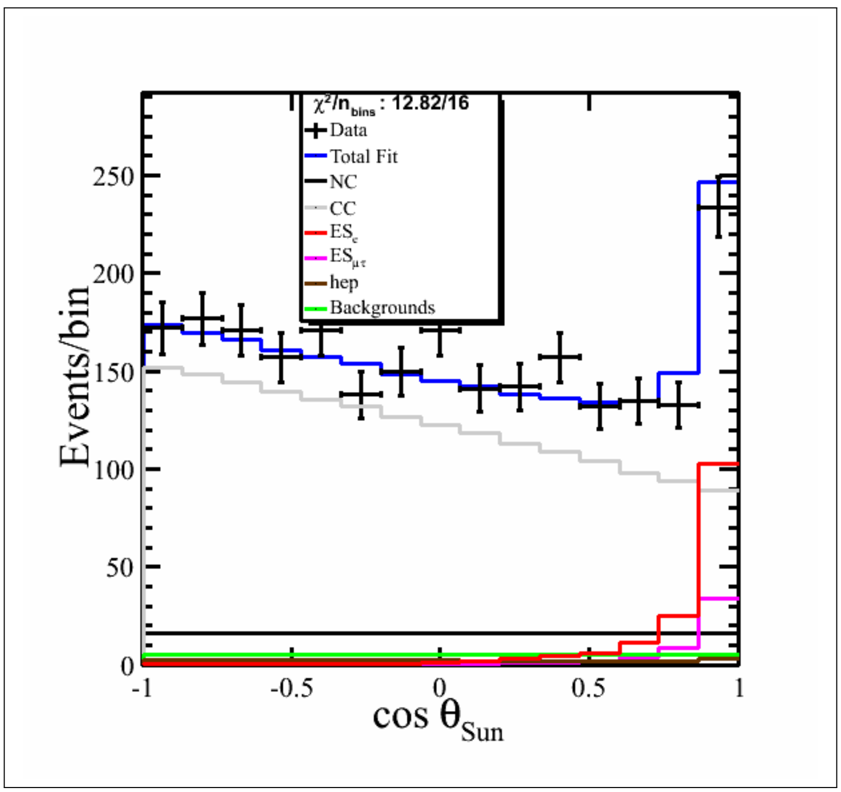


Figure 13.2: One-dimensional projection of the fit in direction ($\cos\theta_{Sun}$) on the Čerenkov data of the NCD phase from the binned-histogram signal extraction with the individual signals separated into three neutrino interactions (ES is split into ES_e and $ES_{\mu\tau}$), backgrounds, and hep neutrino events. Figure also shows χ^2/data points of the fit.

Parameter	**Best-fit**
^{8}B	9.212e-01±3.560e-02
Pee p_0	3.206e-01±1.971e-02
Pee p_1	5.009e-03±8.169e-03
Pee p_2	-1.389e-03±3.331e-03
Day-night asymmetry a_0	4.963e-02±3.469e-02
Day-night asymmetry a_1	-1.754e-02±2.756e-02
ncdpd	1.030e+00±3.164e-01
k2pd	9.963e-01±1.573e-01
k5pd	1.003e+00±1.564e-01
d2opd	9.934e-01±1.573e-01
ex	9.400e-01±4.327e-01
Atmospheric neutrons	1.008e+00±1.961e-01
cc	1.836e+03±3.842e+01
es	1.569e+02±4.410e+00
$es_{\mu\tau}$	5.090e+01±3.478e+00
NC flux-to-event ratio (NCDs)	1.765e+00±3.938e-02
EX day-night asymmetry	-1.977e-02±1.118e-02
D_2OPD day-night asymmetry	-3.460e-02±1.118e-01
NC flux-to-event ratio (PMTs) ϵ_{pmt}	4.620e-01±1.085e-02
Cos θ Resolution Direct	-9.026e-03±6.829e-02
Cos θ Resolution	5.643e-02±1.033e-01
Energy Scale	3.560e-03±6.444e-03
Energy Resolution Direction	5.337e-04±1.221e-02
Energy Resolution	1.243e-02±1.080e-02
X Shift	1.361e+00±3.898e+00
Y Shift	-1.026e+00±3.478e+00
Z Shift	3.332e-01±3.807e+00
Vertex scale	-1.346e-03±4.555e-03
Z scale	5.182e-05±1.348e-03
XY Resolution - constant term	6.487e-02±2.905e-02
XY Resolution - linear term	-5.712e-05±6.146e-05
XY Resolution - quadratic term	3.946e-07±1.972e-07
Z resolution - constant term	7.634e-02±2.808e-02
Z resolution - linear term	1.148e-04±8.171e-05
Vertex diurnal scale	-1.965e-04±1.407e-03
Vertex direction scale	-1.472e-04±1.755e-03
Energy scale diurnal	8.978e-04±3.692e-03
Energy scale direction	2.093e-03±9.283e-03
Energy scale source	-8.458e-04±2.965e-03
Energy-dependent fiducial volume	-3.182e-03±6.486e-03
Energy non-linearity	1.190e-03±6.927e-03
Winter uncertainty	-1.061e-01±9.722e-01

Table 13.2: Fit result of the final analysis. The best-fit is the average of the 68% confidence intervals.

Parameter	**Best-fit**
^{8}B	9.207e-01±3.423e-02
Pee p_0	3.199e-01±1.822e-02
Pee p_1	4.095e-03±7.025e-03
Pee p_2	-1.507e-03±3.109e-03
Day-night a_0	4.932e-02±3.402e-02
Day-night a_1	-1.450e-02±2.718e-02
ncdpd	9.933e-01±3.383e-01
k2pd	9.857e-01±1.576e-01
k5pd	9.908e-01±1.728e-01
d2opd	9.885e-01±1.508e-01
ex	9.355e-01±4.674e-01
Atmospheric neutrons	9.987e-01±1.928e-01
cc	1.833e+03±3.628e+01
es	1.567e+02±4.094e+00
$es_{\mu\tau}$	5.079e+01±3.370e+00
NC flux-to-event ratio (NCDs)	1.762e+00±4.019e-02
EX day-night asymmetry	-1.900e-02±1.125e-02
D_2OPD Day-night asymmetry	-3.714e-02±1.116e-01
NC flux-to-event (PMTs)	4.629e-01±9.015e-03

Table 13.3: Fit result of the final analysis with fixed systematic uncertainties. The best-fit is the average of the 68% confidence intervals.

	^{8}B	p_0	p_1	p_2	a_0	a_l
^{8}B	1.000	-0.729	0.278	-0.117	0.067	-0.042
p_0	-0.729	1.000	-0.318	-0.386	-0.378	0.148
p_1	0.278	-0.318	1.000	-0.139	0.280	-0.666
p_2	-0.117	-0.386	-0.139	1.000	-0.017	0.011
a_0	0.067	-0.378	0.280	-0.017	1.000	-0.383
a_l	-0.042	0.148	-0.666	0.011	-0.383	1.000

Table 13.4: Correlation matrix for the polynomial survival probability fit from the MCMC fit.

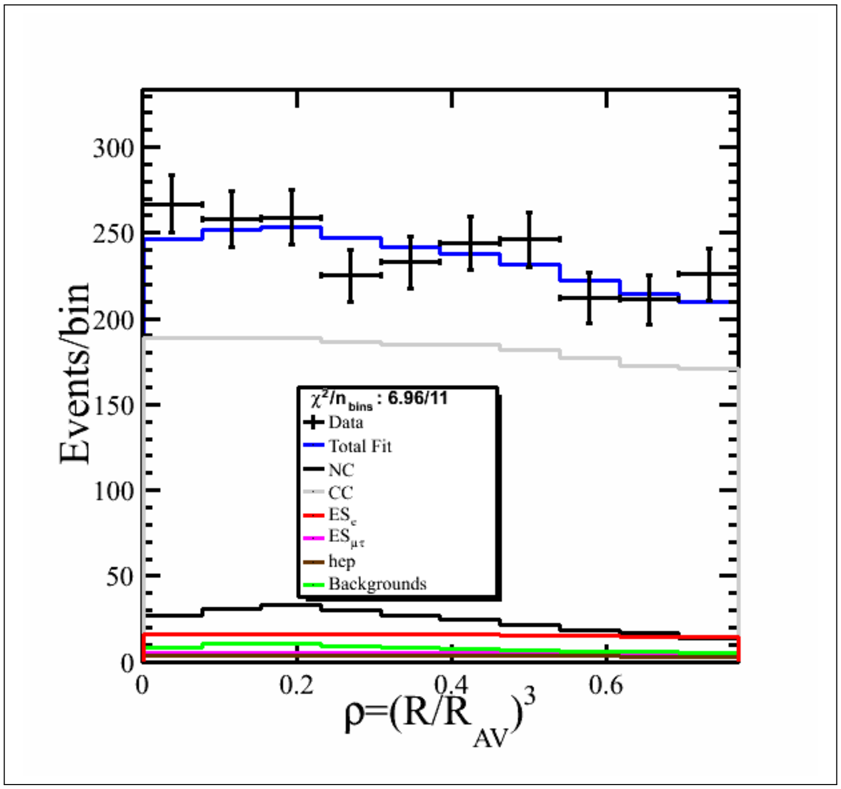


Figure 13.3: One-dimensional projection of the fit in ρ on the Čerenkov data of the NCD phase from the binned-histogram signal extraction with the individual signals separated into three neutrino interactions (ES is split into ES_e and $ES_{\mu\tau}$), backgrounds, and hep neutrino events. Figure also shows χ^2/data points of the fit.

Observable	χ^2 **(data points)**	**Figure Number**
$\cos\theta_{Sun}$	12.82/16	13.2
R^3	6.96/11	13.3
T_{eff}	10.16/14	13.4

Table 13.5: χ^2 from a one-dimensional projections of the fit in three observables. Table also lists number of data points used in the computation of χ^2 along with figure number pointing to the figure which displays the one-dimensional projection.

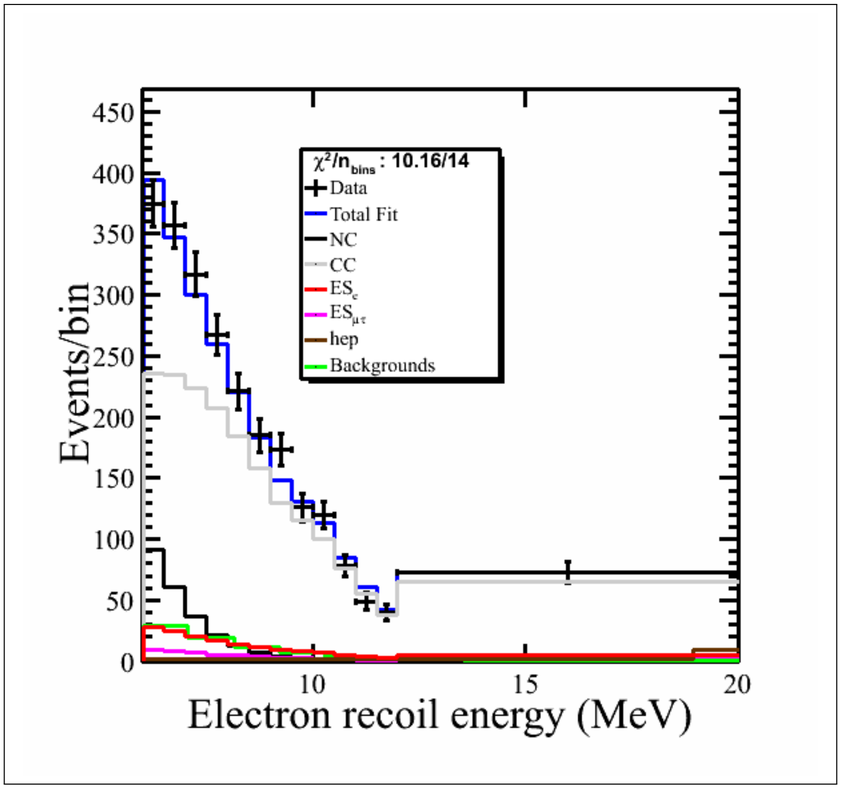


Figure 13.4: One-dimensional projection of the fit in recoil energy (T_{eff}) on the Čerenkov data of the NCD phase from the binned-histogram signal extraction with the individual signals separated into three neutrino interactions (ES is split into ES_e and $ES_{\mu\tau}$), backgrounds, and hep neutrino events. Figure also shows χ^2/data points of the fit.

Backgrounds	**Number of Events**
External neutrons	19.5
D_2O photo-disintegration neutrons	8.25
Atmospheric neutrinos	24.88
K2 photo-disintegration neutrons	9.37
K5 photo-disintegration neutrons	12.10
NCD photo-disintegration neutrons	6.12
hep ν events	34.21
Total data events	2381

Table 13.6: Number of background events in the Čerenkov data of the NCD phase.

13.2 Plotting energy-dependent day-night survival probabilities and day-night asymmetry

The bands in figures 13.5 were computed from $\pm\sigma$ of the posterior distributions representing equations 13.1 to 13.3 where E_ν ranges from 6 MeV to 14 MeV and it is incremented by 0.025 MeV. The bands include statistical and systematic uncertainties as well as correlation between the parameters. MCMC fit makes no assumption about the shape of neutrino energy distribution except that it is a smooth function and varies slowly over the range to which SNO is sensitive.

$$P_{ee}^{day} = p_0 + p_1(E_\nu - 10.0\text{ MeV}) + p_2(E_\nu - 10.0\text{ MeV})^2 \tag{13.1}$$

The regeneration of ν_e in the Earth at night is modelled as a linear perturbation to the daytime ν_e survival probability described in equation (13.1).

$$A = a_0 + a_1(E_\nu - 10.0\text{ MeV}) \tag{13.2}$$

$$P_{ee}^{night} = P_{ee}^{day}\frac{(2+A)}{(2-A)} \tag{13.3}$$

where E_ν is neutrino energy shown on the X axis in figure 13.5.

13.3 Extraction of CC and ES energy spectra

The energy spectra of ^{8}B flux is computed via charged current interactions on deuterium and elastic scattering interactions on electrons. The bands, shown in figure 13.6, represent both statistical and systematic uncertainties and were computed taking into account the uncertainties in the fit parameters and the correlation among the fit parameters. For each sampling of the MCMC output, the 3D PDFs – separated into the day and night PDFs – were reconstructed after smearing the observables in the Monte Carlo with the systematic uncertainties, and then calculating the number of events – split into the day and night events – belonging to CC and ES. The day and night 3D PDFs were scaled according to the number of events for day and night respectively. The 1D projection on the recoil electron energy was used to plot the number of events belonging to each energy bin. From the distribution of the number of events in each energy bin, 68% confidence intervals were computed. The best-fit is the average of 68% CL (equation (11.3)) and the bands represent the average of difference of the 68% confidence levels (equation (11.4)). The spectra is presented in tabular form in Appendix C.

The day and night live times — 176.59 days and 208.85 nights — used in the extraction of the day and night energy spectra were determined by splitting the data on solar zenith angle; day events are when the Sun is above the horizon and night events are when the Sun is below the horizon. From SNOMAN, the number of electrons and deuterium in the fiducial volume are 6.023×10^{31} and 3.0115×10^{32} respectively.

13.4 Comparison between QSigEx and MCMC

Table 13.7 compares 3-phase P_{ee} day/night fit result from QSigEx (column 2) to MCMC (column 3). The relative difference, shown in the last column is the difference between the fitted parameter values in terms of the average total fit uncertainties. Since the maximum relative difference between the two analysis is only 0.3σ, QSigEx result was used for the extraction of global (solar+KamLAND) 3-flavour oscillation parameters.

Parameter	QSigEx	MCMC	Relative Difference
^{8}B scale	0.921 ± 0.035	0.921 ± 0.036	-0.018
p_0	0.319 ± 0.018	0.321 ± 0.020	-0.103
p_1	0.002 ± 0.008	0.005 ± 0.008	+0.119
p_2	-0.001 ± 0.003	-0.002 ± 0.003	-0.081
a_0	0.044 ± 0.034	0.048 ± 0.035	+0.165
a_1	-0.017 ± 0.027	-0.015 ± 0.028	-0.151

Table 13.7: Comparing 3-phase P_{ee} day/night fit result from MCMC to QSigEx. Table from [90].

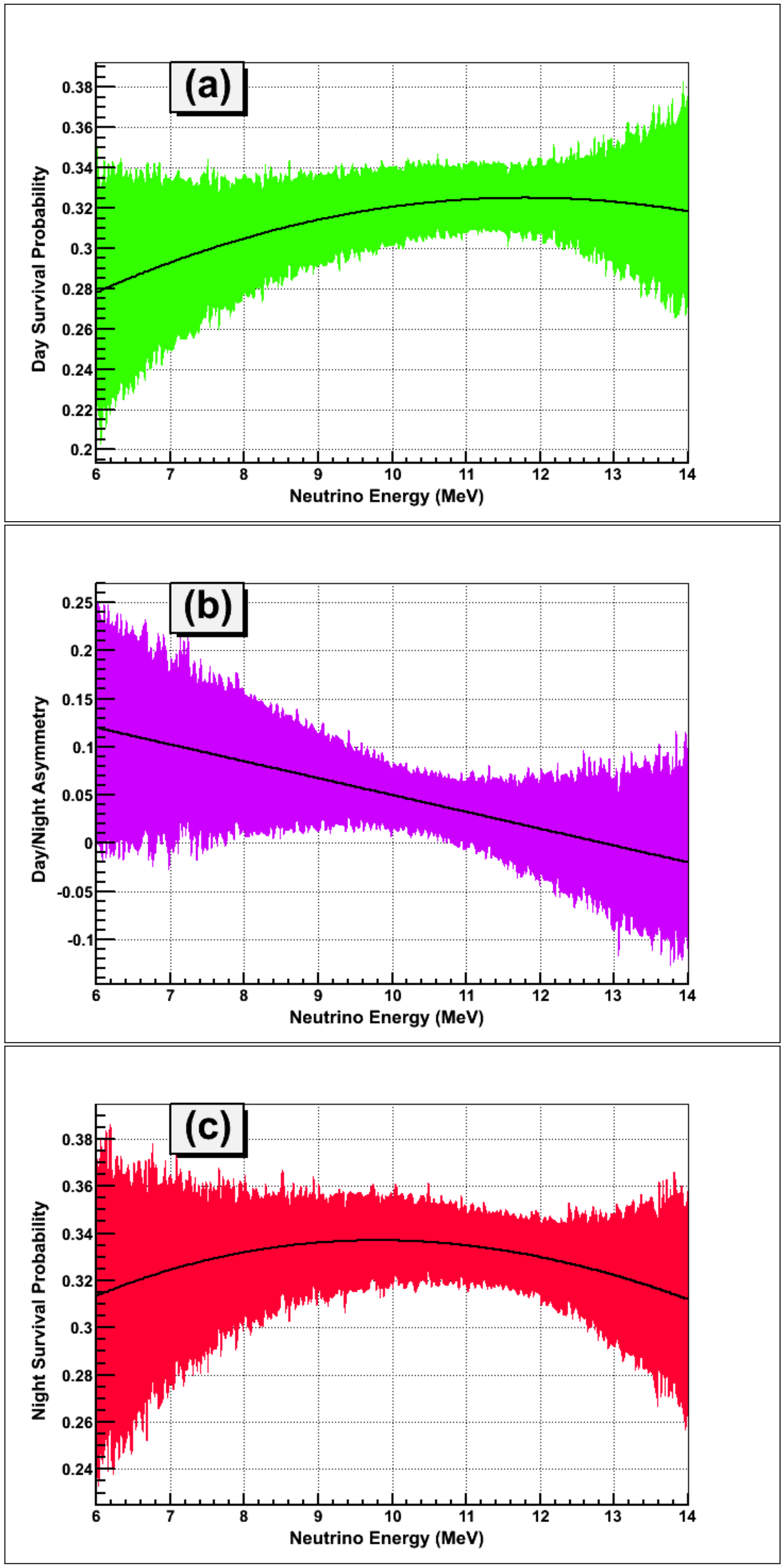


Figure 13.5: Using best-fit and its uncertainty of 68% confidence intervals; (a) showing equation (13.1), (b) showing equation (13.2) and (c) showing equation (13.3).

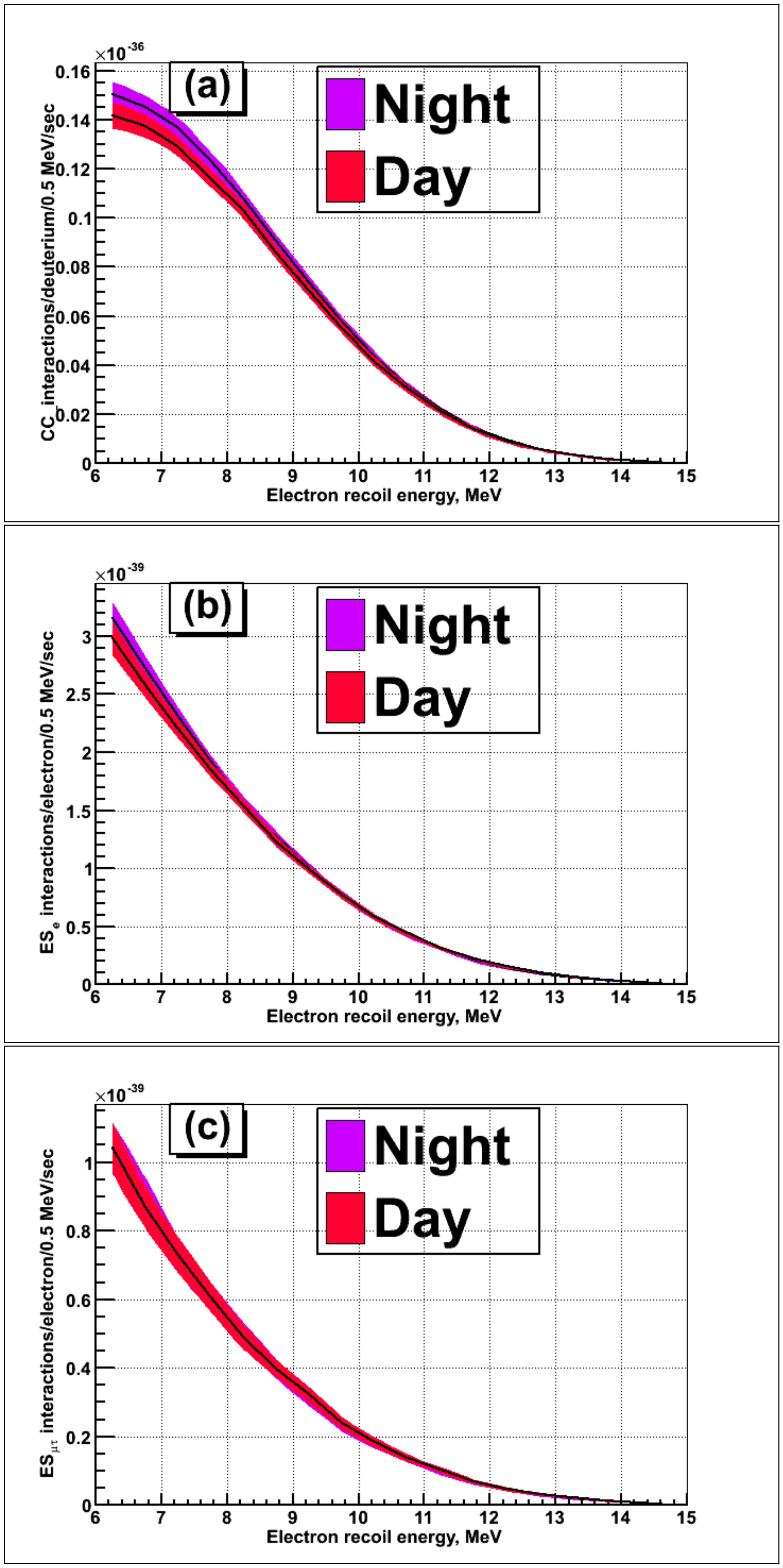


Figure 13.6: Figures (a), (b) and (c) show extracted CC spectra , ES (ν_e) and ES (ν_μ and ν_τ) spectra respectively. Bands include both systematic and statistical uncertainties in the MCMC fit.

Chapter 14

Conclusion

In the 1930s, Hans Bethe [96] and Carl Friedrich von Weizsäcker ([97] and [98]) postulated the Sun's source of energy to be from fusion reactions in its core. For decades, there was no way to directly test the hypothesis. The detection of solar neutrinos by Ray Davis's experiment (a tank of dry cleaning fluid deep in the Homestake mine at Lead, South Dakota) was a strong indication that the nuclear theory of the Sun was correct though the two-third deficit from the prediction caused a suspicion. Since then SuperKamiokande experiment in Japan and SNO experiment in Canada have proved beyond doubt that neutrinos oscillate (one flavour of ν transforms into another flavour and back again.). Using the global solar and the KamLAND result, the parameter space of oscillation is reduced to LMA region (figure 1.7).

Result from each phase of SNO was published separately. To improve the result, the data from the first two phases (D_2O and Salt) were combined into a low-energy-threshold-analysis (LETA). Due to various improvements carried out for LETA, the uncertainty on the total flux of active flavour neutrinos from ^{8}B decay in the Sun, measured via the neutral current interactions, was more than a factor of 2 smaller than previously published results [61]. The role of this thesis is to further narrow it down. Table 14.1 compares the result from MCMC to the published LETA result. A MCMC fit was performed on the

Čerenkov data of the NCD phase using the extended likelihood function. Three observables were used to separate different event types: the effective electron kinetic energy (T_{eff}), the event direction with respect to the vector of the Sun ($\cos\theta_{\odot}$), and the normalized cube of the radial position in the detector (R^3). To take into account correlations between observables, 3-dimensional PDFs $P(R^3, \cos\theta_{\odot}, T_{\text{eff}})$ were used for all the event types. Uncertainties in the distribution of the observables were treated as parameters of the fit for the distortions of the Monte Carlo PDF shapes. All the systematic uncertainties were allowed to vary in the fit. The type of events included in the fit were three ν interactions (CC, NC and ES which was split into ES_e and $ES_{\mu\tau}$) and 6 backgrounds.

The MCMC fit directly extracted the energy-dependent ν_e survival probability which was parametrized as a polynomial function (section 4.10.7) and applied as distortion to the ^{8}B neutrino energy spectrum. The parameters of the polynomial (p_0, p_1, p_2, a_0 and a_1) were varied in the fit for each step, the ^{8}B ν energy spectrum was distorted and the shapes of CC and ES energy spectra were recomputed from the distorted ^{8}B spectrum for each step in the fit. The extracted survival probability also incorporated the uncertainty in the shape of the undistorted ^{8}B energy spectrum by treating the uncertainty as a systematic uncertainty and floating it in the MCMC fit.

The shapes of energy spectra of NC and radioactive backgrounds were only floated within their systematic uncertainties as these have no dependence on the ν oscillation model. Signals and backgrounds are covered in detail in chapter 4.

The flux of solar neutrinos was assumed to be constant during the operation of the three phases of SNO (1999-2006). Although SNO was primarily sensitive to the ^{8}B chain of solar neutrinos, a fixed contribution of solar hep neutrinos (9.3×10^3 cm$^{-2}s^{-1}$ from [99]) was included in the fit.

The author of this thesis performed MCMC fit on the Čerenkov data of the NCD phase. The data, from the NCDs of the NCD phase, are included as a PSA constraint and data, from the D_2O and Salt phases, were combined in the fit by using constraints from the LETA output. The result presented in the LETA paper [61] is:

$$
\begin{aligned}
{}^{8}\text{B Flux } \phi^{\text{binned}} &= 5.14^{+0.160}_{-0.158}(stat)^{+0.132}_{-0.117}(syst) \times 10^{6}\ \text{cm}^{-2}s^{-1} \\
{}^{8}\text{B Flux } \phi^{\text{kernel}} &= 5.17^{+0.159}_{-0.158}(stat)^{+0.132}_{-0.114}(syst) \times 10^{6}\ \text{cm}^{-2}s^{-1}
\end{aligned}
$$

For the LETA fit which combined data from the D_2O and Salt phases, two independent techniques were used to extract the fit parameters. One method used binned PDFs and the other method used unbinned, "kernel estimation" approach. For more details, refer to the LETA paper.

For the fit that combined data from all three phases, two independent techniques were also used for the signal extraction. One method is the Maximum Likelihood Estimation (MLE) method and another is the Markov Chain Monte Carlo (MCMC) method. The strength of MLE is point estimations and the strength of MCMC is posterior distributions for the extraction of confidence intervals.

In the MCMC fit, the systematics are floated along with the number of events belonging to the event classes. The uncertainty in the fit includes both statistical and systematic uncertainties. The result from fitting the MCMC fit on the 1/3 NCD-only data is:

$$
{}^{8}\text{B Flux } \phi = 5.12 \pm 0.50 \times 10^{6}\ \text{cm}^{-2}s^{-1}
$$

There are a number of options[1] available to pick as best-fit from the MCMC fit. We are using the average of 68% confidence intervals. The result from the

[1] Mean, mode and median of the posterior distribution, average of confidence intervals, and fitting the posterior distribution to a Gaussian function to extract the mean μ and σ as best-fit and its uncertainty respectively.

1/3 fit using the constraints from LETA is:

$$^{8}\text{B Flux } \phi = 5.11 \pm 0.29 \times 10^{6} \text{ cm}^{-2} s^{-1}$$

The use of constraint from LETA reduced the total uncertainty on ^{8}B flux from 0.5 $\times 10^{6}$ cm$^{-2}s^{-1}$ to 0.29 $\times 10^{6}$ cm$^{-2}s^{-1}$. From the BS05(OP) model, ^{8}B flux is predicted to be 5.69×10^{6} cm^{-2} s^{-1} with $\pm 16\%$ theoretical uncertainty. The prediction of ^{8}B flux from BS05(AGS,OP) model is $4.51 \times 10^{6}\, cm^{-2}\, s^{-1} \pm 16\%$. The model assumed a lower heavy element abundance in the Sun's surface compared to BS05(OP) model. From the MCMC fit on the full data,

$$^{8}\text{B Flux } \phi = (5.28 \pm 0.20) \times 10^{6}\, \text{cm}^{-2}\, s^{-1},$$

with total uncertainty of 3.78% which agrees well with both models.

The combined three-phase fit of the ν_e energy-dependent survival probability yielded the constant term of the survival probability as 0.3206 ± 0.0197, linear term as 0.005 ± 0.008, the quadratic term as -0.0014 ± 0.0033. On the day-night asymmetry, the constant term is 0.0496 ± 0.0347 and the linear term is -0.0175 ± 0.0276. Another goal of the thesis was to observe a nonzero day-night asymmetry. So far no experiment (References [27], [100]) has measured a significant nonzero day-night asymmetry. The best-fit on day-night asymmetry from floating all the systematic uncertainties is 0.0496 ± 0.0347. To find the effect of statistics, the systematic uncertainties were kept fixed and not "floated" in a fit consisting of 750,000 steps. The result is 0.0493 ± 0.0340. This clearly shows that uncertainty from statistics (0.0340) dominate in the total uncertainty (0.0347) from a fit which included both statistics and systematics. The day-night asymmetry is 1.5σ away from zero which means that if SNO experiment was repeated 100 times, the asymmetry would be greater than zero 87 times[2]

[2] Area under the normal curve between $\pm n\sigma$ is erf$(n/\sqrt{(2)})$ where n is the number of σ and **erf** is the error function.

Table 14.1, compares the result from MCMC to the published result from LETA. The addition of two new systematic uncertainties and more conservative application of other systematic uncertainties resulted in changing the **central values** of some of the fit parameters which negate the advantage of adding the data from the NCD phase to the combined data from the D_2O and salt phases. Treatment of uncertainties for D_2O+salt has been improved for the 3-phase analysis, resulting in larger uncertainties. These include separate floating of PMT $\beta - \gamma$ events in to day and night events instead of floating the total number of events, evaluation of the uncertainties due to finite statistics in Monte Carlo simulation, addition of the ^{8}B Winter uncertainty for neutral current[3], removal of the positive bound for salt acrylic vessel neutrons and merging of the day/night background constraints instead of doubling the penalty terms *etc.*. Additionally for the NCD phase, the external neutrons are not "subtracted". The external neutrons are handled as a separate class of event in the fit, which is floated like any other class of event. When statistical separation or a pull, from other phases, exists for either the ^{8}B scale parameter or the external neutrons or anything else that has neutron-like PDFs, the fitted rate for the external neutrons in the NCD phase is free to vary and will not necessarily fit at the central value of the constraint [101]. In MCMC fit, each neutron-type background is floated separately. In LETA, not only external neutrons but all neutrons other than from the neutral current interactions were subtracted after the fit, based on results from Monte Carlo studies. This should not add uncertainties compared to the published LETA fit, although it can slightly affect the central values of the fit parameters and this is what we observed. The constraint, applied on the 3-phase fit on the total number of neutrons, from pulse shape analysis of data from the NCDs is 7.0% which was

[3]Winter uncertainty was applied on CC and ES in LETA but its effect was not considered on NC.

not applied to the 2-phase (LETA) fit. The shift ((0.9212-0.8868)=0.0344) in the central value of the ^{8}B scale from LETA to MCMC, in terms of average uncertainty $\sigma = (0.0356 + 0.0341)/2$, is $\approx 1.0\sigma$.

The recipe to calculate the total uncertainty, shown in the last column in table 14.1, is

- the asymmetric uncertainties, from statistics and systematics, were converted into symmetric uncertainties using the following equation:

$$\sigma = \sqrt{(\sigma_m^2 + \sigma_p^2 - \sigma_m \sigma_p)} \tag{14.1}$$

  where σ_m is minus uncertainty and σ_p is the plus uncertainty.

- statistical and systematic uncertainties are added in quadrature to obtain the total uncertainty as:

$$\sigma = \sqrt{\sigma_{stat}^2 + \sigma_{syst}^2} \tag{14.2}$$

From the table 14.1, the total uncertainty in MCMC (column 3), compared to LETA (column 8), has clearly improved for the 5 parameters but not for ^{8}B because of the conservative application of the systematics uncertainties. The fractional uncertainty (column 4 for MCMC and column 9 for LETA) increased for those parameters which were nearly zero – p_1, p_2 and a_1. The huge difference in the relative uncertainty is clearly due to the shift of the central values and not the actual variation in the size of the uncertainties. For instance, for p_2, MCMC has $-0.00139 \pm 3.33e-03$ but LETA has $-0.00206 \pm 3.43e-03$. The difference in the central values (0.00067) is ≈ 7 times bigger than the difference in the uncertainties (0.0001). The interpretation of relative uncertainties makes sense when the uncertainties are small compared to the central values, but it is certainly not the case for many of the parameters that we are measuring, hence interpretation is tricky using relative uncertainties to compare LETA and the 3-phase results.

There is an improvement in the survival probability of electron neutrinos (p_0) and day-night asymmetry (a_0). For p_0, the fractional uncertainty in LETA was 0.067 while in MCMC the uncertainty is 0.061. The improvement in fractional uncertainty of a_0 is bigger; LETA has 1.23 while MCMC had 0.699.

14.1 Physics Interpretations

This section presents an interpretation of MCMC results in terms of neutrino oscillations. Nuno Fiuza de Barros performed a scan of the MSW oscillation plane using results from the three-phase analysis. The plots, shown in this section, are from his doctoral dissertation. For more detail on the extraction of oscillation parameters, refer to [102].

Allowed regions of the ν mixing parameters were determined in three configurations, namely:

- "SNO only" (MCMC) result for a two-flavour analysis is shown in figure 14.1 and outlined in table 14.2.
- SNO result (QSigEx) combined with other solar experiments (global solar) for a two-flavour analysis is shown in figure 14.2. Result is outlined in the first row in table 14.3.
- Global solar result, combined with the result from the KamLAND reactor antineutrino experiment, for a three-flavour analysis is shown in figure 14.3. Result is outlined in the second row in table 14.3.
- Results from three-phase analysis are combined with global solar experiments for a three-flavour analysis. The best-fit of the oscillation parameters are outlined in table 14.4.

The "SNO only" result shows degenerate minima in both LMA and LOW oscillation regions in the top plot in figure 14.1, therefore table 14.2 shows the best-fit points and respective uncertainties for each of the local minimum. SNO result when combined with all solar experiments confines the allowed oscillation parameters to the LMA region. The bottom plot in figure 14.1 concentrates on the LMA region. The star, shown in the following figures, represents the best-fit point – a point with the maximum value of the likelihood in the signal extraction fit [102]. The points on the contour – represented by ($\mathcal{L}(\mu_i) - \mathcal{L}_{max}(\mu_i) = -{n^2}/{2}$) – are points for which the log likelihood decreased by ${n^2}/{2}$ from the global maximum $\mathcal{L}_{max}$. The contours represent the n^{th} signal bound on the oscillation parameters where n=1,2,3.

Comparing uncertainties of the best-fit values of oscillation parameters from "SNO only" result (table 14.2) between LETA+NCD and MCMC, the χ^2/dof is better for LETA than MCMC. Uncertainty on $\tan^2\theta_{12}$ is same in the LMA region but improves in MCMC in the LOW region. On the $\Delta^2 m_{21}$, MCMC has reduced uncertainty compared to LETA+NCD result.

The uncertainties from the global analysis (table 14.3) on $\tan^2\theta_{12}$ have improved from the LETA fit 0.0383 to the MCMC fit 0.0314. The total uncertainty on Δm^2_{21} increased from LETA (0.21) to MCMC (0.38). Comparing the global solar three-flavour analysis in table 14.4, it is clear that an improvement in the accuracy of oscillating parameters is gained by combining the data from all three phases as both MCMC and QSigEx showed improvements. Though it is not surprising that the best result is from the maximum likelihood method where the best fit is obtained by maximizing the likelihood function against **all** the parameters in the fit where as in MCMC the best-fit of parameter **x** is obtained by maximising the likelihood function against the parameter **x** and the rest of the parameters are marginalised.

Although SNO is not as sensitive to $\sin^2\theta_{13}$, as other experiments which were designed to measure $\sin^2\theta_{13}$, for instance, Tokai-to-Kamioka (T2K) [103] experiment in Japan and Main Injector Neutrino Oscillation Search (MINOS) experiment in USA. From the global analysis of neutrino data including the latest result from T2K and MINOS, Fogli *et al.*obtained more than 3σ evidence of non-zero θ_{13} [105]. Their result is 0.025 ± 0.007 which agrees with the SNO result (0.020 ± 0.019) although with a much better uncertainty.

	MCMC			Published LETA				
Parameter	**Value**	**Total Error**	**Fractional Error**	**Value**	**Statistical Error**	**Systematics Error**	**Total Error**	**Fractional Error**
^{8}B Scale	0.9212	0.0356	3.86e-02	0.8868	$^{+0.0279}_{-0.0267}$	$^{+0.0188}_{-0.0216}$	0.0341	3.84e-02
p_0	0.3206	1.97e-02	6.14e-02	0.3435	$^{+0.0205}_{-0.0197}$	$^{+0.0122}_{-0.0089}$	2.23e-02	6.62e-02
p_1	0.0050	8.17e-03	1.63	0.00795	$^{+0.00780}_{-0.00745}$	$^{+0.00388}_{-0.00412}$	8.62e-03	1.08
p_2	-0.00139	3.33e-03	2.395	-0.00206	$^{+0.00302}_{-0.00311}$	$^{+0.00159}_{-0.00148}$	3.43e-03	1.67
a_0	0.04963	3.47e-02	6.99e-01	0.0325	$^{+0.0366}_{-0.0360}$	$^{+0.0157}_{-0.0174}$	3.99e-02	1.23
a_1	-0.01754	2.76e-02	1.57	-0.0311	$^{+0.0279}_{-0.0292}$	$^{+0.0174}_{-0.0141}$	3.28e-02	1.05

Table 14.1: Comparing result from MCMC to the published LETA result.

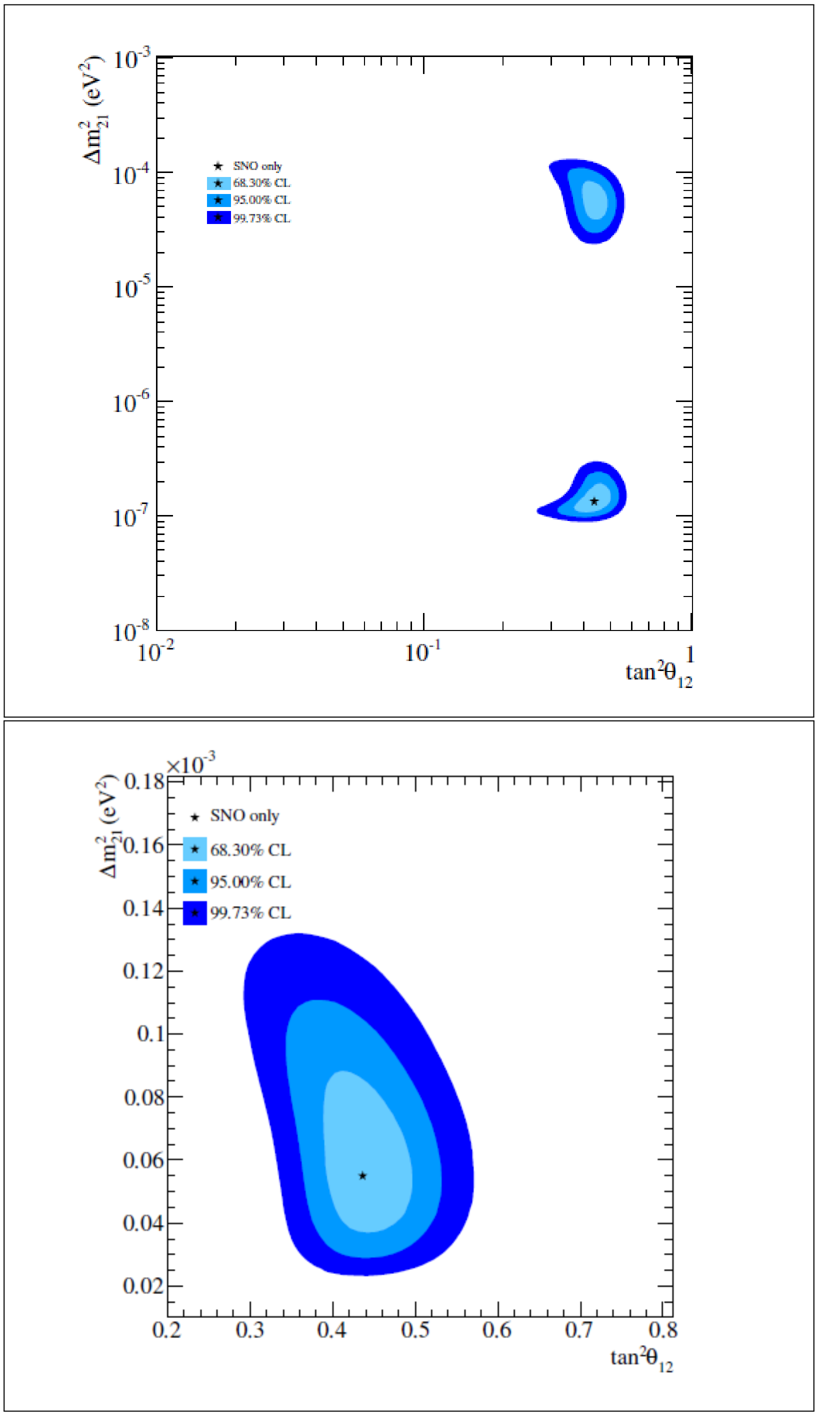


Figure 14.1: Contour of allowed oscillation parameters from the MCMC result in the full region (top plot) of oscillation parameters. The bottom plot shows details of the LMA region (bottom plot). Plots are from [102].

Analysis	**Region**	$\tan^2\theta_{12}$	**stat $\oplus$ stat** $\tan^2\theta_{12}$	$\Delta m^2_{21}\, eV^2$	**stat $\oplus$ stat** $\Delta m^2_{21}\, eV^2$	χ^2/dof
MCMC	LMA	$0.436^{+0.042}_{-0.038}$	0.040	$5.50^{+2.09}_{-1.33} \times 10^{-5}$	$1.832{\times}10^{-5}$	$^{1.69}/_{6}$
	LOW	$0.437^{+0.042}_{-0.056}$	0.050	$1.35^{+0.31}_{-0.23} \times 10^{-7}$	$0.279{\times}10^{-7}$	$^{1.23}/_{6}$
LETA+NCD	LMA	$0.457^{+0.038}_{-0.042}$	0.040	$5.50^{+2.21}_{-1.62} \times 10^{-5}$	$1.982{\times}10^{-5}$	$^{8.20}/_{9}$
	LOW	$0.437^{+0.058}_{-0.058}$	0.058	$1.15^{+0.38}_{-0.18} \times 10^{-7}$	$0.329{\times}10^{-7}$	$^{6.80}/_{9}$

Table 14.2: The best-fit point along with its uncertainty from SNO only solutions of the oscillation parameter space; first two rows show result from MCMC and the last two rows from LETA+NCD; dof is degrees of freedom.

Analysis	$\tan^2\theta_{12}$	$\Delta m^2_{21} \times 10^{-5}\,(eV^2)$	$\sin^2\theta_{13}(\times 10^{-2})$
2-flavour (QSigEx)	$0.422^{+0.026}_{-0.026}$	$7.38^{+0.44}_{-0.22}$	
3-flavour (QSigEx)	$0.440^{+0.034}_{-0.028}$	$7.38^{+0.44}_{-0.22}$	$2.50^{+2.00}_{-1.50}$
			< 5.00(95% C.L.)
LETA+NCD	$0.468^{+0.042}_{-0.033}$	$7.59^{+0.21}_{-0.21}$	$2.00^{+2.09}_{-1.63}$
			< 8.10(95% C.L.)

Table 14.3: Table with best-fit values of the oscillation parameters from two and three flavour analysis of global solar+KamLAND data. Last row shows result from published LETA paper. Uncertainties are $\pm\sigma$.

Analysis	$\tan^2\theta_{12}$	$\Delta m^2_{21} \times 10^{-5}\,(eV^2)$	$\sin^2\theta_{13}(\times 10^{-2})$	χ^2/dof
QSigEx	$0.436^{+0.044}_{-0.042}$	$5.40^{+1.76}_{-1.32}$	< 5.00(95% C.L.)	$^{108.25}/_{126}$
MCMC	$0.434^{+0.054}_{-0.045}$	$5.45^{+1.98}_{-1.98}$	< 5.83(95% C.L.)	$^{109.32}/_{126}$
LETA+NCD	$0.468^{+0.052}_{-0.050}$	$6.31^{+2.49}_{-2.58}$	< 8.10(95% C.L.)	$^{67.4}/_{89}$

Table 14.4: Extracted parameters from a three-flavour neutrino oscillation analysis over QSigEx and MCMC results and other solar neutrino experiments. Constraint from KamLAND data is not used for this analysis. For comparison the last line outlines the corresponding result published in the LETA paper. Uncertainties are $\pm\sigma$.

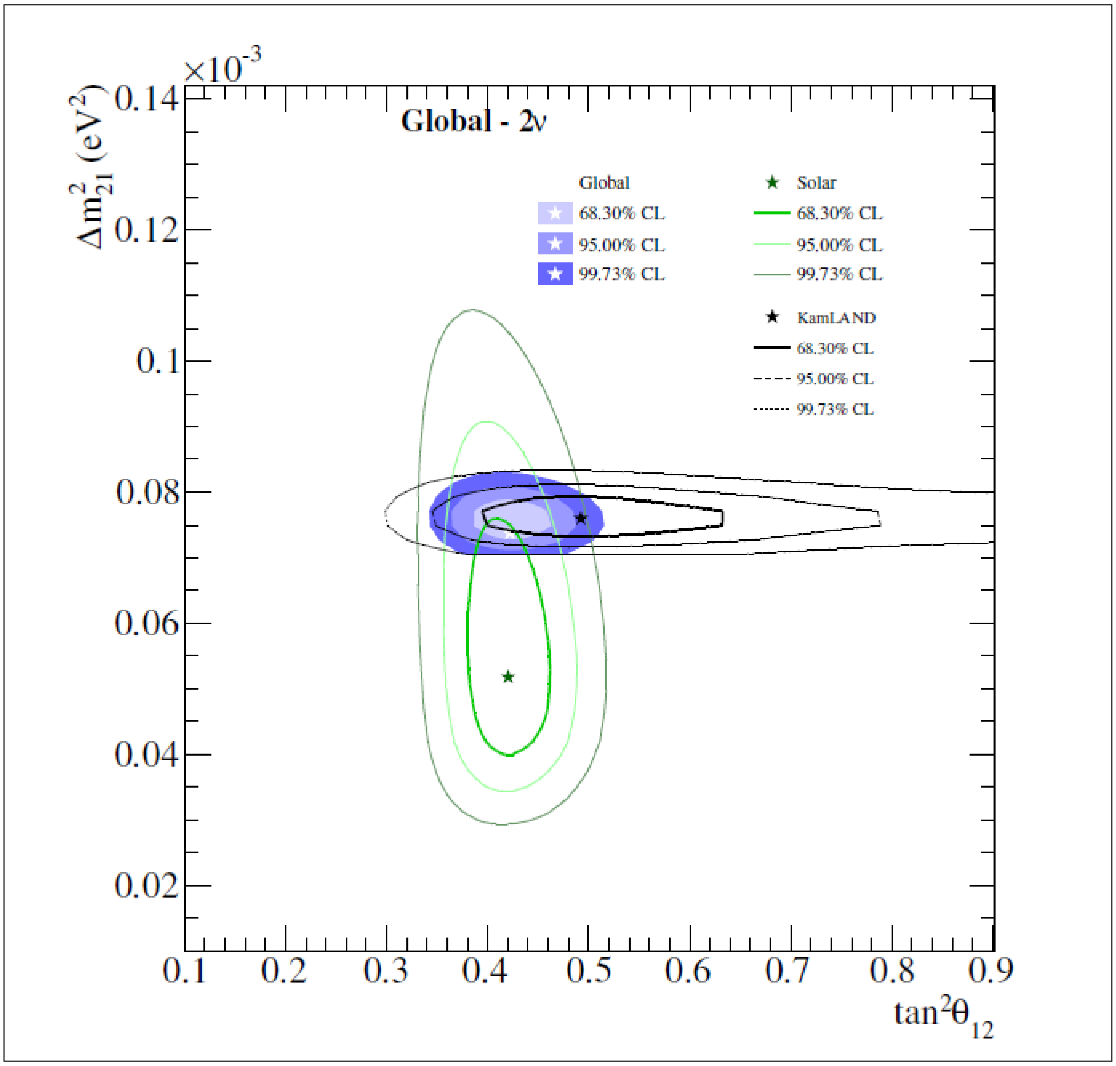


Figure 14.2: Global (all solar + KamLAND) two-flavour oscillation parameter space. Figure from [102].

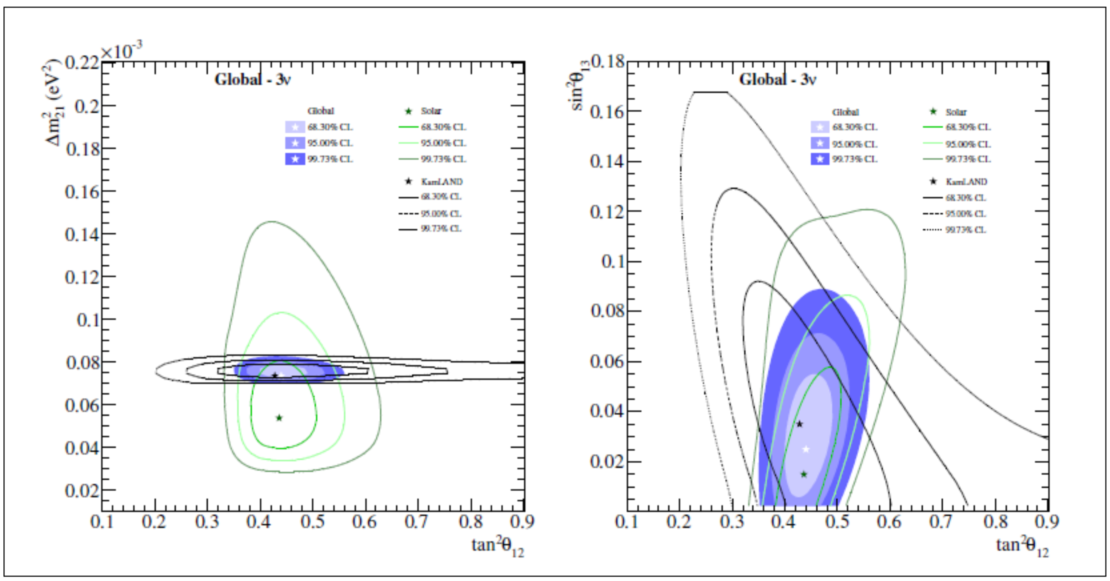


Figure 14.3: Global (all solar + KamLAND) three-flavour oscillation parameter space. Figure from [102].

14.2 Summary

This thesis describes the result of combined three-phase fit using Markov Chain Monte Carlo technique. The ^{8}B flux of $\phi = (5.28 \pm 0.20)$ (stat $\oplus$ syst) $\times 10^6\,\mathrm{cm}^{-2}\,s^{-1}$ measured by the NC reaction where $\oplus$ refers to the total uncertainty including both systematics and statistics. Even with a more conservative application of the systematic uncertainties, the result is comparable to the published LETA result. The CC and ES electron spectra show no signs of spectral distortions. MCMC fitted the energy-dependent ν_e survival probability to the SNO data with the assumption of unitarity of the ν mixing matrix and that the underlying ν spectrum follows a smoothly distorted ^{8}B shape. The survival probability is parametrized as a second-order polynomial and a linear energy-dependent asymmetry between day and night spectra. MCMC fit saw no evidence of either a significant spectral distortion or a significant non-zero day/night asymmetry. The result was used to generate contours showing the allowed regions of the mixing parameters which for "SNO only" result was LOW region. Adding result from other solar experiments and using KamLAND data to constrain the oscillation parameters confined the allowed region to the LMA region. From the 2-flavour fit using MCMC result, the best-fit point in the LMA region is at $\tan^2\theta_{12} = 0.436^{+0.042}_{-0.038}$ and $\Delta m^2_{21} = 5.50^{+2.09}_{-1.33} \times 10^{-5}\,eV^2$.

Bibliography

[1] http://www-sldnt.slac.stanford.edu/alr/standard_model.htm

[2] D. Larson *et al.*, *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP1) Observations: Power Spectra and WMAP-Derived Parameters*, May 28, 2010 arXiv:1001.4635V2, astro-ph.CO

[3] Laura M. Brown, *The Idea of the Neutrino*, Physics Today, **31**, September (1978) Includes an English translation of W. Pauli's December, 1930 letter

[4] Goldhaber M *et al.*1958 Phys. Rev. 109 1015

[5] David Griffiths, *Introduction to Elementary Particles*, John Wiley and Sons, Incorporated (1987) Page 68

[6] J.N. Bahcall, M. Kamionkowski, and A. Sirlin. *Sollar Models: current epoch and time dependences, neutrinos, and helioseismological properties.*, The Astrophysical Journal, Preprint:astro-ph/0010346 on http://arXiv.org, March 2001

[7] http://arxiv.org/PS_cache/astro-ph/pdf/0212/0212331v1.pdf

[8] John N. Bahcall, *Neutrino Astrophysics.*, Cambridge University Press, Cambridge, 1989

[9] R. Davis, *Solar Neutrinos. II. Experimental**, Physical Review Letters, 12 (1964) page 303

[10] SNO Collaboration. *Direct evidence for neutrino flavor transformation from neutral-current interaction in the Sudbury Neutrino Observatory*, Physical Review Letters, 89(011301), 2002

[11] Pasquela D. Serpico, *Cosmological neutrino mass detection: The best probe of neutrino lifetime*, FERMILAB-PUB-07-015-A

[12] John N. Bahcall, Sarbani Basu, M.H.Pinsonnealt, *How uncertain are solar neutrino predictions?*, Physics Letters B 433 (1998) 1-8

[13] http://www.sns.ias.edu/~jnb/SNviewgraphs/Opacities/colorbahcallserenellibs05OP.pdf

J.N.Bachall, A.M.Serenelli, and S.Basu, Astrophys. J. 621, L85 (2005)

[14] A.S.Brun, S.Turck-Chiéze and P. Morel. *Astrophys. J.* 506 (1998)

[15] J.Bonn *et al.*, Nucl. Phys. (Proc. B Suppl.),**91**, 273 (2001)

[16] http://www.sns.ias.edu/~jnb/

[17] Mark Thomson, *Neutrino Oscillations*, University of Cambridge (2002) (published online)

[18] Super-Kamiokande Collaboration, *Evidence for Oscillation of Atmospheric Neutrinos*, Physical Review Letters, 81:1562-1567, 1998

[19] http://hep.bu.edu/ superk/solar.html

[20] The KamLAND Collaboration, *Precision Measurement of Oscillation Parameters by KamLAND*, arXiv:0801.4589v3 [hep-ex]

[21] The SNO Collaboration, The Sudbury Neutrino Observatory Nuclear Instruments and Methods in Physics Research, A449 (2000),

[22] The SNO Collaboration, The Sudbury Neutrino Observatory PACS Numbers: 29.40.Ka, 29.40.Cs, 26.65, 14.60.Pq

[23] http://physicsworld.com/cws/article/news/3259

[24] http://www.lbl.gov/Science-Articles/Archive/NSD-salt-n-SNO.html

[25] http://www.sno.phy.queensu.ca/sno/sno2.html

[26] SNO Collaboration, *Measurement of the Rate of* $\nu_e + d \to p + p + e^-$ *interactions Produced by the* ^8B *Solar Neutrinos at the Sudbury Neutrino Observatory*, Physical Review Letters 87, 071301 (2001).

[27] SNO Collaboration, *Electron Energy Spectra, Fluxes, and Day-Night Asymmetries of* ^8B *Solar Neutrinos from the 391-Day Salt Phase SNO Data Set*, arXiv:nucl-ex/0502021 25 Feb 2005

[28] SNO Collaboration, *Energy energy spectra, fluxes, and day-night asymmetries of* ^8B *solar neutrinos from measurements with NaCl dissolved in the heavy-water detector at the Sudbury Neutrino Observatory*, Physical Review C **72**, 055502 (2005)

[29] SNO Collaboration, *Independent Measurement of the Total Active* ^8B *Solar Neutrino Flux using an Array of 3 He Proportional Counters at the Sudbury Neutrino Observatory*, Physical Review Letters, 101:111301, 2008

[30] Herbert H. Chen, *Direct Approach to Resolve the Solar-Neutrino Problem*, Physical Review Letters, Volume 55 # 14, 30 Sept. 1985

[31] http://www.physics.upenn.edu/balloon/cerenkov_radiation.html

[32] Sudbury Neutrio Observatory - ProposalSNO-87-12, 1987 October

[33] H.H. Staub, *Experimental Nuclear Physics*, volume 1, edited by E. Segré, John Wiley and Sons, New York, (1953), page 72.

[34] *Stopping Powers for Electrons and Positrons*, International Commission on Radiation Units and Measurements, Bethesda, Md, Report 37, (1984)

[35] http://jp.hamamatsu.com/resources/products/etd/eng/html/pmt_001.html

[36] Xiongxin Dai, Etienne Rollin, Alain Bellerive, Cliff Hargrove, David Sinclair, Cathy Miffin, Feng Zhang, *Wavelength shifters for water Čerenkov detector*, Nuclear Instruments and Methods in Physics Research A 589 (2008) 290-295

[37] Glenn F. Knoll, *Radiation Detection and Measurement*, Third Edition, John Wiley and Sons Incorporated, 2000

[38] W.R. Nelson *et al.*, SLAC report **265**, (1985) 1.

[39] Monica Dunford, *Measurement of the ^8B Solar Neutrino Energy Spectrum at the Sudbury Neutrino Observatory.*, PhD Dissertation, University of Pennsylvania, 2006

[40] http://hep.bu.edu/ superk/solar.html

[41] http://hyperphysics.phy-astr.gsu.edu/hbase/particles/neutrino3.html

[42] M.Butler, J.-W. Chen, and X. Kong, *Neutrino-deuteron scattering in effective field theory at next-to-next-to-leading order*, Physical Review C, 63:035501-1-035501-17, February 2001

[43] `http://sunearthday.nasa.gov/2007/locations/ttt_sunlight.php`

[44] arXiv:0810.0466v1 astro-ph

[45] Katsuhiko Sato and Hideyuki Suzuki, *Analysis of neutrino burst from the supernova 1987A in the Large Magellanic Cloud*, Physical Review Letters, 58, 2722 - 2725 (1987)

[46] http://hep.bu.edu/ superk/atmnu/

[47] Shmuel Nussinov and Robert Shrock, *Upper limits on a possible gluon mass*, Phys. Rev. D 82, 034031 (2010)

[48] Nicolas Arnaud *et al.*, *Detection of a close supernova gravitational wave burst in a network of interferometers, neutrino and optical detectors*, Astroparticle Physics, Volume 21, Issue 2, May 2004, Pages 201-221

[49] S.M. Bilinkey, *The History of Neutrino Oscillations*, Physica Scripta. Volume T121, 17-22, 2005

[50] L. Wolfenstein., *Neutrino oscillations in matter*, Physical Review D, 17(9), 1978

[51] S.P. Mikheyev and A.Yu. Smirnov. *Nuovo Cimento*, 9C(17), 1986

[52] Michael B Smy for Super-Kamiokande Collaboration, *Low Energy Neutrino Physics at Super-Kamiokande* Topics in Astroparticle and Underground Physics (TAUP 2009)

[53] Bruno Rossi and David B. Hall,*Variation of the Rate of Decay of Mesotrons with Momentum* Phys. Rev. 59, 223ï£¡228 (1941)

[54] The IceCube Collaboration, *First Year Performance of IceCube Neutrino Telescope*, Astropart.Phys.26:155-173,2006

[55] `http://www.sns.ias.edu/~jnb/SNviewgraphs/Nobel04/colortheoryvsexp.pdf`

[56] Richard James Ford, *Calibration of SNO for the Detection of ^{8}B Neutrinos*, PhD Dissertation, Queen's University, 1999

[57] V. Gribov and B. Pontecorvo, *Neutrino Astronomy and Lepton Charge*, Physics Letters B (1969) 28 (7), p. 493

[58] L. Wolfenstein. *Phys. Rev.*, D17:2369, 1987

[59] M. Apollonio *et al.*(CHOOZ), Phys. Lett. B466, 415:430 (1999)

[60] http://axpd24.pd.infn.it/NO-VE/transparencies/lipari.pdf

[61] SNO Collaboration, *Low-energy-threshold analysis of the Phase I and Phase II data sets of the Sudbury Neutrino Observatory*, Physical Review C 81, 05504 (2010)

[62] K.Zuber, *Neutrino Physics* (2004) Institute of Physics Publishing.

[63] J.N.Bachall, M.H. Pinsonneault, and S. Basu, *Solar Models: current epoch and time dependences, neutrinos, and helioseismological properties*, Preprint:astro-ph/0010346 on http://arXiv.org, March 2001

[64] http://www.physics.carleton.ca/research/sno/introduction/snodetect.html

[65] Helen Mary O'Keefe, *Low Energy Background in the NCD Phase of the Sudbury Neutrino Observatory*, PhD Dissertation, Lincoln College, Oxford 2008

[66] S.R. McGee *et al.*, *Limit on Neutron Production from Activity on NCD Outer Surfaces*, 2008 SNO Internal Report

[67] C.W. Nally, *Atmospheric Neutrino induced backgrounds in the solar neutrino energy range*, SNO Internal Report

[68] A.D. Mario, *Muon-induced spallation products in the SNO detector.* SNO Internal Report, Lawrence Berkley National Laboratory, May 2000

[69] P. Vogel and J.F Beacom, *Angular distribution of neutron inverse beta decay, $\bar{\nu}_e + p \rightarrow e^+ + n$*, Physical Review D **60**, 053003 (1999)

[70] Neil K. McCaulay, *Producing a Background Free Data Set for Measurement of the Charge Current and Day-Night Asymmetry at the Sudbury Neutrino Observatory.*, PhD Dissertation, University of Oxford, 2001

[71] Noah S. Oblath, *A Measurement of Neutral Current Neutrino Interactions at the Sudbury Neutrino Observatory with an Array of ^{3}He Proportional Counters*, PhD Dissertation, University of Washington, 2009

[72] Ryan Martin,*An Analysis of the 3-He Proportional Counter Data from the Sudbury Neutrino Observatory Using Pulse Shape Discrimination*, PhD Dissertation, Queen's University, 2009

[73] Gabriel D. Orebi Gann, *An Improved Measurement of the ^{8}B Solar Neutrino spectrum at the Sudbury Neutrino Observatory*, PhD Dissertation, University of Oxford, 2008

[74] Stanley Reid Seibert, *A Low Energy Measurement of the ^8B Solar Neutrino spectrum at the Sudbury Neutrino Observatory*, PhD Dissertation, University of Texas at Austin, 2008

`http://ewiserver.npl.washington.edu/sno/ncd_ineternal/unidoc/`

[75] http://www-sk.icrr.u-tokyo.ac.jp/sk/index-e.html

[76] Atmospheric neutrino oscillation analysis with sub-leading effects in Super-Kamiokande I, II and IIInew The Super-Kamiokande Collaboration, Phys. Rev. D 81, 092004 (2010) arXiv:1002.3471

[77] http://hep.bu.edu/ superk/atmnu/

[78] The Super-Kamiokande Collaboration, *Study of the atmospheric neutrino flux in the multi-GeV energy range*, hep-ex/9805006 Physical Review Letters, B436(1998)

[79] Metropolis, N., Rosenbluth, A.W, Rosenbluth, M.N., Teller, A.H and Teller, E, *Equations of state calculations by fast computing machine*, J. Chem. Phys., **21**, 1087-1091 (1953)

[80] Gilks, W.R, Richardson, S. and Spiegelhalter, D.J., *Markov Monte Carlo In Practice*, Published by Chapman and Hall, 1996

[81] Roberts, G.O., Gelman, A., Gilks, W.R., *Weak convergence and optimal scaling of random walk Metropolis algorithms*, Ann. Appl. Probab. 7 (1): 110-120 (1997)

[82] Christophe Andrieu, Nnado De Freitas, Arnaud Doucet and Michael I. Jordon *An Introduction to MCMC for Machine Learning*, Machine Learning, 50, 5ï£¡43, 2003

[83] Joanna Dunkley *et al.*, *Fast and reliable MCMC for cosmological parameters estimation*, arXiv:astro-ph/0405462v1, 24 May 2004

[84] Eric B, Ford, *Improving the Efficiency of Markov Chain Monte Carlo for Analysing the orbits of extrasolar planets*, The Astrophysical Journal, 642:505-522, 2006 May 1

[85] Box, G.E.P., and Jenkins,G. *Time Series Analysis: Forecasting and Control*, Holden-Day, 1976

[86] http://root.cern.ch/root/html/TMinuit.html

[87] `http://root.cern.ch/root/html/TF1.html`

[88] `http://root.cern.ch/root/html/TF3.html`

[89] Rice, John (1995), *Mathematical Statistics and Data Analysis*, Second Edition, Duxbury Press, ISBN 0-534-20934-3

[90] Pierre-Luc Drouin, *Three-Phase Extraction of the Electron Neutrino Survival Probability at the Sudbury Neutrino Observatory*, PhD Dissertation to be published, Carleton University

[91] Suggested by Pierre-Luc Drouin in 3 phase meeting in February 2010

[92] The SNO Collaboration, paper to be published

[93] http://root.cern.ch/root/html/TMinuit.html

[94] Geoffrey Jones, *Markov Chain Monte Carlo Estimation for the Two-Component Model*, Technometrics, Volume 46, No. 1, Feb., 2004

[95] Winter *et al.*, *The ^8B Neutrino Energy Spectrum*, PRC 73, 025503 (2006)

[96] Hans A. Bethe, *Energy production in stars*, Physical Review, volume 55, pages 434-456 (1939)

[97] C.F. von Weizsäcker *Über Elementumwandlungen im Innern der Sterne. I (On transformations of elements in the interiors of stars. I)*, Physikalische Zeitschrift, volume 38, pages 176-191 (1937)

[98] C.F. von Weizsäcker *Über Elementumwandlungen im Innern der Sterne. II (On transformations of elements in the interiors of stars. II)*, Physikalische Zeitschrift, volume 39, pages 633-646. (1938)

[99] J.N. Bahchall, H.M.Pinsonneault and S. Basu,*Solar Models: Current Epoch and Time Dependences, Neutrinos, and Helioseismological Properties*, Astrophysical Journal, **555**, 990 (2001)

[100] SuperKamiokande Collaboration, *Solar 8B and hep Neutrino Measurements from 1258 Days of Super-Kamiokande Data*, Physical Review Letters **86**, 2001

[101] Private communications with Pierre-Luc Drouin

[102] Nuno Fluza De Barros, *Precision Measurement of Neutrino Oscillation Parameters: Combined Three-Phase Results of the Sudbury Neutrino Observatory*, Lisbon University, to be published

[103] T2K Collaboration, *Indication of Electron Neutrino Appearance from an Accelerator-produced Off-axis Muon Neutrino Beam*, Physical Review Letters, 2011; (submitted)

[104] DOE/Fermi National Accelerator Laboratory, *Fermilab experiment weighs in on neutrino mystery.*, ScienceDaily, 27 June 2011

[105] Fogli *et al.*, *Evidence of θ_{13} from global neutrino data analysis*, arXiv:1106.6028v1

[106] L.C. Stonehill, *Deployment and Background Characterization of the Sudbury Neutrino Observatory Neutral Current Detectors*, PhD Dissertation, University of Washington, 2005

[107] `http://nobelprize.org/nobel_prizes/physics/laureates/1995/`

[108] `http://nobelprize.org/nobel_prizes/physics/laureates/2002/`

[109] C.M. Lederer, J.M. Hollander, I. Perlman, Table of Isotopes, 6th edition, Wiley & Sons, New York 1968

[110] Keith Rielage, *Proposal to Remove NCD Strings From Final NC Analysis, Version 2*, URL/MANN-6ZBQDN, June 7, 2007

[111] Table 4.4 in [106]

[112] `http://www-pnp.physics.ox.ac.uk/~west/sno/web_snoman/doc/html/companion_frames.html`

[113] Table #3 in *NCD MC Readiness Report for the first NCD Paper*, Version 3.1, April 14, 2008

[114] Page # 145 of [106]

[115] Page #125 of [106]

[116] Figure #7.6 in [106]

[117] Figure #7.8 of [106]

[118] Figure #7.7 of [106]

[119] Alexander Joseph Wright, *Robust Signal Extraction Methods and Monte Carlo Sensitivity Studies for the Sudbury Neutrino Observatory and SNO+ Experiments*, PhD Dissertation, Queen's University, September 2009

[120] `http://www.westgrid.ca/resources_services/compute_facilities`

Appendix A

Determination of Thorium/Uranium content in Neutral Current Detectors (NCD) by Time Coincidence Study

A.1 Introduction

This document describes the time coincidence study employed to quantify the impurity levels in NCDs which consist of ^{232}Th, ^{238}U and all the progeny in their decay chains (See tables A.1 and A.2). The analysis was based on efficiencies determined from Stonewall's simulation [106]. She analysed the combined open[1] and commissioning datasets[2] for her time coincidence study. The later dataset was taken between June 5, 2004 and November 15, 2004 while the former was taken between November 27, 2004 and January 3, 2005. I based my study on the complete NCD data. This study compares the results obtained from analysing the complete data to the results obtained by Stonehill's time coincidence study. This is a cross-check, as a part of box opening procedure, to confirm that the values for the NCD bulk activity from the *in situ* analysis [65] are good.

A.2 Coincidence analysis

A.2.1 Definition of coincidence events

Time coincidence looks for alpha events that are correlated in time. Listed below are three possible coincidence signals, observed in the dataset, that can be utilized to gauge the NCD bulk activity.

- The short-lived alpha emitters are employed for time coincidence study because in the thorium decay chain (Table A.1), the decay of ^{220}Rn ($T_{1/2}$=55.6 seconds) and ^{216}Po ($T_{1/2}$=0.145 second) will be correlated

[1]First month of open data

[2]From now on it will be addressed as combined dataset.

in time. Hence it is possible to define a triple coincidence between the alphas from the ^{224}Ra → ^{220}Rn → ^{216}Po portion of the ^{232}Th decay chain. A triple coincidence is defined as two events with a time difference of less than 111.2 seconds followed by a third event within 0.3 second. All events with energy less than 0.1 MeV were rejected to reduce electronic noise. The sizes of the coincident windows are selected to be twice the half-lives - $T_{1/2}$=55.6 seconds (^{220}Rn) and $T_{1/2}$=0.145 second (^{216}Po) - in thorium decay chain.

- Short double coincident pair consist of ^{220}Rn and ^{216}Po alphas. Time window selected for the analysis is 0.3 second, twice the half-life of ^{216}Po. The $\alpha - \alpha$ coincidence of ^{216}Po also make up the double. Events were selected if the energy deposited is greater than 0.1 MeV.

- Long double coincidence is when two events, from both the uranium chain (Table A.2) and the thorium chain, happen in a coincidence window of 111.2 seconds. To reduce accidental coincidences between low-energy events, the energy threshold was increased to 1 MeV.

Nuclide	decay mode	$T_{1/2}$	Energy (MeV)	Decay Product
^{232}Th	α	1.405×10^{10} a	4.081	^{228}Ra
^{228}Ra	β^-	5.75 a	0.046	^{228}Ac
^{228}Ac	β^-	6.25 h	2.124	^{228}Th
^{228}Th	α	1.9116 a	5.520	^{224}Ra
^{224}Ra	α	3.6319 d	5.789	^{220}Rn
^{220}**Rn**	α	**55.6 s**	**6.404**	^{216}**Po**
^{216}**Po**	α	**0.145 s**	**6.906**	^{212}**Pb**
^{212}Pb	β^-	10.64 h	0.570	^{212}Bi
^{212}Bi	β^- 64.06%	60.55 min	2.252	^{212}Po
	α 35.94%		6.208	^{208}Tl
^{212}Po	α	299 ns	8.955	^{208}Pb
^{208}Tl	β^-	3.053 min	4.999	^{208}Pb
^{208}Pb	stable			

Table A.1: Thorium Series [109]

A.2.2 Chance Coincidence

Accidental/chance coincidence is defined as uncorrelated events that happen close enough in time as to fall within the coincidence window. The method to compute the expected number of accidentals was described in Stonehill's dissertation [114]. For triples, the number of accidentals was calculated according to equation (A.1).

Nuclide	decay mode	$T_{1/2}$	Energy (MeV)	Decay Product
^{238}U	α	4.468×10^{9} a	4.270	^{234}Th
^{234}Th	β^-	24.10 d	0.273	^{234}Pa
^{234}Pa	β^-	6.70 h	2.197	^{234}U
^{234}U	α	245500 a	4.859	^{230}Th
^{230}Th	α	75380 a	4.770	^{226}Ra
^{226}Ra	α	1602 a	4.871	^{222}Rn
^{222}**Rn**	α	**3.8235 d**	**5.590**	^{218}**Po**
^{218}Po	α 99.98%	3.10 min	6.115	^{214}Pb
	β^- 0.02%		0.265	^{218}At
^{218}At	α 99.90%	1.5 s	6.874	^{214}Bi
	β^- 0.1%		2.883	^{218}Rn
^{218}Rn	α	35 ms	7.263	^{214}Po
^{214}Pb	β^-	26.8 min	1.024	^{214}Bi
^{214}Bi	β^- 99.98%	19.9 min	3.272	^{214}Po
	α 0.02%		5.617	^{210}Tl
^{214}Po	α	0.1643 ms	7.883	^{210}Pb
^{210}Tl	β^-	1.30 min	5.484	^{210}Pb
^{210}Pb	β^-	22.3 a	0.064	^{210}Bi
^{210}Bi	β^- 99.99987 %	5.013 d	1.426	^{210}Po
	α 0.00013%		5.982	^{206}Tl
^{210}Po	α	138.376 d	5.407	^{206}Pb
^{206}Tl	β^-	4.199 min	1.533	^{206}Pb
^{206}Pb	stable			

Table A.2: Uranium Series [109]

$$N_a = r_s^3\, t_1\, t_2\, L \tag{A.1}$$

where r_s is the singles rate on that string, t_1 and t_2 are the sizes of the coincidence windows – 0.3 and 111.2 seconds respectively – and L is the lifetime[3] of the dataset. For the doubles, the equation for the number of accidentals was $N_a = r_s^2\, t\, L$ where t is the size of the coincident window. These equations are valid provided that the single's rate times the size of a coincidence window is small compared to unity ($r_s t \ll 1$). To get the correct count of the coincident events (N_c) in a string, the accidentals (N_a) were subtracted from the number

[3]The time that the detector was actively collecting data.

of observed coincident events (N_{oc}) and if accidentals were more than the coincident events, the number of coincident events in the string was set equal to zero. This happened only with string #23.

$$N_c = N_{oc} - N_a \tag{A.2}$$

$$N_a > N_{oc} \text{ then } N_{oc} = 0 \tag{A.3}$$

A.2.3 All strings or good strings only?

For comparison, two studies were performed, once with all the strings and then with certain strings (0, 1, 3, 10, 18, 20, 26, 27, 30, and 31) removed ([110] and [111]). The later will be known as restricted study from now on. Stonehill did not analyse strings 3, 7, 8 and 20, hence her study was also restricted. Even though not all strings were included, the numbers reported in column one of Table A.11 from the restricted study were corrected to represent the full array with the assumption that coincident events are fairly distributed over all the strings that were analysed.

A.2.4 True coincidences

Selection of coincidence candidate was based on the time difference between each event in the dataset and its immediate predecessor. if the time difference is within the size of a time window then the events are considered as the coincidence events. Accidental coincidences are backgrounds in this study. The number of true coincidences is calculated by subtracting the expected numbers of accidentals from the observed number of coincidences (equation (A.2)). Table A.3 is the account of coincident and accidental events. The *italics* text describes Stonehill's study (lifetime of the dataset analysed, number of coincidences and number of chance coincidences) and the rest describes this study. The column four/six of the table suggests that long double (111.2 seconds) is the least accurate coincident event to estimate the contamination in NCDs [See Figure A.1]. Corrected number (N_c shown in red in Figure A.1 and described by equation (A.2).) is the observed number (N_{oc} shown in blue in Figure A.1) of coincident events minus the accidentals (N_a). Considering the accidentals, it seems that triple-coincident event is the most accurate one. The reason for no accidentals for triples is due to a tight constrain: (1) two coincident events within 111.2 seconds of each other followed closely by a third event within 0.3 second (2) all three events occurring in the same string and (3) all three events depositing energy in excess of 0.1 MeV. The tight constrain also limits the statistics. Short doubles (0.3 second), on the other hand, has limited accidentals on account of a narrow time window and has good statistics (349 ± 1.43).

A.3 Data Cleaning Cuts

Table A.4 lists the cuts applied on the analysis to remove non-physics events which include high-voltage discharges evident in some NCD sections. These discharges were introduced inadvertently by the welding process. Even though steps were undertaken to resolve the situation, some strings still have these problems, hence these were removed from the analysis. The cuts applied by Stonehill are not similar to the cuts that were applied for this study, because

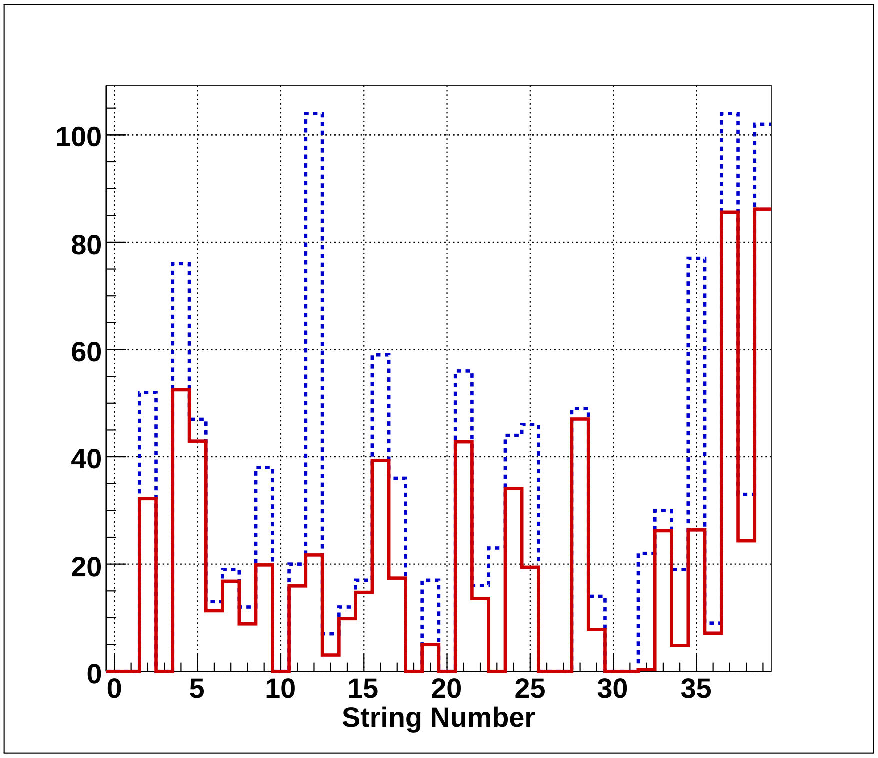


Figure A.1: Distribution of uncorrected and corrected number of long (111.2 seconds) doubles. Strings, not shown, (0, 1, 3, 10, 18, 20, 26, 27, 30 and 31) were not included.(Restricted Study).

the cuts were still work-in-progress when Stonehill performed her analysis. Tables A.13 and A.14 break down the Data Analysis Mask Numbers (DAMN cuts) reported in Table A.4 into bits. For a detailed description of DAMN cuts, consult [112].

A.4 Fits to obtain lifetimes $T_{1/2}$=55.6 seconds (^{220}Rn) and $T_{1/2}$=0.145 second (^{216}Po)

The time difference between Poisson-distributed events was fitted with exponential distribution to see if the lifetime from the fit was the half-life of ^{216}Po (0.145 second) and ^{220}Rn (55.6 seconds). Figure A.2 shows the fit result for short doubles when the size of window is extended to 1.5 seconds from 0.3 second. It shows with χ^2 19.62 and Number of Degrees of Freedom (ndf) 21, that the half-life is 0.1358 ± 0.0089 which is 1.0 σ away from 0.145 second. Figure A.3 shows a fit on the long doubles with χ^2 of 38 for 45 degrees of freedom. Lifetime from the fit is 47.25 ± 11.64 which is 0.7 σ away from 55.6 seconds. To avoid $\alpha - \alpha$ coincidences from ^{216}Po ($T_{1/2}$=0.145-second) in the thorium chain, the range for the fit was restricted between 7.0-222.45 seconds. Figures A.2 and A.3 also display the equation that was fitted to the data. The goodness of a fit is described by the χ^2/ndf but to ascertain it, the fitted background (41.21±3.44) was compared to the accidentals for the long doubles of 222.2-seconds window size. From the fit (Figure A.3), the background comes up to be in the range of 1889 to 2233 [(41.21-3.44)=37.77 × 50 to (41.21+3.44)=44.65 × 50 where 50 is the number of bins] which compares very well with the number of accidentals, 2126, calculated from Poisson distribution.

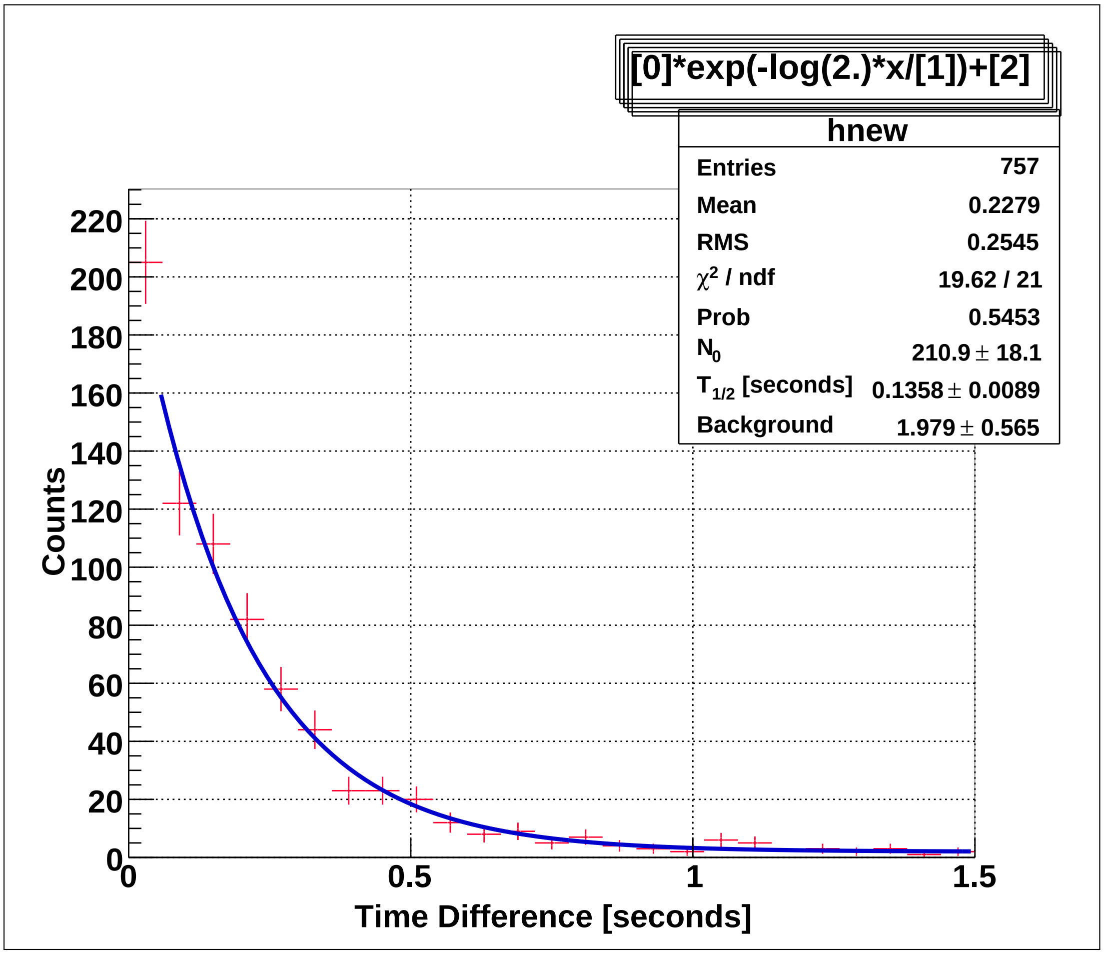


Figure A.2: Time difference distribution of short doubles, extended to 1.5 seconds, yields half-life of ^{216}Po. All strings were included in the analysis.

A.5 Energy Distribution

To verify furthermore the fact that triples and short doubles are short-lived alpha emitters from thorium decay chain, the energy distribution of the events is plotted in Figures A.5 and A.7. The initial energies of ^{222}Ra (From Table A.2), ^{220}Rn and ^{216}Po (From Table A.1), involved in the triples and short doubles, are 5.59 MeV, 6.40 MeV and 6.90 MeV respectively. The energy distributions, as seen in Figures (A.4, A.5, A.6, A.7) range from zero to the initial (maximum energy in the figure) energies of alphas. From Figures A.4 and A.6 it appears

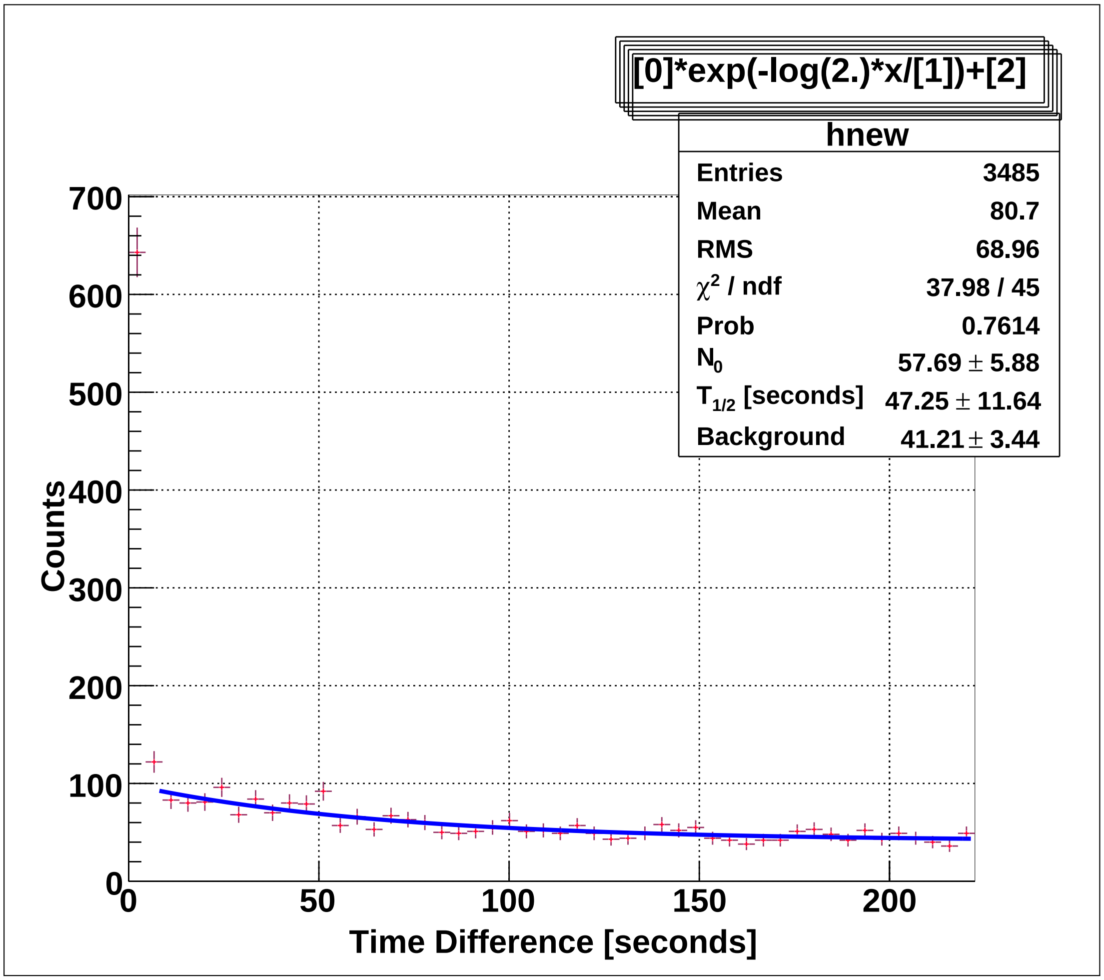


Figure A.3: Time difference distribution of long doubles, extended to 220 seconds, yields half-life of ^{220}Rn.

that the impurity is a bulk impurity and not a surface contamination. Alphas starting within the bulk will lose energy due to collisions within nickel walls while the one originating from the surface enters into NCD active volume with its energy intact. The spectra A.5 and A.8 show that there are significant number of events that are not consistent with bulk-like activity.

A.6 Model dependency of impurity composition [65]

This section touches the subject of model dependency of impurity level calculated from time coincidence study. The energy distribution in Figures A.5 and A.7 is very informative and indicative of the combination of following factors:

1. Alphas originating deeper in the NCD walls and loosing energy escaping from the nickel walls so that the active NCD region detects only partial energy (Bulk model).

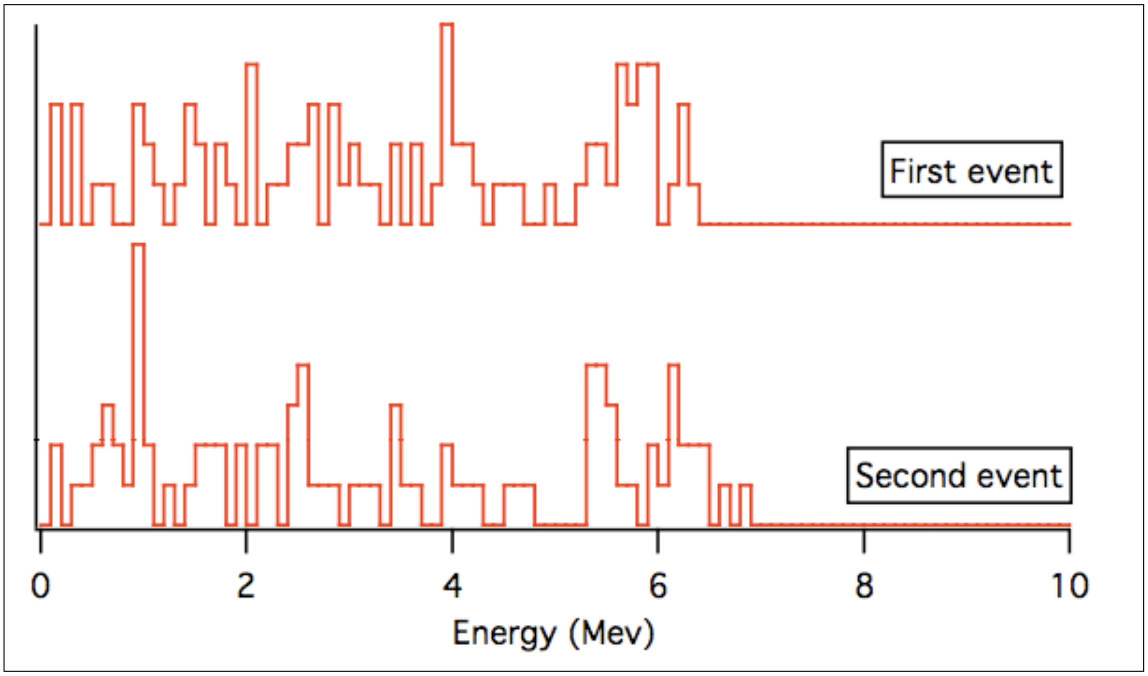


Figure A.4: Energy distribution of the first and second event in the 89 0.3 second double coincidences from Stonehill's analysis [117].

2. Alphas originating from or near the surface (Surface-fixed model) and depositing the entire energy in the NCD active region (see the peaks in the Figures A.5 and A.7) or

3. Alphas originating from or near the surface (Surface-fixed model) but striking another part of the wall and depositing only partial energy, thereby mimicking the bulk-like activity. See the tails of the peaks in the Figures A.5 and A.7.

4. One possibility might be the escape of radon (a noble gas) into the active region of the NCDs which alters the potential prospects for observing coincidence. Ejection of the ^{220}Rn in the gas will result in its decay in the gas since it has a 56 s half life. Thus ^{220}Rn escape model will result in higher detection efficiency of ^{220}Rn and its daughter 0.145 s ^{216}Po compared to the surface-fixed model. In this scenario, the leading alpha will not be observed but the two alphas in the double will definitely be observed.

5. Last possibility is that the origin of alpha is the surface contamination which has a non negligible thickness, consequently the energy peak is smeared out and gives the appearance of true bulk activity.

The ratio of number of short doubles to triples gives an indication as to whether the activity originates from the surface or is uniformly distributed in the bulk of the walls. If the alpha starts from the surface then there are two possibilities for it to travel - towards (T) the active region or away (A) from it. Hence, in the case of triples, there are $2 \times 2 \times 2 = 8$ possibilities [(1) AAA,(2) AAT,(3)

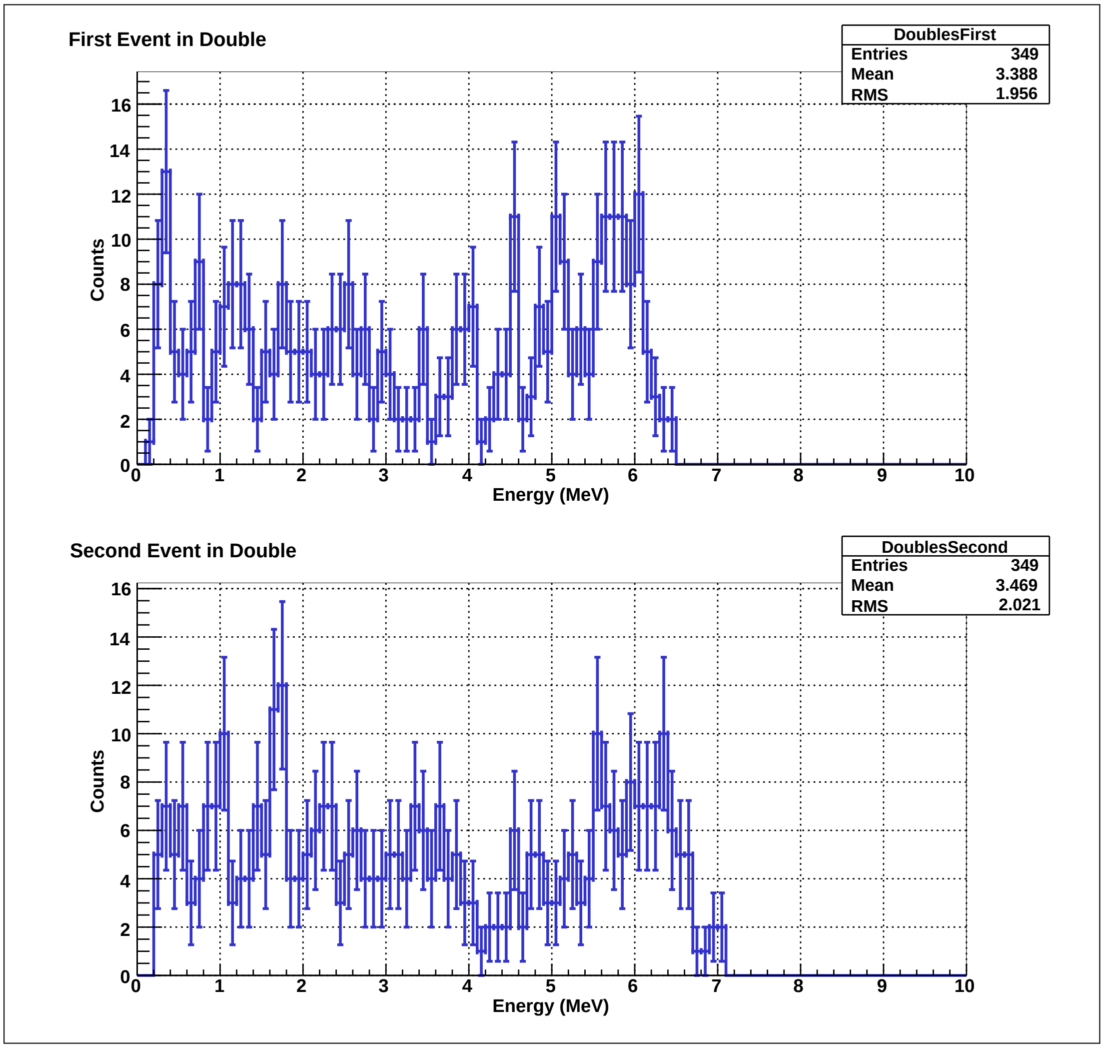


Figure A.5: Energy distribution of the first and second event in the 349 0.3 second double coincidences.

ATT,(4) TTT,(5) TTA,(6) TAA,(7) TAT,(8) ATA] therefore the probability of detection is $(1/8)\times 100 = 12.5\%$. Out of eight possible outcome, only one will be observed as a triple and another will be a double missing a leading alpha. In the case of short doubles there are $2 \times 2 = 4$ possibilities therefore there is a $(1/4)\times 100 = 25\%$ probability of detecting both the ^{220}Rn and ^{216}Po alphas that constitute the short doubles. Hence the ratio of triples to doubles is $12.5/25.0 = 1/2 = \mathbf{0.5}$ in case the impurity is fixed on the surface.

If the impurity is bulk in nature then there are four possibilities for a alpha to travel[4]. Consequently there are $4 \times 4 \times 4 = 64$ and $4 \times 4 = 16$ set of possibilities for a triple and a double coincident pair to be detected respectively. Therefore, for bulk contamination, the ratio of triple to short

[4]If 100 alphas are uniformly distributed to a depth equal to an α range then 25 alphas will escape the wall. Detection efficiency for a single α from the bulk activity is 25% as opposed to 50% from the surface activity.

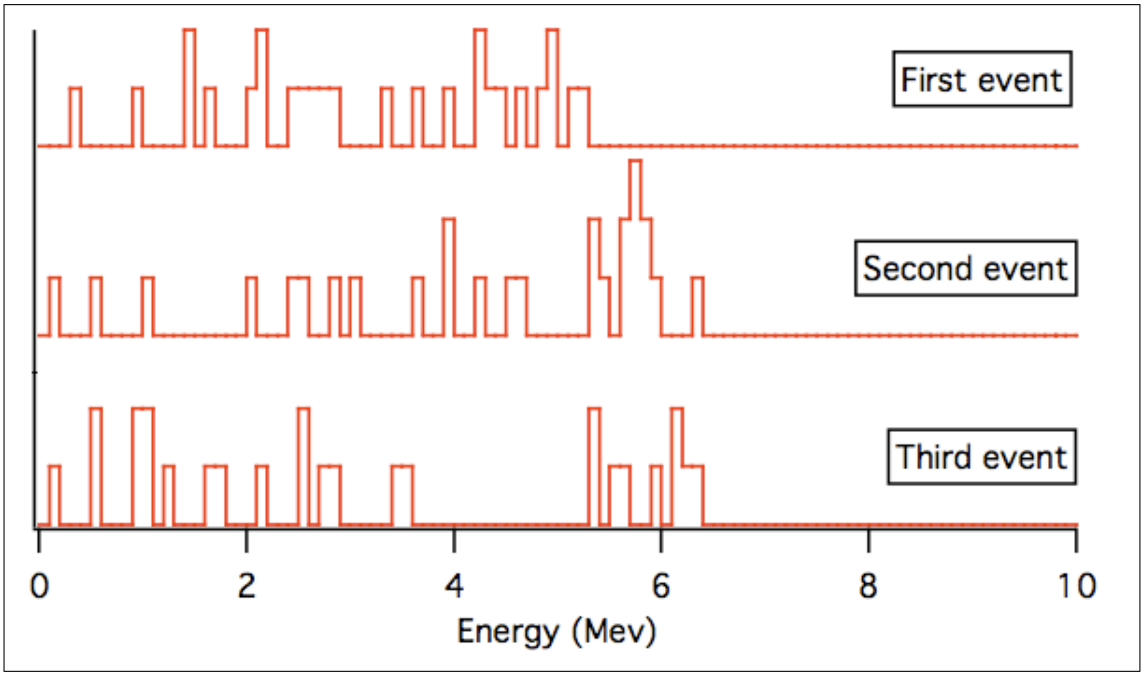


Figure A.6: Energy distribution in the first, second and the third event in 26 triple coincidences from Stonehill's analysis [116].

double would be $\frac{1/64}{1/16} = 1/4 = \mathbf{0.25}$.

As mentioned earlier, in the radon escape model, there is 50% chance that the leading alpha will be observed and the resulting daughter ^{220}Rn will be on/near the surface. The odds of observing a double, originating from the surface, is 25% hence the probability of observing a triple is $(50\% \times 25\%) \times 100 = 12.5\%$. There are two components in the calculation of the likelihood of observing a double in the radon escape model. First when radon's track is towards and the other in which it is away from the active region. In the former case radon escapes into the active region therefore the double ($^{220}Rn \rightarrow {}^{216}Po \rightarrow {}^{210}Pb$) will always be observed and in the later (50% of the time) there is 25% chance of observing a double hence the probability is summed as $(50\% + 50\% \times 25\%) \times 100.0 = 62.5\%$. Therefore the ratio of triples to doubles is 12.5%/62.5%=0.2 for the radon escape model. From the last line of Table A.5, it seems that the ratio spans from 0.2 to 0.5. The efficiency to detect a coincident pair spans from 6.25% to 62.5% – a ten fold change – resulting in a wide deviation in the amount of ^{232}Th derived from the doubles measurement.

Table A.5 summarizes the above discussion. The ratio of triples to doubles is 0.25 (bulk activity), 0.5 (surface activity) and 0.2 (radon escape model). Table A.9 shows that the ratio ranges from $167/575 \approx 0.29$ (All Strings) to $106/349 = 0.30$ (Restricted) instead of 0.25 or 0.5 which means that both bulk and surface-fixed models play a role. The efficiencies discussed here are based on geometric arguments and do not include efficiency ϵ of coincident window to detect the coincident pair.

$$\mathrm{Eff} = 1.0 - e^{-t\,\log(2)/T_{1/2}} \tag{A.4}$$

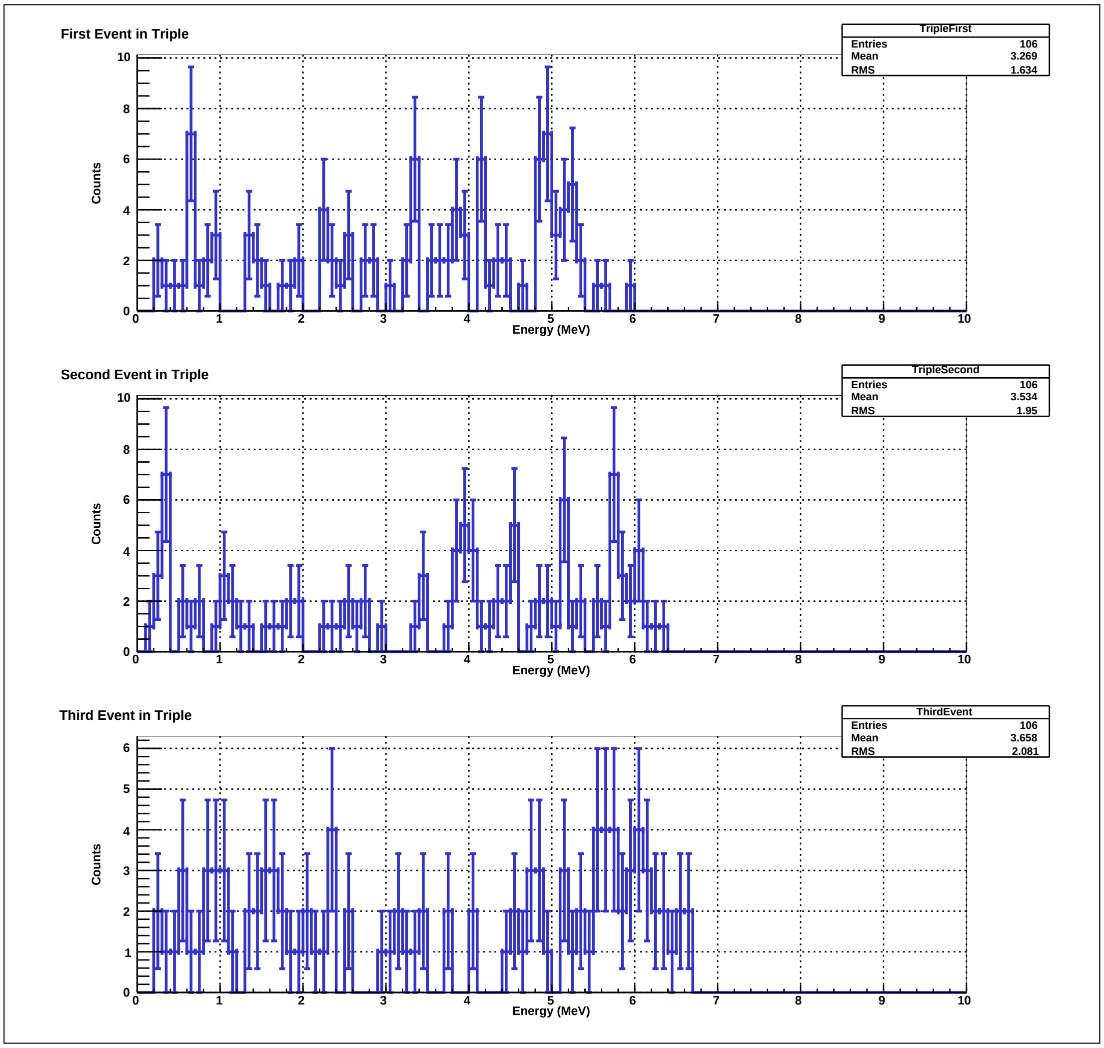


Figure A.7: Energy distribution in the first, second and the third event in 106 triple coincidences.

where t=0.3 seconds is size of the coincident window and $T_{1/2}$=0.15 seconds is the half-life of ^{216}Po. Taking into account efficiency of the coincident window ϵ and the ratio of the number of triples to doubles, the number of triples is calculated and compared to the observed number of triples to figure out the model that best describes the data. Table A.6 lists the result of the computation when all strings are analysed and when only good strings are considered.

$$N_{tpl} = \text{Ratio} \times N_{dbl} \times \text{Eff} \tag{A.5}$$

where Ratio (from last row of Table A.5) is a ratio of number of triples to number of short double coincident pairs obtained from geometrical consideration, N_{dbl} is 349 from restricted study and 575 when all the strings are considered in the fit, and N_{tpl} calculated is shown in columns 3 to 5 in Table A.6.

From Table A.6 it appears that the impurity is combination of bulk and surface models, hence ignoring radon escape model, the composition of impu-

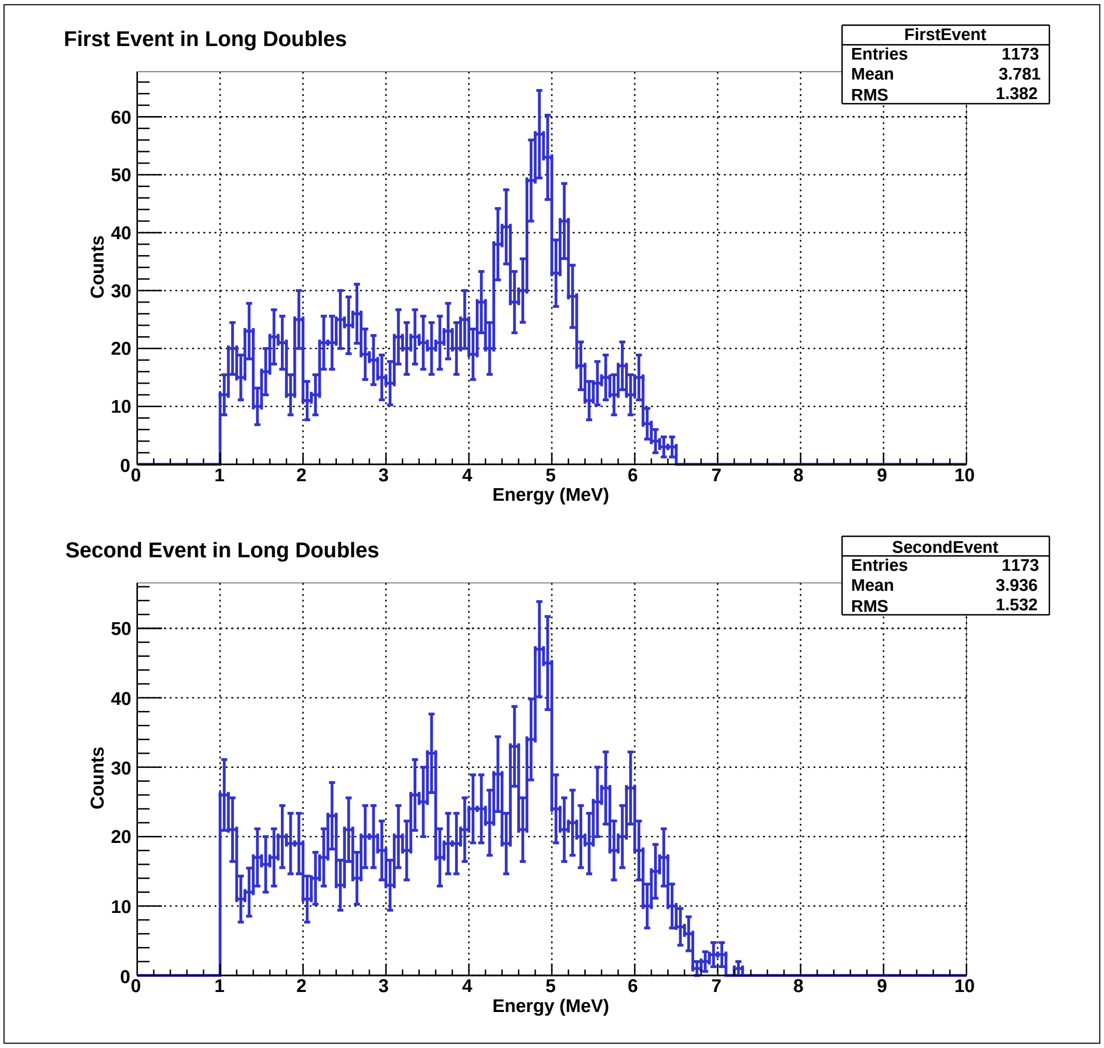


Figure A.8: Energy distribution of the first and second event in the 1173 111.2 second double coincidences.

rity into bulk **y** and surface **x** is determined as:

$$x + y = 1 \tag{A.6}$$

$$130x + 65(1 - x) = 106;\ x \simeq 0.63 \tag{A.7}$$

$$215.6x + 107(1 - x) = 167;\ x \simeq 0.55 \tag{A.8}$$

The values are used from Table A.6. Equation (A.7) is for the restricted study involving only good strings and equation (A.8) is for all strings. The average $(0.63 + 0.55) * 0.5 \times 100 = 59\%$ agrees very well with 58% of [113].

A.6.1 Features of Simulation

For the simulation, 10 million decays were simulated each for uranium and thorium. The alpha emitted by ^{212}Po in the thorium chain is the highest energy alpha with a range of 20 μm in nickel, hence to reduce the computational

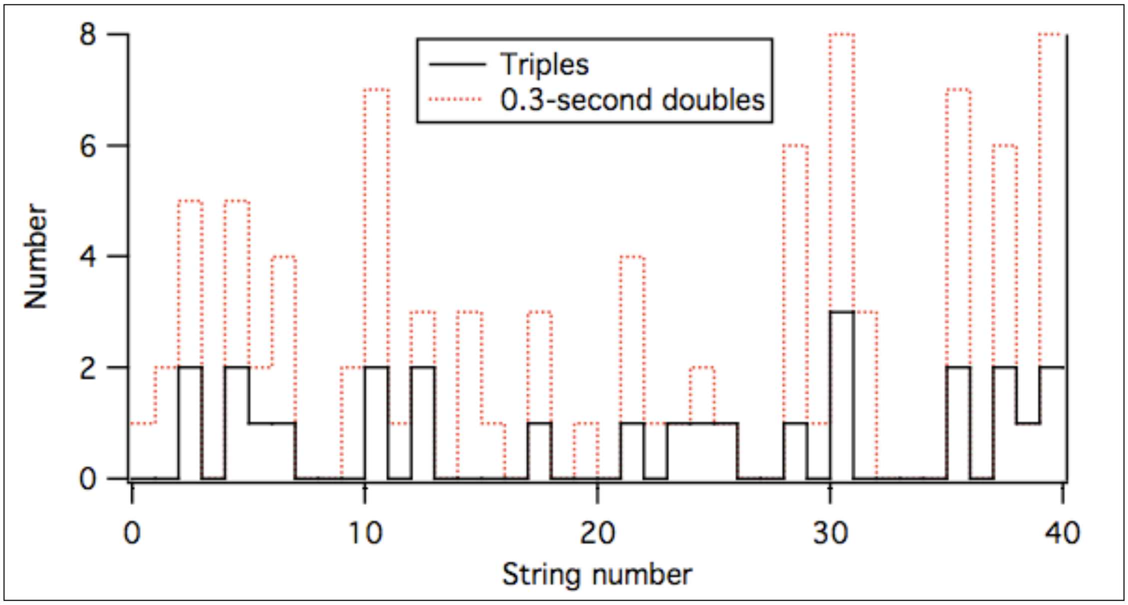


Figure A.9: String distribution of triples and short doubles from Stonehill's analysis. Strings, not shown, (3, 7, 18 and 20) were not analysed [118].

time, alphas were uniformly distributed up to 22 μm in the bulk of the nickel walls. Figures A.4 and A.6 testifies to the validity of this assumption but not Figures A.5 and A.7. In the simulation, the uranium and thorium chains were broken at radium because it is readily dissolved in water [115]. Disequilibrium model was used because of the complication of the CVD (Chemical-Vapour-Deposited) process to deposit nickel on aluminium mandrel to make NCD bodies. The equilibrium was also disturbed by electropolishing and acid etching of the nickel tubes to remove adhered ^{210}Po from the surfaces of nickel tubes.

A.7 Calculation of Thorium/Uranium Content

To calculate the impurity level, accidentals were calculated, using equation (A.1), for each string based on the singles data above 0.1 MeV for triples and short doubles and above 1 MeV for long doubles. Using the number of accidentals, the rate of impurity level R from a coincident event, is expressed as:

$$R = \frac{(N - N_a)}{(A\,L\,f_{\text{scope}}\,\varepsilon)} \tag{A.9}$$

where N and N_a are the numbers of coincident events and accidentals respectively, L, as mentioned earlier, is the lifetime of the dataset, f_{scope} is scope life fraction, and ε is the efficiency from Table A.7. The simulation efficiency ε is the probability of seeing a double or a triple coincidence from a decay chain. For example, in the case of triples, out of 100 simulated events in bulk thorium, only 0.961 triple coincidence was observed. The lifetime was multiplied by scope life fraction (From [119] 95.9±0.3(stat)±0.01(sys))% to account for the deadtime of the scope, which is relevant[5], because the analysis considers only correlated Shaper/ADC and scope events. Area[6] is calculated from the number of stings N_s in which the number of coincident events is greater than zero – $A = (\frac{63.52}{40.0})\, N_s$. Substituting in equation (A.9) results in

$$R = \frac{40\ (N - N_a)}{(63.52\, N_s\, L\, f_{scope}\, \varepsilon)} \tag{A.10}$$

This equation also extends the result obtained from the restricted analysis to the full array of NCDs assuming that the coincident events are distributed over all the strings that were analysed (See Figure A.10). The exception is string #37 in which 43 short doubles and 14 triples occurred – well above the average of 15 and 5 for short doubles and triples respectively. String #37 has problems that were mentioned in [110]. String #37 also has more than its share of long doubles (See Figure A.1) from the rest of the strings. The result of the calculations is shown in column one of Table A.11. As seen in Table A.7, thorium contamination was estimated by counting all three coincident events but the rate of uranium decay was estimated by counting the 111.2 seconds doubles which can have its origin either in thorium or uranium decay chain.

Sources of triples and short doubles are alphas from thorium chain, but long coincident pairs (111.2 seconds) are complicated. There are three sources, that is, alphas from uranium (^{218}Po), alphas from thorium (^{220}Rn) and accidentals. In order to distinguish the contributions from each source, Stonehill applied simultaneous ROOT maximum-likelihood fit to the energy distributions of first and second events and obtained the result shown in Table A.8. Alphas from uranium and thorium is calculated as:

$$N_x = \frac{N_{ld}\, f_x}{\epsilon_x} \tag{A.11}$$

where N_x is the number of atoms of x which can be either ^{218}Po or ^{220}Rn, f_x (from Table A.8) is the fraction of long doubles belonging to uranium (^{218}Po) or thorium (^{220}Rn) or accidentals and ϵ_x (from Table A.7) is an efficiency of seeing a double coincidence where two alphas deposit energy that exceeds 1 MeV and the time difference is within 111.2 seconds.

Number of observed long doubles N_{ld} in the dataset – in terms of N_x calculated from equation (A.11) – is expressed as:

$$N_{ld} = \epsilon_{218}\, N_{218} + \epsilon_{220}\, N_{220} + N_a \tag{A.12}$$

Table A.11 compares the thorium contents determined from the complete dataset to the ones from Stonehill's study.

A.7.1 Calculation of Errors

Statistical errors were calculated based on the $\sqrt{N}$ statistics. Various systematic uncertainties enter the analysis. List include sources:

- The uncertainty in the lifetime (385.17 ± 0.14 [29]) of the dataset.
- Uncertainty from the scope life fraction [95.9±0.3(stat)±0.01(sys)]%

[5] The MUX deadtime is $\sim$ 1 millisecond after every event.

[6] There are forty NCD strings with total surface area equal to 63.52 m^2.

- Fractional uncertainty in the determination of fractions of uranium, thorium and accidentals in the long doubles (Table A.8).
- The uncertainties in simulation efficiencies from Table A.7.
- The dependence of time coincidence analysis on energy calibration, due to energy threshold for the selection of events (0.1 MeV for the short doubles and triples and 1.0 MeV for the long doubles), leads to a systematic uncertainty.
- The use of three *different* coincidence methods to estimate the level of thorium impurity introduces a significant uncertainty and since the three dissimilar methods extracted the results from the similar dataset, the uncertainties are also correlated.

Both systematic and statistic errors were added in quadrature to get the total uncertainty reported in tables A.11, A.10 and A.12.

A.7.2 Inconsistency in column one of Table A.11

Stonehill's calculation of thorium contaminations from all three events are consistent with each other. The three row entries are not consistent in column one of Table A.11, nonetheless, from the last row, it seems the average rates from both columns are consistent within uncertainties of each other. The discrepancy is due to the fact that bulk model does not represent the complete data but adequately represents the dataset that Stonehill analysed.

A.7.3 Cross-check

As a cross-check comparison of the impurity levels, when all strings are considered and when hot strings are removed (Table A.10) was made and it shows that the count rate is higher for the former because of the inclusion of string #10 (Figure A.11) which has higher number of events and accidentals than any other string. This is what we expected. Nevertheless, the average rate (row three in Table A.10) of decay from triples and short doubles from all strings (0.795 ± 0.067) is consistent within uncertainties with the one calculated from the restricted dataset (0.680 ± 0.067).

A.7.4 Thorium/Uranium Content

The number of short doubles observed in the NCD array is 575$\pm$24 (Table A.9). With the efficiency of 3.557%, appropriate for bulk-like activity, the number of ^{232}Th decays is 16165$\pm$675. This decay would be produced[7] by 0.083$\pm$0.005 μg of ^{232}Th in the NCD walls. If, instead, the efficiency for surface-like activity (18.75%) is used in the above calculations, then the decay would be produced by $0.023 \pm 0.0009\,\mu$g of ^{232}Th on the inner surfaces of the NCD walls. The spread reflects uncertainty in the efficiency (model dependency) to convert the activity to actual Thorium content.

[7] $N_0 = N_d/(1-2^{-t/T_{1/2}})$ where N_d is the number of decays of ^{232}Th or ^{238}U and t is the lifetime of the dataset. Mass is related to the number of nuclei by $m = (m_A/N_A)\,N_0$ - m_A is 232 or 238 gm and N_A is the Avagadro number 6.02×10^{23}.

The average of the decay rates of uranium from the complete NCD data (All strings and restricted to good strings) in Table A.12 was 0.527 ± 0.765 which is consistent with the decay rate that Stonehill obtained ($0.50\,^{0.53}_{0.39}$). To calculate the actual uranium content, responsible for the observed activity, the number of long doubles – 2314 (Table A.9) – is used. Following the calculation mentioned above with the bulk model, gives the uranium content as $0.034 \pm 0.034\,\mu$g.

A.8 Conclusion

Using the bulk model of impurity in the time coincidence study, the impurity level of thorium – in the NCD array – from the complete dataset (0.57 ± 0.09 decays/m^2/day) is consistent within the uncertainties with the value obtained by Stonehill ($0.47\pm$ 0.12 decays/m^2/day). The study considered various models of contamination and 63% surface and 37% bulk best fitted the number of triples observed. Based on the time coincidence study, the NCD array contains between 0.083 ± 0.005 to $0.023 \pm 0.0009\,\mu$g of ^{232}Th. Concerning the uncertainties in the uranium calculation, the impurity in the NCD array is dominated by thorium.

Events	**Exposure area × time (m^2 days)**	**Counts N_{oc}**	**Accidentals N_a**	**Exposure area × time (m^2 days)**	**Counts N_{oc}**	**Accidentals N_a**
Triples		*26*	*0.023*	14680	106	0.015
Short Doubles	*5457*	*89*	*0.9*	17738	349	1.43
Long Doubles		*444*	233 ± 14	17738	1173	436

Table A.3: Comparing number of coincident events (observed N_{oc} and chance N_a).

Cut	Applied By Stonehill	Applied for this study
DAMN 0	0XC8440001	0XCC440001
DAMN 1	0X1BE	0X00FFDDFE

Table A.4: Cuts applied. See tables A.13 and A.14 for further details.

	Bulk %	**Surface %**	**Radon Escape %**
Efficiency of detecting triples	100/64=1.56	12.5	12.5
Efficiency of detecting doubles	100/16=6.25	25.0	62.5
Ratio of triples to doubles	0.25	0.5	0.2

Table A.5: Ratio of triples to doubles for different models of contamination.

	Observed	**Surface**	**Bulk**	**Radon Escape**
All Strings	167	215.6	107	86.25
Restricted	106	130	65	52.35

Table A.6: Comparison of observed and expected number of triples from three different models.

Events	**Efficiency ε %**
Thorium triple	0.961 ± 0.003
Thorium short doubles	3.557 ± 0.006
Thorium long double ϵ_{220}	8.606 ± 0.009
Uranium long double ϵ_{218}	1.164 ± 0.003

Table A.7: Efficiencies from Monte Carlo simulation performed by Laura C. Stonehill [106] and employed in equations (A.10) and (A.11).

	Counts	**Fraction of long doubles**
Accidentals f_a	211 ± 25	0.47 ± 0.06
Thorium f_{220}	201 ± 27	0.45 ± 0.06
Uranium f_{218}	32 ± 33	0.07 ± 0.07

Table A.8: Outcome of Stonehill's maximum-likelihood fits to determine the composition of long doubles.

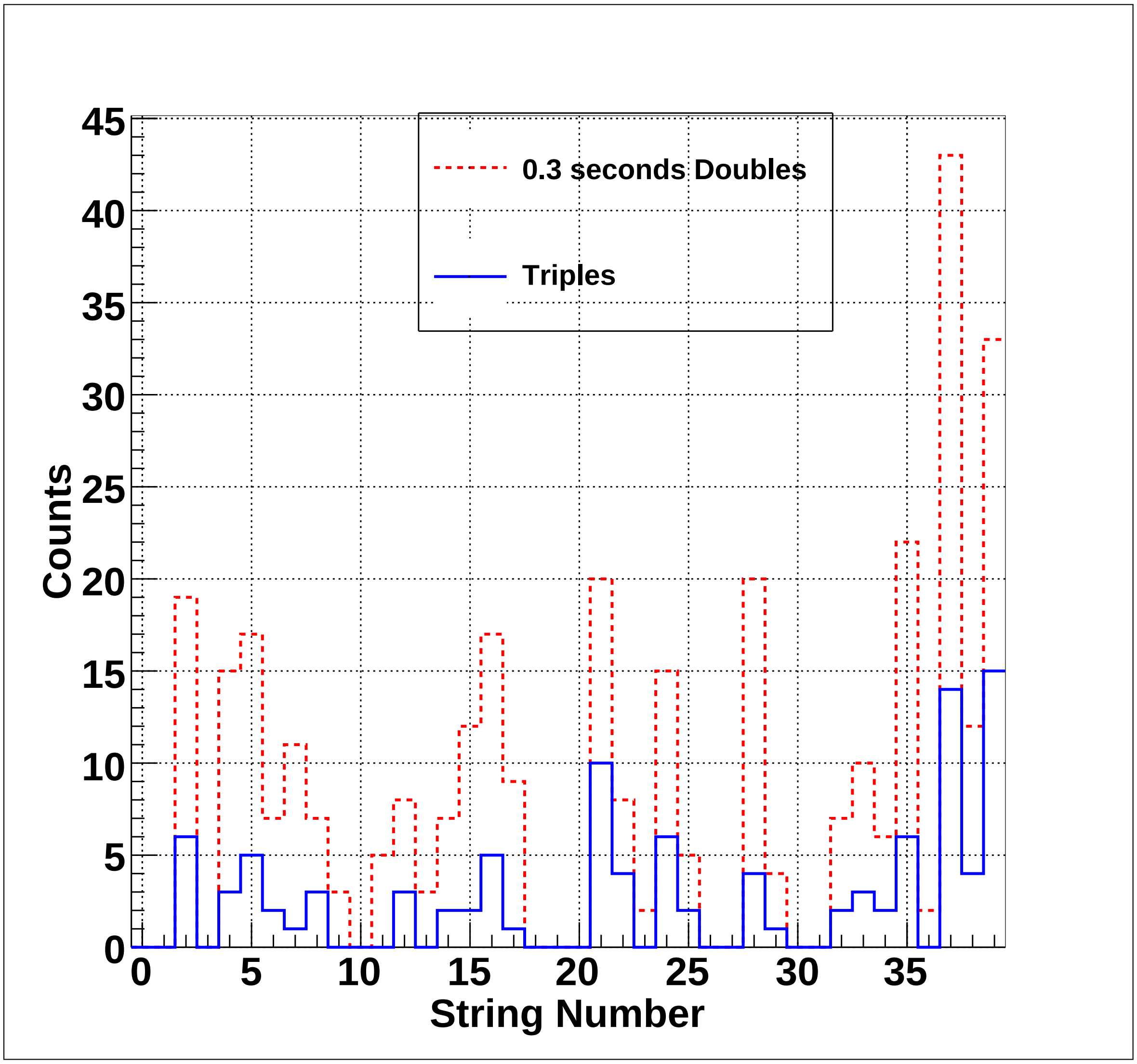


Figure A.10: String distribution of triples and short doubles. Strings, not shown, (0, 1, 3, 10, 18, 20, 26, 27, 30 and 31) were not analysed.

	Restricted		**All Strings**	
Events	**Counts**	**Accidentals**	**Counts**	**Accidentals**
Triples	106	0.015	167	0.07
Short Doubles	349	1.43	575	3.49
Long Doubles	1173	436	2314	1064

Table A.9: Comparing counts when all strings were included to when only good strings were analysed.

	All Strings	**Restricted**
Triples	0.89 ± 0.06	0.78 ± 0.06
Short Doubles	0.70 ± 0.03	0.58 ± 0.03
Average from Triples and Short Doubles	0.795 ± 0.067	0.680 ± 0.067
Long Doubles	0.53 ± 0.07	0.36 ± 0.05

Table A.10: Comparing thorium decay rate (decays/m^2/day) from two fits – one using all the strings and another using only good strings.

Events	**Complete NCD dataset Restricted**	**Stonehill**
Triples	0.78 ± 0.06	0.52 ± 0.10
Short Doubles	0.58 ± 0.03	0.45 ± 0.05
Long Doubles	0.36 ± 0.05	0.43 ± 0.06
Average rate	0.57 ± 0.09	0.47 ± 0.12

Table A.11: Comparing thorium decay rate (decays/m^2/day) between two analysis.

	decays/m^2/day
Complete NCD data (All strings)	0.627 ± 0.064
Complete NCD data (Restricted)	0.427 ± 0.439
Stonehill (Restricted)	$0.50^{+0.53}_{-0.39}$

Table A.12: Uranium content from long double coincident events (111.2 seconds).

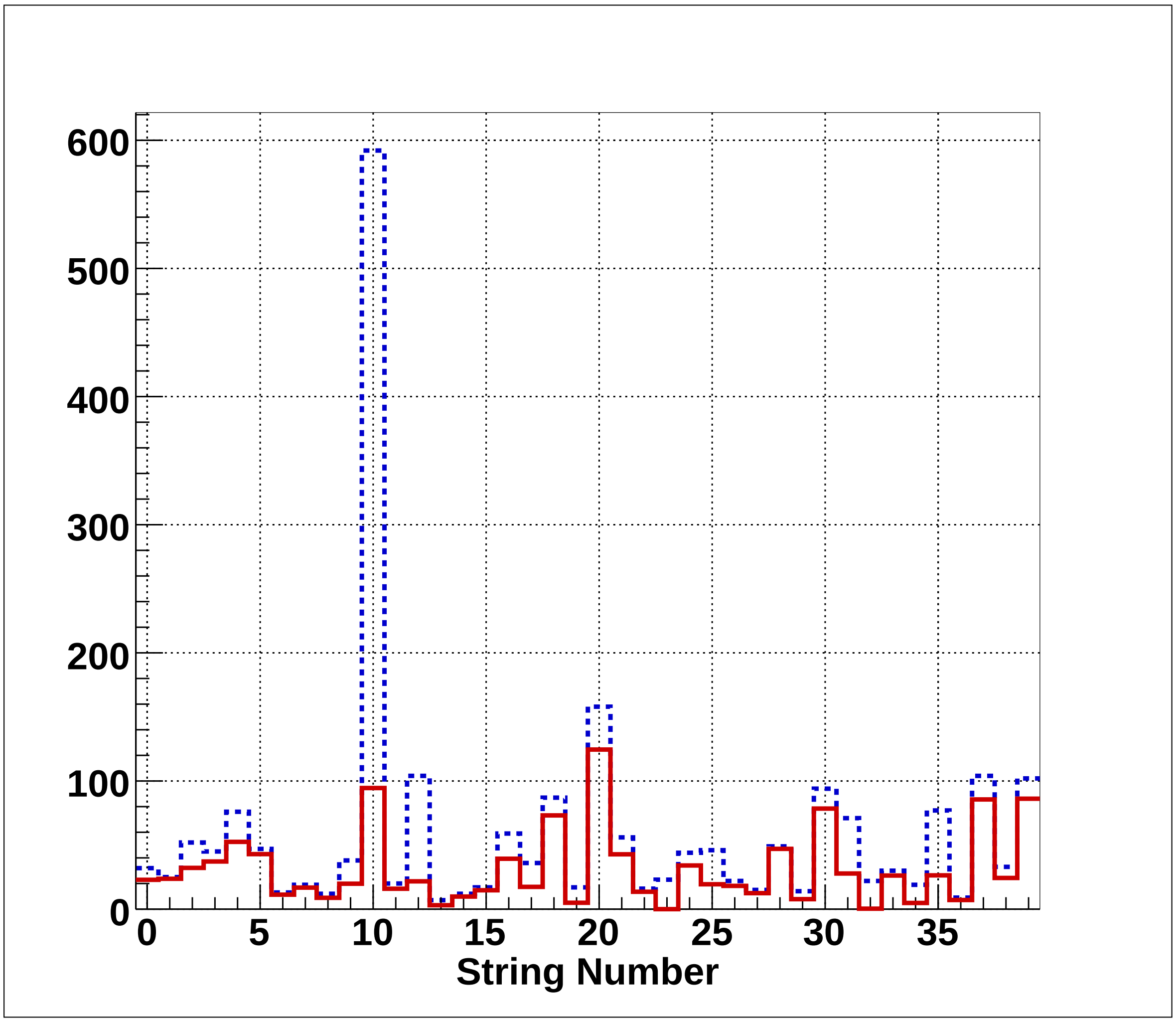


Figure A.11: String distribution of 111.2 seconds doubles; blue and red corresponds to observed and corrected number of 111.2 seconds doubles respectively.

DAMN 0 Bit#	Description
0	Retrigger
18	Muon follower short
22	NHIT burst
26	Muon follower blindness
27	Missed muon follower short
30	Shaper burst
31	Mux burst
DAMN 1 Bit#	Description
1	Reverse fork
2	Fork
3	NCD oscillatory
4	Flat trace
5	Narrow pulse
6	Run boundary
7	NCD pulses
8	Shaper overflow
10	NCD Mux-Shaper correlation
11	NCD correlation time
12	Multi NCD
13	Third reflection
14	Multiple large peak
15	NCD positive signal
16	NCD frequency domain flatness
17	NCD frequency domain fork cut
18	NCD frequency-domain fork cut
19	NCD spike-area cut
20	NCD frequency-domain symmetry cut
21	NCD frequency-domain oscillation cut
22	NCD NRE pulse tag cut
23	NCD general record cut

Table A.13: Breaking down DAMN cuts 0XCC440001 and 0X00FFDDFE into bits. These cuts were applied to remove non-physics events from the current analysis.

DAMN 0 Bit#	Description
0	Retrigger
18	Muon follower short
22	NHIT burst
27	Missed muon follower short
30	Shaper burst
31	Mux burst
DAMN 1 Bit#	**Description**
13	Muon hit all crates
14	Third reflection
15	Multiple large peak
16	NCD positive signal
17	NCD frequency-domain flatness cut
19	NCD spike area cut
20	NCD frequency-domain symmetry cut

Table A.14: Breaking down DAMN cuts 0XC8440001 and 0X1BE into bits. These cuts were applied by Stonehill to remove non-physics events from the analysis.

Appendix B

Plots for testing the bias in the number of events belonging to neutral current

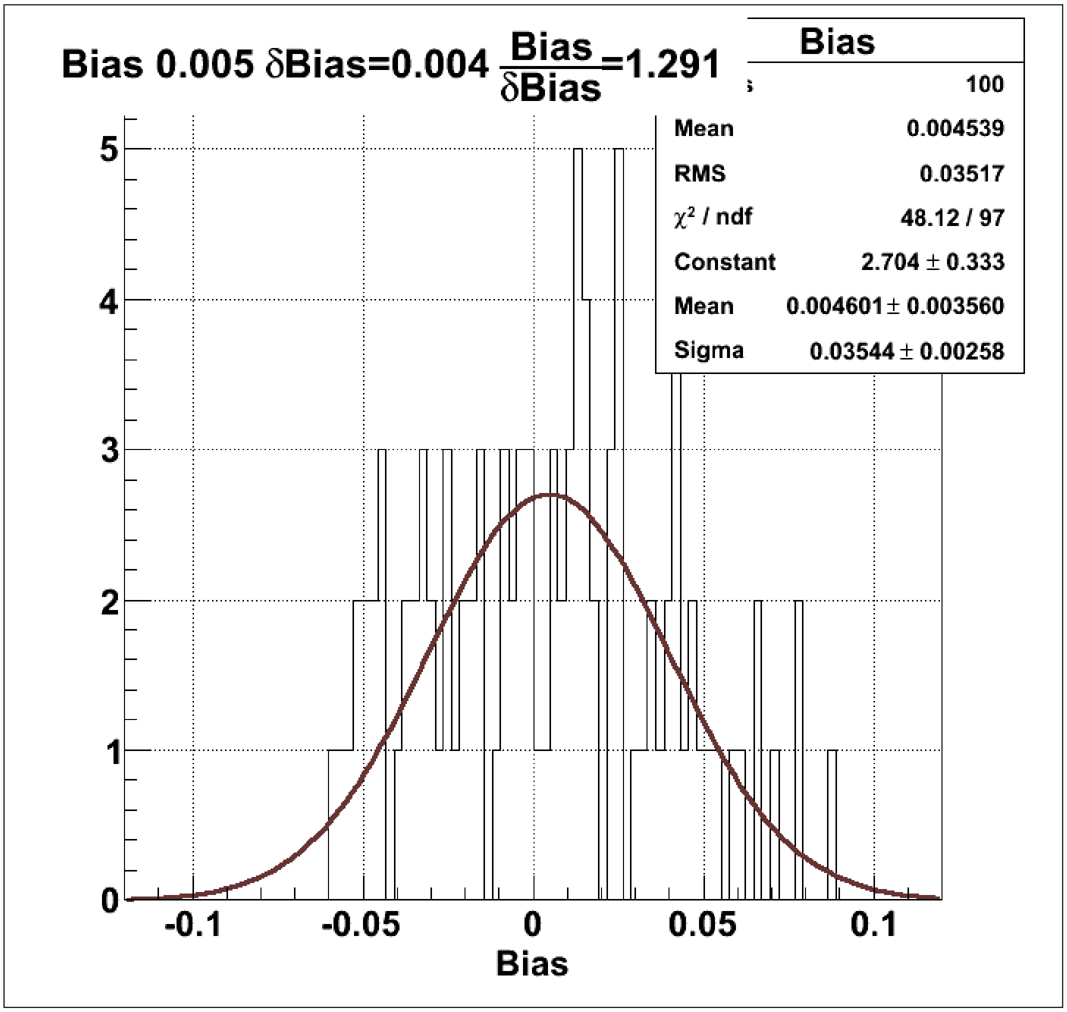


Figure B.1: NC fit result of the 1/3 simulated datasets for the step 1. The expected value and its uncertainty, required to calculate the bias, is the mean and RMS of the posterior distribution after taking out the burn-in period 40,000 out of 75,000.

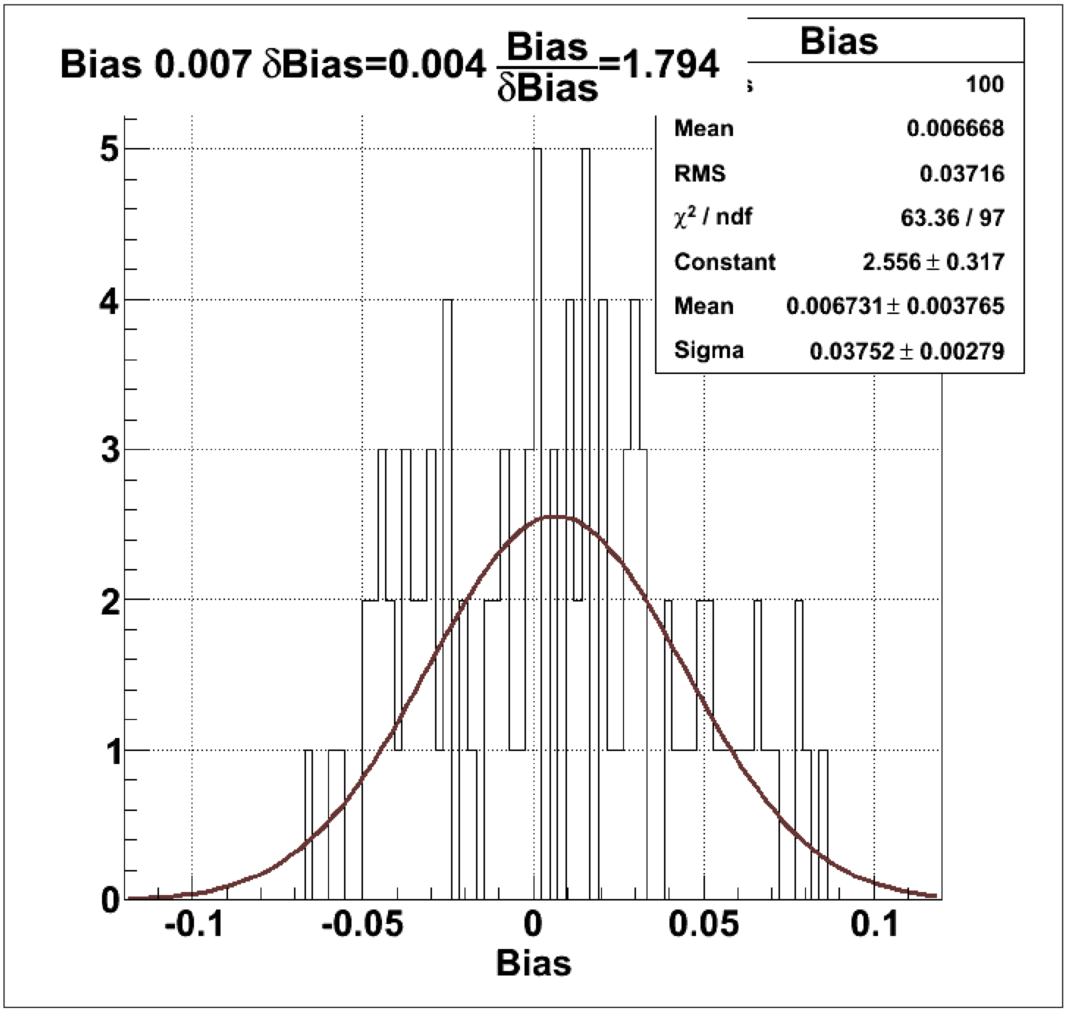


Figure B.2: NC fit result of the 1/3 simulated datasets for the step 2. The expected value and its uncertainty, required to calculate the bias, is the mean and RMS of the posterior distribution after taking out the burn-in period 40,000 out of 75,000.

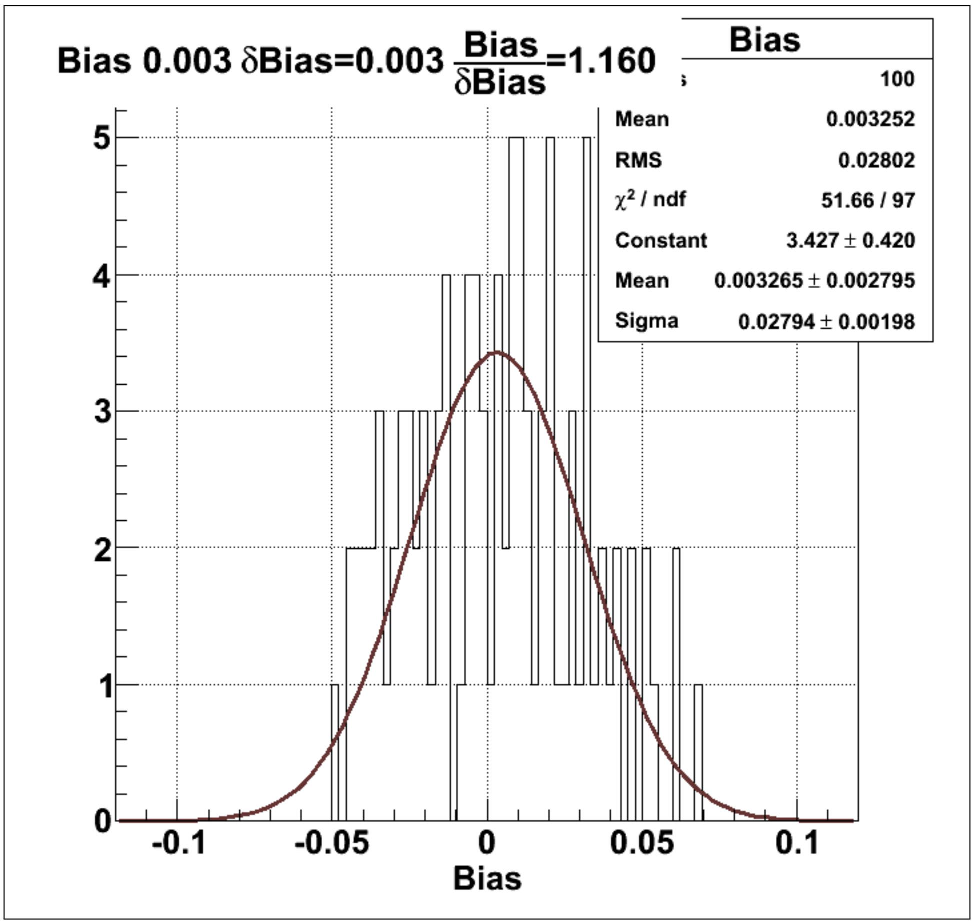


Figure B.3: NC fit result of the 1/3 simulated datasets for the step 3. The expected value and its uncertainty, required to calculate the bias, is the mean and RMS of the posterior distribution after taking out the burn-in period 40,000 out of 75,000.

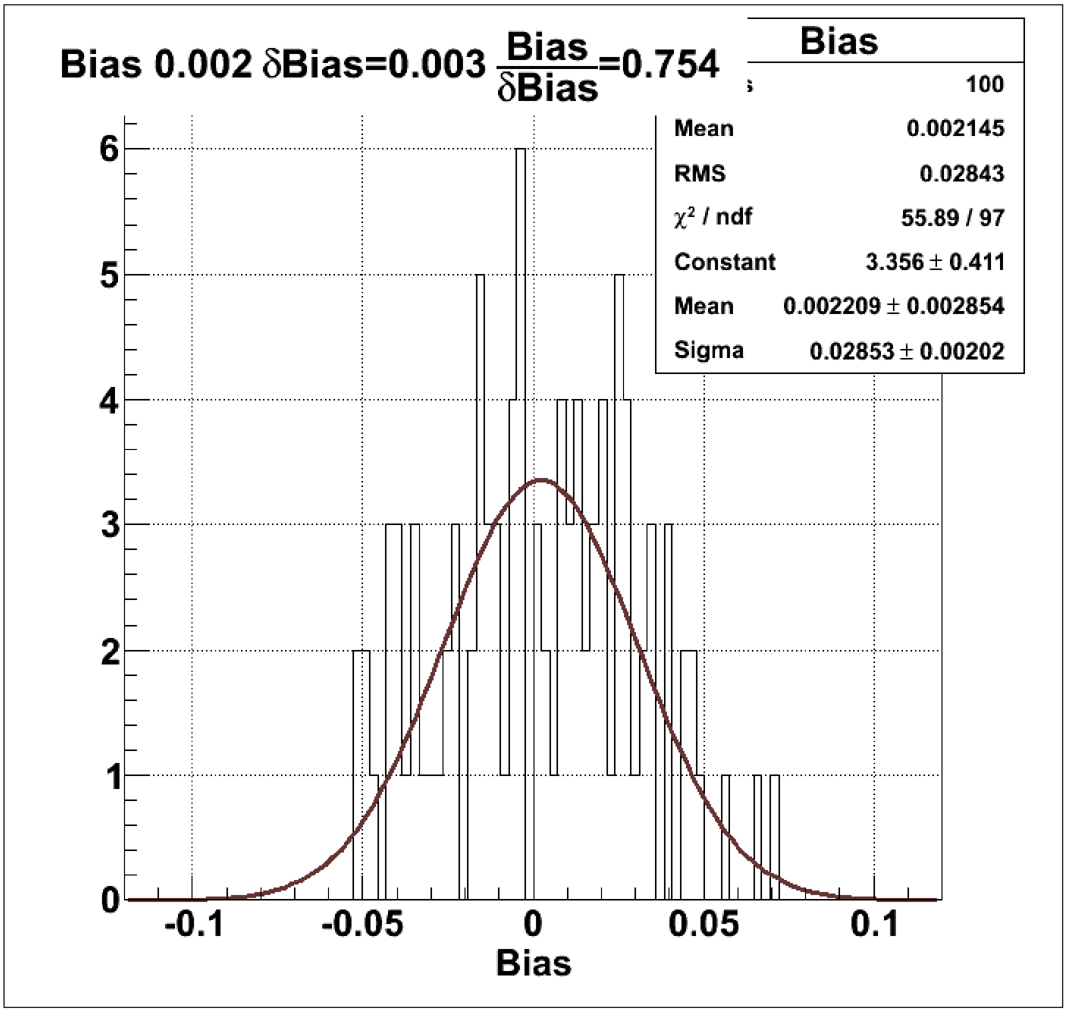


Figure B.4: NC fit result of the 1/3 simulated datasets for the step 4. The expected value and its uncertainty, required to calculate the bias, is the mean and RMS of the posterior distribution after taking out the burn-in period 40,000 out of 75,000.

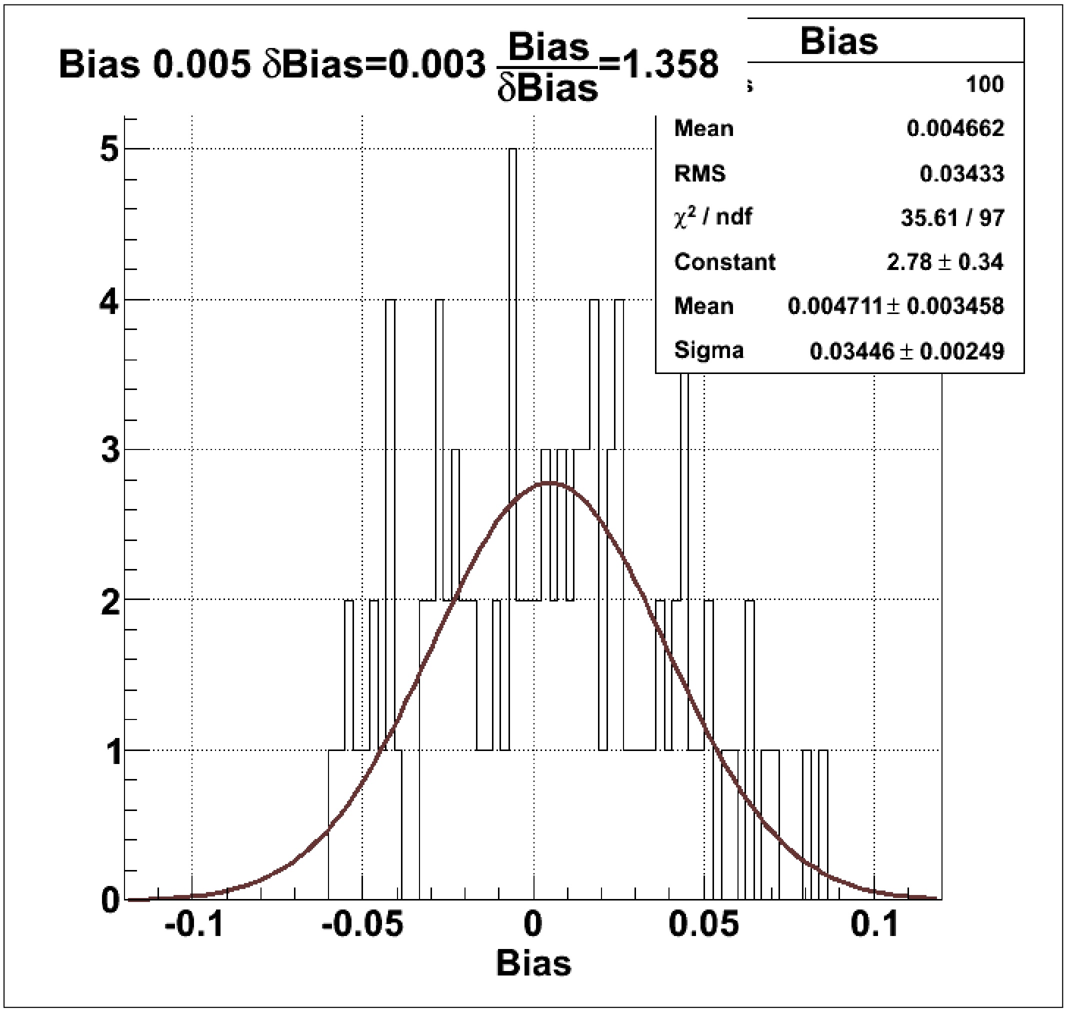


Figure B.5: NC fit result of the 1/3 simulated datasets for the step 5. The expected value and its uncertainty, required to calculate the bias, is the mean and RMS of the posterior distribution after taking out the burn-in period 40,000 out of 75,000.

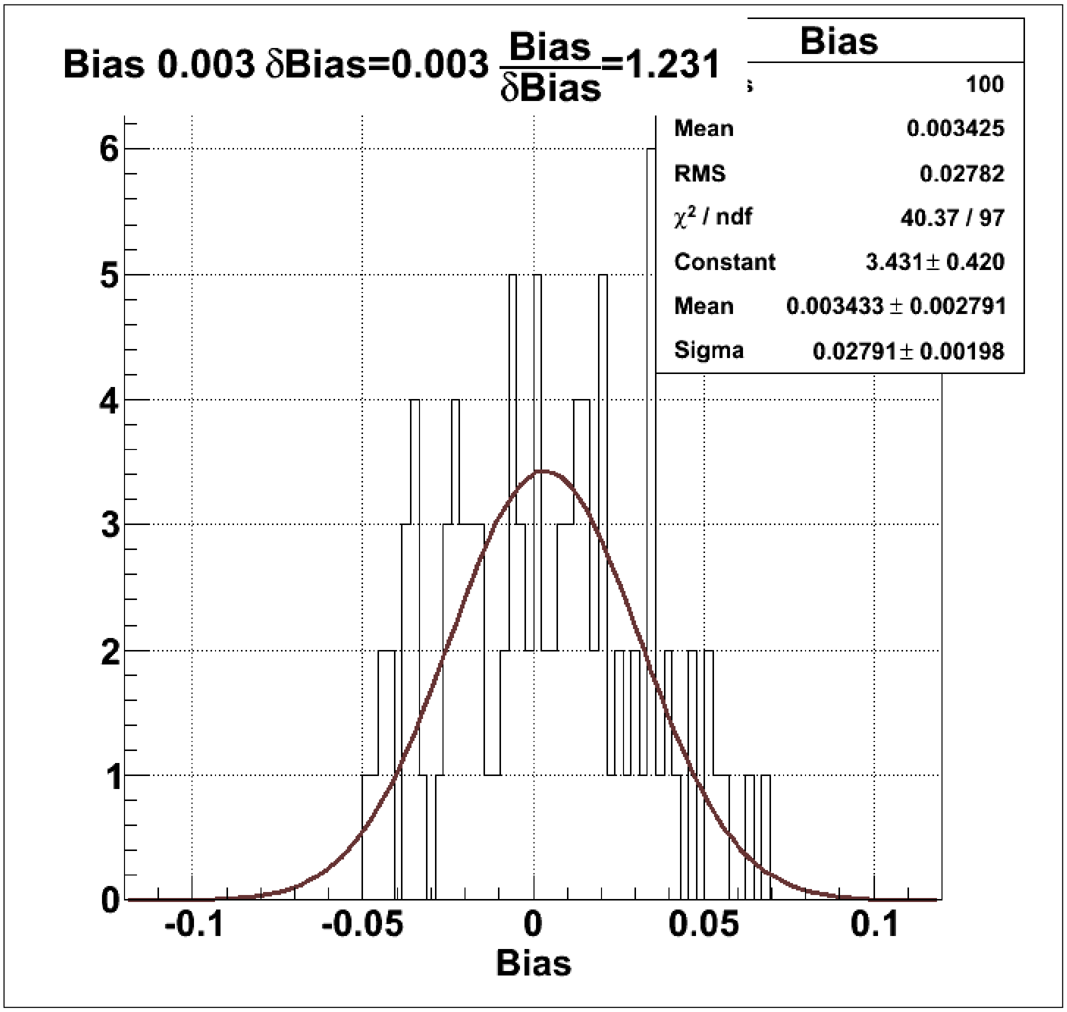


Figure B.6: NC fit result of the 1/3 simulated datasets for the step 6. The expected value and its uncertainty, required to calculate the bias, is the mean and RMS of the posterior distribution after taking out the burn-in period 40,000 out of 75,000.

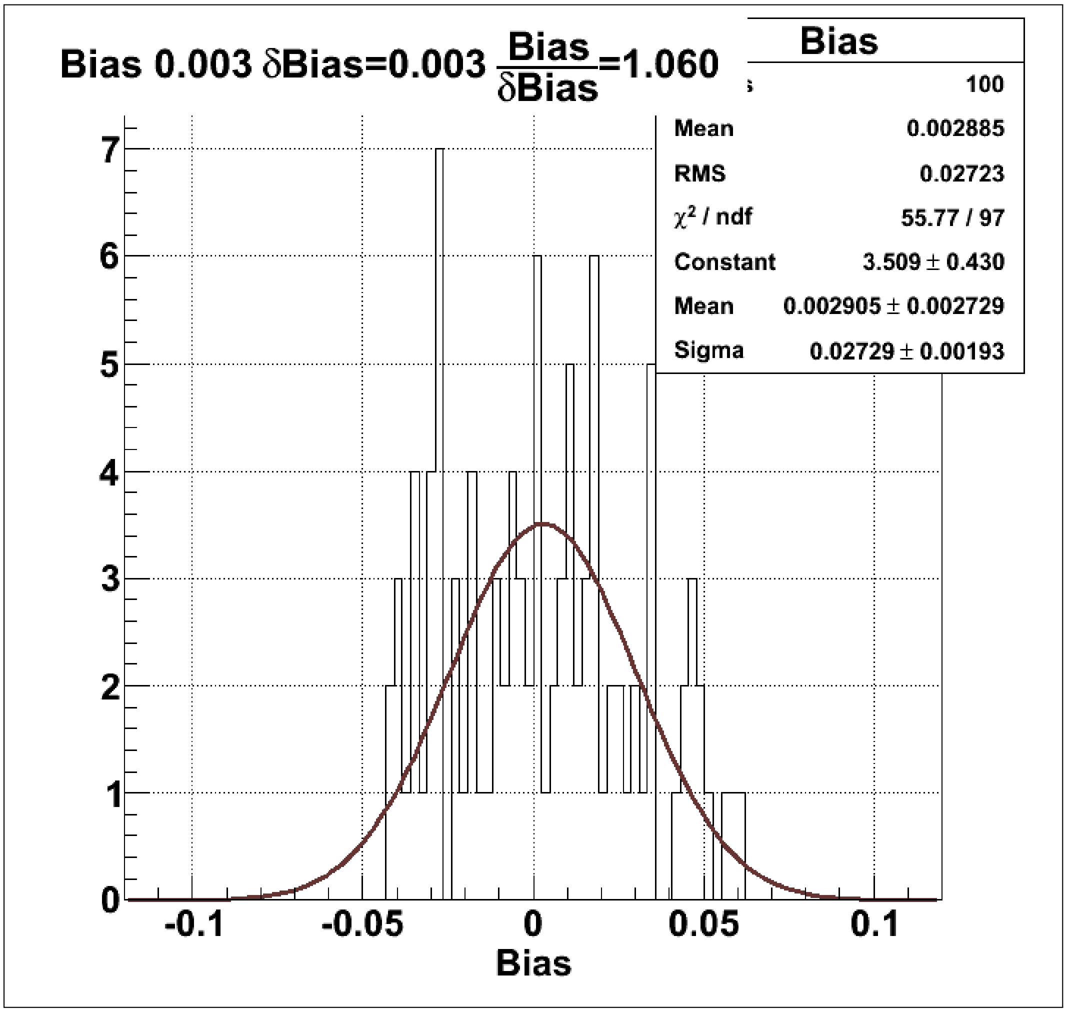


Figure B.7: NC fit result of the 1/3 simulated datasets for the step 7. The expected value and its uncertainty, required to calculate the bias, is the mean and RMS of the posterior distribution after taking out the burn-in period 40,000 out of 75,000.

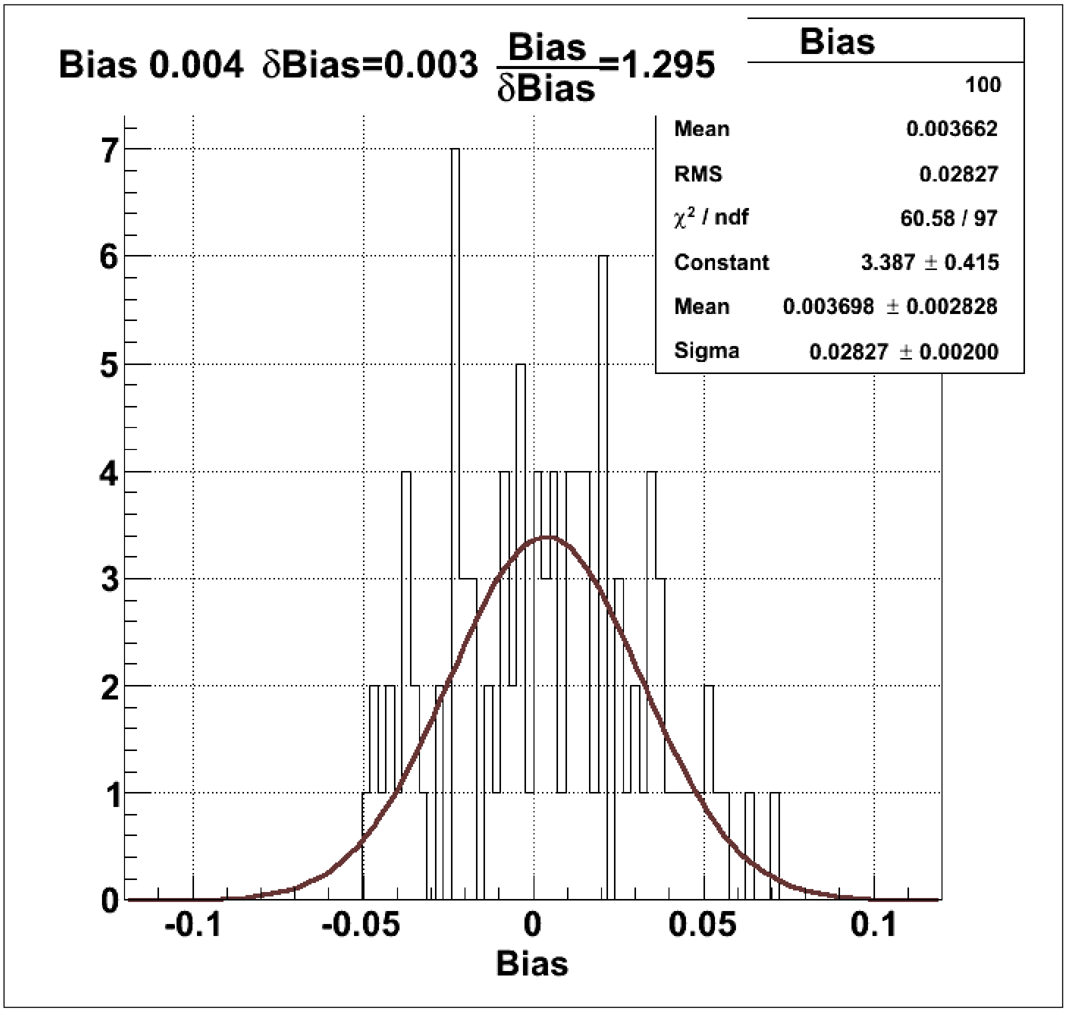


Figure B.8: NC fit result of the 1/3 simulated datasets for the step 8. The expected value and its uncertainty, required to calculate the bias, is the mean and RMS of the posterior distribution after taking out the burn-in period 40,000 out of 75,000. The brown line is a Gaussian fit on the bias distribution. The bias, uncertainty on the bias and bias in terms of number of σ from the Gaussian fit is displayed on the title of the histogram.

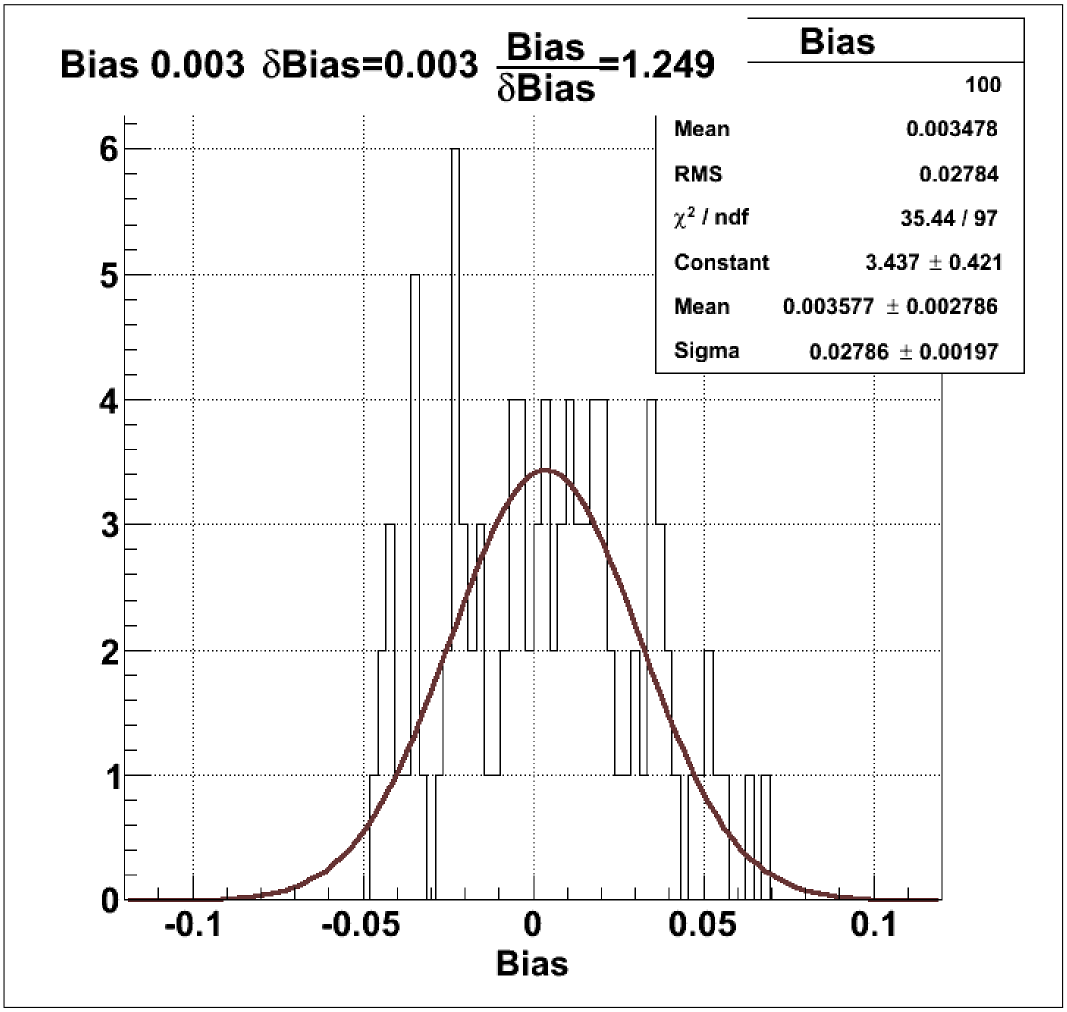


Figure B.9: NC fit result of the 1/3 simulated datasets for the step 9. The expected value and its uncertainty, required to calculate the bias, is the mean and RMS of the posterior distribution after taking out the burn-in period 40,000 out of 75,000. The brown line is a Gaussian fit on the bias distribution. The bias, uncertainty on the bias and bias in terms of number of σ from the Gaussian fit is displayed on the title of the histogram.

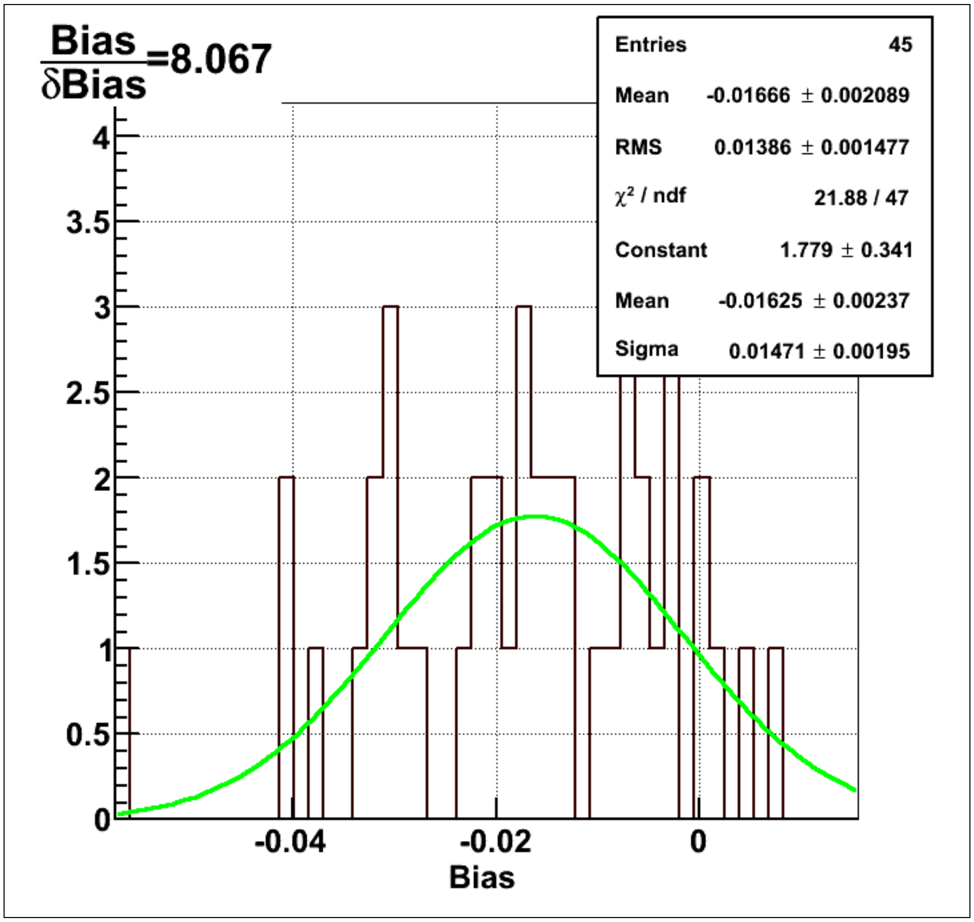


Figure B.10: NC fit result of the 1/3 simulated datasets removing NCDPD background. Green line shows the Gaussian fit. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.

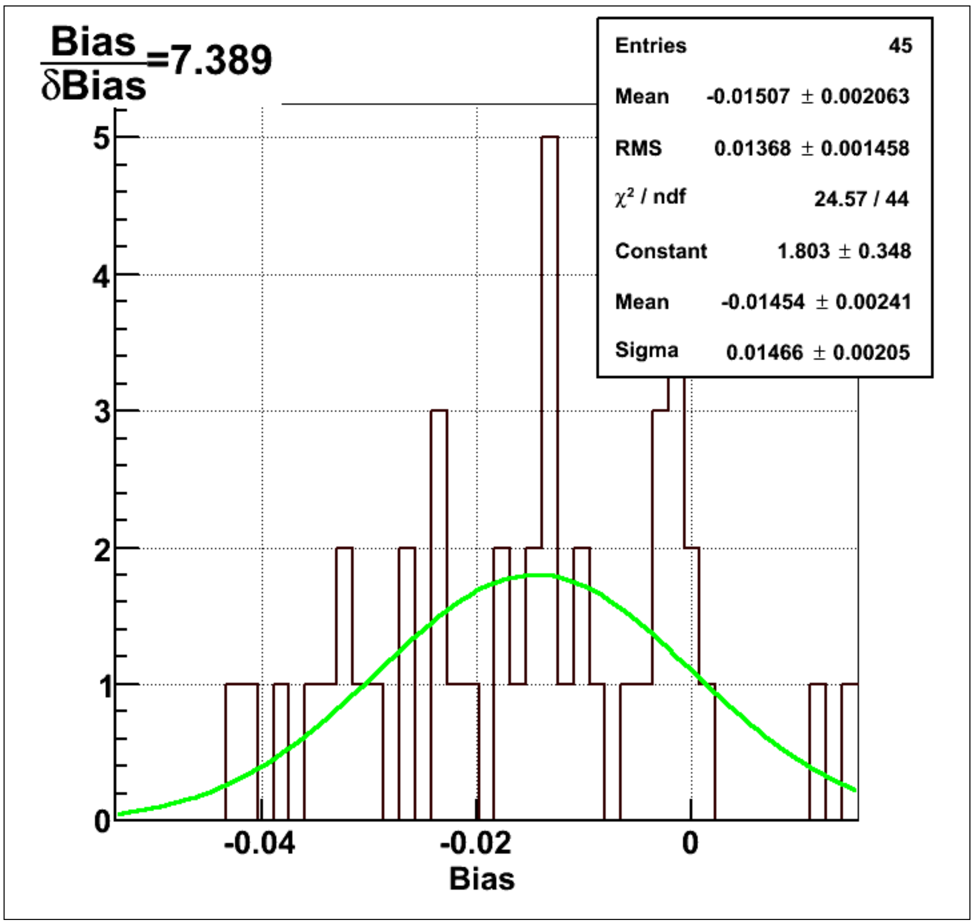


Figure B.11: NC fit result of the 1/3 simulated datasets after removing k5pd background. The 7σ bias on NC demonstrates that k5pd is not the culprit which caused the bias in the neutral current.

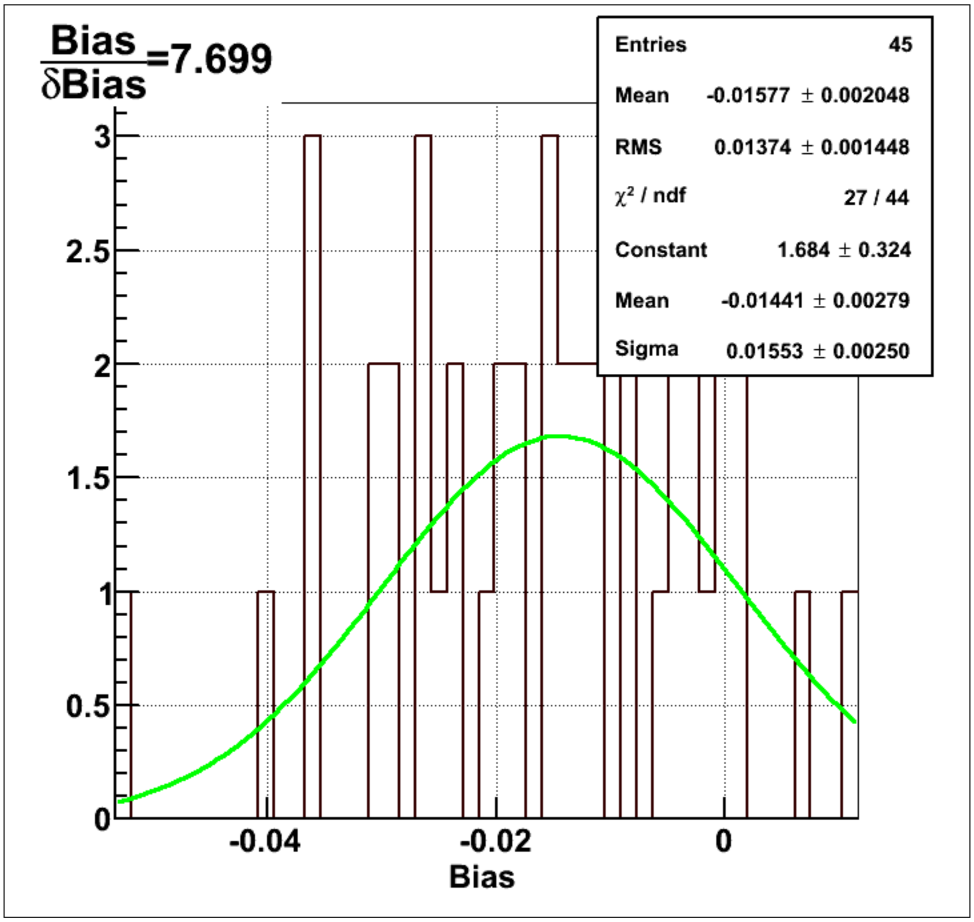


Figure B.12: NC fit result of the 1/3 simulated datasets after removing k2pd background. The 8σ bias on NC demonstrates that k2pd is not the culprit which caused the bias in the neutral current. Green line shows the Gaussian fit. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.

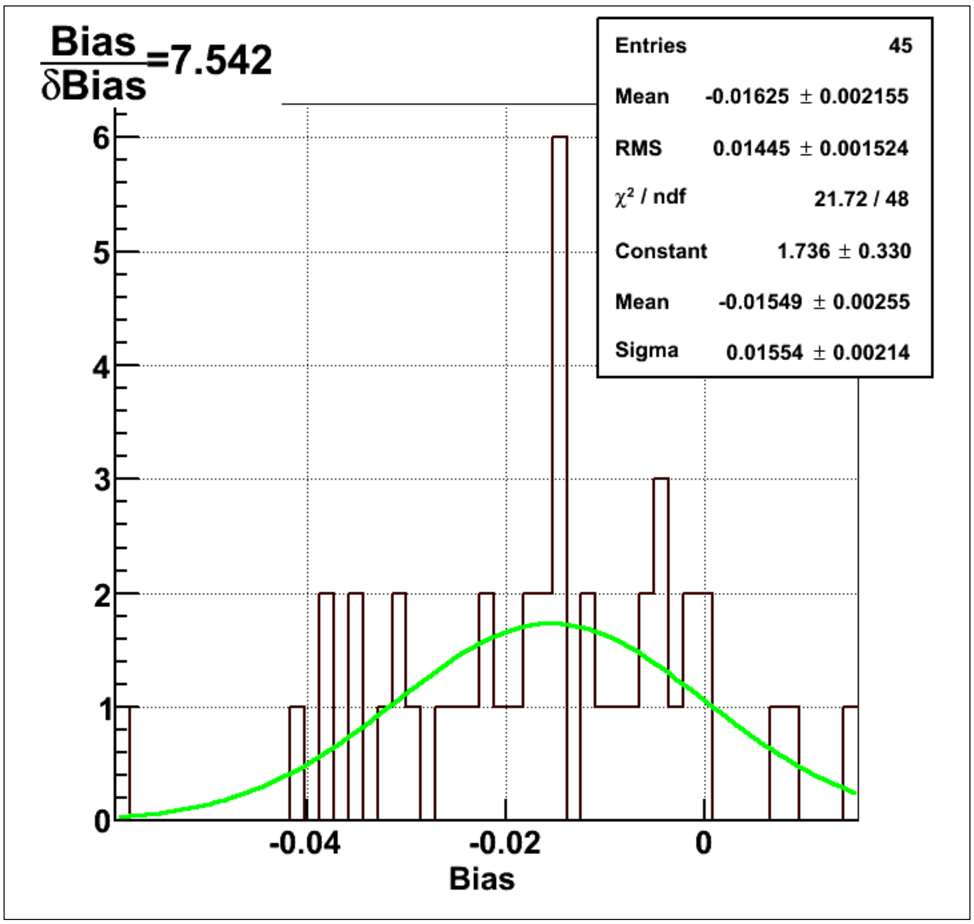


Figure B.13: NC fit result of the 1/3 simulated datasets after removing $d_2$opd background. The bias on NC demonstrates that d2opd is not the culprit which caused the bias in the neutral current.

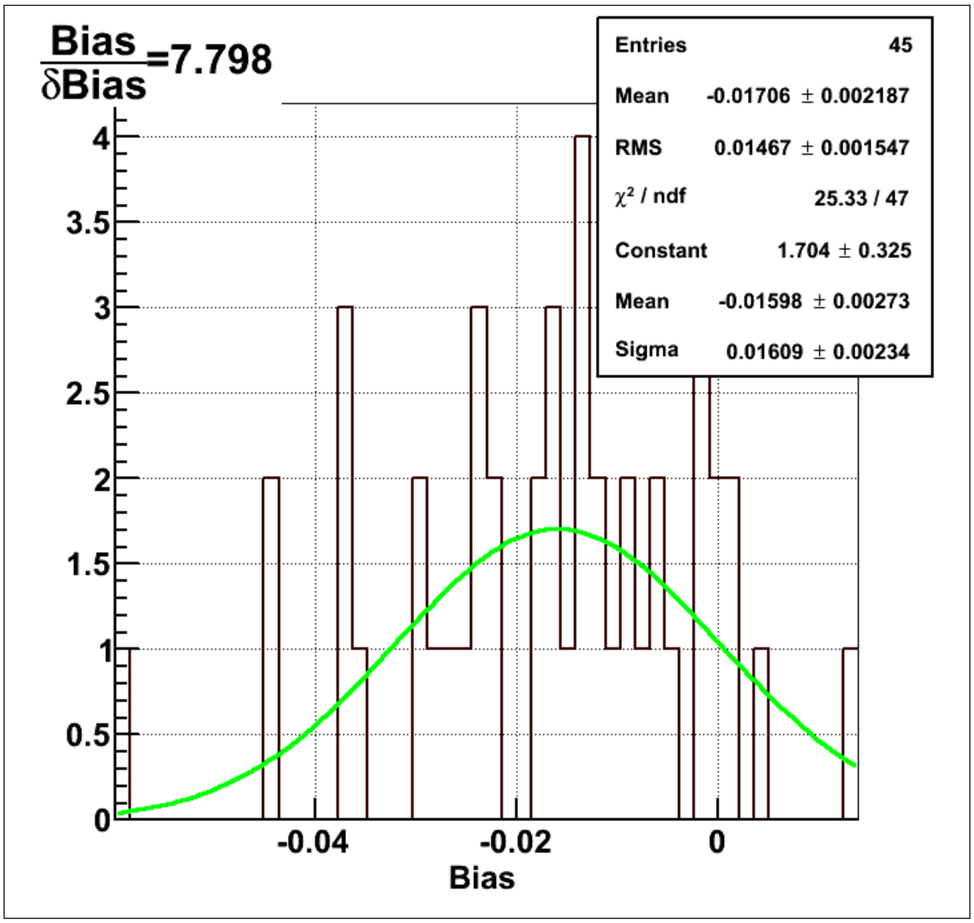


Figure B.14: NC fit result of the 1/3 simulated datasets after removing hep background. The expected value and its uncertainty, required to calculate the bias, is μ and σ from fitting a Gaussian function on the posterior distribution after taking out the burn-in period 40,000 out of 350,000. The green line is a Gaussian fit on the bias distribution. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.

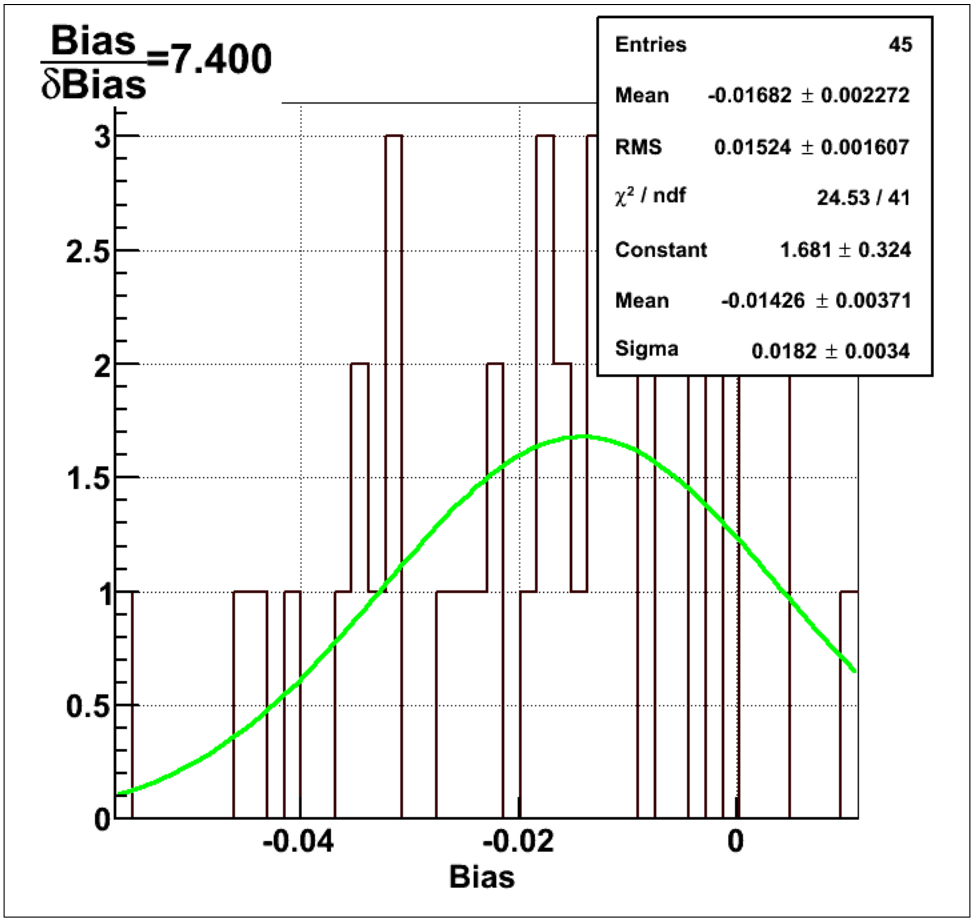


Figure B.15: NC fit result of the 1/3 simulated datasets after removing Atmospheric neutrons. The green line is a Gaussian fit on the bias distribution. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.

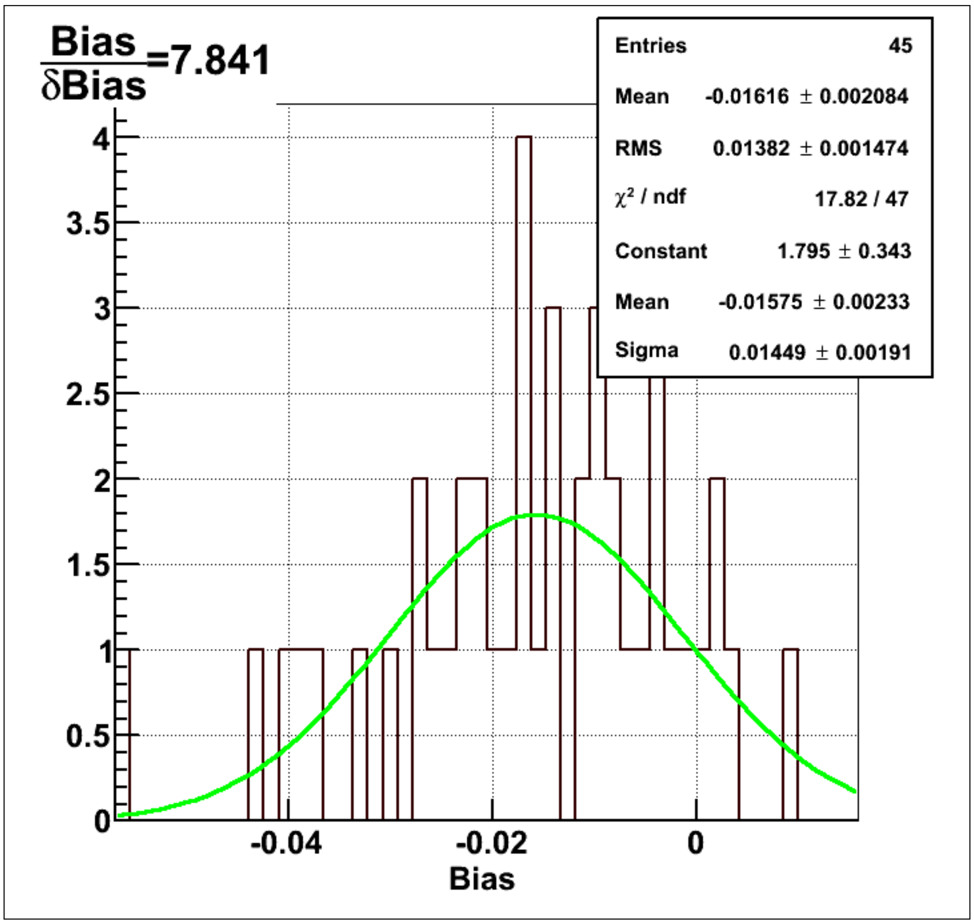


Figure B.16: NC fit result of the 1/3 simulated datasets after removing external neutrons. The green line is a Gaussian fit on the bias distribution. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.

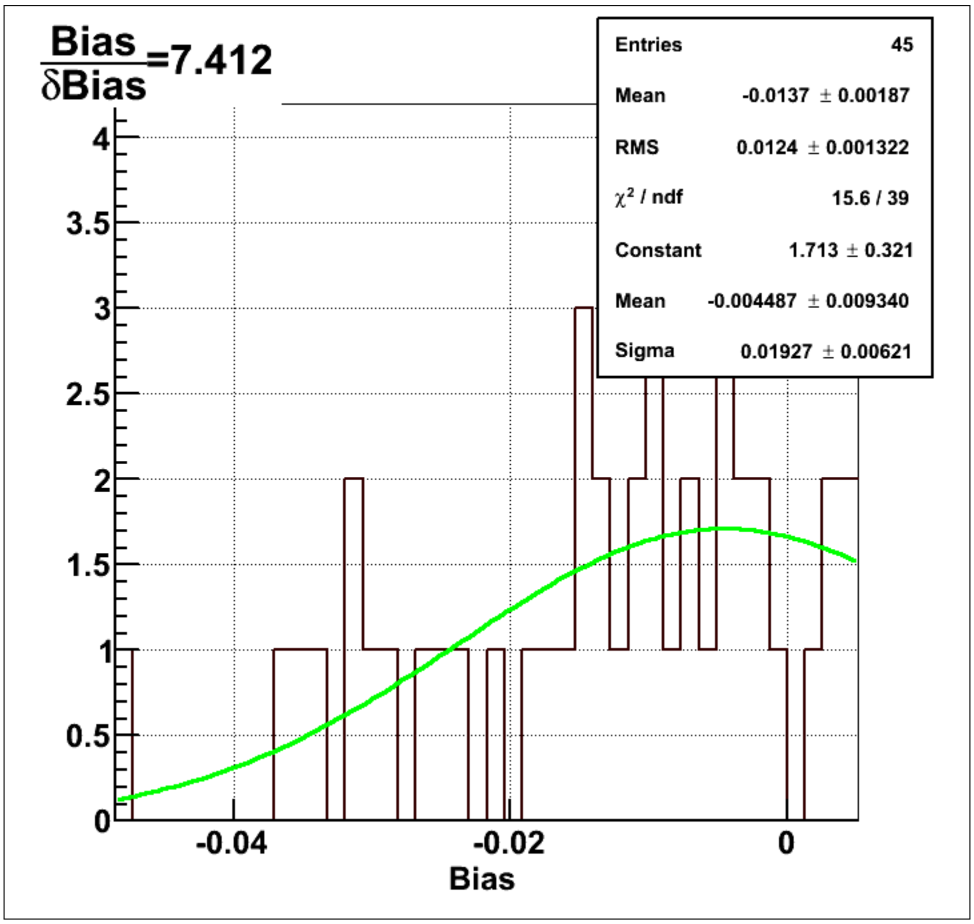


Figure B.17: NC fit result of the 1/3 simulated datasets with signals only (CC, ES, $ES_{\mu\tau}$, NC and EX). The green line is a Gaussian fit on the bias distribution. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.

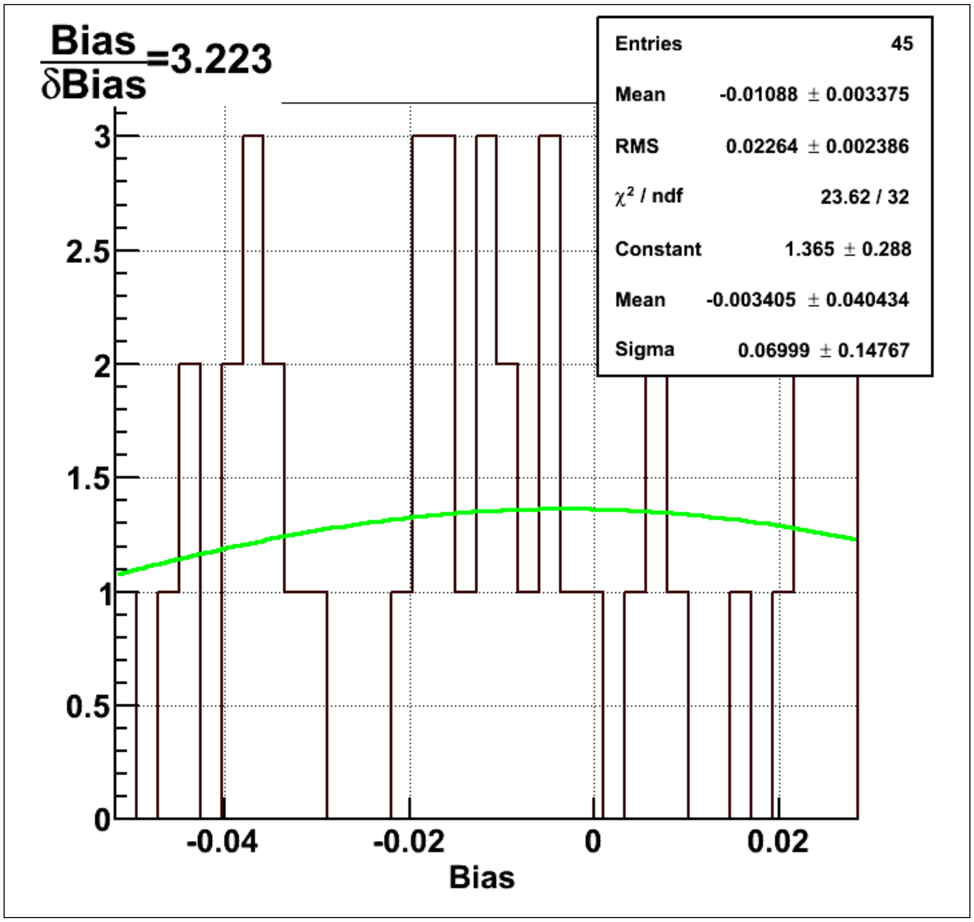


Figure B.18: NC fit result of the 1/3 simulated datasets with NC and p_{ee} parameters floating. All other parameters are fixed. The best-fit and its uncertainty is obtained, after fitting the posterior distribution from each run, to a Gaussian function. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.

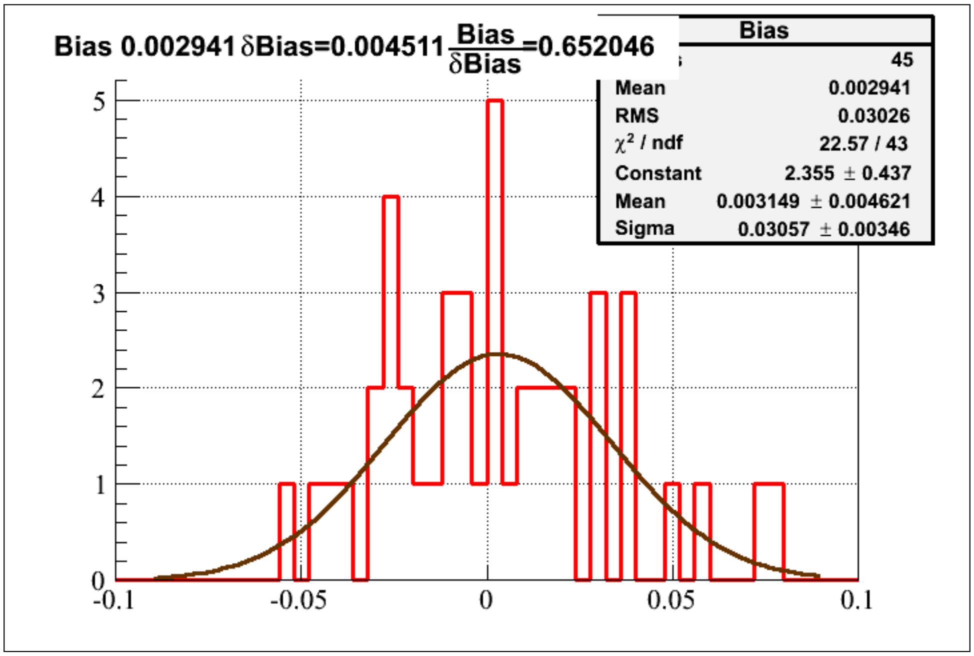


Figure B.19: NC fit result of the 1/3 simulated datasets with only NC floating. The brown line shows a Gaussian fit on the distribution. The test proves that NC is not causing the bias in itself. X axis shows bias distribution.

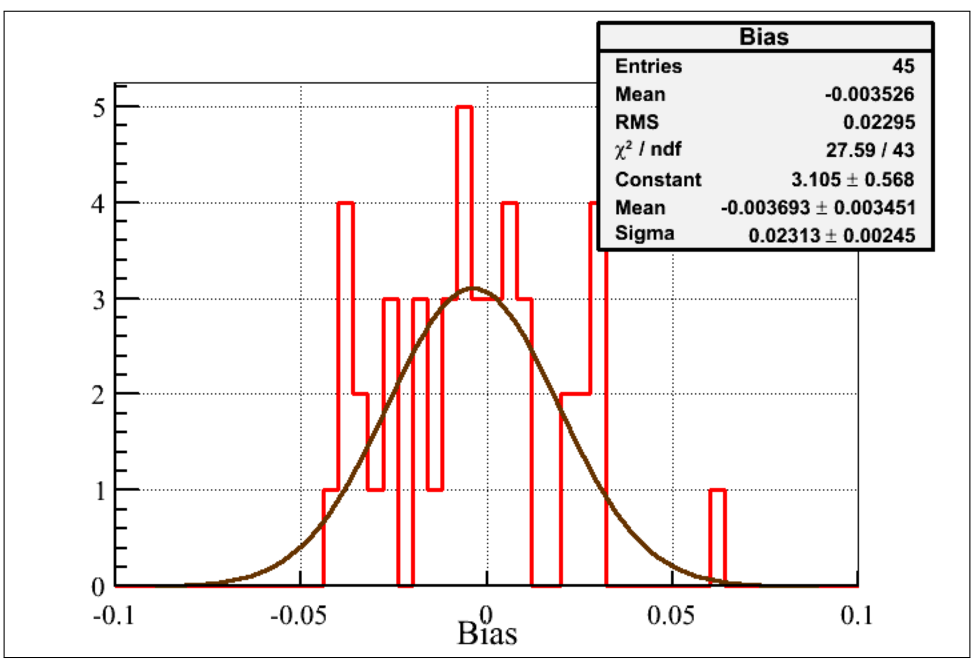


Figure B.20: NC fit result of the 1/3 simulated datasets with fixed p_1 and p_2 from the p_{ee} parameters.

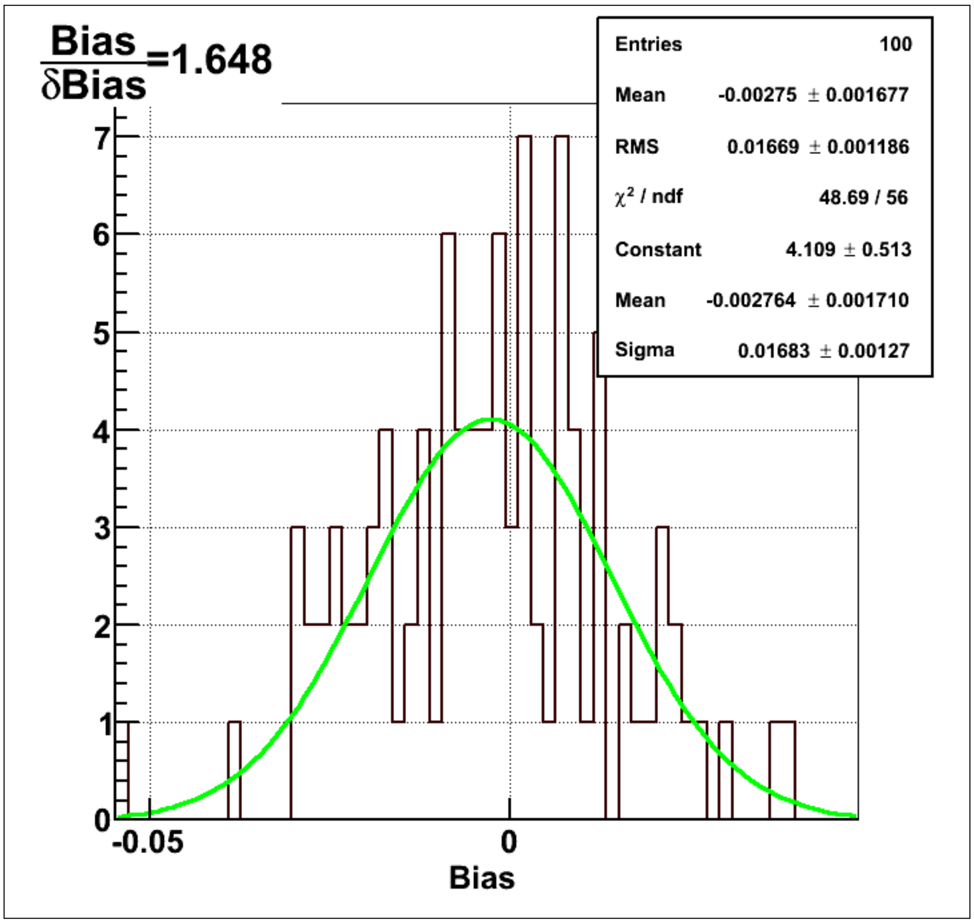


Figure B.21: NC fit result of the 1/3 simulated datasets with only p_0 floating from the p_{ee} parameters. The result indicates that p_0 does cause the bias in NC. The green line is a Gaussian fit on the bias distribution. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.

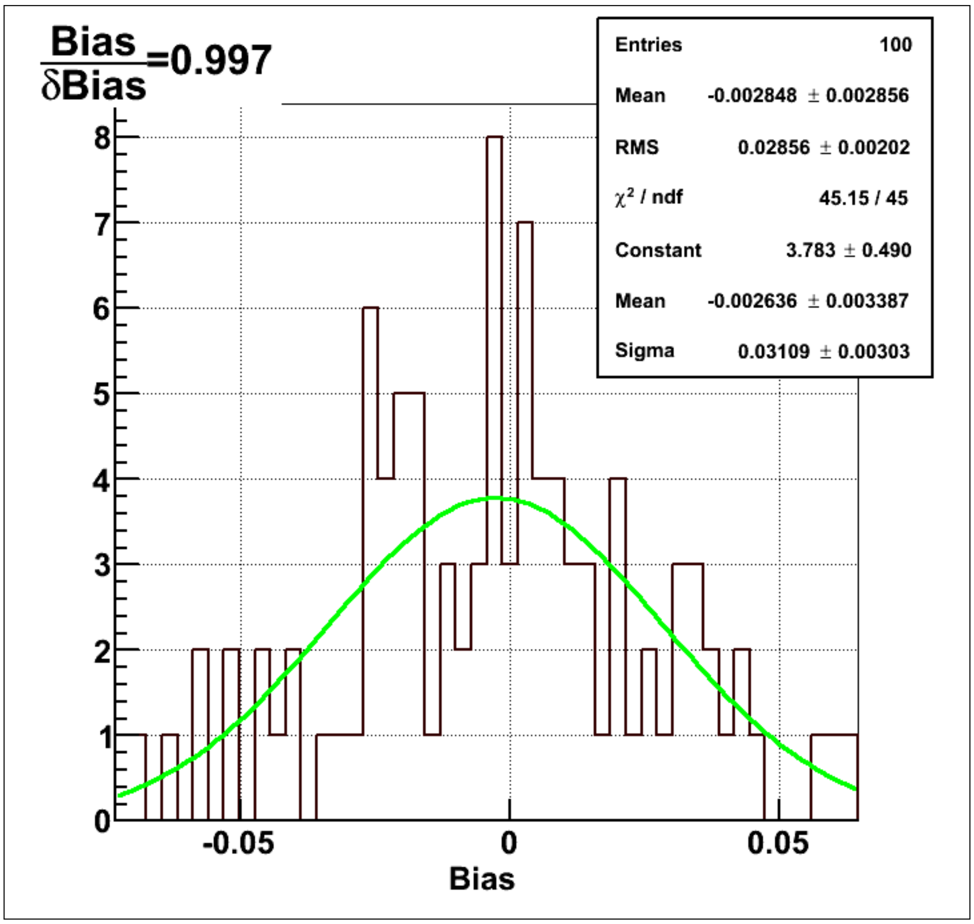


Figure B.22: NC fit result of the 1/3 simulated datasets with only p_1 floating from the p_{ee} parameters. The green line is a Gaussian fit on the bias distribution. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.

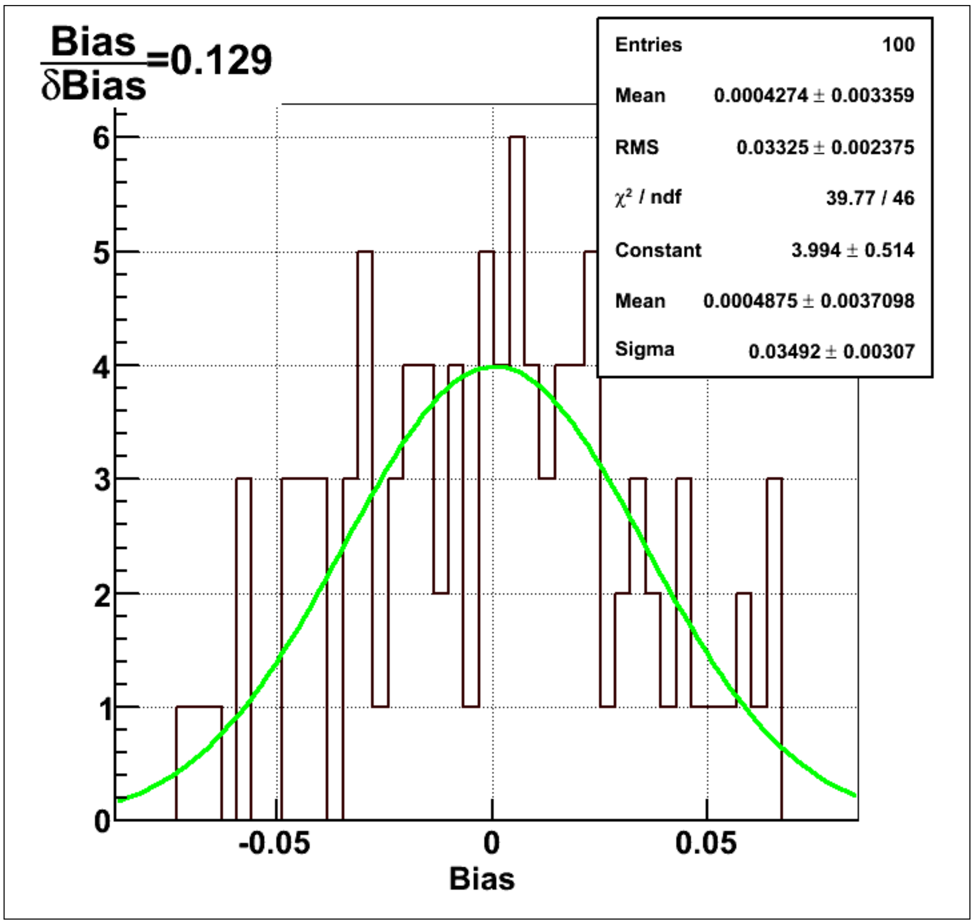
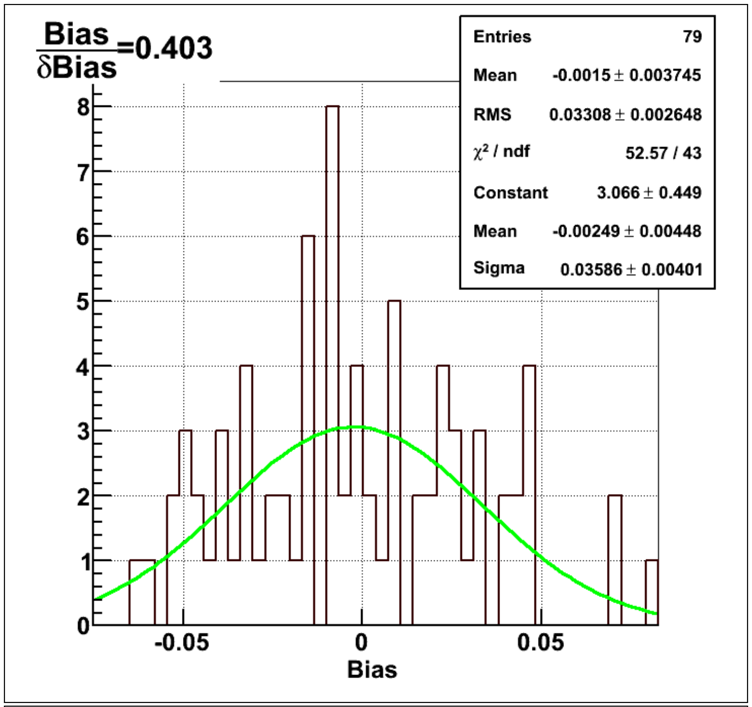


Figure B.23: NC fit result of the 1/3 simulated datasets with only p_2 floating from the p_{ee} parameters. The green line is a Gaussian fit on the bias distribution. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.



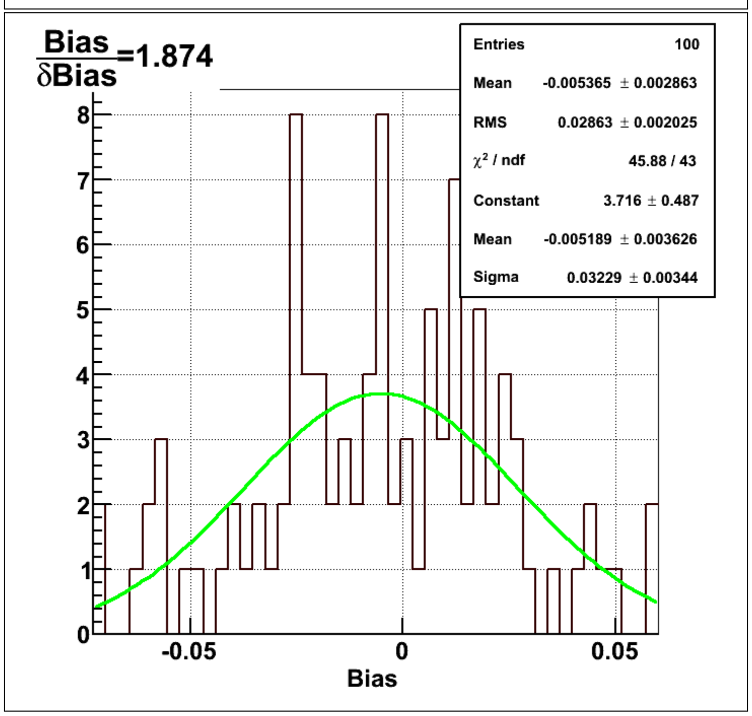


Figure B.24: NC fit result of the 1/3 simulated datasets with only a_0 (Top), a_1 (Bottom) floating from the p_{ee} parameters. The green line is a Gaussian fit on the bias distribution. The bias and uncertainty on the bias corresponds to the mean and uncertainty on the mean of the distribution as shown in the legend.

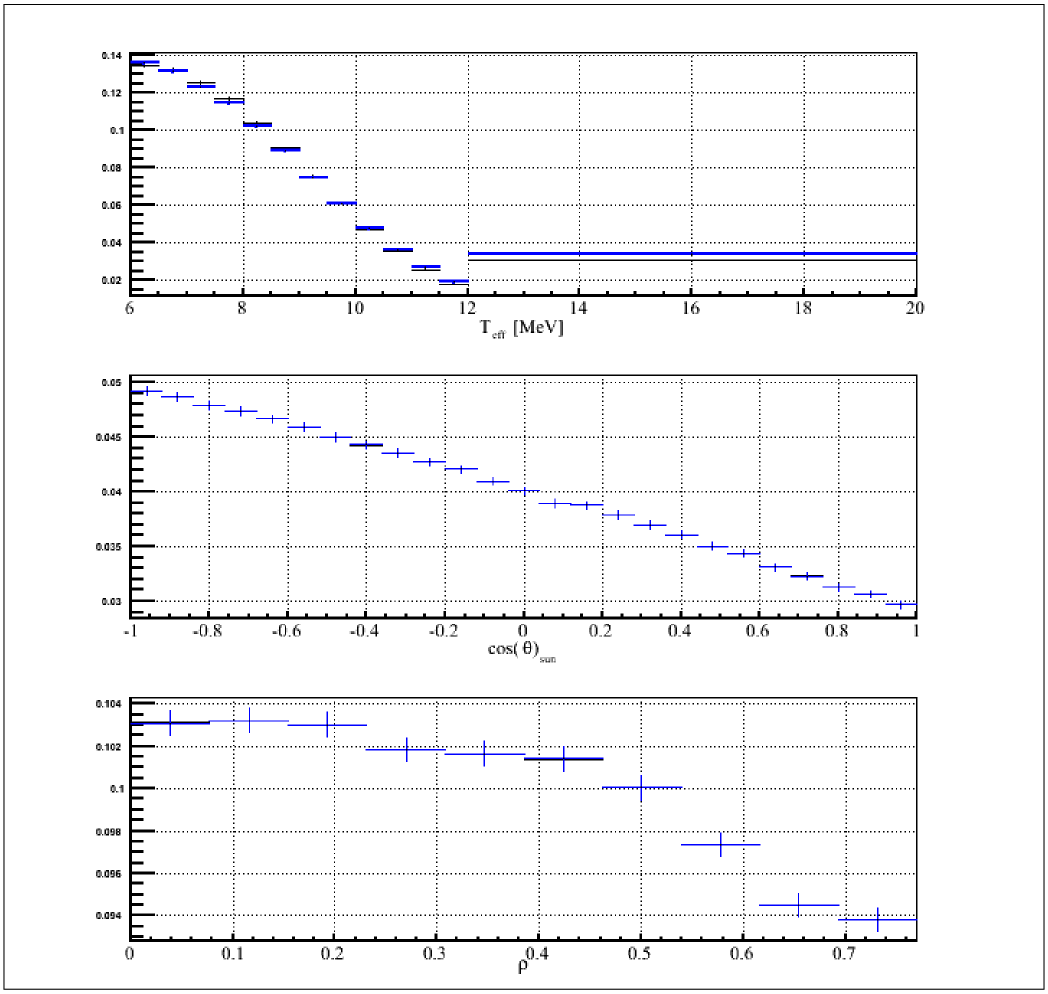


Figure B.25: Comparing three projections (energy, $\cos\theta_{\odot}$, ρ) of the distorted 3D PDFs from MCMC to the QSigEx using the nominal values of p_{ee}. The 3D PDF is for the CC Day class.

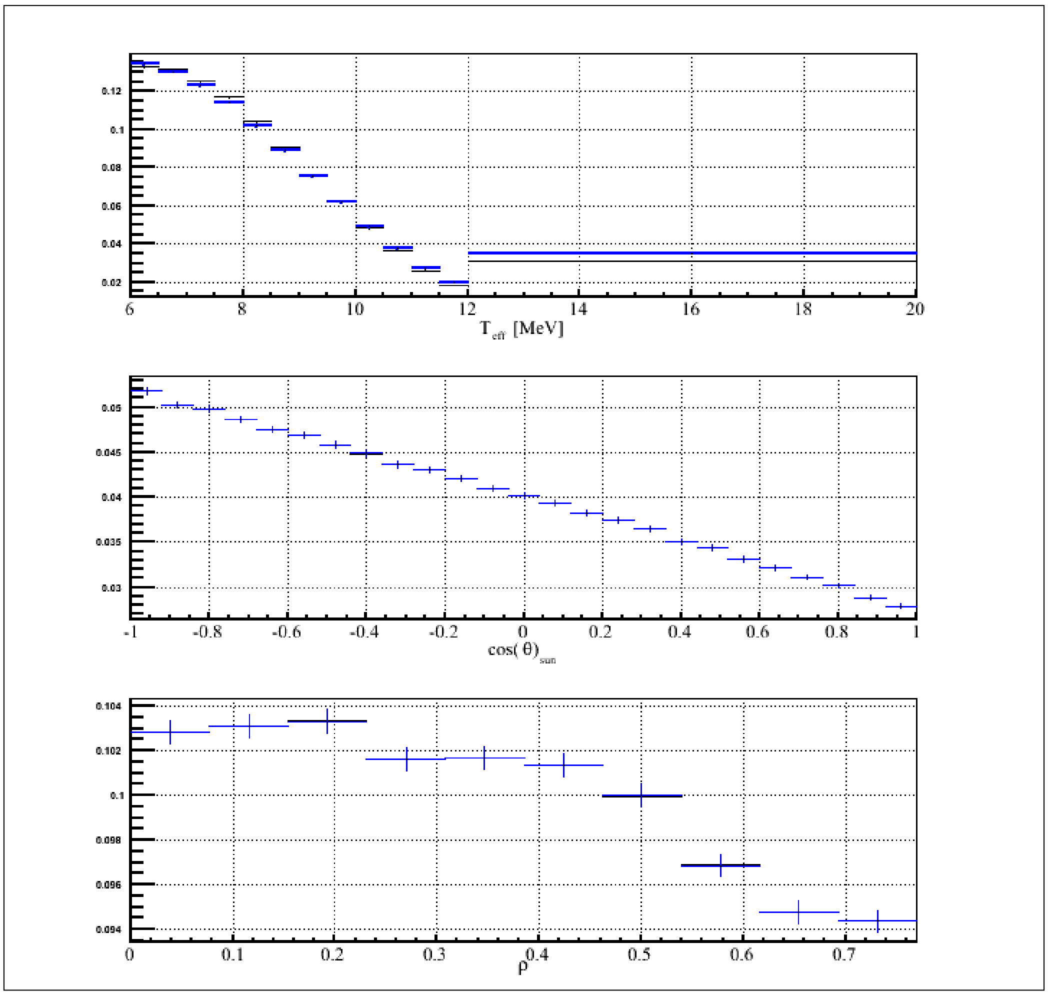


Figure B.26: Comparing three projections (energy, cos $\theta_{\odot}$, ρ) of the distorted 3D PDFs from MCMC to the QSigEx using the nominal values of p_{ee}. The 3D PDF is for the CC Night class.

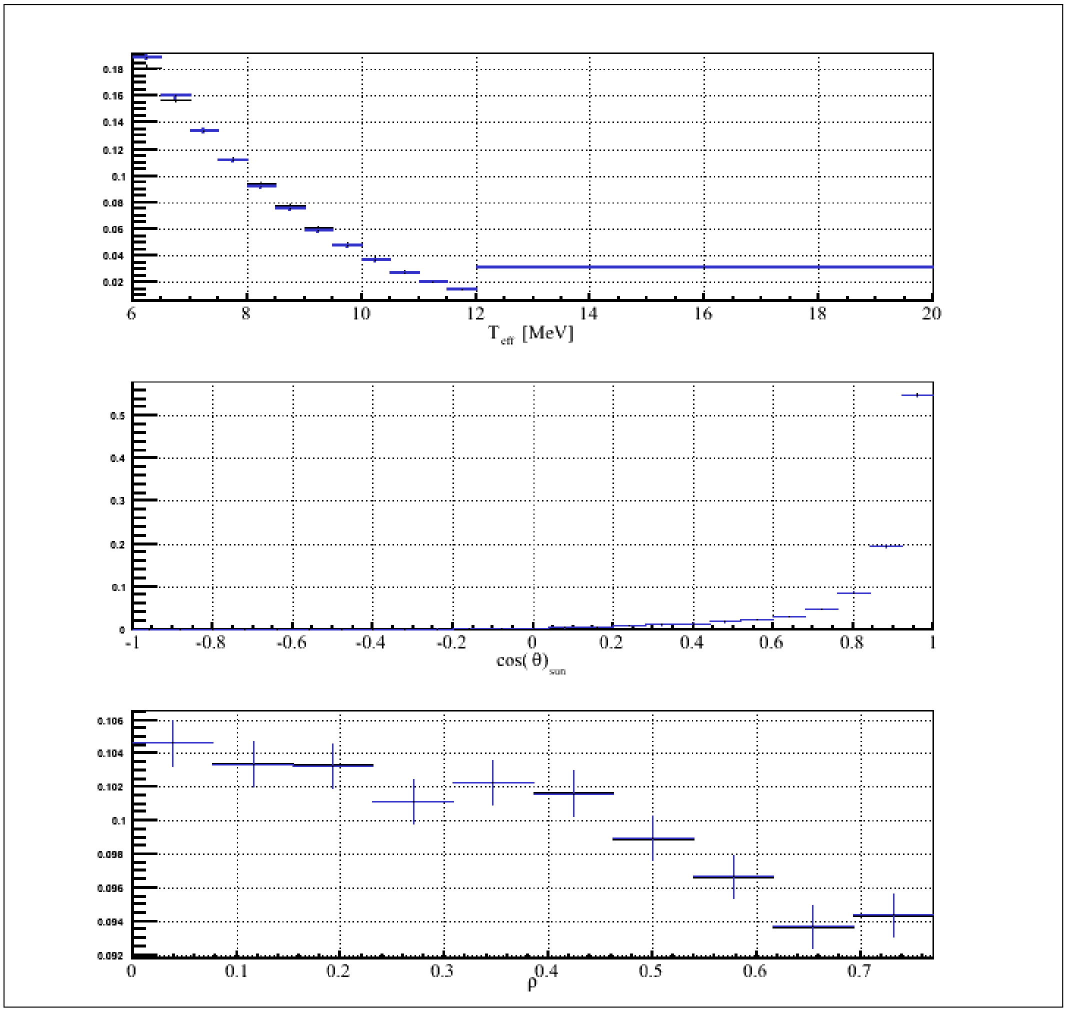


Figure B.27: Comparing three projections (energy, $\cos\theta_{\odot}$, ρ) of the distorted 3D PDFs from MCMC to the QSigEx using the nominal values of p_{ee}. The 3D PDF is for the ES Day class.

Figures B.32, B.34 and B.36 show that there is a difference in the last bin especially for the night PDFs so the next step was to fit MCMC from 6-12 MeV instead of 6-20 MeV.

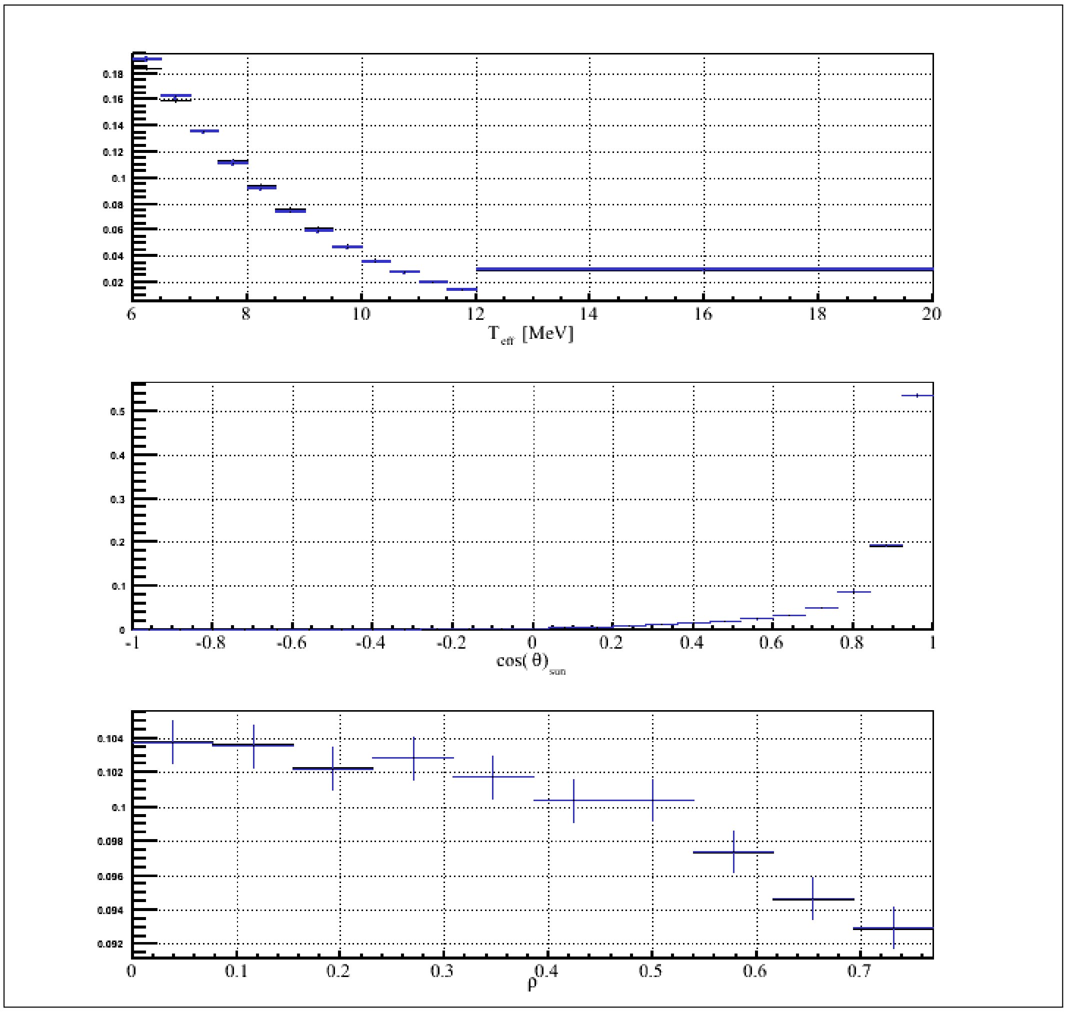


Figure B.28: Comparing three projections (energy, $\cos\theta_{\odot}$, ρ) of the distorted 3D PDFs from MCMC to the QSigEx using the nominal values of p_{ee}. The 3D PDF is for the ES Night class.

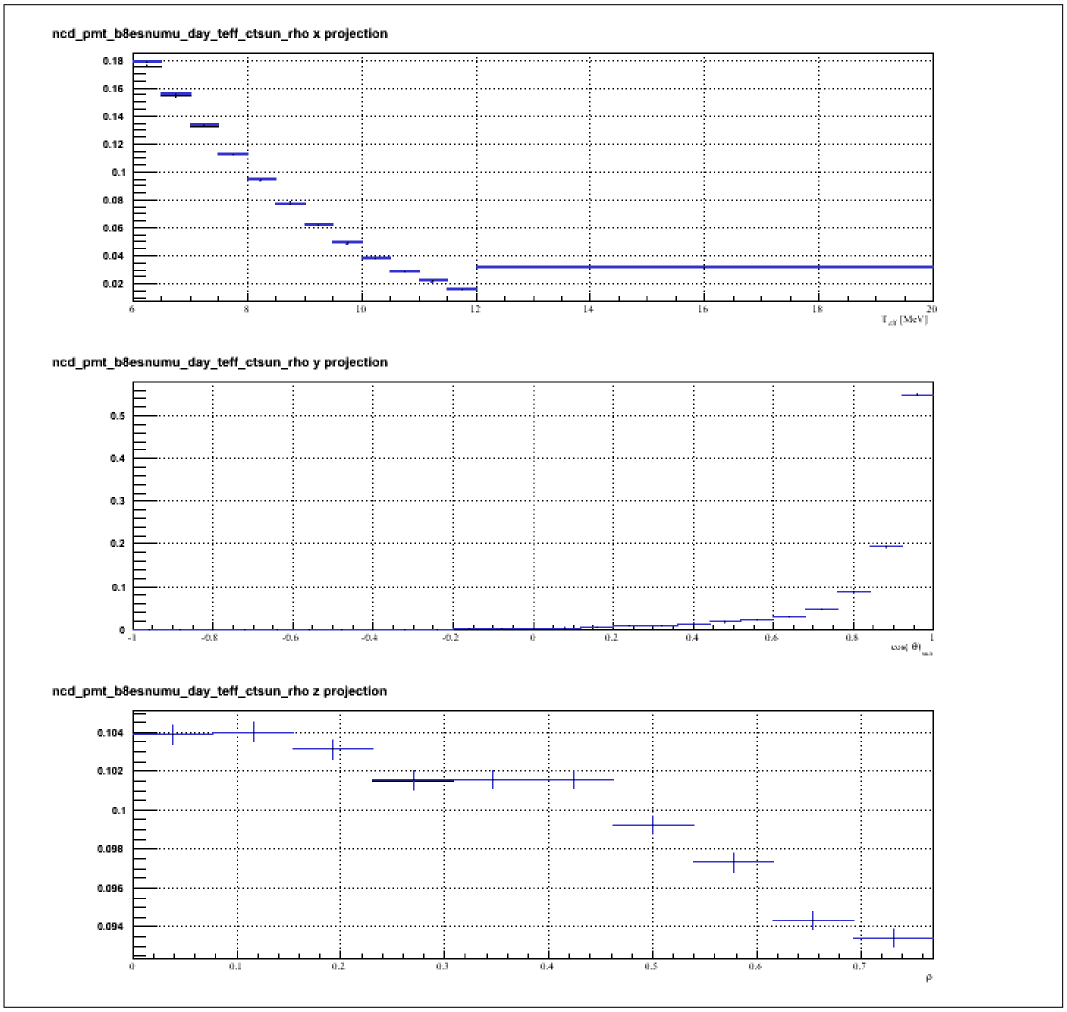


Figure B.29: Comparing three projections (energy, cos $\theta_{\odot}$, ρ) of the distorted 3D PDFs from MCMC to the QSigEx using the nominal values of p_{ee}. The 3D PDF is for the $ES_{\mu\tau}$ Day class.

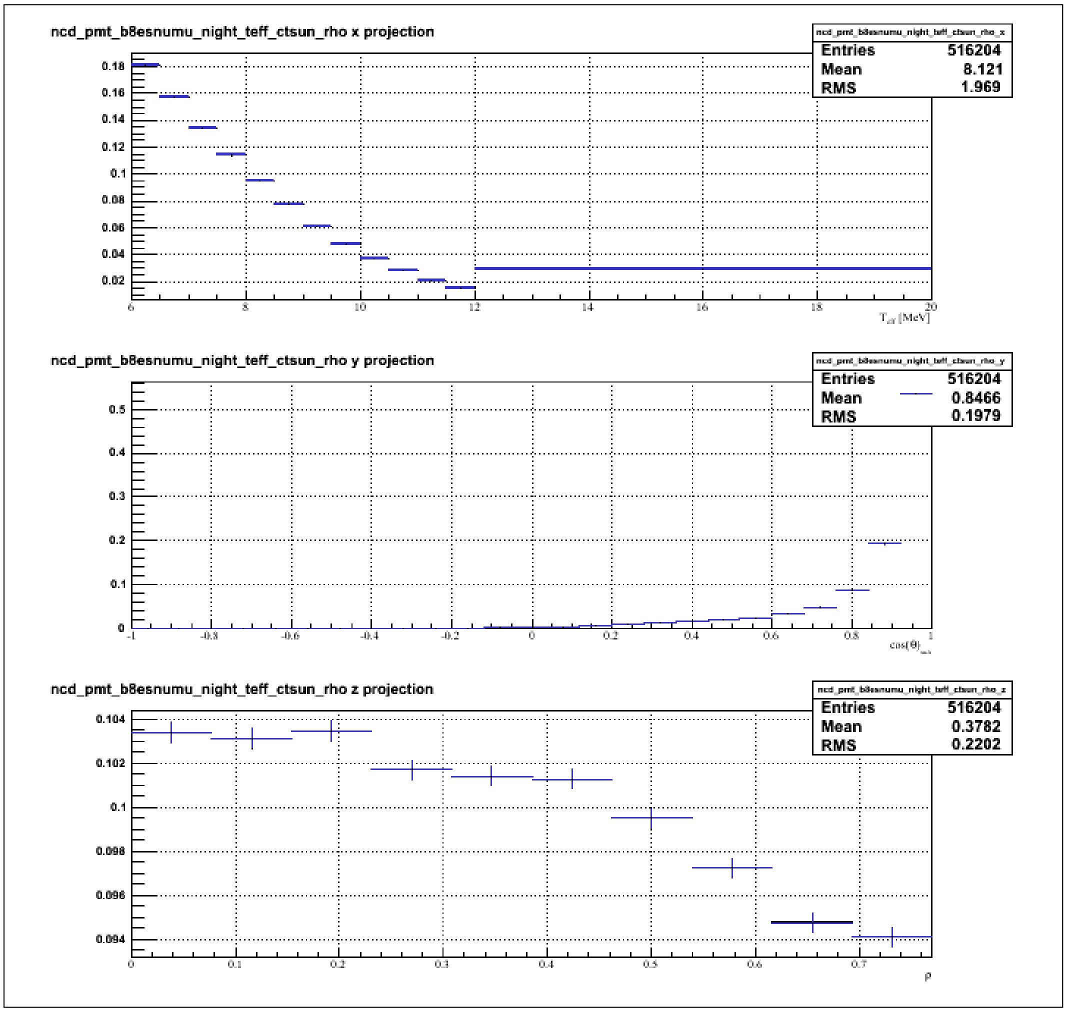


Figure B.30: Comparing three projections (energy, cos $\theta_{\odot}$, ρ) of the distorted 3D PDFs from MCMC to the QSigEx using the nominal values of p_{ee}. The 3D PDF is for the $ES_{\mu\tau}$ Night class.

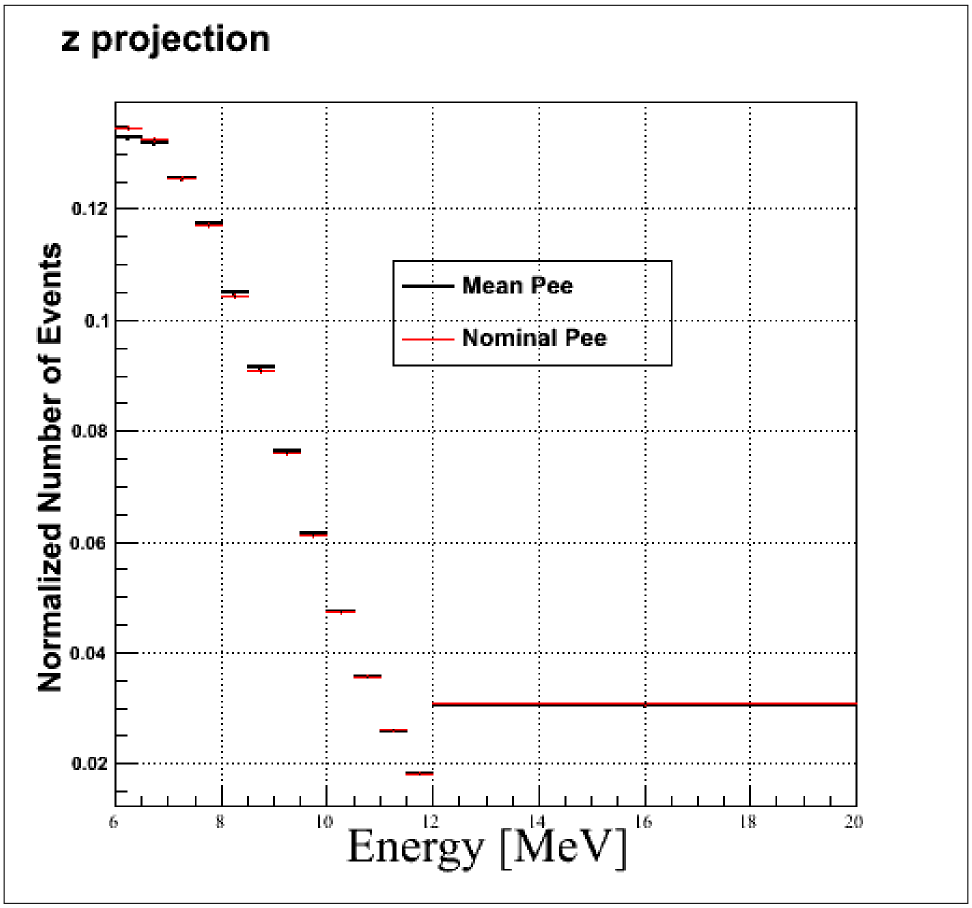


Figure B.31: Comparing energy distribution distorted using nominal values of p_{ee} to the distribution distorted using p_{ee} values, obtained from the fit as listed in Table 11.3, for the CC Day.

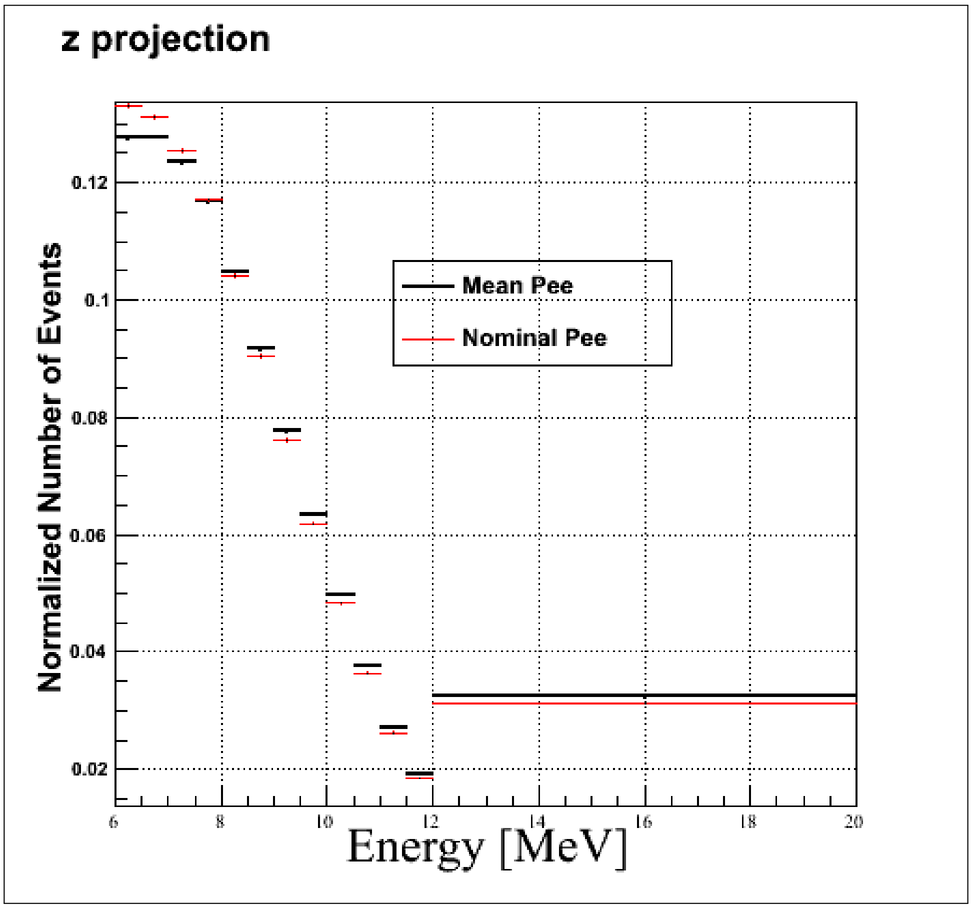


Figure B.32: Comparing energy distribution distorted using nominal values of p_{ee} to the distribution distorted using p_{ee} values, obtained from the fit as listed in Table 11.3, for the CC Night.

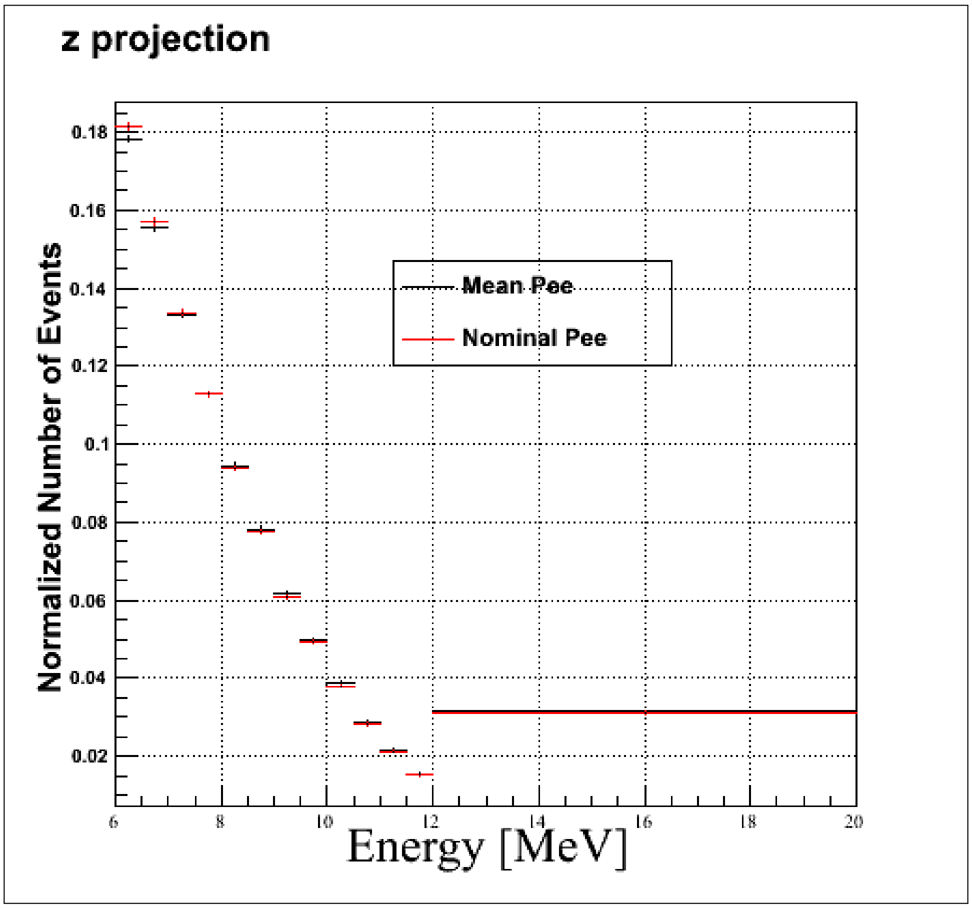


Figure B.33: Comparing energy distribution distorted using nominal values of p_{ee} to the distribution distorted using p_{ee} values, obtained from the fit as listed in Table 11.3, for the ES Day.

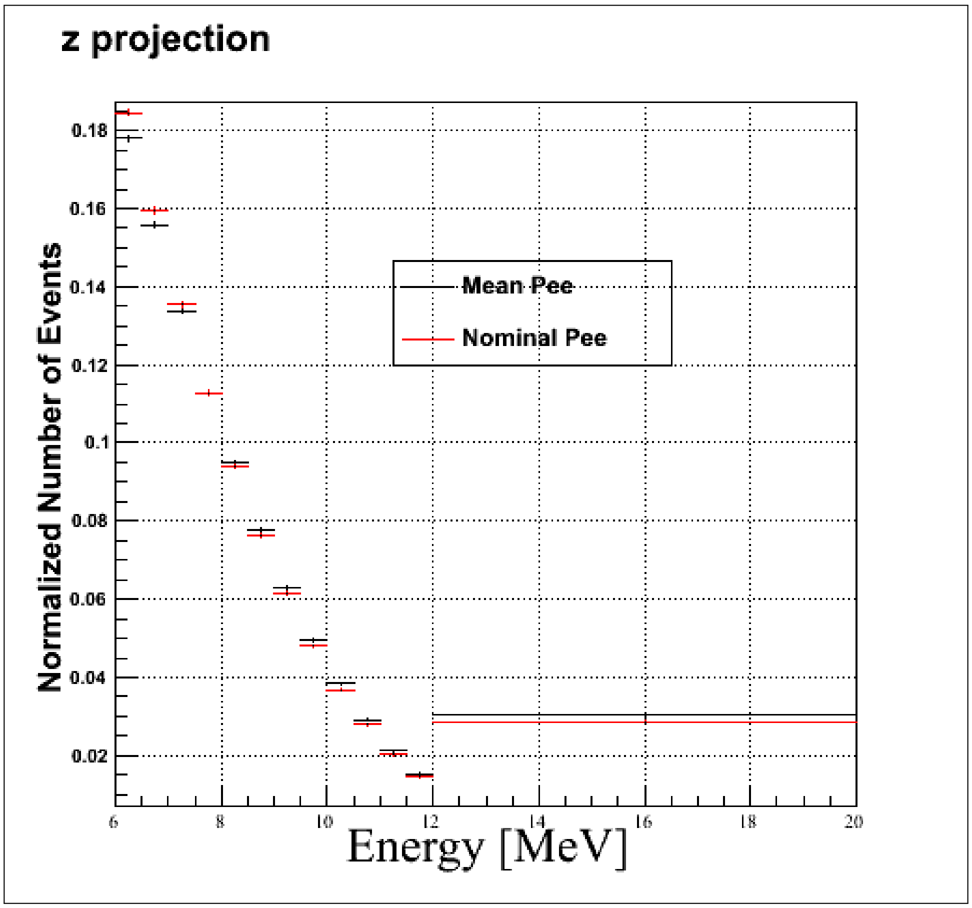


Figure B.34: Comparing energy distribution distorted using nominal values of p_{ee} to the distribution distorted using p_{ee} values, obtained from the fit as listed in Table 11.3, for the ES Night.

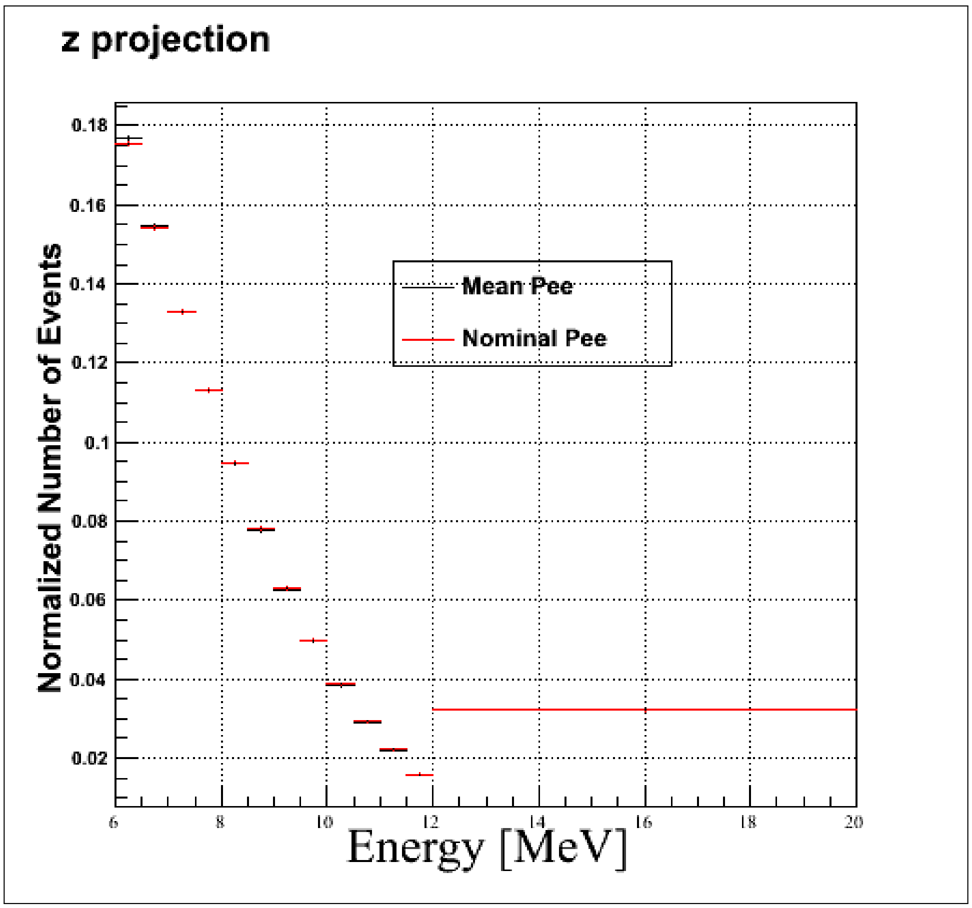


Figure B.35: Comparing energy distribution distorted using nominal values of p_{ee} to the distribution distorted using p_{ee} values, obtained from the fit as listed in Table 11.3, for the $\mathrm{ES}_{\mu\tau}$ Day.

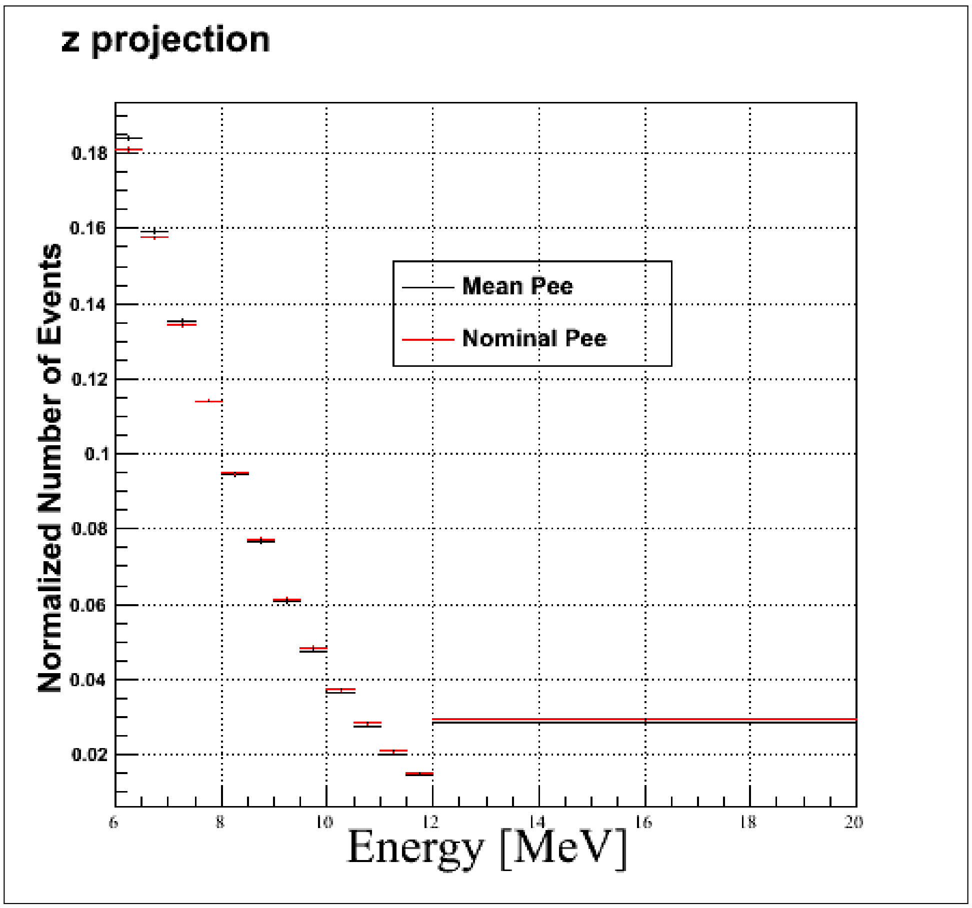


Figure B.36: Comparing energy distribution distorted using nominal values of p_{ee} to the distribution distorted using p_{ee} values, obtained from the fit as listed in Table 11.3, for the $\mathrm{ES}_{\mu\tau}$ Night.

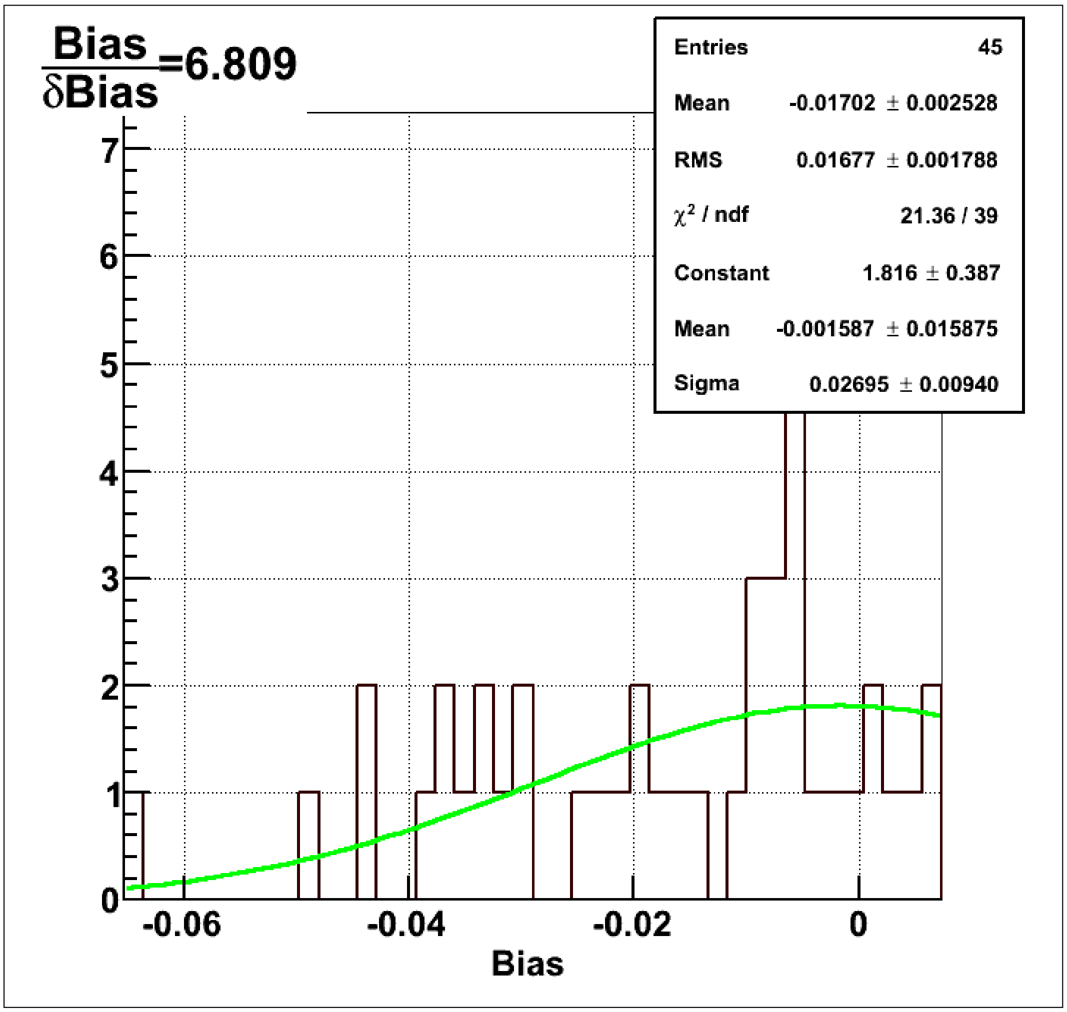


Figure B.37: The MCMC fit with energy range reduced from 6 to 12 MeV instead of 6 to 20 MeV. The green line is a Gaussian fit of the histogram.

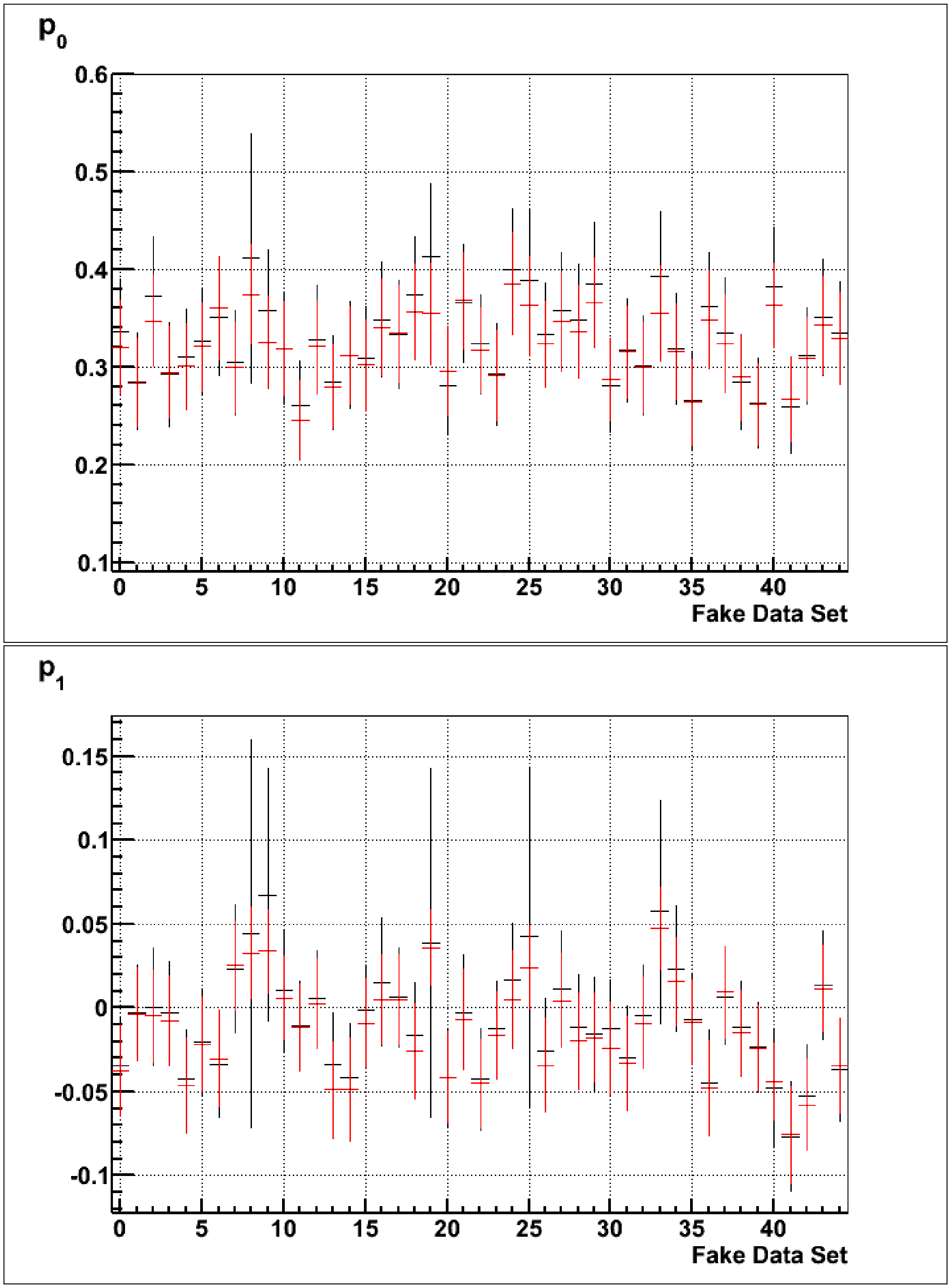


Figure B.38: Comparing P_{ee} parameters (p_0 and p_1) from file to file. Forty five regular simulated datasets were fitted. The X axis shows the file number. Red shows results from QSigEx and black shows from MCMC. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were not floated.

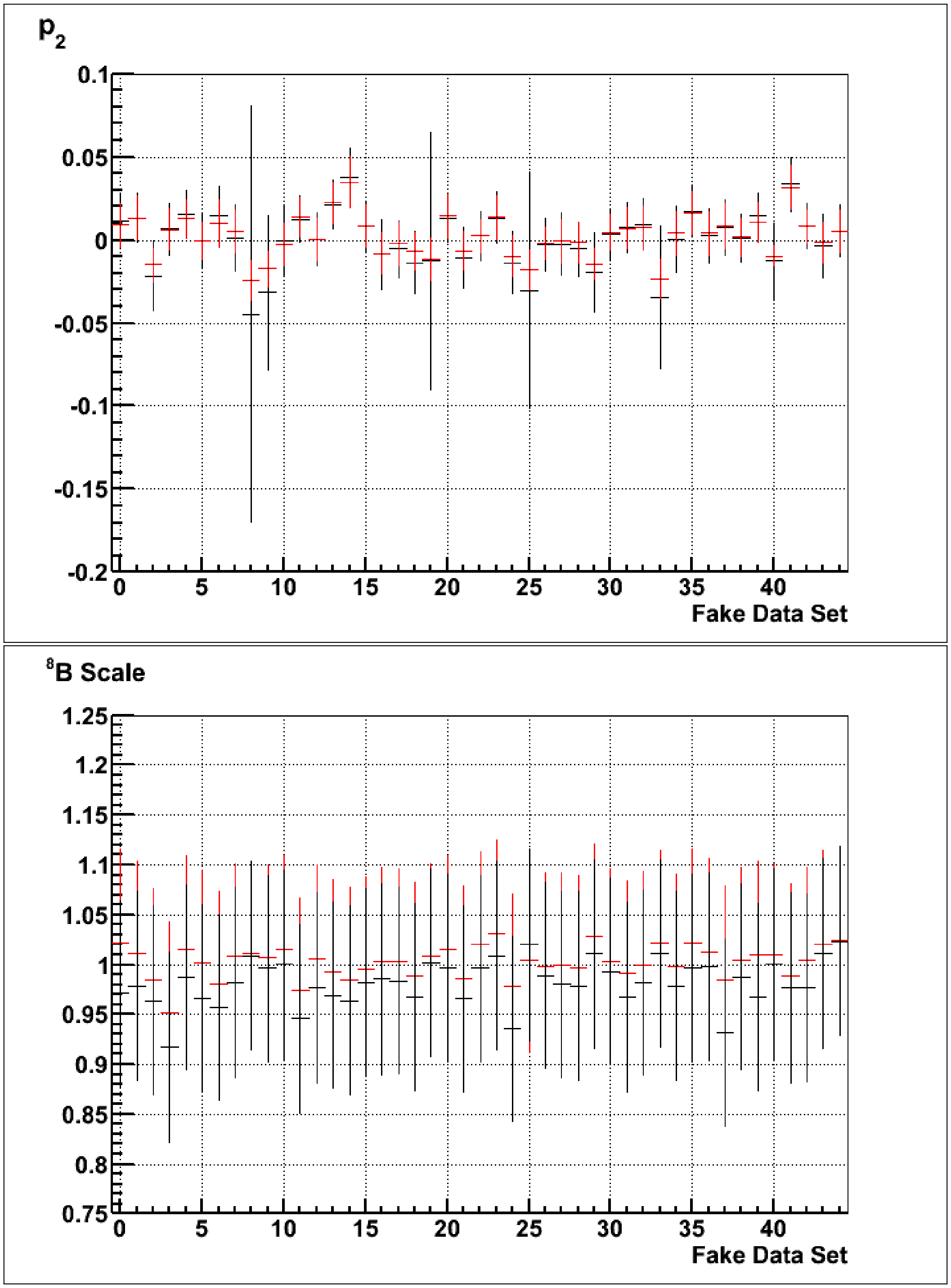


Figure B.39: Comparing P_{ee} parameter p_2 and ^8B scale from file to file. Forty five regular simulated datasets were fitted. The X axis shows the file number. Red shows results from QSigEx and black shows from MCMC. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were not floated.

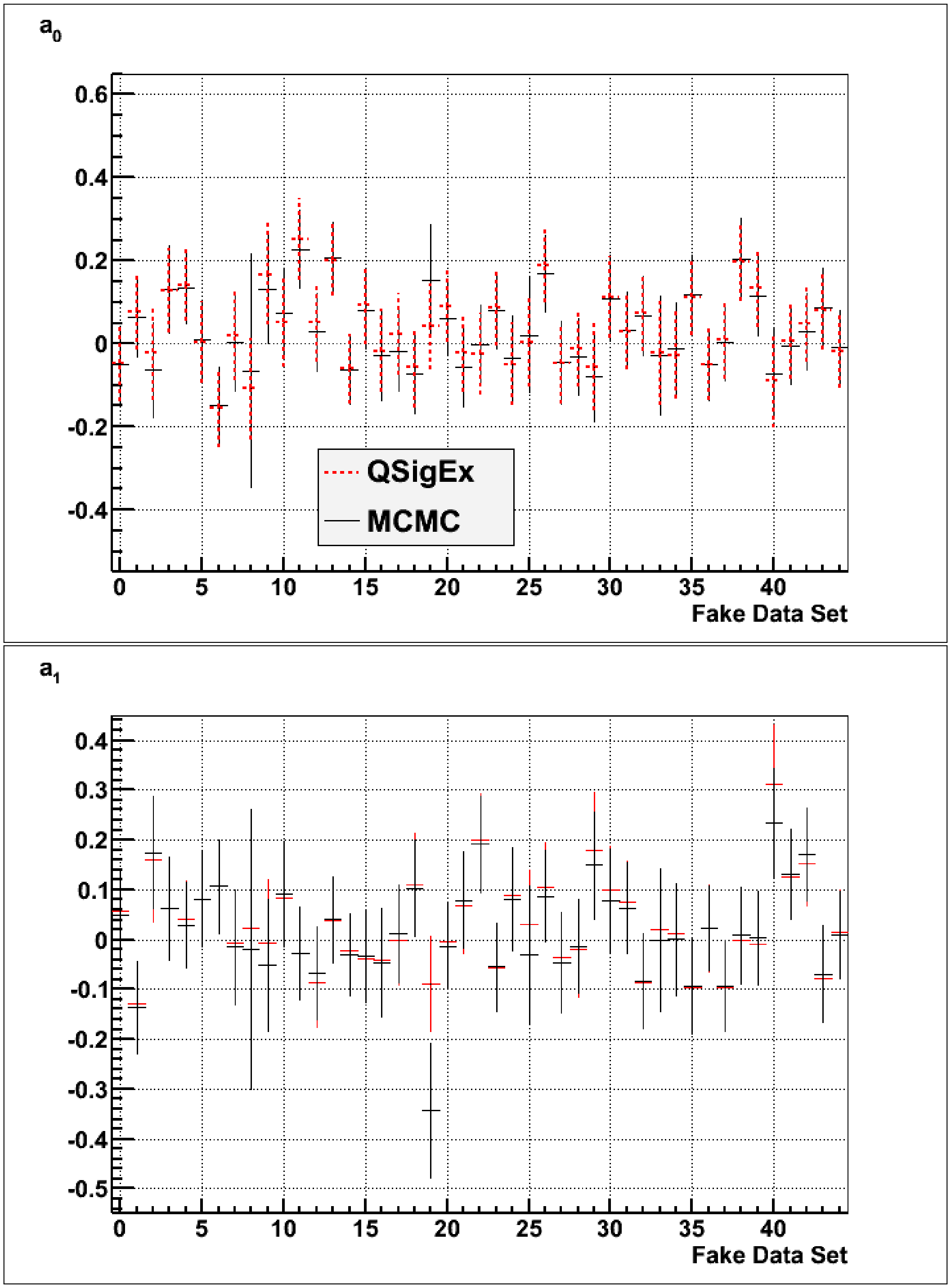


Figure B.40: Comparing day-night asymmetries from file to file. Forty five regular simulated datasets were fitted. The X axis shows the file number. Red shows results from QSigEx and black shows from MCMC. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were not floated.

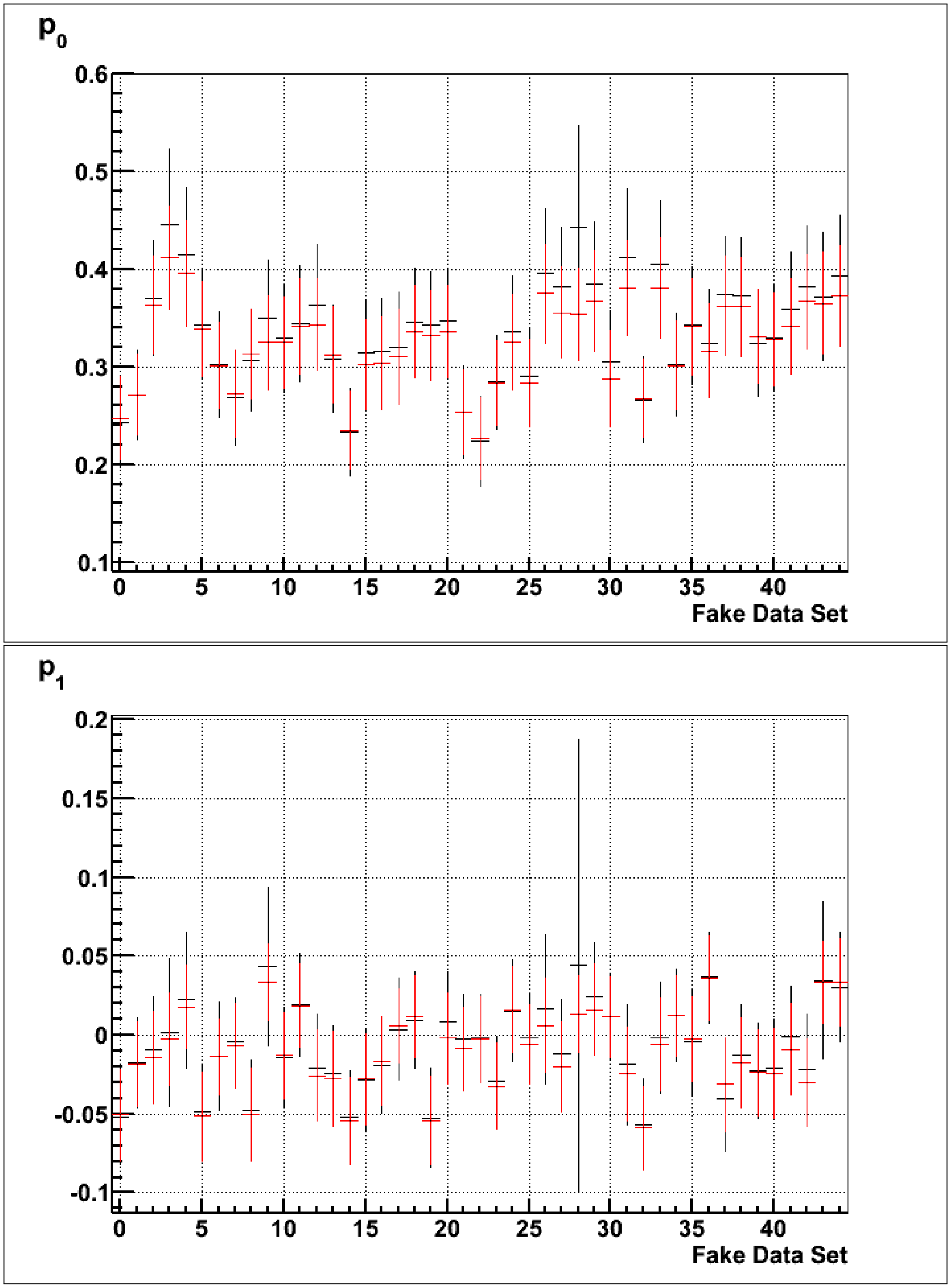


Figure B.41: Comparing P_{ee} (p_0 and p_1) parameters from file to file. Forty five alternate simulated datasets were fitted. The X axis shows the file number. The best-fit from QSigEx along with its uncertainty is shown in red and black shows corresponding result from the MCMC fit. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were not floated.

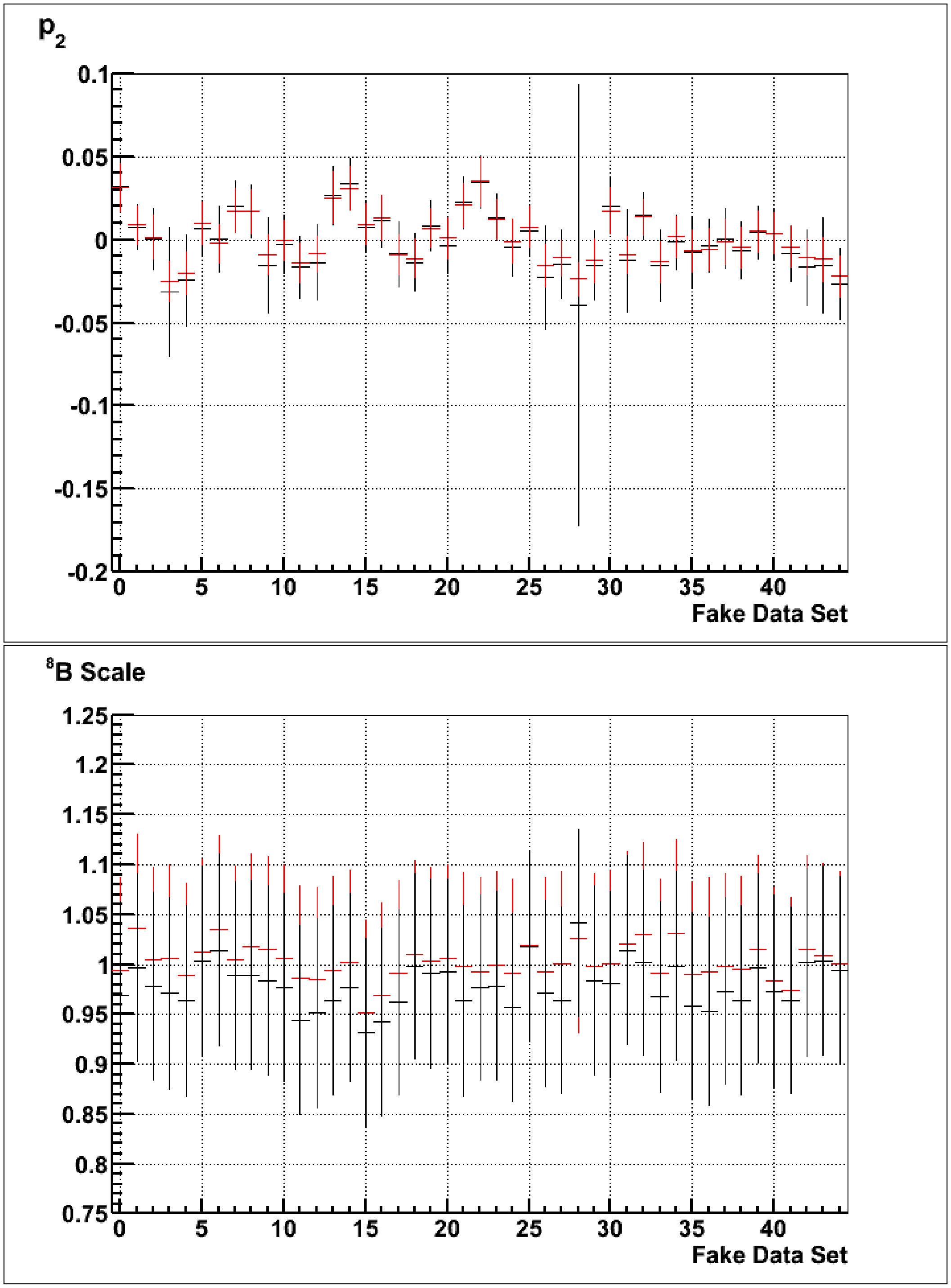


Figure B.42: Comparing P_{ee} parameter (p_2) and ^{8}B scale from file to file. Forty five alternate simulated datasets were fitted. The X axis shows the file number. The best-fit from QSigEx along with its uncertainty is shown in red and black shows corresponding result from MCMC fit. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Systematics were not floated.

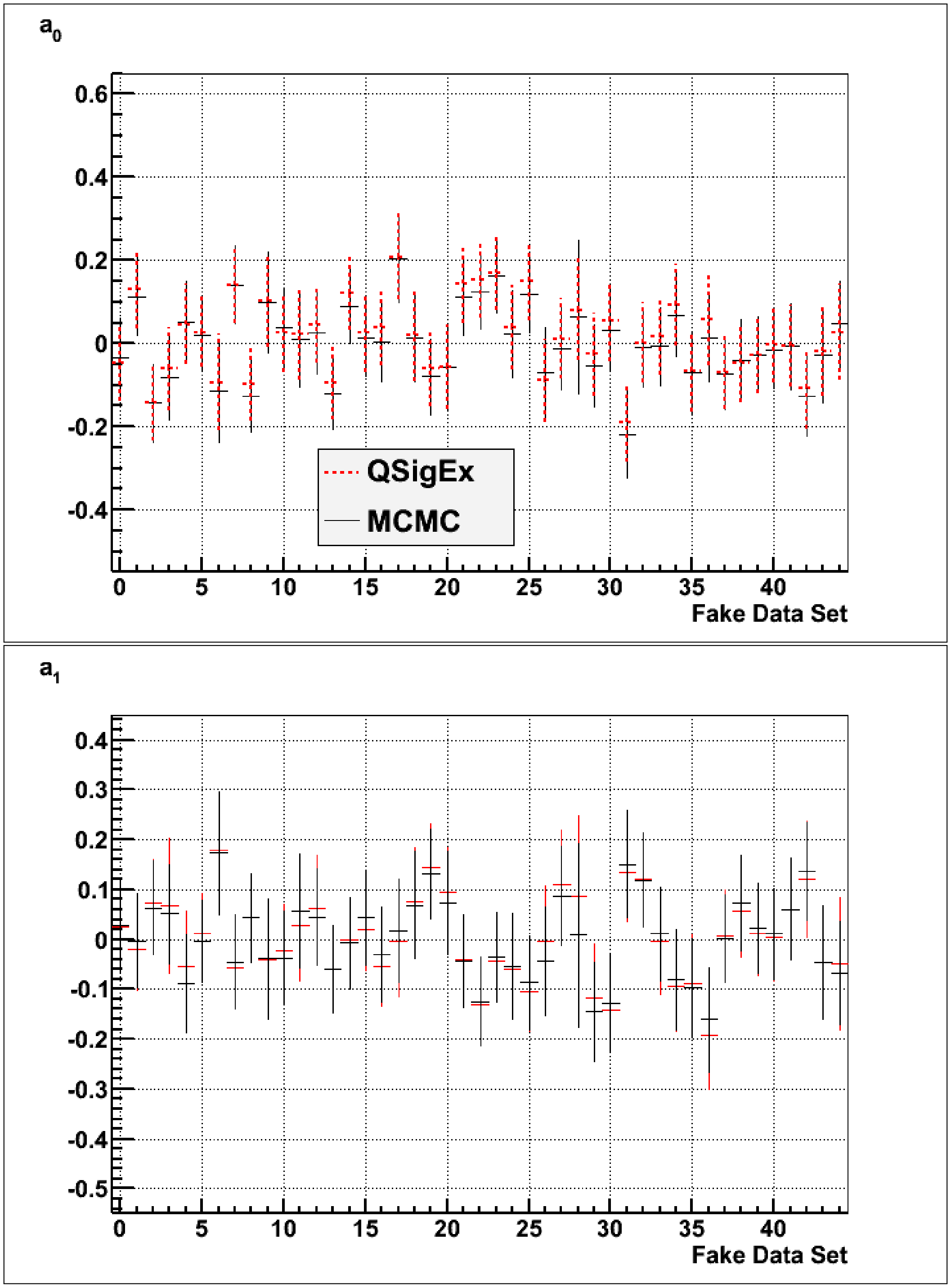
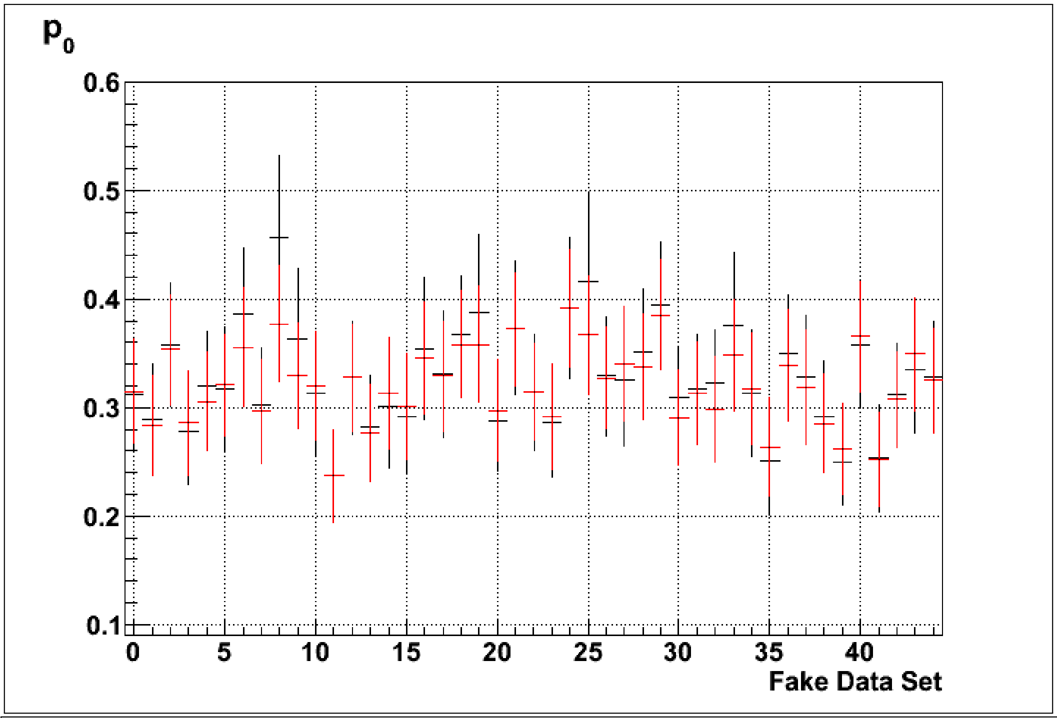


Figure B.43: Comparing day-night asymmetries from file to file. Forty five alternate simulated datasets were fitted. The X axis shows the file number. The best-fit from QSigEx along with its uncertainty is shown in red and black shows corresponding result from MCMC fit. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were not floated.



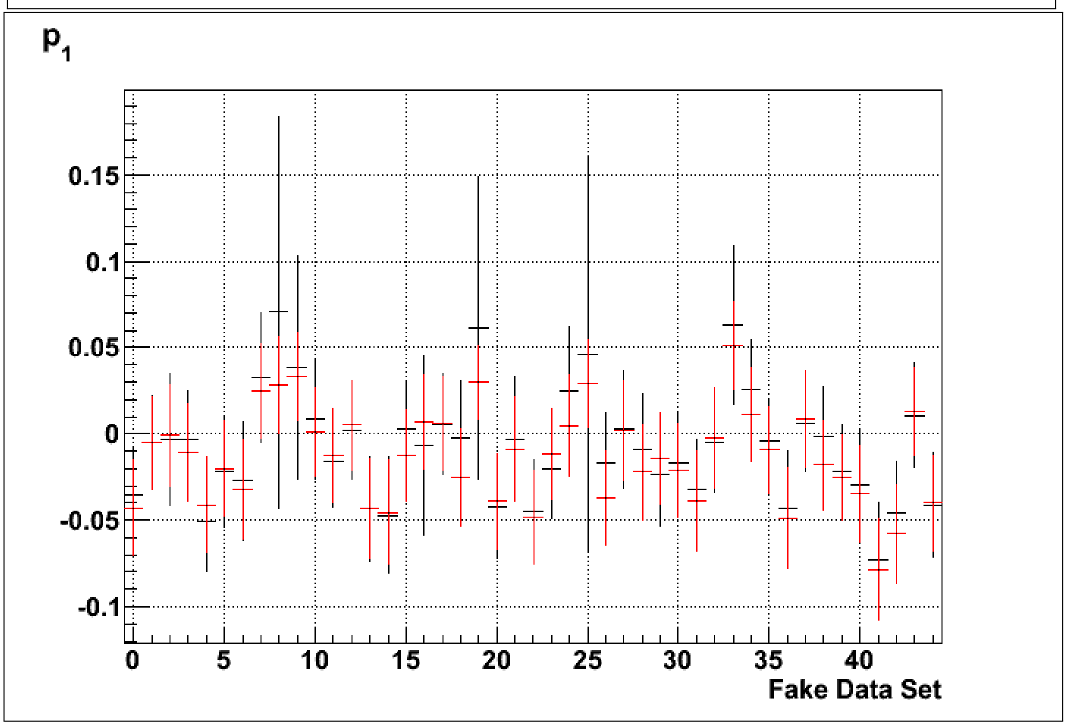


Figure B.44: Comparing P_{ee} parameters (p_0 and p_1) from file to file. Forty five regular simulated datasets were fitted. The X axis shows the file number. The best-fit from QSigEx along with its uncertainty is shown in red and black shows corresponding result from MCMC fit. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were floated.

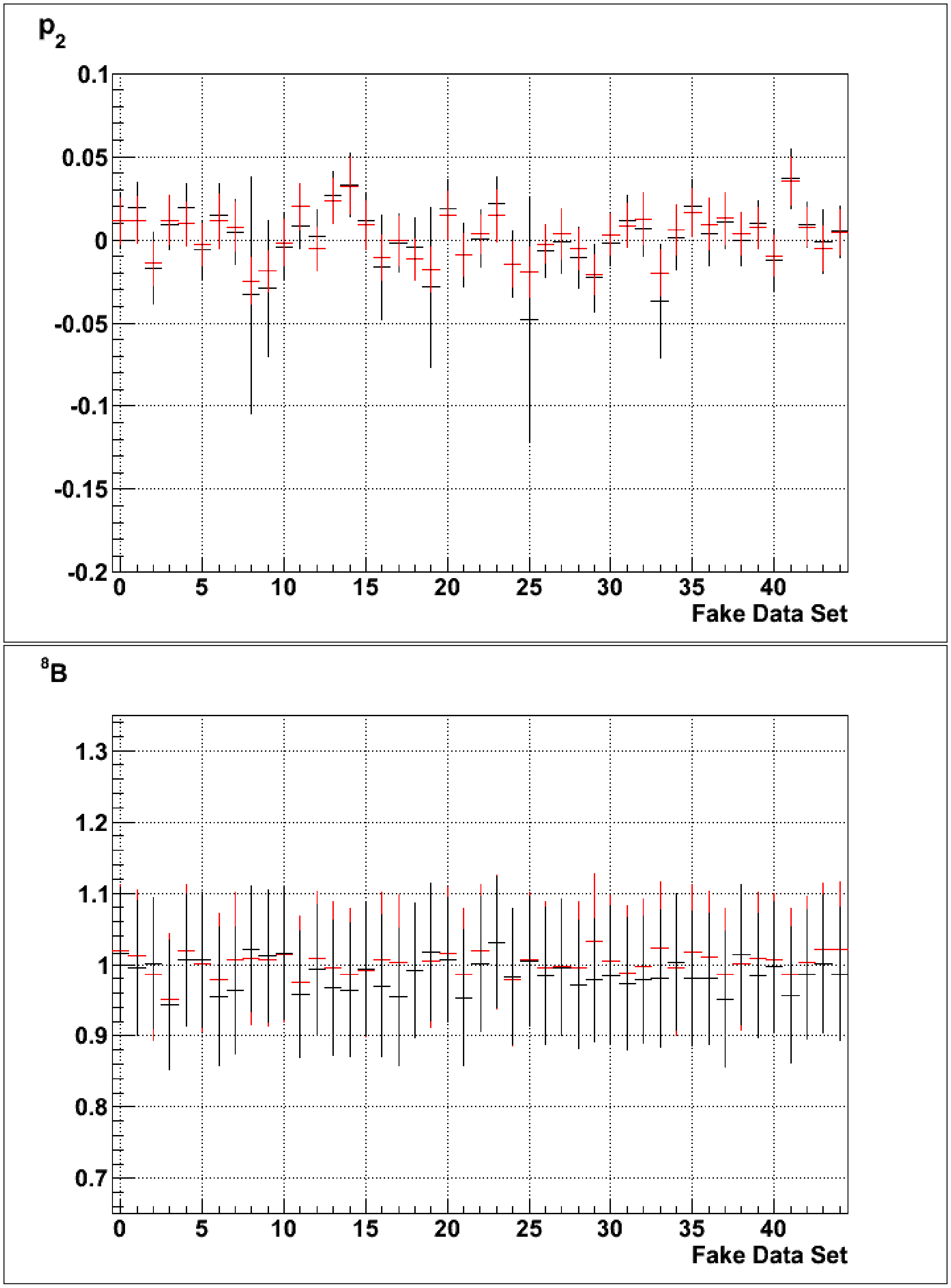
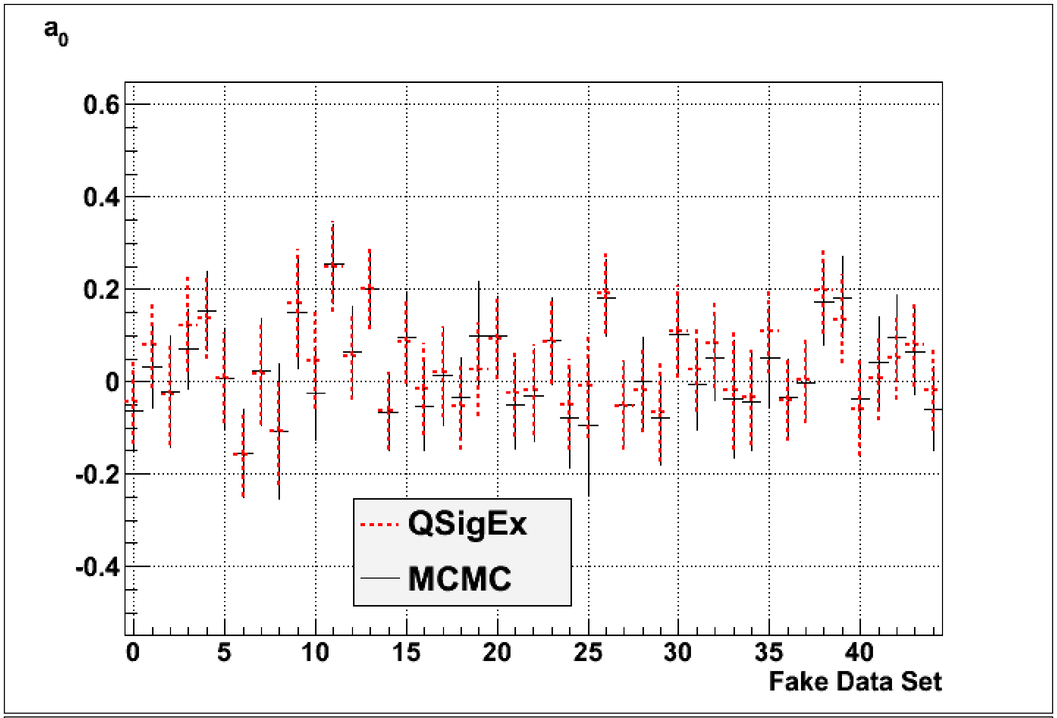


Figure B.45: Comparing P_{ee} parameter (p_2) and ^{8}B scale from file to file. Forty five regular simulated datasets were fitted. The X axis shows the file number. The best-fit from QSigEx along with its uncertainty is shown in red and black shows corresponding result from MCMC fit. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were floated.



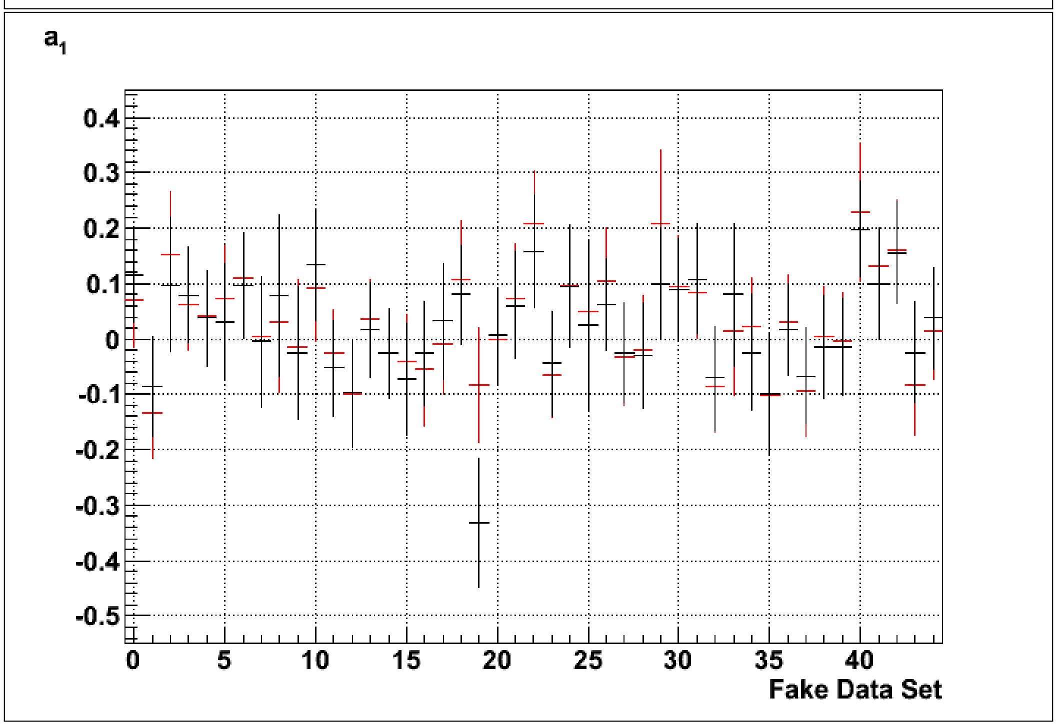


Figure B.46: Comparing day-night asymmetries from file to file. Forty five regular simulated datasets were fitted. The X axis shows the file number. The best-fit from QSigEx along with its uncertainty is shown in red and black shows corresponding result from MCMC fit. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were floated.

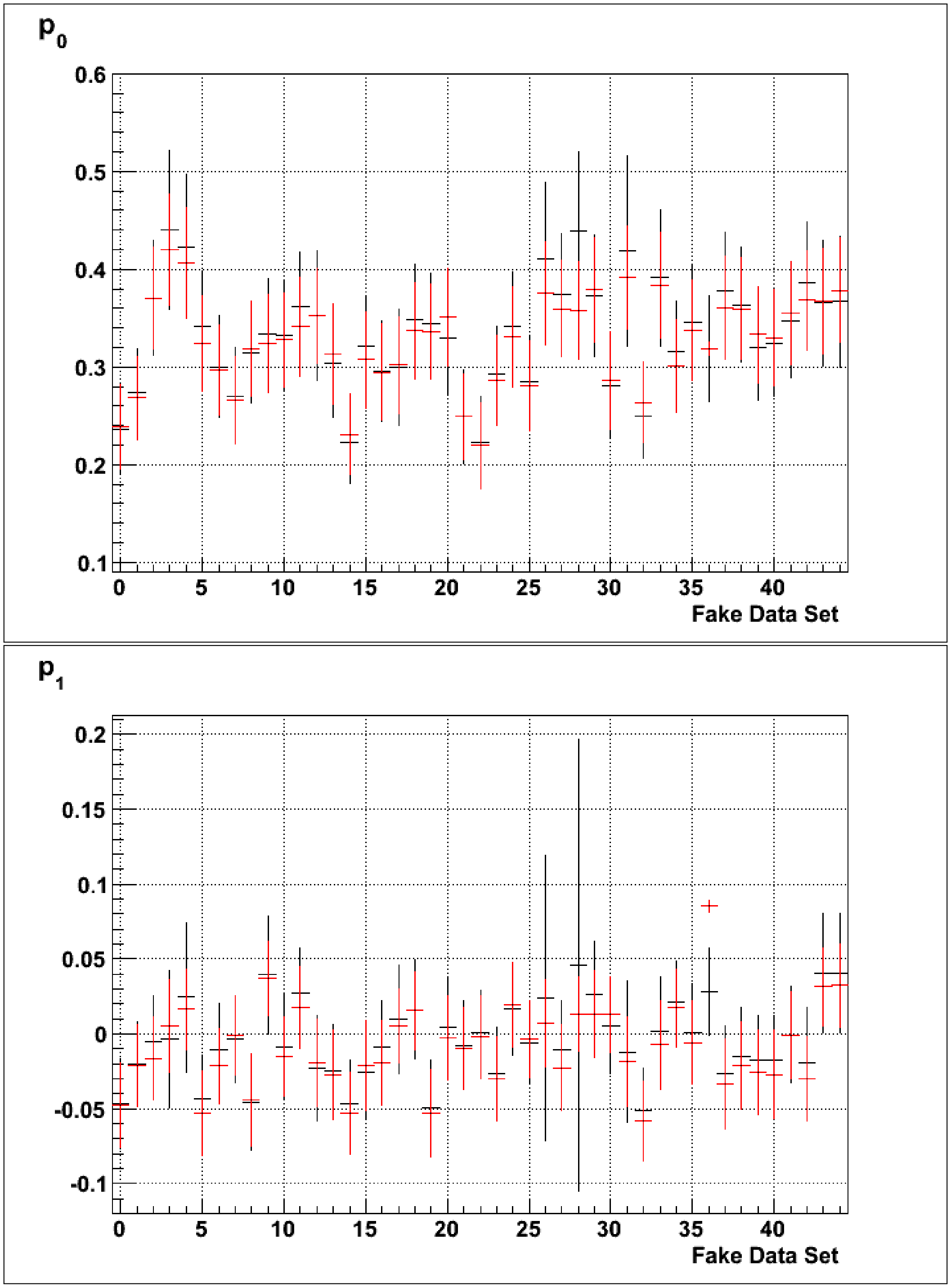


Figure B.47: Comparing P_{ee} parameters (p_0 and p_1) from file to file. Forty five alternate simulated datasets were fitted. The X axis shows the file number. The best-fit from QSigEx along with its uncertainty is shown in red and black shows corresponding result from MCMC fit. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were floated.

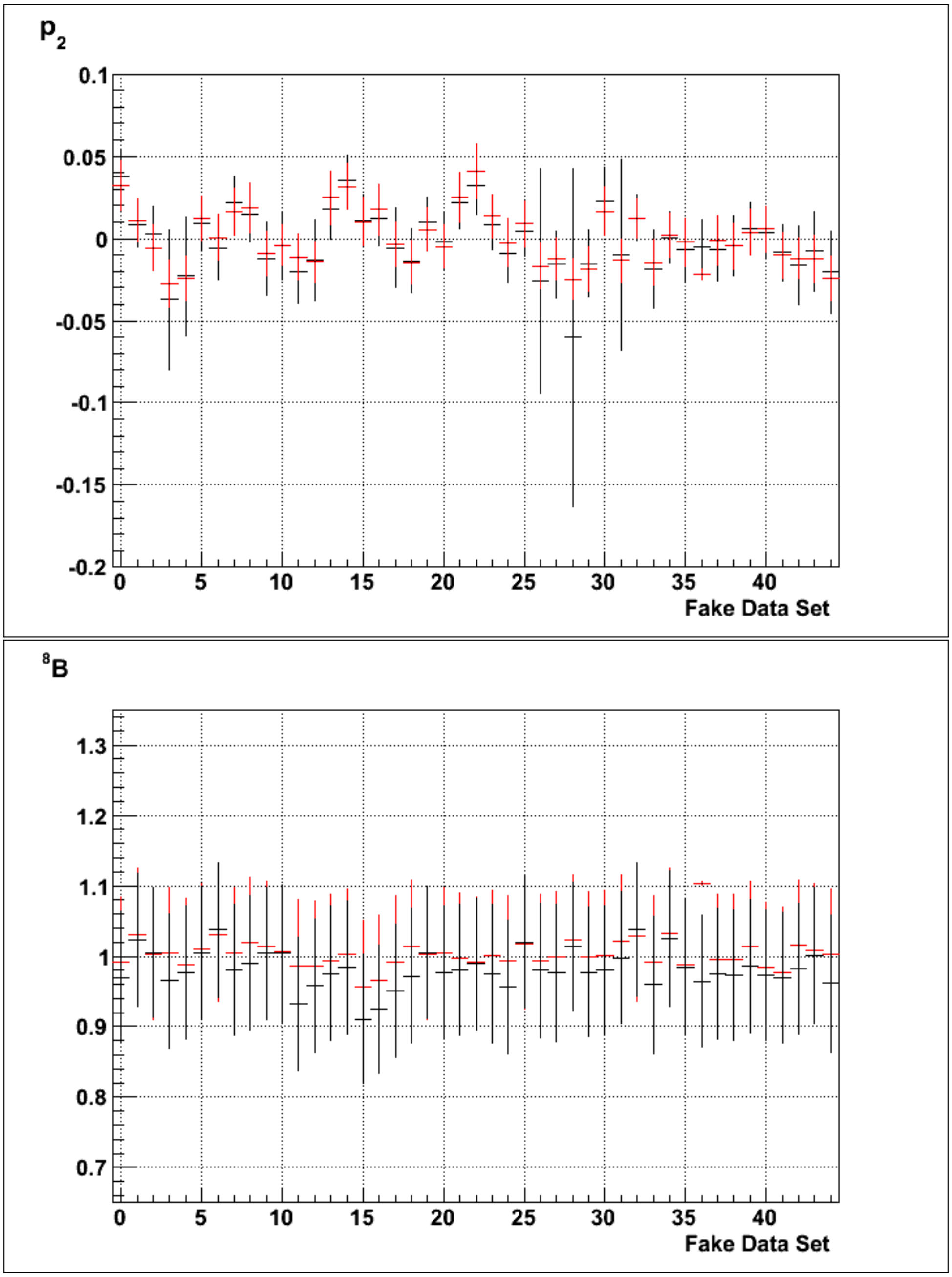


Figure B.48: Comparing P_{ee} parameter (p_2) and ^{8}B scale from file to file. Forty five alternate simulated datasets were fitted. The X axis shows the file number. The best-fit from QSigEx along with its uncertainty is shown in red and black shows corresponding result from MCMC fit. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were floated.

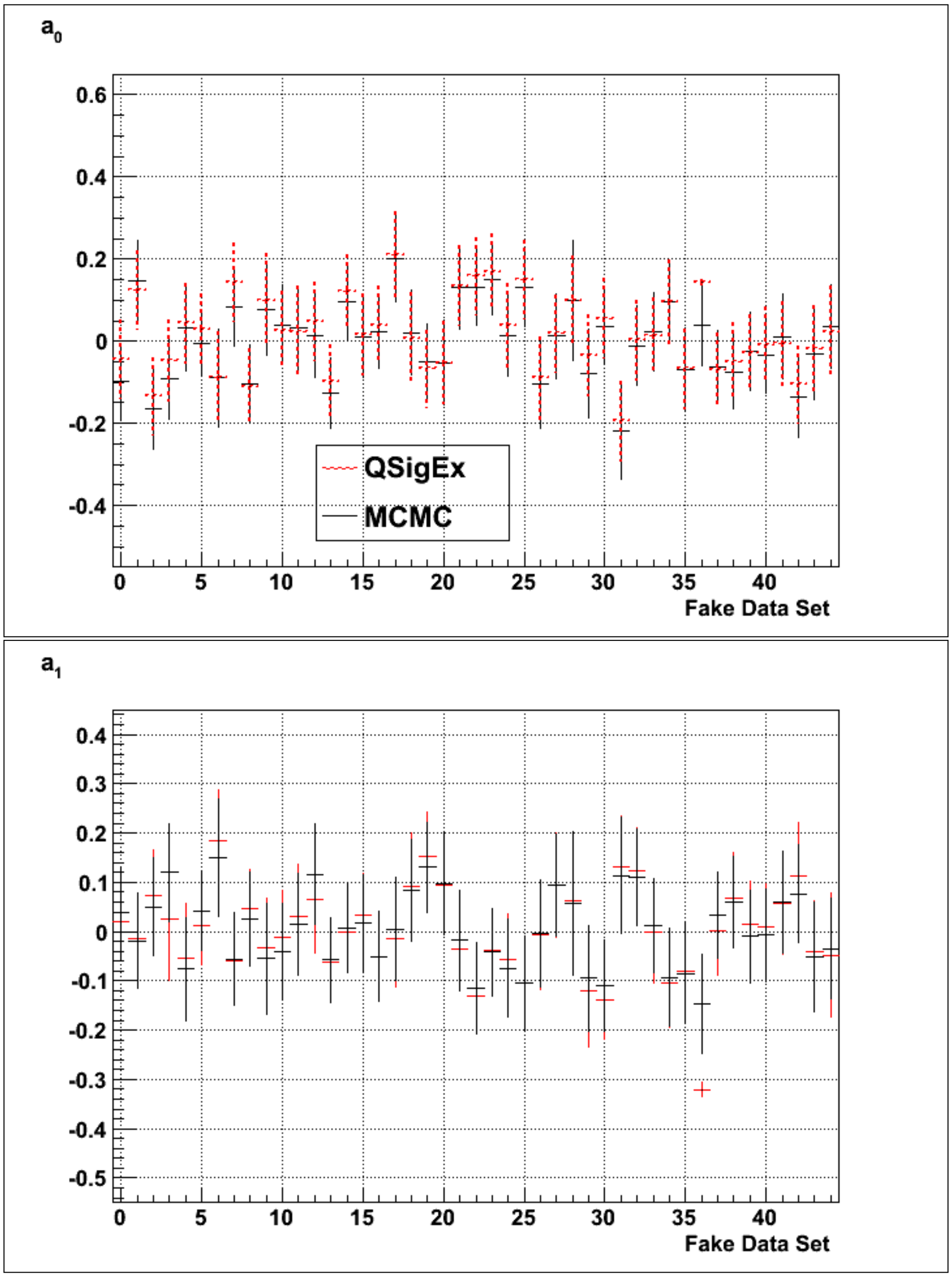


Figure B.49: Comparing day-night asymmetries from file to file. Forty five alternate simulated datasets were fitted. The X axis shows the file number. The best-fit from QSigEx along with its uncertainty is shown in red and black shows corresponding result from MCMC fit. The peak and the RMS of the posterior distribution of the MCMC fit were compared to the result from QSigEx. Eight systematics were floated.

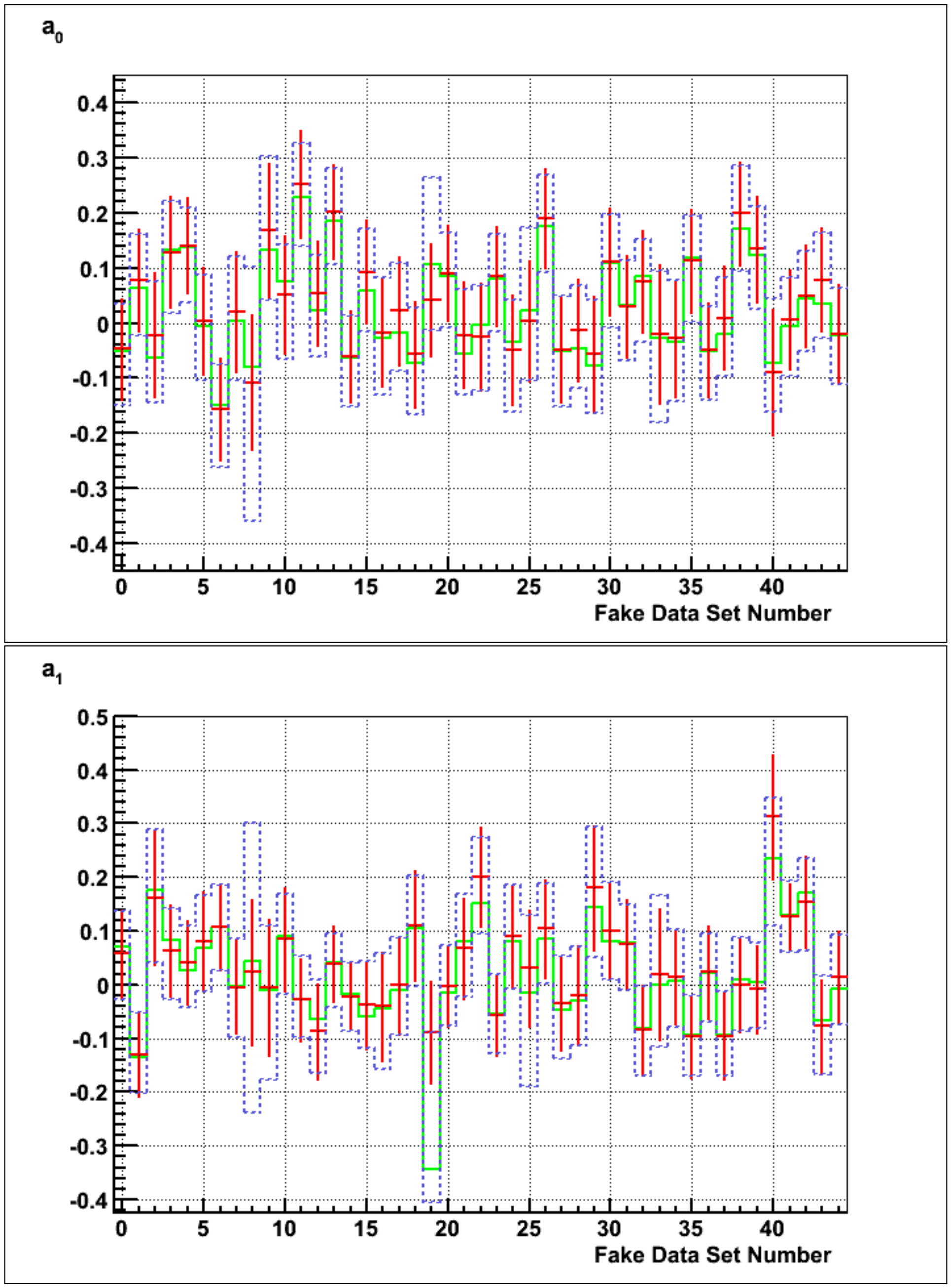


Figure B.50: Showing best-fit result in green color for day-night asymmetries (a_0 and a_1) for each of the 45 fitted regular simulated datasets shown in the X axis. Systematics were not floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red.

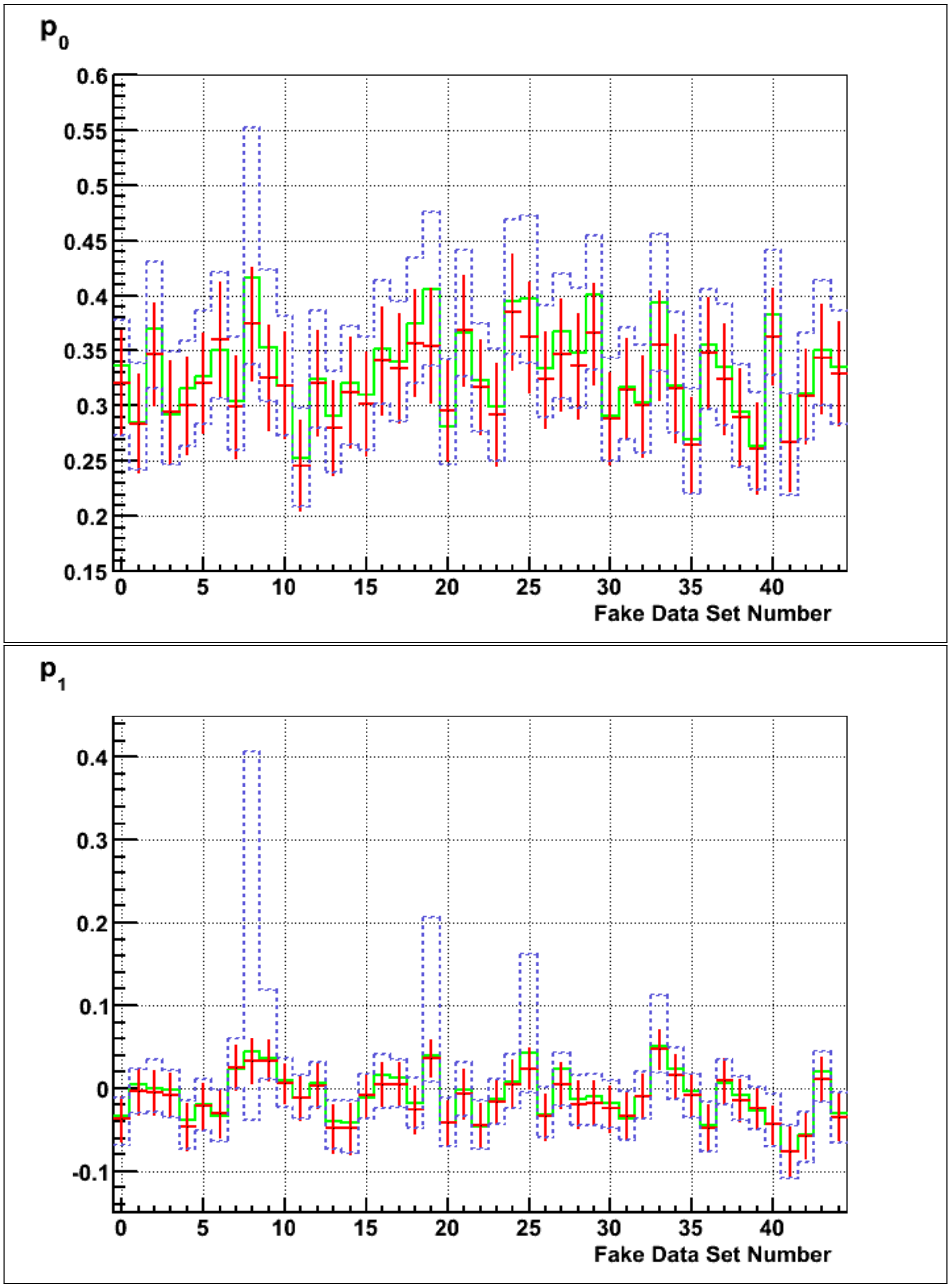


Figure B.51: Showing best-fit MCMC result in green for P_{ee} parameters (p_0 and p_1) for each of the 45 fitted regular simulated datasets shown in the X axis. Systematics were not floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red.

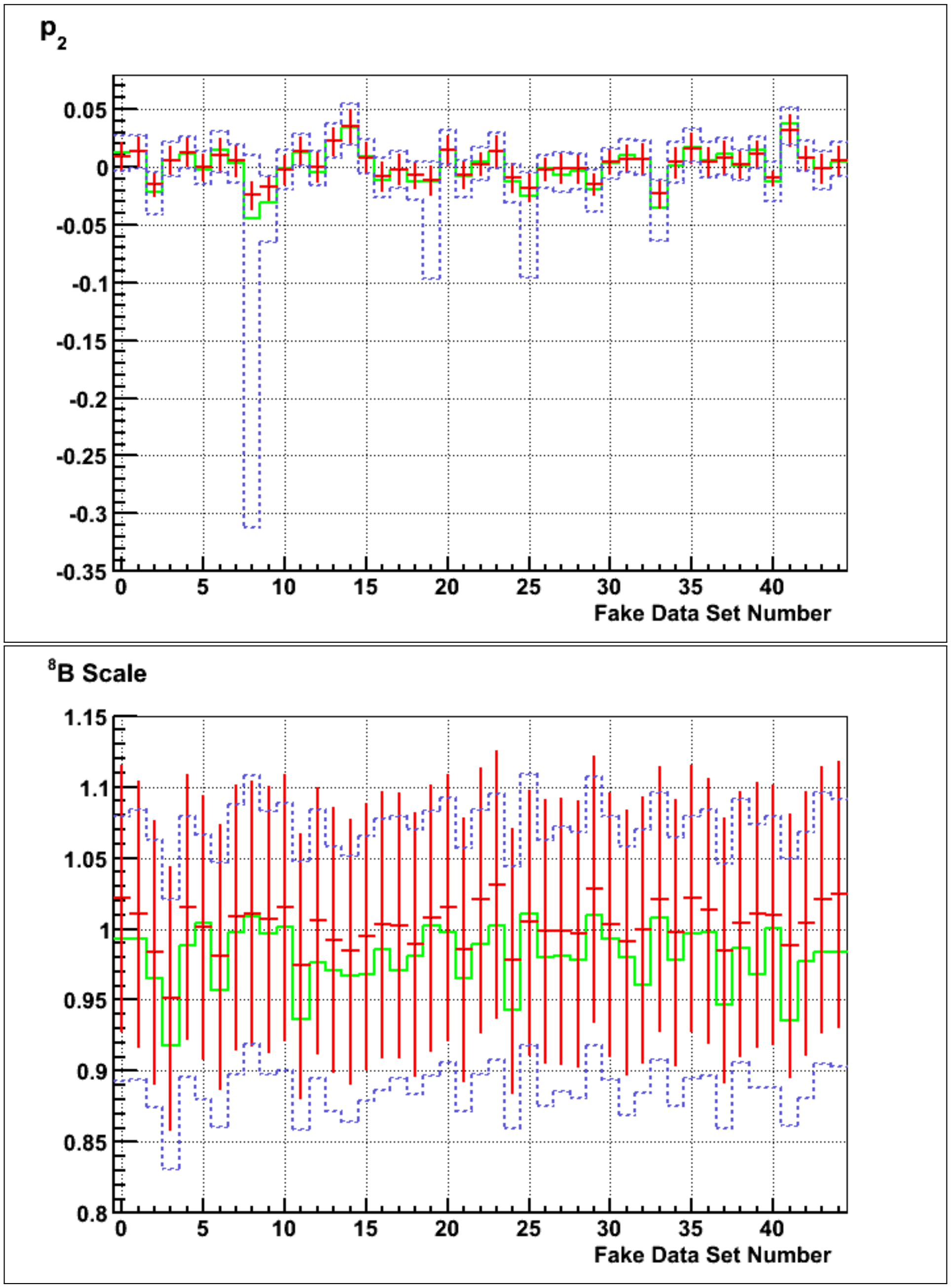


Figure B.52: Showing best-fit MCMC result (green color) of the ^{8}B Scale and P_{ee} parameter (p_2) for each of the 45 fitted regular simulated datasets shown in the X axis. Systematics were not floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red.

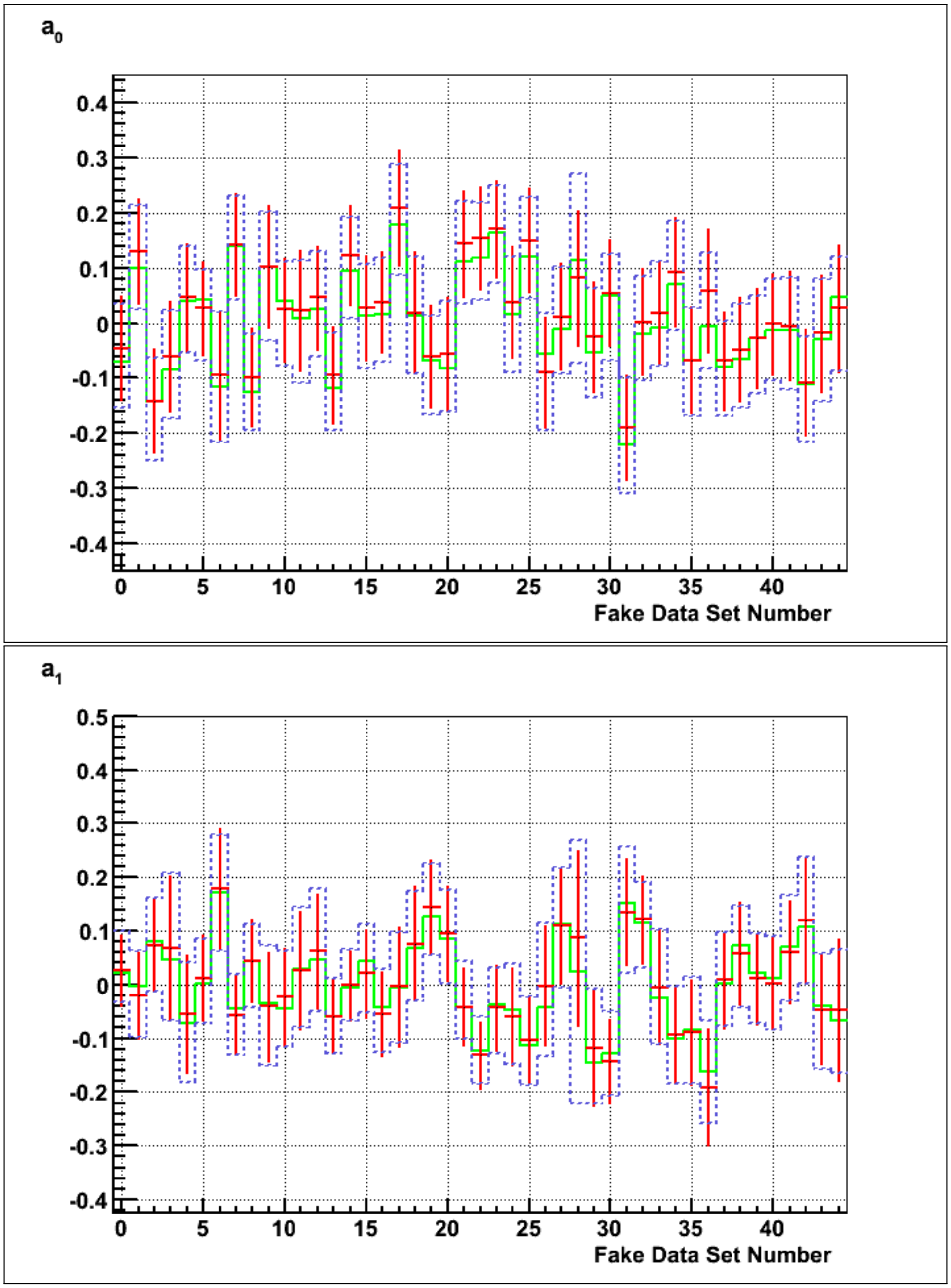
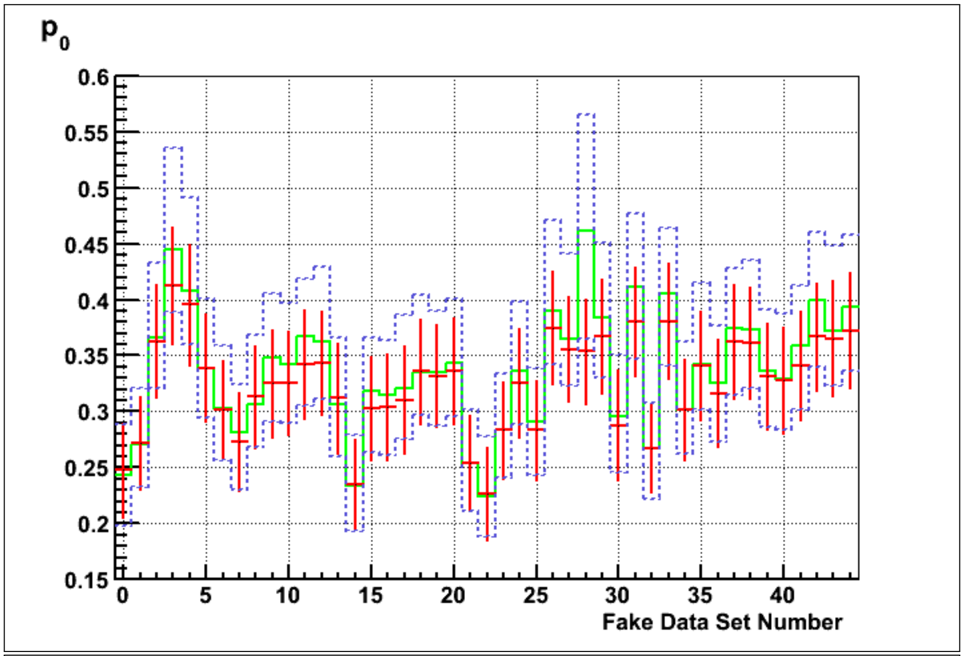


Figure B.53: Showing best-fit result in green color for day-night asymmetries (a_0 and a_1) for each of the 45 fitted alternate simulated datasets shown in the X axis. Systematics were not floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red.



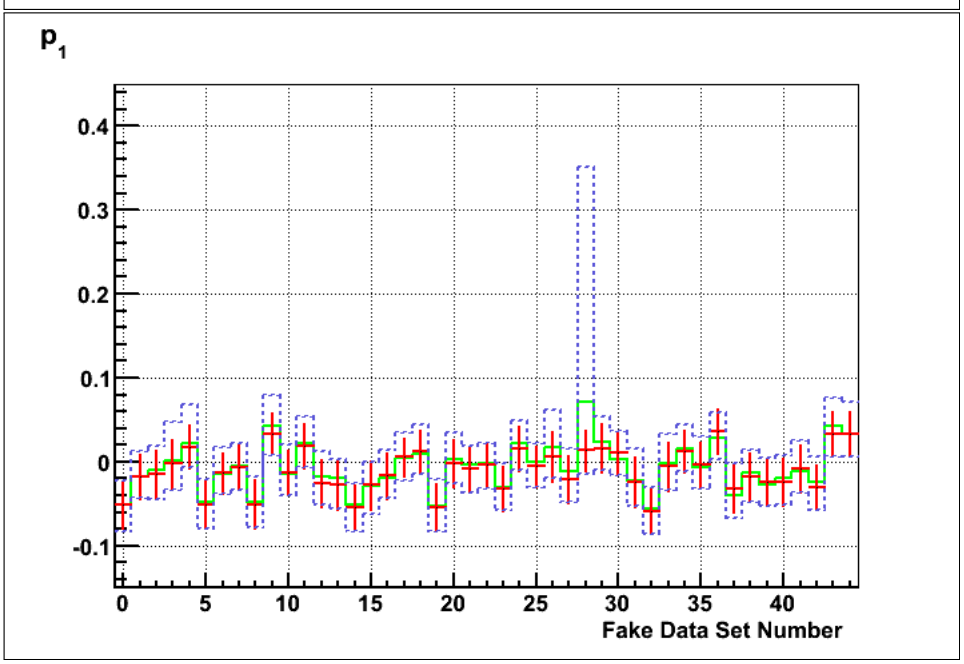


Figure B.54: Showing best-fit MCMC result in green for P_{ee} parameters (p_0 and p_1) for each of the 45 fitted alternate simulated datasets shown in the X axis. Systematics were not floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red.

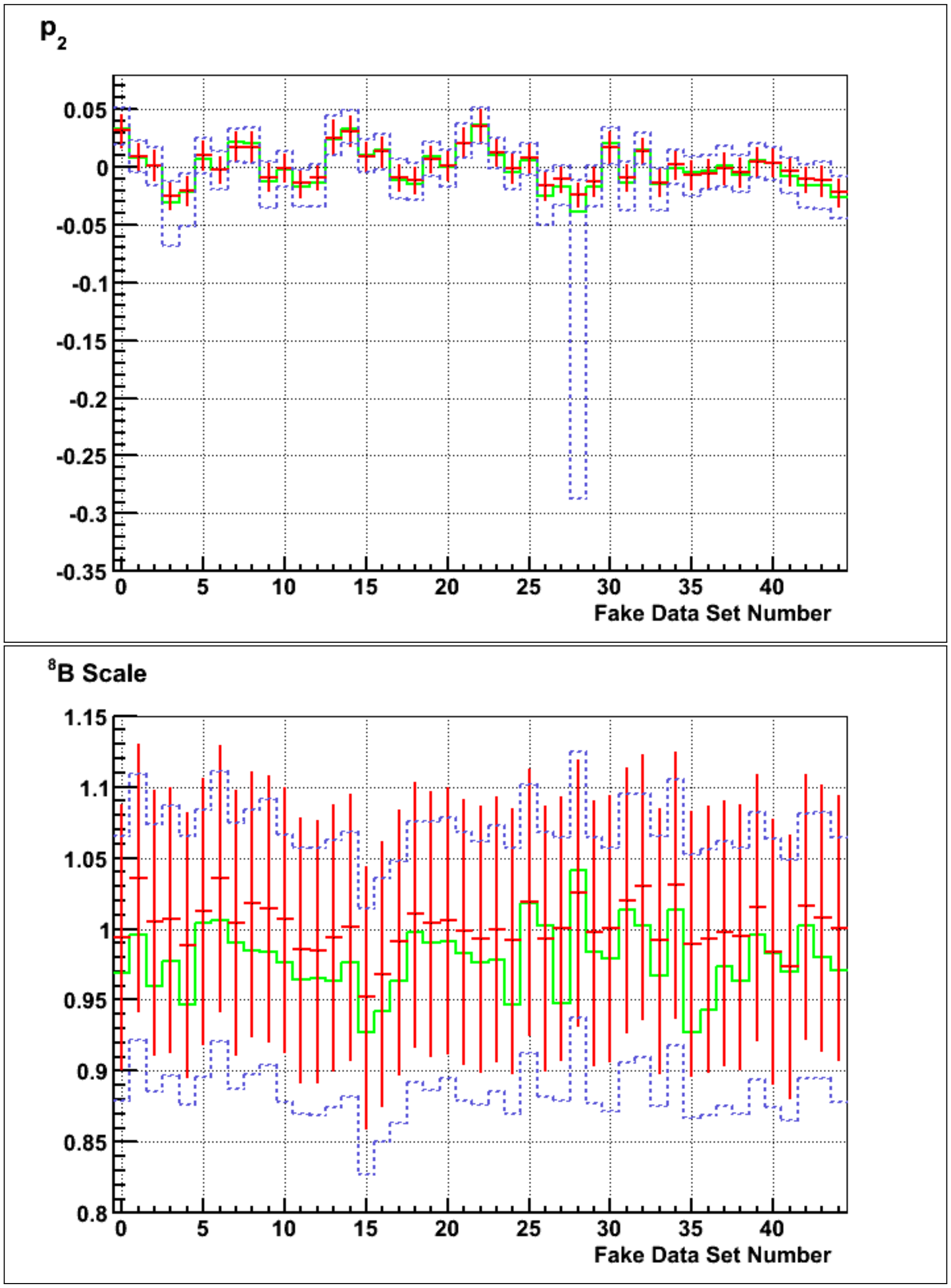


Figure B.55: Showing best-fit MCMC result (green color) of the ^{8}B Scale and P_{ee} parameter (p_2) for each of the 45 fitted alternate simulated datasets shown in the X axis. Systematics were not floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red.

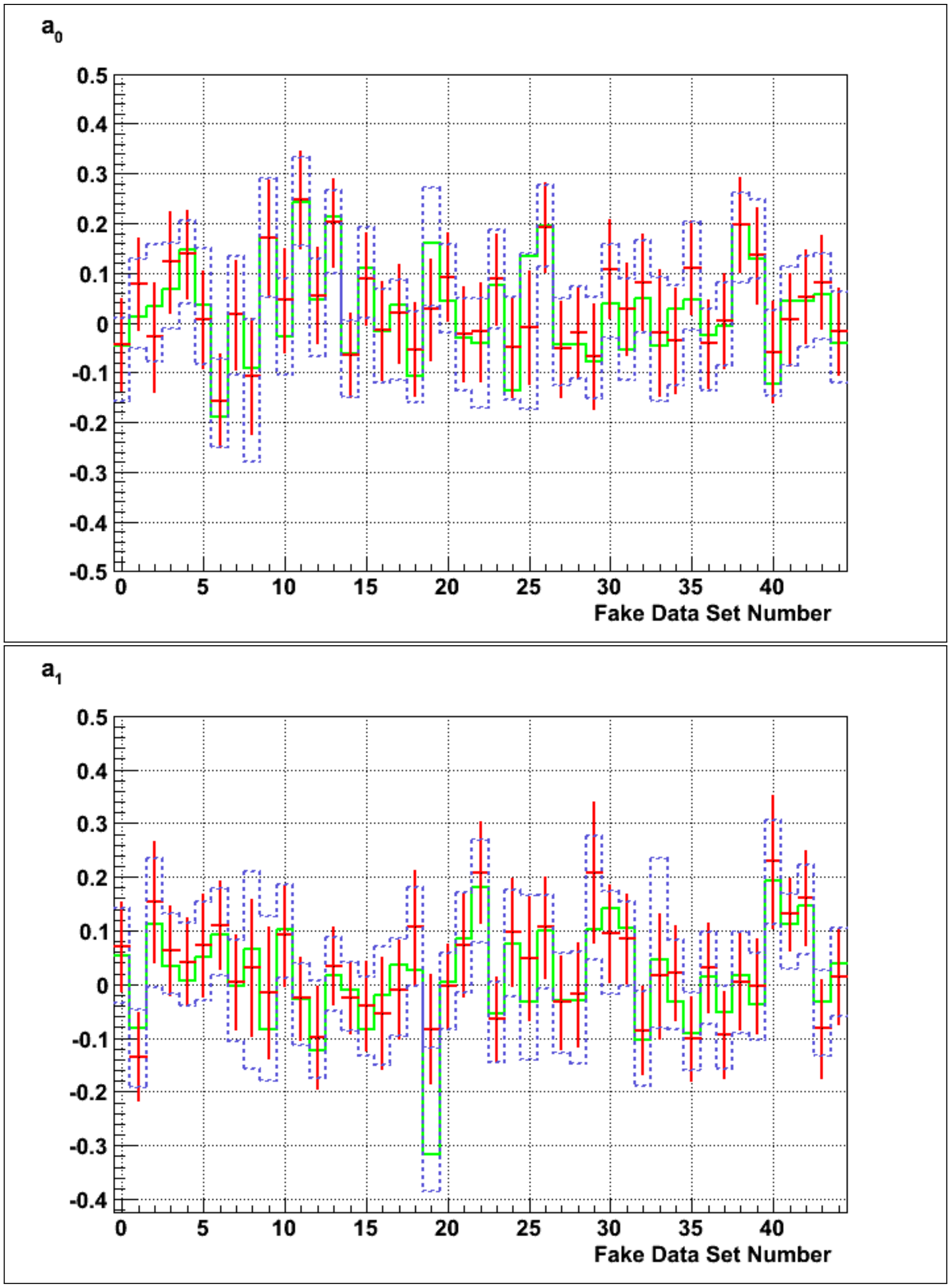
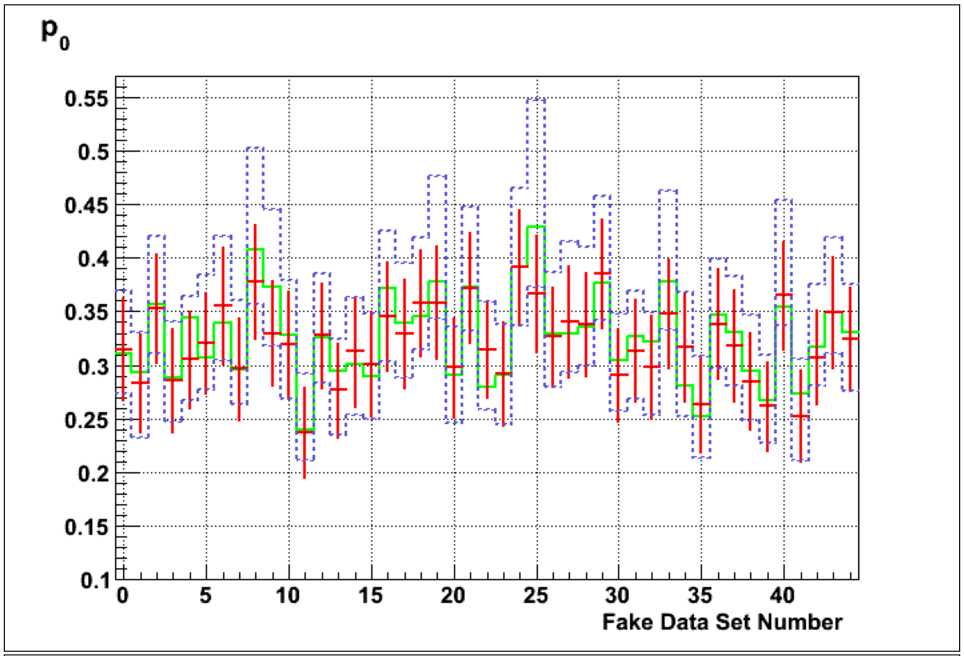


Figure B.56: Showing best-fit result in green color for day-night asymmetries (a_0 and a_1) for each of the 45 fitted simulated datasets shown in the X axis. For this fit systematics were floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red. Tables B.1 (for top plot) and B.2 (for bottom plot) show the data, in tabular form, used to make these plots.



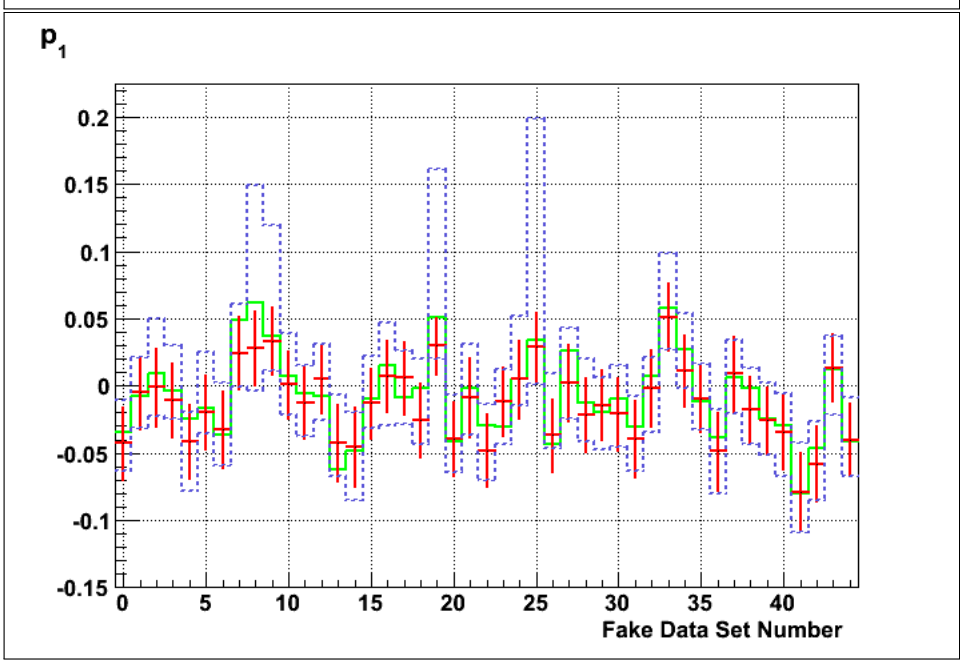


Figure B.57: Showing best-fit MCMC result in green for P_{ee} parameters (p_0 and p_1) for each of the 45 fitted simulated datasets shown in the X axis. For this fit systematics were floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red. Tables B.3 (for top plot) and B.4 (for bottom plot) show the data, in tabular form, used to make these plots.

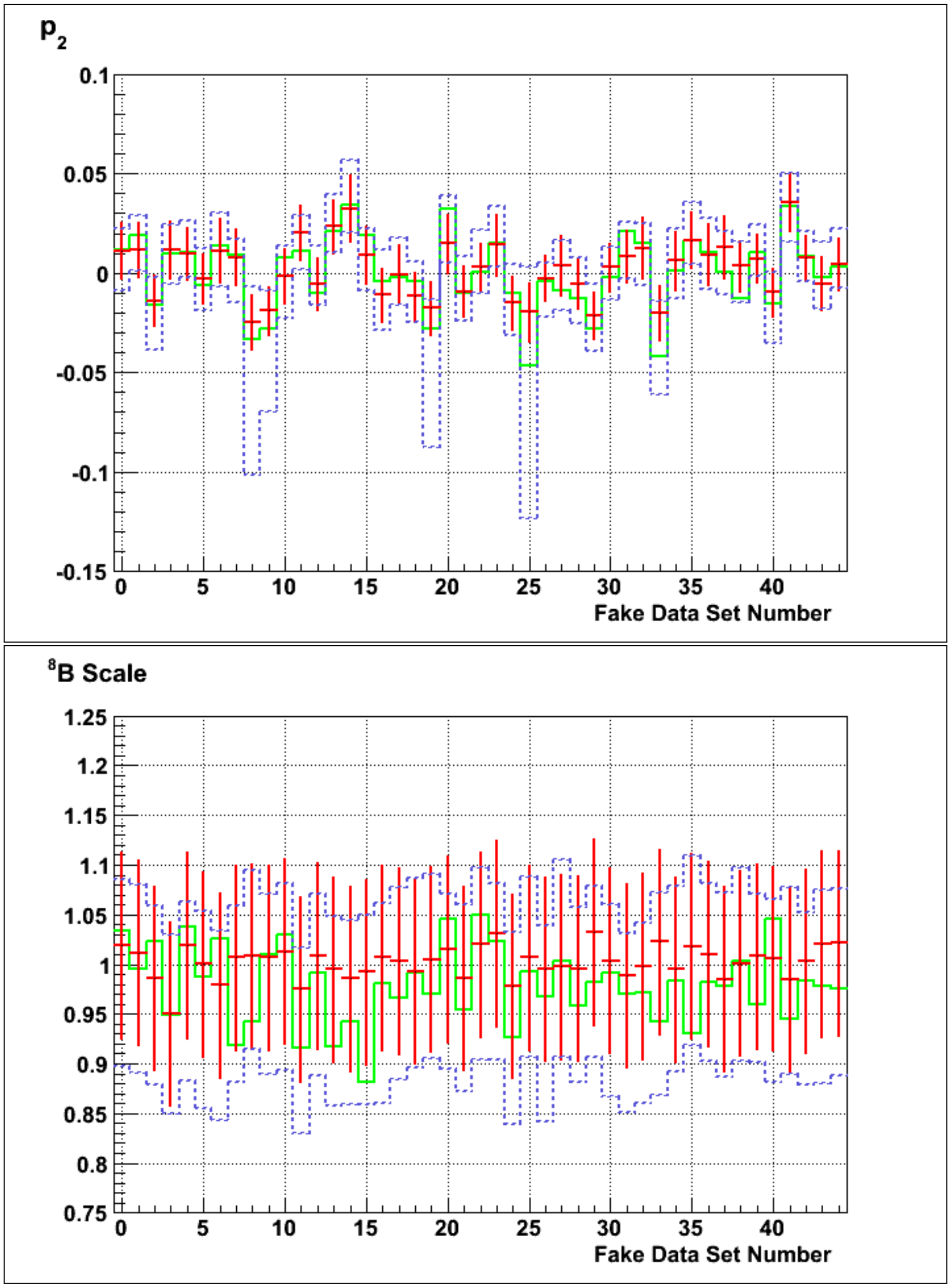


Figure B.58: Showing best-fit MCMC result (green color) of the ^{8}B Scale and P_{ee} parameter (p_2) for each of the 45 fitted simulated datasets shown in the X axis. For this fit systematics were floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red. Tables B.5 (for top plot) and B.6 (for bottom plot) show the data, in tabular form, used to make these plots.

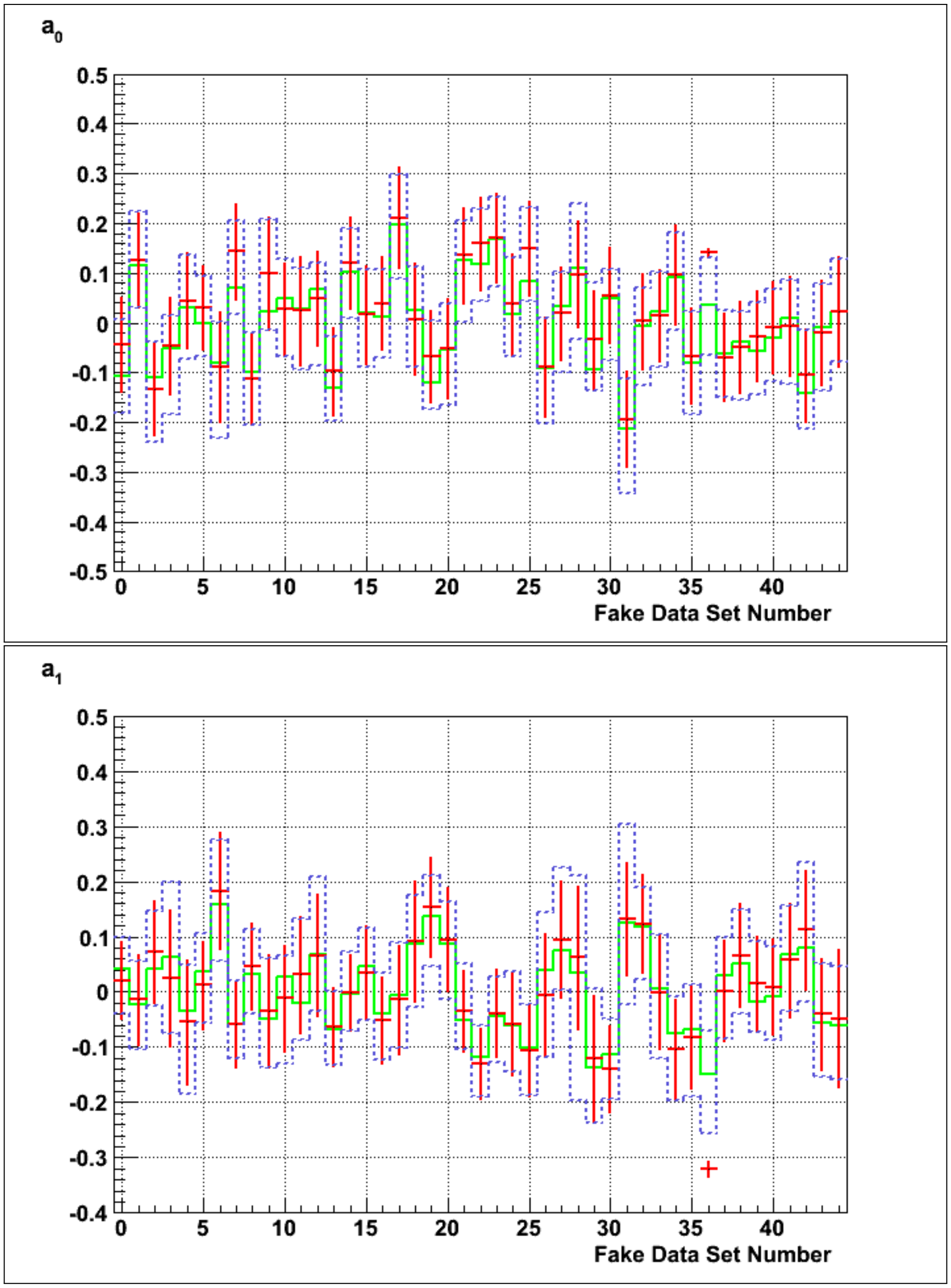


Figure B.59: Showing best-fit result in green color for day-night asymmetries (a_0 and a_1) for each of the 45 fitted alternate simulated datasets shown in the X axis. For this fit systematics were floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red. Tables B.7 (for top plot) and B.8 (for bottom plot) show the data, in tabular form, used to make these plots.

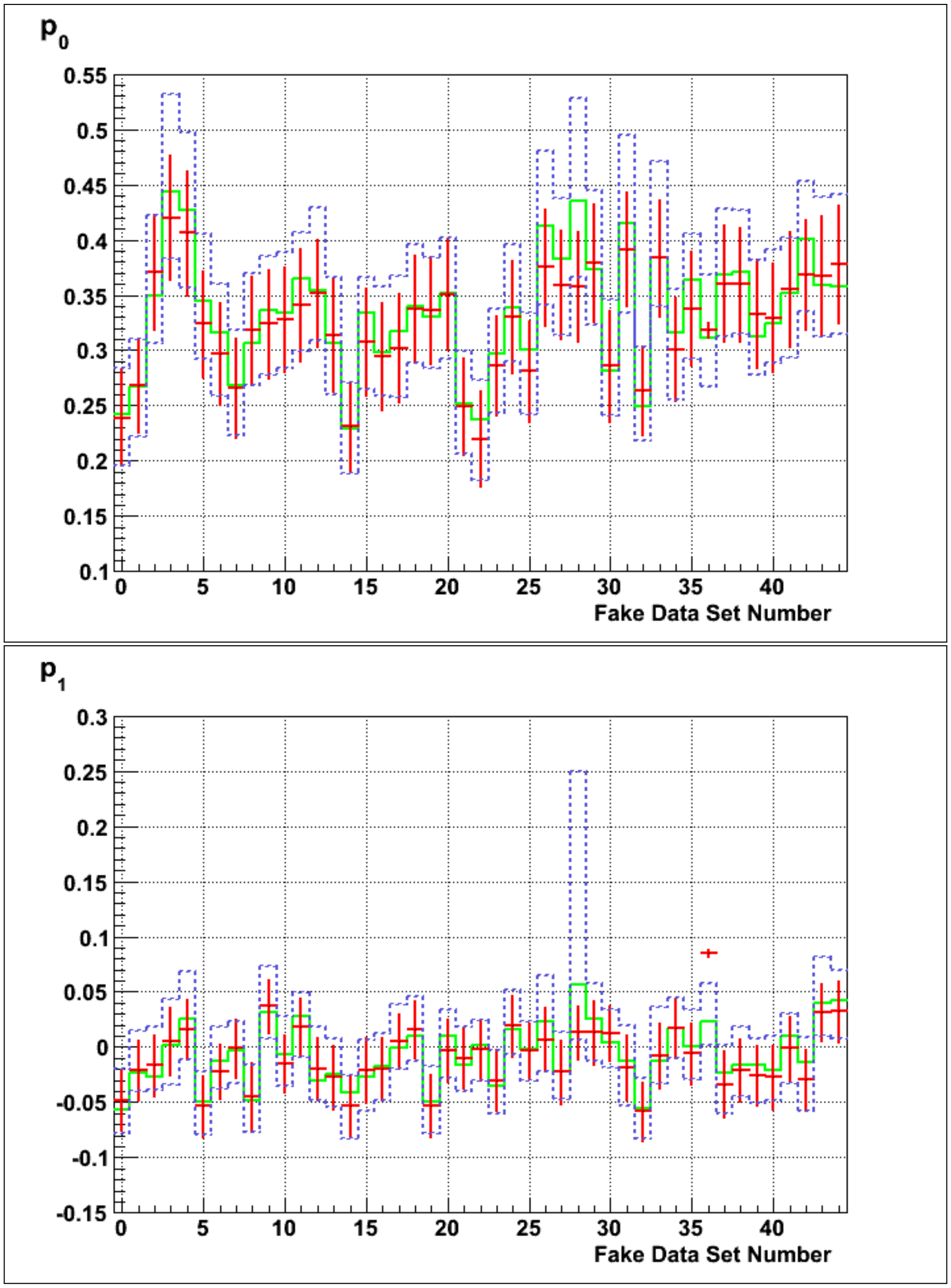


Figure B.60: Showing best-fit MCMC result in green for P_{ee} parameters (p_0 and p_1) for each of the 45 fitted alternate simulated datasets shown in the X axis. For this fit systematics were floated. The blue dotted lines show $\pm\sigma$ confidence intervals from the MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red. Tables B.9 (for top plot) and B.10 (for bottom plot) show the data, in tabular form, used to make these plots.

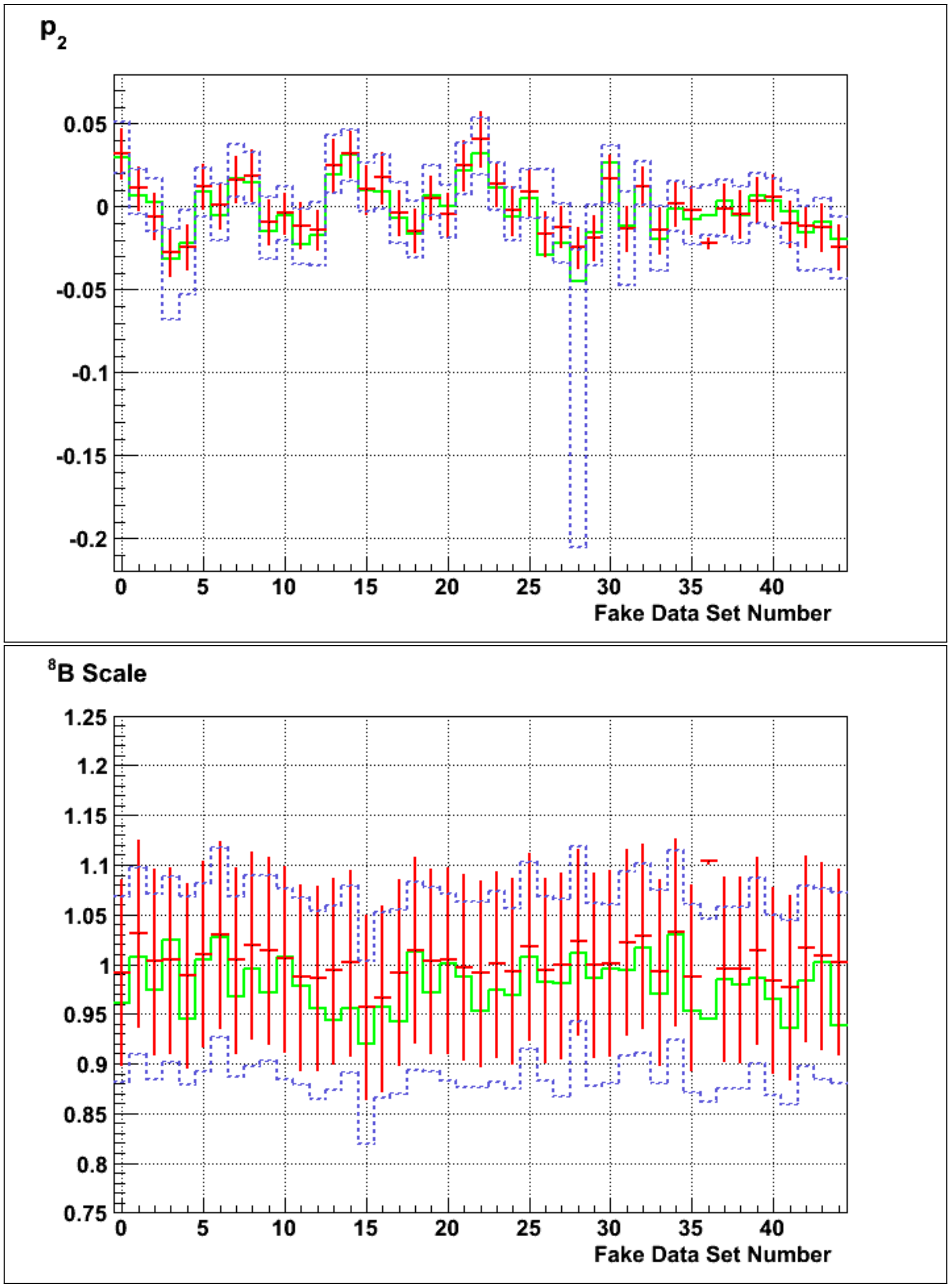


Figure B.61: Showing best-fit MCMC result (green color) of the ^{8}B Scale and P_{ee} parameter (p_2) for each of the 45 fitted alternate simulated datasets shown in the X axis. For this fit systematics were floated. The blue dotted lines show $\pm\sigma$ confidence intervals from MCMC fit. The best-fit from QSigEx along with its uncertainty is shown in red. Tables B.11 (for top plot) and B.12 (for bottom plot) show the data, in tabular form, used to make these plots.

Tables B.7 to B.12 are for the alternate simulated datasets.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	-0.044$\pm$0.092	-0.074$\pm$0.083	-0.070$\pm$0.085
1	0.079$\pm$0.093	0.038$\pm$0.089	0.049$\pm$0.090
2	-0.029$\pm$0.111	0.039$\pm$0.118	-0.006$\pm$0.119
3	0.123$\pm$0.102	0.075$\pm$0.086	0.090$\pm$0.086
4	0.138$\pm$0.088	0.123$\pm$0.084	0.116$\pm$0.086
5	0.007$\pm$0.098	0.033$\pm$0.116	-0.007$\pm$0.104
6	-0.157$\pm$0.094	-0.161$\pm$0.089	-0.160$\pm$0.094
7	0.016$\pm$0.110	0.014$\pm$0.119	0.019$\pm$0.117
8	-0.108$\pm$0.116	-0.137$\pm$0.143	-0.117$\pm$0.145
9	0.170$\pm$0.118	0.171$\pm$0.120	0.162$\pm$0.119
10	0.045$\pm$0.105	-0.008$\pm$0.098	-0.003$\pm$0.099
11	0.247$\pm$0.099	0.245$\pm$0.089	0.259$\pm$0.086
12	0.055$\pm$0.096	0.030$\pm$0.098	0.042$\pm$0.097
13	0.201$\pm$0.088	0.184$\pm$0.083	0.181$\pm$0.085
14	-0.064$\pm$0.085	-0.073$\pm$0.077	-0.075$\pm$0.080
15	0.088$\pm$0.093	0.101$\pm$0.090	0.100$\pm$0.099
16	-0.016$\pm$0.100	-0.027$\pm$0.094	-0.014$\pm$0.093
17	0.019$\pm$0.099	-0.014$\pm$0.100	-0.022$\pm$0.105
18	-0.053$\pm$0.095	-0.069$\pm$0.091	-0.061$\pm$0.088
19	0.027$\pm$0.101	0.151$\pm$0.119	0.117$\pm$0.115
20	0.092$\pm$0.089	0.069$\pm$0.088	0.085$\pm$0.087
21	-0.024$\pm$0.095	-0.044$\pm$0.093	-0.039$\pm$0.096
22	-0.018$\pm$0.099	-0.062$\pm$0.110	-0.021$\pm$0.101
23	0.088$\pm$0.092	0.087$\pm$0.099	0.074$\pm$0.094
24	-0.050$\pm$0.100	-0.052$\pm$0.103	-0.039$\pm$0.110
25	-0.010$\pm$0.115	-0.017$\pm$0.157	-0.051$\pm$0.151
26	0.192$\pm$0.090	0.195$\pm$0.081	0.193$\pm$0.082
27	-0.052$\pm$0.097	-0.038$\pm$0.088	-0.054$\pm$0.090
28	-0.019$\pm$0.092	-0.019$\pm$0.093	-0.028$\pm$0.094
29	-0.068$\pm$0.106	-0.049$\pm$0.102	-0.058$\pm$0.100
30	0.108$\pm$0.100	0.063$\pm$0.094	0.086$\pm$0.092
31	0.027$\pm$0.093	-0.013$\pm$0.102	0.023$\pm$0.098
32	0.082$\pm$0.096	0.073$\pm$0.093	0.072$\pm$0.093
33	-0.020$\pm$0.128	-0.032$\pm$0.124	-0.034$\pm$0.128
34	-0.036$\pm$0.105	-0.025$\pm$0.102	-0.038$\pm$0.104
35	0.110$\pm$0.094	0.094$\pm$0.109	0.091$\pm$0.108
36	-0.042$\pm$0.088	-0.054$\pm$0.081	-0.054$\pm$0.083
37	0.004$\pm$0.096	-0.003$\pm$0.084	-0.003$\pm$0.086
38	0.197$\pm$0.095	0.170$\pm$0.090	0.171$\pm$0.094
39	0.136$\pm$0.096	0.169$\pm$0.080	0.155$\pm$0.086
40	-0.059$\pm$0.103	-0.060$\pm$0.085	-0.066$\pm$0.086
41	0.007$\pm$0.093	0.014$\pm$0.099	0.016$\pm$0.100
42	0.052$\pm$0.094	0.043$\pm$0.092	0.040$\pm$0.091
43	0.082$\pm$0.094	0.053$\pm$0.085	0.050$\pm$0.091
44	-0.017$\pm$0.089	-0.029$\pm$0.091	-0.031$\pm$0.091

Table B.1: Comparison of the best-fit of day-night asymmetry a_0 in regular datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	0.070$\pm$0.083	0.053$\pm$0.089	0.066$\pm$0.089
1	-0.134$\pm$0.081	-0.120$\pm$0.072	-0.119$\pm$0.073
2	0.153$\pm$0.113	0.115$\pm$0.120	0.128$\pm$0.122
3	0.063$\pm$0.083	0.057$\pm$0.074	0.060$\pm$0.075
4	0.042$\pm$0.082	0.038$\pm$0.078	0.036$\pm$0.079
5	0.073$\pm$0.096	0.062$\pm$0.092	0.073$\pm$0.089
6	0.111$\pm$0.083	0.097$\pm$0.080	0.098$\pm$0.081
7	0.004$\pm$0.088	-0.011$\pm$0.095	-0.013$\pm$0.097
8	0.031$\pm$0.128	0.027$\pm$0.184	0.040$\pm$0.171
9	-0.014$\pm$0.122	-0.026$\pm$0.153	-0.020$\pm$0.145
10	0.092$\pm$0.095	0.099$\pm$0.087	0.098$\pm$0.091
11	-0.026$\pm$0.077	-0.037$\pm$0.076	-0.037$\pm$0.080
12	-0.098$\pm$0.096	-0.084$\pm$0.090	-0.085$\pm$0.091
13	0.035$\pm$0.072	0.020$\pm$0.069	0.018$\pm$0.070
14	-0.025$\pm$0.067	-0.022$\pm$0.064	-0.022$\pm$0.064
15	-0.039$\pm$0.083	-0.059$\pm$0.075	-0.049$\pm$0.075
16	-0.053$\pm$0.104	-0.039$\pm$0.111	-0.075$\pm$0.117
17	-0.009$\pm$0.091	-0.005$\pm$0.091	0.022$\pm$0.087
18	0.107$\pm$0.106	0.089$\pm$0.092	0.089$\pm$0.092
19	-0.083$\pm$0.102	-0.251$\pm$0.133	-0.194$\pm$0.137
20	-0.002$\pm$0.078	-0.012$\pm$0.071	-0.010$\pm$0.074
21	0.074$\pm$0.096	0.079$\pm$0.094	0.076$\pm$0.098
22	0.209$\pm$0.095	0.174$\pm$0.095	0.196$\pm$0.098
23	-0.064$\pm$0.078	-0.070$\pm$0.075	-0.054$\pm$0.075
24	0.097$\pm$0.101	0.077$\pm$0.099	0.054$\pm$0.105
25	0.049$\pm$0.115	0.013$\pm$0.153	0.056$\pm$0.158
26	0.106$\pm$0.094	0.079$\pm$0.086	0.084$\pm$0.092
27	-0.033$\pm$0.087	-0.035$\pm$0.093	-0.030$\pm$0.099
28	-0.018$\pm$0.097	-0.043$\pm$0.104	-0.044$\pm$0.102
29	0.209$\pm$0.131	0.162$\pm$0.115	0.198$\pm$0.126
30	0.095$\pm$0.091	0.078$\pm$0.096	0.089$\pm$0.099
31	0.085$\pm$0.084	0.075$\pm$0.079	0.076$\pm$0.080
32	-0.085$\pm$0.082	-0.101$\pm$0.088	-0.083$\pm$0.085
33	0.016$\pm$0.117	0.077$\pm$0.157	0.046$\pm$0.145
34	0.022$\pm$0.088	-0.001$\pm$0.083	0.006$\pm$0.088
35	-0.101$\pm$0.079	-0.088$\pm$0.073	-0.087$\pm$0.074
36	0.031$\pm$0.084	0.013$\pm$0.086	0.013$\pm$0.086
37	-0.094$\pm$0.081	-0.080$\pm$0.078	-0.088$\pm$0.082
38	0.005$\pm$0.089	0.003$\pm$0.094	-0.002$\pm$0.094
39	-0.003$\pm$0.088	-0.022$\pm$0.081	-0.023$\pm$0.081
40	0.229$\pm$0.124	0.209$\pm$0.097	0.208$\pm$0.104
41	0.131$\pm$0.068	0.098$\pm$0.071	0.115$\pm$0.071
42	0.161$\pm$0.089	0.140$\pm$0.083	0.140$\pm$0.084
43	-0.082$\pm$0.092	-0.053$\pm$0.079	-0.060$\pm$0.085
44	0.014$\pm$0.087	0.023$\pm$0.082	0.022$\pm$0.085

Table B.2: Comparison of the best-fit of day-night asymmetry a_1 in regular datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	0.315±0.048	0.321±0.048	0.327±0.050
1	0.284±0.046	0.282±0.049	0.296±0.050
2	0.353±0.051	0.365±0.055	0.371±0.056
3	0.286±0.048	0.294±0.046	0.301±0.048
4	0.306±0.045	0.316±0.048	0.319±0.050
5	0.321±0.047	0.331±0.053	0.340±0.056
6	0.355±0.055	0.363±0.058	0.369±0.059
7	0.297±0.048	0.312±0.049	0.319±0.053
8	0.377±0.053	0.429±0.073	0.440±0.075
9	0.329±0.049	0.382±0.063	0.382±0.064
10	0.320±0.050	0.324±0.055	0.338±0.055
11	0.237±0.042	0.252±0.041	0.250±0.040
12	0.328±0.049	0.335±0.051	0.335±0.052
13	0.277±0.045	0.280±0.045	0.288±0.047
14	0.313±0.052	0.308±0.054	0.319±0.056
15	0.301±0.049	0.299±0.050	0.305±0.052
16	0.346±0.051	0.364±0.061	0.367±0.065
17	0.329±0.051	0.342±0.054	0.342±0.056
18	0.358±0.050	0.366±0.052	0.371±0.055
19	0.358±0.053	0.410±0.066	0.410±0.072
20	0.298±0.047	0.290±0.045	0.295±0.046
21	0.372±0.052	0.390±0.058	0.396±0.062
22	0.314±0.045	0.309±0.050	0.323±0.053
23	0.292±0.049	0.292±0.047	0.303±0.049
24	0.391±0.054	0.401±0.064	0.409±0.064
25	0.367±0.054	0.460±0.087	0.446±0.083
26	0.327±0.046	0.333±0.054	0.348±0.054
27	0.341±0.053	0.354±0.061	0.361±0.060
28	0.338±0.049	0.355±0.056	0.369±0.059
29	0.385±0.051	0.400±0.058	0.397±0.058
30	0.291±0.044	0.303±0.046	0.312±0.047
31	0.314±0.048	0.319±0.050	0.321±0.050
32	0.298±0.049	0.302±0.048	0.310±0.048
33	0.348±0.051	0.398±0.065	0.401±0.067
34	0.317±0.051	0.310±0.058	0.327±0.058
35	0.264±0.045	0.261±0.048	0.267±0.049
36	0.339±0.051	0.348±0.051	0.346±0.055
37	0.319±0.052	0.332±0.051	0.342±0.057
38	0.285±0.045	0.298±0.049	0.305±0.051
39	0.262±0.042	0.268±0.041	0.262±0.040
40	0.365±0.050	0.395±0.059	0.385±0.058
41	0.252±0.043	0.259±0.048	0.266±0.049
42	0.307±0.044	0.328±0.047	0.322±0.047
43	0.349±0.052	0.365±0.055	0.368±0.059
44	0.325±0.048	0.325±0.050	0.335±0.051

Table B.3: Comparison of the best-fit of P_{ee} p_0 in regular datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	-0.043$\pm$0.028	-0.037$\pm$0.026	-0.037$\pm$0.028
1	-0.005$\pm$0.027	-0.005$\pm$0.026	-0.007$\pm$0.027
2	-0.001$\pm$0.029	0.014$\pm$0.036	0.013$\pm$0.039
3	-0.011$\pm$0.028	0.003$\pm$0.027	-0.009$\pm$0.028
4	-0.041$\pm$0.028	-0.049$\pm$0.029	-0.044$\pm$0.029
5	-0.020$\pm$0.028	-0.005$\pm$0.030	-0.013$\pm$0.032
6	-0.032$\pm$0.029	-0.029$\pm$0.031	-0.031$\pm$0.033
7	0.025$\pm$0.027	0.030$\pm$0.031	0.036$\pm$0.037
8	0.028$\pm$0.028	0.073$\pm$0.076	0.119$\pm$0.113
9	0.033$\pm$0.026	0.066$\pm$0.054	0.081$\pm$0.064
10	0.001$\pm$0.025	0.008$\pm$0.030	0.012$\pm$0.034
11	-0.013$\pm$0.027	-0.011$\pm$0.027	-0.010$\pm$0.026
12	0.005$\pm$0.026	0.003$\pm$0.028	0.003$\pm$0.028
13	-0.043$\pm$0.029	-0.037$\pm$0.030	-0.042$\pm$0.030
14	-0.045$\pm$0.030	-0.052$\pm$0.033	-0.053$\pm$0.034
15	-0.013$\pm$0.026	-0.004$\pm$0.027	-0.008$\pm$0.028
16	0.007$\pm$0.027	0.009$\pm$0.038	0.024$\pm$0.052
17	0.006$\pm$0.027	-0.001$\pm$0.027	-0.002$\pm$0.029
18	-0.025$\pm$0.028	-0.012$\pm$0.032	-0.020$\pm$0.033
19	0.030$\pm$0.021	0.091$\pm$0.071	0.120$\pm$0.087
20	-0.039$\pm$0.028	-0.036$\pm$0.029	-0.042$\pm$0.030
21	-0.009$\pm$0.030	-0.003$\pm$0.034	-0.002$\pm$0.036
22	-0.048$\pm$0.027	-0.042$\pm$0.028	-0.045$\pm$0.029
23	-0.012$\pm$0.026	-0.015$\pm$0.028	-0.019$\pm$0.028
24	0.005$\pm$0.029	0.018$\pm$0.033	0.020$\pm$0.037
25	0.029$\pm$0.026	0.100$\pm$0.099	0.141$\pm$0.113
26	-0.037$\pm$0.027	-0.019$\pm$0.028	-0.022$\pm$0.029
27	0.002$\pm$0.029	0.009$\pm$0.034	0.010$\pm$0.034
28	-0.022$\pm$0.028	-0.011$\pm$0.030	-0.011$\pm$0.032
29	-0.014$\pm$0.027	-0.021$\pm$0.027	-0.018$\pm$0.029
30	-0.021$\pm$0.027	-0.015$\pm$0.030	-0.015$\pm$0.030
31	-0.039$\pm$0.029	-0.035$\pm$0.028	-0.038$\pm$0.029
32	-0.002$\pm$0.029	-0.006$\pm$0.028	-0.007$\pm$0.029
33	0.051$\pm$0.026	0.063$\pm$0.036	0.071$\pm$0.046
34	0.011$\pm$0.027	0.026$\pm$0.028	0.025$\pm$0.029
35	-0.009$\pm$0.025	-0.008$\pm$0.024	-0.009$\pm$0.025
36	-0.049$\pm$0.029	-0.049$\pm$0.031	-0.051$\pm$0.032
37	0.009$\pm$0.028	0.007$\pm$0.027	0.009$\pm$0.028
38	-0.018$\pm$0.026	-0.015$\pm$0.028	-0.011$\pm$0.028
39	-0.025$\pm$0.025	-0.025$\pm$0.027	-0.023$\pm$0.027
40	-0.035$\pm$0.028	-0.036$\pm$0.031	-0.033$\pm$0.033
41	-0.079$\pm$0.029	-0.076$\pm$0.034	-0.072$\pm$0.034
42	-0.058$\pm$0.028	-0.056$\pm$0.029	-0.057$\pm$0.030
43	0.013$\pm$0.025	0.008$\pm$0.029	0.011$\pm$0.030
44	-0.040$\pm$0.028	-0.038$\pm$0.029	-0.039$\pm$0.030

Table B.4: Comparison of the best-fit of P_{ee} p_1 in regular datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	0.011$\pm$0.014	0.007$\pm$0.016	0.010$\pm$0.016
1	0.012$\pm$0.014	0.015$\pm$0.014	0.017$\pm$0.015
2	-0.014$\pm$0.013	-0.020$\pm$0.018	-0.025$\pm$0.021
3	0.012$\pm$0.015	0.009$\pm$0.015	0.010$\pm$0.015
4	0.010$\pm$0.013	0.012$\pm$0.014	0.013$\pm$0.015
5	-0.003$\pm$0.012	-0.003$\pm$0.015	-0.005$\pm$0.017
6	0.011$\pm$0.016	0.012$\pm$0.019	0.013$\pm$0.019
7	0.008$\pm$0.014	0.001$\pm$0.016	0.001$\pm$0.019
8	-0.025$\pm$0.014	-0.054$\pm$0.047	-0.087$\pm$0.071
9	-0.019$\pm$0.012	-0.039$\pm$0.030	-0.054$\pm$0.041
10	-0.002$\pm$0.014	-0.004$\pm$0.018	-0.003$\pm$0.020
11	0.020$\pm$0.014	0.015$\pm$0.013	0.015$\pm$0.014
12	-0.005$\pm$0.013	-0.001$\pm$0.015	-0.002$\pm$0.016
13	0.023$\pm$0.014	0.025$\pm$0.015	0.025$\pm$0.015
14	0.033$\pm$0.017	0.039$\pm$0.018	0.040$\pm$0.019
15	0.009$\pm$0.015	0.007$\pm$0.016	0.013$\pm$0.015
16	-0.011$\pm$0.013	-0.008$\pm$0.020	-0.018$\pm$0.031
17	-0.001$\pm$0.015	0.001$\pm$0.017	0.003$\pm$0.017
18	-0.012$\pm$0.012	-0.010$\pm$0.015	-0.010$\pm$0.017
19	-0.018$\pm$0.013	-0.051$\pm$0.037	-0.058$\pm$0.048
20	0.015$\pm$0.015	0.022$\pm$0.017	0.022$\pm$0.017
21	-0.009$\pm$0.013	-0.008$\pm$0.016	-0.011$\pm$0.019
22	0.003$\pm$0.012	0.006$\pm$0.016	0.003$\pm$0.017
23	0.014$\pm$0.016	0.019$\pm$0.015	0.019$\pm$0.016
24	-0.015$\pm$0.014	-0.013$\pm$0.018	-0.017$\pm$0.020
25	-0.019$\pm$0.015	-0.060$\pm$0.063	-0.094$\pm$0.073
26	-0.003$\pm$0.011	-0.008$\pm$0.013	-0.008$\pm$0.015
27	0.004$\pm$0.015	-0.001$\pm$0.018	-0.005$\pm$0.019
28	-0.005$\pm$0.013	-0.009$\pm$0.017	-0.011$\pm$0.018
29	-0.021$\pm$0.012	-0.022$\pm$0.017	-0.027$\pm$0.020
30	0.003$\pm$0.012	-0.000$\pm$0.013	-0.000$\pm$0.014
31	0.009$\pm$0.013	0.012$\pm$0.014	0.010$\pm$0.015
32	0.013$\pm$0.015	0.009$\pm$0.016	0.009$\pm$0.016
33	-0.020$\pm$0.014	-0.038$\pm$0.024	-0.048$\pm$0.034
34	0.006$\pm$0.015	0.005$\pm$0.018	0.003$\pm$0.019
35	0.017$\pm$0.014	0.020$\pm$0.015	0.021$\pm$0.016
36	0.009$\pm$0.015	0.010$\pm$0.018	0.009$\pm$0.019
37	0.013$\pm$0.016	0.005$\pm$0.016	0.007$\pm$0.016
38	0.004$\pm$0.013	0.001$\pm$0.015	0.000$\pm$0.016
39	0.007$\pm$0.012	0.012$\pm$0.013	0.012$\pm$0.013
40	-0.010$\pm$0.012	-0.017$\pm$0.018	-0.014$\pm$0.018
41	0.036$\pm$0.014	0.033$\pm$0.017	0.035$\pm$0.018
42	0.008$\pm$0.011	0.008$\pm$0.013	0.009$\pm$0.013
43	-0.005$\pm$0.014	-0.001$\pm$0.017	-0.006$\pm$0.019
44	0.004$\pm$0.013	0.007$\pm$0.015	0.007$\pm$0.016

Table B.5: Comparison of the best-fit of P_{ee} p_2 in regular datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	1.019$\pm$0.094	0.992$\pm$0.094	1.004$\pm$0.093
1	1.012$\pm$0.094	0.986$\pm$0.095	0.984$\pm$0.094
2	0.986$\pm$0.093	0.969$\pm$0.090	0.966$\pm$0.091
3	0.951$\pm$0.093	0.939$\pm$0.090	0.934$\pm$0.090
4	1.019$\pm$0.094	0.973$\pm$0.090	0.983$\pm$0.091
5	1.000$\pm$0.093	0.955$\pm$0.099	0.983$\pm$0.095
6	0.979$\pm$0.093	0.938$\pm$0.095	0.943$\pm$0.096
7	1.007$\pm$0.094	0.970$\pm$0.089	0.980$\pm$0.089
8	1.008$\pm$0.093	1.004$\pm$0.090	1.003$\pm$0.089
9	1.007$\pm$0.094	0.980$\pm$0.090	0.985$\pm$0.093
10	1.013$\pm$0.094	0.987$\pm$0.094	0.989$\pm$0.094
11	0.975$\pm$0.093	0.924$\pm$0.093	0.951$\pm$0.088
12	1.009$\pm$0.093	0.979$\pm$0.091	0.987$\pm$0.091
13	0.995$\pm$0.093	0.953$\pm$0.095	0.965$\pm$0.094
14	0.986$\pm$0.093	0.952$\pm$0.093	0.956$\pm$0.093
15	0.992$\pm$0.093	0.955$\pm$0.095	0.966$\pm$0.093
16	1.007$\pm$0.094	0.961$\pm$0.101	0.987$\pm$0.098
17	1.003$\pm$0.094	0.981$\pm$0.097	0.980$\pm$0.097
18	0.992$\pm$0.093	0.992$\pm$0.095	0.977$\pm$0.094
19	1.005$\pm$0.093	0.998$\pm$0.093	0.996$\pm$0.097
20	1.015$\pm$0.094	0.983$\pm$0.088	0.990$\pm$0.087
21	0.986$\pm$0.093	0.966$\pm$0.094	0.958$\pm$0.095
22	1.020$\pm$0.093	1.001$\pm$0.096	1.001$\pm$0.095
23	1.031$\pm$0.094	0.992$\pm$0.089	0.988$\pm$0.091
24	0.978$\pm$0.093	0.936$\pm$0.097	0.957$\pm$0.096
25	1.007$\pm$0.094	0.997$\pm$0.091	1.003$\pm$0.091
26	0.995$\pm$0.093	0.940$\pm$0.099	0.966$\pm$0.096
27	0.997$\pm$0.094	1.006$\pm$0.099	0.985$\pm$0.096
28	0.996$\pm$0.093	0.969$\pm$0.088	0.966$\pm$0.089
29	1.032$\pm$0.094	0.993$\pm$0.086	1.012$\pm$0.085
30	1.004$\pm$0.093	0.964$\pm$0.096	0.977$\pm$0.096
31	0.989$\pm$0.093	0.941$\pm$0.090	0.960$\pm$0.092
32	0.998$\pm$0.094	0.951$\pm$0.090	0.971$\pm$0.089
33	1.023$\pm$0.094	0.970$\pm$0.102	0.989$\pm$0.097
34	0.995$\pm$0.094	0.986$\pm$0.093	0.984$\pm$0.095
35	1.018$\pm$0.094	1.014$\pm$0.096	1.005$\pm$0.095
36	1.010$\pm$0.093	0.992$\pm$0.089	0.990$\pm$0.092
37	0.985$\pm$0.093	0.980$\pm$0.093	0.963$\pm$0.095
38	1.001$\pm$0.093	1.000$\pm$0.097	0.988$\pm$0.097
39	1.008$\pm$0.093	0.986$\pm$0.085	0.983$\pm$0.087
40	1.006$\pm$0.093	0.973$\pm$0.092	0.980$\pm$0.091
41	0.985$\pm$0.093	0.983$\pm$0.094	0.961$\pm$0.094
42	1.003$\pm$0.093	0.965$\pm$0.087	0.968$\pm$0.089
43	1.020$\pm$0.094	0.978$\pm$0.097	0.994$\pm$0.096
44	1.022$\pm$0.093	0.982$\pm$0.094	1.000$\pm$0.093

Table B.6: Comparison of the best-fit of ^{8}B Scale in regular datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	-0.043$\pm$0.095	-0.087$\pm$0.094	-0.072$\pm$0.094
1	0.126$\pm$0.095	0.127$\pm$0.097	0.120$\pm$0.096
2	-0.134$\pm$0.093	-0.138$\pm$0.101	-0.146$\pm$0.099
3	-0.047$\pm$0.098	-0.084$\pm$0.099	-0.081$\pm$0.098
4	0.044$\pm$0.096	0.032$\pm$0.104	0.039$\pm$0.105
5	0.030$\pm$0.085	0.014$\pm$0.082	0.014$\pm$0.080
6	-0.088$\pm$0.111	-0.115$\pm$0.117	-0.130$\pm$0.119
7	0.143$\pm$0.096	0.112$\pm$0.094	0.116$\pm$0.094
8	-0.111$\pm$0.092	-0.112$\pm$0.094	-0.106$\pm$0.095
9	0.100$\pm$0.113	0.097$\pm$0.111	0.096$\pm$0.113
10	0.028$\pm$0.094	0.031$\pm$0.097	0.028$\pm$0.098
11	0.024$\pm$0.110	0.009$\pm$0.102	0.001$\pm$0.103
12	0.048$\pm$0.096	0.017$\pm$0.103	0.025$\pm$0.102
13	-0.098$\pm$0.088	-0.112$\pm$0.085	-0.101$\pm$0.086
14	0.120$\pm$0.092	0.099$\pm$0.090	0.107$\pm$0.090
15	0.017$\pm$0.099	0.009$\pm$0.098	0.011$\pm$0.098
16	0.040$\pm$0.093	0.018$\pm$0.089	0.022$\pm$0.091
17	0.212$\pm$0.102	0.194$\pm$0.104	0.195$\pm$0.105
18	0.007$\pm$0.112	0.012$\pm$0.101	0.008$\pm$0.103
19	-0.067$\pm$0.093	-0.086$\pm$0.089	-0.087$\pm$0.090
20	-0.052$\pm$0.100	-0.063$\pm$0.103	-0.054$\pm$0.104
21	0.136$\pm$0.097	0.103$\pm$0.101	0.123$\pm$0.101
22	0.159$\pm$0.093	0.136$\pm$0.093	0.141$\pm$0.093
23	0.170$\pm$0.090	0.163$\pm$0.089	0.162$\pm$0.089
24	0.037$\pm$0.103	0.033$\pm$0.099	0.028$\pm$0.099
25	0.150$\pm$0.095	0.137$\pm$0.094	0.131$\pm$0.096
26	-0.089$\pm$0.101	-0.097$\pm$0.106	-0.110$\pm$0.109
27	0.019$\pm$0.094	0.001$\pm$0.100	-0.007$\pm$0.103
28	0.098$\pm$0.108	0.103$\pm$0.137	0.091$\pm$0.146
29	-0.034$\pm$0.099	-0.028$\pm$0.110	-0.038$\pm$0.108
30	0.055$\pm$0.097	0.017$\pm$0.091	0.032$\pm$0.093
31	-0.194$\pm$0.097	-0.228$\pm$0.115	-0.235$\pm$0.118
32	0.003$\pm$0.097	-0.028$\pm$0.097	-0.020$\pm$0.098
33	0.015$\pm$0.092	0.007$\pm$0.096	0.001$\pm$0.095
34	0.096$\pm$0.101	0.084$\pm$0.098	0.087$\pm$0.099
35	-0.067$\pm$0.097	-0.081$\pm$0.103	-0.077$\pm$0.101
36	0.142$\pm$0.008	0.033$\pm$0.097	0.032$\pm$0.101
37	-0.070$\pm$0.089	-0.062$\pm$0.087	-0.077$\pm$0.088
38	-0.049$\pm$0.094	-0.066$\pm$0.090	-0.056$\pm$0.092
39	-0.027$\pm$0.091	-0.051$\pm$0.093	-0.040$\pm$0.094
40	-0.008$\pm$0.093	-0.025$\pm$0.092	-0.027$\pm$0.093
41	-0.007$\pm$0.102	-0.019$\pm$0.104	-0.015$\pm$0.103
42	-0.106$\pm$0.095	-0.114$\pm$0.099	-0.115$\pm$0.100
43	-0.020$\pm$0.105	-0.028$\pm$0.108	-0.021$\pm$0.111
44	0.023$\pm$0.111	0.025$\pm$0.104	0.033$\pm$0.102

Table B.7: Comparison of the best-fit of day-night asymmetry a_0 in alternative datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	0.021±0.070	0.031±0.069	0.029±0.071
1	-0.014±0.083	-0.025±0.080	-0.025±0.083
2	0.072±0.094	0.062±0.086	0.069±0.089
3	0.026±0.124	0.062±0.137	0.091±0.142
4	-0.055±0.112	-0.068±0.118	-0.059±0.120
5	0.012±0.080	0.025±0.081	0.021±0.084
6	0.183±0.105	0.166±0.109	0.199±0.123
7	-0.059±0.077	-0.049±0.071	-0.054±0.073
8	0.047±0.079	0.038±0.076	0.037±0.077
9	-0.034±0.102	-0.038±0.099	-0.047±0.102
10	-0.012±0.096	-0.033±0.098	-0.045±0.100
11	0.031±0.105	0.023±0.110	0.040±0.115
12	0.066±0.110	0.087±0.122	0.086±0.131
13	-0.063±0.072	-0.066±0.067	-0.065±0.068
14	-0.000±0.069	0.002±0.071	-0.000±0.070
15	0.034±0.085	0.032±0.083	0.031±0.083
16	-0.052±0.078	-0.044±0.078	-0.047±0.080
17	-0.014±0.098	-0.006±0.095	-0.004±0.097
18	0.092±0.109	0.073±0.101	0.071±0.104
19	0.153±0.090	0.129±0.082	0.139±0.085
20	0.095±0.094	0.074±0.088	0.083±0.093
21	-0.035±0.074	-0.026±0.077	-0.034±0.078
22	-0.131±0.064	-0.125±0.063	-0.124±0.063
23	-0.039±0.080	-0.049±0.078	-0.048±0.082
24	-0.058±0.094	-0.054±0.092	-0.065±0.093
25	-0.105±0.084	-0.104±0.083	-0.100±0.088
26	-0.005±0.111	0.013±0.130	0.007±0.130
27	0.095±0.106	0.114±0.111	0.116±0.116
28	0.063±0.130	0.007±0.203	0.053±0.202
29	-0.120±0.115	-0.116±0.121	-0.111±0.117
30	-0.140±0.080	-0.121±0.073	-0.130±0.079
31	0.132±0.102	0.141±0.163	0.220±0.194
32	0.124±0.089	0.106±0.083	0.112±0.085
33	-0.000±0.104	-0.008±0.113	-0.001±0.117
34	-0.104±0.090	-0.103±0.094	-0.100±0.095
35	-0.082±0.094	-0.088±0.101	-0.085±0.106
36	-0.321±0.014	-0.163±0.093	-0.161±0.093
37	0.002±0.091	0.007±0.092	0.007±0.093
38	0.067±0.095	0.055±0.094	0.057±0.096
39	0.015±0.087	0.010±0.082	0.010±0.085
40	0.009±0.087	-0.002±0.084	0.002±0.086
41	0.057±0.103	0.061±0.095	0.058±0.095
42	0.112±0.109	0.109±0.127	0.122±0.127
43	-0.040±0.102	-0.052±0.103	-0.053±0.107
44	-0.048±0.126	-0.056±0.103	-0.054±0.105

Table B.8: Comparison of the best-fit of day-night asymmetry a_1 in alternative datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	0.239±0.043	0.240±0.044	0.245±0.046
1	0.268±0.042	0.266±0.044	0.277±0.045
2	0.371±0.052	0.364±0.058	0.380±0.059
3	0.420±0.057	0.457±0.075	0.473±0.081
4	0.406±0.057	0.427±0.070	0.435±0.073
5	0.324±0.048	0.349±0.056	0.350±0.056
6	0.297±0.046	0.310±0.051	0.313±0.052
7	0.266±0.045	0.271±0.048	0.279±0.049
8	0.319±0.049	0.319±0.051	0.324±0.051
9	0.324±0.050	0.332±0.054	0.343±0.057
10	0.328±0.048	0.336±0.053	0.346±0.056
11	0.341±0.051	0.353±0.053	0.364±0.056
12	0.352±0.049	0.369±0.060	0.385±0.066
13	0.313±0.051	0.313±0.053	0.322±0.055
14	0.231±0.041	0.230±0.041	0.235±0.042
15	0.307±0.049	0.316±0.051	0.321±0.052
16	0.295±0.049	0.308±0.049	0.313±0.051
17	0.302±0.050	0.313±0.055	0.323±0.059
18	0.337±0.049	0.342±0.054	0.353±0.057
19	0.337±0.049	0.334±0.050	0.344±0.051
20	0.351±0.050	0.347±0.055	0.357±0.056
21	0.249±0.044	0.253±0.047	0.259±0.048
22	0.220±0.044	0.228±0.046	0.234±0.047
23	0.286±0.045	0.290±0.047	0.294±0.049
24	0.331±0.051	0.343±0.053	0.348±0.055
25	0.281±0.046	0.288±0.046	0.293±0.048
26	0.375±0.053	0.411±0.070	0.430±0.078
27	0.360±0.050	0.375±0.062	0.390±0.063
28	0.357±0.050	0.447±0.081	0.459±0.082
29	0.379±0.053	0.384±0.061	0.395±0.062
30	0.286±0.050	0.293±0.053	0.299±0.055
31	0.392±0.052	0.415±0.080	0.445±0.097
32	0.264±0.041	0.260±0.042	0.266±0.043
33	0.383±0.054	0.405±0.065	0.416±0.069
34	0.301±0.047	0.306±0.050	0.314±0.052
35	0.338±0.052	0.349±0.056	0.360±0.059
36	0.318±0.007	0.318±0.050	0.327±0.054
37	0.361±0.053	0.371±0.058	0.381±0.060
38	0.360±0.052	0.371±0.056	0.380±0.058
39	0.333±0.049	0.330±0.052	0.338±0.053
40	0.330±0.050	0.340±0.051	0.347±0.053
41	0.355±0.052	0.347±0.054	0.360±0.058
42	0.368±0.050	0.394±0.059	0.406±0.062
43	0.368±0.054	0.376±0.064	0.391±0.064
44	0.378±0.053	0.378±0.063	0.389±0.067

Table B.9: Comparison of the best-fit of P_{ee} p_0 in alternative datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	-0.048$\pm$0.028	-0.049$\pm$0.029	-0.053$\pm$0.030
1	-0.021$\pm$0.028	-0.013$\pm$0.028	-0.014$\pm$0.028
2	-0.017$\pm$0.028	-0.010$\pm$0.029	-0.013$\pm$0.031
3	0.005$\pm$0.031	0.004$\pm$0.039	0.012$\pm$0.045
4	0.016$\pm$0.027	0.028$\pm$0.040	0.042$\pm$0.050
5	-0.053$\pm$0.028	-0.051$\pm$0.028	-0.052$\pm$0.029
6	-0.022$\pm$0.025	-0.010$\pm$0.028	-0.009$\pm$0.030
7	-0.001$\pm$0.027	-0.005$\pm$0.028	-0.007$\pm$0.028
8	-0.044$\pm$0.031	-0.046$\pm$0.030	-0.049$\pm$0.032
9	0.037$\pm$0.025	0.040$\pm$0.032	0.048$\pm$0.039
10	-0.015$\pm$0.027	-0.004$\pm$0.032	-0.003$\pm$0.035
11	0.018$\pm$0.027	0.020$\pm$0.029	0.024$\pm$0.031
12	-0.019$\pm$0.029	-0.015$\pm$0.033	-0.020$\pm$0.035
13	-0.027$\pm$0.030	-0.023$\pm$0.030	-0.027$\pm$0.030
14	-0.053$\pm$0.028	-0.054$\pm$0.029	-0.055$\pm$0.030
15	-0.021$\pm$0.030	-0.025$\pm$0.032	-0.029$\pm$0.032
16	-0.019$\pm$0.028	-0.018$\pm$0.031	-0.019$\pm$0.031
17	0.005$\pm$0.025	0.009$\pm$0.029	0.011$\pm$0.036
18	0.016$\pm$0.026	0.017$\pm$0.029	0.017$\pm$0.033
19	-0.053$\pm$0.029	-0.048$\pm$0.030	-0.052$\pm$0.031
20	-0.003$\pm$0.028	0.003$\pm$0.031	0.002$\pm$0.032
21	-0.010$\pm$0.028	-0.010$\pm$0.029	-0.012$\pm$0.029
22	-0.002$\pm$0.028	-0.003$\pm$0.028	-0.004$\pm$0.028
23	-0.030$\pm$0.028	-0.031$\pm$0.029	-0.032$\pm$0.030
24	0.019$\pm$0.028	0.023$\pm$0.029	0.022$\pm$0.031
25	-0.004$\pm$0.026	-0.003$\pm$0.027	-0.006$\pm$0.028
26	0.007$\pm$0.029	0.021$\pm$0.043	0.047$\pm$0.095
27	-0.023$\pm$0.029	-0.017$\pm$0.030	-0.015$\pm$0.033
28	0.013$\pm$0.025	0.128$\pm$0.122	0.189$\pm$0.151
29	0.013$\pm$0.029	0.022$\pm$0.035	0.021$\pm$0.035
30	0.013$\pm$0.025	0.008$\pm$0.027	0.011$\pm$0.032
31	-0.018$\pm$0.030	-0.017$\pm$0.036	-0.008$\pm$0.047
32	-0.058$\pm$0.026	-0.056$\pm$0.028	-0.056$\pm$0.028
33	-0.007$\pm$0.030	0.002$\pm$0.034	0.005$\pm$0.036
34	0.017$\pm$0.026	0.017$\pm$0.027	0.018$\pm$0.028
35	-0.006$\pm$0.028	0.002$\pm$0.031	-0.000$\pm$0.033
36	0.085$\pm$0.004	0.030$\pm$0.028	0.030$\pm$0.029
37	-0.034$\pm$0.030	-0.029$\pm$0.031	-0.030$\pm$0.032
38	-0.021$\pm$0.029	-0.013$\pm$0.031	-0.014$\pm$0.032
39	-0.026$\pm$0.028	-0.021$\pm$0.029	-0.024$\pm$0.030
40	-0.027$\pm$0.029	-0.019$\pm$0.029	-0.023$\pm$0.029
41	-0.001$\pm$0.029	-0.001$\pm$0.031	-0.002$\pm$0.032
42	-0.030$\pm$0.028	-0.024$\pm$0.033	-0.018$\pm$0.036
43	0.031$\pm$0.026	0.046$\pm$0.036	0.050$\pm$0.040
44	0.032$\pm$0.028	0.038$\pm$0.031	0.045$\pm$0.040

Table B.10: Comparison of the best-fit of P_{ee} p_1 in alternative datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	0.032$\pm$0.015	0.035$\pm$0.016	0.037$\pm$0.016
1	0.011$\pm$0.013	0.009$\pm$0.013	0.010$\pm$0.014
2	-0.006$\pm$0.014	0.001$\pm$0.016	-0.000$\pm$0.017
3	-0.027$\pm$0.014	-0.041$\pm$0.027	-0.055$\pm$0.043
4	-0.024$\pm$0.014	-0.028$\pm$0.025	-0.040$\pm$0.036
5	0.012$\pm$0.014	0.009$\pm$0.015	0.009$\pm$0.016
6	0.001$\pm$0.014	-0.003$\pm$0.017	-0.006$\pm$0.020
7	0.017$\pm$0.014	0.022$\pm$0.016	0.023$\pm$0.016
8	0.019$\pm$0.015	0.017$\pm$0.016	0.019$\pm$0.017
9	-0.009$\pm$0.013	-0.013$\pm$0.018	-0.017$\pm$0.022
10	-0.004$\pm$0.013	-0.004$\pm$0.016	-0.006$\pm$0.020
11	-0.011$\pm$0.014	-0.018$\pm$0.017	-0.021$\pm$0.019
12	-0.014$\pm$0.012	-0.016$\pm$0.019	-0.022$\pm$0.025
13	0.025$\pm$0.016	0.026$\pm$0.018	0.027$\pm$0.018
14	0.032$\pm$0.014	0.031$\pm$0.015	0.035$\pm$0.016
15	0.010$\pm$0.015	0.012$\pm$0.015	0.012$\pm$0.015
16	0.018$\pm$0.015	0.015$\pm$0.016	0.015$\pm$0.016
17	-0.004$\pm$0.014	-0.004$\pm$0.018	-0.008$\pm$0.024
18	-0.015$\pm$0.013	-0.014$\pm$0.017	-0.015$\pm$0.019
19	0.005$\pm$0.013	0.010$\pm$0.015	0.009$\pm$0.015
20	-0.005$\pm$0.013	-0.003$\pm$0.016	-0.004$\pm$0.017
21	0.025$\pm$0.015	0.023$\pm$0.015	0.025$\pm$0.016
22	0.041$\pm$0.017	0.036$\pm$0.017	0.039$\pm$0.017
23	0.014$\pm$0.013	0.012$\pm$0.014	0.013$\pm$0.015
24	-0.002$\pm$0.014	-0.004$\pm$0.016	-0.005$\pm$0.017
25	0.009$\pm$0.014	0.008$\pm$0.015	0.008$\pm$0.015
26	-0.017$\pm$0.014	0.008$\pm$0.015	-0.051$\pm$0.068
27	-0.012$\pm$0.012	-0.016$\pm$0.018	-0.020$\pm$0.020
28	-0.025$\pm$0.012	-0.116$\pm$0.090	-0.157$\pm$0.103
29	-0.019$\pm$0.014	-0.017$\pm$0.018	-0.020$\pm$0.020
30	0.017$\pm$0.015	0.019$\pm$0.017	0.019$\pm$0.020
31	-0.013$\pm$0.013	-0.019$\pm$0.028	-0.043$\pm$0.058
32	0.012$\pm$0.012	0.014$\pm$0.013	0.015$\pm$0.014
33	-0.014$\pm$0.014	-0.019$\pm$0.019	-0.023$\pm$0.024
34	0.002$\pm$0.013	0.001$\pm$0.015	-0.001$\pm$0.016
35	-0.002$\pm$0.015	-0.006$\pm$0.017	-0.009$\pm$0.019
36	-0.022$\pm$0.003	-0.002$\pm$0.015	-0.002$\pm$0.016
37	-0.001$\pm$0.015	-0.001$\pm$0.017	-0.003$\pm$0.019
38	-0.004$\pm$0.014	-0.005$\pm$0.017	-0.006$\pm$0.018
39	0.004$\pm$0.014	0.005$\pm$0.015	0.006$\pm$0.016
40	0.006$\pm$0.014	0.003$\pm$0.015	0.003$\pm$0.016
41	-0.010$\pm$0.014	-0.006$\pm$0.016	-0.008$\pm$0.017
42	-0.012$\pm$0.012	-0.019$\pm$0.019	-0.025$\pm$0.024
43	-0.012$\pm$0.014	-0.016$\pm$0.021	-0.021$\pm$0.024
44	-0.024$\pm$0.013	-0.025$\pm$0.018	-0.029$\pm$0.025

Table B.11: Comparison of the best-fit of P_{ee} p_2 in alternative datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Dataset	QSigEx	Mean of $\pm\sigma$	Mean$\pm$RMS of ML
0	0.992±0.093	0.975±0.093	0.977±0.094
1	1.031±0.094	1.003±0.094	1.017±0.094
2	1.003±0.093	0.977±0.093	0.975±0.092
3	1.004±0.093	0.995±0.093	0.988±0.095
4	0.989±0.093	0.973±0.094	0.973±0.094
5	1.010±0.093	0.986±0.095	0.991±0.094
6	1.030±0.094	1.022±0.095	1.024±0.096
7	1.004±0.094	0.977±0.091	0.979±0.092
8	1.019±0.094	0.993±0.096	0.990±0.095
9	1.014±0.094	0.996±0.094	0.996±0.093
10	1.006±0.093	0.980±0.096	0.983±0.098
11	0.987±0.093	0.972±0.094	0.975±0.094
12	0.986±0.093	0.959±0.095	0.958±0.094
13	0.994±0.093	0.966±0.092	0.966±0.094
14	1.002±0.093	0.985±0.094	0.984±0.094
15	0.957±0.093	0.911±0.092	0.918±0.092
16	0.966±0.093	0.959±0.094	0.947±0.091
17	0.992±0.093	0.962±0.092	0.963±0.094
18	1.014±0.094	0.988±0.095	0.987±0.095
19	1.003±0.093	0.985±0.093	0.982±0.094
20	1.005±0.093	0.977±0.094	0.986±0.095
21	0.997±0.093	0.969±0.093	0.976±0.092
22	0.991±0.094	0.969±0.093	0.973±0.093
23	1.000±0.093	0.977±0.096	0.980±0.098
24	0.993±0.093	0.966±0.091	0.965±0.094
25	1.018±0.094	1.009±0.094	1.004±0.093
26	0.994±0.093	0.976±0.093	0.982±0.095
27	0.999±0.093	0.966±0.099	0.975±0.098
28	1.023±0.093	1.030±0.088	1.033±0.090
29	0.999±0.093	0.969±0.092	0.975±0.092
30	1.001±0.094	0.970±0.090	0.969±0.090
31	1.022±0.093	1.001±0.093	1.006±0.093
32	1.028±0.093	1.006±0.095	1.015±0.094
33	0.992±0.093	0.977±0.098	0.967±0.097
34	1.032±0.094	1.020±0.096	1.013±0.096
35	0.987±0.093	0.966±0.095	0.967±0.096
36	1.104±0.002	0.954±0.092	0.963±0.094
37	0.995±0.093	0.966±0.091	0.971±0.093
38	0.995±0.093	0.966±0.092	0.970±0.093
39	1.014±0.094	0.993±0.094	0.993±0.094
40	0.984±0.093	0.959±0.090	0.963±0.092
41	0.976±0.093	0.951±0.093	0.959±0.093
42	1.016±0.093	0.988±0.091	0.992±0.092
43	1.009±0.094	0.980±0.096	0.989±0.096
44	1.002±0.093	0.976±0.096	0.982±0.097

Table B.12: Comparison of the best-fit of ^{8}B Scale in alternative datasets between QSigEx and MCMC. From MCMC the best-fit is mean of 68% confidence intervals.

Appendix C

Tables for Energy Spectra

Energy spectra was binned in 18-binned histograms. Following tables C.1 to C.3 list number of events and the uncertainties in the number of events for each energy bin for charged current interactions, elastic scattering interactions on electrons initiated by ν_e and elastic scattering interactions initiated by ν_μ and ν_τ. Each interaction is split into day and night event. Tables C.4 to C.9 give bin-by-bin correlation matrix for each interactions.

Table C.1: Day and night spectra for charged current interactions. The number of day events were converted in to number of interactions/deuterium/sec/0.5 MeV by dividing them with $6.023 \times 10^{31} \times 176.59 \times 3600.0 \times 24.0$ number and the night events by dividing them by $6.023 \times 10^{31} \times 208.85 \times 3600.0 \times 24.0$ number.

Electron Recoil Energy (MeV)	**Number of day Events**	**Number of night Events**
6.25	128.161±5.77613	164.446±6.38619
6.75	130.157±4.91089	163.539±5.61196
7.25	126.292±4.34212	157.663±5.06666
7.75	118.718±3.72874	148.702±4.52998
8.25	106.837±3.40636	134.07±3.97687
8.75	94.6407±3.07815	117.276±3.50696
9.25	78.9907±2.61173	98.3046±3.03786
9.75	64.4415±2.30342	80.2439±2.69206
10.25	50.6013±2.02562	62.7512±2.36483
10.75	38.0521±1.71383	47.4103±1.95994
11.25	27.847±1.40866	34.1766±1.59022
11.75	19.4181±1.12626	23.9929±1.28792
12.25	12.8989±0.956992	15.8856±1.08833
12.75	8.30016±0.629247	10.3212±0.738847
13.25	5.1404±0.417694	6.27387±0.555073
13.75	3.12103±0.287746	3.88789±0.34504
14.25	1.68727±0.182381	2.22981±0.249763
14.75	0.973839±0.112694	1.18125±0.149017

Table C.2: Day and night spectra for elastic scattering (ν_e) interactions. The number of day events were converted in to number of interactions/electron/sec/0.5 MeV by dividing them with $3.0115 \times 10^{32} \times 176.59 \times 3600 \times 24.0$ number and night events by $3.0115 \times ^{32} \times 208.85 \times 3600 \times 24$ number.

Electron Recoil Energy (MeV)	**Number of day Events**	**Number of night Events**
6.25	15.5156±0.863691	19.355±1.01362
6.75	13.6872±0.657437	17.1689±0.738539
7.25	11.8438±0.495311	14.8158±0.551523
7.75	10.1729±0.379486	12.5923±0.422607
8.25	8.5329±0.293008	10.4636±0.327038
8.75	7.10183±0.249055	8.5381±0.299595
9.25	5.66037±0.210368	6.93217±0.236234
9.75	4.53923±0.166478	5.46893±0.19781
10.25	3.54031±0.147186	4.19611±0.176436
10.75	2.67704±0.118661	3.14829±0.138553
11.25	2.02177±0.103066	2.3119±0.11574
11.75	1.45822±0.0850116	1.68172±0.096359
12.25	1.03263±0.0633336	1.13624±0.0782564
12.75	0.711275±0.0520842	0.780388±0.0575016
13.25	0.477284±0.0397303	0.514336±0.0472819
13.75	0.317251±0.0321526	0.336964±0.0291887
14.25	0.197213±0.020552	0.209081±0.0257504
14.75	0.116167±0.0139827	0.124081±0.0138656

Table C.3: Day and night spectra for Elastic scattering (ν_μ,ν_τ) interactions. The number of day events were converted in to number of interactions/electron/sec/0.5 MeV by dividing them with $3.0115 \times 10^{32} \times 176.59 \times 3600 \times 24.0$ number and night events by $3.0115 \times^{32} \times 208.85 \times 3600 \times 24$ number.

Electron Recoil Energy (MeV)	**Number of day Events**	**Number of night Events**
6.25	5.41638±0.41152	6.53605±0.483418
6.75	4.78631±0.342195	5.64621±0.409934
7.25	3.99489±0.284223	4.84724±0.346718
7.75	3.36413±0.240244	3.9754±0.275416
8.25	2.79096±0.199861	3.3126±0.226457
8.75	2.26029±0.170749	2.71455±0.187946
9.25	1.83393±0.124608	2.16615±0.152532
9.75	1.47884±0.107	1.69401±0.1225
10.25	1.10656±0.0852571	1.26621±0.0985134
10.75	0.846701±0.066899	0.974217±0.075257
11.25	0.634212±0.0534467	0.733762±0.0594344
11.75	0.469661±0.0430702	0.530463±0.047662
12.25	0.3192±0.0333371	0.359059±0.0355644
12.75	0.217508±0.0237888	0.254486±0.0266218
13.25	0.151988±0.017512	0.167864±0.0207501
13.75	0.0973477±0.0128699	0.109937±0.0131184
14.25	0.0571667±0.00864015	0.0647462±0.00903563
14.75	0.038622±0.00615517	0.0424741±0.0066879

1.000	0.972	0.871	0.685	0.443	0.213	0.029	-0.097	-0.187	-0.246	-0.286	-0.307	-0.319	-0.319	-0.311	-0.296	-0.268	-0.265
0.972	1.000	0.960	0.831	0.627	0.411	0.221	0.081	-0.026	-0.104	-0.163	-0.203	-0.246	-0.252	-0.246	-0.243	-0.231	-0.226
0.871	0.960	1.000	0.952	0.814	0.635	0.461	0.320	0.204	0.112	0.037	-0.019	-0.095	-0.106	-0.104	-0.116	-0.126	-0.120
0.685	0.831	0.952	1.000	0.952	0.837	0.699	0.574	0.463	0.367	0.282	0.214	0.116	0.095	0.090	0.054	0.009	0.017
0.443	0.627	0.814	0.952	1.000	0.964	0.881	0.789	0.697	0.611	0.529	0.458	0.352	0.324	0.312	0.254	0.177	0.184
0.213	0.411	0.635	0.837	0.964	1.000	0.974	0.922	0.858	0.789	0.719	0.654	0.554	0.521	0.504	0.430	0.328	0.335
0.029	0.221	0.461	0.699	0.881	0.974	1.000	0.985	0.950	0.904	0.851	0.797	0.711	0.680	0.661	0.583	0.470	0.475
-0.097	0.081	0.320	0.574	0.789	0.922	0.985	1.000	0.989	0.963	0.926	0.885	0.815	0.785	0.765	0.684	0.565	0.568
-0.187	-0.026	0.204	0.463	0.697	0.858	0.950	0.989	1.000	0.991	0.970	0.940	0.886	0.860	0.840	0.760	0.640	0.642
-0.246	-0.104	0.112	0.367	0.611	0.789	0.904	0.963	0.991	1.000	0.992	0.975	0.937	0.916	0.899	0.826	0.709	0.712
-0.286	-0.163	0.037	0.282	0.529	0.719	0.851	0.926	0.970	0.992	1.000	0.994	0.969	0.956	0.941	0.879	0.772	0.773
-0.307	-0.203	-0.019	0.214	0.458	0.654	0.797	0.885	0.940	0.975	0.994	1.000	0.987	0.981	0.970	0.918	0.821	0.821
-0.319	-0.246	-0.095	0.116	0.352	0.554	0.711	0.815	0.886	0.937	0.969	0.987	1.000	0.991	0.979	0.922	0.821	0.821
-0.319	-0.252	-0.106	0.095	0.324	0.521	0.680	0.785	0.860	0.916	0.956	0.981	0.991	1.000	0.995	0.963	0.887	0.887
-0.311	-0.246	-0.104	0.090	0.312	0.504	0.661	0.765	0.840	0.899	0.941	0.970	0.979	0.995	1.000	0.980	0.916	0.918
-0.296	-0.243	-0.116	0.054	0.254	0.430	0.583	0.684	0.760	0.826	0.879	0.918	0.922	0.963	0.980	1.000	0.976	0.976
-0.268	-0.231	-0.126	0.009	0.177	0.328	0.470	0.565	0.640	0.709	0.772	0.821	0.821	0.887	0.916	0.976	1.000	0.997
-0.265	-0.226	-0.120	0.017	0.184	0.335	0.475	0.568	0.642	0.712	0.773	0.821	0.821	0.887	0.918	0.976	0.997	1.000

Table C.4: Bin-by-bin correlation matrix of day charged current spectra.

1.000	0.970	0.858	0.663	0.425	0.202	0.030	-0.090	-0.178	-0.227	-0.265	-0.283	-0.288	-0.288	-0.288	-0.281	-0.265	-0.254
0.970	1.000	0.957	0.822	0.619	0.408	0.229	0.095	-0.015	-0.083	-0.143	-0.179	-0.221	-0.221	-0.231	-0.223	-0.220	-0.212
0.858	0.957	1.000	0.951	0.815	0.641	0.474	0.337	0.215	0.132	0.053	-0.001	-0.079	-0.081	-0.106	-0.100	-0.118	-0.117
0.663	0.822	0.951	1.000	0.952	0.841	0.709	0.589	0.469	0.381	0.291	0.225	0.121	0.114	0.070	0.071	0.024	0.013
0.425	0.619	0.815	0.952	1.000	0.965	0.887	0.799	0.700	0.622	0.535	0.467	0.354	0.344	0.288	0.282	0.212	0.190
0.202	0.408	0.641	0.841	0.965	1.000	0.976	0.926	0.857	0.795	0.721	0.659	0.551	0.539	0.473	0.464	0.374	0.343
0.030	0.229	0.474	0.709	0.887	0.976	1.000	0.985	0.946	0.903	0.846	0.795	0.700	0.687	0.620	0.606	0.508	0.471
-0.090	0.095	0.337	0.589	0.799	0.926	0.985	1.000	0.986	0.962	0.920	0.880	0.803	0.789	0.719	0.703	0.597	0.556
-0.178	-0.015	0.215	0.469	0.700	0.857	0.946	0.986	1.000	0.993	0.971	0.945	0.886	0.875	0.815	0.799	0.698	0.656
-0.227	-0.083	0.132	0.381	0.622	0.795	0.903	0.962	0.993	1.000	0.991	0.974	0.932	0.922	0.864	0.848	0.746	0.704
-0.265	-0.143	0.053	0.291	0.535	0.721	0.846	0.920	0.971	0.991	1.000	0.995	0.968	0.963	0.921	0.906	0.817	0.777
-0.283	-0.179	-0.001	0.225	0.467	0.659	0.795	0.880	0.945	0.974	0.995	1.000	0.986	0.984	0.950	0.937	0.855	0.817
-0.288	-0.221	-0.079	0.121	0.354	0.551	0.700	0.803	0.886	0.932	0.968	0.986	1.000	0.996	0.961	0.946	0.864	0.825
-0.288	-0.221	-0.081	0.114	0.344	0.539	0.687	0.789	0.875	0.922	0.963	0.984	0.996	1.000	0.978	0.969	0.899	0.865
-0.288	-0.231	-0.106	0.070	0.288	0.473	0.620	0.719	0.815	0.864	0.921	0.950	0.961	0.978	1.000	0.997	0.968	0.947
-0.281	-0.223	-0.100	0.071	0.282	0.464	0.606	0.703	0.799	0.848	0.906	0.937	0.946	0.969	0.997	1.000	0.978	0.961
-0.265	-0.220	-0.118	0.024	0.212	0.374	0.508	0.597	0.698	0.746	0.817	0.855	0.864	0.899	0.968	0.978	1.000	0.995
-0.254	-0.212	-0.117	0.013	0.190	0.343	0.471	0.556	0.656	0.704	0.777	0.817	0.825	0.865	0.947	0.961	0.995	1.000

Table C.5: Bin-by-bin correlation matrix of night charged current spectra.

1.000	0.991	0.962	0.881	0.762	0.565	0.355	0.176	0.029	-0.071	-0.131	-0.165	-0.210	-0.192	-0.182	-0.157	-0.158	-0.140
0.991	1.000	0.989	0.934	0.837	0.662	0.462	0.285	0.133	0.027	-0.040	-0.083	-0.150	-0.129	-0.117	-0.091	-0.093	-0.075
0.962	0.989	1.000	0.973	0.907	0.758	0.576	0.407	0.252	0.142	0.065	0.013	-0.072	-0.053	-0.043	-0.022	-0.022	-0.011
0.881	0.934	0.973	1.000	0.974	0.882	0.737	0.588	0.447	0.336	0.261	0.204	0.089	0.118	0.132	0.155	0.152	0.161
0.762	0.837	0.907	0.974	1.000	0.959	0.860	0.741	0.609	0.509	0.426	0.362	0.240	0.260	0.264	0.271	0.273	0.267
0.565	0.662	0.758	0.882	0.959	1.000	0.966	0.895	0.801	0.717	0.648	0.587	0.457	0.479	0.482	0.482	0.481	0.466
0.355	0.462	0.576	0.737	0.860	0.966	1.000	0.976	0.924	0.867	0.811	0.760	0.641	0.658	0.655	0.640	0.641	0.612
0.176	0.285	0.407	0.588	0.741	0.895	0.976	1.000	0.977	0.945	0.903	0.859	0.764	0.768	0.755	0.725	0.724	0.683
0.029	0.133	0.252	0.447	0.609	0.801	0.924	0.977	1.000	0.986	0.969	0.944	0.864	0.874	0.865	0.838	0.834	0.792
-0.071	0.027	0.142	0.336	0.509	0.717	0.867	0.945	0.986	1.000	0.987	0.971	0.917	0.920	0.904	0.859	0.866	0.815
-0.131	-0.040	0.065	0.261	0.426	0.648	0.811	0.903	0.969	0.987	1.000	0.993	0.949	0.960	0.950	0.917	0.914	0.869
-0.165	-0.083	0.013	0.204	0.362	0.587	0.760	0.859	0.944	0.971	0.993	1.000	0.968	0.981	0.974	0.942	0.937	0.895
-0.210	-0.150	-0.072	0.089	0.240	0.457	0.641	0.764	0.864	0.917	0.949	0.968	1.000	0.985	0.957	0.899	0.898	0.833
-0.192	-0.129	-0.053	0.118	0.260	0.479	0.658	0.768	0.874	0.920	0.960	0.981	0.985	1.000	0.988	0.955	0.953	0.910
-0.182	-0.117	-0.043	0.132	0.264	0.482	0.655	0.755	0.865	0.904	0.950	0.974	0.957	0.988	1.000	0.980	0.980	0.953
-0.157	-0.091	-0.022	0.155	0.271	0.482	0.640	0.725	0.838	0.859	0.917	0.942	0.899	0.955	0.980	1.000	0.981	0.979
-0.158	-0.093	-0.022	0.152	0.273	0.481	0.641	0.724	0.834	0.866	0.914	0.937	0.898	0.953	0.980	0.981	1.000	0.977
-0.140	-0.075	-0.011	0.161	0.267	0.466	0.612	0.683	0.792	0.815	0.869	0.895	0.833	0.910	0.953	0.979	0.977	1.000

Table C.6: Bin-by-bin correlation matrix of day elastic scattering spectra for ν_e.

1.000	0.990	0.953	0.867	0.736	0.542	0.356	0.202	0.082	-0.004	-0.051	-0.072	-0.089	-0.086	-0.070	-0.077	-0.052	-0.057
0.990	1.000	0.986	0.926	0.819	0.643	0.466	0.312	0.184	0.092	0.036	0.008	-0.032	-0.027	-0.010	-0.014	0.007	0.004
0.953	0.986	1.000	0.974	0.901	0.754	0.595	0.447	0.316	0.219	0.156	0.119	0.056	0.062	0.078	0.077	0.090	0.090
0.867	0.926	0.974	1.000	0.968	0.880	0.754	0.625	0.506	0.409	0.344	0.302	0.217	0.225	0.244	0.242	0.253	0.253
0.736	0.819	0.901	0.968	1.000	0.948	0.867	0.764	0.648	0.563	0.490	0.440	0.338	0.344	0.341	0.349	0.323	0.335
0.542	0.643	0.754	0.880	0.948	1.000	0.971	0.915	0.842	0.772	0.715	0.674	0.567	0.577	0.585	0.587	0.569	0.577
0.356	0.466	0.595	0.754	0.867	0.971	1.000	0.979	0.934	0.885	0.840	0.800	0.701	0.706	0.699	0.703	0.664	0.673
0.202	0.312	0.447	0.625	0.764	0.915	0.979	1.000	0.982	0.956	0.922	0.893	0.806	0.811	0.797	0.806	0.749	0.765
0.082	0.184	0.316	0.506	0.648	0.842	0.934	0.982	1.000	0.989	0.973	0.956	0.890	0.894	0.884	0.886	0.836	0.849
-0.004	0.092	0.219	0.409	0.563	0.772	0.885	0.956	0.989	1.000	0.990	0.980	0.932	0.935	0.914	0.920	0.853	0.870
-0.051	0.036	0.156	0.344	0.490	0.715	0.840	0.922	0.973	0.990	1.000	0.992	0.962	0.964	0.947	0.953	0.892	0.903
-0.072	0.008	0.119	0.302	0.440	0.674	0.800	0.893	0.956	0.980	0.992	1.000	0.977	0.981	0.968	0.970	0.912	0.925
-0.089	-0.032	0.056	0.217	0.338	0.567	0.701	0.806	0.890	0.932	0.962	0.977	1.000	0.996	0.973	0.970	0.906	0.917
-0.086	-0.027	0.062	0.225	0.344	0.577	0.706	0.811	0.894	0.935	0.964	0.981	0.996	1.000	0.984	0.983	0.928	0.940
-0.070	-0.010	0.078	0.244	0.341	0.585	0.699	0.797	0.884	0.914	0.947	0.968	0.973	0.984	1.000	0.988	0.977	0.980
-0.077	-0.014	0.077	0.242	0.349	0.587	0.703	0.806	0.886	0.920	0.953	0.970	0.970	0.983	0.988	1.000	0.959	0.972
-0.052	0.007	0.090	0.253	0.323	0.569	0.664	0.749	0.836	0.853	0.892	0.912	0.906	0.928	0.977	0.959	1.000	0.988
-0.057	0.004	0.090	0.253	0.335	0.577	0.673	0.765	0.849	0.870	0.903	0.925	0.917	0.940	0.980	0.972	0.988	1.000

Table C.7: Bin-by-bin correlation matrix of night elastic scattering spectra for ν_e.

1.000	0.996	0.984	0.961	0.932	0.887	0.836	0.785	0.726	0.656	0.595	0.512	0.429	0.405	0.382	0.322	0.310	0.260
0.996	1.000	0.995	0.981	0.959	0.920	0.874	0.828	0.770	0.701	0.642	0.557	0.470	0.446	0.422	0.358	0.344	0.293
0.984	0.995	1.000	0.994	0.980	0.950	0.913	0.872	0.819	0.754	0.697	0.613	0.526	0.500	0.477	0.410	0.395	0.342
0.961	0.981	0.994	1.000	0.995	0.977	0.949	0.917	0.872	0.814	0.762	0.681	0.596	0.571	0.548	0.481	0.465	0.412
0.932	0.959	0.980	0.995	1.000	0.991	0.972	0.947	0.909	0.857	0.809	0.732	0.649	0.625	0.601	0.533	0.517	0.462
0.887	0.920	0.950	0.977	0.991	1.000	0.994	0.980	0.954	0.914	0.876	0.810	0.735	0.713	0.691	0.627	0.611	0.561
0.836	0.874	0.913	0.949	0.972	0.994	1.000	0.995	0.980	0.951	0.921	0.865	0.797	0.778	0.757	0.697	0.680	0.632
0.785	0.828	0.872	0.917	0.947	0.980	0.995	1.000	0.993	0.974	0.951	0.903	0.843	0.825	0.806	0.747	0.730	0.685
0.726	0.770	0.819	0.872	0.909	0.954	0.980	0.993	1.000	0.993	0.978	0.943	0.896	0.880	0.865	0.813	0.797	0.755
0.656	0.701	0.754	0.814	0.857	0.914	0.951	0.974	0.993	1.000	0.995	0.975	0.941	0.928	0.916	0.873	0.860	0.821
0.595	0.642	0.697	0.762	0.809	0.876	0.921	0.951	0.978	0.995	1.000	0.990	0.965	0.957	0.946	0.908	0.894	0.864
0.512	0.557	0.613	0.681	0.732	0.810	0.865	0.903	0.943	0.975	0.990	1.000	0.989	0.986	0.980	0.956	0.944	0.923
0.429	0.470	0.526	0.596	0.649	0.735	0.797	0.843	0.896	0.941	0.965	0.989	1.000	0.997	0.996	0.981	0.972	0.954
0.405	0.446	0.500	0.571	0.625	0.713	0.778	0.825	0.880	0.928	0.957	0.986	0.997	1.000	0.997	0.987	0.979	0.965
0.382	0.422	0.477	0.548	0.601	0.691	0.757	0.806	0.865	0.916	0.946	0.980	0.996	0.997	1.000	0.992	0.986	0.973
0.322	0.358	0.410	0.481	0.533	0.627	0.697	0.747	0.813	0.873	0.908	0.956	0.981	0.987	0.992	1.000	0.993	0.990
0.310	0.344	0.395	0.465	0.517	0.611	0.680	0.730	0.797	0.860	0.894	0.944	0.972	0.979	0.986	0.993	1.000	0.981
0.260	0.293	0.342	0.412	0.462	0.561	0.632	0.685	0.755	0.821	0.864	0.923	0.954	0.965	0.973	0.990	0.981	1.000

Table C.8: Bin-by-bin correlation matrix of day elastic scattering spectra for ν_μ and ν_τ.

1.000	0.996	0.985	0.964	0.933	0.893	0.839	0.797	0.729	0.664	0.611	0.524	0.451	0.422	0.360	0.358	0.306	0.268
0.996	1.000	0.996	0.982	0.958	0.925	0.876	0.837	0.772	0.709	0.655	0.567	0.490	0.461	0.395	0.393	0.337	0.298
0.985	0.996	1.000	0.995	0.979	0.954	0.912	0.878	0.817	0.758	0.706	0.617	0.540	0.509	0.442	0.439	0.382	0.339
0.964	0.982	0.995	1.000	0.994	0.979	0.947	0.920	0.867	0.814	0.765	0.681	0.603	0.573	0.506	0.502	0.444	0.401
0.933	0.958	0.979	0.994	1.000	0.995	0.975	0.955	0.913	0.866	0.823	0.745	0.671	0.641	0.576	0.571	0.515	0.470
0.893	0.925	0.954	0.979	0.995	1.000	0.991	0.979	0.947	0.909	0.871	0.799	0.728	0.702	0.636	0.633	0.573	0.532
0.839	0.876	0.912	0.947	0.975	0.991	1.000	0.996	0.980	0.954	0.926	0.867	0.807	0.781	0.725	0.718	0.666	0.625
0.797	0.837	0.878	0.920	0.955	0.979	0.996	1.000	0.992	0.973	0.950	0.898	0.844	0.821	0.765	0.760	0.708	0.668
0.729	0.772	0.817	0.867	0.913	0.947	0.980	0.992	1.000	0.994	0.982	0.946	0.903	0.884	0.838	0.832	0.787	0.751
0.664	0.709	0.758	0.814	0.866	0.909	0.954	0.973	0.994	1.000	0.996	0.974	0.940	0.927	0.886	0.883	0.839	0.811
0.611	0.655	0.706	0.765	0.823	0.871	0.926	0.950	0.982	0.996	1.000	0.988	0.966	0.954	0.922	0.917	0.881	0.853
0.524	0.567	0.617	0.681	0.745	0.799	0.867	0.898	0.946	0.974	0.988	1.000	0.987	0.986	0.964	0.963	0.931	0.916
0.451	0.490	0.540	0.603	0.671	0.728	0.807	0.844	0.903	0.940	0.966	0.987	1.000	0.993	0.986	0.976	0.965	0.943
0.422	0.461	0.509	0.573	0.641	0.702	0.781	0.821	0.884	0.927	0.954	0.986	0.993	1.000	0.989	0.992	0.967	0.962
0.360	0.395	0.442	0.506	0.576	0.636	0.725	0.765	0.838	0.886	0.922	0.964	0.986	0.989	1.000	0.991	0.991	0.982
0.358	0.393	0.439	0.502	0.571	0.633	0.718	0.760	0.832	0.883	0.917	0.963	0.976	0.992	0.991	1.000	0.979	0.985
0.306	0.337	0.382	0.444	0.515	0.573	0.666	0.708	0.787	0.839	0.881	0.931	0.965	0.967	0.991	0.979	1.000	0.981
0.268	0.298	0.339	0.401	0.470	0.532	0.625	0.668	0.751	0.811	0.853	0.916	0.943	0.962	0.982	0.985	0.981	1.000

Table C.9: Bin-by-bin correlation matrix of night elastic scattering spectra for ν_μ and ν_τ.