Three-Phase Extraction of the Electron Neutrino Survival Probability at the Sudbury Neutrino Observatory

by

Pierre-Luc Drouin

A thesis submitted to the

Faculty of Graduate and Postdoctoral Affairs

in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

Department of Physics

Carleton University

Ottawa-Carleton Institute for Physics

Ottawa, Ontario

December, 2011

"Are you suggesting coconuts migrate?"

Unnamed castle guard – Monty Python and the Holy Grail

Abstract

Unique in its ability to solve the solar neutrino problem, the Sudbury Neutrino Observatory (SNO) was a heavy water Čerenkov experiment designed to measure the total flux of neutrinos from the Sun, as well as the electron neutrino component of this flux. This thesis presents the analysis used to produce the definitive SNO results, by performing a combined fit of the data from the three phases of the experiment. The SNO experiment is first motivated and introduced, followed by a discussion of the methods used to reconstruct observables used to discriminate between the different signals and backgrounds. A thorough derivation of the likelihood function used to perform the final fit is then carried out, including the development of a method intended to fit the solar neutrino data in a totally model-independent way for both active and non-active neutrino states. Verifications of the signal extraction process are then presented, based on the results from multiple ensemble tests.

As published, a fit to the SNO data that used a second-order polynomial to directly describe the survival probability of electron-type neutrinos produced through the ^{8}B decays in the Sun led to a value of $5.25^{+0.16}_{-0.16}(\text{stat.})^{+0.11}_{-0.13}(\text{syst.}) \times 10^6$ cm^{-2}s^{-1} for the total flux of ^{8}B neutrinos, assuming only active neutrino flavours. The same fit resulted in a fitted survival probability value of $0.317^{+0.016}_{-0.016}(\text{stat.})^{+0.009}_{-0.009}(\text{syst.})$ for the zeroth-order polynomial parameter, which represents the best SNO rejection of a hypothesis predicting no neutrino oscillation. A first-order day/night asymmetry polynomial was extracted along with the survival probability function and yields to $0.046^{+0.031}_{-0.031}(\text{stat.})^{+0.014}_{-0.014}(\text{syst.})$ at the zeroth order. The alternate hypotheses that there were no spectral distortions of the electron-type

neutrino survival probability and that there were no day/night distortions of this probability could not be rejected at the 95% confidence level.

A fit allowing non-active neutrino flavours was also performed. When applied to the BS05(OP) model, this fit resulted in zeroth-order measurements of $0.292^{+0.067}_{-0.039}$ and $0.12^{+0.14}_{-0.23}$ for the electron-type neutrino survival probability and the probability of conversion of these neutrinos to a non-active flavour, respectively. This result is not inconsistent with the assumption of only three neutrino states.

Acknowledgements

Je veux remercier spécialement Étienne Côté, Raymond Villemure et Jacques Carel, ces merveilleux enseignants de la polyvalente et du cégep de Lévis dont la passion des sciences a grandement marqué mon apprentissage.

À mes parents, merci de votre présence et de constamment vous soucier de mon bien-être. Merci de toujours m'avoir encouragé à poursuivre mes passions.

Above all, thanks to Kate, for your love and affection, for your understanding and patience when I was staying up until the middle of the night to debug my code or analyse results. Thanks also for hooking me on espresso and for the frozen lasagna, which both contributed to the timely completion of this thesis!

Statement of Originality

Except when otherwise mentioned, the content of this thesis presents the results of original work from the author. Information originating from other sources is referenced in the text. Since the Sudbury Neutrino Observatory (SNO) was an experiment that was already well in motion at the time the author joined its collaboration, which included many scientists, it is important to situate the contributions from the author and to relate them to the different sections of this thesis, when appropriate.

The initial contact of the author with the experiment started in 2002 and 2003 when, as an undergraduate summer student, he designed and implemented most of a signal extraction package for the collaboration that was used by Darren Grant and Kathryn Miknaitis as a flexible tool to perform verifications in the context of the D_2O and salt phase analyses [1, 2]. This package was developed in C/C++ and relied on a variety of classes from the ROOT data analysis framework [3]. As a new graduate student in 2004, the author first worked on the optimisation of the configuration of probability density functions (PDFs) to reduce biases when performing SNO signal extraction, while contributing to the data collection of the experiment, through some detector operation and helping at the deployment of calibration sources.

The author's contributions that are directly related to the content of this thesis began when he was tasked to develop an efficient event position and direction reconstruction algorithm for the NCD phase, which led to the development and implementation of FTN, the tool used to reconstruct the position and direction of the SNO data as well as of the

simulation and calibration events in this third phase of the experiment. The author also studied the different systematic uncertainties related to position and direction reconstruction in the NCD phase and the reconstructed event direction systematic uncertainties for LETA [4], which involved the analytical modeling of the PDF for the first interaction point of the ^{16}N calibration source (Section A.2). Section 2.1.1.2 of this thesis summarises the functioning of FTN and presents the results from the study of the related systematic uncertainties.

The main contribution of the author to the SNO analysis consists of the final extraction of the SNO signal using a combined analysis of the whole dataset of the experiment. This endeavour was accomplished through multiple steps, including the design and implementation of a new signal extraction library whose performance could make such an analysis possible, the modeling of a likelihood function for the three-phase analysis, the implementation of this function in a signal extraction package, verifications of the tool via thorough ensemble testing, the reproduction of the signal extraction results from the previous publications, verifications of the LETA results and the final fit to the data of the SNO experiment. This effort led to many improvements of the final analysis.

The QSigEx library was entirely designed and implemented by the author. It consists of a framework that regroups multiple tools used to evaluate likelihood functions, such as differently optimised histogram/PDF classes, multi-threaded processing units and multi-threaded data arrays that use access-based memory management. It depends on the ROOT framework [3], mainly through the MINUIT minimisation package, as well as different GNU [5] and FreeBSD [6] libraries. The QSigEx library, written in C/C++ and x86 assembly languages, gave to the three-phase extraction package the performance and

flexibility that was necessary to achieve all the tests and verifications that ultimately led to the final fit.

In parallel to the development of the QSigEx library, the author worked at modeling likelihood functions for the three-phase analysis. Chapters 3 and 4 present the derivations of these functions, as independently obtained by the author. In Chapter 4, the likelihood function for an unconstrained reconstructed electron energy fit (Section 4.1) had been used by previous analyses, such as LETA. Also, the usage of a P_{ee} polynomial function to extract the SNO data was an original idea of Alain Bellerive and the usage of an unbinned version of this method, as presented in Section 4.2.3, had been previously attempted. The binned version of the P_{ee} polynomial fit (Section 4.2.4), is an original idea of the author and was used to produce the results of the LETA publication [4]. All other methods derived in Chapter 4, including the improved binned P_{ee} polynomial fit of Section 4.3.2 as well the $P_{ee} + P_{ea}$ methods, were designed and modeled by the author. Finally, the author developed the P_{ee} polynomial integration algorithm of Section 4.4.2.

The three-phase analysis is based on the work that had been previously done by the collaboration for LETA [4] and the analysis of the NCD phase data [7]. Most of the systematic uncertainty study, external measurement, detector calibration, data selection and data simulation work that was performed for these previous results was incorporated in the three-phase analysis. The main new elements that were introduced to the three-phase analysis by the author, in addition to the aforementioned improvements to the likelihood function, are the evaluation of the effects of the statistical PDF uncertainties (Section 3.8), the reevaluation of the rate-to-events factors for the PMT events in the NCD phase, the development of more consistent scaling of these factors for all phases as a function of the systematic uncertainties (Section 3.2), the removal of the positive

constraints for the background rates (Section 3.4), the fine-tuning of the evaluation of the background rate constraints (Section 3.5.4), the introduction of the $\mathrm{ES}_{\mu\tau}$ signal, the improvements to the propagation of the systematic uncertainty for the shape of the ^{8}B spectrum (Section D.3.6.1), the treatment of the backgrounds for other neutrons, atmospheric and instrumental events as event classes in the likelihood function, the fit of asymmetrical Gaussian functions while performing scans of the systematic uncertainties, as well as the zoom-in procedure (Section 5.4.2), the evaluation of systematic effects through the usage of multiple random shifts of the systematic parameters (Section 5.4.3) and the combination of the asymmetrical systematic uncertainty effects using a Monte Carlo simulation (Section 5.4.4).

Following the completion of the signal extraction package for the three-phase analysis, the author spent considerable effort at submitting QSigEx to different tests in order to verify the quality of the results. A multitude of ensemble tests were performed for the different SNO phases independently, as well as for the three-phase analysis, and using the different supported likelihood configurations, such that the signal extraction mechanism be tested at an unprecedented level for the SNO experiment. A fraction of the ensemble test results are presented in Chapter 6. Also, the results from the NCD phase analysis [7] were reproduced, followed by the results for the unconstrained fit of LETA [4]. Finally, several months of work were spent at reproducing, improving and verifying the results from the P_{ee} fit of LETA.

This thesis was written using the LaTeX 2_ε document preparation system [8], with the help of the Vim [9] text editor. Figure 2.3 was generated by the author using XSnoed, the SNO event display software. Feynman diagrams were generated using the feynmp MetaPost package [10]. The FTN representation of the SNO detector in Figure 2.6 was

produced by the author using the ROOT graphical libraries [3]. All other figures, unless otherwise noted, were produced using PSPlot, a plotting package written by the author using the T$_{\!E}$X [11] typesetting system and the PSTricks macros [12].

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

AV	Acrylic Vessel.
CC	Charged Current reaction.
DWK	One of the PSA methods.
ES	Electron Scattering reaction.
FTK	Energy reconstruction algorithm developed for LETA.
FTN	Vertex reconstruction algorithm used for the NCD phase.
FTP	Path fitter used by the D_2O and salt phases to perform vertex reconstruction.
GT	Global Trigger of the SNO detector.
GTID	Global Trigger Identification Number.
LETA	Low-Energy Threshold Analysis of the D_2O and salt phases.
MC	Monte Carlo simulation.
MSW	Mikheyev-Smirnov-Wolfenstein effect.
NAP	One of the PSA methods.
NC	Neutral Current reaction.
NCD	Neutral Current Detector.
NCD-NCD	NCD-side of the data for the NCD phase.
NCD-PMT	PMT-side of the data for the NCD phase.
OWL PMT	Outward Looking PMT.
PDF	Probability Density Function.
PMF	Probability Mass Function.
PMT	PhotoMultiplier Tube.
PSA	Pulse-Shape Analysis of the NCD data.
QGF	One of the PSA methods.
QSigEx library	Library of optimised, multi-threaded tools used to compute likelihood functions.
QSigEx	Signal extraction package used to perform the three-phase analysis.
RMS	Root Mean Square.
RSP	Energy reconstruction algorithm used indirectly for LETA and as the main algorithm for the NCD phase.
SNO	Sudbury Neutrino Observatory.
SNOMAN	SNO Monte carlo and ANalysis program.
SSM	Solar Standard Model.
TAC	Time-to-Amplitude Converter.
P_{ee}	Electron neutrino survival probability.

A_{ee}	Day/night asymmetry for P_{ee}.
P_{ea}	Probability that an electron neutrino transforms into another active flavour.
A_{ea}	Day/night asymmetry for P_{ea}.
P_{es}	Probability that an electron neutrino transforms into a sterile flavour.
$i, j, k, \ldots$	Integers (indices) or values of discrete random variables.
$u, v, \ldots, x, y$	Real numbers or values of continuous random variables.
$\alpha, \beta, \theta, \phi, \ldots$	Parameters.
$\vec{x}$	Vector.
$x_1, x_2, \ldots$	Specific values for x, value of a vector component.
$\mathbf{A}$	Matrix.
A_{ij}	Matrix element.
$\vec{A}_j$	j^{th} column of a matrix.
$b(x)$	If x values are binned in a set of disjoint intervals (or bins) $[x_i, x_{i+1})$, where i is called the bin index, $b(x)$ is defined as the function returning the bin index associated to a given x value.
$\Delta_x(i)$	Bin width of bin i along the x axis defined as $\Delta_x(i) \equiv x_{i+1} - x_i$.
$P(i)$	PMF $P(i) \equiv \Pr(I = i)$, where i is the value taken by the discrete random variable I, such that $P(i)$ is the probability of occurrence of i. $\sum_{\forall i \in I} P(i) = 1$.
$f(x)$	PDF where x is the value taken by the continuous random variable X, such that $f(x)$ is the density of X at $X = x$. $\int_{-\infty}^{+\infty} f(x)dx = 1$.
$f_b(x), \ f_b(i)$	Binned PDF defined as a PDF which value has been averaged for a set of bins, such that $f_b(x) = \frac{1}{\Delta_x(b(x))} \int_{x_{b(x)}}^{x_{b(x)+1}} f(x')dx'$. $f_b(i)$ is equivalent to $f_b(x_i)$. The probability of x to be in bin i follows a PMF which values are given by $P(i) = \int_{x_i}^{x_{i+1}} f(x)dx = \int_{x_i}^{x_{i+1}} f_b(x)dx = f_b(i)\Delta_x(i)$.
$H(x), \ H(i)$	Histogram of data or simulation, for which the bin content corresponds to a number of events. $H(i)$ is equivalent to $H(x_i)$.
$P_i(i)$	Marginal PMF for i. Given a 2D PMF $P(i, j)$, $P_i(i) = \sum_{\forall j \in J} P(i, j)$.
$f_x(x)$	Marginal PDF for x. Given a 2D PDF $f(x, y)$, $f_x(x) = \int_{-\infty}^{+\infty} f(x, y)dy$.
$P(i, j)$	Joint PMF. $\sum_{\substack{\forall i \in I \\ \forall j \in J}} P(i, j) = 1$.
$f(x, y)$	Joint PDF. $\iint_{-\infty}^{+\infty} f(x, y)dxdy = 1$.
$f(x, i)$	Joint PDF. $\sum_{\forall i \in I} \int_{-\infty}^{+\infty} f(x, i)dx = 1$.
$P(i\|j)$	Conditional PMF of i given j. $P(i\|j) = \frac{P(i,j)}{P(j)}$.
$f(x\|y)$	Conditional PDF of x given y. $f(x\|y) = \frac{f(x,y)}{f(y)}$.
$P(i\|x)$	Conditional PMF of i given x. $P(i\|x) = \frac{f(i,x)}{f(x)}$.
$f(x\|i)$	Conditional PDF of x given i. $f(x\|i) = \frac{f(x,i)}{P(i)}$.
$f(x\|y, i)$	Conditional PDF of x given y and i.

$E[x]$, $\langle x \rangle$	Expectation value of x.
$\mathsf{Var}[x]$, σ_x^2	Variance of x.
$\mathsf{Cov}[x, y]$	Covariance between x and y.
$\rho[x, y]$	Correlation factor between x and y.
$\overline{x}$	Sample mean of x_i values.
s_x^2	Sample variance of x_i values.
$\hat{\theta}$	Estimator for θ.
ν	Poisson parameter.
$\delta_{()}$	Kronecker delta which evaluates to 1 if the condition expressed as a subscript is true and evaluates to 0 otherwise.
$L(\vec{\theta})$, $L(\mathbf{x}|\vec{\theta})$	Likelihood function.
$LL(\vec{\theta})$	Log-likelihood function.
$ELL(\vec{\theta})$	Extended log-likelihood function.

Chapter 1

Introduction

The neutrino (ν) adventure debuted in 1930, when Pauli proposed [13] the existence of these neutral particles to preserve the conservation of energy and momentum laws in nuclear beta decays,

$$n \to p + e + \bar{\nu}_e, \tag{1.1}$$

whose electron energy spectrum had been discovered to be continuous by Chadwick in 1914 [14]. More than 25 years later, electron neutrinos (more accurately, electron anti-neutrinos) were directly observed for the first time by the Cowan-Reines neutrino experiment [15]. Muon neutrinos were then detected in 1962 by Danby *et al.* [16]. In 2000, the DONUT Collaboration announced the first observation of the tau neutrino [17], which confirmed the existence of the third active neutrino flavour predicted by the Standard Model.

1.1 Solar Neutrinos

1.1.1 The Standard Solar Model

The Sun generates its energy through the nuclear fusion of light elements into heavier elements, which produces a large quantity of neutrinos. In fact, on Earth, trillions of neutrinos produced by the Sun pass through our bodies every second. The primary neutrino-generating process in the Sun is the proton-proton reaction (pp), which can be expressed as

$$p + p \rightarrow {}^2\text{H} + e^+ + \nu_e. \tag{1.2}$$

Although this process generates the highest neutrino flux, it releases 400 keV of energy or less, which makes it extremely difficult to detect on Earth, because many common radioactive decay processes release similar energies. However, other fusion processes in the Sun produce more energetic neutrinos, as presented in Figure 1.1. This figure lists the solar neutrino reactions that are part of the pp chain, whose initial reaction is shown above, and of the CNO cycle that uses either carbon, nitrogen or oxygen as a catalyst. These reactions are part of the Standard Solar Model [18] (SSM), whose development was initiated by John Bahcall in 1962 [19]. The figure thus shows that the most energetic solar neutrinos are the ^{8}B and hep neutrinos. These neutrinos are produced through the reactions

$$^7\text{Be} + p \rightarrow {}^8\text{B} + \gamma$$
$$^8\text{B} \rightarrow {}^8\text{Be}^* + e^+ + \nu_e \tag{1.3}$$
$$^8\text{Be}^* \rightarrow 2\,^4\text{He}$$
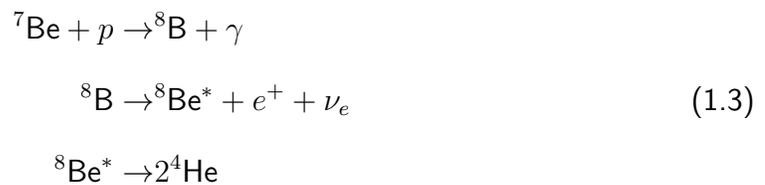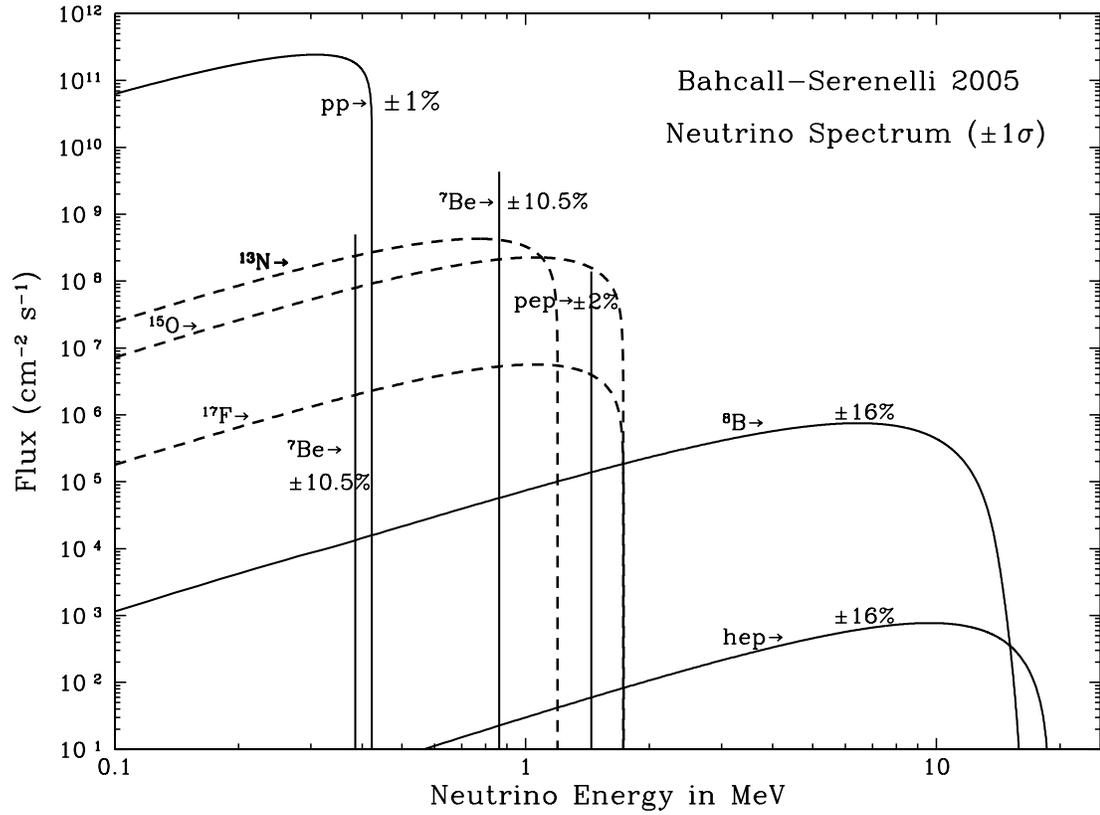
Figure 1.1: Solar neutrino flux spectra at the Earth surface, along with flux uncertainties [18]. The pp chain reactions are shown using solid black lines, while the CNO reactions are shown using dashed lines.

and

$$^{3}\mathrm{He} + p \rightarrow {}^{4}\mathrm{He} + e^{+} + \nu_{e}, \tag{1.4}$$

respectively. The predicted neutrino fluxes for these reactions at the Earth's surface, as provided by the BS05(OP) model [18], are $5.69(1.00 \pm 0.16) \times 10^{6}$ cm^{-2}s^{-1} for ^{8}B and $7.93(1.00 \pm 0.16) \times 10^{3}$ cm^{-2}s^{-1} for hep.

1.1.2 The Solar Neutrino Problem

The creation of the SSM by John Bahcall followed a request by Ray Davis, who was interested to know if the detection of solar neutrinos would be possible using a large tank filled with a chlorine-rich liquid [19]. Through the Homestake experiment [20], Davis and collaborators measured the flux of solar neutrinos using 100 000 gallons (380 m^3) of perchloroethylene, a cleaning fluid. This detector measured neutrinos via the reaction

$$\nu_e + {}^{37}\text{Cl} \rightarrow e^- + {}^{37}\text{Ar}. \tag{1.5}$$

Unexpectedly, from 1968 to 1994, this measurement consistently led to results which were approximately one-third of the SSM predictions. The upper limit of the ^{8}B flux placed by the final results of the Homestake experiment is $(2.25 \pm 0.21) \times 10^6$ cm^{-2}s^{-1}, assuming the entire solar neutrino signal in their detector is from ^{8}B decays [20]. This discrepancy between the predicted and measured solar neutrino fluxes was named the "solar neutrino problem".

1.1.3 Neutrino Oscillation

Although the discrepancy between the observed and theoretical solar neutrino fluxes was rather surprising, it could be explained by a phenomenon that had already been predicted. One should note that solar neutrinos are always produced as electron-type neutrinos in the Sun. Also, experiments such as Homestake measured the solar neutrino flux using interactions only sensitive to the electron neutrino flavour. If neutrinos happened to "oscillate" to undetectable states during their journey to the detector, the measured flux

would become lower than the produced flux. Bruno Pontecorvo had proposed neutrino-antineutrino transitions in 1957 [21]. Pontecorvo's idea had then been used to develop the theory of neutrino oscillation by Maki, Nakagawa and Sakata in 1962 [22].

According to the theory that explains neutrino oscillations, the individual neutrino quantum eigenstates that are involved in weak interactions are different superpositions of the neutrino mass eigenstates. Solar neutrinos are thus created in a superposition of eigenstates that can be expressed as

$$|\nu_e\rangle = \sum_{i=1}^{n} U_{ei} |\nu_i\rangle , \qquad (1.6)$$

where $|\nu_e\rangle$ represents the electron-type neutrino flavour eigenstate, $|\nu_i\rangle$ the neutrino mass eigenstates and U_{ei} the transformation between both bases. The individual mass states then propagate as the plane wave solutions

$$|\nu_i(\vec{x}, t)\rangle = e^{-i(E_i t - \vec{p}_i \cdot \vec{x})} |\nu_i(\vec{0}, 0)\rangle, \qquad (1.7)$$

where E_i and $\vec{p}_i$ are the energy and the momentum of the $|\nu_i\rangle$ state, respectively, and where t and $\vec{x}$ are the time and position of evaluation of the wave relative to the initial $|\nu_i(\vec{0}, 0)\rangle$. If different masses are associated to the different states $|\nu_i\rangle$, their phases vary differently as a function of time and positions, due to the relation $E^2 = p^2 + m^2$. At the ultrarelativistic limit,

$$E_i \approx p_i + \frac{m_i^2}{2p_i}, \qquad (1.8)$$

where m_i are the different neutrino masses. If neutrino energy is given by E_ν, the probability of detection for electron-type neutrinos is thus given by

$$P_{e \to e} = |\langle \nu_e | \nu_e(\vec{x}, t) \rangle|^2 \approx \left| \sum_{i=1}^{n} U_{ei}^* U_{ei} e^{-i \frac{m_i^2 L}{2E_\nu}} \right|^2$$

$$= P_{e \to e}(E_\nu, L), \tag{1.9}$$

where L is defined as the travelled distance by the neutrinos, given by $L = |\vec{x}| \approx t$.

1.1.4   Neutrino Interaction with Matter

As mentioned in the previous section, a theoretical model which involves massive neutrinos and flavour eigenstates that differ from mass eigenstates could explain neutrino oscillation in vacuum. If such neutrino mixing occurs and if neutrinos travel through matter, additional effects would affect oscillations. This phenomenon, named the Mikheyev-Smirnov-Wolfenstein (MSW) effect [23, 24], is due to the coherent forward scattering of the neutrinos with electrons and quarks as they travel through matter. While muon-type and tau-type neutrinos can only undergo coherent forward scattering through a neutral current reaction, where a Z boson is exchanged, electron-type neutrinos can additionally scatter with electrons through a charged current reaction via the exchange of a W boson. Since any weak interaction could cause a neutrino state to transform into one of the flavour eigenstates, neutrino oscillations are thus affected by propagation in matter. This occurs in an energy-dependent fashion, notably because of the energy-dependence of the coherent scattering cross-sections. Such an MSW effect would be significant for solar

neutrinos as they propagate through the very large electron densities in the Sun. Additionally, this could affect the neutrinos that are detected during the night for terrestrial experiments, due to their passage through the Earth.

1.1.5 Sterile Neutrinos

In principle, the Standard Model can accommodate an arbitrary number of neutrino flavours. However, the number of "active" neutrino eigenstates was constrained by many experiments. From the ALEPH, DELPHI, L3 and OPAL experiments, the number of light neutrino types that interact weakly, that is the number of neutrino types having a mass smaller than $m_Z/2$, has been measured to be 2.9840 ± 0.0082 [25]. However, there could potentially exist "sterile" neutrinos that do not interact weakly, but only through gravity, which would make them very difficult to detect directly. If these sterile neutrinos can mix with the other active neutrinos, it might be however possible to detect their presence by comparing a measurement of the energy spectrum and flux of active solar neutrinos to the SSM prediction for the electron-type neutrino production.

1.1.6 Neutrino Flavour Survival Probabilities

Through this thesis, the probability of an electron-type neutrino to be detected as an electron-type neutrino, $P_{e \to e}$, will be referred to as the "electron neutrino survival probability", or P_{ee}. Note that P_{ee} does depend on the energy of the initial neutrinos. The probability of an electron-type neutrino to be detected as one of the two remaining active flavours, ν_μ or ν_τ, will be referred to as P_{ea}. For a theoretical model that does not include

sterile neutrinos, this means that

$$P_{ee} + P_{ea} = 1. \tag{1.10}$$

For sterile neutrino models, the probability of an electron-type neutrino to be a sterile neutrino when reaching the detector is referred to as P_{es}, such that

$$P_{ee} + P_{ea} + P_{es} = 1. \tag{1.11}$$

1.2 The SNO Experiment

The Sudbury Neutrino Observatory (SNO) was specially designed to detect neutrinos with $E_\nu \geq 4$ MeV and thus to measure the flux of ^{8}B solar neutrinos produced through the ^{8}B solar reaction (c.f. Figure 1.1) and to study their possible oscillations. In 2001, the first published SNO results [26] presented the first clear evidence of solar neutrino oscillations for an undistorted solar energy spectrum. Subsequent publications provided increasingly more accurate measurements of the ^{8}B flux and of the solar neutrino energy spectrum. The next sections give an overview of the SNO experiment.

1.2.1 The SNO Detector

The SNO detector was located in Vale's Creighton mine near Sudbury, Ontario, Canada. The centre of the detector's spherical geometry being at a depth of 2092 m (or 5890 ± 94 m water equivalent), the rock overburden provided a very efficient shield against cosmic-ray

muons; the rate of these muons entering the detector was approximately three per minute, which is over six orders of magnitude lower than at sea level. As shown in Figure 1.2, the detector was composed of a 6 m radius transparent acrylic vessel (AV) that contained 10^6 kg of 99.92% isotopically pure heavy water (D_2O). The AV was located in a barrel-shaped cavern which was filled with over 7×10^6 kg of light water (H_2O) to shield the detector against the natural radioactivity of the rock. Surrounding the vessel was a 17.8 m diameter stainless steel geodesic structure (PSUP) which supported 9456 inward-facing 20 cm photomultiplier tubes that could detect Čerenkov radiation produced in the D_2O and H_2O.

The SNO PMTs, Hamamatsu model R1408, were made using Schott glass to minimise their radioactivity and had a time resolution of 1.5 ns. As depicted in Figure 1.3, the PMTs were housed in light-concentrating assemblies, which allowed to increase the effective light coverage to almost 55%. The quantum efficiency of the PMTs was 21.5% at 440 nm [27]. To help veto cosmic-ray muon events, 91 additional outward looking (OWL) PMTs were attached to the PSUP. The 1.7×10^6 kg of H_2O between the PSUP and the AV attenuated the radioactive backgrounds from the PSUP and the PMTs. H_2O and D_2O radioactivity was maintained at an extremely low level using advanced purification systems [28]. To study the response of the SNO detector during the whole duration of the experiment, many calibration sources (isotropic light source, ^{16}N, ^{8}Li, pT, ^{24}Na and Rn injection, ^{252}Cf, AmBe, ^{238}U and ^{232}Th) were used [29, 30, 31, 32].

In SNO, Čerenkov radiation resulting from neutrino interactions could be detected by the PMTs. The signal from a PMT could then trigger a discriminator which activated a time-to-amplitude converter (TAC) having a timeout of approximately 400 ns. The discriminators initiated the counting of firing PMTs during a 100 ns time window. The
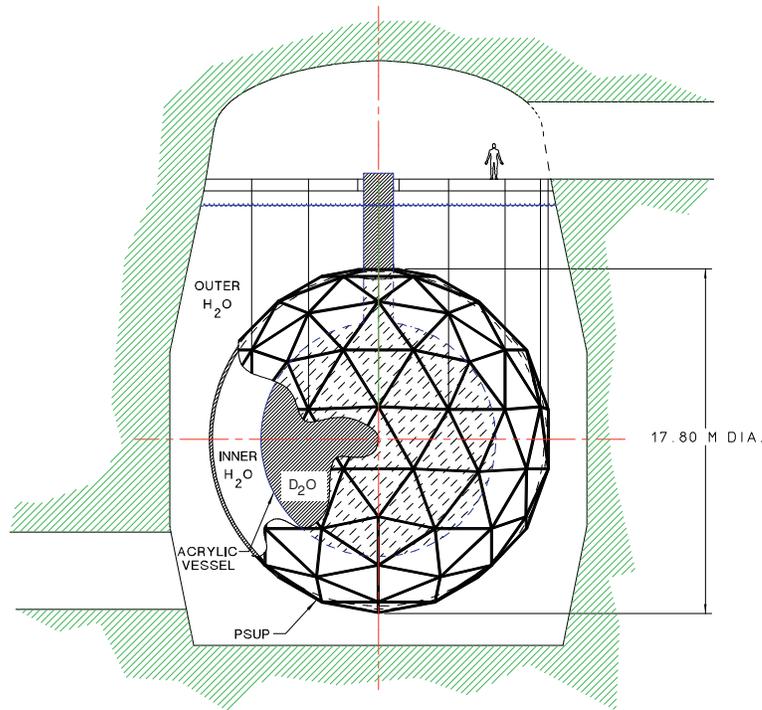
Figure 1.2: Schematic representation of the SNO detector, as presented in [34].

resulting SNO events were recorded when the number of fired PMTs within this time window exceeded some threshold value (normally around 17 PMT hits) which generated a global trigger (GT). For each event, three digitised charges were recorded for each fired PMT, along with a global trigger identification number (GTID) obtained using 10 MHz and 50 MHz clocks. The number of PMT hits versus the total electron energy in the detector was about $(-8.12 + 8.10$ MeV$^{-1}E_e)$ hits [33].
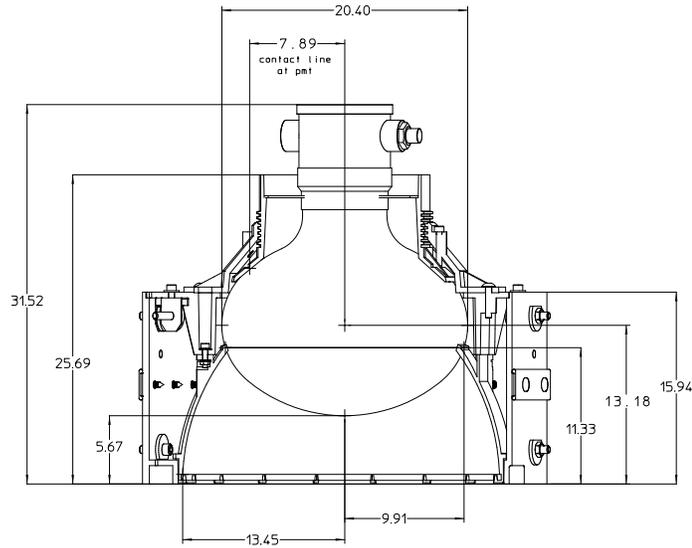
Figure 1.3: Drawing of a SNO PMT, mounted in its light-concentrator housing, as shown in [34]. Dimensions are in cm.

1.2.2   Types of Signal Interactions

One of the main characteristics that made the SNO experiment unique in its ability to confirm or repudiate the oscillation of solar neutrinos was its capacity to measure separately the flux of electron-type neutrinos and the total flux of active neutrinos. This was possible due to the usage of heavy water, as proposed by Herb Chen [35]. This medium allowed SNO to capture neutrinos through three different reactions, namely the charged current (CC), electron scattering (ES) and neutral current (NC) reactions, as presented in the following sections.

1.2.2.1   Charged Current

In SNO, the CC reaction is only sensitive to electron-type neutrinos. It consists of the t-channel interaction whose Feynman diagram is shown in Figure 1.4. In this interaction,
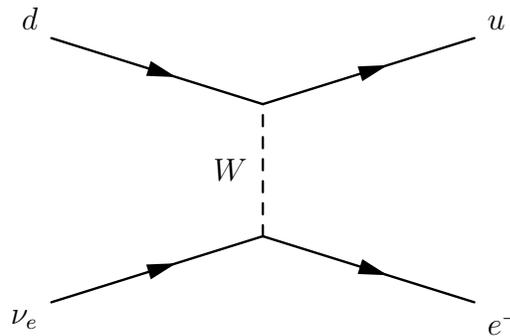
Figure 1.4: Feynman diagram of the charged current interaction. The reaction transforms one of the down quarks of the neutron from a deuteron atom into an up quark.

a deuteron is struck by an electron-type neutrino to produce an electron and a pair of protons. Since the outgoing electron carries most of the energy of the initial neutrino, the measurement of its energy allows one to probe the energy spectrum of the solar electron-type neutrinos. Also, for the CC signal, the direction of the produced electron is slightly correlated with the initial direction of the neutrino.

1.2.2.2 Electron Scattering

As indicated by its name, the electron scattering interaction occurs when a neutrino scatters with an electron in the detector. This interaction involves two different Feynman diagrams, as shown in Figure 1.5. Because of the diagram involving the exchange of a virtual Z boson, the ES interaction is sensitive to all neutrino flavours. Since both diagrams participate in the interaction for electron-type neutrinos, while only the left one is allowed for muon-type and tau-type neutrinos, the ES cross-section is considerably larger for ν_e. In fact, the average ES cross-section for ν_e is about 6.2 times larger than for ν_μ and ν_τ. Throughout this thesis, the ES interactions involving electron-type neutrinos will be referred to as ES_e interactions, while the one involving the other neutrino flavours will be referred to as $ES_{\mu\tau}$ interactions. Due to the kinematics of the ES interaction, the
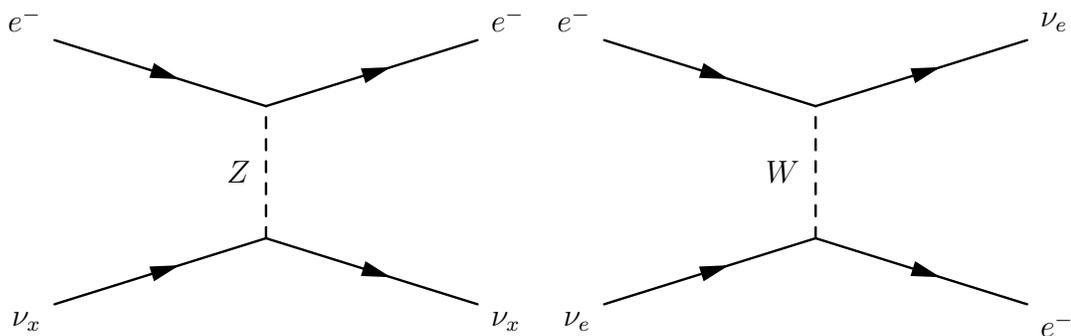
Figure 1.5: Feynman diagrams for the electron scattering interaction.

direction of the scattered electrons is essentially the same as the initial neutrino direction. Also, the initial neutrino energy is shared between the two outgoing particles, such that the average energy of the scattered electrons is lower compared to the ones produced by the CC interaction.

1.2.2.3   Neutral Current

The third signal interaction in the SNO experiment was the NC interaction. As shown in Figure 1.6, this interaction occurred in the detector when a neutrino scattered with a quark, which resulted in the breakage of a deuteron into its neutron and proton constituents. The NC interaction was unique to the SNO experiment and allowed the measurement of the total ^{8}B solar neutrino flux, because it has exactly the same cross-section for all neutrino flavours. The NC signal was detected by SNO after the outgoing neutron thermalised and was then captured by a nucleus. Because the energy ultimately measured by the SNO detector was thus characteristic of the exothermicity of the neutron capture interaction rather than the initial neutron energy, the detected energy from the NC interaction was uncorrelated to the initial neutrino energy. The next section provides more details regarding the processes involved in the capture of these neutrons.
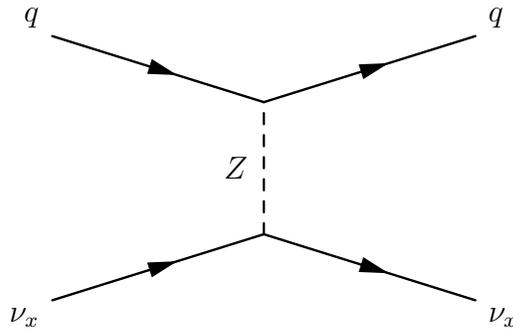
Figure 1.6: Feynman diagram of the neutral current interaction.

| Phase | Start Date | End Date | Live Time [Days] | |
			Day	Night
D_2O	11/1999	05/2001	119.9	157.4
salt	07/2001	08/2003	176.5	214.9
NCD	11/2004	11/2006	176.6	208.6

Table 1.1: The three phases of the SNO experiment, along with their respective day and night live times.

1.2.3   Experimental Phases

The SNO experiment was divided in three different phases, these phases being mainly differentiated by distinct methods of observing neutrons, through different detector configurations. Spanning from November 1999 to November 2006, the live times of these three phases, namely the D_2O, salt and NCD phases, are presented in Table 1.1. The three-phase analysis extracts information regarding the ^8B solar neutrinos collected during the three phases of the SNO experiment. The previous analysis [4] consisted of a combination of the results from the D_2O and salt phases using a lower energy threshold compared to earlier analyses. Since the three-phase analysis integrates this low-energy-threshold analysis (LETA), the D_2O and salt phases will be referred to as the LETA phases throughout this thesis, when appropriate. The following sections provide more information regarding the three phases of SNO.

1.2.3.1   D$_2$O Phase

In the first phase of the experiment, the detecting medium of SNO consisted only of pure D$_2$O. To observe the neutrons produced by the NC reaction, this phase depended on the capture of the neutrons by the deuterons of the D$_2$O molecules. This reaction releases a single 6.25 MeV gamma ray, which can create Compton electrons as well as electron-position pairs. In the SNO detector, these relativistic particles can create Čerenkov radiation, which can then be detected by the PMTs.

1.2.3.2   Salt Phase

For the second phase of the SNO experiment, 2×10^3 kg of NaCl were added to the D$_2$O. Since Cl nuclei have a neutron capture cross-section almost 10^5 times larger than D$_2$O nuclei, this allowed to increase the neutron detection efficiency significantly. In addition, the capture of neutrons on chlorine yields to the emission of multiple gamma rays, which allows to identify neutrons more easily due to a higher isotropy of this type of event compared to electron and single gamma events. The energy released during this reaction (8.6 MeV) is higher compared to the single photon energy of the D$_2$O phase. The combined effects of the improved neutron capture efficiency and the increase of the energy produced by these events allowed the detection efficiency for the NC events to be larger by a factor of 2.6 in the salt phase when compared to the D$_2$O phase.

1.2.3.3   NCD Phase

For the third phase of SNO, an array of 40 vertical neutral current detectors (NCDs) was deployed in the AV after removing the NaCl from the D_2O. These NCDs, whose geometry is pictured in Figure 1.7, were made using highly pure nickel tubes that were welded together to produce detector "strings". Since the NCDs had a lower average density than heavy water, they were fixed via Vectran braids to specially designed anchors at the bottom of the AV. Out of the 40 NCDs, 36 were filled with ^{3}He gas, which has a neutron cross-section about 120 times larger than chlorine. Such a large cross-section thus made the neutron capture efficiency by the NCDs much larger than that of the D_2O in the NCD phase. Compared to the D_2O phase, the capture efficiency of neutrons of the D_2O in the NCD phase was over five times smaller. In the NCD phase, neutrons were captured on ^{3}He through the reaction

$$^3\text{He} + n \rightarrow {}^3\text{H} + p, \tag{1.12}$$

where the produced triton and proton have kinetic energies of 191 keV and 573 keV, respectively. These charged particles then produced ionisation electrons while drifting toward the electrodes of the proportional counters. The voltage induced by the collected charges was logarithmically amplified, to increase the dynamic range of the signal. The NCD electronics allowed to record digitised integrated charges and digitised pulse shapes. This was accomplished using multiplexed analog-to-digital converters and oscilloscopes, respectively, with both systems generating independent triggers. The SNO master trigger card used the two NCD triggers as additional inputs to generate GTs, which allowed the

combination of the PMT and NCD data since the NCD system could access the GTID register to identify its events.

To study the backgrounds of the ^{3}He strings, four strings filled with ^{4}He, which were insensitive to neutrons, were also deployed in the detector. This allowed to differentiate alpha events from neutron events, using pulse shape discrimination. The NCDs were calibrated using ^{252}Cf and AmBe point sources [36], as well as a distributed ^{24}Na source [32]. More information regarding the NCDs can be found in [37].

For the first SNO results that included data from the NCDs [7], NCD event information was analysed along with the PMT data. A detailed description of the analysis leading to these results can be found in [38]. For the three-phase analysis, a pulse shape analysis (PSA) of the NCD data is performed independently, and the results from this analysis are then used to provide a constraint for the analysis of the PMT data. Section 2.2 provides some information regarding the PSA. More details about PSA can also be found in [39].

1.2.4   Backgrounds

Although great care was taken during the design, construction and runtime of SNO to minimise the sources of backgrounds that could affect the measurement of the flux of solar neutrinos, the nature of the solar neutrino interactions is such that backgrounds could not be avoided. Most of the SNO background events can be divided in two different types: "electron-like" and neutron events. Generally, the electron-like events consisted of radioactive decays to either beta particles or gamma rays whose reconstructed energies were sufficiently large to fall above the energy threshold of the analysis. The decays of ^{214}Bi (3.27 MeV β) and of ^{208}Tl (2.62 MeV γ) from the ^{238}U and ^{232}Th decay chains,

Readout cable

Cable endcap
with acrylic spacer

Pinch-off
fill tube

Counter body
(^{3}He-CF$_4$ gas mix)

Resistive coupler
(cable end only)

Fused-silica
insulator

Nickel endcap
body

Delay line
termination

Vectran braid

Anchor balls
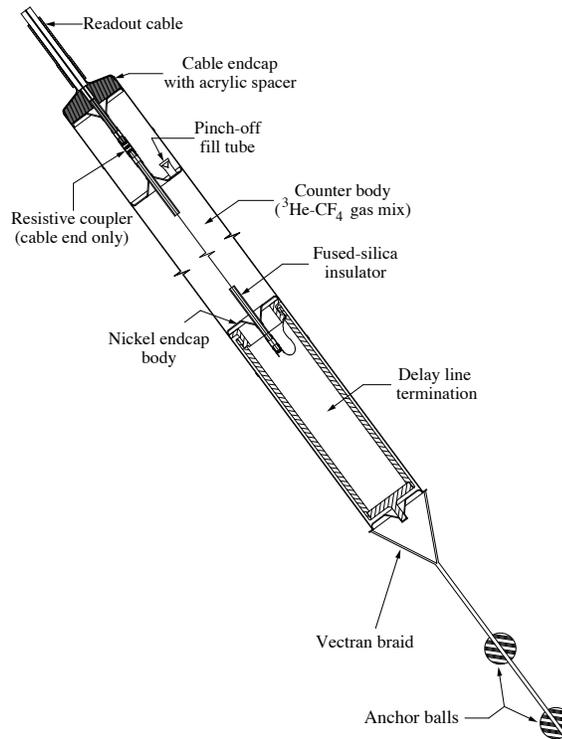
Figure 1.7: Schematic of a SNO neutral current detector, as shown in [34].

respectively, constitute major sources for such backgrounds. Neutron events could be created by the photodisintegration of deuterons by the aforementioned gamma decays having energies larger than 2.2 MeV. Here photodisintegration refers to the dissociation of the deuteron into a neutron and a proton by a photon. Also, neutrons could be created through (α, n) reactions in the AV. Notably, radon progeny that deposited on the AV surface during the construction of the detector could produce alpha particles that trigger such reactions. Because of their different event signature, the radioactive decays from Bi and TI traces in the different regions of the detector (D$_2$O, PSUP, AV and H$_2$O) were treated separately by LETA. The decay events originating in the PSUP region are named PMT β-γ in LETA. In the case of the NCD phase, the events in all regions of

the detector except the D_2O were regrouped in an event class called "external neutrons", while radioactive decays in the heavy water are referred to as "D_2O photodisintegrations". This grouping is partially motivated by the higher energy threshold for the analysis of the NCD phase data that practically eliminates any gamma event from the dataset, such that these backgrounds all appear as neutron events.

In the salt phase, the dissolved NaCl was at the origin of an additional source of gammas and neutrons, because of the neutron activation of ^{23}Na after the deployment of a neutron calibration source in the detector. The resulting ^{24}Na atoms eventually emitted some gammas with an energy of 2.75 MeV, which could photodisintegrate deuterons. In the NCD phase, decays from the traces of radioactive elements in the NCD strings and cables could also occur. "Hot spots" of radioactivity were discovered on two of the strings, so they were consequently included as distinct background sources in the analysis. This latter type of events is referred to as K2 and K5 photodisintegrations, after the names of the affected strings.

Other classes of minor background events are taken into account in the analysis. Among these are the hep solar neutrinos as well as the neutrinos produced by the decay of particles in the atmosphere. Other sources of neutrons, such as cosmogenic events and anti-neutrinos are also considered. Here, cosmogenic events refer to decay processes originating from isotopes that are created after cosmic rays interact with matter. Finally, a small number of instrumental events, referred to as "Leslie" events are included for LETA.

In order to either constrain or to estimate the rates of the different sources of backgrounds that are considered in the SNO ^{8}B analysis, numerous *ex-situ* and *in-situ* measurements were performed. More details regarding the evaluation and simulation of the different backgrounds can be found in [4, 38]. The nominal number of background events for the three SNO phases, as well as the external constraints on their rates can be found in Appendix D.

1.2.5 PMT Reconstructed Observables

Although neutrino flavours interact via different reactions in the SNO detector, as discussed in Section 1.2.3, most of these interactions are ultimately detected by the same PMTs. Hence, CC, ES and NC events cannot simply be counted to directly infer solar neutrino fluxes. Also, uncertainties on some of the background events are such that a simple subtraction of these classes of events from the dataset would translate to very large uncertainties for the measured signals. For these reasons, SNO relies on the statistical separation of the different signals and backgrounds to measure the quantities of interest. To perform such a "signal extraction", differentiating variables must be defined. This was done by identifying observable quantities whose distributions differed the most between the classes of events that composed the SNO dataset. This section presents the different PMT observables used for the three-phase analysis.

1.2.5.1 Reconstructed Kinetic Energy

The reconstructed kinetic energy (T_{eff}) is an observable of primary interest for the SNO experiment. As mentioned in Section 1.2.2, the kinetic energy distributions for the particles

produced by the different signal interactions differ for a given initial neutrino energy, due to the nature of these interactions. The energy distributions for both CC and ES depend on neutrino energy, but the mean for the CC distribution is larger, because electrons produced through this reaction carry most of the neutrino energy, while the energy is shared for ES events. In the case of NC, the detectable kinetic energy is characteristic of the neutron capture process and is thus uncorrelated to neutrino energy. Regarding the backgrounds, their energy spectra can be classified in two different groups. Neutron backgrounds, either directly produced by background processes, or resulting from the photodisintegration of deuterons by gamma rays having energies above 2.2 MeV, produced kinetic energy distributions which were identical to NC events after they thermalised and then captured. On the other hand, backgrounds that produced gamma rays and/or beta particles with energies above the analysis threshold had a signature that resembled more electron-like events in the detector, although the associated energy distributions had lower average energies compared to the SNO signals. Figure 1.8 presents an example of SNO observable probability density functions (PDFs) for the three signals and the two categories of background mentioned above. It also shows how the NC signal (and the neutron background) differs between the D_2O and salt phases, due to the different neutron capture interaction.

1.2.5.2 Normalised Cubed Radius

The normalised cubed radius is defined as

$$\rho = \left(\frac{R}{600 \text{ cm}}\right)^3 = \left(\frac{\sqrt{x^2 + y^2 + z^2}}{600 \text{ cm}}\right)^3, \tag{1.13}$$
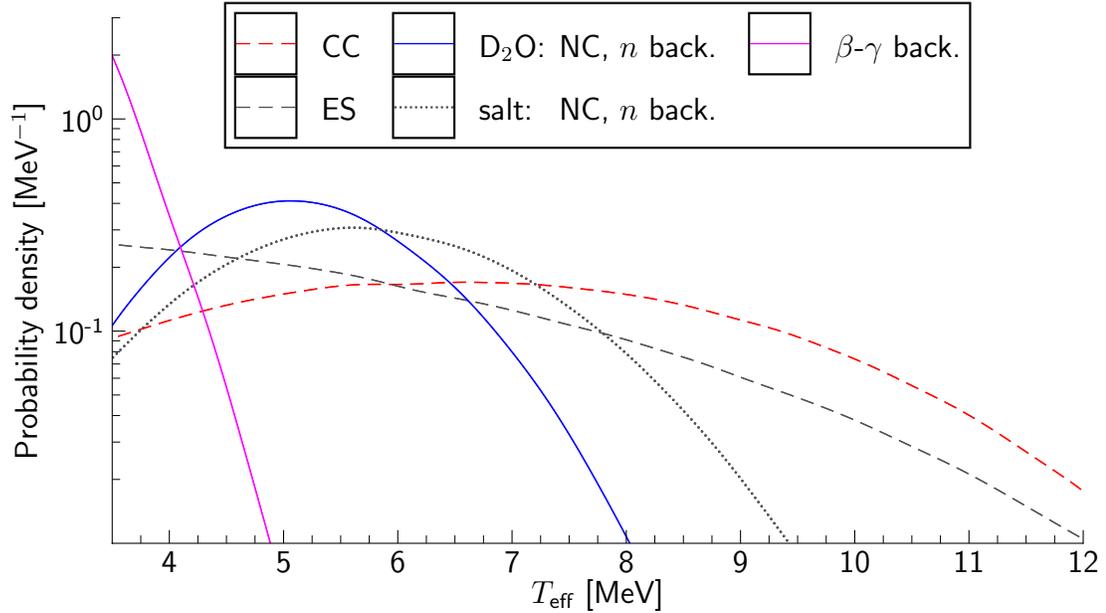
Figure 1.8: Example of reconstructed energy (T_{eff}) PDFs for the SNO signals and backgrounds.

where the z axis in the SNO detector is vertical and has its origin in the centre of the AV and where $\rho < 1$ defines the region inside the AV. The coordinates x, y and z refer to the position of the reconstructed neutrino interaction. This ρ variable is very important, since it was used to determine the effective volume of heavy water used for the analysis. Because of the natural radioactivity from the materials located outside the D_2O, including the H_2O, the PSUP, the PMTs and the AV, events whose reconstructed position is located within 50 cm of the AV are excluded from the analysis. The remaining region, called the "fiducial volume", is thus used to calculate the conversion factors between the incoming neutrino flux and the expected numbers of detected events. The knowledge of ρ is thus crucial in determining these factors.

Because external background events are exponentially attenuated as they propagate inside the detector, while the neutrino interactions and the internal β-γ background events
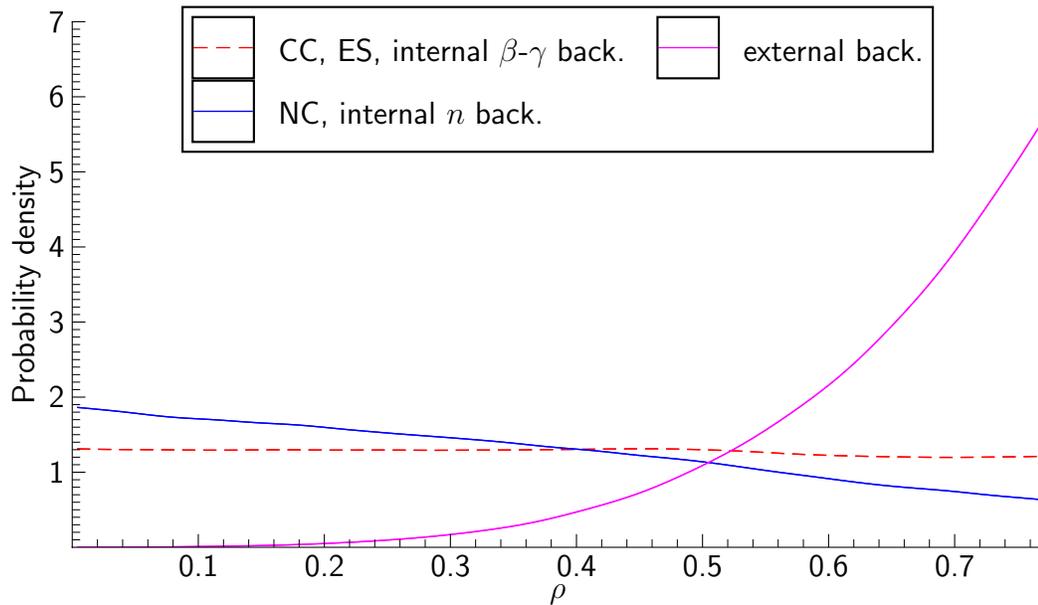
Figure 1.9: Example of normalised cubed radius (ρ) PDFs for the SNO signals and backgrounds.

are homogeneously distributed in the D$_2$O, PDFs of the ρ observable are very useful to statistically separate these external events, as shown in Figure 1.9. Finally, because the neutrons produced by the NC interaction and by the background processes can wander outside the heavy water before being captured, the ρ PDF for NC decreases with larger radii, such that it provides a way to differentiate these events from the CC events.

1.2.5.3   Direction

The reconstructed event direction with respect to the direction of neutrinos coming from the centre of the Sun is an observable which is used almost exclusively to separate ES interactions from the other event classes. Figure 1.10 presents PDFs of $\cos\theta_\odot$, the cosine of the angle between the direction of the reconstructed electron and the vector pointing
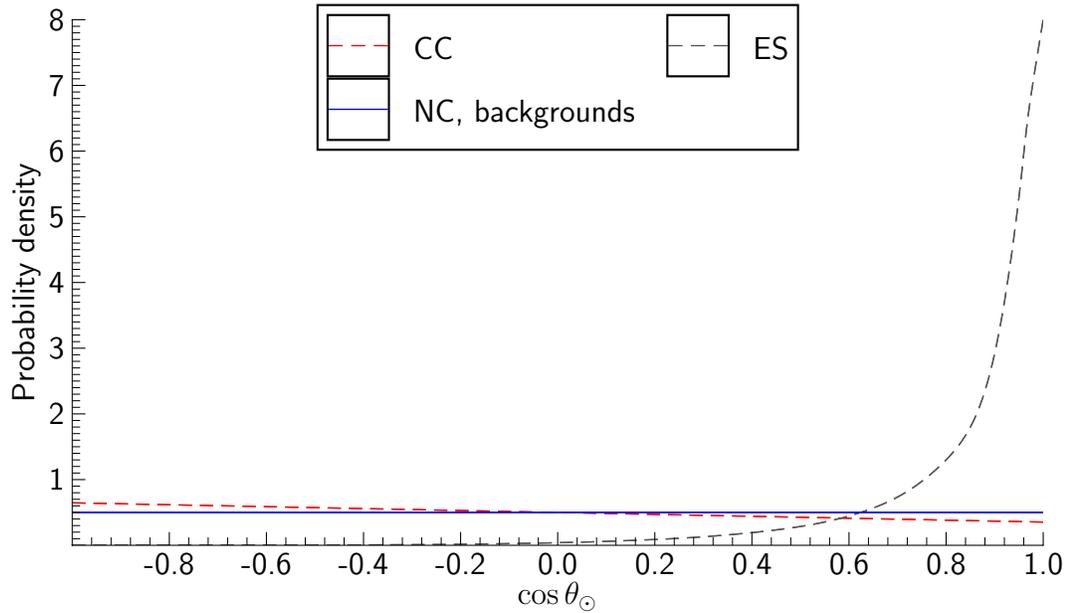
Figure 1.10: Example of $\cos\theta_\odot$ PDFs for the SNO signals backgrounds.

from the direction of the Sun. It shows how the PDF for ES is sharply forward-peaked, due to the kinematics of this reaction. By contrast, the NC PDF is completely flat, because the direction of the gamma rays emitted after the neutron capture is completely isotropic and hence uncorrelated to the initial neutrino direction. The CC interaction tends to slightly favour negative $\cos\theta_\odot$ values, although it is a mild effect. Finally, since all major SNO backgrounds are due to natural radioactivity, the direction of their events is completely uncorrelated to the direction of the Sun.
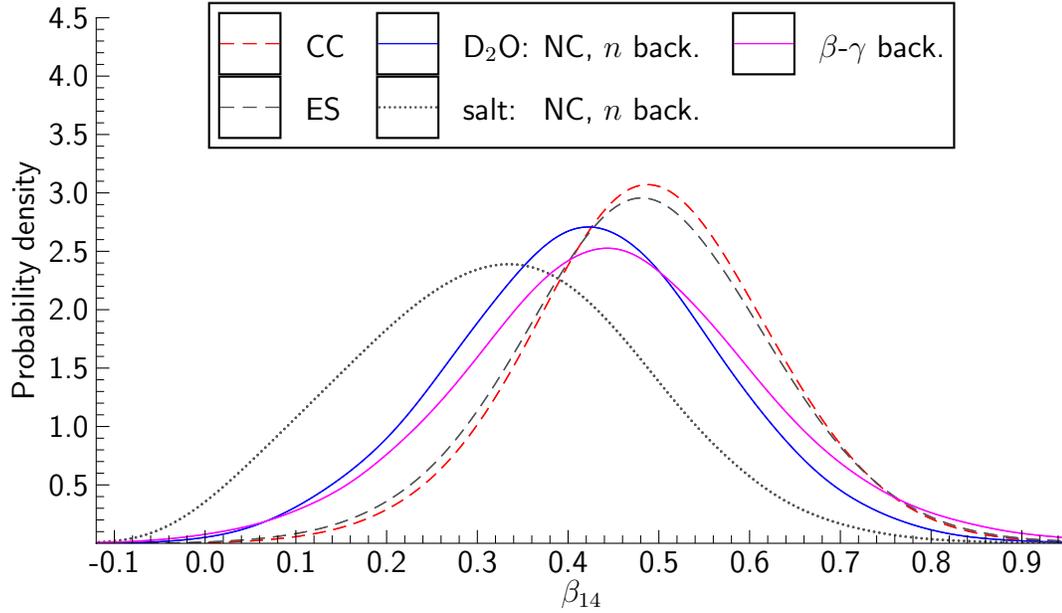
Figure 1.11: Example of β_{14} PDFs for the SNO signals and backgrounds.

1.2.5.4 Isotropy

The fourth observable used for SNO signal extraction is an indicator of event isotropy, called β_{14}. This variable is defined as

$$\beta_{14} \equiv \beta_1 + 4\beta_4, \tag{1.14}$$

where β_k is computed as

$$\beta_k \equiv \frac{2}{n_{\text{hits}}(n_{\text{hits}} - 1)} \sum_{i=1}^{n_{\text{hits}}-1} \sum_{j=i+1}^{n_{\text{hits}}} P_k(\cos\theta_{ij}). \tag{1.15}$$

In the above expression, n_{hits} is the number of PMT hits for a given event, θ_{ij} is the angle between PMTs i and j relative to the reconstructed event position and $P_k(x)$ is the Legendre polynomial of order k. The β_{14} variable was originally used for the initial

SNO results of the salt phase [2] to separate NC events from CC events. As mentioned in Section 1.2.3.2, the capture of neutrons on the chlorine atoms in the salt phase led to the emission of multiple gamma rays. These events thus tend to be more isotropic than Čerenkov events that originate from a single electron, such as the ones created by the CC interaction. This results in larger values for the θ_{ij} variable, on average, for the NC events when compared to the CC events. The choice of the first- and fourth-order Legendre polynomials for Equation (1.14) was determined empirically through Monte Carlo studies. Figure 1.11 shows how the NC β_{14} PDF differs between the D_2O and salt phases. The NC PDF for the D_2O phase is much more similar to the CC PDF, since neutron capture on a deuteron leads to the emission of a single gamma ray. Finally, the isotropy PDF for the β-γ backgrounds resembles a mixture of the PDFs for the NC events in the D_2O phase and for the CC events, because the β-γ events typically create a few energetic electrons in the detector.

1.3   Thesis Outline

As mentioned earlier, the SNO experiment was unique in its ability to measure the total flux of solar neutrinos, as well as the flux of electron neutrinos, through the usage of heavy water in which charged current and neutral current reactions occurred. Ultimately, in addition to the measurement of the ^{8}B flux, the variable of interest is the fraction of that flux that consists of electron-type neutrinos, since it represents the survival probability for ν_e. Due to the correlation between the neutrino energy and the reconstructed electron energy for CC and ES events, SNO also has a handle on the solar neutrino energy spectrum.

Because of the important overlap between the observable distributions for the three SNO interactions, as well as between the signal and background distributions, a statistical method must be used to discriminate between the different signals. The ability to accurately measure the neutrino fluxes thus depends on the differences between the corresponding observable distributions for the different signals and backgrounds, in addition to the existing statistical uncertainties due to the finite size of the dataset and the systematic uncertainties affecting the analysis.

Although the usage of the electron energy distributions can be useful to differentiate the SNO signals from some of the backgrounds, other observables must be used to differentiate the three signals in the context of an analysis that seeks to measure the energy spectra for CC and ES. Looking at the different observable distributions in the previous section, it is clear that it can be quite challenging to discriminate between CC and NC in the D_2O phase because of the similarity between the ρ, $\cos\theta_\odot$ and β_{14} distributions. The situation is also similar for the PMT data in the NCD phase because the signal observable distributions of that phase were very similar to those of the first phase. In the salt phase, however, the addition of the NaCl to the heavy water had a significant effect on the NC distributions. The higher degree of isotropy and larger reconstructed energy of this signal for the salt phase, due to the emission of multiple gamma rays following the capture of neutrons on chlorine atoms, allows a better differentiation between CC and NC β_{14} distributions. Also, because the NC signal is more energetic in the salt phase, the number of NC events falling in the analysis window is improved, which allows to reduce statistical uncertainties. The analysis window refers to the region of the event space which is defined by all analysis cuts. In the NCD phase, the pulse shape analysis also provides a quasi-independent measurement of the NC rate. The determination of the CC and ES

energy spectra in SNO is thus reliant on to the ability to measure NC. Once the NC signal is isolated, all phases contribute to measuring the energy spectra. The extraction of the SNO signals thus involves interplay between all the classes of events that are present in the dataset of all phases. In order to obtain the most from the SNO data while minimising biases, it is thus important to carefully define the statistical model used to represent the dataset.

In Chapter 2, the algorithms used to reconstruct SNO observables are described. Chapter 3 derives the general form of the likelihood function used by the three-phase analysis, while a detailed derivation of the terms of the function that are related to the SNO signals can be found in Chapter 4. The signal extraction procedure is described in Chapter 5 and results from ensemble testing of this procedure are shown and discussed in Chapter 6. Chapter 7 presents the results from the three-phase signal extraction and Chapter 8 summarises the analysis.

Chapter 2

Pattern Recognition

In the previous chapter, the experimental observables of SNO were presented, along with their distributions for the different classes of events that constitute the data sample under investigation. These distributions play a primary role in signal extraction, as their differences allow to statistically differentiate the different signals and also some of the backgrounds from the signals, as well as to separate the CC, ES and NC signal classes. The shapes of the observable distributions are determined by the nature of the processes generating the types of events, but also by the response of the detector to these events and by the reconstruction of the observables from the acquired data. At the analysis level, the first two are immutable, but one must achieve the best performance from reconstruction, notably to minimise the widening of the distributions caused by the reconstruction resolution which ultimately results in a loss of separating power between the event classes. This chapter thus describes the different algorithms used to perform event reconstruction via pattern recognition for the three phases of the SNO experiment. It is divided in two main sections that describe the event reconstruction for the PMTs and for the NCDs.

2.1 Pattern Recognition for the PMTs

2.1.1 Vertex Reconstruction

Vertex reconstruction represents the first stage of the event reconstruction process, as the reconstructed position and directions are used to determine the event energy. The knowledge of the event vertices is very important for different reasons. First of all, signal extraction is performed using events which are reconstructed within a spherical fiducial volume (having a radius of 550 cm) inside the acrylic vessel. The number of events found in the analysis is used directly to infer the flux of neutrinos coming from the Sun, such that a systematic uncertainty of 1% in the measurement of the radius translates into a 3% uncertainty in the measured flux. Also, many types of external background events have spatial distributions which increase drastically with radius, such that a good position resolution allows a better rejection of these events. Finally, as explained in the previous chapter, the capture of the neutrons produced by the NC reaction occur less uniformly than CC and ES events inside the fiducial volume, due to the higher mean free path of neutrons in heavy water. An accurate position measurement thus gives a better handle to separate NC from other signal events. The angular distribution of events with respect to the direction of the Sun is very important, mainly because the ES events are so forward-peaked (see Figure 1.10), compared to any other type of events. Also, the cross-section for CC events causes the resulting electron to be emitted backward with a slightly higher probability, differentiating them from NC and from all classes of background events that have a flat angular distribution.

2.1.1.1 D$_2$O and Salt Phases

For the D$_2$O and salt phases, two different algorithms were used to reconstruct event vertices: FTU, which is a fitter using PMT hit time to determine event position and direction in two stages, and FTP, a more advanced algorithm which performs both tasks simultaneously while considering different paths for the light generated by the events.

2.1.1.1.1 FTU

FTU is an all-purpose and very fast algorithm. To determine the position of an event in the SNO detector, FTU uses the hit times of all the PMTs involved in the event and maximises the likelihood of these hit times by varying the time and position of the event. To achieve this, some noisy PMT hits are first eliminated by computing the median hit time for the event and by eliminating the hits that are located outside a window of ± 50 ns around this time. The event position is then found by defining a time residual variable as a time-of-flight corrected time for each hit PMT,

$$t_{\text{res}} = t_{\text{PMT}} - t_{\text{fit}} - \frac{|\vec{x}_{\text{fit}} - \vec{x}_{\text{PMT}}|}{c_{\text{avg}}}, \tag{2.1}$$

where t_{PMT} is the PMT hit time, t_{fit} and $\vec{x}_{\text{fit}}$ are respectively the fitted time and position of the event, $\vec{x}_{\text{PMT}}$ the position of the PMT and c_{avg} the average group velocity of the light in the detector (approximately 2.19×10^8 m/s), and then by maximising the likelihood function (see Section 3.1 for an introduction to likelihood functions)

$$L = \prod_{i=1}^{n_{\text{hits}}} f(t_{\text{res}}(\vec{x}_{\text{PMT}_i}, t_{\text{PMT}_i}, \vec{x}_{\text{fit}}, t_{\text{fit}})|\text{hit } i), \tag{2.2}$$
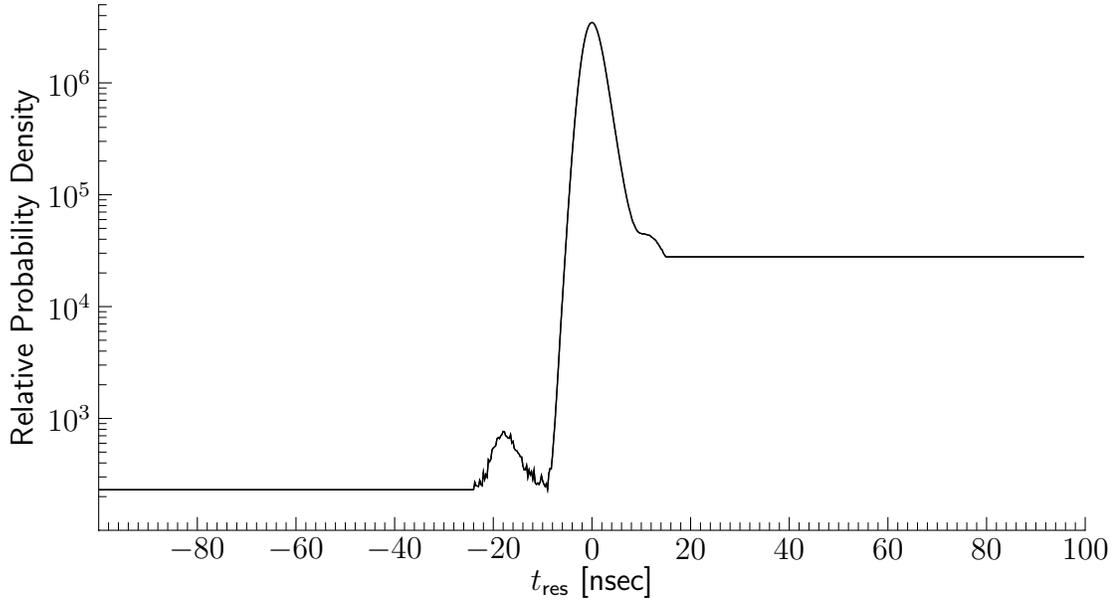
Figure 2.1: Time residual probability density function used by FTU to reconstruct event position, in relative units.

where $f(t_\text{res}(\vec{x}_{\text{PMT}_i}, t_{\text{PMT}_i}, \vec{x}_\text{fit}, t_\text{fit})|\text{hit } i)$ is the probability density that a firing PMT i fire at time t_res. The profile of the probability density $f(t_\text{res}|\text{hit})$ is shown in Figure 2.1, in relative units. It was produced using both Monte Carlo simulations and optical calibration data. The figure shows PMT pre-pulsing, followed by the main peak, whose width is defined by the PMT time resolution (approximately 1.5 ns), then by PMT after-pulsing. Flat region probability densities are determined by the PMT white noise and late light.

FTU uses the Levenberg-Marquardt [40, 41] algorithm iteratively to minimise the likelihood function of Equation (2.2). For each iteration, a random position is drawn uniformly from a sphere having a radius of 10 m and a random time is drawn uniformly from a window of ± 50 ns around the median PMT hit time. The likelihood is then minimised with the Levenberg-Marquardt algorithm using these spatial and temporal coordinates as the starting point, and the values of the fitted position and likelihood function are recorded.

The process is then repeated until either the same position or a higher likelihood function minimum is returned for six consecutive iterations, after which the global minimum is assumed to have been found.

As mentioned previously, FTU estimates the event direction in a second stage. To achieve this task, it minimises the likelihood function

$$
\begin{aligned}
L &= \prod_{i=1}^{n_{\text{sel. hits}}} f(\theta_i(\vec{x}_{\text{fit}}, \hat{u}_{\text{fit}}), \text{hit } i | \vec{x}_{\text{fit}}) = \prod_{i=1}^{n_{\text{sel. hits}}} f(\theta_i | \vec{x}_{\text{fit}}, \text{hit } i) P(\text{hit } i | \vec{x}_{\text{fit}}) \\
&\propto \prod_{i=1}^{n_{\text{sel. hits}}} f(\theta_i) \Omega_i(\vec{x}_{\text{fit}}) = \prod_{i=1}^{n_{\text{sel. hits}}} f(\theta_i) \frac{\cos \phi_i(\vec{x}_{\text{fit}})}{(\vec{x}_{\text{PMT}_i} - \vec{x}_{\text{fit}})^2},
\end{aligned}
\tag{2.3}
$$

where $\theta_i(\vec{x}_{\text{fit}}, \hat{u}_{\text{fit}})$ is the angle between the direction of the event, $\hat{u}_{\text{fit}}$, and the direction of the PMT i with respect to the fitted event position. $P(\text{hit } i | \vec{x}_{\text{fit}})$ is the probability of a photon to hit PMT i, given the event position $\vec{x}_{\text{fit}}$ but no directional information, which is proportional to the solid angle effect $\Omega_i(\vec{x}_{\text{fit}})$ given by $\frac{\cos \phi_i(\vec{x}_{\text{fit}})}{(\vec{x}_{\text{PMT}_i} - \vec{x}_{\text{fit}})^2}$, where $\phi_i(\vec{x}_{\text{fit}})$ is the angle of incidence of the light on the PMT i. The likelihood function is computed using the events selected with a time residual window of ± 10 ns, which avoids complications resulting from scattered light. The probability density function $f(\theta)$ has been determined from Monte Carlo simulations and is shown in Figure 2.2. It peaks at $41.6°$, which corresponds to the Čerenkov angle in D_2O for 8 MeV electrons.

2.1.1.1.2   FTP

Although time fitters like FTU can provide satisfying results, the separation of the position and direction reconstruction in two separate processes has some disadvantages. Due to the nature of Čerenkov light, which results in an azimuthal symmetry of the PMT hits around
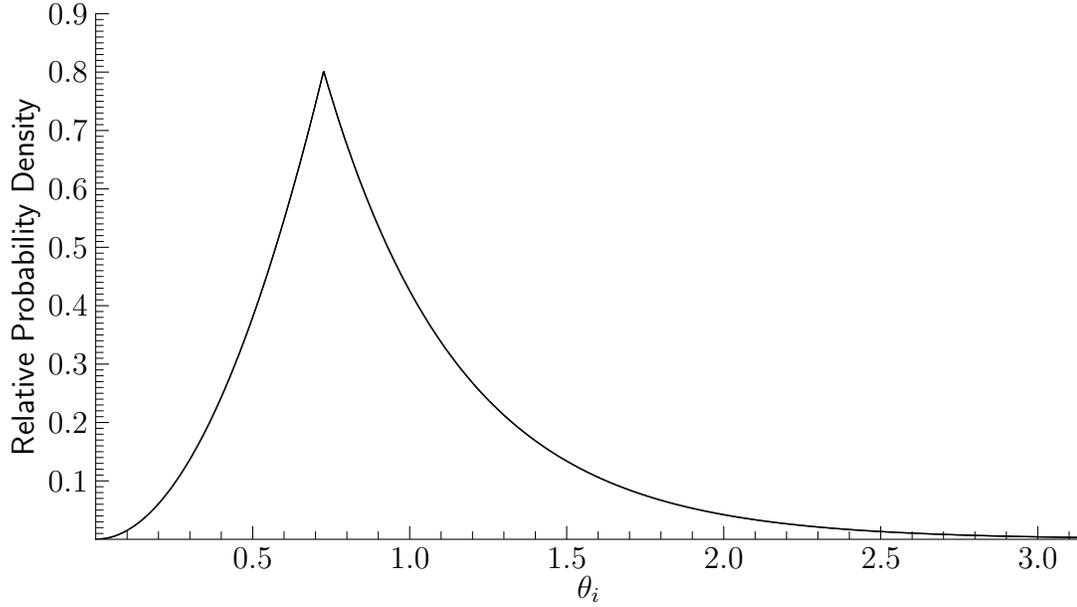
Figure 2.2: Angular probability density function $f(\theta)$ used by FTU to reconstruct event direction, in relative units.

the direction of the event, to ignore angular information while fitting event position can lead to larger reconstruction uncertainties since it is possible to move events forward and backward along their direction without causing large fluctuations of likelihood values. To help reduce this undesirable effect, the Path Fitter (FTP) determines the event position and direction simultaneously by minimising the likelihood function

$$
\begin{aligned}
L &= \prod_{i=1}^{n_{\text{sel. hits}}} f(\theta_i(\vec{x}_{\text{fit}}, \hat{u}_{\text{fit}}), t_{\text{res}}(\vec{x}_{\text{PMT}_i}, t_{\text{PMT}_i}, \vec{x}_{\text{fit}}, t_{\text{fit}}), \text{hit } i) \\
&= \prod_{i=1}^{n_{\text{sel. hits}}} [f(\theta_i, t_{\text{res}}, \text{hit } i, \text{direct}) + f(\theta_i, t_{\text{res}}, \text{hit } i, \text{other})] \\
&= \prod_{i=1}^{n_{\text{sel. hits}}} [P(\text{direct})f(\theta_i, t_{\text{res}}, \text{hit } i|\text{direct}) + P(\text{other})f(\theta_i, t_{\text{res}}, \text{hit } i|\text{other})], \quad (2.4)
\end{aligned}
$$

where $P(\text{direct})$ and $P(\text{other})$ are the probabilities for direct and other light (reflections, scattering, PMT noise), respectively, whose values are estimated to be $P(\text{direct}) = 0.879$ and $P(\text{other}) = 0.121$ using an isotropic light source, and where the other variables are defined as in the previous section. The probability density functions $f(\theta_i, t_{\text{res}}, \text{hit } i|\text{direct})$ and $f(\theta_i, t_{\text{res}}, \text{hit } i|\text{other})$ are defined using the products of the time residual and angular PDFs. The time residual PDFs are found using an isotropic light source. The direct component contains PMT pre-pulsing, the prompt peak, late-pulsing and after-pulsing, while the PDF for other light is a step function around the prompt peak. The direct angular PDF has been determined from Monte Carlo simulation similarly to what was done for FTU. The corresponding angular PDF for other light is simply a uniform function.

Although the inclusion of the angular information while determining event position was *a priori* expected to lead to less uncertainty of the reconstructed position along the direction of the event, it was observed that it tends to create a systematic shift forward along that same direction, due to late scattered light. A "drive" correction is thus applied to the fit position determined by the procedure described above and the event direction is then fitted again.

The reconstruction technique used by FTP significantly improves vertex resolution compared to the FTU fitter, as seen in Table 2.1. FTP is however much more computationally demanding (it is about 8 times slower than FTU), which limits its usage with high statistics datasets. More information about FTP can be found in [42].

| Phase | Vertex Resolution (cm) | |
	FTU	FTP
D_2O	22.7	19.3
Salt	23.3	20.0

Table 2.1: Average FTP and FTU vertex resolutions for D_2O and salt phases measured using reconstruction of ^{16}N calibration data events.

2.1.1.2 NCD Phase

For the NCD phase, FTN, a single multi-purpose algorithm developed by the author of this thesis was selected to perform vertex reconstruction. Having to be used for the reconstruction of all Monte Carlo simulations and calibration data in addition to neutrino runs, such an algorithm was required to be very fast. An approach based on the FTU fitter has been chosen, which performs position and direction reconstruction in two separate steps while addressing the challenges created by the addition of the NCD array in the detector. This section explains the functioning of FTN and presents comparisons of its performance with the previous algorithms.

2.1.1.2.1 Time Residual PDFs

As mentioned in Section 2.1.1.1.1, with the original FTU fitter, vertex reconstruction is performed by maximising the likelihood function

$$L = \prod_{i=1}^{n_{\text{hits}}} f(t_{\text{res}}(\vec{x}_{\text{PMT}_i}, t_{\text{PMT}_i}, \vec{x}_{\text{fit}}, t_{\text{fit}})|\text{hit } i), \tag{2.2}$$

where $f(t_{\text{res}}|\text{hit})$ is a constant PDF for all PMTs. For the NCD phase, this fixed PDF may lead to poor fit results because some PMTs are partially or completely shadowed by
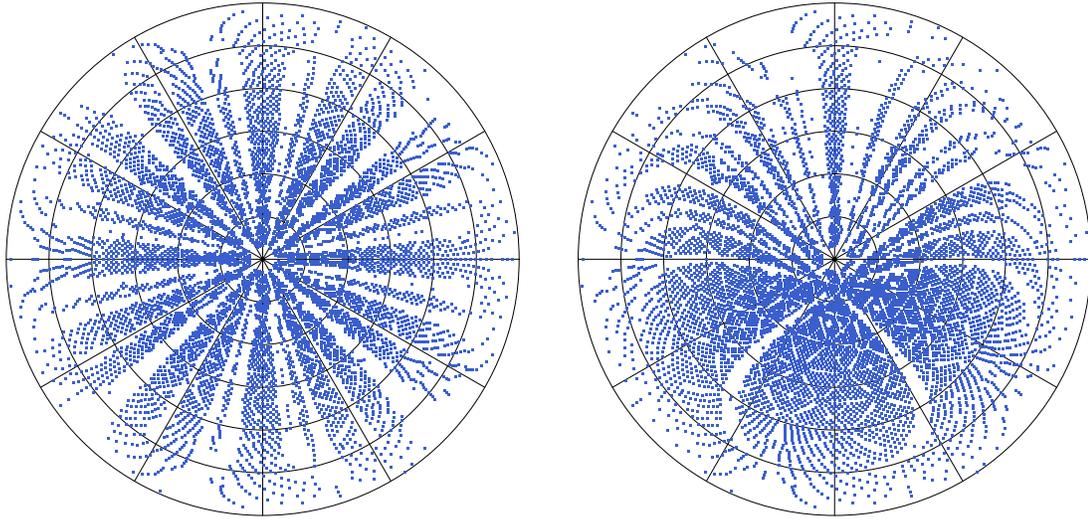
Figure 2.3: Non-shadowed PMTs for a point light source located at the centre of the detector (left) and 300 cm off-centre in the (x, y) plane (right) as seen from the top of the detector.

one or many NCD strings. The shadow affects a large fraction of the PMTs, as seen in Figure 2.3. This results in a high dependency of the time residual PDF on event position:

$$f(t_{\text{res}}|\text{hit}) \overset{\text{NCD}}{\to} f(t_{\text{res}}|\text{hit}, \vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}}) \tag{2.5}$$

To break this dependency, the time residual distribution can be separated into different components that do not depend directly on PMT and light source positions. If the following types of light paths are defined:

$$d : \text{source} \to \text{PMT}$$

$$s : \text{source} \to \text{shadowing NCD} \to \dots \to \text{PMT},$$

$f(t_{\text{res}}|\text{hit}, \vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}})$ can be expressed as

$f(t_{\text{res}}|\text{hit}, \vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}})$

$$= f(t_{\text{res}}, d|\text{hit}, \vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}}) + f(t_{\text{res}}, s|\text{hit}, \vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}})$$

$$\approx \left[1 - P(s|\vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}})\right] f(t_{\text{res}}|\text{hit}, d) + P(s|\vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}}) f(t_{\text{res}}|\text{hit}, s)$$

$$= f(t_{\text{res}}|\text{hit}, d) - P(s|\vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}}) \left[f(t_{\text{res}}|\text{hit}, d) - f(t_{\text{res}}|\text{hit}, s)\right], \qquad (2.6)$$

where $P(s|\vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}})$ is the probability that a photon leaving from a source at position $\vec{x}_{\text{fit}}$ in the direction of a PMT at position $\vec{x}_{\text{PMT}}$ hits a shadowing NCD. $f(t_{\text{res}}|\text{hit}, d)$ is the time residual PDF for non-shadowed PMTs and $f(t_{\text{res}}|\text{hit}, s)$ is the corresponding PDF for completely shadowed PMTs. In the above equation, the effects of NCD reflections have been neglected. NCD reflections are estimated to have a magnitude of about 0.04% relative to direct light intensity for a source at the centre of the detector. Most acrylic vessel and PMT reflections are removed by using the same prompt time cut as FTU.

The previous expression describes a low-energy photon point source. For real events in the SNO detector, photons are usually emitted along a track. Since FTU fits event position and direction in two stages, the fitter does not have information about track direction when fitting for position. To take this effect into account, Equation (2.6) has to be modified to allow for the associated smearing. The actual implementation of FTN accomplishes this by redefining $f(t_{\text{res}}|\text{hit}, d)$ and $f(t_{\text{res}}|\text{hit}, s)$, and by using a mapping function between the fraction of shadowing computed analytically and an effective fraction of shadowing described by the probability function $P'(s|\vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}})$. The effective time

residual PDF is thus approximated by the expression:

$$f(t_\mathsf{res}|\mathsf{hit}|\vec{x}_\mathsf{PMT}, \vec{x}_\mathsf{fit})$$

$$\approx f(t_\mathsf{res}|\mathsf{hit}, d') - P'(s|\vec{x}_\mathsf{PMT}, \vec{x}_\mathsf{fit}) \left[f(t_\mathsf{res}|\mathsf{hit}, d') - f(t_\mathsf{res}|\mathsf{hit}, s') \right] \quad (2.7)$$

To generate the two time residual PDFs, 10^7 6 MeV and 10^7 10 MeV electrons have been uniformly and isotropically generated in the detector. For each event, the shadowing of all hit PMTs has been computed analytically, using the position of the generated electron as the source position. The firing time of completely shadowed PMTs has then been used to generate $f(t_\mathsf{res}|\mathsf{hit}, s')$ and the time of non-shadowed PMTs to produce $f(t_\mathsf{res}|\mathsf{hit}, d')$, using a weight of 70% for 6 MeV electrons and a weight of 30% for 10 MeV electrons, for a better representation of the energies for solar neutrino data. The resulting PDFs are shown in Figure 2.4. The probability density for late light has been set to a constant as done with the original FTU fitter. The mapping function is generated by computing the height of the prompt peak for a set of time residual PDFs generated using different ranges of shadowing. The relation between the effective and the computed NCD shadows can be seen in Figure 2.5.

2.1.1.2.2   Shadow Computation

To compute PMT shadowing as described in the previous section, FTN relies on a set of algorithms. These functions are used to compute the fraction of a PMT solid angle that is shadowed by the NCD strings, as seen from a point light source located at a given position in the detector. To achieve this, a given hit PMT is tested for NCD shadow by first discarding most non-shadowing strings using a fast 2D algorithm. If some strings
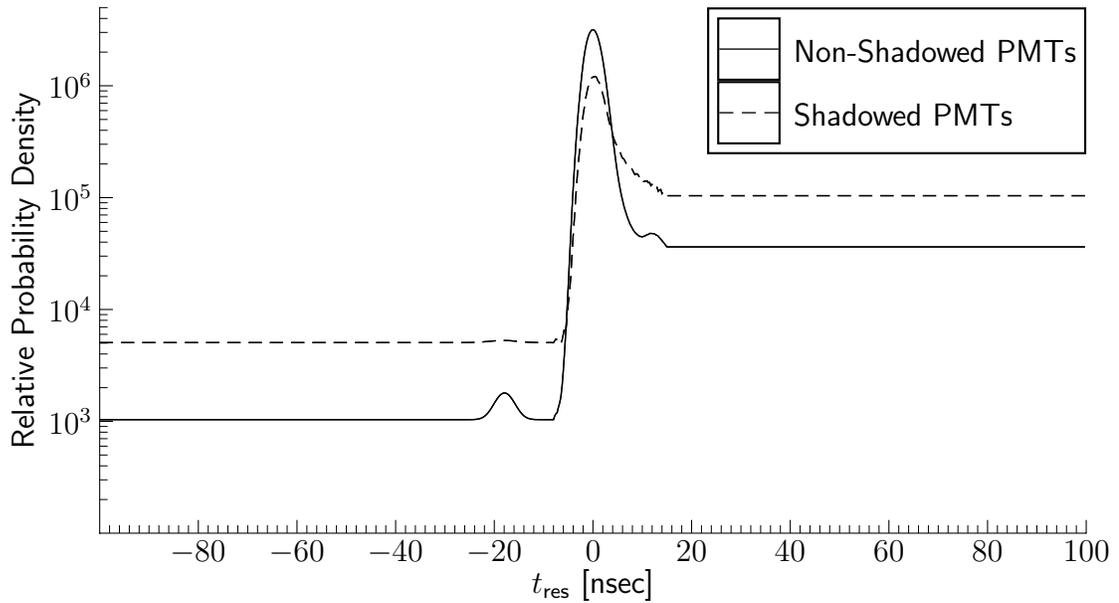
Figure 2.4: Time residual probability density functions for non-shadowed (continuous line) and completely shadowed (dashed line) PMTs used by FTN to reconstruct event position, in relative units.
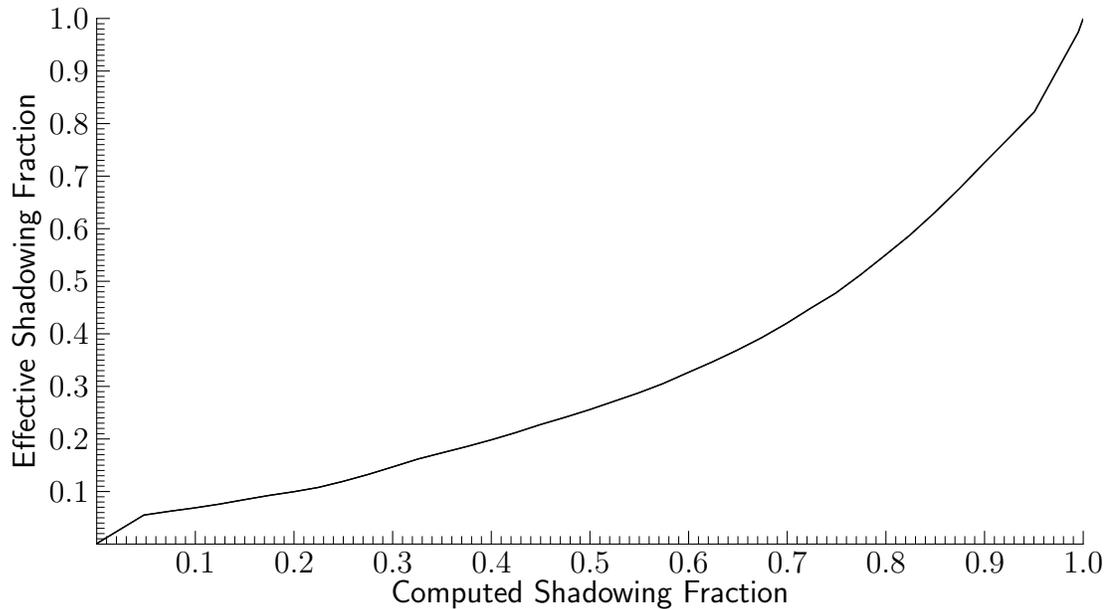
Figure 2.5: Mapping function between computed and effective fraction of NCD shadow used by FTN.
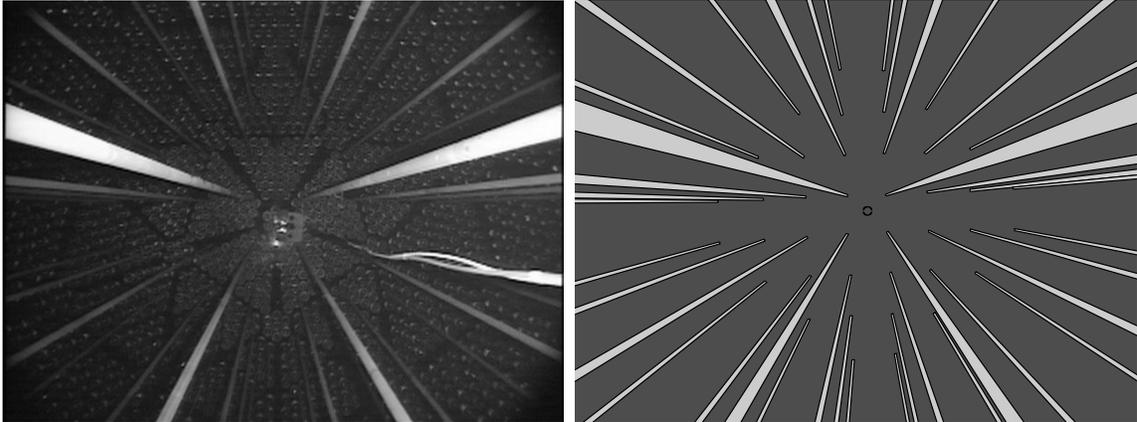
Figure 2.6: Picture of the completed NCD array taken from the neck area during strings deployment (left) and representation of the NCD strings and of a single PMT as seen by FTN (right). Position of observation is approximated.

remain after this step, the centre point of the PMT and three points around the reflector are projected on the inner surface of the acrylic vessel by fitting the refracted path of the light from the source to the PMT. These points are then projected in perspective on a plane whose normal direction is a straight line linking the source to the projected PMT centre, and an ellipse is fitted to the three other points. The sides of the non-discarded NCD strings are projected on the same plane. The resulting quadrilaterals that cross the PMT ellipse identify the shadowing NCD strings. Overlapping quadrilaterals are merged and the area of the intersection between these "virtual" strings and the PMT ellipse is computed analytically. The resulting value is finally divided by the total PMT ellipse area to get the fraction of shadow. Figure 2.6 compares a view of the SNO detector as seen by the algorithm that computes NCD shadow to an actual picture of the detector taken during string deployment.

2.1.1.2.3 Fitting Procedure

The fitting procedure used by FTN is very similar to the one used by FTU, which is described in Section 2.1.1.1.1. The differences reside in some CPU time optimisation algorithms and special treatment of the events that are reconstructed inside the NCD array region. Indeed, the value of $P'(s|\vec{x}_{\text{PMT}}, \vec{x}_{\text{fit}})$ that is computed by FTN for events located inside the NCD array cannot be very reliable, since in this region the electron track length is of the same order of magnitude as the distance between the initial position of the electron and the surrounding NCD strings. The effective shadowing of PMTs is thus strongly correlated to the direction and energy of the event and to the distribution of the generated photons along the track. Discrete position shifts around the NCD strings have been observed when using shadow information to fit events located in the NCD array. FTN avoids these systematic position shifts by not using the shadow information for these events, that is to say by using $f(t_{\text{res}}|\text{hit}) = f(t_{\text{res}}|\text{hit}, d)$ when the event is reconstructed inside the NCD array ($|x| + |y| < 450$ cm). To speed up the fitting procedure, FTN starts by fitting all events using $f(t_{\text{res}}|\text{hit}, d)$ as the time residual PDF. As done with the original FTU fitter, this algorithm is called many times using a different seed vertex as the starting position until it converges to the same minimum six times. Then FTN repeats this procedure, using shadow information and starting seed vertices that are located at a maximum distance of 50 cm from that position. During this second pass, PMT shadowing is not computed when the reconstructed position is more than 800 cm from the centre of the detector. Finally, event direction is reconstructed using the same algorithm as FTU. This is done to avoid unnecessary complexification of the model expressed by Equation (2.3), which would not lead to significant improvements,

due to the point source approximation used by the PMT shadowing algorithm and to the two-stage nature of the FTN vertex reconstruction process.

2.1.1.2.4   Fitting Time

To minimise the CPU cycles needed by vertex reconstruction, many efforts have been directed at optimising the FTN code. Several flags have also been implemented to allow the selection of different algorithms depending on the user's priority (accuracy versus CPU time). One of the most time-consuming tasks is the computation of light refraction through the AV, which can be turned off. The effects of AV refraction on FTN reconstruction resolution seem to be very small compared to the improvements from other parameters and are not worth the extra CPU time. The average time needed to fit an event has been measured using a set of 2500 5.5 MeV electrons generated uniformly within the acrylic vessel. Without AV refraction activated, FTN took 41% less time to reconstruct the events. Some runs have been also reconstructed with FTU, FTN and FTP for reconstruction time comparison. Results can be seen in Table 2.2. It shows that FTN algorithm is only 1.2 to 2 times slower than FTU while being 3.4 to 7.1 times faster than FTP. These results have been generated using three types of event simulations. The first line presents the reconstruction time of simulated ^{16}N events, which produce photons having energies between 5 MeV and 7 MeV. The second line shows results when simulating a ^{24}Na source, which produces both electron-like and neutron-like events. Finally, results from simulated solar neutrino events are also presented.

Run	Average Fitting Time [s/event]		
	FTU	FTN	FTP
63890 (^{16}N)	0.032	0.053	0.18
55781 (^{24}Na)	0.014	0.017	0.12
63636 (ν)	0.046	0.092	0.48

Table 2.2: Average time required to fit single events for given ^{16}N, ^{24}Na and ν runs reconstructed with FTU, FTN and FTP.

2.1.1.2.5   Position Shift

To evaluate the differences between the true and reconstructed event positions, or "position shift", the average reconstructed event position of ^{16}N calibration data relative to the source position was compared to that computed from simulated data. Figure 2.7 shows the difference between the data and Monte Carlo position shift as a function of the source position for scans along the main axes of the detector. It shows a spread of about 4 cm in the three directions, and this value was taken as the vertex shift uncertainty. This uncertainty was found to be correlated to the PMT timing calibration of the detector. An overall offset of 5 cm was also observed in the z direction.

2.1.1.2.6   Position Scale

The systematic uncertainty on "position scale", a position-dependent inward or outward shift of reconstructed position, can have a direct effect on the fiducial volume for events detected by the PMT array. It was measured by determining the range of the slope of a first-order polynomial that allowed the inclusion of 68% of the data points in Figure 2.7. This uncertainty was believed to be caused by physical factors such as a mismatch of the
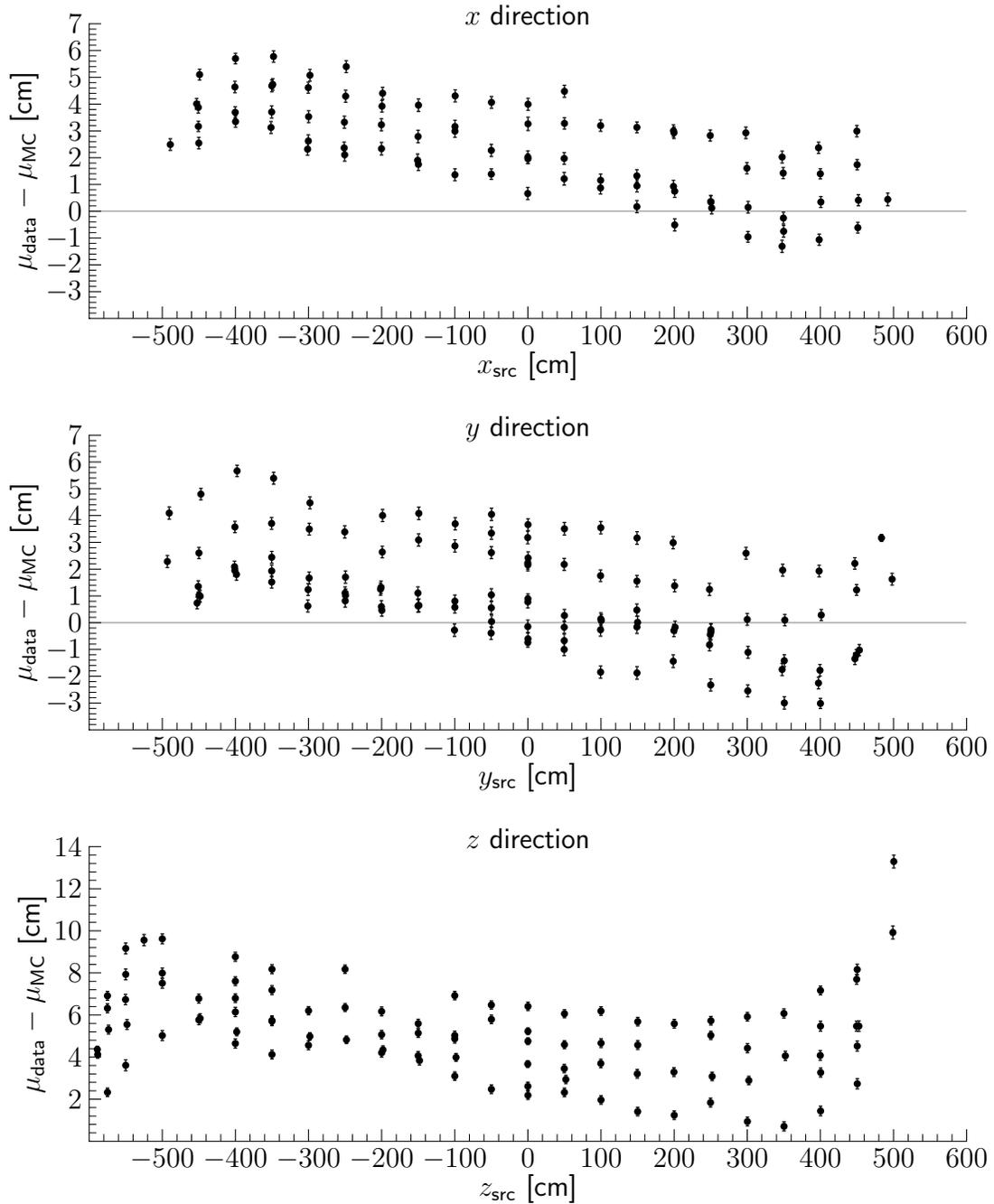
Figure 2.7: Difference between data and MC position shifts in the three axis directions as a function of source position along the main axes of the detector. The top, middle and bottom panels represent the shifts in x, y and z coordinates, respectively. For each position, there are several points corresponding to different ^{16}N calibration scans taken over different time periods.

speed of light in different media and was expected to be the same in all directions. It was estimated to be 0.9% of the Cartesian coordinates.

2.1.1.2.7   Position Resolution

To evaluate the position resolution of FTN, the function

$$R(x) = \frac{1-\alpha}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} + \frac{\alpha}{2\tau} e^{\frac{-|x-\mu|}{\tau}} \tag{2.8}$$

has been used. Equation (2.8) will be referred to as the four-parameter vertex position resolution function. The use of an exponential component for the resolution function allows for the addition of tails to the Gaussian function [43]. When position reconstruction performance is evaluated for an analysis that involves the use of a calibration source, the distribution of reconstructed position for the relativistic electrons that are ultimately produced by that source is described by the convolution of the fitter resolution function with the distribution of the true position of the relativistic electrons. This is notably the case with the ^{16}N source. The theoretical distribution of the true position is presented in Appendix A.2.

Volume-weighted position resolution has been obtained by generating isotropically and uniformly 10^5 mono-energetic electrons within the acrylic vessel. Both a simple Gaussian and the four-parameter resolution functions have been fitted to the distribution of reconstructed event position (using a range of ± 80 cm). A total of 33 simulations have been performed with different electron energies. Figure 2.8 shows the values of the fitted width parameter of a simple Gaussian and the computed RMS as a function of energy. Figure 2.9 shows the values of fitted parameters α, σ and τ for the four-parameter

Direction	a_0 [cm]	a_1 $[\times 10^{-2}]$	a_2 $[\times 10^{-5}$ cm$^{-1}]$
x, y	1.19 ± 0.52	-0.10 ± 0.11	0.71 ± 0.36
z	1.29 ± 0.51	0.21 ± 0.15	

Table 2.3: Fitted values for the parameters in the polynomial used to evaluate the systematic uncertainty on vertex resolution. The parameters a_0, a_1 and a_2 are the zeroth, first and second order coefficients of the polynomial, respectively.

resolution function (fitted means are not shown) and the computed RMS as a function of energy. Equivalent results for the FTU fitter are also shown.

Position resolution uncertainty can also affect the fiducial volume. It was assessed by taking the difference between the data and the MC fitted position resolution for all ^{16}N calibration data taken inside the acrylic vessel. Figure 2.10 shows that this uncertainty varies significantly as a function of the z position of the source. Some smaller fluctuations were also observed for ^{16}N calibration scans in the (x, y) plane and were incorporated in the analysis. The apparent similarity of the results in x and y directions and the cylindrical symmetry of the detector in the third phase suggested the use of the same parameterisation of the systematic effect for these two directions. The uncertainty on the vertex resolution was expressed as a second-order polynomial for x and y directions and as a first-order polynomial for the z direction. Tables 2.3 and 2.4 present the fitted values for the parameters of these polynomials, along with their associated correlation matrices.
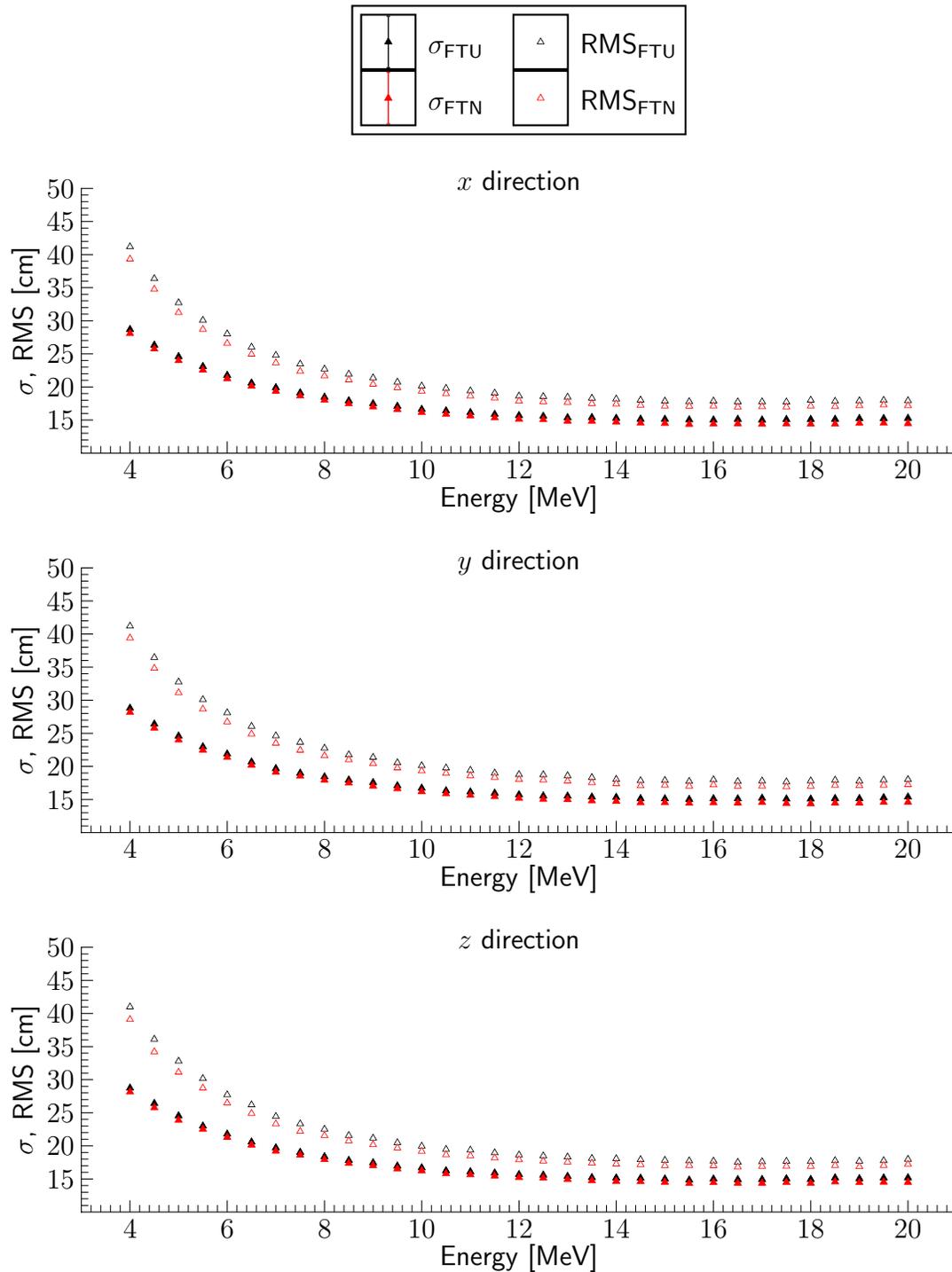
Figure 2.8: Fitted simple Gaussian resolution function parameter and computed RMS values in the three axis directions as a function of generated Monte Carlo electron energy.
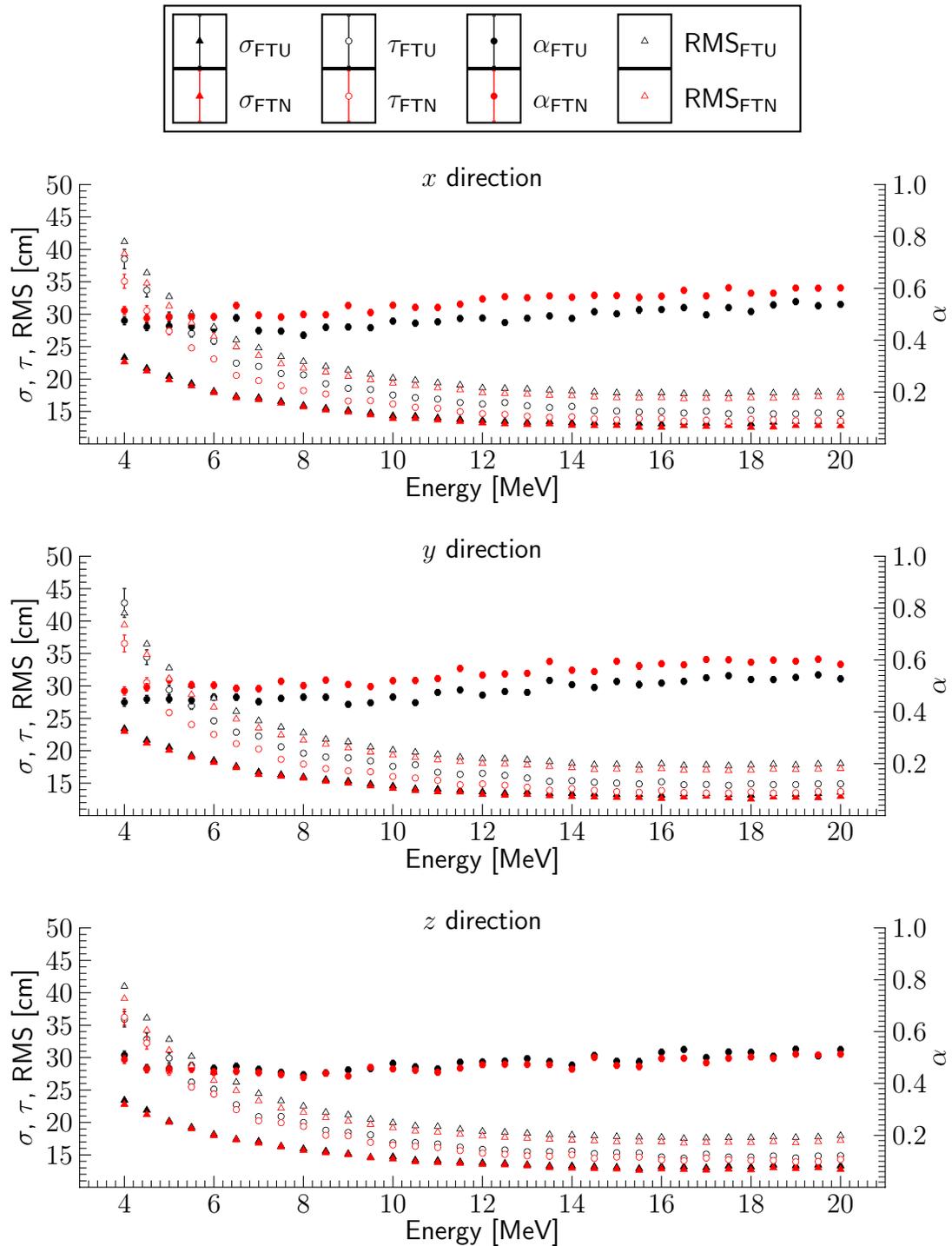
Figure 2.9: Fitted four-parameter resolution function parameters and computed RMS values in the three axis directions as a function of generated Monte Carlo electron energy.
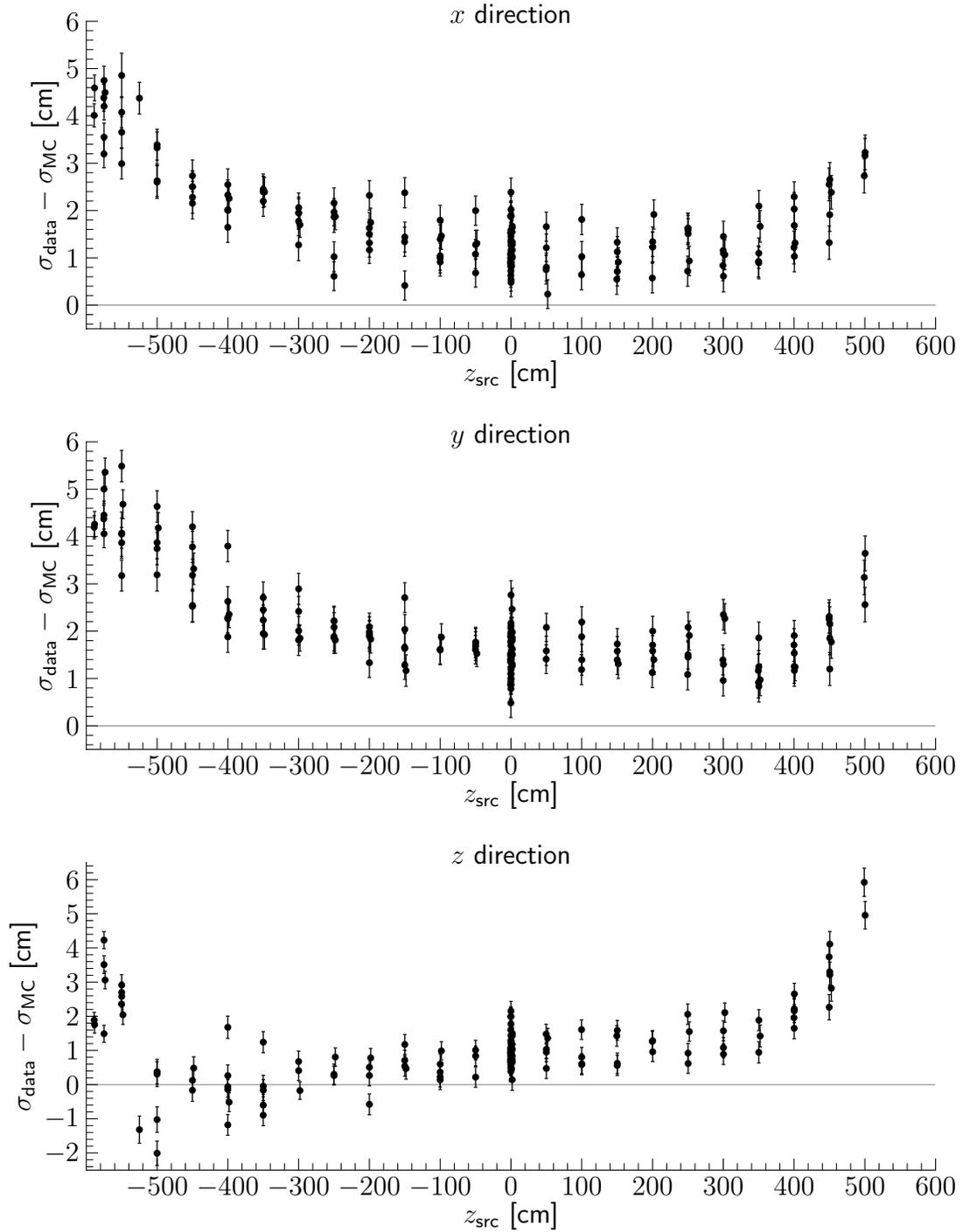
Figure 2.10: Difference between data and MC fitted vertex width in the three axis directions (x, y and z, from top to bottom) as a function of the position of the source (z_{src}) for ^{16}N runs taken along the z axis of the detector.

x and y directions			
ρ	a_0	a_1	a_2
a_0	1.00	-0.13	-0.74
a_1	-0.13	1.00	0.31
a_2	-0.74	0.31	1.00

z direction		
ρ	a_0	a_1
a_0	1.00	0.15
a_1	0.15	1.00

Table 2.4: Correlation matrices for the fitted parameters of the polynomials used to evaluate the systematic uncertainty on vertex resolution.

2.1.1.2.8 Angular Resolution

The angular resolution of FTN is evaluated through the function

$$R(\cos\theta) = \alpha_\mathsf{M}\frac{\beta_\mathsf{M}e^{\beta_\mathsf{M}(\cos\theta-1)}}{1-e^{-2\beta_\mathsf{M}}} + (1-\alpha_\mathsf{M})\frac{\beta_\mathsf{S}e^{\beta_\mathsf{S}(\cos\theta-1)}}{1-e^{-2\beta_\mathsf{S}}}. \tag{2.9}$$

In this equation, θ is the angle between the true and the reconstructed event directions. Two components are used for the angular resolution function: one for the main peak (β_S) and another for the multiple scattering (β_M), as recommended in [44]. Since the $^{16}\mathsf{N}$ source emits γ rays whose first interaction point can be quite far from the source position, the reconstructed event position can be used to obtain a good approximation of the true event's direction. Here $\cos\theta$ is thus approximated using the expression

$$\cos\theta \approx \frac{\vec{x}_\mathsf{fit} - \vec{x}_\mathsf{source}}{|\vec{x}_\mathsf{fit} - \vec{x}_\mathsf{source}|} \cdot \hat{u}_\mathsf{fit}. \tag{2.10}$$
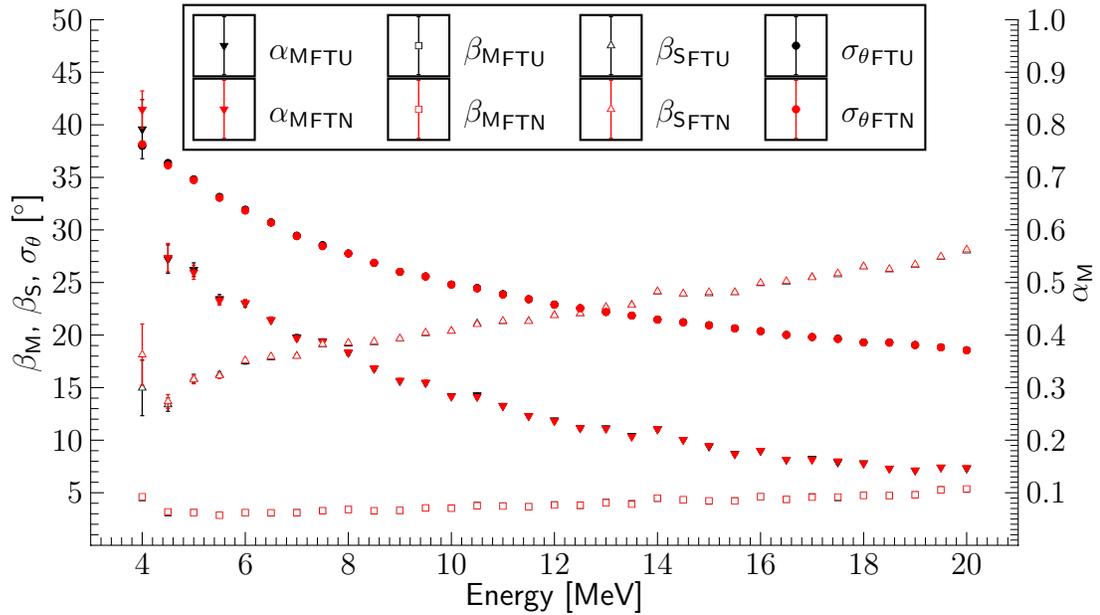
Figure 2.11: Fitted unitless angular resolution function parameters (α_M, β_M, β_S) and computed angular resolution (σ_θ) as a function of generated Monte Carlo electron energy, using sets of 10^5 isotropically and uniformly generated mono-energetic electrons. The points for the FTU and FTN algorithms are indistinguishable.

Events reconstructed less than 120 cm away from the source position are not considered in order to have a good estimate of the true event's direction.

The same Monte Carlo simulations that were presented in the previous section have been used to evaluate the volume-weighted angular resolution of FTN and FTU. Figure 2.11 shows the values of fitted parameters along with the computed angular resolution, corresponding to the angle between the initial electron direction and the fit direction that contains 68.27% of the angular distribution, as a function of energy. Error bars are the statistical uncertainties from the fits. Equivalent results for the FTU fitter are also shown. Since both fitters use the same algorithm to fit event direction, no significant improvements were expected with FTN for these fits.

For ES events, the systematic uncertainty on the angular resolution function of Equation (2.9) is propagated using the expression

$$\cos \theta'_{\odot} = 1 + [1 \pm \Delta_\theta](\cos \theta_{\odot} - 1), \tag{2.11}$$

where $\Delta_\theta = 0.12$ is the relative uncertainty on β_M and β_S parameters. It has been shown [45] that the CC $\cos \theta_{\odot}$ PDF, although having a slight negative slope, does not fluctuate significantly when varying angular resolution. The $\cos \theta_{\odot}$ PDFs for the other classes of events have flat distributions, except for hep CC and hep ES which have negligible effects for signal extraction. The above expression is derived in Appendix B.

2.1.1.2.9   Resolution Comparison with Old Algorithm

Different tests have been performed to evaluate the performance of FTN for the NCD phase. Events from a Monte Carlo simulation of 10^6 5.5 MeV electrons generated uniformly and isotropically within the acrylic vessel have been reconstructed by FTU and FTN. Table 2.5 shows a comparison between FTU and FTN regarding migration in and out of the fiducial volume. The four-parameter vertex position resolution function of Equation (2.8) has also been fitted to these distributions as a function of generated event radius. Figure 2.12 shows the ratio FTN/FTU of the parameters in the three axis directions as a function of generated event radius. These electrons have been reconstructed twice by FTN: with and without AV refraction calculations. Shown in Figure 2.12 are the results obtained by neglecting the effect of refraction through the AV, because the results of the fits with and the fits without AV refraction have been found to be very similar. From the ratios $\sigma_{FTN}/\sigma_{FTU}$ and τ_{FTN}/τ_{FTU} depicted in Figure 2.12, it shows that the new time fitter

R_{MC}	R_{fit}	
	< 550 cm	> 550 cm
< 550 cm	710449 / 708330	52711 / 54828
> 550 cm	29035 / 29675	199004 / 198363

Table 2.5: Number of FTN / FTU events generated inside and outside the fiducial volume that have been reconstructed inside and outside the fiducial volume, for a Monte Carlo simulation of 10^6 5.5 MeV electrons generated uniformly and isotropically within the acrylic vessel. Upper-right cell corresponds to migration of events generated inside the fiducial to a reconstructed position outside the fiducial volume. Lower-left cell corresponds to contamination from events generated outside the fiducial volume. Default FTN settings have been used.

$|x_{MC}| + |y_{MC}|$ [cm]	σ_{FTU} [cm]	τ_{FTU} [cm]	RMS_{FTU} [cm]
< 450	19.29 ± 0.05	26.6 ± 0.2	30.02
> 450	18.90 ± 0.04	25.0 ± 0.1	28.39

Table 2.6: RMS values and fitted four-parameter resolution function parameters in the x direction for 10^6 5.5 MeV electrons generated uniformly and isotropically inside and outside the NCD array. Events have been reconstructed using FTU.

$|x_{MC}| + |y_{MC}|$ [cm]	σ_{FTN} [cm]	τ_{FTN} [cm]	RMS_{FTN} [cm]
< 450	19.02 ± 0.05	25.2 ± 0.1	28.52
> 450	18.50 ± 0.06	22.2 ± 0.1	27.10

Table 2.7: RMS values and fitted four-parameter resolution function parameters in the x direction for 10^6 5.5 MeV electrons generated uniformly and isotropically inside and outside the NCD array. Events have been reconstructed using FTN.

for the NCD phase performed better by considering the NCD shadowing. Tables 2.6 to 2.7 show the values of the same parameters for events generated inside and outside the NCD array and reconstructed using FTU and FTN. The number of events reconstructed 300 cm away or more from the generated position is shown in Table 2.8.

The two-parameter resolution function has been fitted for the ^{16}N data runs taken
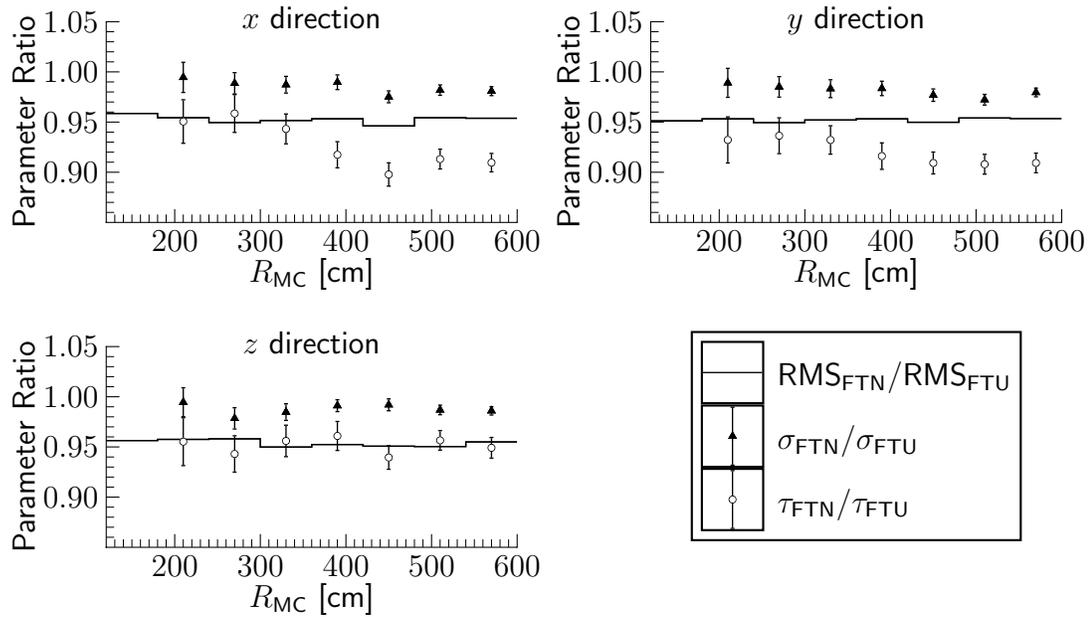
Figure 2.12: Ratios of RMS values and of fitted four-parameter resolution function parameters in the three axis directions as a function of generated events radius, for 10^6 5.5 MeV electrons generated uniformly and isotropically within the acrylic vessel. Default FTN settings have been used.

Direction	FTU	FTN
x	2484	2081
y	2488	2058
z	2508	2081

Table 2.8: Number of events reconstructed 300 cm away or more from the generated position, for 10^6 5.5 MeV electrons generated uniformly and isotropically within the acrylic vessel. Events have been reconstructed using FTU and FTN.

during the NCD phase with sufficient statistics. In the reconstructed position histograms, more than 5 events for all bins located in the range $[-150, +150]$ cm was enforced. Each run has been reconstructed with both the FTU and the FTN fitters. Figures 2.13 and 2.14 show the difference between fitted resolution parameters, weighted using statistical uncertainties from the fits. Figure 2.14 shows a clear improvement of vertex resolution
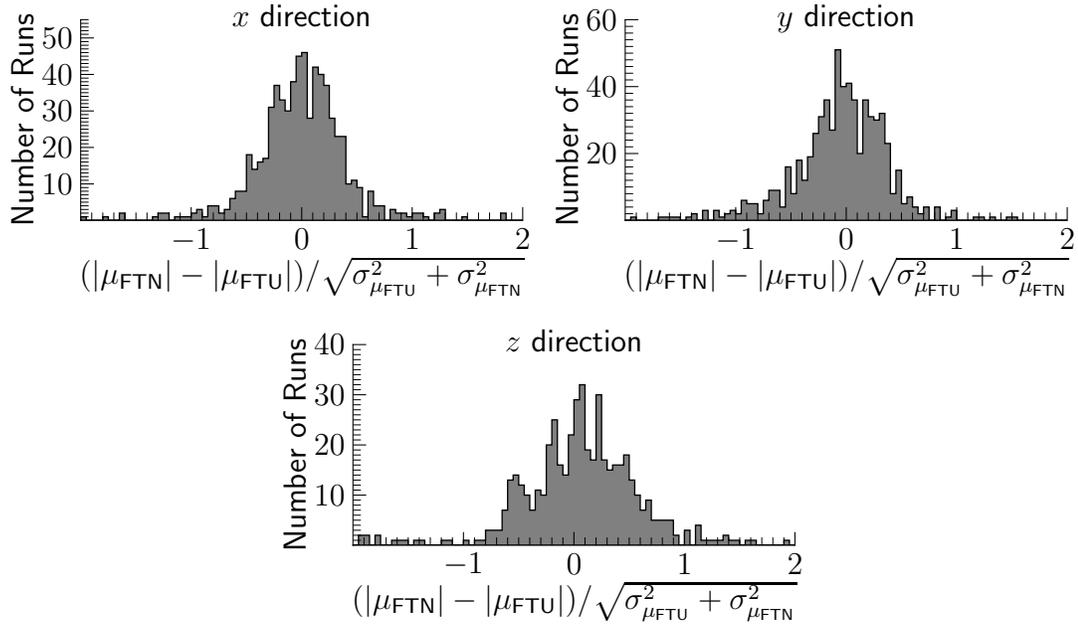
Figure 2.13: Weighted difference between FTN and FTU fitted two-parameter resolution function mean in the three axis directions for ^{16}N calibration runs taken during the NCD phase.

when comparing FTN to FTU. This is particularly obvious with data where higher statistics significantly reduce statistical uncertainties. Figures 2.15 and 2.16 show the corresponding distributions for the absolute difference between resolution parameters.

The four-parameter resolution function has also been fitted for the ^{16}N data runs taken during the NCD phase with sufficient statistics. In the reconstructed position histograms, more than 5 events for all bins located in the range $[-180, +180]$ cm was enforced. The increased number of parameters and their high correlation requires an extended range for the fit. The α parameter has been fixed to the values shown in Table 2.9 corresponding to the interpolated values at 5.63 MeV in Figure 2.9. Figures 2.17 to 2.19 show the difference between fitted resolution parameters. Figure 2.18 presents an average improvement of ≈ 1.25 cm for the Gaussian parameter in the x and y direction and of ≈ 0.6 cm for the
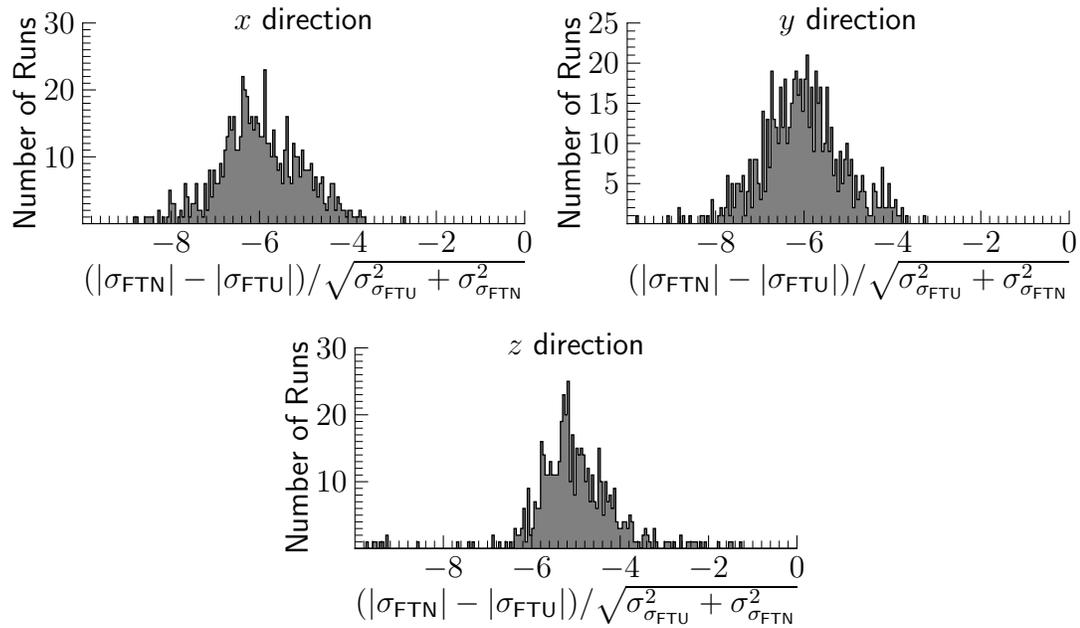
Figure 2.14: Weighted difference between FTN and FTU fitted two-parameter resolution function σ parameter in the three axis directions for ^{16}N calibration runs taken during the NCD phase.
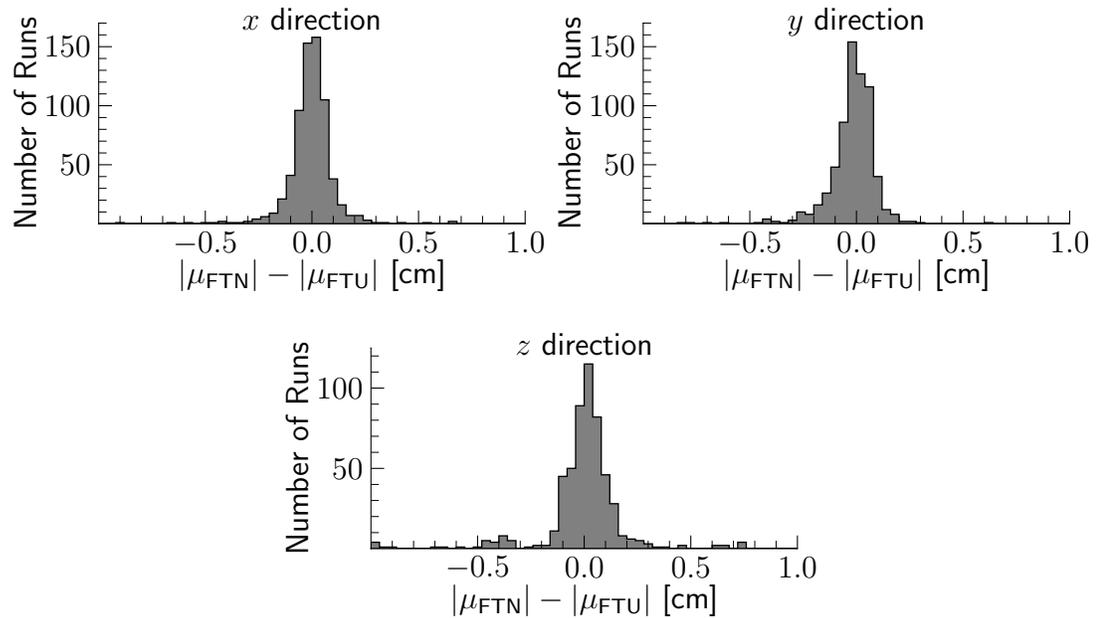
Figure 2.15: Difference between FTN and FTU fitted two-parameter resolution function mean in the three axis directions for ^{16}N calibration runs taken during the NCD phase.
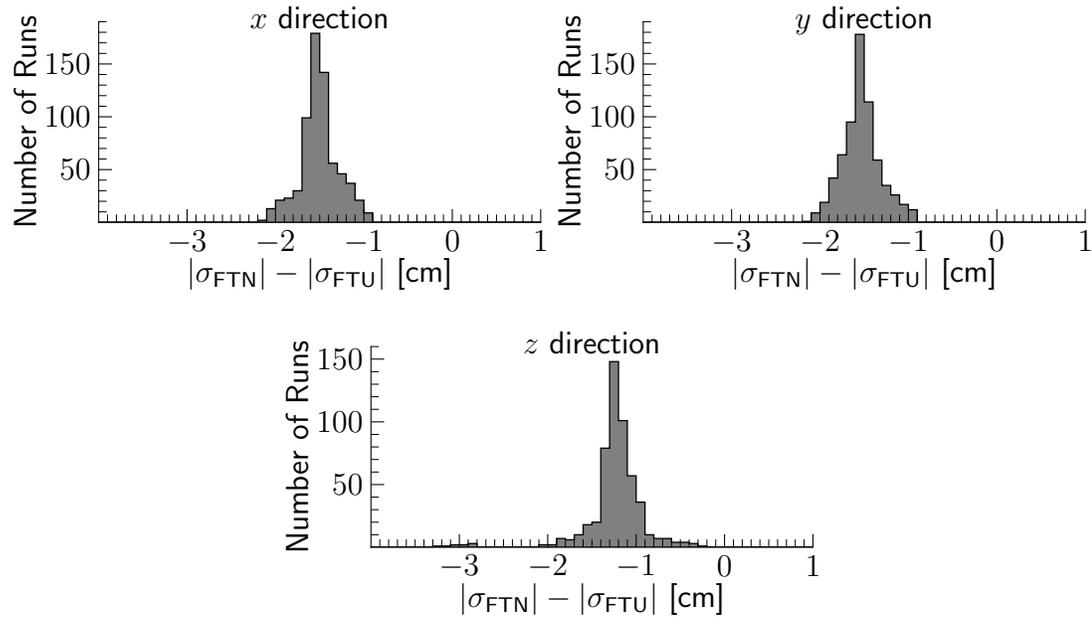
Figure 2.16: Difference between FTN and FTU fitted two-parameter resolution function σ parameter in the three axis directions for ^{16}N calibration runs taken during the NCD phase.

Parameter	FTU	FTN
α_x, α_y	0.448	0.496
α_z	0.467	0.453

Table 2.9: Values of α used to fit the convoluted four-parameter resolution function on ^{16}N data and MC reconstructed events position distributions.

same parameter in the z direction. Figure 2.19 shows an average improvement of ≈ 3 cm for the exponential parameter τ in the x and y direction and of ≈ 1.2 cm in the z direction.
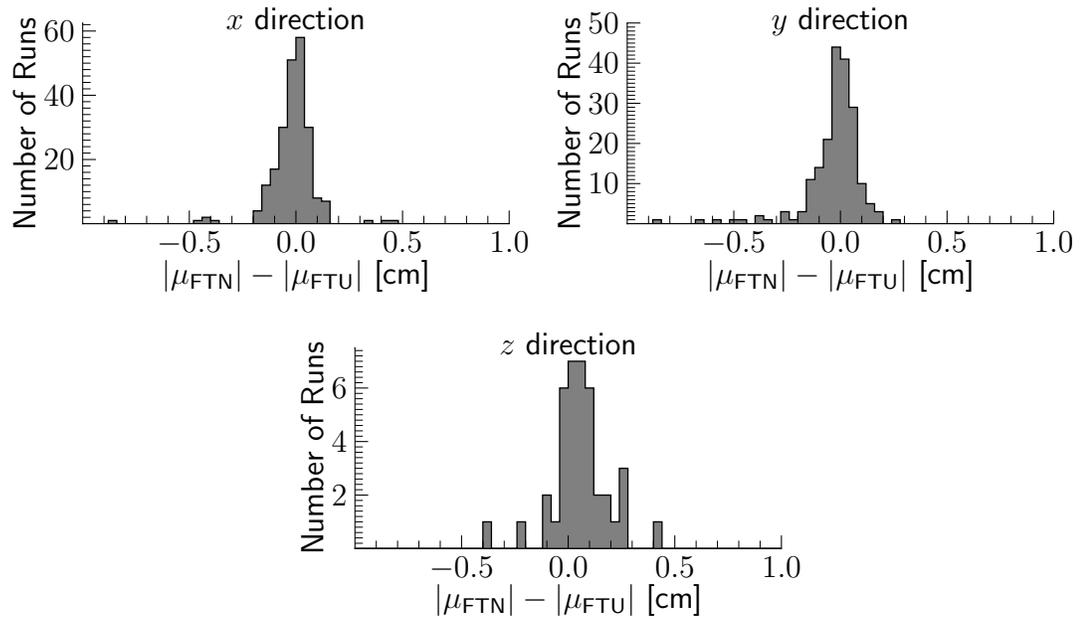
Figure 2.17: Difference between FTN and FTU fitted four-parameter resolution function mean in the three axis directions for ^{16}N calibration runs taken during the NCD phase.
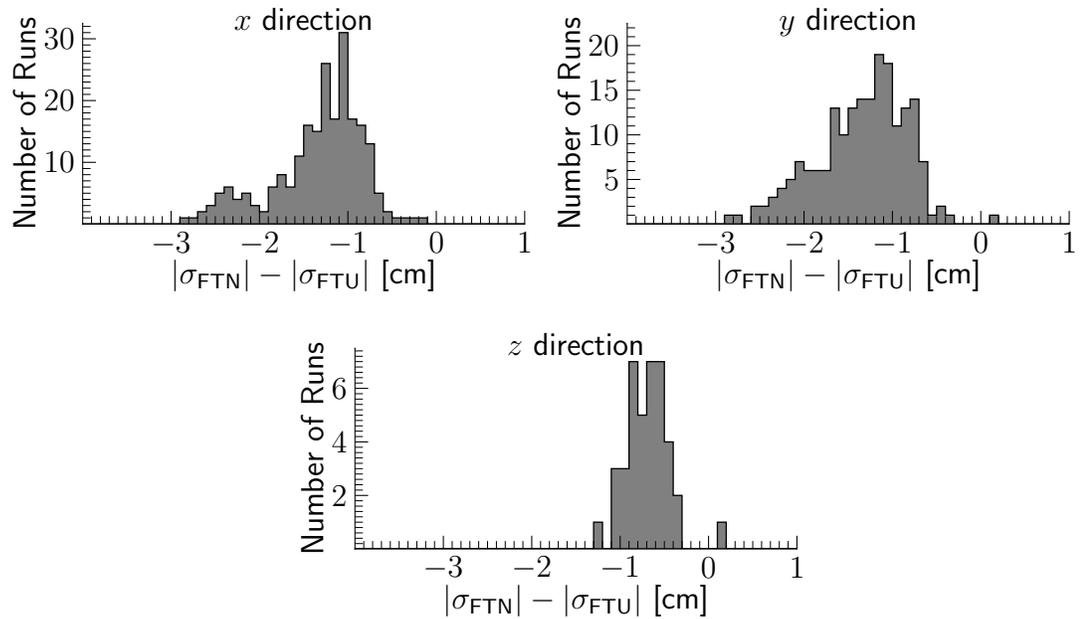
Figure 2.18: Difference between FTN and FTU fitted four-parameter resolution function σ parameter in the three axis directions for ^{16}N calibration runs taken during the NCD phase.
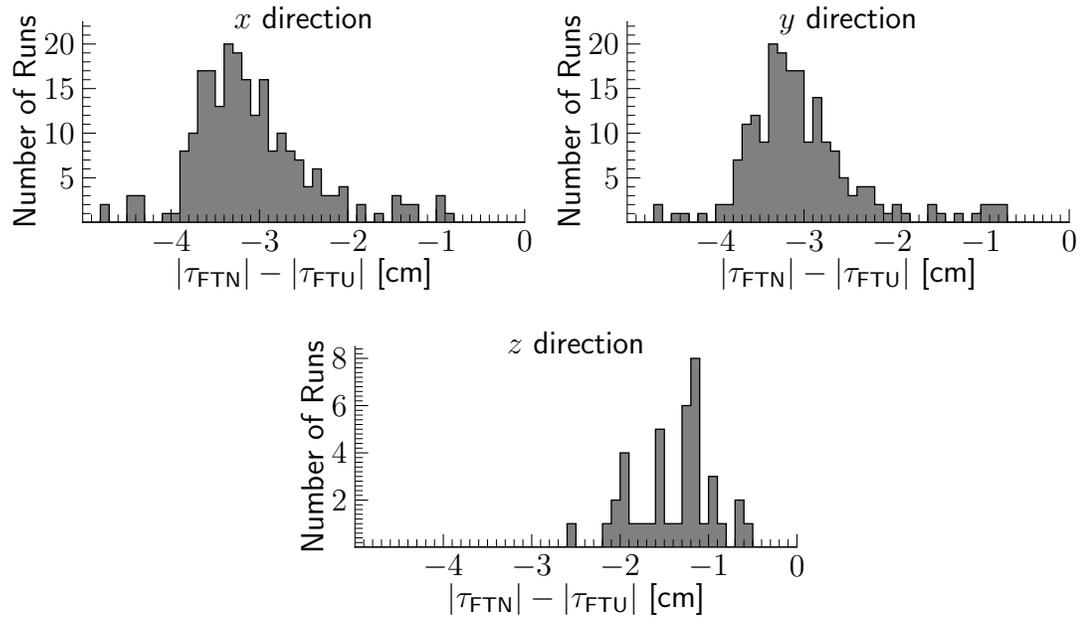
Figure 2.19: Difference between FTN and FTU fitted four-parameter resolution function τ parameter in the three axis directions for ^{16}N calibration runs taken during the NCD phase.

2.1.2   Energy Reconstruction

After the reconstruction of the event radius, the estimation of the event energy constitutes the most important observable for the SNO signal extraction. This observable allows to distinguish the different neutrino reactions, but also the signals from low-energy backgrounds. The electron energy is also very strongly correlated to the neutrino energy for CC and ES reactions, such that this observable can give a handle on the extraction of the energy dependence of the electron neutrino survival probability. Needless to say, SNO energy reconstruction algorithms must thus be properly designed to provide an unbiased energy estimator.

2.1.2.1 D$_2$O and Salt Phases

2.1.2.1.1 RSP

RSP is a fast energy reconstruction algorithm that estimates energy by determining the deposited energy necessary to generate the estimated number of direct hits recorded by the PMTs. The technique used by RSP aims at depending as weakly as possible on Monte Carlo simulations in order to minimise the systematic uncertainties associated with the simulation modelling for the extracted energy. The number of direct hits are estimated by only considering the hits in a prompt peak window of ± 10 ns around the time residual value determined from vertex reconstruction. This allows to eliminate reflected and scattered light, which could not be taken into account without relying on simulations. For the D$_2$O and salt phases, RSP bases its calculations using vertex information provided by FTU.

2.1.2.1.2 FTK

Albeit the qualities of the RSP estimator, the usage of only the direct light can be quite limiting in the context of a low-energy-threshold analysis, where the energy resolution is very important to minimise the effects of the background at low energy on the extracted signals. The usage of all the available light, which increases the number of included hits per MeV by 12%, is estimated to improve the energy resolution by 6%, which is expected to reduce the number of events for these backgrounds by up to 60% [46]. For LETA, a new energy estimator, FTK, has thus been developed to circumvent this limitation.

To perform energy reconstruction, FTK maximises the likelihood function

$$L(E) = \sum_{n_\gamma=1}^{\infty} \frac{\nu_{\text{hits}}(n_\gamma)^{n_{\text{hits}}} e^{-\nu_{\text{hits}}(n_\gamma)}}{n_{\text{hits}}!} P(n_\gamma|E),$$ (2.12)

where $P(n_\gamma|E)$ is the probability that n_γ Čerenkov photons be emitted given an energy E for the event and where n_{hits} is the number of observed hits that follows a Poisson distribution with parameter $\nu_{\text{hits}}(n_\gamma)$ for the expected number of PMT hits given a number of emitted photons.

The challenge of FTK is the modelling of $\nu_{\text{hits}}(n_\gamma)$ that has to be performed with an accuracy better than 1%, while being three orders of magnitude faster than performing a full Monte Carlo simulation. FTK achieves this task by splitting $\nu_{\text{hits}}(n_\gamma)$ into direct and scattered light, AV and PMT reflections, plus PMT noise components. Each type of event is computed separately, using optimised analytical and/or MC methods. More details about FTK can be found in [46].

2.1.2.2   NCD Phase

For the NCD phase, the RSP fitter has been modified to correct the reconstructed energy for PMT shadowing by the NCDs. Since RSP uses the light from the prompt peak to establish the event energy, this correction is achieved by evaluating the fraction of the light which is incident on the NCDs and by assuming that this light is then lost. As Čerenkov light is emitted along a track having a non-negligible length, shadowing cannot be computed very efficiently because of the requirement to integrate its probability along the track. The approach taken by RSP is to use a lookup table to provide average

shadowing as a function of radius and direction (the latter being measured with respect to the radial vector). The RSP fitter for the NCD phase is described in further detail in [47].

2.2 Pattern Recognition for the NCDs: The Pulse Shape Analysis

During the NCD phase, the neutral current detectors provide an almost completely independent measurement of the number of neutrons produced in the detector. In 2008, SNO released a letter [7] in which results from the third phase of the experiment were presented. In this first analysis, the number of NC events from the proportional counters was obtained using the statistical separation between the probability density functions of the energy spectra for neutrons, alpha particles and instrumental events detected by the counters.

For the three-phase combined analysis of the SNO experiment, a different method is used to extract the signal from the NCD strings. Rather than relying solely on event energy, this new method relies also on the information contained in the shape of the waveforms generated by the strings to distinguish the different types of events. This gives an additional lever in discriminating between the signal and the backgrounds, because neutron interactions lead to the production of two particles, a proton and a triton, while this is not the case for alpha particles. For events whose proton-triton pair is not parallel to the wire of the proportional counter, a double-peaked pulse can thus be observed. This section briefly presents three different methods (DWK, NAP and QGF) to determine

the total number of neutrons detected by the NCDs, followed by the technique used to combine them.

2.2.1   DWK

The DWK method relies on the computation of two moments of the measured pulses to separate the alpha particles from the neutrons. The separation is achieved by deconvolving the ion mobility from the initial pulses, then by computing the skewness and the kurtosis of the resulting waveforms, using a range defined by the continuous region where the voltage is at least 20% of its maximum value. The kurtosis evaluates the peakedness of the distribution and is higher on average for alpha particles, because they generate waveforms that are single-peaked. The skewness measures the asymmetry of the pulses. Neutrons generally produce waveforms with larger, positive skewness, due to space charge effects and the relatively different energies of the produced neutrons and tritons. To select neutrons, DWK uses a bidimensional cut in a skewness-kurtosis plane.

2.2.2   NAP

Compared to DWK, the approach taken by NAP is significantly different. To differentiate neutrons and alpha particles, this second method uses libraries of pulses that are generated using Monte Carlo simulations. Each library consists of a multi-dimensional grid where pulses are sorted depending on Monte Carlo generation parameters. The neutron library is generated by varying particle position and the z position on the NCD string, the proton and triton energies being fixed at 574 keV and 191 keV, respectively. For the alpha

library, the initial direction, the z position and the particle energy are varied, while the initial radius is fixed at either the wall inner radius or the wire outer radius. To separate the events, NAP starts by performing Pearson χ^2 tests to compare the data pulses to the pulses contained in the libraries. This parameter is computed by using a range defined by the region located between the position where the rising edge of the pulse reaches 10% of the pulse maximum and the position where the falling edge of the pulse crosses 30% of the maximum. The higher threshold for the falling edge is used to avoid using information from the ion-drift tail, which is similar for neutrons and alpha particles. The variance used in the calculation of the χ^2 is based on different components, which include the electronic noise, the digitisation uncertainty, the sparseness of the Monte Carlo library and the systematic uncertainties of the simulation. The NAP algorithm computes χ^2 values in order to find which neutron and alpha pulses from the libraries are the most similar to a given data pulse. Once the pair of pulses is identified, it then computes the selection variable

$$\Delta \log(\chi^2) \equiv \log(\chi^2_\alpha) - \log(\chi^2_n), \tag{2.13}$$

where χ^2_α and χ^2_n are the best χ^2 values that have been computed using the alpha and neutron libraries, respectively. A value $\Delta \log(\chi^2)_{\text{corr}}$ is then computed, after correcting the original value for temporal changes in the response of the neutral current detectors. From the definition of $\Delta \log(\chi^2)_{\text{corr}}$, neutrons are thus selected by considering the events whose value of the parameter falls above a certain threshold value.

2.2.3   QGF

Similarly to NAP, QGF computes χ^2 values between data pulses and libraries of neutron and alpha events. QGF's libraries are however based on calibration data rather than simulation events. The upper limit of the range used to compute the χ^2 values for this method corresponds to the point where the falling edge of the pulse crosses 40% of the maximum amplitude. Also, the variance used is determined using the root mean square (RMS) of the noise in the last 3 μs of the data pulse. The selection parameter which is used for the QGF method is χ_α^2/χ_n^2 and can thus be interpreted very similarly to the parameter $\Delta\log(\chi^2)$ of the NAP method.

2.2.4   Constraint on Neutron Events

To provide a measurement for the total number of detected neutrons by the neutral current detectors during the NCD phase, the pulse shape analysis (PSA) uses a combination of a cut-based selection algorithm and a statistical separation method.

The PSA cut is defined as a pair of 2D cuts. The first one is based on the NAP and QGF methods and thus defines a contour which surrounds the alpha-rich region in the $\Delta\log(\chi^2)_{\text{corr}}$-$\chi_\alpha^2/\chi_n^2$ plane. The second 2D cut uses the skewness and kurtosis variables from the DWK method. The overall cut is very efficient at eliminating alpha particles, rejecting over 98% of them between 0.4 MeV and 0.9 MeV for the ^{4}He strings.

Following the application of the PSA cut, the analysis proceeds to the extraction of the number of neutrons that are contained within the remaining events, using an energy fit between 0.4 MeV and 1.4 MeV. The fit is performed while floating the shape of an

analytical PDF for the alpha particles, which is defined by

$$f(E|\alpha) = a_0 \left[P_0(E) + \sum_{i=1}^{n} a_i P_i(E) \right],$$ (2.14)

where $P_i(E)$ is the Legendre polynomial of order i. The maximum polynomial order n is determined using the method described in [48].

This section has thus summarised the pulse shape analysis that is used to provide a measurement for the total number of neutrons in the NCD phase. More details regarding this analysis can be found in [39].

Chapter 3

Likelihood Function

To perform the three-phase extraction of the electron neutrino survival probability, the SNO experiment relies on the statistical separation of the signals with respect to the backgrounds when comparing their expected PDFs. The idea behind the maximum likelihood parameter estimation of the survival probability is to determine the parameter values of the model that maximise the likelihood of the SNO data. The construction of the likelihood function is quite complex for the current analysis, due to the number of different observables, backgrounds, phases, constraints and to the direct extraction of the survival probability function. This chapter presents different aspects of the likelihood function used by the package developed by the author of this thesis to perform a maximum likelihood fit of the SNO data.

3.1 Likelihood Function Overview

For frequentists, a likelihood function represents a model that provides the probability of a given data sample, for specific model parameter values. These data obviously include the observables for all events in the analysis windows, but can also incorporate external measurements of parameters, such as background rates and parameters of the systematic

uncertainties. In this document, the former will be referred to simply as "data" and the latter as "measurements". Using simply the definition of the conditional probability, the likelihood function can thus be expressed as

$$L = f(\text{data}, \text{measurements})$$
$$= \underbrace{f(\text{data}|\text{measurements})}_{\equiv L_{\text{data}}} \times \underbrace{f(\text{measurements})}_{\equiv L_{\text{measurements}}}. \tag{3.1}$$

The probability density for the data is the joint probability of the observed values $\mathbf{x}$ and the number of observed events n_{events}:

$$L_{\text{data}} = f(\text{data}) = f(\mathbf{x}, n_{\text{events}}) = P(n_{\text{events}})f(\mathbf{x}|n_{\text{events}}). \tag{3.2}$$

To simplify the notation, the conditionality of L_{data} on the measurements will not always be mentioned in this document, although it is always implied. The probability of observing a given number of events, $P(n_{\text{events}})$, follows a Poisson distribution with parameter ν that depends on parameters such as the generated event rate, the overall efficiency of the detector, the live time of the experiment and the predefined analysis window. Given that the individual observed events are considered as independent, the probability density $f(\mathbf{x}|n_{\text{events}})$ can be expressed as the product of the probability densities for the individual events. Equation (3.2) can thus be rewritten as

$$L_{\text{data}} = \frac{\nu^{n_{\text{events}}}}{n_{\text{events}}!}e^{-\nu} \prod_{i=0}^{n_{\text{events}}} f(\vec{x}_i), \tag{3.3}$$

where $\vec{x}_i$ is the multi-dimensional vector which represents the set of observed values associated to event i.

Since maximising a likelihood function is equivalent to maximising its logarithm, the logarithm of Equation (3.1) can be computed, giving

$$LL = LL_{\text{data}} + LL_{\text{measurements}},\tag{3.4}$$

where

$$LL_{\text{data}} \equiv \log(L_{\text{data}}) = -\nu + \sum_{i=1}^{n_{\text{events}}} \log\left[\nu f(\vec{x}_i)\right] - \log(n_{\text{events}}!)\tag{3.5}$$

and $LL_{\text{measurements}} \equiv \log(L_{\text{measurements}})$. The last term in Equation (3.5) does not depend on any parameter in the model, such that it contributes to a constant term in the log-likelihood function; hence it can be dropped. The resulting expression is the well-known general form for an unbinned extended log-likelihood function given by

$$ELL_{\text{data}} = -\nu + \sum_{i=1}^{n_{\text{events}}} \log\left[\nu f(\vec{x}_i)\right].\tag{3.6}$$

If a fit consists of determining the number of events for n_{classes} mutually independent classes of events by statistical separation of their PDFs, $f(\vec{x}_i)$ in Equation (3.6) can be rewritten as

$$f(\vec{x}_i|\vec{\theta}) = \sum_{j=1}^{n_{\text{classes}}} f(\vec{x}_i, j|\theta_j) = \sum_{j=1}^{n_{\text{classes}}} P(j|\theta_j) f(\vec{x}_i|j)$$
$$= \sum_{j=1}^{n_{\text{classes}}} \theta_j f(\vec{x}_i|j),\tag{3.7}$$

where $\vec{\theta}$ introduces the first parameters to the statistical model. θ_j, the jth element of vector $\vec{\theta}$, is the probability of an event to be in class j, which must satisfy

$$\sum_{j=1}^{n_{\text{classes}}} \theta_j = 1. \tag{3.8}$$

Equation (3.7) was obtained using $f\left(\vec{x}_i, j | \vec{\theta}\right) = f(\vec{x}_i, j | \theta_j)$, because the probability density of an event i to have observable values $\vec{x}_i$ and to be in the class j depends only on the probability to be in this particular class of events. Also, $f(\vec{x}_i | \theta_j, j) = f(\vec{x}_i | j)$ since the probability for an event known to be in class j to have observable values $\vec{x}_i$ does not depend on θ_j and $\theta_j = P(j | \theta_j)$ by definition. Defining $\nu_j \equiv \theta_j \nu$, Equations (3.6) and (3.7) can be combined to give

$$ELL_{\text{data}} = -\sum_{j=1}^{n_{\text{classes}}} \nu_j + \sum_{i=1}^{n_{\text{events}}} \log\left[\sum_{j=1}^{n_{\text{classes}}} \nu_j f(\vec{x}_i | j)\right]. \tag{3.9}$$

In the above expression, the parameters ν_j constitute independent Poisson parameters which are not subject to a constraint equivalent to Equation (3.8). The total number of defined parameters, equal to n_{classes}, has thus been conserved.

Although Equation (3.9) allows to fit for the number of events for each class j, its parameterisation is such that the values of the estimators $\hat{\nu}_j$ will be influenced by systematic uncertainties in an undesirable way (remembering that the model constituted by ELL_{data} can implicitly depend on parameters of the systematic uncertainties). For example, regarding the CC, ES and NC signal classes for the SNO PMTs, the estimators are normally used to measure neutrino fluxes, rather than the number of detected events. Additionally, given a multi-phase fit, one will want to constrain the fluxes to be the same for

all phases. However, if simulations are used to compute the conversion factors between neutrino fluxes and the number of events in the analysis and if these simulations are affected by different systematic uncertainties, the conversion factors will not only depend on live time, overall detector efficiency and analysis window, but also on all the systematic uncertainties affecting the observables. Additionally, the number of background events is often constrained by independent measurements which measure the background source strength, rather than the number of events within the analysis window. To convert these measurements to the expected number of detected events, a simulation of these events is used. It is thus useful to define the parameterisation such that the estimators of interest be independent from the systematic uncertainties affecting the simulation.

Let $\vec{\eta}_j$ represent the value of all systematic uncertainty parameter (also often referred to as nuisance parameters) shifts for class j. If $\nu_{j_{\text{exp}}}(\vec{\eta}_j)$ is defined as the expected number of observed j-type events for the experiment, as obtained with a simulation modified using systematic shifts $\vec{\eta}_j$, a new set of parameters α_j, each having an expected value of 1, can be defined as $\alpha_j \equiv \frac{\nu_j}{\nu_{j_{\text{exp}}}}$, such that Equation (3.9) can be rewritten as

$$ELL_{\text{data}} = -\sum_{j=1}^{n_{\text{classes}}} \alpha_j \nu_{j_{\text{exp}}}(\vec{\eta}_j) + \sum_{i=1}^{n_{\text{events}}} \log\left[\sum_{j=1}^{n_{\text{classes}}} \alpha_j \nu_{j_{\text{exp}}}(\vec{\eta}_j) f(\vec{x}_i|j, \vec{\eta}_j)\right]. \qquad (3.10)$$

In the case of a multi-phase (multi-detector) fit, a combined fit can be performed by simply summing the extended log-likelihood functions which are each given by Equation (3.10) for the different phases (detectors), as long as the events from the different phases (detectors) can be considered as independent. If some classes of events are common to multiple phases (detectors) and the source for such events is consistent for all these phases (detectors), the associated α_j parameters can simply be forced to have the same

value. The same argument applies to systematic uncertainty parameters. Equation (3.10) thus represents a very general form of an extended log-likelihood which aims to provide unbiased estimators for the production rates (in units of the expected rates) of n_{classes} independent classes of events.

3.2 Probability Density Function Renormalisation

Continuing to use the general form of the extended log-likelihood when fitting the number of events of mutually independent classes of events, n_{classes}, it will now be shown that the normalisation of PDFs can be changed by a constant factor A_i, or in other words, by a factor that can be different for every event, but that is the same for every class, without changing the results of a fit. The substitution $f(\vec{x}_i|j) \rightarrow A_i f(\vec{x}_i|j)$ in Equation (3.9) gives

$$
\begin{aligned}
ELL'_{\text{data}} &= -\sum_{j=1}^{n_{\text{classes}}} \nu_j + \sum_{i=1}^{n_{\text{events}}} \log\left[\sum_{j=1}^{n_{\text{classes}}} \nu_j A_i f(\vec{x}_i|j)\right] \\
ELL'_{\text{data}} &= -\sum_{j=1}^{n_{\text{classes}}} \nu_j + \sum_{i=1}^{n_{\text{events}}} \log\left[A_i \sum_{j=1}^{n_{\text{classes}}} \nu_j f(\vec{x}_i|j)\right] \\
ELL'_{\text{data}} &= -\sum_{j=1}^{n_{\text{classes}}} \nu_j + \sum_{i=1}^{n_{\text{events}}} \log\left[\sum_{j=1}^{n_{\text{classes}}} \nu_j f(\vec{x}_i|j)\right] + \sum_{i=1}^{n_{\text{events}}} \log A_i \\
ELL'_{\text{data}} &= ELL_{\text{data}} + \sum_{i=1}^{n_{\text{events}}} \log A_i.
\end{aligned}
\tag{3.11}
$$

Here, the last term does not depend in any way on the fit parameters. Such a term can be dropped, showing the equivalence between the two expressions, since it does not affect

the maximisation of the likelihood function for a given dataset, as it only shifts the value of the log-likelihood by a constant.

This property of the likelihood function has a very useful application. In signal extraction, it is quite common to estimate the PDFs $f(\vec{x}_i|j, \vec{\eta}_j)$ using binned PDFs $f_b(\vec{x}_i|j, \vec{\eta}_j)$ that are generated from Monte Carlo samples using the systematic uncertainty parameters $\vec{\eta}_j$. To compute one of these binned PDFs, a histogram $H(b(\vec{x}_i)|j, \vec{\eta}_j)$ with linearized bin index b is filled with $N(\vec{\eta}_j)$ Monte Carlo events located in the analysis window. Then the PDF is normalised using

$$f_b(\vec{x}_i|j, \vec{\eta}_j) = f_b(b(\vec{x}_i)|j, \vec{\eta}_j) = \frac{H(b(\vec{x}_i)|j, \vec{\eta}_j)}{N(\vec{\eta}_j)V(b)}, \qquad (3.12)$$

where $V(b)$ is the width/area/volume/hypervolume of the bin b. If the PDFs for all the classes of events are binned PDFs generated using the exact same binning, the normalisation of all the histograms $H(b(\vec{x}_i)|j, \vec{\eta}_j)$ shares common factors $V(b(\vec{x}_i))$ which depend on the event index but not on the event class. From Equation (3.11), it means that the substitution

$$f_b(\vec{x}_i|j, \vec{\eta}_j) \rightarrow \frac{1}{N(\vec{\eta}_j)}H(b(\vec{x}_i)|j, \vec{\eta}_j) = \frac{1}{N(\vec{\eta}_j)}H(\vec{x}_i|j, \vec{\eta}_j) \qquad (3.13)$$

leads to an equivalent likelihood function. If the likelihood function is now parameterised as Equation (3.10), that is to say by fitting parameters $\alpha_j \equiv \frac{\nu_j}{\nu_{j\,\text{exp}}(\vec{\eta}_j)}$ rather than ν_j, one can simplify the likelihood function even further. Effectively, the dependency of $\nu_{j\,\text{exp}}(\vec{\eta}_j)$ on the systematic uncertainty parameters can be estimated from Monte Carlo samples by

using the relation

$$\nu_{j\,\mathrm{exp}}(\vec{\eta}_j) = \nu_{j\,\mathrm{nom}}\frac{N(\vec{\eta}_j)}{N_{\mathrm{MC}j}} = \frac{N(\vec{\eta}_j)}{M_{\mathrm{MC}j}},$$ (3.14)

where $\nu_{j\,\mathrm{nom}}$ and $N_{\mathrm{MC}j}$ are, respectively, the nominal, expected Poisson parameters and the number of Monte Carlo events in the analysis window for class j when default systematic shifts are applied. Also, $M_{\mathrm{MC}j}$ is defined as the Monte Carlo multiplicity with respect to $\nu_{j\,\mathrm{nom}}$. Finally, from Equations (3.13) and (3.14), when using binned Monte Carlo as an approximation for the PDFs $f(\vec{x}_i|j,\vec{\eta}_j)$, an equivalent expression for Equation (3.10) is given by

$$ELL'_{\mathrm{data}} = -\sum_{j=1}^{n_{\mathrm{classes}}} \alpha_j \frac{N(\vec{\eta}_j)}{M_{\mathrm{MC}j}} + \sum_{i=1}^{n_{\mathrm{events}}} \log\left[\sum_{j=1}^{n_{\mathrm{classes}}} \frac{\alpha_j}{M_{\mathrm{MC}j}} H(\vec{x}_i|j,\vec{\eta}_j)\right],$$ (3.15)

which holds as long as the same observable binning is used for all classes of events. This expression thus allows to evaluate the likelihood function without having to normalise the Monte Carlo histogram at all, avoiding thus a time-consuming operation when systematic uncertainty parameters are floated in the fit. Under the listed conditions, such an optimisation is always possible. It should be noticed though that this technique allows to only absorb one set of observable bin widths/areas/volumes/hypervolumes and one number of Monte Carlo events. One must thus proceed with care when applying it to joint PDFs composed of multiple marginal PDFs.

3.3 Binned versus Unbinned Likelihood Functions

Ideally, unbinned PDFs should be used in a likelihood function in order to maximise statistical separation. In practice, this is not always an easy task to do however, due

to the complexity involved in the correlations between the observables and/or in the parameterisation of the systematic uncertainties. It is thus common to create binned PDFs generated from Monte Carlo samples. The binning of such PDFs should normally be chosen to minimise the total uncertainty on the parameters of interest. It is obtained when there is a good balance between the statistical uncertainties that increase and the systematic uncertainties associated to finite Monte Carlo statistics that decrease, when increasing the size of the bins. To avoid biases, the same binning should also be used for the different classes of events involved, such that the probability of a given region delimited by some bin boundaries corresponds to the same probability as the unbinned PDFs. Using a different binning depending on the class of events effectively violates this rule, unless bin grouping is performed for known flat regions.

It will now be shown that using binned PDFs in an unbinned extended log-likelihood is mathematically equivalent to using a binned extended log-likelihood function. Replacing the unbinned PDFs $f(\vec{x}_i|j)$ in Equation (3.9) by binned PDFs $f_b(\vec{x}_i|j)$ with linearized bin index $b(\vec{x}_i)$ gives

$$
\begin{aligned}
ELL_{\text{data}} &= -\sum_{j=1}^{n_{\text{classes}}} \nu_j + \sum_{i=1}^{n_{\text{events}}} \log\left[\sum_{j=1}^{n_{\text{classes}}} \nu_j f_b(b(\vec{x}_i)|j)\right] \\
&= -\sum_{j=1}^{n_{\text{classes}}} \nu_j + \sum_{i=1}^{n_{\text{events}}} \left\{\log\left[\sum_{j=1}^{n_{\text{classes}}} \nu_j f_b(b|j)\right]\right\}_{b=b(\vec{x}_i)} \\
&= -\sum_{j=1}^{n_{\text{classes}}} \nu_j + \sum_{b=1}^{n_{\text{bins}}} \sum_{i=1}^{n_{\text{events}}} \log\left[\sum_{j=1}^{n_{\text{classes}}} \nu_j f_b(b|j)\right] \delta_{(\vec{x}_i \in b)} \\
&= -\sum_{j=1}^{n_{\text{classes}}} \nu_j + \sum_{b=1}^{n_{\text{bins}}} n_b \log\left[\sum_{j=1}^{n_{\text{classes}}} \nu_j f_b(b|j)\right],
\end{aligned}
\tag{3.16}
$$

where n_{bins} is the number of bins in the PDFs and where n_b is the number of data events

in bin b. By the virtue of the properties of the likelihood function demonstrated in the previous section, the binned PDF $f_b(b|j)$ can be replaced by the probability mass function (PMF) $P(b|j)$ (which differs by a factor given by the width/area/volume/hypervolume of bin b) to lead to an equivalent function which is more similar to the common expression of an unbinned extended log-likelihood.

In QSigEx, fits are always performed using a binned extended log-likelihood which also uses unnormalised Monte Carlo histograms. From Equations (3.15) and (3.16), the general expression for the extended log-likelihood, when fitting n_{classes} mutually independent classes of events and when using binned PDFs built from Monte Carlo samples with the same binning for all classes of events), can be written as

$$ELL'_{\text{data}} = -\sum_{j=1}^{n_{\text{classes}}} \alpha_j \frac{N(\vec{\eta}_j)}{M_{\text{MC}j}} + \sum_{b'=1}^{n_{\text{filled\ bins}}} n_{b'} \log \left[\sum_{j=1}^{n_{\text{classes}}} \frac{\alpha_j}{M_{\text{MC}j}} H(b'|j, \vec{\eta}_j) \right], \qquad (3.17)$$

where the index b' runs over bins that contain some events in the data.

3.4   Fitted Number of Events in an Extended Log-Likelihood

As seen in the previous section, when fitting for the number of events of n_{classes} mutually independent classes of events, the general form of the extended log-likelihood is given by

$$ELL_{\text{data}} = -\sum_{j=1}^{n_{\text{classes}}} \nu_j + \sum_{i=1}^{n_{\text{events}}} \log \left[\sum_{j=1}^{n_{\text{classes}}} \nu_j f(\vec{x}_i|j) \right]. \qquad (3.9)$$

This expression evaluates to a real value as long as $\sum_{j=1}^{n_{\text{classes}}} \nu_j f(\vec{x}_i|j) > 0$. The proper behaviour of a minimisation algorithm can thus be normally ensured by constraining positive values of this expression. Since an empty dataset is three standard deviations away from its expectation value when the latter is only nine events, such a constraint normally does not represent a significant distortion of the likelihood space. A stricter constraint could consist of forcing all ν_j parameters to be positive (physically meaningful). However, this leads automatically to biased estimators for the true number of events in every class if the data contain a sufficiently small fraction of events in a given class. Effectively, if many similar experiments were generated, an unbiased estimator would naturally lead to negative and positive number of events, averaging to the true, physically meaningful value.

To enforce positive values thus tends to bias negatively the fitted number of events for the other classes. It should be ensured instead that a negative fitted value for a number of events is compatible with 0 within the total uncertainties. A negative value that is too far off would indicate a problem with the modelling of the data (including underestimation of the systematic uncertainties). In this analysis, the ν_j parameters are not thus forced to be positive.

3.5   Likelihood Penalty Terms

As shown in Equation (3.4), the likelihood function of SNO is composed of a series of terms representing the data and other terms associated to external measurements that are referred to as "penalty terms". Although these measurements are usually assumed

to follow Gaussian-like statistics, the expressions of their penalty terms can differ quite significantly depending on the assumptions that are made about how these measurements affect the number of events in LL_data. The following sections address the different types of penalty terms used in the extraction of the survival probability.

3.5.1   Simple Univariate Penalty Terms

These penalty terms are used when a measurement is assumed to directly constrain a fit parameter. This type is normally used to constrain individual parameters such as average background rates, day/night asymmetries and systematic uncertainty parameters. They are simply expressed as

$$L = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_- + \sigma_+} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma(x)}\right]^2}, \quad \sigma(x) \equiv \begin{cases} \sigma_- & x < \mu \\ \sigma_+ & x \geq \mu \end{cases}, \tag{3.18}$$

where σ_- and σ_+ allow for asymmetrical uncertainties. Note that for all types of penalty terms presented in this section, μ refers to the peak of the probability density which does not correspond to its mean unless $\sigma_- = \sigma_+$. The previous expression simplifies to

$$LL = -\frac{1}{2}\left[\frac{x-\mu}{\sigma(x)}\right]^2 \tag{3.19}$$

for the log-likelihood term after dropping the constants.

3.5.2 Simple Multivariate Penalty Terms

These penalty terms are used instead of univariate terms when the external measurement provides correlations between the parameters. The general form for k correlated parameters is given by

$$L = \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1}(\vec{x}-\vec{\mu})}, \tag{3.20}$$

or

$$LL = -\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1}(\vec{x} - \vec{\mu}) \tag{3.21}$$

for the log-likelihood term after dropping the constants, where Σ is the covariance matrix and $|\Sigma|$ its determinant. For the two-dimensional case, this gives

$$L = \frac{1}{2\pi\sqrt{\sigma_x^2\sigma_y^2 - \text{cov}^2}} e^{-\frac{1}{2(\sigma_x^2\sigma_y^2 - \text{cov}^2)}\left[\sigma_y^2(x-\mu_x)^2 + \sigma_x^2(y-\mu_y)^2 - 2(x-\mu_x)(y-\mu_y)\text{cov}\right]}, \tag{3.22}$$

where cov is the covariance between the parameters x and y. The associated log-likelihood term is thus given by

$$LL = -\frac{1}{2(\sigma_x^2\sigma_y^2 - \text{cov}^2)}\left[\sigma_y^2(x - \mu_x)^2 + \sigma_x^2(y - \mu_y)^2 - 2(x - \mu_x)(y - \mu_y)\text{cov}\right]. \tag{3.23}$$

3.5.3   Penalty Terms on a Sum with Fixed Constraint

This type of penalty term is used to constrain a sum of events when the external measurement provides directly a number of events. It is expressed as

$$L = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_- + \sigma_+} e^{-\frac{1}{2}\left[\frac{\sum_i \alpha_i \nu_{i\exp}(\vec{\eta}) - \mu}{\sigma}\right]^2}, \quad \sigma \equiv \begin{cases} \sigma_- & \sum_i \alpha_i \nu_{i\exp}(\vec{\eta}) < \mu \\ \sigma_+ & \sum_i \alpha_i \nu_{i\exp}(\vec{\eta}) \geq \mu \end{cases}, \quad (3.24)$$

where α_i are fit parameters as defined in Section 3.1 and where $\nu_{i\exp}(\vec{\eta})$ are the expected Poisson parameter values for systematic uncertainty parameter values $\vec{\eta}$. The corresponding log-likelihood penalty term is given by

$$LL = -\frac{1}{2}\left[\frac{\sum_i \alpha_i \nu_{i\exp}(\vec{\eta}) - \mu}{\sigma}\right]^2. \quad (3.25)$$

3.5.4   Penalty Terms on a Sum with Floating Constraint

These penalty terms are used to constrain a sum of events when the external measurement provides rate measurements. They are notably used to constrain the number of fitted day and night events when not using a day/night asymmetry. In this particular case, the single external measurement is assumed to effectively provide an average background rate. This type of constraint is expressed as

$$L = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma_{r-} + \sigma_{r+}} e^{-\frac{1}{2}\left[\frac{\left(\frac{\sum_i \alpha_i \nu_{i\exp}(\vec{\eta})}{\sum_i \nu_{i\exp}(\vec{\eta})}\right) - \mu_r}{\sigma_r}\right]^2}, \quad (3.26)$$

with

$$\sigma_r \equiv \begin{cases} \sigma_{r-} & \sum_i \alpha_i \nu_{i\exp}(\vec{\eta}) < \mu_r \sum_i \nu_{i\exp}(\vec{\eta}) \\ \sigma_{r+} & \sum_i \alpha_i \nu_{i\exp}(\vec{\eta}) \geq \mu_r \sum_i \nu_{i\exp}(\vec{\eta}) \end{cases}, \tag{3.27}$$

where μ_r and σ_r are the relative mean and variance, respectively, and where α_i and $\nu_{i\exp}(\vec{\eta})$ parameters are defined as before. μ_r is usually 1. The associated log-likelihood term is given by

$$LL = -\frac{1}{2} \left[\frac{\left(\frac{\sum_i \alpha_i \nu_{i\exp}(\vec{\eta})}{\sum_i \nu_{i\exp}(\vec{\eta})} \right) - \mu_r}{\sigma_r} \right]^2. \tag{3.28}$$

3.6 PSA Constraint

The PSA provides a penalty on the total number of neutrons detected by the NCD during the NCD phase. The parameters provided by the PSA are the measured total number of neutrons passing the PSA cuts (n_{PSA}) along with an uncertainty ($\sigma_{n\mathsf{PSA}}$) and the PSA cut efficiency ($\epsilon_{\mathsf{PSA}} \pm \sigma_{\epsilon\mathsf{PSA}}$). This efficiency can be viewed as a conversion factor between the number of neutrons detected by the NCDs in the energy window of the spectral analysis and the number of neutrons passing the PSA cut. The equivalent number of neutrons for a spectral analysis, n_n, is thus given by

$$n_n = \frac{n_{\mathsf{PSA}}}{\epsilon_{\mathsf{PSA}}}. \tag{3.29}$$

If one wants to penalise n_n, the PSA uncertainties can be combined to give a relative uncertainty δ_{nn} on this parameter, given by

$$\delta_{nn} \equiv \frac{\sigma_{nn}}{n_n} = \frac{1}{n_n}\sqrt{\left(\frac{\partial n_n}{\partial n_{\mathsf{PSA}}}\right)^2 \sigma_{n\mathsf{PSA}}^2 + \left(\frac{\partial n_n}{\partial \epsilon_{\mathsf{PSA}}}\right)^2 \sigma_{\epsilon\mathsf{PSA}}^2}$$
$$= \sqrt{\delta_{n\mathsf{PSA}}^2 + \delta_{\epsilon\mathsf{PSA}}^2}, \tag{3.30}$$

where

$$\delta_{n\mathsf{PSA}} \equiv \frac{\sigma_{n\mathsf{PSA}}}{n_{\mathsf{PSA}}} \tag{3.31}$$

and

$$\delta_{\epsilon\mathsf{PSA}} \equiv \frac{\sigma_{\epsilon\mathsf{PSA}}}{\epsilon_{\mathsf{PSA}}}, \tag{3.32}$$

that is to say the relative uncertainties on n_{PSA} and ϵ_{PSA}, respectively. Since the PSA provides a number of detected events and not a constraint on a generated rate, its associated penalty term is thus a particular case of a constraint on a sum when using a fixed constraint, as shown in Section 3.5.3. Explicitly, it is

$$LL_{\mathsf{PSA}} = -\frac{1}{2}\left(\frac{\left[\sum_{j\in\{n \text{ classes}\}} \alpha_j \nu_{j\,\mathrm{exp}}(\vec{\eta})\right] - n_n}{n_n \delta_{nn}}\right)^2, \tag{3.33}$$

where $\nu_{j\,\mathrm{exp}}$ are the expected Poisson parameters for the number of neutron events detected by the NCDs and α_j are the fit parameters for all the classes of events representing neutrons.

3.7   Day/Night Background Asymmetries

In the context of a day/night analysis, it is important to allow for day/night asymmetries of the backgrounds when such fluctuations are possible. For example, due to the circulation of the water in the SNO detector, it is possible that D_2O and H_2O background rates vary between day and night. Some measurements of these asymmetries have thus been performed. The day/night asymmetry on a background rate is defined as

$$A \equiv 2\frac{r_n - r_d}{r_n + r_d}, \tag{3.34}$$

where r_d and r_n are the background rates for day and night, respectively. If the average background rate is defined as

$$r \equiv \frac{r_n + r_d}{2}, \tag{3.35}$$

the day/night background rates can be expressed as

$$r_n = r\left(1 + \frac{A}{2}\right). \tag{3.36}$$

$$r_d = r\left(1 - \frac{A}{2}\right). \tag{3.37}$$

The day/night Poisson parameters are thus computed using

$$\nu_n = r_n f_n(\vec{\eta}) \tag{3.38}$$

$$\nu_d = r_d f_d(\vec{\eta}), \tag{3.39}$$

where $f_n(\vec{\eta})$ and $f_d(\vec{\eta})$ are the rate-to-events conversion factors that depend on the systematic uncertainty parameters $\vec{\eta}$. These factors can be computed using

$$f_{d,n}(\vec{\eta}) = \frac{N_{d,n}(\vec{\eta})}{R_{\mathsf{MC}}}, \tag{3.40}$$

where R_{MC} is the Monte Carlo rate and where $N_{d,n}(\vec{\eta})$ is the number of day/night Monte Carlo events in the analysis window when it is smeared by the current systematic uncertainties. Given the fit parameters r, A and $\vec{\eta}$, the night and day Poisson parameters are thus given respectively by

$$\nu_n = r\left(1 + \frac{A}{2}\right)\frac{N_n(\vec{\eta})}{R_{\mathsf{MC}}} \tag{3.41}$$

$$\nu_d = r\left(1 - \frac{A}{2}\right)\frac{N_d(\vec{\eta})}{R_{\mathsf{MC}}}. \tag{3.42}$$

As the nominal number of background events in the data (number of expected events in the analysis window obtained using default values for the systematic uncertainty parameters) is computed using

$$\nu_{\mathsf{nom}} = \frac{r_{\mathsf{nom}}}{R_{\mathsf{MC}}}N_{\mathsf{MC}}, \tag{3.43}$$

where r_{nom} is the nominal background data rate (a value that usually corresponds to the most likely value from an external measurement) and where N_{MC} is the number of Monte Carlo background events in the analysis window when using the default systematic uncertainty parameters, R_{MC} can be thus expressed using the constants r_{nom} and ν_{nom} along with the value of N_{MC}. Redefining the fit parameter associated to the average rate as

$$\alpha = \frac{r}{r_{\mathsf{nom}}}, \tag{3.44}$$

Equations (3.41), (3.42), (3.43) and (3.44) finally give

$$\nu_n = \alpha \frac{\nu_{\text{nom}}}{N_{\text{MC}}} \left(1 + \frac{A}{2} \right) N_n(\vec{\eta}) = \alpha \frac{N_n(\vec{\eta})}{M_{\text{MC}}} \left(1 + \frac{A}{2} \right) \tag{3.45}$$

$$\nu_d = \alpha \frac{\nu_{\text{nom}}}{N_{\text{MC}}} \left(1 - \frac{A}{2} \right) N_d(\vec{\eta}) = \alpha \frac{N_d(\vec{\eta})}{M_{\text{MC}}} \left(1 - \frac{A}{2} \right), \tag{3.46}$$

where M_{MC} is defined as the Monte Carlo multiplicity with respect to ν_{nom} and where α has a value of 1 for a fit returning an average background rate corresponding to the rate from the external measurement.

Under the assumption that external measurements for backgrounds effectively constrain the background rates and not the number of PMT events inside the analysis window, α and A should be constrained directly rather than by applying constraints on numbers of events that depend on the systematic uncertainty parameters $\vec{\eta}$.

3.8 Evaluation of the Finite Monte Carlo Statistics Uncertainties

As mentioned in Section 3.3, the binned PDFs used in the maximum likelihood fit are commonly generated from Monte Carlo samples. Ideally, a quasi-infinite number of Monte Carlo events would be produced to generate these PDFs, such that the statistical uncertainty on the shape of these PDFs tend to zero. However, computing power restrictions when generating Monte Carlo samples and also when smearing the resulting events with systematic uncertainties during the fit limit the statistics that can be handled. Although

the resulting PDFs, within the known systematic uncertainties, are considered to represent the "true" distributions, there are some uncertainties associated with the finite Monte Carlo statistics used to generate them. That can effectively translate into additional systematic uncertainties. The relative importance of these uncertainties grows as the number of bins in the PDFs increases, particularly when adding an extra dimension. The effects are also larger when the Monte Carlo statistics are split into sub-classes, as in day/night analyses. Finally, the finite Monte Carlo statistics uncertainties are expected to increase when the Monte Carlo is smeared by systematic shifts which increase observable distribution widths since the statistics become distributed over a larger number of bins. These uncertainties should thus be evaluated after all the systematic shifts have been applied.

The finite Monte Carlo statistics also have a second-order effect because the flux-to-event and the rate-to-event factors that are used to provide the expected number of events in the likelihood function are also computed from Monte Carlo samples. This comes from the fact that external measurements provide rate constraints and that Monte Carlo is thus required to convert these rates into the actual number of events in the analysis window. Direct usage of the number of Monte Carlo events in the likelihood is shown in Equation (3.17). Since the number of Monte Carlo events in the analysis follows Poisson statistics, there is thus an effective uncertainty on the conversion factors given a finite number of Monte Carlo events. This has virtually no effect for independent classes of background events that are not constrained by a penalty term, but it can have an effect for tightly constrained classes with low Monte Carlo statistics.

In order to assess these two aspects of the finite Monte Carlo statistics' effects, a technique has been developed to evaluate the resulting uncertainties. It consists of performing

multiple minimisations of the likelihood function by Poisson-fluctuating the Monte Carlo samples around what is assumed to generate the "true" distributions, that is to say the whole Monte Carlo samples smeared with all the systematic shifts. The procedure is the following:

1. The whole Monte Carlo is read and smeared by all the fitted systematic shifts.

2. The signal extraction cuts are applied.

3. Multi-dimensional histograms $H(b|j, \vec{\eta}_j)$ are defined using all of the variables and the finest binning used by all of the PDFs and filled with the smeared Monte Carlo.

4. All of the required PDFs are generated by properly projecting the multi-dimensional histogram and by normalising it when necessary.

5. A new fit is performed and the resulting fit values are assumed to be the central values for the evaluation of the effects of finite Monte Carlo statistics.

6. New (empty) histograms $H'(b|j, \vec{\eta}_j)$ are created using the same binning as H.

7. A histogram H' is filled using Poisson fluctuations of H, meaning that it is filled using Poisson random variables v_b ($H'(b|j, \vec{\eta}_j) = v_b$) whose associated Poisson parameters ν_b are given by the bin content of $H(b|j, \vec{\eta}_j)$. H' has thus the following characteristics:

   - $E\left[H'(b|j, \vec{\eta}_j)\right] = H(b|j, \vec{\eta}_j)$.

   - $\mathrm{Var}\left[H'(b|j, \vec{\eta}_j)\right] = H(b|j, \vec{\eta}_j)$.

8. New PDFs are generated using the appropriate projections of H' and by normalising them, when required.

9. New flux-to-event or rate-to-event factors are effectively computed in the likelihood function due to their integration in the function via unnormalised histograms as shown in Equation (3.17).

10. The fit is performed using the PDFs created from H'.

11. The differences between the fit values obtained when using H and H' are stored.

12. Steps 6 to 11 are repeated 1000 times, each time using a different seed for the Poisson random variables.

13. Using an unbinned log-likelihood, an asymmetrical Gaussian function is fitted to the set of fit parameter shifts (with respect to the central fit) for every fit parameter. The peak of this function is forced to be at zero. If all of the fit values obtained from H' are either smaller or greater than the ones obtained initially with H, the RMS around 0 is computed using the one-sided shifts and the uncertainty on the other side is approximated to be the same.

Chapter 4

Signals and the SNO Likelihood Function

In the previous chapter, an optimised form of a binned extended log-likelihood function, parameterised for the extraction of production rates, was derived. Although this statistical model is perfectly sufficient for the SNO backgrounds and the NC signal, it needs to be extended for the other signals. If solar neutrinos are mixing on their way to the SNO detector, this can affect CC and ES signals, since the survival probability of a given neutrino flavour can become a function of the neutrino energy and these two signals have different cross-sections for the different neutrino flavours. Integral rates, as previously used in the expression of the likelihood function, thus cannot model CC and ES signals properly.

4.1 Unconstrained Reconstructed Electron Energy Fit

One way to account for neutrino spectral distortion in the signal extraction procedure is to perform a binned, unconstrained fit of the reconstructed electron energy, since this

observable is the most sensitive probe of the neutrino energy spectrum. In a binned, unconstrained fit, each electron energy bin is allowed to float as a free parameter.

4.1.1 Derivation

For an unconstrained fit in reconstructed electron energy, the factors $\alpha_j \nu_{j_{\exp}}(\vec{\eta}_j) f(\vec{x}|j)$ of Equation (3.10), with $j \in \{\mathsf{CC}, \mathsf{ES}\}$, can be expressed as

$$
\begin{aligned}
\alpha_j \nu_{j_{\exp}} f(\vec{x}|j, \vec{\phi}_j) &= \alpha_j \nu_{j_{\exp}} \sum_{k=1}^{n_{\mathrm{unconstr}}} f(\vec{x}, k|j, \vec{\phi}_j) \\
&= \alpha_j \nu_{j_{\exp}} \sum_{k=1}^{n_{\mathrm{unconstr}}} P(k|j, \vec{\phi}_j) f(\vec{x}|j, k),
\end{aligned}
\tag{4.1}
$$

where n_{unconstr} is the number of reconstructed electron energy bins used for the unconstrained fit and where $P(k|j, \vec{\phi}_j)$ is the probability of an event of class j to be in energy bin k, given a statistical model with a spectral shape defined by $\vec{\phi}_j$ parameters. $P(k|j, \vec{\phi}_j)$ can be expressed as

$$
P(k|j, \phi_{jk}) = \phi_{jk} P(k|j),
\tag{4.2}
$$

where $P(k|j)$ is the corresponding probability in the case of an undistorted spectrum and where ϕ_{jk} are parameters whose values are unity for the undistorted case and which must satisfy

$$
\sum_{k=1}^{n_{\mathrm{unconstr}}} \phi_{jk} P(k|j) = 1.
\tag{4.3}
$$

This constraint can be included by defining $\alpha_{jk} \equiv \alpha_j \phi_{jk}$, where α_{jk} are not subject to such a constraint. This gives for Equation (4.1)

$$\alpha_j \nu_{j_{\text{exp}}} f(\vec{x}|j, \vec{\phi}_j) = \sum_{k=1}^{n_{\text{unconstr}}} \alpha_{jk} \nu_{j_{\text{exp}}} P(k|j) f(\vec{x}|j, k) \tag{4.4}$$

$$= \sum_{k=1}^{n_{\text{unconstr}}} \alpha_{jk} \nu_{jk_{\text{exp}}} f(\vec{x}|j, k), \tag{4.5}$$

where $\nu_{jk_{\text{exp}}} \equiv \nu_{j_{\text{exp}}} P(k|j)$. If Equation (4.5) is substituted in Equation (3.10), it appears that doing an unconstrained fit for CC or ES is equivalent to fitting for n_{unconstr} mutually independent classes of events with PDFs $f(\vec{x}|j, k)$ and Poisson parameters $\alpha_{jk} \nu_{jk_{\text{exp}}}$.

Due to limited statistics in the Monte Carlo samples and in order to limit the number of parameters in the fit, it can be desired to generate a number of conditional PDFs $f(\vec{x}|j, k)$ different from the number of bins in reconstructed electron energy used for NC and backgrounds. However, to avoid biased fit parameters, it is required to use the same binning in reconstructed electron energy for all signal and background PDFs. A different energy binning with index m is thus defined to represent the binning common to all signals and backgrounds and $f(\vec{x}|j, k)$ is expressed as

$$f(\vec{x}|j, k) = \sum_{m \subseteq k} f(\vec{x}, m|j, k) = \sum_{m \subseteq k} P(m|j, k) f(\vec{x}|j, k, m). \tag{4.6}$$

The above expression implies that bins with index k contain integer numbers of bins with index m. If $\vec{y}$ is defined such that $\vec{x} = (T_{\mathrm{eff}}, \vec{y})$, Equation (4.6) gives

$$f(\vec{x}|j,k) = \sum_{m \subseteq k} P(m|j,k)f(T_{\mathrm{eff}}|j,k,m)f(\vec{y}|j,k,m,T_{\mathrm{eff}}) \tag{4.7}$$

$$= \sum_{m \subseteq k} f(T_{\mathrm{eff}},m|j,k)f(\vec{y}|j,k,m,T_{\mathrm{eff}}). \tag{4.8}$$

Since the PDFs in the above equation are binned PDFs generated from Monte Carlo samples, Equation (4.8) can be simplified. Here the PDF for T_{eff} is binned using bins with index m, such that $f(T_{\mathrm{eff}},m|j,k)$ can be expressed as

$$f(T_{\mathrm{eff}},m|j,k) = f(T_{\mathrm{eff}}|j,k)P(m|j,k,T_{\mathrm{eff}})$$

$$= f(T_{\mathrm{eff}}|j,k)\delta_{T_{\mathrm{eff}} \in m} \overset{\mathrm{binning}}{\approx} f_b(m|j,k)\delta_{T_{\mathrm{eff}} \in m}. \tag{4.9}$$

Regarding $f(\vec{y}|j,k,m,T_{\mathrm{eff}})$, it is generated from the events in the energy bin k, leading to the approximation

$$f(\vec{y}|j,k,m,T_{\mathrm{eff}}) \overset{\mathrm{binning}}{\approx} f(\vec{y}|j,k). \tag{4.10}$$

From Equations (4.9) and (4.10), Equation (4.8) can be rewritten as

$$f(\vec{x}|j,k) = \sum_{m \subseteq k} f_b(m|j,k)f(\vec{y}|j,k)\delta_{T_{\mathrm{eff}} \in m}. \tag{4.11}$$

From Equations (4.1) and (4.11),

$$\alpha_j \nu_{j\,\text{exp}} f(\vec{x}|j, \vec{\phi}_j) = \sum_{k=1}^{n_{\text{unconstr}}} \alpha_{jk} \nu_{j\,\text{exp}} P(k|j) \sum_{m \subseteq k} f_b(m|j, k) f(\vec{y}|j, k) \delta_{T_{\text{eff}} \in m}$$

$$= \sum_{k=1}^{n_{\text{unconstr}}} \sum_{m \subseteq k} \alpha_{jk} \nu_{j\,\text{exp}} f_b(m|j) f(\vec{y}|j, k) \delta_{T_{\text{eff}} \in m}$$

$$= \alpha_{j,k(T_{\text{eff}})} \nu_{j\,\text{exp}} f_b(m(T_{\text{eff}})|j) f(\vec{y}|j, k(T_{\text{eff}})), \tag{4.12}$$

where $k(T_{\text{eff}}) \equiv k : T_{\text{eff}} \in k$ and $m(T_{\text{eff}}) \equiv m : T_{\text{eff}} \in m, \ m \subseteq k$. Variables $k(T_{\text{eff}})$ and $m(T_{\text{eff}})$ thus refer to the indices that correspond to the value of T_{eff} defined in $\vec{x}$.

4.1.2 Summary

For the event classes $j \in \{\text{CC}, \text{ES}\}$, the product of the Poisson parameter with the PDF $f(\vec{x}|j)$ can be approximated by

$$\alpha_j \nu_{j\,\text{exp}}(\vec{\eta}_j) f(\vec{x}|j, \vec{\eta}_j) = \alpha_{j,k(T_{\text{eff}})} \nu_{j\,\text{exp}}(\vec{\eta}_j) f_b(m(T_{\text{eff}})|j, \vec{\eta}_j) f(\vec{y}|j, k(T_{\text{eff}}), \vec{\eta}_j), \tag{4.13}$$

where $f_b(m(T_{\text{eff}})|j, \vec{\eta}_j)$ is the undistorted reconstructed electron energy binned PDF for class j with systematic shifts $\vec{\eta}_j$, whose bins are defined such that $m(T_{\text{eff}}) \subseteq k(T_{\text{eff}})$, and where $f(\vec{y}|j, k(T_{\text{eff}}), \vec{\eta}_j)$ is the PDF in the observable space $\vec{y}$ for the unconstrained energy bin k of class j with systematic shifts $\vec{\eta}_j$. The fit parameters α_{jk} are defined such that the fitted Poisson parameter for the unconstrained energy bin k of class j in the model is given by

$$\nu_{jk}(\alpha_{jk}, \vec{\eta}_j) \equiv \alpha_{jk} \nu_{jk\,\text{exp}}(\vec{\eta}_j) = \nu_{j\,\text{exp}}(\vec{\eta}_j) \alpha_{jk} P(k|j, \vec{\eta}_j), \tag{4.14}$$

where $\nu_{j_{\text{exp}}}(\vec{\eta}_j)$ is the expected number of detected j events, given systematic shifts $\vec{\eta}_j$ and where $P(k|j, \vec{\eta}_j)$ is the PMF of the undistorted unconstrained energy bin for the associated class with systematic shifts $\vec{\eta}_j$. The Poisson parameter for class j in the model is given by

$$\nu_j(\vec{\alpha}_j, \vec{\eta}_j) = \sum_{k=1}^{n_{\text{unconstr}}} \nu_{jk}(\alpha_{jk}, \vec{\eta}_j), \tag{4.15}$$

where

$$\vec{\alpha}_j \equiv (\alpha_{j1}, \ldots, \alpha_{jn_{\text{unconstr}}}). \tag{4.16}$$

4.2   Fits using Neutrino Energy Spectrum Distortion

Although an unconstrained fit in reconstructed electron energy allows to perform a three-phase fit in a way that does not depend on the oscillation model while using only the observable space distributions, it can be more difficult to apply it in a multi-phase fit where the detector response can vary between the phases for a given signal class. In addition, when the ultimate goal of an analysis is to measure the parameters of a given oscillation model, the large number of parameters in this type of fit can lead to an increase of the statistical uncertainties. One way to try avoiding these problems is to work in neutrino energy space. Starting from a space that includes both the observable space and the neutrino energy, one can retrieve the observable space distribution for event class j by simply using the projection

$$f(\vec{x}|j, \varphi_j(E_\nu)) = \int_{-\infty}^{+\infty} f(E_\nu, \vec{x}|j, \varphi_j(E_\nu)) dE_\nu \tag{4.17}$$

$$= \int_{-\infty}^{+\infty} f(E_\nu|j, \varphi_j(E_\nu)) f(\vec{x}|j, E_\nu) dE_\nu, \tag{4.18}$$

where the function $\varphi_j(E_\nu)$ parameterises the distortion of the E_ν spectrum for class j. In the above equation, the identity $f(\vec{x}|j, \varphi_j(E_\nu), E_\nu) = f(\vec{x}|j, E_\nu)$ was used. Here $\varphi_j(E_\nu)$ can be defined such that $f(E_\nu|j, \varphi_j(E_\nu))$ is expressed as

$$f(E_\nu|j, \varphi_j(E_\nu)) = \varphi_j(E_\nu)f(E_\nu|j), \tag{4.19}$$

where $f(E_\nu|j)$ is the corresponding E_ν PDF in the case of an undistorted spectrum and φ_j are parameters whose values are unity for the undistorted case. Finally, from Equations (4.18) and (4.19),

$$f(\vec{x}|j, \varphi_j(E_\nu)) = \int_{-\infty}^{+\infty} \varphi_j(E_\nu)f(E_\nu|j)f(\vec{x}|j, E_\nu)dE_\nu \tag{4.20}$$

$$= \int_{-\infty}^{+\infty} \varphi_j(E_\nu)f(E_\nu, \vec{x}|j)dE_\nu \tag{4.21}$$

4.2.1 Binned Neutrino Energy Fit

In a binned neutrino energy fit, E_ν is binned using n_ν bins in neutrino energy with index l, leading to the approximations

$$f(E_\nu|j) \overset{\text{binning}}{\approx} f_b(l(E_\nu)|j) \tag{4.22}$$

$$f(\vec{x}|j, E_\nu) \overset{\text{binning}}{\approx} f(\vec{x}|j, l(E_\nu)). \tag{4.23}$$

Substituting into Equation (4.20), this gives

$$
\begin{aligned}
f(\vec{x}|j, \varphi_j(E_\nu)) &= \int_{-\infty}^{+\infty} \varphi_j(E_\nu) f_b(l(E_\nu)|j) f(\vec{x}|j, l(E_\nu)) dE_\nu \\
&= \sum_{l=1}^{n_\nu} \left[\int_l \varphi_j(E_\nu) dE_\nu \right] f_b(l|j) f(\vec{x}|j, l) \\
&= \sum_{l=1}^{n_\nu} \frac{1}{\Delta_{E_\nu}(l)} \left[\int_l \varphi_j(E_\nu) dE_\nu \right] P(l|j) f(\vec{x}|j, l),
\end{aligned}
\qquad (4.24)
$$

where $\Delta_{E_\nu}(l)$ is the bin width of the neutrino energy bin l. If the parameters ϕ_{jl} are defined as

$$
\phi_{jl} \equiv \frac{1}{\Delta_{E_\nu}(l)} \left[\int_l \varphi_j(E_\nu) dE_\nu \right],
\qquad (4.25)
$$

which have an expectation value of 1 for an undistorted spectrum and which, from Equation (4.19), have to satisfy the constraint

$$
\sum_{l=1}^{n_\nu} \phi_{jl} P(l|j) = 1,
\qquad (4.26)
$$

the product of the Poisson parameter with the observable space PDF $f(\vec{x}|j, \varphi_j(E_\nu))$ can be expressed as

$$
\alpha_j \nu_{j\,\text{exp}} f(\vec{x}|j, \varphi_j(E_\nu)) = \sum_{l=1}^{n_\nu} \alpha_j \phi_{jl} \nu_{j\,\text{exp}} P(l|j) f(\vec{x}|j, l),
\qquad (4.27)
$$

where the PDFs $f(\vec{x}|j, l)$ can be approximated using a similar configuration as the PDF $f(\vec{x}|j, \vec{\phi}_j)$ in Section 4.1. From Equation (4.13), the undistorted $f(\vec{x}|j, \vec{\phi}_j)$ PDF in reconstructed electron energy is given by

$$
f(\vec{x}|j, \vec{\eta}_j) = f_b(m(T_{\text{eff}})|j, \vec{\eta}_j) f(\vec{y}|j, k(T_{\text{eff}}), \vec{\eta}_j).
\qquad (4.28)
$$

The conditional PDF in neutrino energy $f(\vec{x}|j, l)$ can thus be written as

$$f(\vec{x}|j, l, \vec{\eta}_j) = f_b(m(T_{\mathsf{eff}})|j, l, \vec{\eta}_j) f(\vec{y}|j, l, k(T_{\mathsf{eff}}), \vec{\eta}_j) \qquad (4.29)$$

and the product of the Poisson parameter with $f(\vec{x}|j, \varphi_j(E_\nu))$ becomes

$$\alpha_j \nu_{j\,\mathsf{exp}} f(\vec{x}|j, \varphi_j(E_\nu)) = \alpha_j \nu_{j\,\mathsf{exp}} \sum_{l=1}^{n_\nu} \phi_{jl} P(l|j) f_b(m(T_{\mathsf{eff}})|j, l) f(\vec{y}|j, l, k(T_{\mathsf{eff}})). \quad (4.30)$$

4.2.2   Unconstrained Neutrino Energy Fit

In an unconstrained neutrino energy fit, the bins in neutrino energy are fitted independently by defining a set of parameters $\alpha_{jl} \equiv \alpha_j \phi_{jl}$ as it was done in Section 4.1. The product of the Poisson parameter with the PDF $f(\vec{x}|j, \vec{\alpha}_j)$ in Equation (4.27) thus becomes

$$\begin{aligned}
\alpha_j \nu_{j\,\mathsf{exp}} f(\vec{x}|j, \vec{\alpha}_j) &= \sum_{l=1}^{n_\nu} \alpha_{jl} \nu_{j\,\mathsf{exp}} P(l|j) f(\vec{x}|j, l) \\
&= \sum_{l=1}^{n_\nu} \alpha_{jl} \nu_{jl_{\mathsf{exp}}} f(\vec{x}|j, l),
\end{aligned} \qquad (4.31)$$

where $\nu_{jl_{\mathsf{exp}}} \equiv \nu_{j_{\mathsf{exp}}} P(l|j)$ is the expected Poisson parameter for the neutrino energy bin l in signal j. The above equation has the exact same form as Equation (4.5), such that doing an unconstrained neutrino energy fit for CC or ES is equivalent to fitting for n_ν mutually independent classes of events with PDFs $f(\vec{x}|j, l)$ and Poisson parameters

$\alpha_{jl}\nu_{jl_{\text{exp}}}$. Finally, Equation (4.30) gives

$$\alpha_j \nu_{j_{\text{exp}}}(\vec{\eta}_j) f(\vec{x}|j, \vec{\alpha}_j, \vec{\eta}_j)$$

$$= \nu_{j_{\text{exp}}}(\vec{\eta}_j) \sum_{l=1}^{n_\nu} \alpha_{jl} P(l|j, \vec{\eta}_j) f_b(m(T_{\text{eff}})|j, l, \vec{\eta}_j) f(\vec{y}|j, l, k(T_{\text{eff}}), \vec{\eta}_j)$$

$$= \sum_{l=1}^{n_\nu} \nu_{jl}(\alpha_{jl}, \vec{\eta}_j) f_b(m(T_{\text{eff}})|j, l, \vec{\eta}_j) f(\vec{y}|j, l, k(T_{\text{eff}}), \vec{\eta}_j), \qquad (4.32)$$

where $\nu_{jl}(\alpha_{jl}, \vec{\eta}_j) \equiv \alpha_{jl}\nu_{j_{\text{exp}}}(\vec{\eta}_j) P(l|j, \vec{\eta}_j)$ are the Poisson parameters for the unconstrained neutrino energy bin l for class j.

The main advantage of this method over an unconstrained fit in reconstructed electron energy is to allow a multi-phase fit when the detector response is different across the different phases. However, this technique requires the sub-binning of the MC samples in neutrino energy and, contrary to the binning of an observable, this can lead to biases of the fit parameters.

Effectively, when the observables are binned for all MC samples and for the data using the same binning, this reduces the statistical separation power between the different classes of events, increasing the statistical uncertainties of the fit, but it normally does not create biases, as long as the simulation samples contain sufficient statistics and model the data perfectly. Limited MC statistics can however motivate the use of coarser bins, so the binning of observables should be chosen such that the distortions due to the limited statistics be negligible, or such that the total uncertainties on the fit parameters, including systematic uncertainties due to the statistics in the PDFs, be minimal.

To bin an experimentally unobserved variable, however, effectively affects how well

the simulation can mimic the data, which translates automatically into biases of the fit parameters if the true distribution of the variable is different from the one of the MC events. To minimise these biases, it is thus preferable to use bins that are as narrow as possible for the unobserved variable when binning cannot be avoided. It can be quite challenging to achieve this using the type of fit described above in this section due to the high dimensionality of the PDFs in Equation (4.32). Even if the simulation statistics would allow it, the increased number of bins would translate into a higher number of fit parameters having strong correlations and to an increased of the statistical uncertainties.

4.2.3   Unbinned P_{ee} Polynomial Fit

It is actually possible to avoid most of the problems described above by aiming to fit directly the neutrino energy spectrum distortion function $\varphi_j(E_\nu)$. $\varphi_j(E_\nu)$ is a simple function of the electron neutrino survival probability $P_{ee}(E_\nu)$, which is expected to be smooth and well-behaved, when averaged over position and time, such that it can be approximated quite well using a low-order polynomial expansion.

Let $g_j(E_\nu, P_{ee})$ be a weighting function equal to a parameterised P_{ee} function in the case of CC and ES$_e$ and equal to $1 - P_{ee}$ in the case of ES$_{\mu\tau}$. Given the function $g_j(E_\nu, P_{ee})$, the PDF $f(\vec{x}|j, P_{ee}, \vec{\eta}_j)$ can be generated directly by filling a histogram $H(\vec{x}|j, P_{ee}, \vec{\eta}_j)$ using a weighting of the simulated observable values by $g_j(E_\nu, P_{ee})$, and by normalising it appropriately. The Poisson parameter for class j is simply approximated by

$$\alpha_j \nu_{j_{\exp}}(\vec{\eta}_j) \approx \alpha_\text{B} \nu_{j_{\exp}} \frac{\sum_{i'=1}^{N_\text{MC}} g_j(E_{\nu i'}, P_{ee})}{N_\text{MC}}, \tag{4.33}$$

where α_B is a fit parameter for the ^{8}B flux scale and which is equivalent to α_NC for models that do not involve sterile neutrinos, N_MC is the number of simulated events and $E_{\nu i'}$ is the neutrino energy for the simulated event i'.

An unbinned P_{ee} polynomial fit thus allows to fit for the energy spectral distortion while keeping the same PDF dimensionality and configuration as the PDFs used for the unconstrained reconstructed electron energy fit. It also allows to correctly handle a multi-phase fit, uses a reduced number of fit parameters and constrains the relative number of events from the different signal classes, which should considerably reduce the statistical uncertainties. This method is however affected by a major drawback: it requires to regenerate the PDFs using all simulated events as soon as the value of one of the P_{ee} polynomial parameters is modified during the minimisation process, which slows down the fit by many orders of magnitude, to an unpractical level.

4.2.4   Binned P_{ee} Fit

As described in Section 4.2.2, the high correlations between the fit parameters along with the presence of large statistical uncertainties constitute issues with the unconstrained neutrino energy fit when using many neutrino energy bins. One way to improve the statistical separation between the fitted parameters is to fit for α_B and the shape of a binned P_{ee} function.

In the previous section, it was shown that $g_j(E_\nu, P_{ee})$ can be used as weighting functions to generate the distorted neutrino energy spectra. Comparing with Equation (4.19), it is obvious that there exists a directly proportional relation between $g_j(E_\nu, P_{ee})$ and $\varphi_j(E_\nu)$. To ensure the normalisation of both PDFs in Equation (4.19) for any P_{ee} shape,

$\varphi_j(E_\nu)$ has in fact to be defined as

$$\varphi_j(E_\nu) \equiv \frac{g_j(E_\nu, P_{ee})}{\int_{-\infty}^{+\infty} g_j(E_\nu, P_{ee}) f(E_\nu | j) dE_\nu} \tag{4.34}$$

with respect to $g_j(E_\nu, P_{ee})$, such that ϕ_{jl} from Equation (4.25) becomes

$$\phi_{jl} = \frac{1}{\Delta_{E_\nu}(l)} \frac{\left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right]}{\int_{-\infty}^{+\infty} g_j(E_\nu, P_{ee}) f(E_\nu | j) dE_\nu}. \tag{4.35}$$

In a P_{ee} fit, the values of α_j for CC, ES and NC in Equation (3.10) are not allowed to vary freely in the fit, as mentioned in the previous section. They are constrained by the ^{8}B flux and P_{ee}. For CC and ES signal classes, the expression for α_j as a function of α_B and P_{ee} is by definition

$$\alpha_j = \alpha_B \int_{-\infty}^{+\infty} g_j(E_\nu, P_{ee}) f(E_\nu | j) dE_\nu. \tag{4.36}$$

From Equations (4.30), (4.35) and (4.36), the product of the Poisson parameter with $f(\vec{x}|j)$, in the context of the current parameterisation, can be written as

$\alpha_j \nu_{j_{\text{exp}}}(\vec{\eta}_j) f(\vec{x}|j, P_{ee}, \vec{\eta}_j)$

$$= \alpha_B \nu_{j_{\text{exp}}}(\vec{\eta}_j) \sum_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] \frac{P(l|j)}{\Delta_{E_\nu}(l)} f_b(m(T_{\text{eff}})|j, l, \vec{\eta}_j) f(\vec{y}|j, l, k(T_{\text{eff}}), \vec{\eta}_j) \right\}$$

$$= \alpha_B \nu_{j_{\text{exp}}}(\vec{\eta}_j) \sum_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] f_b(l|j) f_b(m(T_{\text{eff}})|j, l, \vec{\eta}_j) f(\vec{y}|j, l, k(T_{\text{eff}}), \vec{\eta}_j) \right\}. \tag{4.37}$$

The Poisson parameter for class j, found by integrating Equation (4.37) over the whole observable space, is given by

$$\alpha_j \nu_{j\,\text{exp}}(\vec{\eta}_j) = \alpha_\text{B} \nu_{j\,\text{exp}}(\vec{\eta}_j) \sum_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] f_b(l|j) \right\}. \qquad (4.38)$$

The integrals in both Equations (4.37) and (4.38) can be computed analytically very easily if P_{ee} is approximated by a polynomial (recall that $g_j(E_\nu, P_{ee})$ is related to a parameterised P_{ee} for j corresponding to CC and ES). One can also decide to fit for a binned P_{ee} function $g_j(l, P_{eel})$ defined as

$$g_j(l, P_{eel}) \equiv \frac{1}{\Delta_{E_\nu}(l)} \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right], \qquad (4.39)$$

with $1 \leq l \leq n_\nu$. This uses only the Equation (1.10) to increase the statistical separation, without assuming a smooth and continuous P_{ee} function. The P_{ee} shapes that can be mimicked by this variant are however limited by the number of P_{eel} parameters.

Although the method described in this section can combine the speed of the unconstrained neutrino energy fit and the smaller statistical uncertainties from the smaller number of fit parameters of the unbinned P_{ee} polynomial fit, it uses the same configuration of PDFs (and high dimensionality) as the former method, such that it is quite challenging to accumulate sufficient MC statistics in the PDFs. Coarse binning of an unobserved variable can also lead to biases in the fit parameters as explained earlier. To increase the number of neutrino energy bins would require either to reduce the number of bins for the observables, thus reducing the statistical separation power between the signals and backgrounds, or to reduce the number of reconstructed electron energy bins

with index k, affecting the mechanism through which the main correlations between the observables are taken into account.

4.2.5 $P_{ee} + P_{ea}$ Fit

As mentioned in the previous section, one of the advantages of a P_{ee} fit is the improvement of statistical separation compared to an unconstrained fit, due to the reduced number of fit parameters. It is however done at the cost of a more model-dependent approach, where neutrinos are assume to oscillate only to other active neutrino flavours (ν_e, ν_μ and ν_τ). A constraint introduced by the P_{ee} fit is the assumption of a model where the ratios of the predicted number of events for CC, ES$_e$, ES$_{\mu\tau}$ and NC are solely determined by the P_{ee} function (excluding systematic uncertainties and ^{8}B spectral shape). This type of fit is thus not ideal for a model where there is one (or many) non-active neutrino flavour(s), since in this scenario a P_{ee} fit would be over-constrained and would be difficult to interpret, because the fit parameter values would have a non-physical meaning.

To avoid the problems described above, a slightly different type of fit based on a more general assumption can be achieved. Instead of using a model where $P_{es} = 0$, that is to say, where

$$P_{ee} + P_{ea} = 1, \tag{1.10}$$

with P_{ea} representing the probability of an electron neutrino converting to other active flavours (ν_μ, ν_τ), one can modify the fit to reflect a more general hypothesis with non-active flavour(s), with a probability of conversion from ν_e given by P_{es} such that

$$P_{ee} + P_{ea} + P_{es} = 1. \tag{1.11}$$

In such fit methods, the components of the likelihood associated to CC and ES_e are exactly the same as shown for the P_{ee} methods described in the previous sections, since these reactions only involve electron neutrinos. $ES_{\mu\tau}$, on the other hand, is by definition limited to the other active flavours. In the context of a $P_{ee} + P_{ea}$ fit, the number of $ES_{\mu\tau}$ events and its PDFs can thus be computed simply by using

$$g_{\mathsf{ES}_{\mu\tau}}(E_\nu, P_{ea}) \equiv P_{ea}(E_\nu) \qquad (4.40)$$

instead of $g_{\mathsf{ES}_{\mu\tau}}(E_\nu, P_{ee}) \equiv 1 - P_{ee}(E_\nu)$ as used in P_{ee} fits.

The main difference between $P_{ee} + P_{ea}$ and P_{ee} fits occurs in the treatment of NC. For P_{ee} fits, the product of the Poisson parameter with $f(\vec{x}|\mathsf{NC})$ is given by

$$\nu_{\mathsf{NC}} f(\vec{x}|\mathsf{NC}) = \alpha_{\mathsf{B}} \nu_{\mathsf{NCexp}} f(\vec{x}|\mathsf{NC}). \qquad (4.41)$$

The simplicity of this relation is due to the fact that the NC reaction involves equally all active neutrino flavours and is thus independent of the survival probability parameters if $P_{es} = 0$ because of the Equation (1.10). It is however not the case if non-active neutrino flavours exist. The expression of likelihood components associated to NC must now be determined similarly to the way it was done for the other signals, with a major simplification: the PDF $f(\vec{x}|\mathsf{NC})$ is totally independent of the survival probability parameters since the energy, direction and isotropy of the emitted photons resulting from this reaction are totally uncorrelated to the incoming neutrino parameters. Using Equation (4.36), the product of the Poisson parameter with the PDF for NC in the case of a $P_{ee} + P_{ea}$ fit is thus given by

$$\nu_{\mathsf{NC}} f(\vec{x}|\mathsf{NC}, \vec{\eta}_{\mathsf{NC}}) = \alpha_{\mathsf{B}} \nu_{\mathsf{NCexp}}(\vec{\eta}_{\mathsf{NC}}) \left\{ \int_{-\infty}^{\infty} g_{\mathsf{NC}}(E_\nu, P_{ee}, P_{ea}) f(E_\nu|\mathsf{NC}) dE_\nu \right\} f(\vec{x}|\mathsf{NC}, \vec{\eta}_{\mathsf{NC}}), \quad (4.42)$$

where $g_{\mathsf{NC}}(E_\nu, P_{ee}, P_{ea})$ is the probability of an electron neutrino to convert to an active flavour,

$$g_{\mathsf{NC}}(E_\nu, P_{ee}, P_{ea}) \equiv P_{ee}(E_\nu) + P_{ea}(E_\nu). \tag{4.43}$$

Here $f(E_\nu|\mathsf{NC})$ can be binned using n_ν bins in energy with index l, such that Equation (4.42) becomes

$$\nu_{\mathsf{NC}} f(\vec{x}|\mathsf{NC}, \vec{\eta}_{\mathsf{NC}}) = \alpha_{\mathsf{B}} \nu_{\mathsf{NC} \exp}(\vec{\eta}_{\mathsf{NC}}) \sum_{l=1}^{n_\nu} \left\{ \left[\int_l g_{\mathsf{NC}}(E_\nu, P_{ee}, P_{ea}) dE_\nu \right] f_b(l|\mathsf{NC}) \right\} f(\vec{x}|\mathsf{NC}, \vec{\eta}_{\mathsf{NC}}). \tag{4.44}$$

Finally, a binned fit can also be performed by defining $g_{\mathsf{NC}}(l, P_{eel}, P_{eal})$ as

$$g_{\mathsf{NC}}(l, P_{eel}, P_{eal}) \equiv P_{eel} + P_{eal} = \frac{1}{\Delta_{E_\nu}(l)} \left[\int_l g_{\mathsf{NC}}(E_\nu, P_{ee}, P_{ea}) dE_\nu \right], \tag{4.45}$$

with n_ν P_{eel} and n_ν P_{eal} fit parameters.

As it has been mentioned, a $P_{ee} + P_{ea}$ fit allows to correctly perform a three-phase fit for a case involving non-active neutrino flavours. However, this type of fit, as described so far, is under-constrained. Effectively, for a given neutrino energy, the fitted number of events for any SNO ^{8}B signal is proportional to the product between α_{B} and either P_{ee}, P_{ea} or $P_{ee} + P_{ea}$. In this case, P_{ee} and P_{ea} can thus always be redefined as

$$P'_{ee} = \frac{\alpha_{\mathsf{B}}}{\alpha'_{\mathsf{B}}} P_{ee} \tag{4.46}$$

$$P'_{ea} = \frac{\alpha_{\mathsf{B}}}{\alpha'_{\mathsf{B}}} P_{ea}, \tag{4.47}$$

such that the best fit values for P_{ee} and P_{ea} parameters and their uncertainties are simply proportional to the inverse of the ^{8}B scale. The correlations between the fit parameters, as extracted using any constant value for the ^{8}B scale, are consequently totally invariant

under a change of α_B. One can thus fix α_B during signal extraction without introducing any approximation at the signal extraction level or any complication for the interpretation.

4.3   Improved Binned Neutrino Energy Fit

Now if a step back is taken and the definition of conditional probability is used with Equation (4.21),

$$f(\vec{x}|j, \varphi_j(E_\nu)) = \int_{-\infty}^{+\infty} \varphi_j(E_\nu) f(E_\nu|j, \vec{x}) f(\vec{x}|j) dE_\nu$$

$$= f(\vec{x}|j) \int_{-\infty}^{+\infty} \varphi_j(E_\nu) f(E_\nu|j, \vec{x}) dE_\nu. \qquad (4.48)$$

If a subset $\vec{z}$ of the observables in the subset $\vec{y}$ has negligible correlations with the neutrino energy for a given reconstructed electron energy, then $f(E_\nu|j, \vec{x})$ can be rewritten as

$$f(E_\nu|j, \vec{x}) = f(E_\nu|j, T_{\text{eff}}, \vec{y}) = f(E_\nu|j, T_{\text{eff}}, \vec{z}, \vec{z'}) \approx f(E_\nu|j, T_{\text{eff}}, \vec{z'}), \qquad (4.49)$$

where $\vec{z'}$ is the subset of observables from $\vec{y}$ that is correlated to neutrino energy for a given reconstructed electron energy. From Equations (4.48) and (4.49),

$$\tilde{f}(\vec{x}|j, \varphi_j(E_\nu)) = f(\vec{x}|j) \int_{-\infty}^{+\infty} \varphi_j(E_\nu) f(E_\nu|j, T_{\text{eff}}, \vec{z'}) dE_\nu. \qquad (4.50)$$

The previous expression is exact as long as $\vec{z}$ is truly uncorrelated to neutrino energy for a given reconstructed electron energy. In the case of an approximation, the expression

$$f(\vec{x}|j, \varphi_j(E_\nu)) = \frac{\tilde{f}(\vec{x}|j, \varphi_j(E_\nu))}{\int_{-\infty}^{+\infty} \tilde{f}(\vec{x}|j, \varphi_j(E_\nu)) dV} \tag{4.51}$$

must be used to ensure the normalisation, where the integral in the denominator is performed over the width/area/volume/hypervolume of the entire observable space. This is why the PDF in Equation (4.50) is denoted using a tilde. In the current form, the PDF in Equation (4.50) would be difficult to use in a likelihood function, since it would have to be generated every time the $\varphi_j(E_\nu)$ function is varied in the fit. If $f(E_\nu|j, T_{\text{eff}}, \vec{z'})$ is binned using n_ν bins in neutrino energy with index l, the expression can be rewritten as

$$\begin{aligned}
\tilde{f}(\vec{x}|j, \varphi_j(E_\nu)) &\approx \int_{-\infty}^{+\infty} \varphi_j(E_\nu) f(\vec{x}|j) f_b(E_\nu|j, T_{\text{eff}}, \vec{z'}) dE_\nu \\
&= \sum_{l=1}^{n_\nu} \left\{ \left[\int_l \varphi_j(E_\nu) dE_\nu \right] f(\vec{x}|j) f_b(l|j, T_{\text{eff}}, \vec{z'}) \right\} \\
&= \sum_{l=1}^{n_\nu} \left\{ \phi_{jl} f(\vec{x}|j) P(l|j, T_{\text{eff}}, \vec{z'}) \right\},
\end{aligned} \tag{4.52}$$

where ϕ_{jl} are defined as in Equation (4.25) and where $f(\vec{x}|j) P(l|j, T_{\text{eff}}, \vec{z'})$ can be computed for every data event before the minimisation process. Substituting the previous expression in Equation (4.51) gives

$$f(\vec{x}|j, \varphi_j(E_\nu)) = \frac{\sum_{l=1}^{n_\nu} \left\{ \phi_{jl} f(\vec{x}|j) P(l|j, T_{\text{eff}}, \vec{z'}) \right\}}{\sum_{l=1}^{n_\nu} \left\{ \phi_{jl} \left[\int_{-\infty}^{+\infty} f(\vec{x}|j) P(l|j, T_{\text{eff}}, \vec{z'}) dV \right] \right\}} \tag{4.53}$$

and the product of the Poisson parameter is thus

$$\alpha_j \nu_{j_{\text{exp}}} f(\vec{x}|j, \varphi_j(E_\nu)) = \alpha_j \nu_{j_{\text{exp}}} \frac{\sum_{l=1}^{n_\nu} \left\{ \phi_{jl} f(\vec{x}|j) P(l|j, T_{\text{eff}}, \vec{z}') \right\}}{\sum_{l=1}^{n_\nu} \left\{ \phi_{jl} \left[\int_{-\infty}^{+\infty} f(\vec{x}|j) P(l|j, T_{\text{eff}}, \vec{z}') dV \right] \right\}}. \tag{4.54}$$

4.3.1 Improved Unconstrained Neutrino Energy Fit

Using the same method as before, α_j and ϕ_{jl} are reparameterised by defining $\alpha_{jl} \equiv \alpha_j \phi_{jl}$. Using the constraint in Equation (4.26), the relation between α_j and α_{jl} is found to be

$$\alpha_j = \sum_{l=1}^{n_\nu} \alpha_{jl} P(l|j). \tag{4.55}$$

From Equation (4.54), the product of the Poisson parameter with the PDF $f(\vec{x}|j, \vec{\alpha}_j, \vec{\eta}_j)$ for the improved unconstrained fit is thus given by

$$\alpha_j \nu_{j_{\text{exp}}}(\vec{\eta}_j) f(\vec{x}|j, \vec{\alpha}_j, \vec{\eta}_j) = \nu_{j_{\text{exp}}}(\vec{\eta}_j) \frac{\sum_{l=1}^{n_\nu} \{\alpha_{jl} P(l|j, \vec{\eta}_j)\} \sum_{l=1}^{n_\nu} \left\{ \alpha_{jl} f(\vec{x}|j, \vec{\eta}_j) P(l|j, T_{\text{eff}}, \vec{z}', \vec{\eta}_j) \right\}}{\sum_{l=1}^{n_\nu} \left\{ \alpha_{jl} \left[\int_{-\infty}^{+\infty} f(\vec{x}|j, \vec{\eta}_j) P(l|j, T_{\text{eff}}, \vec{z}', \vec{\eta}_j) dV \right] \right\}}$$

$$= \nu_j(\vec{\alpha}_j, \vec{\eta}_j) \frac{\sum_{l=1}^{n_\nu} \left\{ \alpha_{jl} f(\vec{x}|j, \vec{\eta}_j) P(l|j, T_{\text{eff}}, \vec{z}', \vec{\eta}_j) \right\}}{\sum_{l=1}^{n_\nu} \left\{ \alpha_{jl} \left[\int_{-\infty}^{+\infty} f(\vec{x}|j, \vec{\eta}_j) P(l|j, T_{\text{eff}}, \vec{z}', \vec{\eta}_j) dV \right] \right\}}, \tag{4.56}$$

where $\nu_j(\vec{\alpha}_j, \vec{\eta}_j) \equiv \nu_{j_{\text{exp}}}(\vec{\eta}_j) \sum_{l=1}^{n_\nu} \alpha_{jl} P(l|j, \vec{\eta}_j)$ is the Poisson parameter for class j and where $f(\vec{x}|j, \vec{\eta}_j)$ is approximated by

$$f(\vec{x}|j, \vec{\eta}_j) \approx f_b(m(T_{\text{eff}})|j, \vec{\eta}_j) f(\vec{y}|j, k(T_{\text{eff}}), \vec{\eta}_j) \tag{4.57}$$

as before. Note that this improved unconstrained fit cannot be represented using a set of n_ν mutually independent classes of events with PDFs $f(\vec{x}|j, l)$ and Poisson parameters

$\alpha_{jl}\nu_{j_{\exp}}$ because of the correction in the denominator of Equation (4.56).

4.3.2 Improved Binned P_{ee} Fit

For a binned P_{ee} fit, it has been found that ϕ_{jl} is given by Equation (4.35). The exact expression for α_j is given by Equation (4.36), but in practice the PDF $f(E_\nu|j)$ is usually binned. If $f(E_\nu|j)$ is binned using n'_ν neutrino energy bins with index p, α can be approximated by

$$\alpha_j \approx \alpha_{\mathsf{B}} \int_{-\infty}^{\infty} g_j(E_\nu, P_{ee}) f_b(E_\nu|j) dE_\nu$$

$$= \alpha_{\mathsf{B}} \sum_{p=1}^{n'_\nu} \left\{ \left[\int_p g_j(E_\nu, P_{ee}) dE_\nu \right] f_b(p|j) \right\}. \tag{4.58}$$

From Equations (4.35), (4.54) and (4.58), the product of the Poisson parameter with the observable space PDF for CC and ES in the context of this P_{ee} fit can thus be written as

$\alpha_j \nu_{j_{\exp}}(\vec{\eta}_j) f(\vec{x}|j, P_{ee})$

$$= \alpha_{\mathsf{B}} \nu_{j_{\exp}}(\vec{\eta}_j) \frac{\sum\limits_{p=1}^{n'_\nu} \left\{ \left[\int_p g_j(E_\nu, P_{ee}) dE_\nu \right] f_b(p|j) \right\} \sum\limits_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] f(\vec{x}|j) \frac{P(l|j, T_{\mathsf{eff}}, \vec{z'})}{\Delta_{E_\nu}(l)} \right\}}{\sum\limits_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] \left[\int_{-\infty}^{+\infty} f(\vec{x}|j) \frac{P(l|j, T_{\mathsf{eff}}, \vec{z'})}{\Delta_{E_\nu}(l)} dV \right] \right\}}$$

$$= \alpha_{\mathsf{B}} \nu_{j_{\exp}}(\vec{\eta}_j) \frac{\sum\limits_{p=1}^{n'_\nu} \left\{ \left[\int_p g_j(E_\nu, P_{ee}) dE_\nu \right] f_b(p|j) \right\} \sum\limits_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] f(\vec{x}|j) f_b(l|j, T_{\mathsf{eff}}, \vec{z'}) \right\}}{\sum\limits_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] \left[\int_{-\infty}^{+\infty} f(\vec{x}|j) f_b(l|j, T_{\mathsf{eff}}, \vec{z'}) dV \right] \right\}}$$

$$= \nu_j(\alpha_{\mathsf{B}}, P_{ee}, \vec{\eta}_j) \frac{\sum\limits_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] f(\vec{x}|j) f_b(l|j, T_{\mathsf{eff}}, \vec{z'}) \right\}}{\sum\limits_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] \left[\int_{-\infty}^{+\infty} f(\vec{x}|j) f_b(l|j, T_{\mathsf{eff}}, \vec{z'}) dV \right] \right\}}, \tag{4.59}$$

where all PDFs implicitly depend on the systematic shifts $\vec{\eta}_j$, where $f(\vec{x}|j, \vec{\eta}_j)$ is given by Equation (4.57) and where the Poisson parameter for class j has been defined as

$$\nu_j(\alpha_B, P_{ee}, \vec{\eta}_j) \equiv \alpha_B \nu_{j_{\text{exp}}}(\vec{\eta}_j) \sum_{p=1}^{n'_\nu} \left\{ \left[\int_p g_j(E_\nu, P_{ee}) dE_\nu \right] f_b(p|j, \vec{\eta}_j) \right\}, \qquad (4.60)$$

that is to say using n'_ν bins in neutrino energy with index p. It should be possible that n'_ν be much larger than n_ν if the subset $\vec{z'}$ contains some observables. Similarly to what has been done in Section 4.2.4, it is also possible to fit a binned P_{ee} function by using Equation (4.39). However, this restrains n'_ν to be equal to n_ν.

The likelihood expression of the improved P_{ee} fit developed in this section thus allows to fit for P_{ee} parameters with similar speed as the initial P_{ee} fit method described in Section 4.2.4. Additionally, if the observable subspace $\vec{z'}$ is smaller than $\vec{y}$, the new expression allows to reduce the dimensionality of the involved PDFs. In practice, the number of observables in $\vec{z'}$ is small if non-zero. Effectively, for a specific generated electron energy, the correlation between the observables in $\vec{y}$ and the neutrino energy is exactly null. Possible correlations between the observables in $\vec{y}$ and E_ν can thus only come from PMT reconstruction, which constitutes a second-order effect. Given a small $\vec{z'}$ subspace, the low dimensionality of $f_b(l|j, T_{\text{eff}}, \vec{z'})$ should allow to significantly increase the number of n_ν bins while maintaining the observable binning at the level of the unconstrained reconstructed electron energy fit without increasing the systematic uncertainties due to the limited statistics in the MC simulation. This should thus translate into smaller total uncertainties on the fitted parameters.

4.3.3 $P_{ee} + P_{ea}$ Fit

The discussion of Section 4.2.5 still holds for the improved fits. The only required modification to the derived equations for NC is to replace the neutrino energy binning with index l for a binning that uses n'_ν bins with index p.

4.4   Polynomial Survival Probability Function

4.4.1   Parameterisation

Some of the fit methods that were derived in this chapter involve the use of one or many polynomial survival probability function(s). For the region of interest, the P_{ee} function is expected to be smooth and well-behaved, such that it can be approximated using a low-order polynomial. Also, since the main handle on the neutrino energy with SNO comes from the reconstructed electron energy, the combination of the broad resolution in the mapping between the two types of energies (due to the nature of the interactions and to the energy reconstruction resolution) with the finite statistics available motivated the use of a second-order polynomial for the P_{ee} function. The peak of the statistics for SNO being around 10 MeV in neutrino energy, it was decided to expand the polynomial around this energy. In order to aim at measuring Earth matter effects, a different parameterisation was chosen for day and night survival probability curves. The selected parameterisation involves a second-order P_{ee} polynomial for the day curve, which is additionally combined with a first-order asymmetry function for the night curve. Explicitly, the expressions for

P_{ee} are the following:

$$P_{ee\,\text{day}}(E_\nu) = P_{ee0} + P_{ee1}(E_\nu - 10 \text{ MeV}) + P_{ee2}(E_\nu - 10 \text{ MeV})^2 \tag{4.61}$$

$$P_{ee\,\text{night}}(E_\nu) = P_{ee\,\text{day}}(E_\nu) \cdot \frac{2 + A_{ee0} + A_{ee1}(E_\nu - 10 \text{ MeV})}{2 - A_{ee0} - A_{ee1}(E_\nu - 10 \text{ MeV})}, \tag{4.62}$$

such that the day/night asymmetry of the survival probability, $A_{ee}(E_\nu)$, is given by

$$A_{ee}(E_\nu) \equiv A_{ee0} + A_{ee1}(E_\nu - 10 \text{ MeV}) = \frac{2\left[P_{ee\,\text{night}}(E_\nu) - P_{ee\,\text{day}}(E_\nu)\right]}{P_{ee\,\text{night}}(E_\nu) + P_{ee\,\text{day}}(E_\nu)}. \tag{4.63}$$

Regarding $P_{ee} + P_{ea}$ fits, the same types of expressions are used to parameterise P_{ea}, due to the arguments presented above, but also mainly due to the small number of $\text{ES}_{\mu\tau}$ events in the SNO data, which constitutes the only class of events constraining the shape of P_{ea}. For the salt phase, the ratio of the number of events for which the reconstructed electron energy distribution depends on P_{ea} ($\text{ES}_{\mu\tau}$) to the number of events that are instead affected by P_{ee} (CC and ES_e) is roughly a mere 3.5%, which shows how difficult it is to measure the P_{ea} polynomial parameters above the zeroth order; E_ν dependencies easily get buried in the noise caused by the other classes of events. Nonetheless, NC allows to fix the zeroth terms and it is thus mainly these parameters that allow for the additional freedom between the three signals that is required by a sterile model. For P_{ea}, the strategy adopted here is thus to use the parameterisation

$$P_{ea\,\text{day}}(E_\nu) = P_{ea0} + P_{ea1}(E_\nu - 10 \text{ MeV}) \tag{4.64}$$

$$P_{ea\,\text{night}}(E_\nu) = P_{ea\,\text{day}}(E_\nu) \cdot \frac{2 + A_{ea0}}{2 - A_{ea0}}. \tag{4.65}$$

To summarise, from Equation (4.36), the α scales of the SNO signals, as a function of the probability functions, are given by

$$\alpha_{CC} = \alpha_B \int_{-\infty}^{+\infty} P_{ee}(E_\nu) f(E_\nu | CC) dE_\nu \tag{4.66}$$

$$\alpha_{ES_e} = \alpha_B \int_{-\infty}^{+\infty} P_{ee}(E_\nu) f(E_\nu | ES_e) dE_\nu \tag{4.67}$$

$$\alpha_{ES_{\mu\tau}} = \alpha_B \int_{-\infty}^{+\infty} [1 - P_{ee}(E_\nu)] f(E_\nu | ES_{\mu\tau}) dE_\nu \tag{4.68}$$

$$\alpha_{NC} = \alpha_B, \tag{4.69}$$

in the case of the P_{ee} fits, and

$$\alpha_{CC} = \alpha_B \int_{-\infty}^{+\infty} P_{ee}(E_\nu) f(E_\nu | CC) dE_\nu \tag{4.70}$$

$$\alpha_{ES_e} = \alpha_B \int_{-\infty}^{+\infty} P_{ee}(E_\nu) f(E_\nu | ES_e) dE_\nu \tag{4.71}$$

$$\alpha_{ES_{\mu\tau}} = \alpha_B \int_{-\infty}^{+\infty} P_{ea}(E_\nu | ES_{\mu\tau}) f(E_\nu | ES_{\mu\tau}) dE_\nu \tag{4.72}$$

$$\alpha_{NC} = \alpha_B \int_{-\infty}^{+\infty} [P_{ee}(E_\nu) + P_{ea}(E_\nu)] f(E_\nu | NC) dE_\nu \tag{4.73}$$

for the $P_{ee} + P_{ea}$ fits.

4.4.2 Binned Polynomial

In this chapter, different methods involving binned survival probability analytical functions were presented. By definition, and as shown in Equation (4.38), these fit methods require to evaluate definite integrals of the P_{ee} (or P_{ea}) expressions for each bin in neutrino energy. It is straightforward to do in the case of a day curve as expressed in Equation (4.61), where

the indefinite integral of $P_{ee\text{day}}(E_\nu)$, evaluated at an energy $E_\nu = \Delta E_\nu + 10$ MeV, simply gives

$$\left[\int P_{ee\text{day}}(E_\nu)dE_\nu \right]_{\Delta E_\nu + 10 \text{ MeV}} = \Delta E_\nu \left[P_{ee0} + \Delta E_\nu \left(\frac{P_{ee1}}{2} + \Delta E_\nu \frac{P_{ee2}}{3} \right) \right]. \quad (4.74)$$

An analytical expression for the integral of $P_{ee\text{night}}$ is however not as trivial to evaluate numerically. The indefinite integral of $P_{ee\text{night}}(E_\nu)$, evaluated at an energy $E_\nu = \Delta E_\nu + 10$ MeV is given by the expression

$$\left[\int P_{ee\text{night}}(E_\nu)dE_\nu \right]_{\Delta E_\nu + 10 \text{ MeV}} =$$

$$- \frac{1}{6A_{ee1}^3} \left\{ A_{ee1}\Delta E_\nu \left[6\left(P_{ee0}A_{ee1}^2 + 4\left[P_{ee1}A_{ee1} + P_{ee2}(2 - A_{ee0}) \right] \right) \right. \right.$$

$$+ A_{ee1}\Delta E_\nu \left[3(P_{ee1}A_{ee1} + 4P_{ee2}) + 2P_{ee2}A_{ee1}\Delta E_\nu \right] \Big] + 24\log|A_{ee0} + A_{ee1}\Delta E_\nu - 2|$$

$$\cdot \left[A_{ee1}(2P_{ee1} + P_{ee0}A_{ee1}) + 4P_{ee2}(1 - A_{ee0}) + A_{ee0}(P_{ee2}A_{ee0} - P_{ee1}A_{ee1}) \right] \Big\}. \quad (4.75)$$

This expression being undefined at $A_{ee1} = 0$, and noticing that the evaluation of the natural logarithm can easily lead to truncation error in the context of a definite integral, it is natural to develop the expression of the logarithm around $A_{ee1} = 0$. Given an integration range between $\Delta E_{\nu 1} + 10$ MeV and $\Delta E_{\nu 2} + 10$ MeV, this gives

$$\log \left| \frac{A_{ee0} + A_{ee1}\Delta E_{\nu 2} - 2}{A_{ee0} + A_{ee1}\Delta E_{\nu 1} - 2} \right| = -\sum_{n=1}^{\infty} \frac{[\Delta E_{\nu 2}{}^n - \Delta E_{\nu 1}{}^n]}{n(2 - A_{ee0})^n} A_{ee1}{}^n. \quad (4.76)$$

Equation (4.75) has to be finite when $A_{ee1} \to 0$. In fact, since $A_{ee1} = 0$ leads to an asymmetry that does not depend on E_ν, it has to give

$$\lim_{A_{ee1} \to 0} \left[\int P_{ee\text{night}}(E_\nu)dE_\nu \right] = \frac{2 + A_{ee0}}{2 - A_{ee0}} \left[\int P_{ee\text{day}}(E_\nu)dE_\nu \right]. \quad (4.77)$$

The expansion in Equation (4.76) must thus allow to cancel any term in Equation (4.75) that has a power of A_{ee1} less than or equal to -1. Replacing Equation (4.76) in Equation (4.75) (using only the upper limit), expanding for $n \leq 3$ and after doing some algebra and simplification, this gives

$$
\left[\left(\int P_{ee\,\text{night}}(E_\nu) dE_\nu \right) \right]_{\Delta E_\nu + 10 \text{ MeV}} =
$$
$$
\frac{1}{6(A_{ee0} - 2)^3} \left\{ 6P_{ee0} \Big(A_{ee0}[4 + A_{ee0}(2 - A_{ee0})] - 8 \Big) \Delta E_\nu \right.
$$
$$
+ 3 \Big[P_{ee1} \Big(A_{ee0}[4 + A_{ee0}(2 - A_{ee0})] - 8 \Big) - 4P_{ee0}A_{ee1}(2 - A_{ee0}) \Big] \Delta E_\nu^2
$$
$$
+ 2 \Big[P_{ee2} \Big(A_{ee0}[4 + A_{ee0}(2 - A_{ee0})] - 8 \Big) - 4P_{ee1}A_{ee1}(2 - A_{ee0}) - 4P_{ee0}A_{ee1}^2 \Big] \Delta E_\nu^3 \right\}
$$
$$
+ 4[A_{ee1}(2P_{ee1} + P_{ee0}A_{ee1}) + 4P_{ee2}(1 - A_{ee0}) + A_{ee0}(P_{ee2}A_{ee0} - P_{ee1}A_{ee1})]
$$
$$
\cdot \sum_{n=4}^{\infty} \frac{\Delta E_\nu^n}{n(2 - A_{ee0})^n} A_{ee1}^{n-3} + C, \quad (4.78)
$$

where C is simply a constant that contains the terms that cancel out in the context of a definite integral. Finally, if the following variables are defined:

$$
C_1 = 2 - A_{ee0} \tag{4.79}
$$
$$
C_2 = -\frac{1}{6C_1^3} \tag{4.80}
$$
$$
C_3 = A_{ee0}(4 + A_{ee0}C_1) - 8 \tag{4.81}
$$
$$
C_4 = 4A_{ee1}C_1 \tag{4.82}
$$
$$
C_5 = 6P_{ee0}C_3 \tag{4.83}
$$
$$
C_6 = 3(P_{ee1}C_3 - P_{ee0}C_4) \tag{4.84}
$$
$$
C_7 = 2(P_{ee2}C_3 - P_{ee1}C_4 - 4P_{ee0}A_{ee1}^2) \tag{4.85}
$$
$$
C_8 = 4[A_{ee1}(2P_{ee1} + P_{ee0}A_{ee1}) + 4P_{ee2}(1 - A_{ee0}) + A_{ee0}(P_{ee2}A_{ee0} - P_{ee1}A_{ee1})], \tag{4.86}
$$

Equation (4.78) simplifies to

$$\left[\int P_{ee\,\text{night}}(E_\nu)dE_\nu\right]_{\Delta E_\nu+10\ \text{MeV}} =$$

$$C_2\Delta E_\nu(C_5 + \Delta E_\nu(C_6 + C_7\Delta E_\nu)) + C_8\sum_{n=4}^{\infty}\frac{\Delta E_\nu^n}{nC_1^n}A_{ee1}^{n-3} + C. \quad (4.87)$$

This simple expression can typically be computed very quickly and converges after only a few terms. It reduces to Equation (4.77) when $A_{ee1} = 0$.

4.5   Summary

Table 4.1 presents a summary of the different fit configurations discussed in this chapter. $P_{ee} + P_{ea}$ methods are not shown although they are very similar to the corresponding P_{ee} method. The meaning of the different indices and variables is the following:

j : Class of event (CC, ES$_e$, ES$_{\mu\tau}$). $1 \le j \le n_{\text{classes}}$.

$k,\ m$: Reconstructed electron energy (T_{eff}) bins. $1 \le k \le n_{\text{unconstr}}$, $m(T_{\text{eff}}) \subseteq k(T_{\text{eff}})$.

$l,\ p$: Neutrino energy $(E_n u)$ bins. $1 \le l \le n_\nu$, $1 \le p \le n'_\nu$, $n_\nu \le n'_\nu$.

$\vec{y}$: All observables used in parameter extraction, excluding T_{eff}.

$\vec{z'}$: Subset of observables from $\vec{y}$ having significant correlations with E_ν for a given
        T_{eff}.

The index m has been used to replace the reconstructed electron energy T_{eff} in the table, since it is the binning scheme that is chosen when the conditional PDFs are generated from Monte Carlo samples. For the same reason, all the unbinned PDFs are now written as binned PDFs. As it can be seen, except for the particular case of the unbinned P_{ee} polynomial method, all the methods form binned observable PDFs $f_b(\vec{x}|\ldots)$ by using binned

Method	Fit Param. (# param.)	CC and ES PDFs	$\nu f(\vec{x}|j)$
Unconstr. Reconst. T_{eff}	α_{jk} ($n_{\text{unconstrained}} \times n_{\text{classes}}$)	$f_b(m|j)$, $f_b(\vec{y}|j,k)$	(4.13)
Unconstr. E_ν	α_{jl} ($n_\nu \times n_{\text{classes}}$)	$P(l|j)$, $f_b(m|j,l)$, $f_b(\vec{y}|j,l,k)$	(4.32)
Unbinned P_{ee} pol.	P_{ee} pol. pars, α_{B} (pol. order+2)	$f_b(\vec{x}|j, P_{ee})$	
Binned P_{ee} pol.	P_{ee} pol. pars, α_{B} (pol. order+2)	$f_b(l|j)$, $f_b(m|j,l)$, $f_b(\vec{y}|j,l,k)$	(4.37)
Binned P_{ee}	P_{eel}, α_{B} ($n_\nu + 1$)	$P(l|j)$, $f_b(m|j,l)$, $f_b(\vec{y}|j,l,k)$	
Improved unconstr. E_ν	α_{jl} ($n_\nu \times n_{\text{classes}}$)	$P(l|j)$, $P(l|j,m,\vec{z'})$, $f_b(m|j)$, $f_b(\vec{y}|j,k)$	(4.56)
Improved binned P_{ee} pol.	P_{ee} pol. pars, α_{B} (pol. order+2)	$f_b(p|j)$, $f_b(l|j,m,\vec{z'})$, $f_b(m|j)$, $f_b(\vec{y}|j,k)$	(4.59)
Improved binned P_{ee}	P_{eel}, α_{B} ($n_\nu + 1$)	$P(l|j)$, $P(l|j,m,\vec{z'})$, $f_b(m|j)$, $f_b(\vec{y}|j,k)$	

Table 4.1: Summary of the different fit configurations in Chapter 4. For the improved methods, the reconstructed electron energy T_{eff} in the conditional PDFs has been replaced by the associated bin index m.

conditional T_{eff} spectral PDFs $f_b(m|\ldots)$ and binned conditional PDFs $f_b(\vec{y}|\ldots,k)$. Although building the likelihood function using these PDFs would be absolutely correct, it is a simplification of how it is actually done in QSigEx, because of the optimisation mentioned in Section 3.3. In reality, one of these two types of conditional distributions is always a normalised histogram and the other is a PMF, since it is normalised only by the number of Monte Carlo events. As shown, extra distributions are required by the methods that use neutrino energy spectra. For these, as a general rule, the binned fits in E_ν and P_{ee} use PMFs instead of PDFs while the P_{ee} polynomial methods use regular PDFs.

In this chapter, multiple methods were derived to perform spectral extractions of CC and ES SNO signals. In the context of a multi-phase fit that involves varying detector response and where the ultimate goal is to extract the neutrino survival probability, it appears that the best method is the unbinned P_{ee} polynomial fit. This method reflects the sensitivity of the SNO detector and is based on a statistical model which is not underconstrained like unconstrained methods. It does not bin the neutrino energy, a non-observable, and also uses reduced PDF dimensionality; both factors contribute to reduce biases. However, such a method is computationally very challenging to use, because it is extremely slow. The improved binned P_{ee} polynomial method constitutes the best alternative, since it is much faster and can potentially allow for a large number of neutrino energy bins. Both the improved binned P_{ee} polynomial method and the unbinned P_{ee} polynomial method are used for the final fits, as described in Chapter 7. It will be shown in Chapter 6 how the two methods generate very similar results. If the P_{ee} methods are perfectly suited for the extraction of the survival probability parameters for a three neutrino flavour model, it is not as true for cases with non-active flavours for which a less constrained model is required. Corresponding $P_{ee} + P_{ea}$ methods have thus been developed to perform a fit for such cases.

As all the elements of the likelihood function have now been presented, the details of the three-phase extraction with QSigEx will be covered in the next chapter.

Chapter 5

Signal Extraction

Although Chapters 3 and 4 described the general form of the different building blocks of the extended log-likelihood function used by QSigEx, specific details regarding the three-phase signal extraction have not yet been covered. This chapter provides information regarding data selection, PDF parameters and the signal extraction process, information which is relevant for the ensemble tests analysed in Chapter 6 as well as for the real data fits presented in Chapter 7.

5.1 Data Selection

To perform the three-phase extraction, the PMT events are selected using the same criteria as for the previous analyses of these data [7, 4]. In particular, the analysis window that is defined for the LETA observables is given by

$$3.5 \text{ MeV} \leq \quad T_{\text{eff}} \quad < 20.0 \text{ MeV}$$

$$0 \leq \quad \rho \quad < 0.77025$$

$$-0.12 \leq \quad \beta_{14} \quad < 0.95$$

$$-1 \leq \cos\theta_{\odot} < 1$$

and the window for the NCD phase is defined by

$$6.0 \text{ MeV} \leq \quad T_{\text{eff}} \quad < 20.0 \text{ MeV}$$

$$0 \leq \quad \rho \quad < 0.77025$$

$$-1 \leq \cos\theta_{\odot} < 1.$$

In addition to the above criteria, more advanced algorithms referred to as "high level cuts" have been used to enhance background rejection. More information regarding these tests can be found in [7, 4, 49, 50].

5.2   PDF Binning and Configuration

5.2.1   Observable and Monte Carlo Variable Binning

As explained in Section 3.3, for a given dataset, QSigEx uses the same observable binning for the data and the signal extraction PDFs. The binning described in this section thus applies to all classes of events. This also holds for the CC, ES_e and $ES_{\mu\tau}$ PDFs, but in some cases these signals use additional binning of the reconstructed electron energy to generate conditional PDFs. The regular binning of T_{eff} as an observable uses the index m and that for the conditional PDFs uses the index k, as described in Section 4.1. However, for the fit methods selected for the three-phase analysis, the same binning is actually used. Binned P_{ee} polynomial fits also bin the neutrino energy, a non-observable. Two binning schemes are used for E_ν, which are referred to by using indices l and p. Tables 5.1, 5.2 and 5.3 give the binning used for the three-phase analysis.

Observable	# bins	Binning
$T_{\text{eff}}\ (m,\ k)$	21	0.25 MeV-wide bins from 3.5 to 5.0 MeV 0.5 MeV-wide bins from 5.0 to 12.0 MeV 8.0 MeV-wide bin above 12.0 MeV
$\cos\theta_\odot$	8	0.25-wide bins from -1 to +1
ρ	5	0.15405-wide bins from 0 to 0.77025
β_{14}	15	$0.071\bar{3}$-wide bins from -0.12 to +0.95

Table 5.1: Observable binning used for LETA.

Observable	# bins	Binning
$T_{\text{eff}}\ (m,\ k)$	13	0.5 MeV-wide bins from 6.0 to 12.0 MeV 8.0 MeV-wide bin above 12.0 MeV
$\cos\theta_\odot$	25	0.08-wide bins from -1 to +1
ρ	10	0.077025-wide bins from 0 to 0.77025

Table 5.2: Observable binning used for the NCD phase.

Variable	# bins	Binning
$E_\nu\ (l)$	32	4.0 MeV-wide bin from 1.0 to 5.0 MeV 0.3 MeV-wide bins from 5.0 to 14.0 MeV 3.0 MeV-wide bin from 14.0 to 17.0 MeV
$E_\nu\ (p)$	96	3.00 MeV-wide bin from 1.00 to 4.00 MeV 0.75 MeV-wide bin from 4.00 to 4.75 MeV 0.25 MeV-wide bin from 4.75 to 5.00 MeV 0.10 MeV-wide bins from 5.00 to 14.00 MeV 0.25 MeV-wide bin from 14.00 to 14.25 MeV 0.75 MeV-wide bin from 14.25 to 15.00 MeV 2.00 MeV-wide bin from 15.00 to 17.00 MeV

Table 5.3: Monte Carlo variable binning used for the three-phase analysis.

5.2.2   PDF Configuration

5.2.2.1   CC, ES$_e$ and ES$_{\mu\tau}$ Signals

For the CC and ES signals, the different PDF configurations were described almost entirely in Section 4.5. However, in the case of the improved binned P_{ee} polynomial fit, the subset of observables $\vec{z'}$ has not been specified. For this fit the CC, ES$_e$ and ES$_{\mu\tau}$ PDFs were found to be given by

$$
f(\vec{x}|j, P_{ee}) = \frac{\sum\limits_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] f(\vec{x}|j) f_b(l|j, T_{\text{eff}}, \vec{z'}) \right\}}{\sum\limits_{l=1}^{n_\nu} \left\{ \left[\int_l g_j(E_\nu, P_{ee}) dE_\nu \right] \left[\int_{-\infty}^{+\infty} f(\vec{x}|j) f_b(l|j, T_{\text{eff}}, \vec{z'}) dV \right] \right\}}
\tag{5.1}
$$

in Equation (4.59), where $\vec{x}^{(\text{D}_2\text{O,salt})} \equiv (T_{\text{eff}}, \rho, \cos\theta_\odot, \beta_{14})$ for the D$_2$O and salt phases and $\vec{x}^{\text{NCD−PMT}} \equiv (T_{\text{eff}}, \rho, \cos\theta_\odot)$ for the NCD phase. In the same equation, $g_j(E_\nu, P_{ee})$ is given by $P_{ee}(E_\nu)$, $1 - P_{ee}(E_\nu)$ or $P_{ea}(E_\nu)$, depending on the signal j, and the integral of the denominator is performed over the volume/hypervolume of the entire observable space. A vector $\vec{y}$ was defined as the subset of $\vec{x}$ that excludes T_{eff}. The vector $\vec{z'}$ was defined as the subset of $\vec{y}$ whose observables are correlated to the neutrino energy for a given reconstructed electron energy. This partitioning of $\vec{y}$ allowed to significantly reduce the dimensionality of the PDF $f(E_\nu|j, \vec{x})$ such that it could be approximated by $f(E_\nu|j, T_{\text{eff}}, \vec{z'})$ (Equation (4.49)). For the CC and ES signals, the correlations between the observables of the $\vec{y}$ subset and the neutrino energy for a given reconstructed electron energy are expected to be small, because they are only induced through observable reconstruction. Effectively, for a given electron energy in the detector, the true values of the ρ, $\cos\theta_\odot$ and β_{14} observables are not correlated to the neutrino energy.

After using CC and ES Monte Carlo samples to compute the correlations between the reconstructed observables and the neutrino energy, for narrow bins of electron energies, the $\vec{z'}$ subset for D_2O and salt phases was defined as

$$\vec{z'}^{(D_2O,\text{salt})} \equiv (\beta_{14}), \tag{5.2}$$

since this observable has the largest correlation with a maximum correlation coefficient of only about 0.038 among all the defined T_{eff} bins and a correlation of ≈ 0.02 around 10 MeV. For the NCD phase β_{14} is not used in signal extraction, so $\vec{z'}$ is defined as

$$\vec{z'}^{NCD-PMT} \equiv (\cos \theta_\odot) \tag{5.3}$$

instead, since the maximum value for the related correlation is about 0.035.

5.2.2.2   NC Signal and Backgrounds

For the NC signal, the direction of the emitted photons following the neutron capture is totally uncorrelated with the direction of the neutrino, such that the $\cos \theta_\odot$ observable can be marginalised in its own PDF. The PDF configuration for NC is thus

$$f(\vec{x}|NC) = f(T_{\text{eff}}, \rho, \beta_{14}|NC) f(\cos \theta_\odot|NC) \tag{5.4}$$

for LETA, and

$$f(\vec{x}|NC) = f(T_{\text{eff}}, \rho|NC) f(\cos \theta_\odot|NC) \tag{5.5}$$

for the NCD phase.

For backgrounds, the $\cos\theta_\odot$ distributions are also uncorrelated with the other observables, such that they use the same PDF configuration as for the NC signal.

5.3 Likelihood Function Minimisation

To minimise the three-phase negative extended log-likelihood function, the QSigEx signal extraction package relies on TMinuit, the version of the MINUIT minimisation package implemented in ROOT [3]. TMinuit has the capability to handle the large number of floating parameters of the three-phase fit as well as the ability to find the global minimum in a reasonable number of iterations. In QSigEx, the negative extended log-likelihood function is minimised by using the MIGRAD algorithm. To evaluate parameter uncertainties, MIGRAD is called a second time in order to get a better estimate of the covariance matrix and then the MINOS algorithm is used to compute the asymmetrical uncertainty values.

5.4 Treatment of Systematic Uncertainties

The three-phase extraction of the survival probability involves a large number of fit parameters. Excluding the signal parameters, the likelihood function contains 36 parameters for the background rates and 117 systematic uncertainty parameters (nuisance parameters). Since systematic uncertainty parameters represent degrees of freedom in the statistical model of the likelihood function, similar to the signal parameters, the data themselves can constitute an indirect measurement of the systematic uncertainty parameters. The

data thus have the potential to provide additional constraints on the systematic uncertainties, albeit at various degrees, depending on the significance of the changes in the observable distributions, including scaling effects, that occur when these parameters are varied within their nominal range. Hence it can be beneficial to float systematic uncertainty parameters in the fit. These additional constraints contribute to reduce the effects of the systematic uncertainty parameters on the signal parameters. However, many systematic uncertainty parameters affect observable values, allowing Monte Carlo events to fall in different PDF bins depending on the uncertainty parameter value. Due to the finite number of Monte Carlo events, this causes discontinuities in the likelihood function, which prevent minimisation and uncertainty estimation algorithms from working properly. Finally, some systematic uncertainty parameters either have very small effects on the signal parameters, or are not significantly constrained by the data. Varying all these parameters in the fit would significantly increase the required time to evaluate the uncertainties of the signal parameters, without providing any benefit for the quality of the results. For all the above reasons, it is thus preferable to resort to an ensemble of methods to evaluate the effects of systematic uncertainties. The following sections describe the different strategies that are used by QSigEx.

5.4.1   Floated Parameters

Since many systematic uncertainty parameters affect observable values, only a fraction of them can truly be floated without causing minimisation problems such as the ones discussed in the previous section. Among the remaining parameters, most of them have very small effects on signal parameters, such that it is not very useful to float them rather

than using the shift and refit method described below. The only parameters that are truly floated in the three-phase fit are the NC capture efficiency relative uncertainties for the NCD phase, which are defined in Appendix D.

5.4.2   Scanned Parameters

Among the parameters that are not truly floated, some have the potential to be significantly constrained by the data. These include parameters such as energy scales, energy resolution shifts, β_{14} scales and the parameters for the PMT β-γ analytical PDFs. To determine ranges for these uncertainties that are additionally constrained by the data, the approach taken is to iteratively scan the likelihood space along the direction of the individual scanned parameters, each time finding the point in the floated parameter space that minimises the negative extended log-likelihood function. After each scan, the best fit value for the current parameter is updated and the algorithm proceeds to the next parameter. This technique, as well as a more extensive discussion about the selection of the scanned systematic uncertainty parameters, can be found in [49]. Compared to the previous analysis, the technique used in the three-phase analysis is more automated. An algorithm has been developed to search for the negative likelihood minimum when scanning parameters as well as to define scanning ranges for the next iteration, which notably allows to "zoom in" around the negative likelihood minimum to enhance accuracy. This algorithm is the following:

  1. The next systematic uncertainty parameter that needs to be scanned is selected.

2. The parameter is scanned using a specified number of equally spaced parameter values. The scan range is given by the value computed during the previous iteration or by using 2.5 times the nominal systematic uncertainty value in the case of the first iteration. To compute the negative log-likelihood value for each scan point, the other parameters are fixed to their current best fit values and the negative log-likelihood is minimised using the floating parameters.

3. An asymmetrical parabola (that is to say, the negative logarithm of an asymmetrical Gaussian function) is fitted to the resulting curve, after determining a fit range given by the set of scan points around the minimum position that are less than two likelihood units (two sigmas) larger than the minimum value.

4. The updated best fit value for the current scan parameter is given by the fitted value for the parabola minimum. The new scan range on each side of the fitted minimum is computed using the largest value between twice the largest fitted asymmetrical uncertainty value and half the previous scan range.

5. Steps 1 to 4 are repeated until all parameters have been scanned.

6. Steps 1 to 5 are repeated until convergence is achieved, which generally occurs when parameter fit values stop systematically drifting from one iteration to the other.

7. The fitted values and uncertainties for the systematic uncertainty parameters are given by the fitted peak and widths of the fitted asymmetrical Gaussian functions as provided by the last iteration.

The main advantage of the above algorithm, in addition to providing an automated method to scan the systematic uncertainty parameters, is that it allows to evaluate parameters
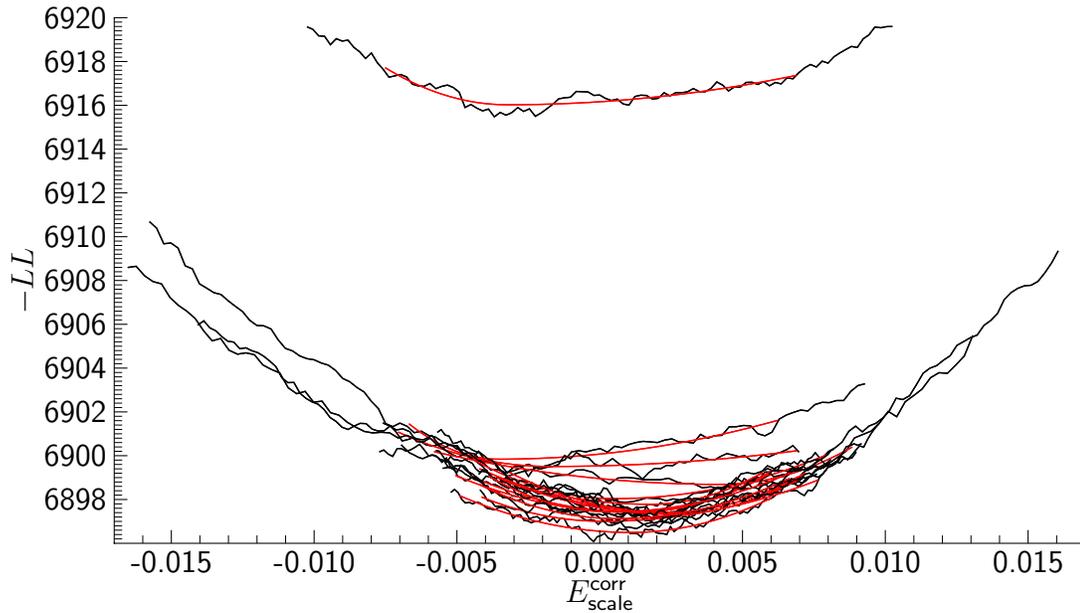
Figure 5.1: Example of scan iterations of the correlated energy scale systematic uncertainty parameter. Shown are the computed likelihood values along with the fitted asymmetrical parabolas.

that would otherwise be quite challenging to scan manually, due to discontinuities in the likelihood space caused by the finite number of Monte Carlo events that are moved to different observable bins, but also in and out of the analysis window, when varying these systematic uncertainty parameters. Figure 5.1 presents an example of scan iterations for the correlated energy scale systematic uncertainty parameter. It shows such discontinuities, along with some drift of the best fit value for the first five iterations.

5.4.3 Shift and Refit Parameters

For the remaining systematic uncertainty parameters, as well as for the parameters whose most probable values and uncertainties are determined by the scan procedure as described

in the previous section, the associated systematic uncertainties on the extracted param-
eters, such as the ^{8}B scale and the P_{ee} parameters, are computed using a shift and refit
procedure. In the previous SNO analyses that used a maximum likelihood procedure,
these systematic effects were typically computed by minimising the negative log-likelihood
function using single one-sigma shifts of the parameters on each side of the most probable
values.

In QSigEx, the shift and refit procedure has been modified in order to improve the
evaluation of the systematic effects in cases where the relation between a systematic uncer-
tainty parameter and the position of the minimum of the negative log-likelihood function
is non-linear. Instead of performing a single shift of a systematic uncertainty parameter
on each side of its most probable value, QSigEx performs 100 random shifts according to
an asymmetrical Gaussian distribution whose parameters are set using the most probable
value and parameter uncertainties. For each shift, the negative log-likelihood function is
minimised within the floating parameter space, as usual. The fitted value for the floating
parameters is recorded for each fit. Finally, to obtain the resulting uncertainties for one
of these parameters, the RMS of its fitted values is computed separately on each side
of the most probable floating parameter value (the value obtained when using the most
probable systematic uncertainty parameter values) and also relatively to this value. This
ensures that the most probable value for the floating parameters correspond to the "best"
parameter value. This procedure is also equivalent to fitting the widths of an asymmetrical
Gaussian function whose peak is forced to be located at the "best" floating parameter
value.

5.4.4   Error Combination

To combine the individual effects of the shifted systematic uncertainty parameters, in-
cluding the parameters that are scanned beforehand, the method followed by the previous
SNO analyses consisted of adding the positive sides of the systematic uncertainty effects
in quadrature, and similarly for the negative sides. The intent of this procedure was to
combine the effects under the assumption that the systematic uncertainties are not corre-
lated. However, this method was not completely correct, because under the no-correlation
assumption, total systematic shifts result from individual effects in random directions.

To more correctly combine the effects of the systematic uncertainties, QSigEx relies on
a simple Monte Carlo simulation. For each floating parameter, it generates a distribution of
total systematic shifts by drawing random individual shifts according to the asymmetrical
Gaussian distributions that are computed during the shift and refit procedure. Each total
shift value in this distribution is computed simply by adding linearly the individual random
shifts. The distribution is populated using 10^6 of these computed shifts. Finally, the
total asymmetrical systematic uncertainties on the floating parameters are obtained by
computing RMS values of the distributions about zero on the positive and negative sides
separately.

5.5   Summary

Combined with the information contained in Appendix D, the elements presented in this chapter represent the more specific aspects of the analysis that are used to perform the three-phase SNO extraction. In the next chapter, results from thorough tests of the extraction procedure are presented.

Chapter 6

Ensemble Testing of SNO

6.1 Introduction

Chapters 3 to 5 have shown the complexity of the SNO signal extraction for the three-phase analysis. In QSigEx, this process involves minimising the likelihood function over 30 000 times, while floating 45 parameters and scanning or shifting randomly over 115 systematic uncertainty parameters, which trigger the regeneration of the PDFs using over 3.5 GB of Monte Carlo simulation data. Although QSigEx's code has been designed to minimise the possibility of unnoticed bugs to find their way into the final fit, notably by making its structure as flexible as possible such that its code can be reused multiple times in many contexts without having to change input values in different locations, it remains essential to ensure that the results it provides are not affected by any significant inconsistency. Although this chapter is quite substantial, its sole purpose can nonetheless be summarised as verifying and proving that the extracted signal parameters are bias-free in the frequentist sense.

To perform ensemble testing, that is to say to produce and analyse fit results obtained using multiple fake datasets generated using known parameters, allows not only to

identify bugs in QSigEx's code, but also to find problems resulting from neglected correlations and/or non-linearities due to the involvement of multiple observables, systematic uncertainty parameters and classes of events in the parameter extraction process. Ensemble testing allows also to tune different parameters, such as observable bin size, in order to minimise the expected total uncertainty on the fitted signal parameters. Finally, this technique is useful to establish criteria of comparison for the real data fit, when this is performed via multiple extraction packages.

In the context of SNO, different factors influence how ensemble testing can be performed. The main factor is the limited Monte Carlo statistics, which allows to generate only a few fake datasets for some of the background events due to the low efficiency with which they can be generated inside the analysis window. Another challenge consists of the computational resources involved in a SNO three-phase fit, which constrain how ensemble testing can be performed while propagating all of the systematic uncertainties of the real data fit.

To reduce the effects of these constraints, two techniques were used for the three-phase analysis. The first one is to generate fake data for one event class at a time, then to combine them using a progressive approach:

1. Using only signal (CC, ES, NC) events.

2. Using signal events and background events with high statistics (that is to say, background classes with high Monte Carlo multiplicity).

3. Using all signal and background events.

4. Using all signal and background events, and varying the main scanned systematic uncertainty parameters.

For each type of dataset, fits were performed using only the corresponding classes of events. This approach allows to perform ensemble testing in decreasing order of both sensitivity and of possible negative consequences. Performing ensemble testing using hundreds of signal-only fake datasets allows to ensure that the QSigEx package builds the likelihood function properly given a ^{8}B flux and survival probability parameters. It also allows to probe the chosen PDF configuration using the best sensitivity possible. Adding high-statistics backgrounds allows to verify the proper behaviour of the code that builds the likelihood for any background while retaining almost the same sensitivity as for signal-only testing. It also gives the chance to test the PDF configuration for the selected backgrounds and to see how these classes of events interact with the signals in the fit. The inclusion of the remaining classes of events allows only to ensure the proper interaction of these backgrounds in the fit, admittedly with significantly less sensitivity. Finally, ensemble testing of full fake sets while varying the main scanned systematic uncertainty parameters is useful to verify the parameter scanning and zooming procedure, to test the proper implementation of these systematic uncertainties at some extent and also to assess their effect on the reliability of the extracted signal parameters. The progressive ensemble testing was performed separately for the LETA component and the NCD phase, and then for the three-phase analysis. The individual LETA and NCD phase fits allow to possibly show problems pertaining to one of the two sub-analyses, which could be suppressed when performing three-phase fits. However, only the results from the three-phase tests are included in this thesis, because of the similarity in the results and the large number of involved plots.

The second technique allows to reduce the effects of the SNO constraints by performing the ensemble tests mentioned above twice; the second time using fake datasets generated with a different random number generator seed. This gives the possibility to generate split sets that, although not statistically independent from each other, allow to appreciate the effects of the limited number of sets on the results from ensemble testing. Thus, if results from the first ensemble test are a source of concern, it is possible to see how much these results change using the second ensemble. If the new results move considerably in the other direction, one can conclude that the initial shifts were statistical fluctuations. Otherwise, more investigation could be necessary.

6.2   Fake Data and PDF Event Generation

As mentioned in the previous section, to perform the multiple ensemble tests presented later in this chapter, fake sets were first generated for each class of events. To correctly model the stochastic nature of event detection, these sets were generated by randomly drawing events from the Monte Carlo simulation data without replacement, with the number of drawn events being Poisson-distributed. The Poisson means were chosen such that they matched the number of expected events in the analysis window, as computed using Monte Carlo samples with event production rates set to known values when available. The production rates for the signals were set to the BP2000 SSM ^{8}B flux [51] and the background production rates were set to the values measured using external assays when available [4, 7]. Otherwise, the values obtained from previous SNO analyses were used to set the background Poisson parameters.

For the ^{8}B signals, the acceptance-rejection method was used to generate events with the required dependency on the survival probability functions. The survival probability parameters were chosen to be

$$P_{ee_{\text{day}}} = 0.325 - 8.88 \times 10^{-3} \text{ MeV}^{-1}(E_\nu - 10 \text{ MeV}) +$$

$$1.22 \times 10^{-3} \text{ MeV}^{-2}(E_\nu - 10 \text{ MeV})^2 \tag{6.1}$$

$$P_{ee_{\text{night}}} = P_{ee_{\text{day}}} \cdot \frac{2 + 2.80 \times 10^{-2} + 4.78 \times 10^{-3} \text{ MeV}^{-1}(E_\nu - 10 \text{ MeV})}{2 - 2.80 \times 10^{-2} - 4.78 \times 10^{-3} \text{ MeV}^{-1}(E_\nu - 10 \text{ MeV})} \tag{6.2}$$

$$P_{ea_{\text{day}}} = 0.650 + 7.92 \times 11^{-3} \text{ MeV}^{-1}(E_\nu - 10 \text{ MeV}) \tag{6.3}$$

$$P_{ea_{\text{night}}} = P_{ea_{\text{day}}} \cdot \frac{2 + 2.23 \times 10^{-2}}{2 - 2.23 \times 10^{-2}}, \tag{6.4}$$

where P_{ea} functions were used only for $P_{ee} + P_{ea}$ fits. The hep events have been generated using 0.35 SSM, 0.47 SSM and 1.00 SSM for hep CC, hep ES and hep NC, respectively. Tables C.1 to C.4 in Appendix C show the content of the fake datasets and the associated PDF events.

6.3   Ensemble Testing

As mentioned earlier, ensemble tests allow to identify possible problems in signal extraction. In order to analyse the ensemble tests, various metrics must be defined. All of the ensemble tests of this chapter have been analysed using the following figures of merit:

Fractional Bias   The fractional bias is defined as

$$\text{Fractional Bias} \equiv \frac{\text{Fit value} - \text{Nominal value}}{\text{Nominal value}}. \tag{6.5}$$

In the bias plots, drawn error bars indicate the uncertainty on the average bias, which corresponds approximately to the sample standard deviation of the bias divided by the square root of the number of samples.

Pull   The pulls are defined as

$$\text{Pull} \equiv \frac{\text{Fit value} - \text{Nominal value}}{\text{Fitted Uncertainty}}. \tag{6.6}$$

In the pull plots, drawn error bars show the sample standard deviation of the pull and not the uncertainty on the average pull. Statistics predict these error bars to have an expectation value of approximately $1 - \frac{1}{4(n-1)}$ and an uncertainty of about $\frac{1}{\sqrt{2(n_{\text{samples}}-1)}}$. The expected uncertainty on the pull average, which is given by $\frac{1}{\sqrt{n_{\text{samples}}}}$ from statistics, is shown on the plots using dashed lines. The expected uncertainty bands for the locations of the ends of the error bars, given the computed averages, are also shown using grey strips.

Relative Uncertainty   The relative uncertainties are defined as

$$\text{Relative Uncertainty} \equiv \frac{\text{Fitted Uncertainty}}{\text{Nominal value}} \cdot 100\%. \tag{6.7}$$

In the associated plots, average relative uncertainties are shown along with the uncertainty on this average, which is the sample standard deviation of the relative uncertainties divided by the square root of the number of samples.

In order to satisfy the assumption of the likelihood function that all data events and measurements are independent, the means for all the penalty terms used during ensemble testing have been randomly fluctuated via Gaussian random values. This simulates multiple measurements for these penalty terms and ensures in particular that the pull uncertainties are not artificially reduced. The following sections show the ensemble test results for the three-phase analysis.

6.4   Three-Phase Testing

This section presents results of the different types of ensemble tests as described in Section 6.1, when using three-phase fake datasets.

6.4.1 P_{ee} Polynomial Fits

6.4.1.1   Signal-Only Testing

In this section, results of ensemble testing of the P_{ee} fit are shown; these results were obtained when using only signal (CC, ES$_e$, ES$_{\mu\tau}$ and NC) events. As mentioned earlier, such tests allow to use a maximum number of fake sets (250 three-phase sets) to look at the test variables with the best sensitivity to eventual problems related to the P_{ee}

parameterisation. This permits to decouple these possible problems from other factors related to the backgrounds.

Figure 6.1 presents the fractional biases obtained from both the improved binned P_{ee} polynomial and the unbinned P_{ee} polynomial fit. As the figure shows, although the bias for P_{ee2} on the top plot might seem to be significant at first sight, the bottom plot, which was generated using the alternate ensemble test, reveals that they are in fact only fluctuations due to the limited statistics in the Monte Carlo samples used to generate the signal extraction PDFs. The same can be said about the ^{8}B scale for the top plot of Figure 6.2, when comparing the mean value to the expected uncertainty on the mean (dashed line) with 250 fake sets. Additionally, this second figure shows that the size of the error bars for the pulls are in good agreement with the expected spread, which is indicated by the grey strips. The two figures thus show that three-phase ensemble testing using only the signals leads to bias-free results that also have expected pull distributions. Finally, Figure 6.3 shows the computed relative uncertainties obtained for both the regular and alternate ensemble tests. For these tests, the relative uncertainty on the ^{8}B scale is $(2.260 \pm 0.003)\%$ and the one on P_{ee0} is $(4.22 \pm 0.01)\%$.

6.4.1.2   Signal with High-Statistics Monte Carlo Backgrounds

Tables C.1 to C.4 show that it is possible to perform three-phase ensemble testing with 250 sets as for the previous section, but this time including backgrounds, except for Bi and Tl events for LETA and neutral counter photodisintegration events from the NCD phase. Since all backgrounds are handled by the same code in QSigEx (except the analytical backgrounds), this allows to test all the components that are used by full set fits, but
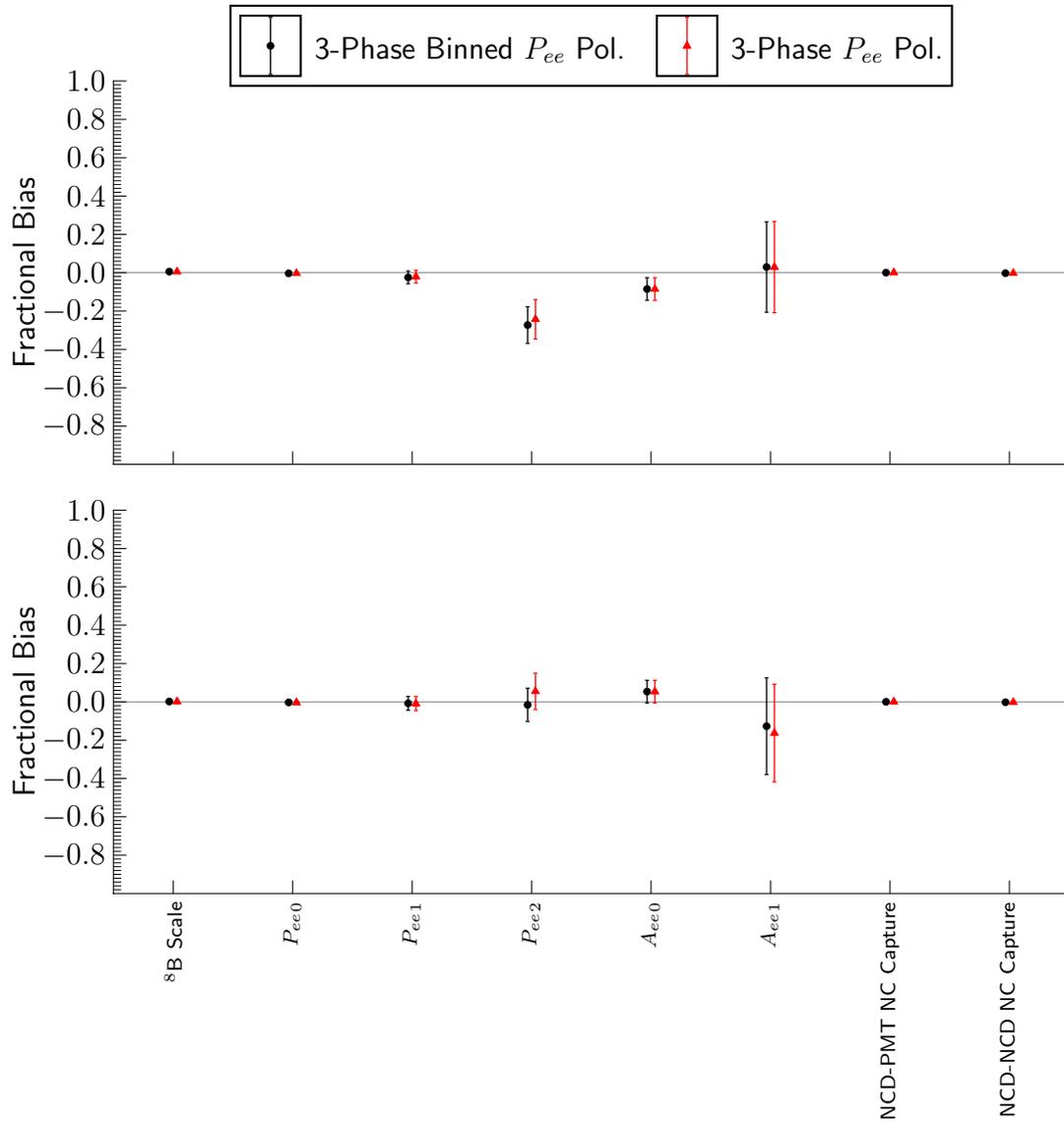
Figure 6.1: Fractional biases for ensemble testing of the three-phase analysis (250 signal-only fake sets). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles.
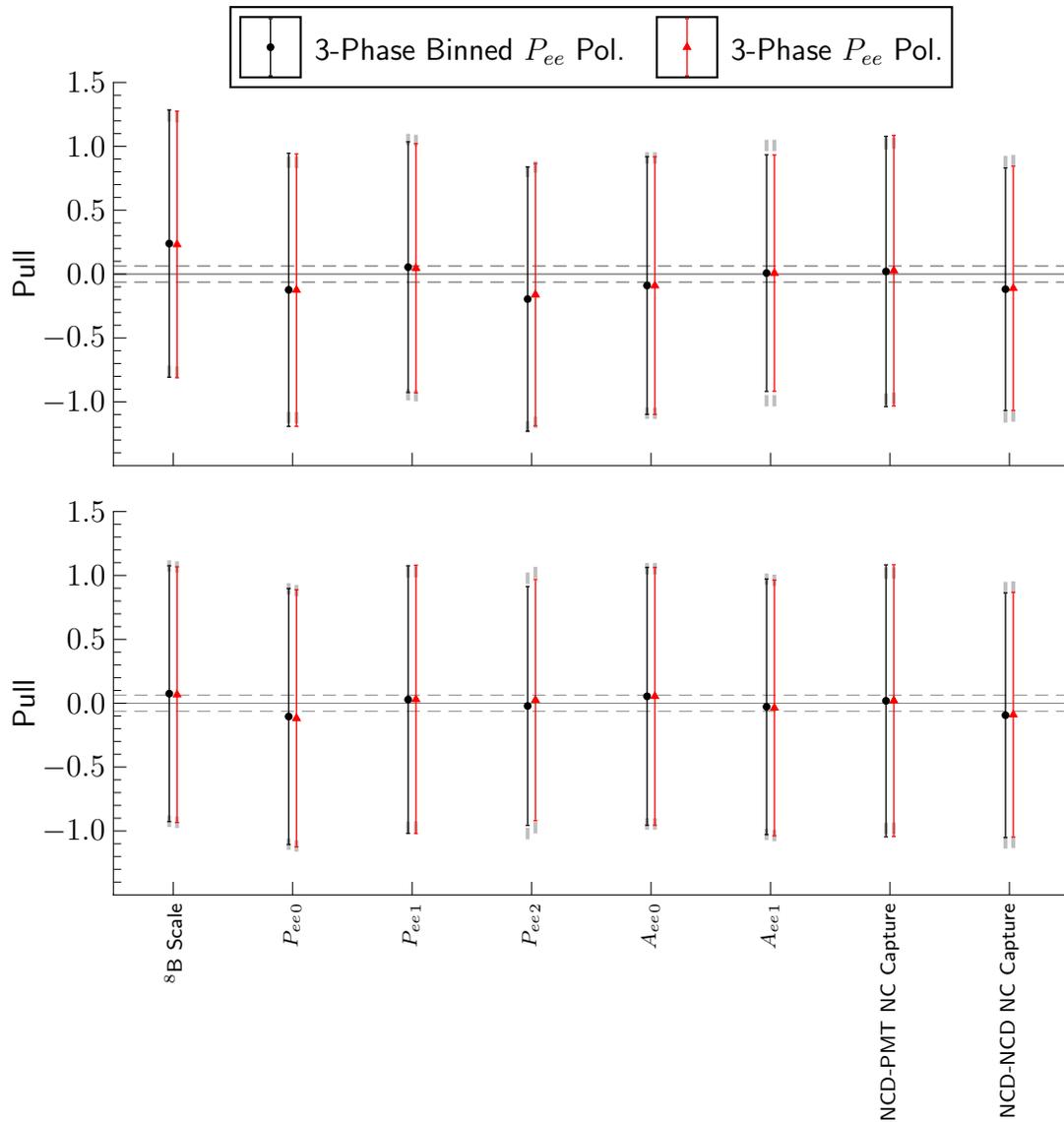
Figure 6.2: Pulls for ensemble testing of the three-phase analysis (250 signal-only fake sets). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles. The one-sigma bands for the pull mean (dashed lines) and the expected spread (grey strips) are also drawn.
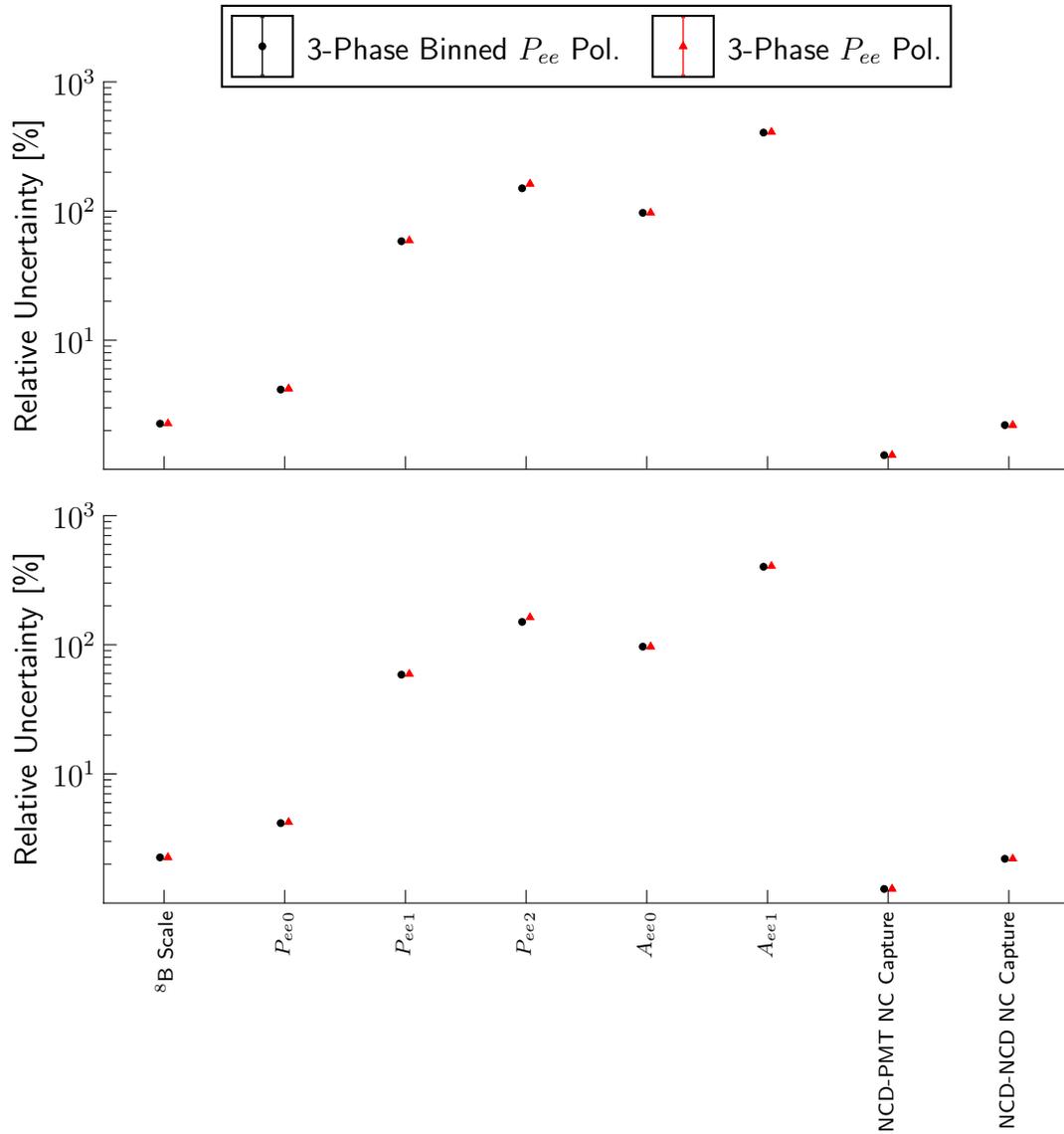
Figure 6.3: Uncertainties for ensemble testing of the three-phase analysis (250 signal-only fake sets). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles.

with a much better sensitivity. Figures 6.4 to 6.6 show the results of this test. Similarly to the results obtained with the signal-only fake sets, these figures show no significant biases for the signal parameters. The pull plots also give results which are within the expected uncertainties.

Since it has now been shown that the inclusion of backgrounds led to good results, it is interesting to look at the fluctuations of the test parameters when computing them using only the first 14 sets from the previous test that was shown to be unbiased, because this is the number of fake sets used when performing full ensemble testing. This gives the results of Figures 6.7 to 6.9. These plots show many more fluctuations between regular and alternate ensemble tests when compared to the previous set, that is to say bias and pull shifts that appear to be further from the origin and more unstable pull spreads compared to the expected fluctuations from Gaussian statistics.

6.4.1.3   Full Fake Set Testing

Ensemble testing using fake sets that include all the backgrounds allows to simulate the real data fit as closely as possible. However, the limited available Monte Carlo statistics, particularly for Bi and Tl events of LETA, allows to generate only 14 fake sets. Such a test is thus expected to provide test variables that fluctuate quite significantly. Since only a few full fake sets can be generated, the uncertainties on these variables are also not expected to behave entirely like Gaussian variables, such that the size of the error bars for the biases and pulls shown in Figures 6.10 and 6.11, as well as the size of the dashed band for the pulls, are not expected to model the real fluctuations perfectly. In such a case the comparison between the regular and alternate ensembles acquires a greater
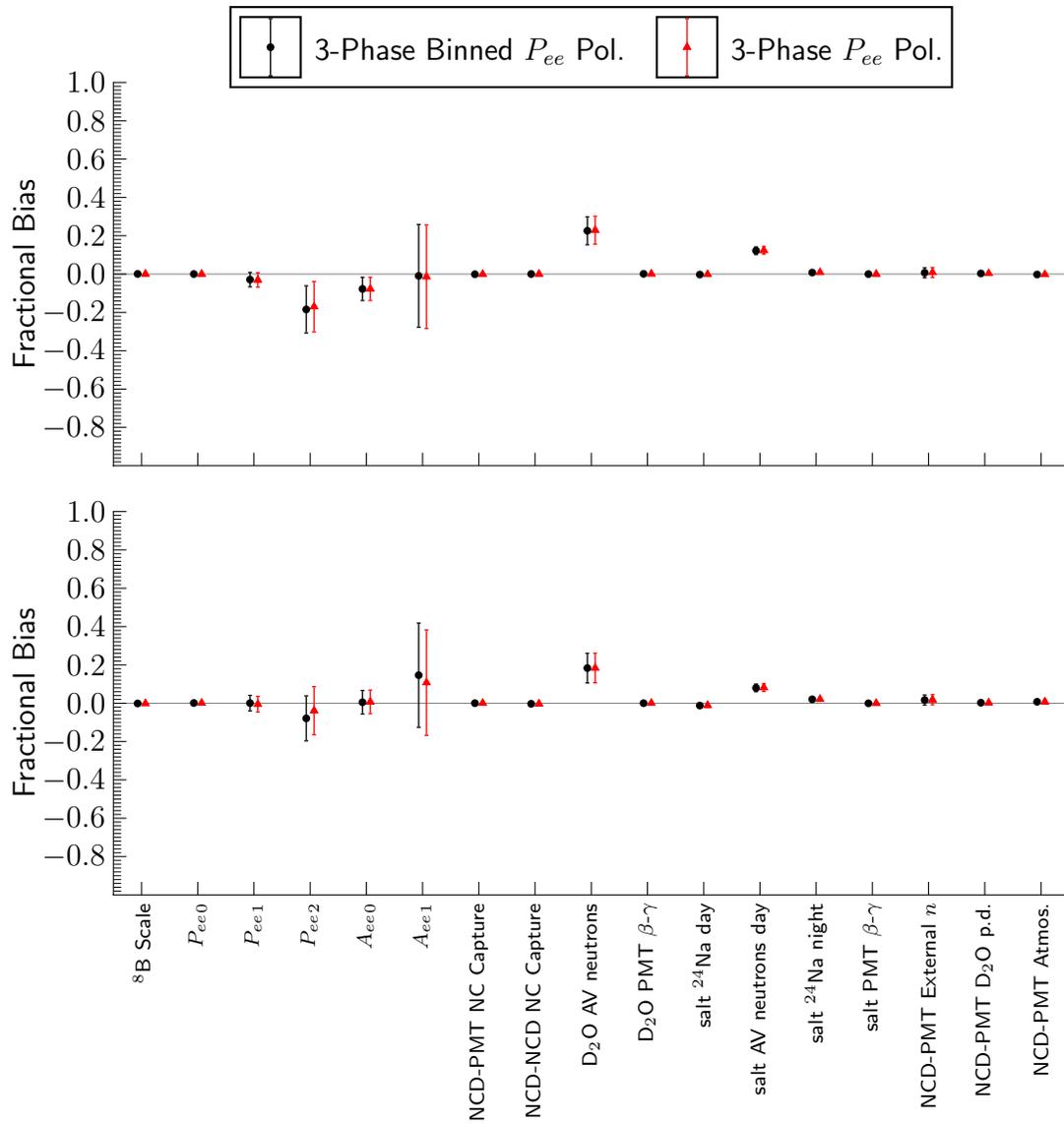
Figure 6.4: Fractional biases for ensemble testing of the three-phase analysis (250 full fake sets excluding Bi and Tl LETA events and neutral counter photodisintegration events). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles.
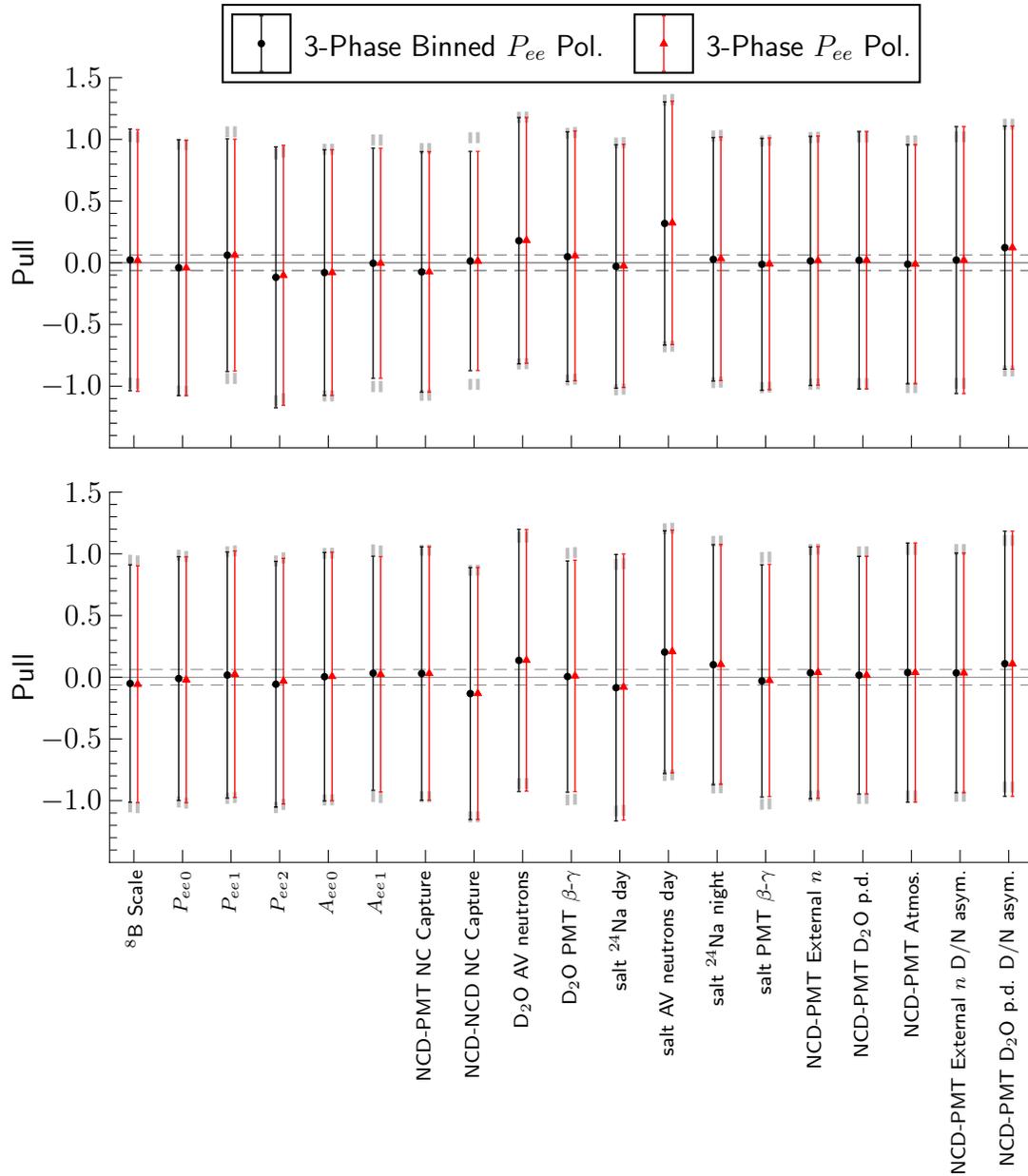
Figure 6.5: Pulls for ensemble testing of the three-phase analysis (250 full fake sets excluding Bi and Tl LETA events and neutral counter photodisintegration events). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles. The one-sigma bands for the pull mean (dashed lines) and the expected spread (grey strips) are also drawn.
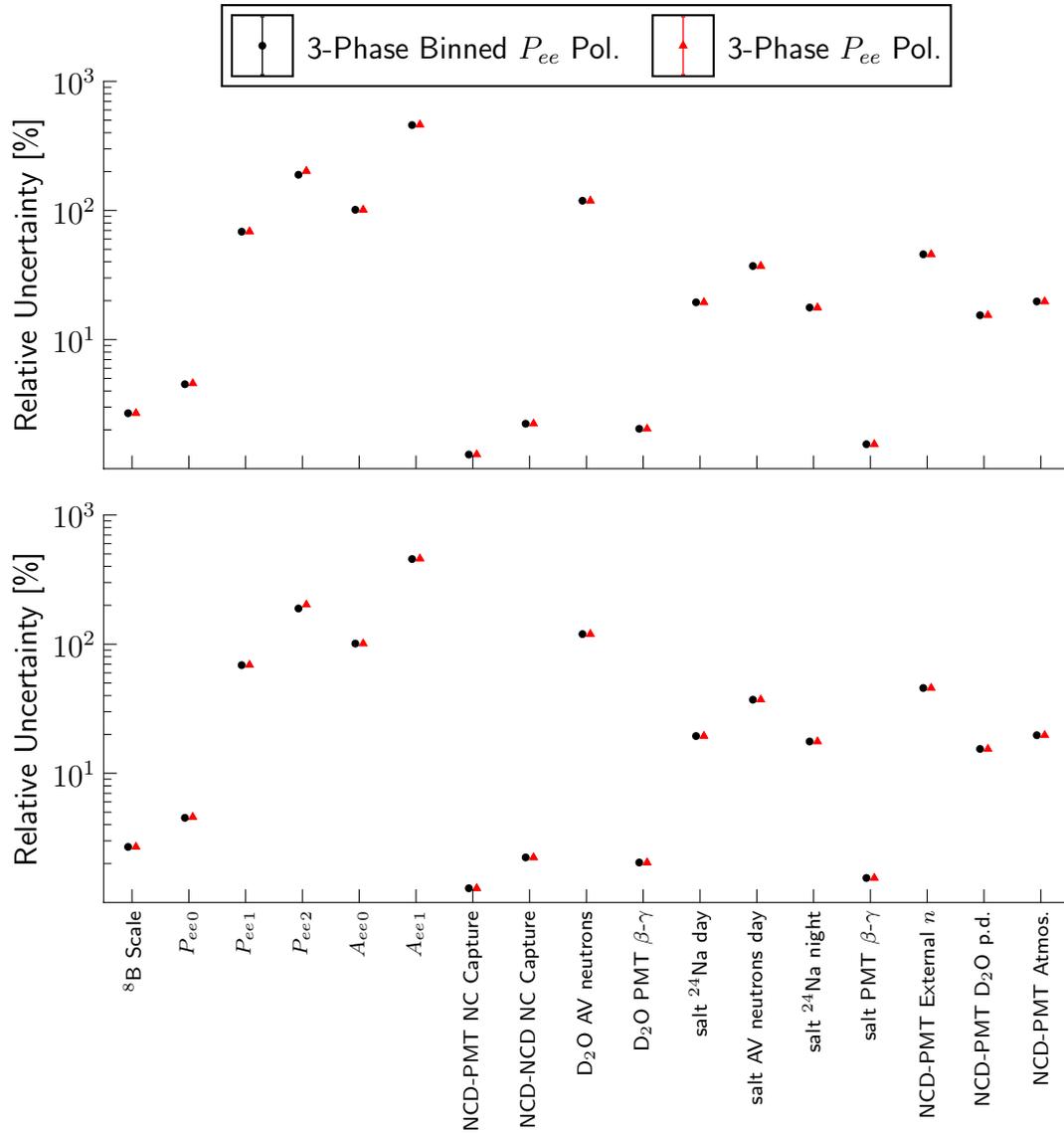
Figure 6.6: Uncertainties for ensemble testing of the three-phase analysis (250 full fake sets excluding Bi and Tl LETA events and neutral counter photodisintegration events). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles.
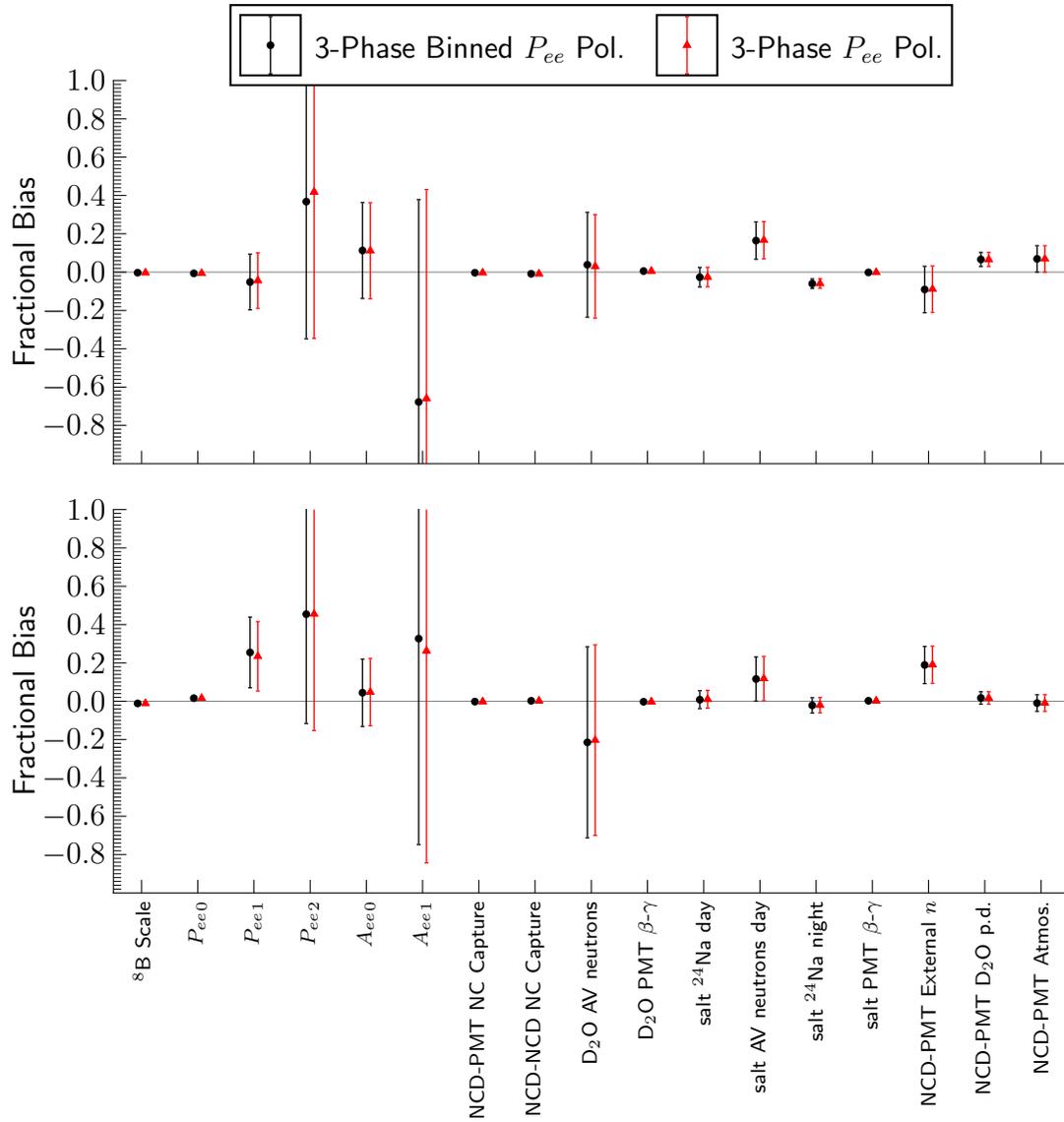
Figure 6.7: Fractional biases for ensemble testing of the three-phase analysis (first 14 full fake sets excluding Bi and Tl LETA events and neutral counter photodisintegration events). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles.
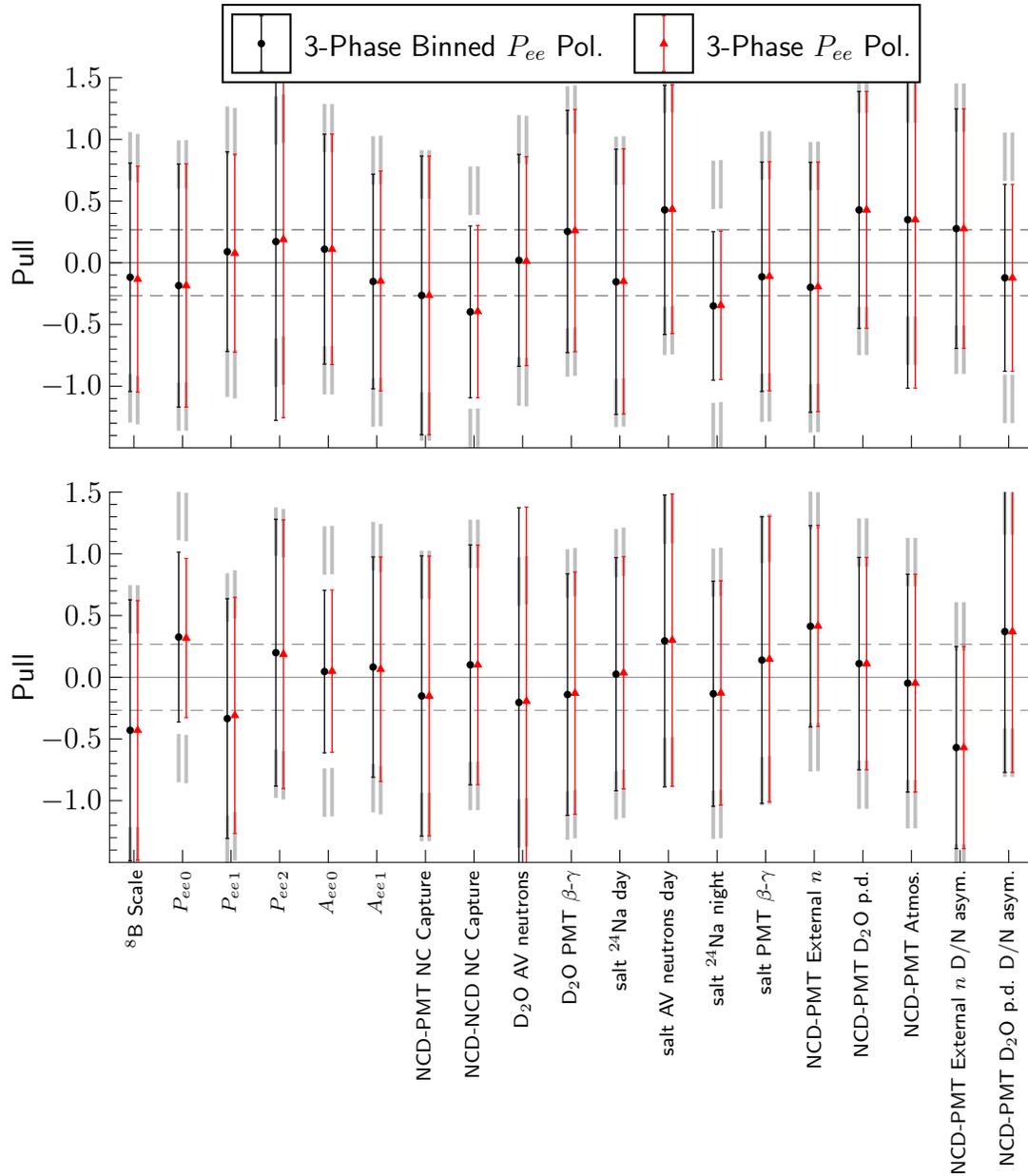
Figure 6.8: Pulls for ensemble testing of the three-phase analysis (first 14 full fake sets excluding Bi and Tl LETA events and neutral counter photodisintegration events). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles. The one-sigma bands for the pull mean (dashed lines) and the expected spread (grey strips) are also drawn.
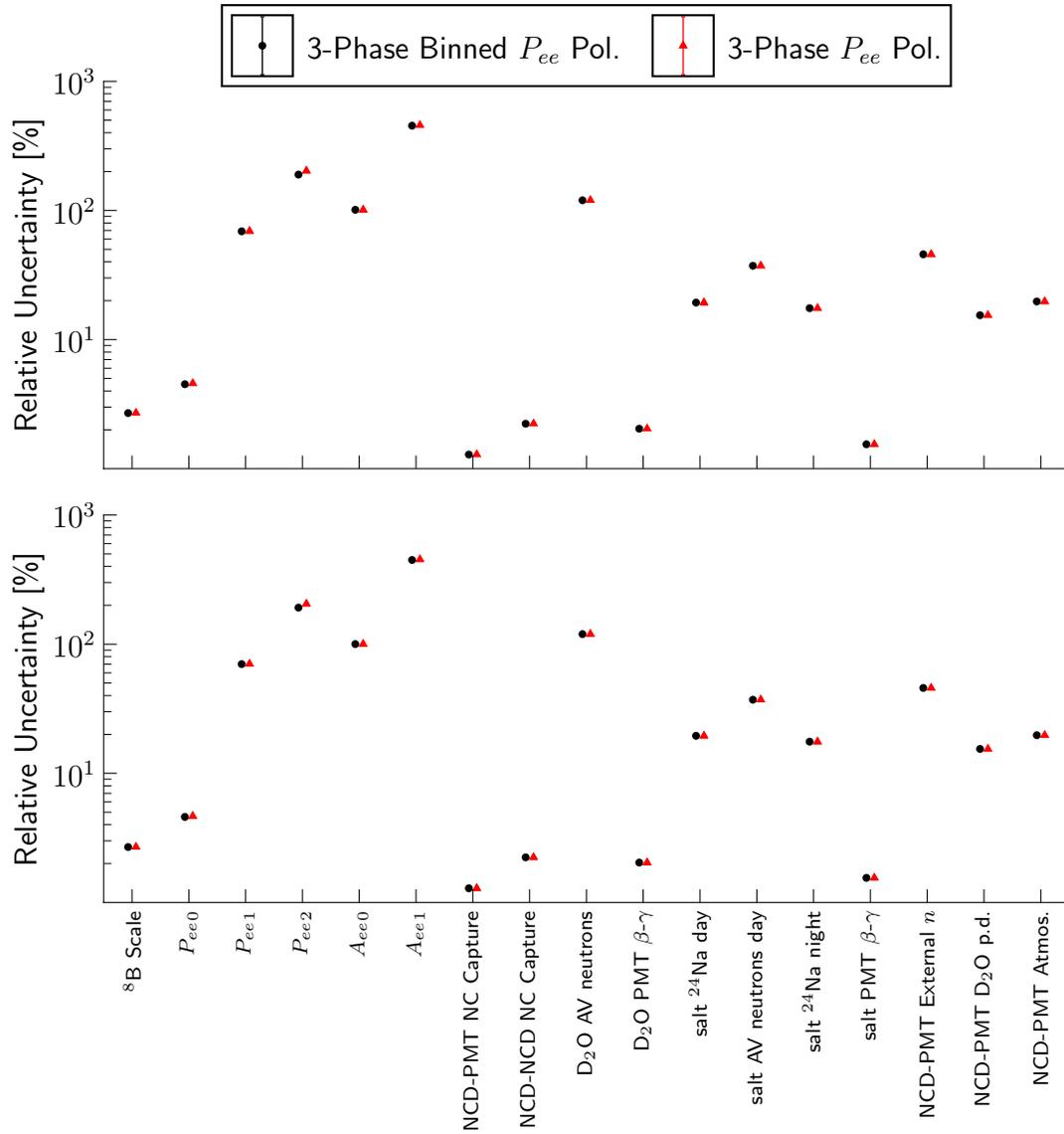
Figure 6.9: Uncertainties for ensemble testing of the three-phase analysis (first 14 full fake sets excluding Bi and Tl LETA events and neutral counter photodisintegration events). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles.

significance. Considering these factors, the results presented for these ensemble tests do not seem to show any significant bias or pull. Overall, ensemble testing with QSigEx thus gives very good results. Both the improved binned and unbinned P_{ee} polynomial fits give similar results that do not show significant biases or pulls. Figure 6.12 presents the corresponding fitted relative uncertainties.

6.4.1.4 Full Fake Testing with Nominal Systematic Uncertainty Shifts and Scanned Systematic Uncertainty Parameters

The ensemble tests performed in this subsection are quite different from the ones that have been presented so far:

1. Although their fake sets were generated using the same P_{ee} and Poisson parameters (that is to say the same expected number of events when using the default skimmed Monte Carlo) as the other ensemble tests, the observables of the sets were smeared by all of the nominal systematic uncertainty shifts (except the day/night asymmetries).

2. Instead of generating the maximum number of sets for each class of events, 14 sets were generated for each class to maximise the Monte Carlo statistics used to build the PDFs.

3. Full scanning and zooming of the systematic uncertainty parameters were performed. The higher statistics in the PDFs ensure smoother scans of the systematic uncertainty parameters, because it is less likely to have a data event that suddenly falls into a bin that is empty for all of the PDFs while scanning.
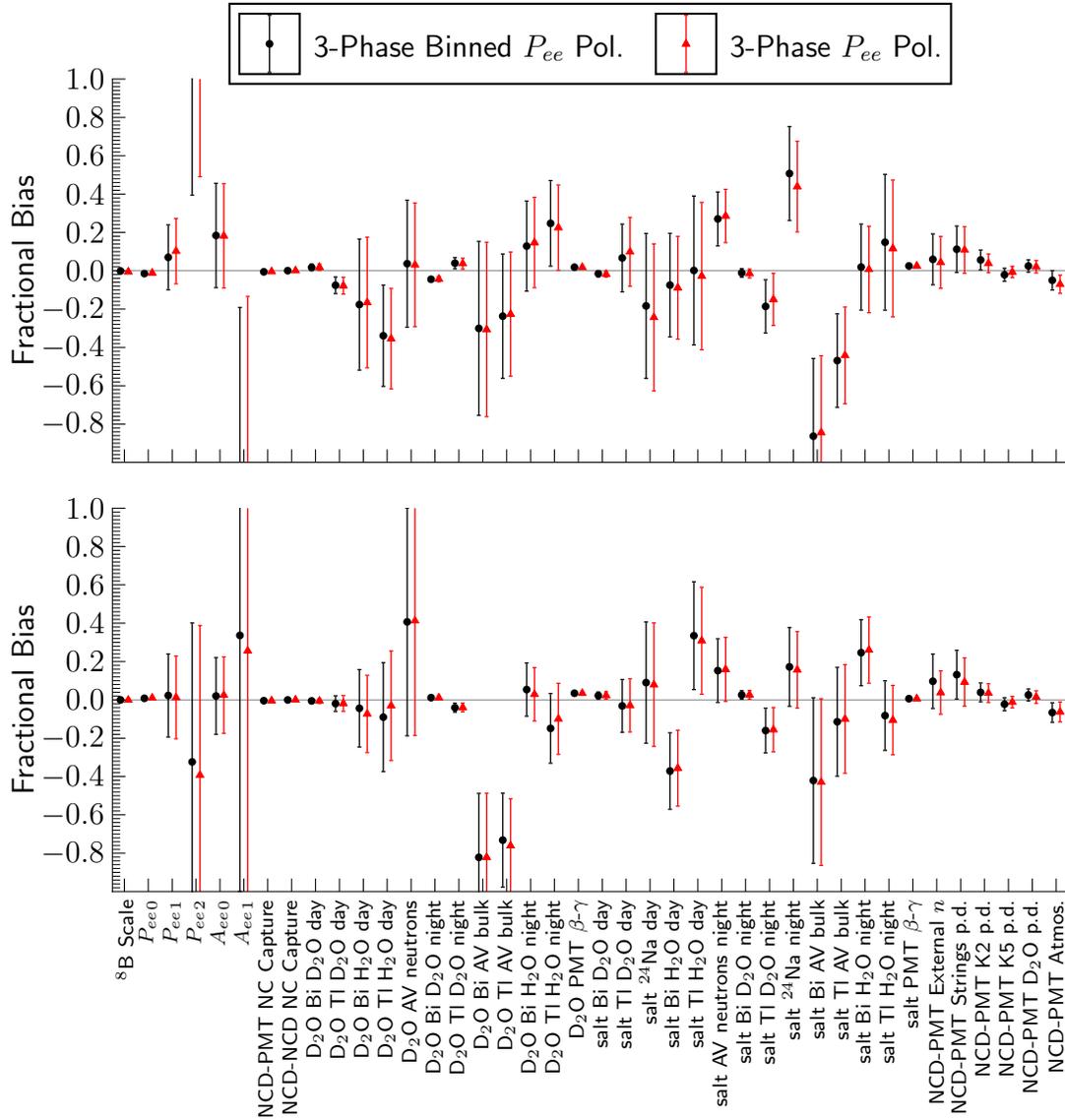
Figure 6.10: Fractional biases for ensemble testing of the three-phase analysis (14 full fake sets). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles.
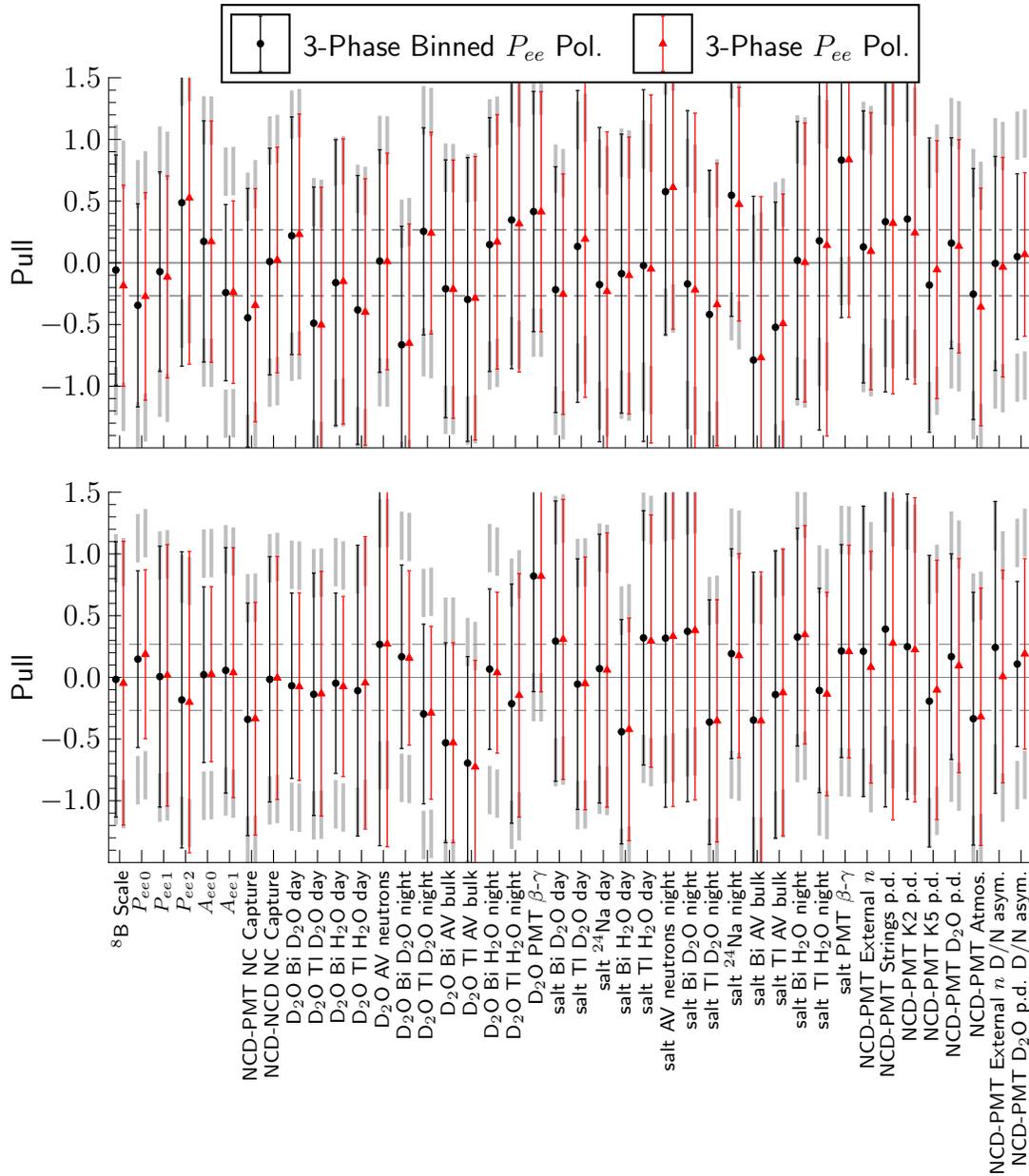
Figure 6.11: Pulls for ensemble testing of the three-phase analysis (14 full fake sets). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles. The one-sigma bands for the pull mean (dashed lines) and the expected spread (grey strips) are also drawn.
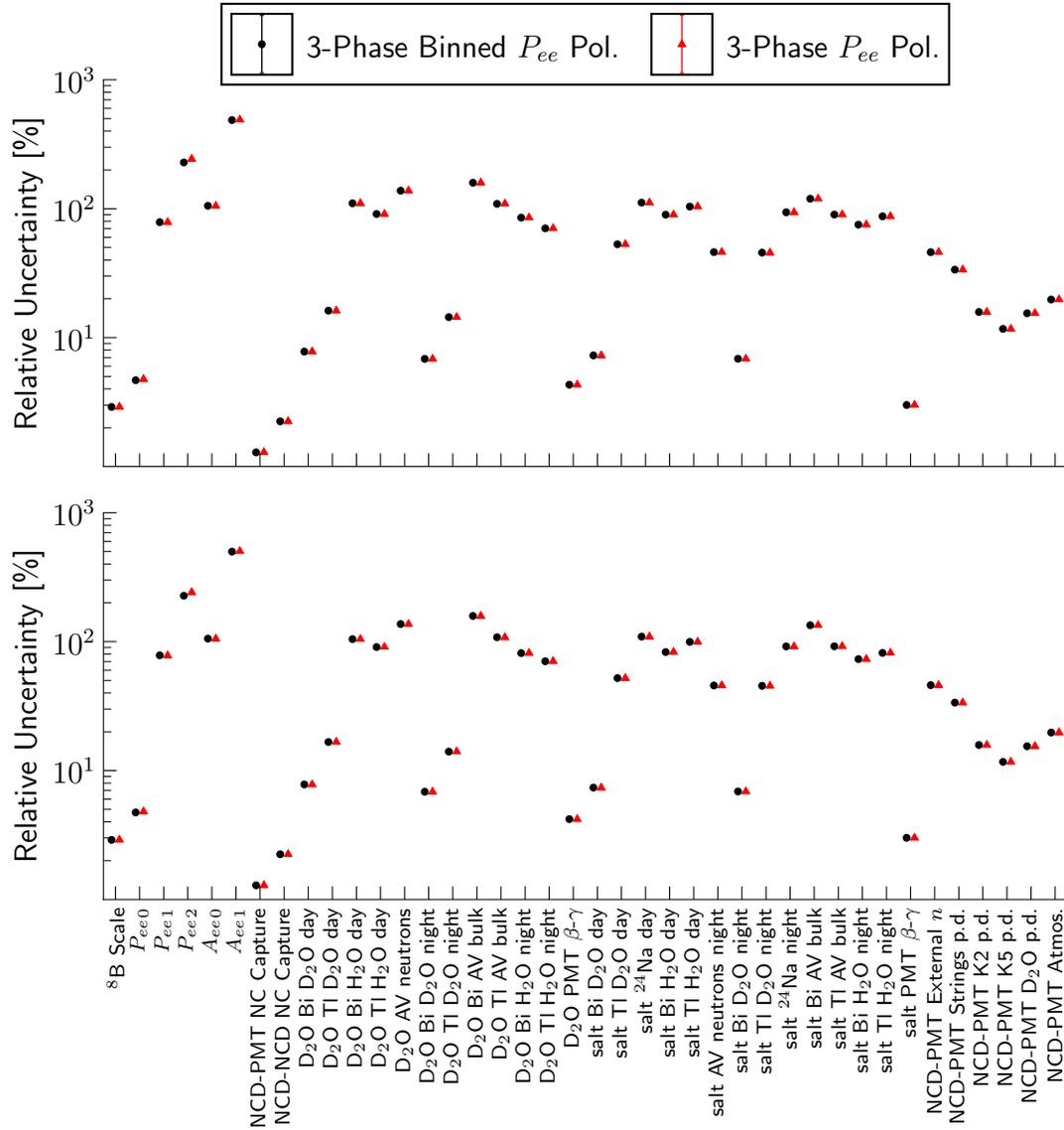
Figure 6.12: Uncertainties for ensemble testing of the three-phase analysis (14 full fake sets). Shown are the results for the improved binned P_{ee} fit and the unbinned P_{ee} fit, for regular (top) and alternate (bottom) ensembles.

4. The means of the penalty terms for the scanned systematic uncertainty parameters were randomly shifted, as were the means of the constraints for the backgrounds and NC detection efficiencies. For a given dataset, these means were shifted consistently for all of the fits involved in the scans. This holds for all of the scanned parameters, except the parameters for the analytical radial PDFs of the LETA PMT β-γ backgrounds, that could easily generate negative probability densities if they were shifted.

5. After scanning and zooming the systematic uncertainty parameters, their effects on the fit parameters were evaluated by a shift-and-refit procedure using the results from the scans.

6. The ensemble test parameters were then computed using total uncertainties computed from the addition in quadrature of the statistical uncertainties and the systematic uncertainties generated by the scanned systematic uncertainty parameters.

7. The whole procedure was repeated twice for the regular and alternate ensemble tests.

This technique should normally lead to ensemble test parameters with the expected distributions, modulo the following factors:

- The means of the 2D penalty terms for the radial PDFs of the LETA PMT β-γ backgrounds are not randomly shifted as mentioned above and this will tend to reduce the pull spreads.

- The test parameters do not have perfectly Gaussian distributions since they are computed using the results from only 14 fits, as is always the case when ensemble testing with the full fake datasets.

- Since for most signals, a large majority of the Monte Carlo statistics are used to generate the PDFs, the comparison between the regular and alternate ensembles must be done while considering that the observed discrepancies are mainly due to the fake data statistics and not to the PDF statistics. To consider the latter effects, one must consider the uncertainties due to the finite Monte Carlo statistics.

- Some systematic uncertainty parameters are constrained in an asymmetrical way by the data, such that the average values for the fit parameters might not correspond to the most likely values (the output from the minimisation of the likelihood function). This could thus create biases that would not indicate problems in signal extraction.

The results for the signal parameters of the ensemble tests described above, when using the improved binned P_{ee} fit method, are shown in Figures 6.13 to 6.15. Overall the results show no significant bias or pull. As expected, the results from the regular and alternate ensembles are more correlated due to the usage of essentially the same PDF statistics. The pull means shown in Figure 6.14 are all reasonable. The pull spreads for the alternate ensemble appear to be a bit small, but this can be explained by the reduced number of sets used to compute the test parameters such that their distributions are less Gaussian.
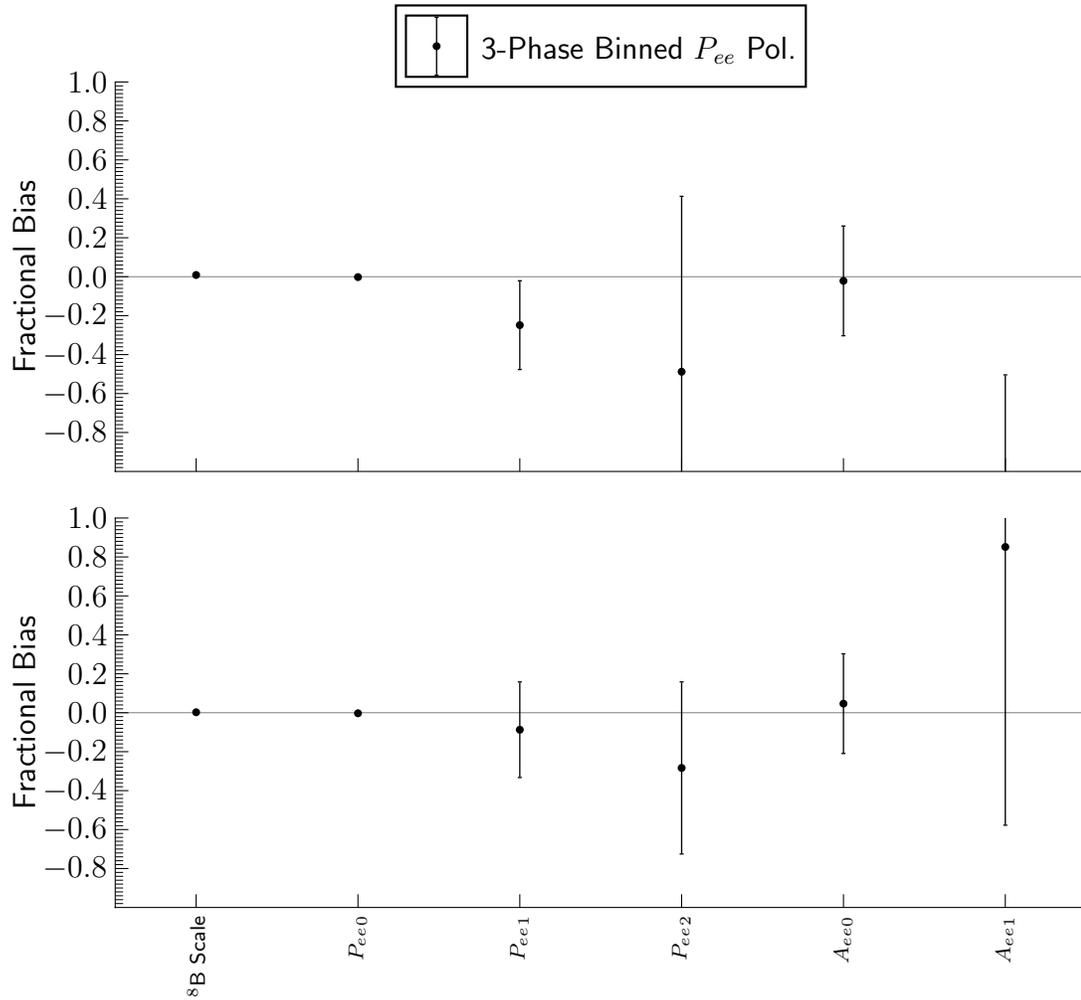
Figure 6.13: Fractional biases for ensemble testing of the three-phase analysis with nominal systematic uncertainty shifts and scanned systematic uncertainty parameters. Shown are the results for the improved binned P_{ee} fit, for regular (top) and alternate (bottom) ensembles.
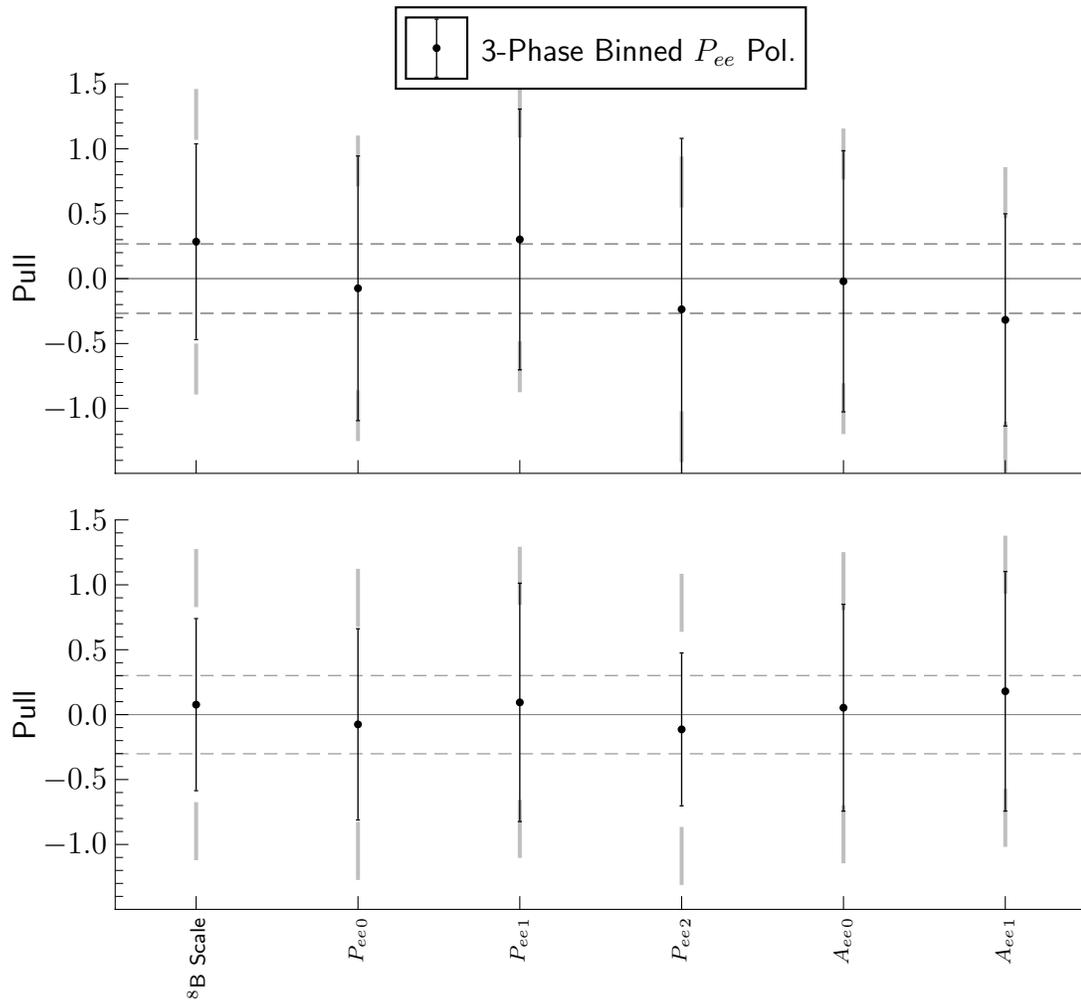
Figure 6.14: Pulls for ensemble testing of the three-phase analysis with nominal systematic uncertainty shifts and scanned systematic uncertainty parameters. Shown are the results for the improved binned P_{ee} fit, for regular (top) and alternate (bottom) ensembles. The one-sigma bands for the pull mean (dashed lines) and the expected spread (grey strips) are also drawn.
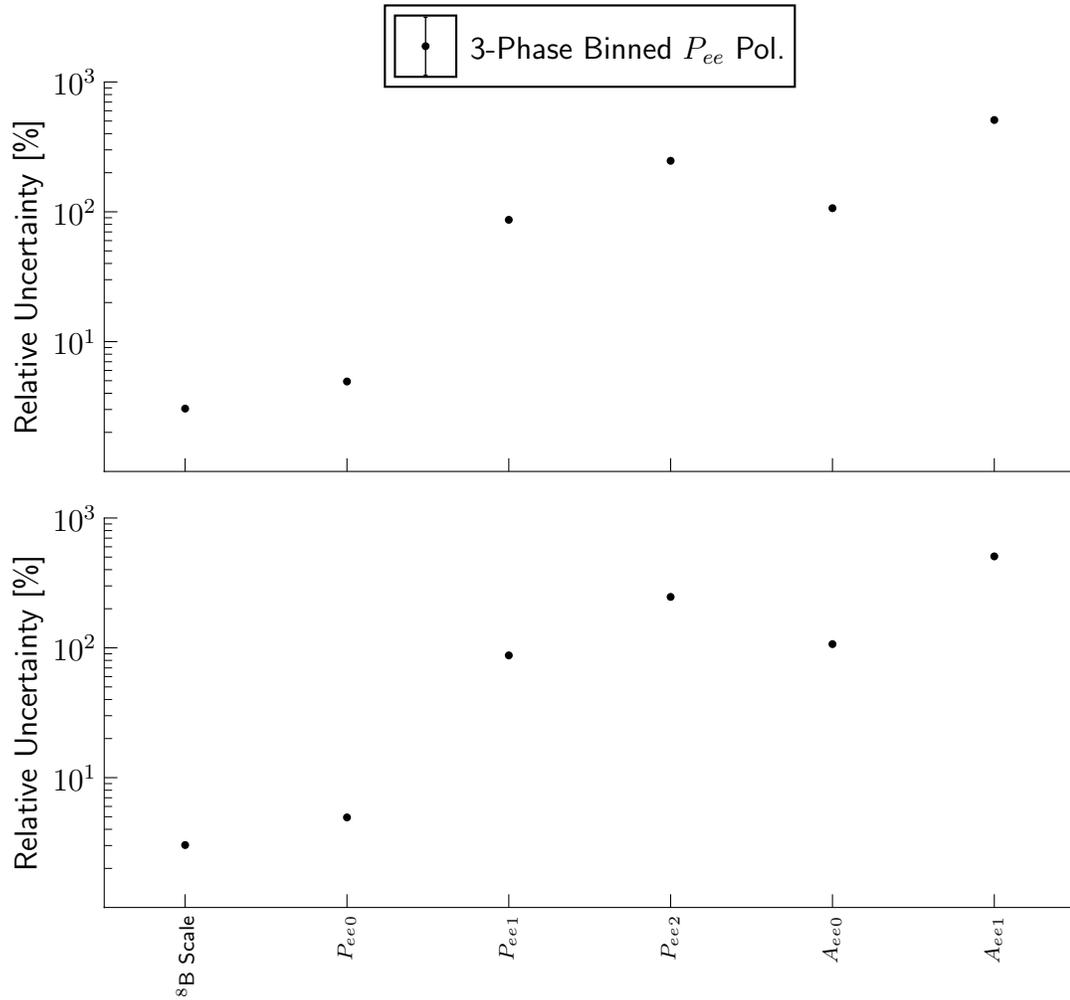
Figure 6.15: Uncertainties for ensemble testing of the three-phase analysis with nominal systematic uncertainty shifts and scanned systematic uncertainty parameters. Shown are the results for the improved binned P_{ee} fit, for regular (top) and alternate (bottom) ensembles.

6.4.2 $P_{ee} + P_{ea}$ Polynomial Fits

This section presents the ensemble test results obtained using the $P_{ee} + P_{ea}$ fit methods.

6.4.2.1 Signal-Only Testing

Figures 6.16 to 6.18 present the results when ensemble testing the three-phase analysis using the improved binned $P_{ee} + P_{ea}$ and the unbinned $P_{ee} + P_{ea}$ fit methods, with 250 signal-only fake sets. Both methods give similar results that do not show significant biases or pulls. As expected, the uncertainties on the P_{ea} parameters are huge (or even off-scale), due to the limited $ES_{\mu\tau}$ statistics, but the ones on P_{ee} are not necessarily increased compared to the P_{ee} fits, although this comparison is not very meaningful since the ${}^8B$ scale has to be fixed in signal extraction when testing a sterile model.

6.4.2.2 Signal with High-Statistics Monte Carlo Backgrounds

As it was the case for the P_{ee} fit, ensemble testing using only backgrounds having high-statistics Monte Carlo was performed for the $P_{ee} + P_{ea}$ fit, in order to test the behaviour of the signal parameters in the presence of backgrounds using the highest possible sensitivity. Figure 6.19 to 6.21 show the results from the regular pair of ensemble tests. The results do not show any anomaly.
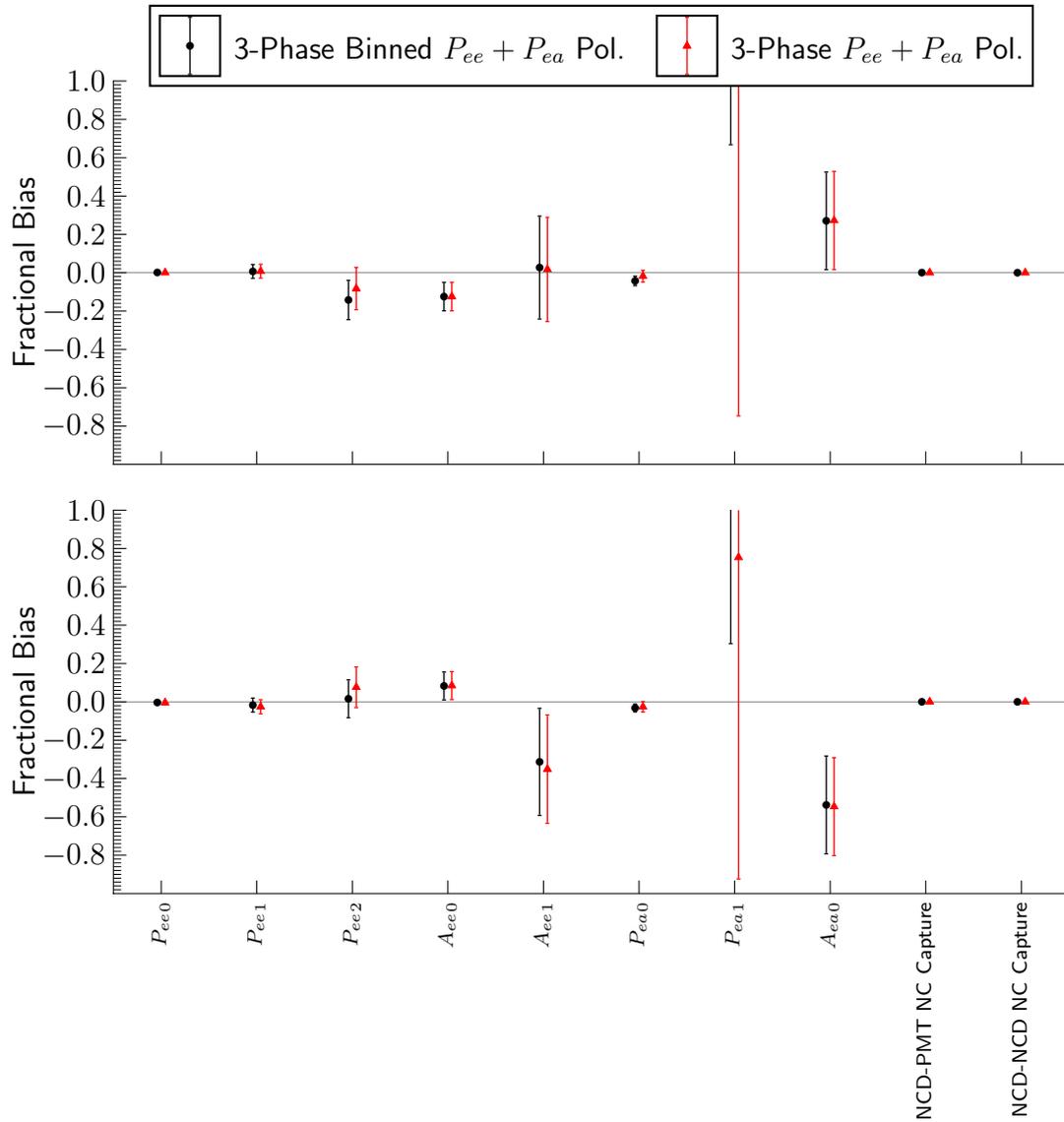
Figure 6.16: Fractional biases for ensemble testing of the three-phase analysis (250 signal-only fake sets). Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit and the unbinned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles.
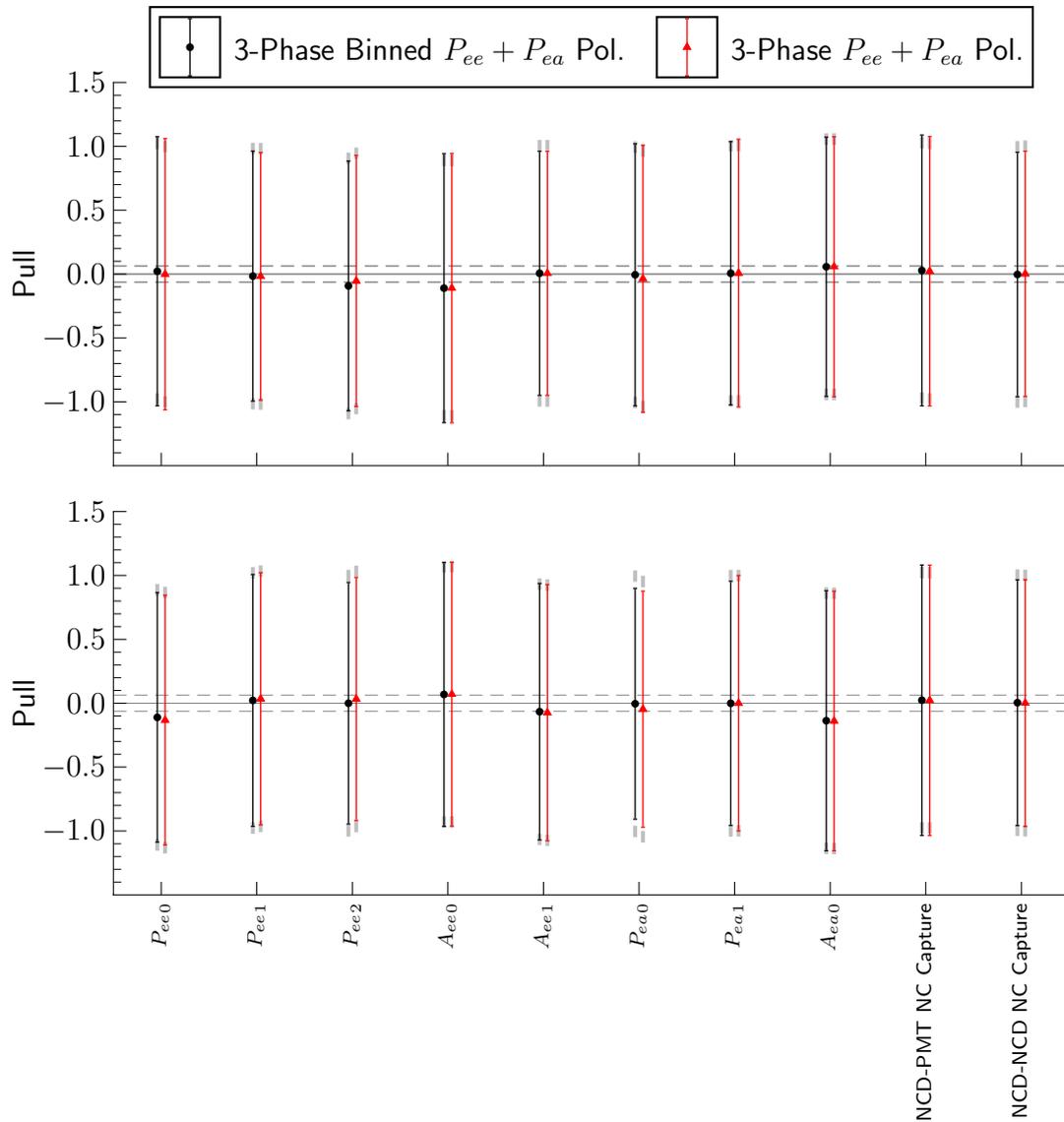
Figure 6.17: Pulls for ensemble testing of the three-phase analysis (250 signal-only fake sets). Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit and the unbinned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles. The one-sigma bands for the pull mean (dashed lines) and the expected spread (grey strips) are also drawn.
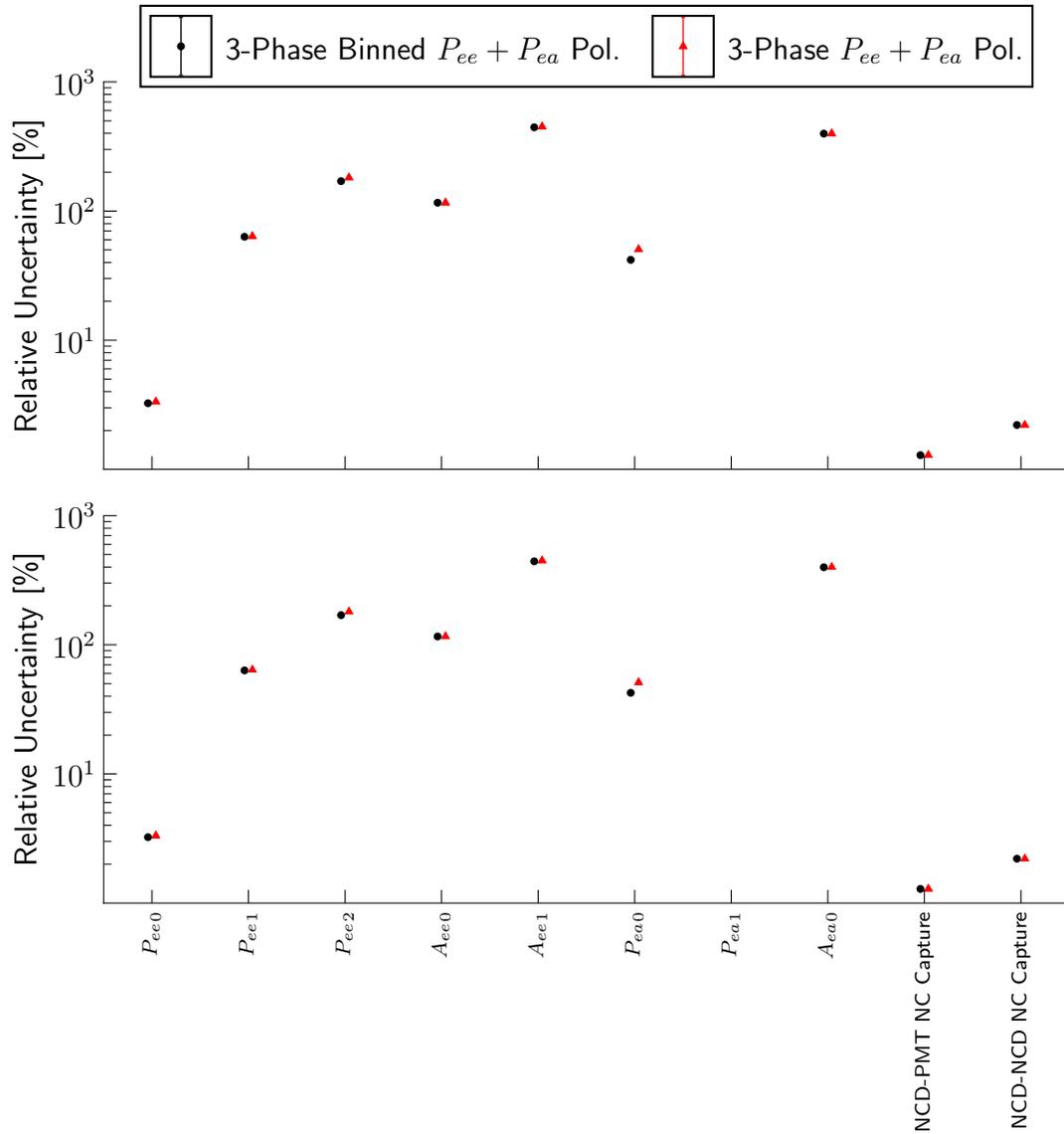
Figure 6.18: Uncertainties for ensemble testing of the three-phase analysis (250 signal-only fake sets). Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit and the unbinned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles.
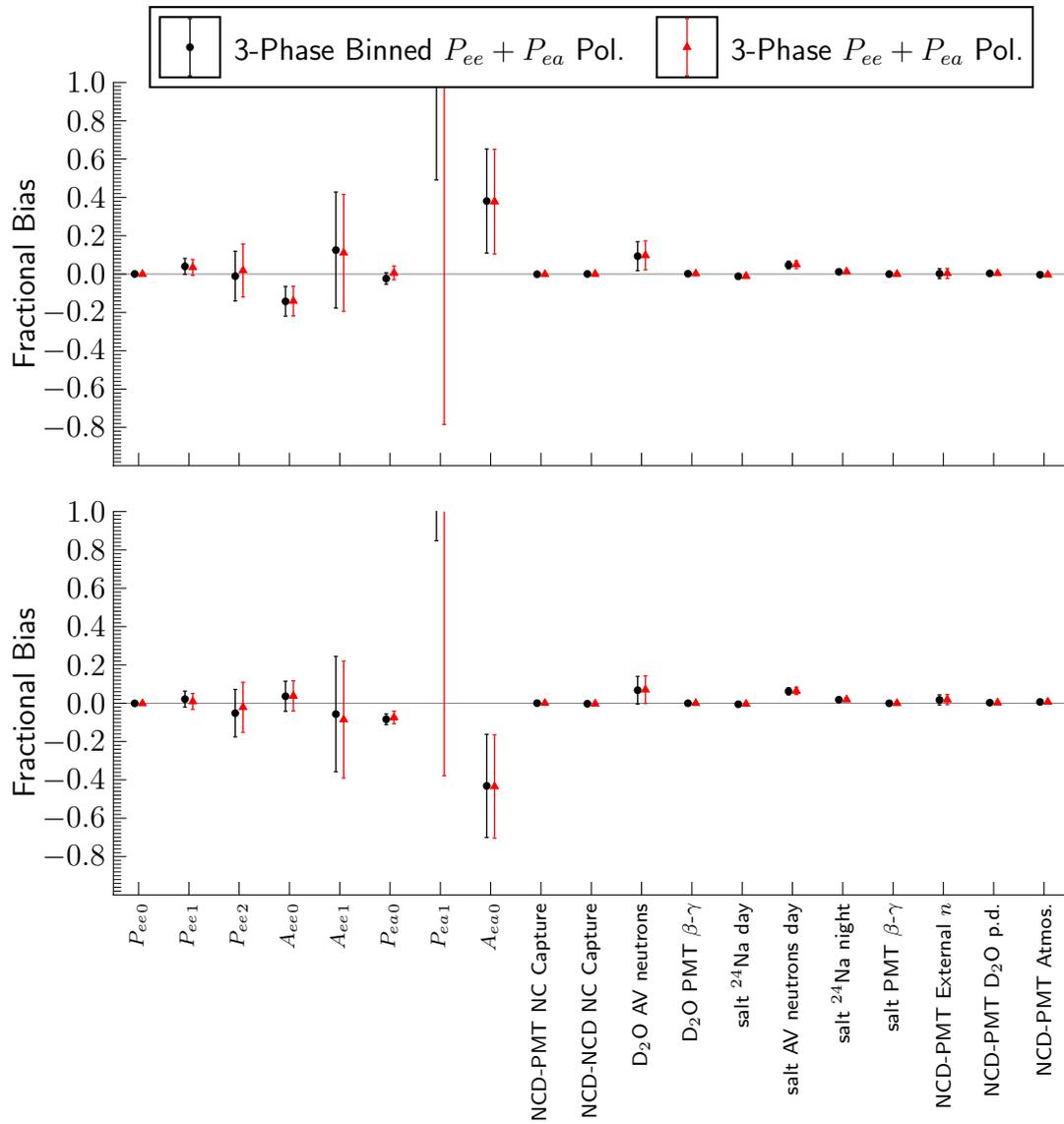
Figure 6.19: Fractional biases for ensemble testing of the three-phase analysis (250 full fake sets excluding Bi and Tl LETA events and neutral counter photodisintegration events). Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit and the unbinned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles.
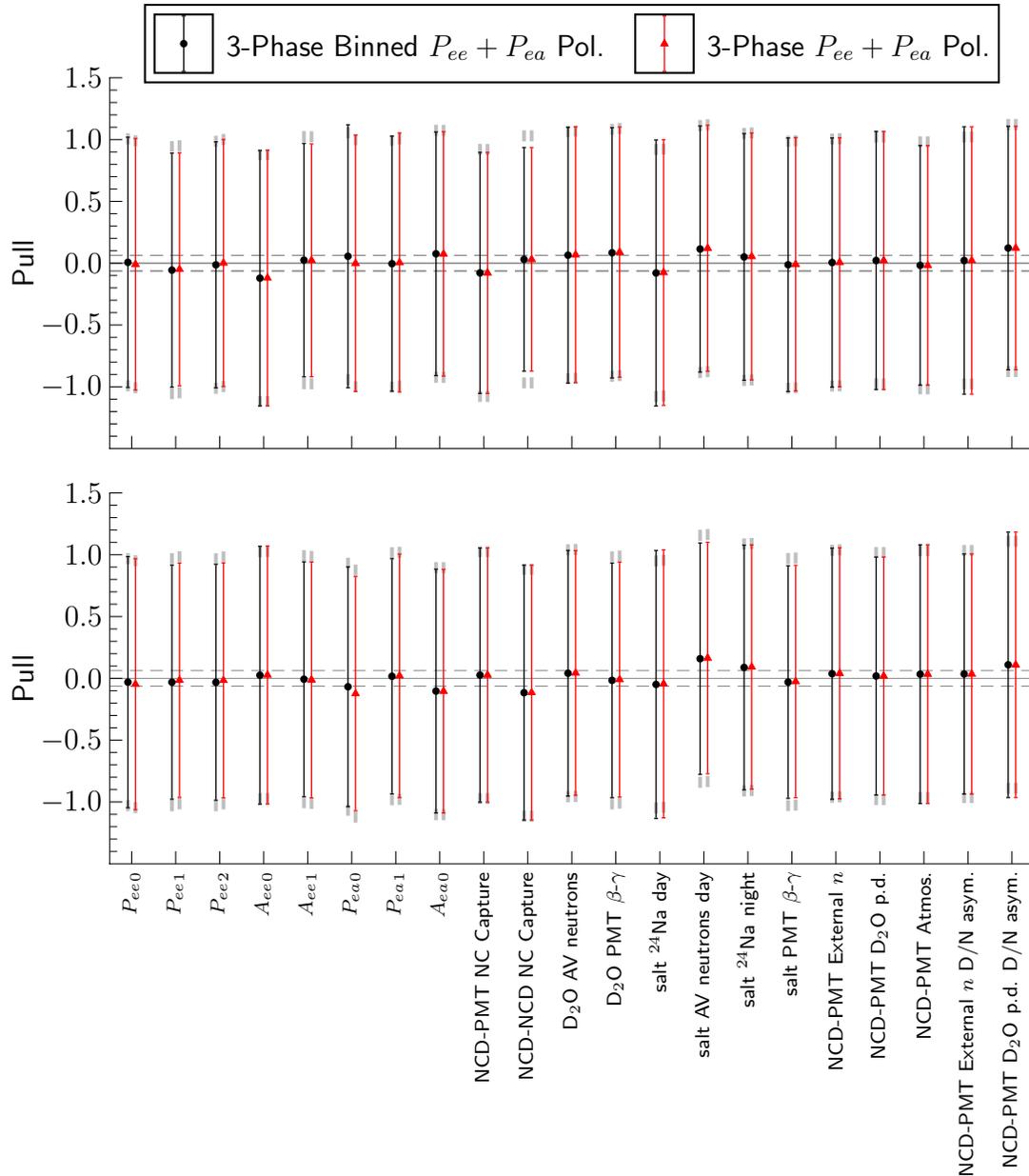
Figure 6.20: Pulls for ensemble testing of the three-phase analysis (250 full fake sets excluding Bi and Tl LETA events and neutral counter photodisintegration events). Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit and the unbinned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles. The one-sigma bands for the pull mean (dashed lines) and the expected spread (grey strips) are also drawn.
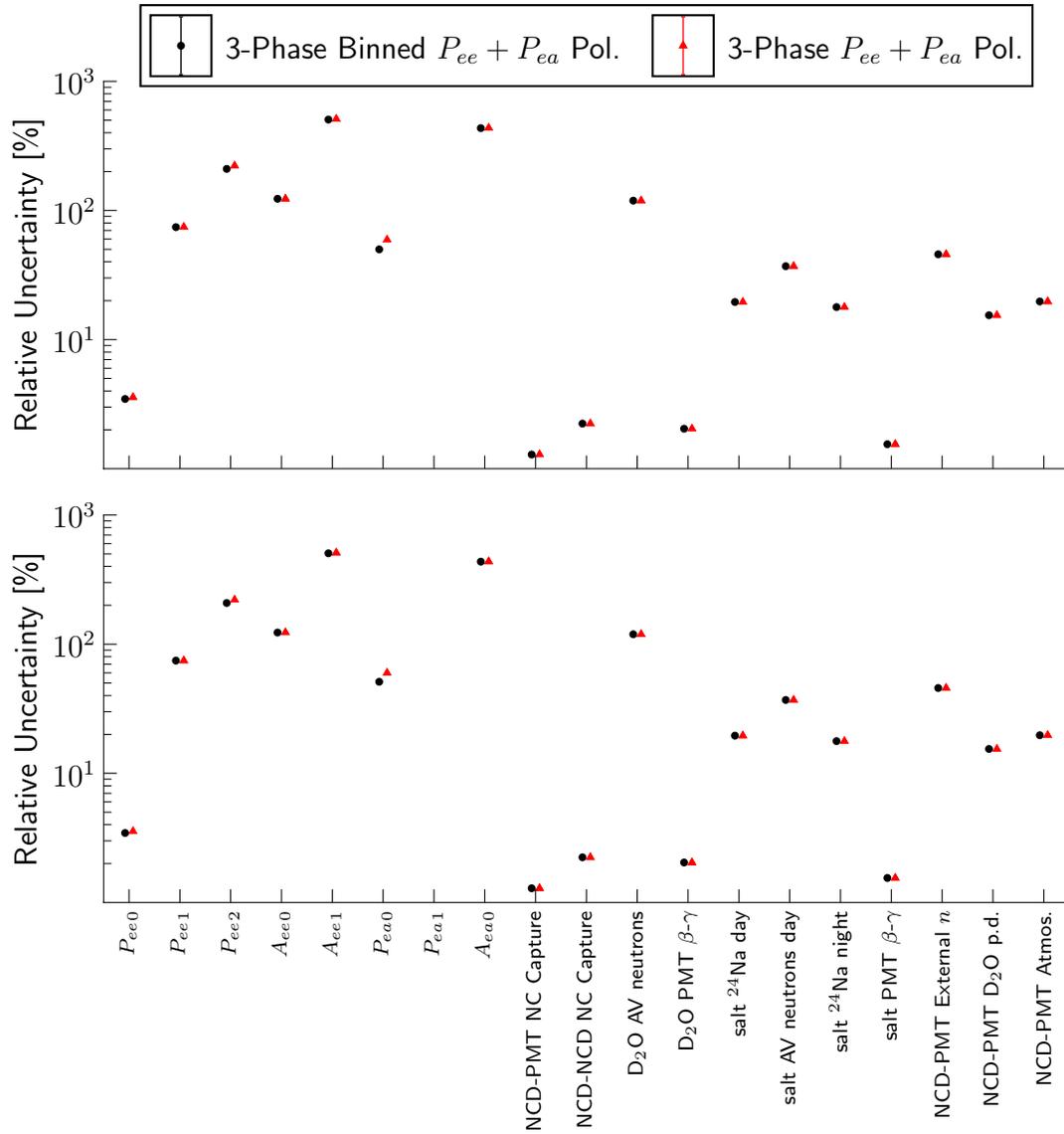
Figure 6.21: Uncertainties for ensemble testing of the three-phase analysis (250 full fake sets excluding Bi and Tl LETA events and neutral counter photodisintegration events). Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit and the unbinned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles.

6.4.2.3 Full Fake Set Testing

Figures 6.22 to 6.24 present the results when ensemble testing the three-phase analysis using the improved binned $P_{ee} + P_{ea}$ and the unbinned $P_{ee} + P_{ea}$ fit methods, with 14 full fake sets. The results for both methods are similar and do not show significant biases or pulls.

6.4.2.4 Full Fake Set Testing with Nominal Systematic Shifts and Scanned Systematic Uncertainties

This section presents results from ensemble testing of $P_{ee} + P_{ea}$ fits where the systematic uncertainty parameters were scanned, using the same procedure described in Section 6.4.1.4. Figures 6.25 to 6.27 present the corresponding results for the signal parameters. These plots do not show significant bias or pull anomalies.
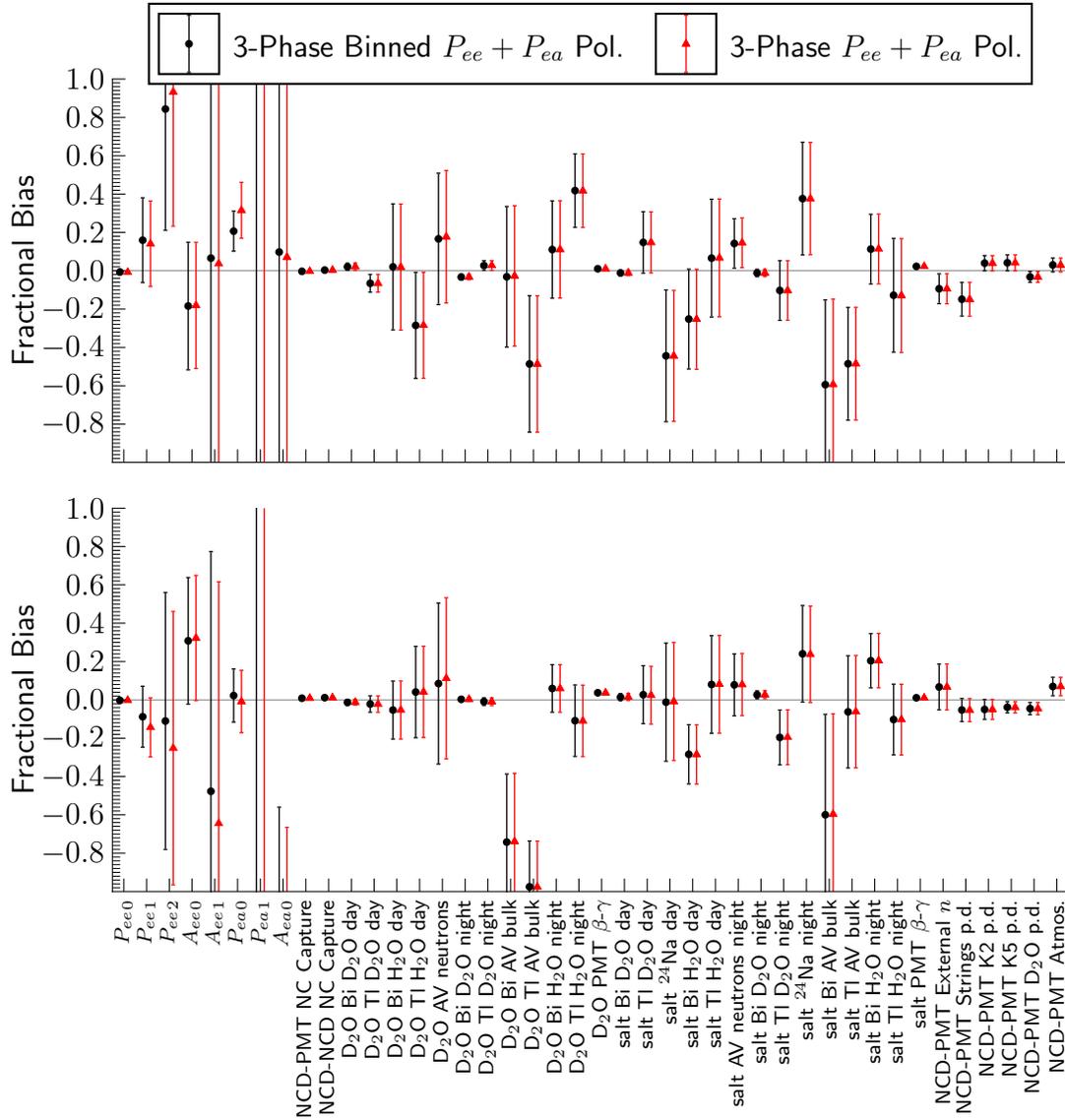
Figure 6.22: Fractional biases for ensemble testing of the three-phase analysis (14 full fake sets). Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit and the unbinned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles.
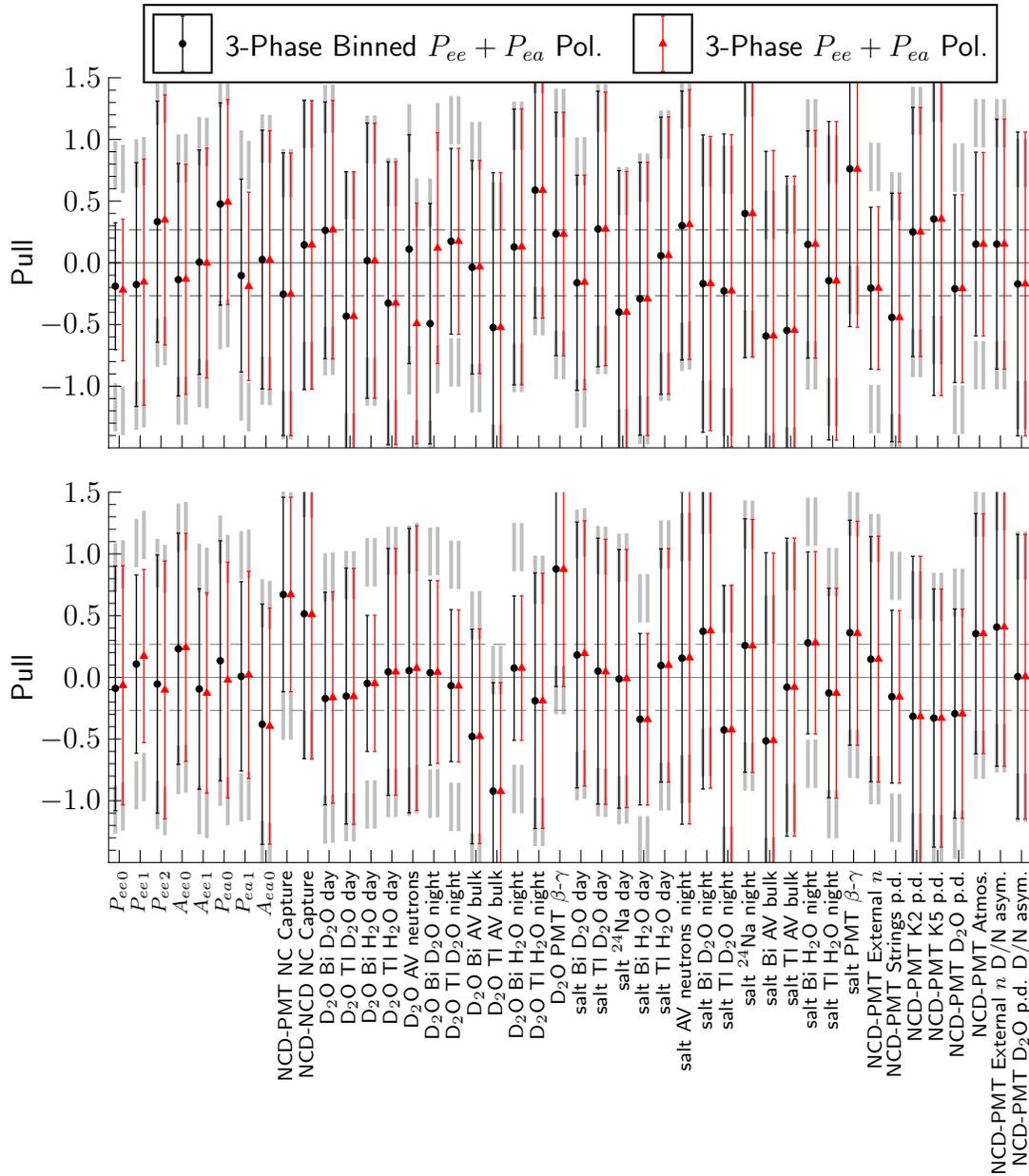
Figure 6.23: Pulls for ensemble testing of the three-phase analysis (14 full fake sets). Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit and the unbinned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles. The one-sigma bands for the pull mean (dashed lines) and the expected spread (grey strips) are also drawn.
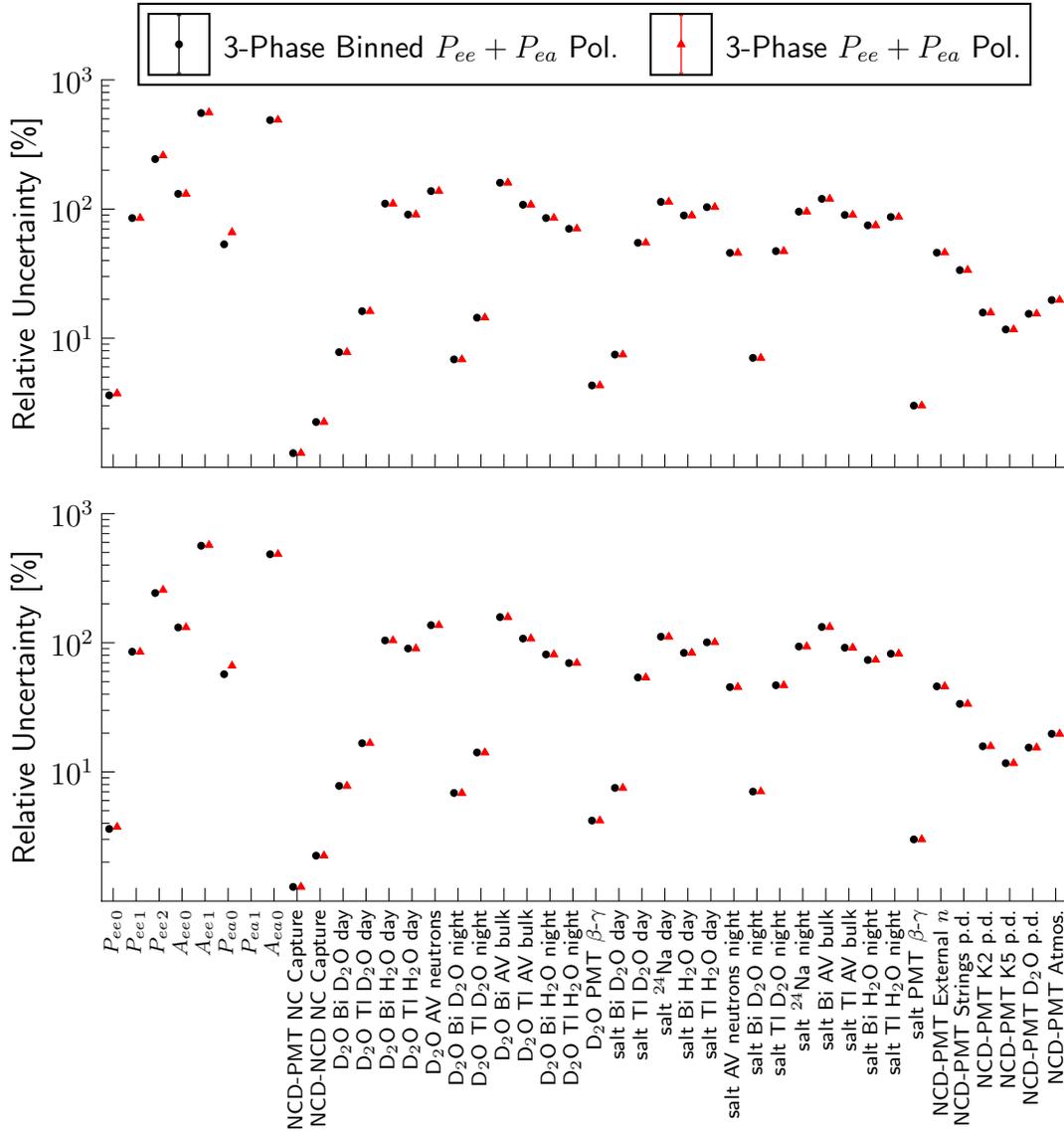
Figure 6.24: Uncertainties for ensemble testing of the three-phase analysis (14 full fake sets). Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit and the unbinned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles.
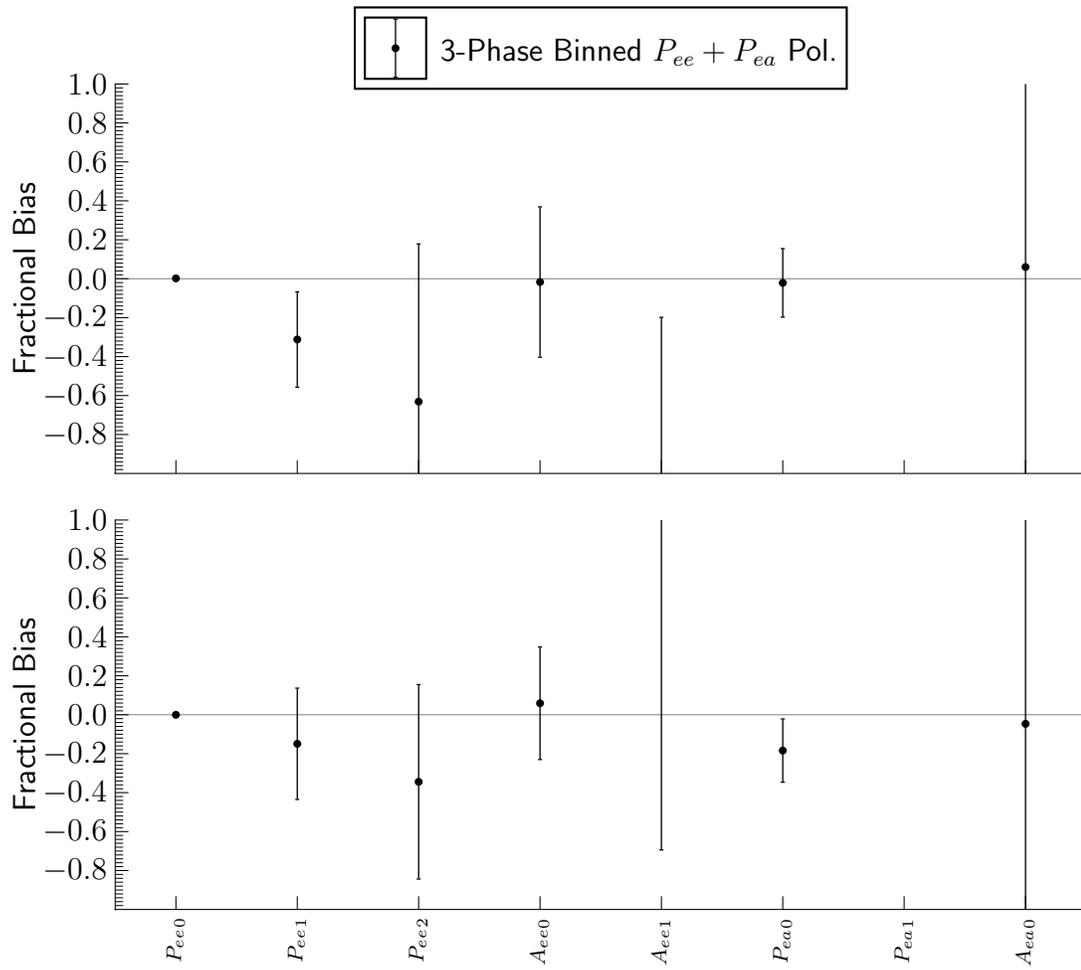
Figure 6.25: Fractional biases for ensemble testing of the three-phase analysis with nominal systematic uncertainty shifts and scanned systematic uncertainty parameters. Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles.
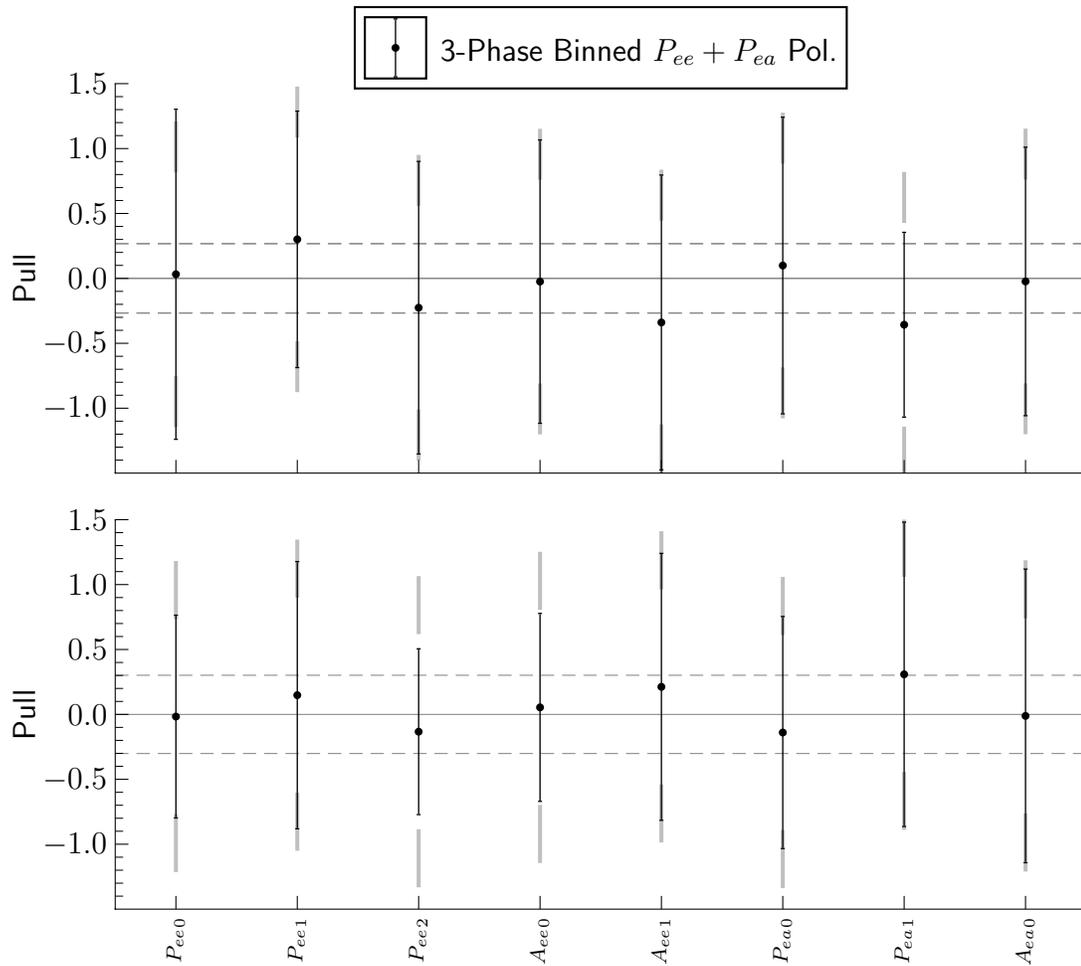
Figure 6.26: Pulls for ensemble testing of the three-phase analysis with nominal systematic uncertainty shifts and scanned systematic uncertainty parameters. Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles. The one-sigma bands for the pull mean (dashed lines) and the expected spread (grey strips) are also drawn.
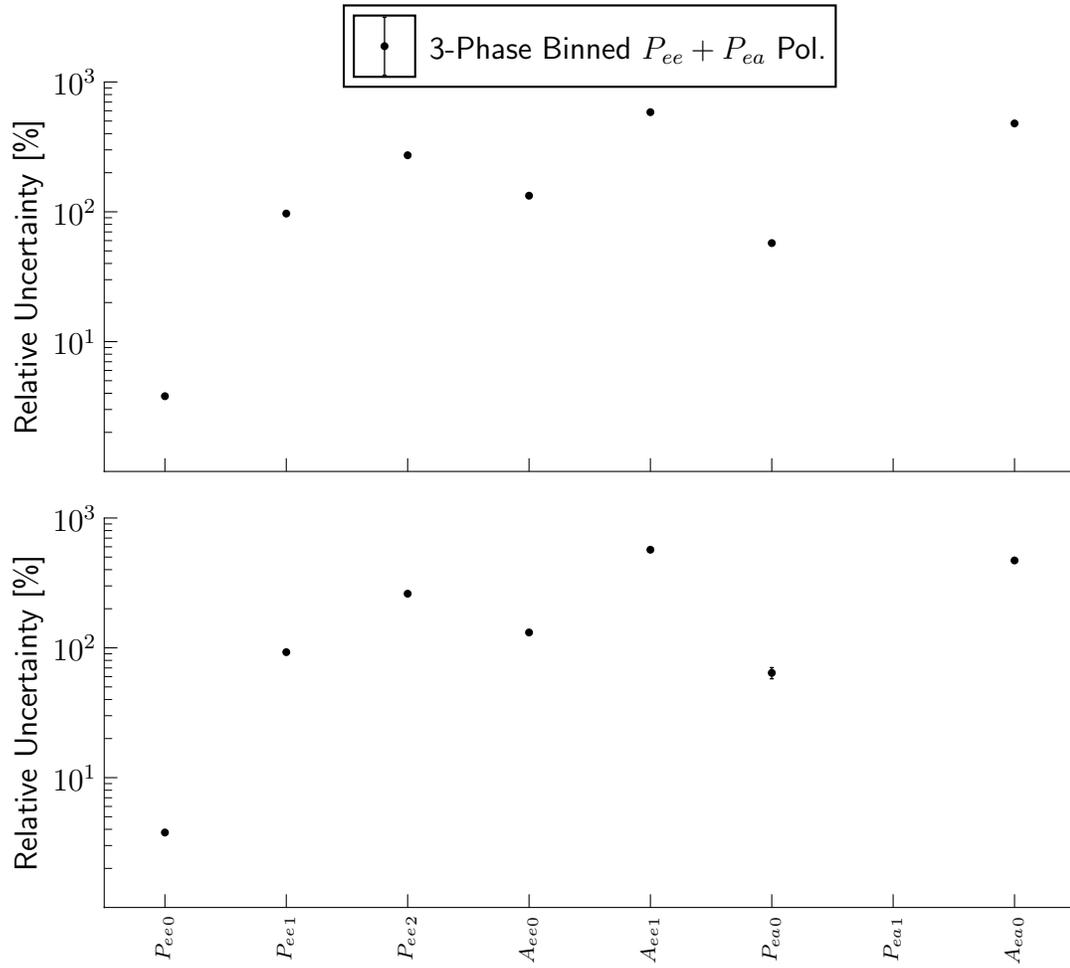
Figure 6.27: Uncertainties for ensemble testing of the three-phase analysis with nominal systematic uncertainty shifts and scanned systematic uncertainty parameters. Shown are the results for the improved binned $P_{ee} + P_{ea}$ fit, for regular (top) and alternate (bottom) ensembles.

Chapter 7

SNO Three-Phase Results

In Chapter 6, the QSigEx three-phase signal extraction procedure was throughly tested and was shown to provide unbiased results with accurate uncertainties. It is now possible to perform three-phase P_{ee} and $P_{ee} + P_{ea}$ fits of the SNO data. Prior to these combined fits, individual fits were performed for LETA and the NCD phase, using configurations of the likelihood function that were very similar to the published results. The resulting fitted parameters were found to be in good agreement with the previous analyses.

This chapter presents the results from the three-phase P_{ee} and $P_{ee} + P_{ea}$ fits, using the configuration described in Appendix D. The configuration for both fits is identical; they only differ in the survival probability parameterisation for the $ES_{\mu\tau}$ and NC signals. The improved binned P_{ee} polynomial method and the unbinned P_{ee} polynomial method have been shown to provide very similar results in Chapter 6, but the former is orders of magnitude faster, while the latter should theoretically be the best. It was thus decided to use the improved binned method to scan the systematic uncertainty parameters and evaluate the systematic effects through the shift and refit procedure, while the unbinned method is reserved for the final fit that extracts the floated parameters and their statistical uncertainties. The parameterisation of the $P_{ee} + P_{ea}$ fit is required to correctly analyse the SNO results using a model that allows a non-active neutrino component. Since the

nominal numbers of detected neutrino events in these fits are based on the prediction
from the BS05(OP) model [18], all ^{8}B scale values in this chapter are relative to the ^{8}B
flux value from this model, which is 5.69×10^6 cm^{-2}s^{-1}. Also, to combine statistical
and systematic uncertainties for analysis purposes, the individual uncertainties are first
symmetrised using

$$\sigma^2 = \sigma_+^2 + \sigma_-^2 - \sigma_+\sigma_-, \tag{7.1}$$

where σ corresponds to the RMS of an asymmetrical Gaussian function when computed
around the function's peak. Total uncertainties are then computed by adding the sym-
metrised statistical and systematic uncertainties in quadrature, which is a standard pro-
cedure [25] that assumes independent statistical and systematic effects.

7.1 P_{ee} Fit

Tables 7.1 and 7.2 present the fitted P_{ee} and ^{8}B scale parameter values and correlation
matrix for the three-phase P_{ee} fit, respectively. In Table 7.1, the "Stat." column presents
the statistical fit uncertainties, which include the effects from all parameters that are
floated. The column "Basic Syst." includes the effects for most of the scanned parameters
as well as the shift and refit systematic uncertainties, except for the following two columns.
Day/night asymmetry parameters effects are shown in the "D/N Syst." column. The
column "MC Effects" shows the uncertainties on the parameters that are due to the finite
Monte Carlo statistics used to generate the PDFs of the likelihood function. Finally, the
column "Total Syst." presents the combined effects from all the systematic uncertainties.
The values of this last column were calculated using the Monte Carlo procedure described

Parameter	Fit Value	Stat.	Basic Syst.	D/N Syst.	MC Effects	Total Syst.
^{8}B Scale	0.924	$^{+0.028}_{-0.028}$	$^{+0.019}_{-0.021}$	$^{+0.002}_{-0.002}$	$^{+0.002}_{-0.005}$	$^{+0.019}_{-0.022}$
P_{ee0}	0.317	$^{+0.016}_{-0.016}$	$^{+0.008}_{-0.010}$	$^{+0.002}_{-0.002}$	$^{+0.002}_{-0.001}$	$^{+0.009}_{-0.009}$
P_{ee1}	0.0039	$^{+0.0065}_{-0.0067}$	$^{+0.0047}_{-0.0038}$	$^{+0.0012}_{-0.0018}$	$^{+0.0004}_{-0.0008}$	$^{+0.0045}_{-0.0045}$
P_{ee2}	-0.0010	$^{+0.0029}_{-0.0029}$	$^{+0.0013}_{-0.0016}$	$^{+0.0002}_{-0.0003}$	$^{+0.0004}_{-0.0002}$	$^{+0.0014}_{-0.0016}$
A_{ee0}	0.046	$^{+0.031}_{-0.031}$	$^{+0.007}_{-0.005}$	$^{+0.012}_{-0.012}$	$^{+0.002}_{-0.003}$	$^{+0.014}_{-0.013}$
A_{ee1}	-0.016	$^{+0.025}_{-0.025}$	$^{+0.003}_{-0.006}$	$^{+0.009}_{-0.009}$	$^{+0.002}_{-0.002}$	$^{+0.010}_{-0.011}$

Table 7.1: Fitted P_{ee} and ^{8}B parameters for the 3-phase P_{ee} fit.

	^{8}B Scale	P_{ee0}	P_{ee1}	P_{ee2}	A_{ee0}	A_{ee1}
^{8}B Scale	1.000	-0.723	0.302	-0.168	0.028	-0.012
P_{ee0}	-0.723	1.000	-0.299	-0.366	-0.376	0.129
P_{ee1}	0.302	-0.299	1.000	-0.206	0.219	-0.677
P_{ee2}	-0.168	-0.366	-0.206	1.000	0.008	-0.035
A_{ee0}	0.028	-0.376	0.219	0.008	1.000	-0.297
A_{ee1}	-0.012	0.129	-0.677	-0.035	-0.297	1.000

Table 7.2: Correlation matrix for the P_{ee} and ^{8}B parameters of the 3-phase P_{ee} fit.

in Section 5.4.4, which correctly combines asymmetrical uncertainties, and cannot be deduced analytically.

For this fit, parameter uncertainties are all dominated by statistical uncertainties, although the effects from the total systematic uncertainties are only smaller than statistical effects by up to a factor of approximately two. Table 7.2 shows important correlations between the different fit parameters. These correlations are expected, since the parameterisation is such that the numbers of CC and ES events in the fit are computed using combinations of the parameters. In particular, the strong anti-correlation between the ^{8}B scale and P_{ee0} is due to the fact that the number of CC events is essentially determined by the product between the ^{8}B scale and P_{ee0}. Since CC and NC signals constitute major components of the SNO data and the ^{8}B scale is mostly determined by the number of

NC events, the scale of P_{ee0} is, to first order, therefore proportional to the number of CC events divided by the ^{8}B scale.

The fitted ^{8}B flux, which is $(5.25 \pm 0.20) \times 10^6$ cm^{-2}s^{-1} after symmetrising the uncertainties and adding them in quadrature, is in good agreement with the BS05(OP) prediction of $(5.69 \pm 0.91) \times 10^6$ cm^{-2}s^{-1} [18]. It also agrees with the more recent BPS09(GS) and BPS09(AGSS09) SSM predictions of $(5.88 \pm 0.65) \times 10^6$ cm^{-2}s^{-1} and $(4.85 \pm 0.58) \times 10^6$ cm^{-2}s^{-1}, respectively [52]. Regarding the P_{ee} parameters, their values are very far from the no oscillation scenario, which corresponds to $P_{ee0} = 1$, $P_{ee1} = 0$, $P_{ee2} = 0$, $A_{ee0} = 0$ and $A_{ee1} = 0$. If the correlation matrix in Table 7.2 is inverted and is used to perform a chi-squared test where the alternate hypothesis is the no oscillation scenario, the resulting χ^2 value is 2420 for five degrees of freedom. This value corresponds to a p-value, the probability of an equal or larger discrepancy, smaller than 1×10^{-300}. The no oscillation hypothesis is thus very strongly rejected. However, the χ^2 value is probably not very accurate, since the tested hypothesis is so far from the central point that the Gaussian approximation for the parameter space might not hold properly. Another interesting test consists of verifying the possibility that there is no spectral distortion in the ν_e survival probability, corresponding to a scenario where only the P_{ee1} parameter is allowed to be different from zero. The χ^2 value for this test is 2.02 , which corresponds to a p-value of 73.2% for the four degrees of freedom. The no spectral distortion alternate hypothesis cannot thus be rejected at the 95% level. Finally, the hypothesis that there is no day/night distortion in the ν_e survival probability ($A_{ee0} = 0$ and $A_{ee1} = 0$) has an associated χ^2 value of 1.92 and thus a p-value of 38.3% with two degrees of freedom, such that it cannot be rejected either using the same confidence level.

If the results presented in Table 7.1 are compared to the published LETA results [4],

it is found that the total relative uncertainty on the fitted ^{8}B scale parameter, 3.77%, is not appreciably smaller than the relative uncertainty on the corresponding parameter for LETA, which was 3.84%. Although the three-phase analysis incorporates the pulse shape analysis and the PMT data from the NCD phase, both of which contribute to constraining the ^{8}B scale, different improvements to the LETA component of the fit tend to increase the uncertainty on the same parameter. In particular, all uncertainty parameters linked to the PMT β-γ events are now evaluated separately for day and night to account for the possible drift of the detector response over time, which can result in effective day/night asymmetries due to the difference between day and night live times. Also, the parameters for the production rates of the background events are not constrained to be positive in the three-phase analysis in order to minimise biases and also to reduce the distortion of the likelihood space, as described in Section 3.4. For the three-phase analysis, the effects of using finite Monte Carlo statistics to generate PDFs are also evaluated, but this was not done in the previous analyses. Finally, improvements to the implementation of existing systematic uncertainties and/or their constraints in the signal extraction procedure also contribute to slightly increase the size of the uncertainties. Since the uncertainty of the PSA constraint is 7.0%, the NCD phase provides a weak constraint on the neutron rate compared to the constraint on the NC signal from LETA. The resulting constraint provided by PSA for the ^{8}B scale is even weaker, due to the uncertainties on the neutron backgrounds and on the neutron capture efficiencies. The additional LETA background and systematic uncertainties mentioned above thus cancel the improvements to the ^{8}B scale uncertainties that would normally result from the inclusion of the data from the NCD phase. Although the uncertainties on the P_{ee} parameters partially depend on the ^{8}B scale uncertainty, due to the interplay between these parameters in the fit, the NCD

phase provides a significant number of CC and ES events which allow to constrain the P_{ee} parameters, this contribution being enhanced when the ^{8}B scale is constrained by all three phases. This results in a 20% improvement of the statistical uncertainty on P_{ee0} when compared to the LETA result. The total relative uncertainty on this parameter is 5.8% for the three-phase analysis, versus 6.1% for LETA.

Figure 7.1 shows the one-sigma uncertainty bands for the $P_{ee\text{day}}$ survival probability polynomial and for the day/night asymmetry function, along with the most likely curves. The curves were generated using the best fit parameter values from Table 7.1. The bands were obtained by symmetrising the uncertainties shown in Table 7.1, by adding the resulting statistical and systematic uncertainties in quadrature and then by generating 10^6 sets of random values for the P_{ee} and A_{ee} parameters using these total uncertainties along with the correlation matrix shown in Table 7.2. Correlated random variables can be generated using the Cholesky decomposition method [25]. For each set of random values, the corresponding $P_{ee}(E_\nu)$ and $A_{ee}(E_\nu)$ functions were computed for 1000 different E_ν values. Using the resulting distributions of random P_{ee} and A_{ee} values for each selected E_ν value, RMS values were computed on each side of the most likely P_{ee} and A_{ee} values separately. This produces uncertainty bands that can be asymmetrical, although it is not the case here because of the linear dependence of the P_{ee} and A_{ee} functions on the P_{ee} and A_{ee} parameters and the symmetrisation of the P_{ee} and A_{ee} parameter uncertainties. The resulting bands provide a more qualitative appreciation of the results from the P_{ee} fit that were described above. It allows to easily visualise how the no spectral distortion and the no day/night asymmetries scenarios are not excluded by the SNO data. Furthermore, it shows without any doubt that the SNO dataset is not compatible with $P_{ee} = 1$. Hence, it provides direct evidence of the observation of solar neutrino oscillation by SNO.
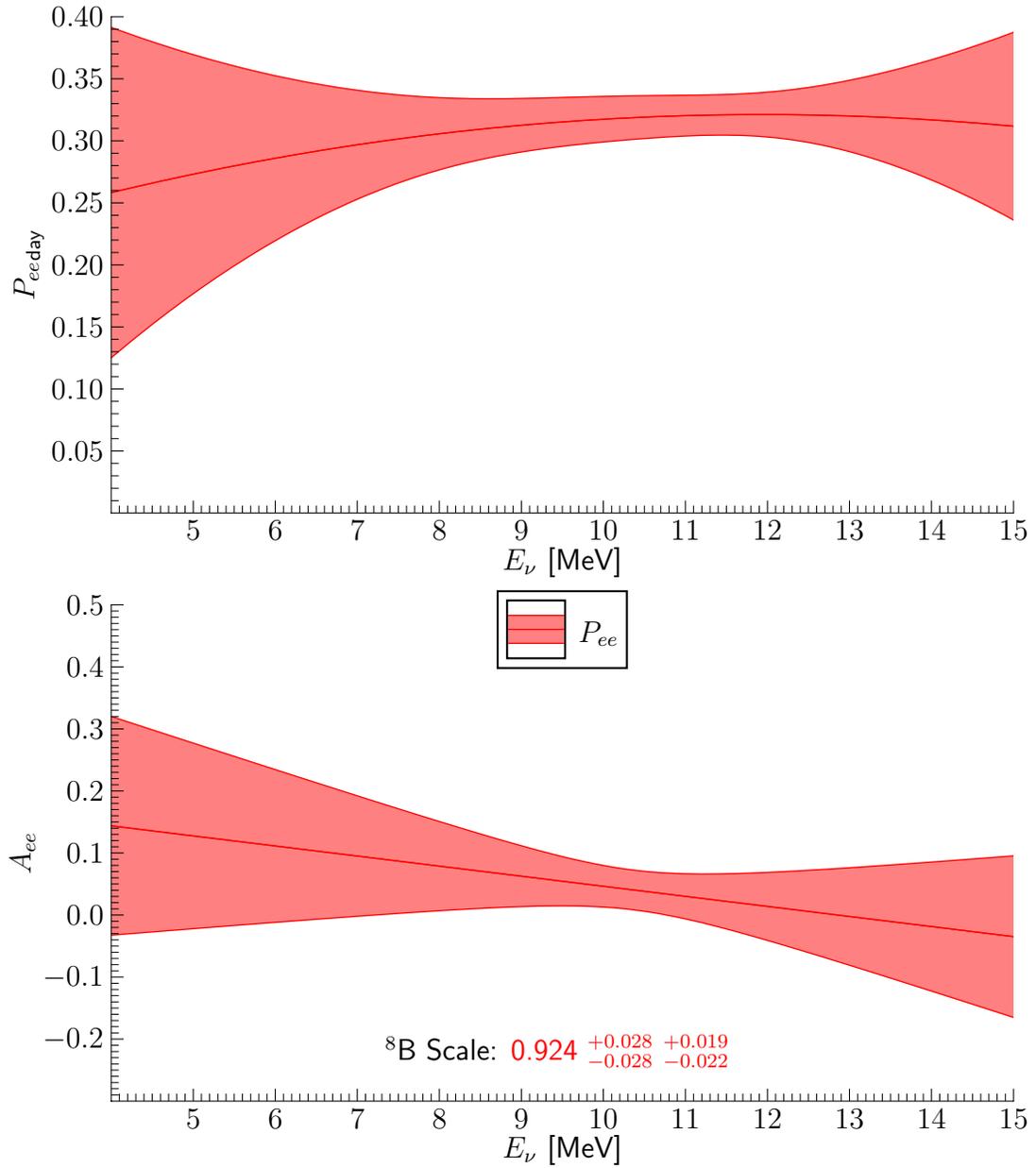
Figure 7.1: One-sigma (symmetrised) uncertainty bands of $P_{ee\mathrm{day}}(E_\nu)$ and $A_{ee}(E_\nu)$ for the three-phase P_{ee} fit, which take into account the correlations between the parameters.

Using the fitted parameter values, including all fitted background production rates and systematic uncertainty parameters, and the PDFs for all classes of events, reconstructed observable distributions have been generated in order to be compared to the SNO data. Figures 7.2 to 7.12 present the resulting day and night distributions, for all SNO phases and observables. The data points are shown, along with the numbers of reconstructed events and the breakdown of the reconstructed results for the different signals and the combined backgrounds. The discontinuities in the T_{eff} distributions are due to the variable bin width used for this observable. Also shown are the χ^2 values between the data and the reconstructed distributions. Note that the calculation of the number of degrees of freedom is not straightforward for these plots, since they present unidimensional projections of results which were obtained following a multidimensional fit. In this multidimensional space, there are 77 floated or scanned parameters that can directly or indirectly influence the fit between the reconstructed distributions and the data. However, there are 54 penalty terms in the likelihood function, each of which represent one additional degree of freedom. The net reduction in the number of degrees of freedom due to the fitted parameters is thus 23. Considering that the numbers of observable bins are $21 \times 8 \times 5 \times 15 = 12600$ for the D$_2$O and salt phases and $13 \times 25 \times 10 = 3250$ for the NCD phase, it corresponds to a fractional reduction of $23/(12600 + 3250) = 0.00145$ in the number of degrees of freedom constituted by the total number of observable bins. The number of bins in the unidimensional projections thus represents a very good approximation for the effective number of degrees of freedom. The χ^2 values for these projections all correspond to p-values greater than 5%, showing a good agreement between the data and the reconstructed distributions, except for the $\cos\theta_\odot$ day projection of the NCD phase in Figure 7.12. For this projection, the discrepancy appears to be due to a statistical
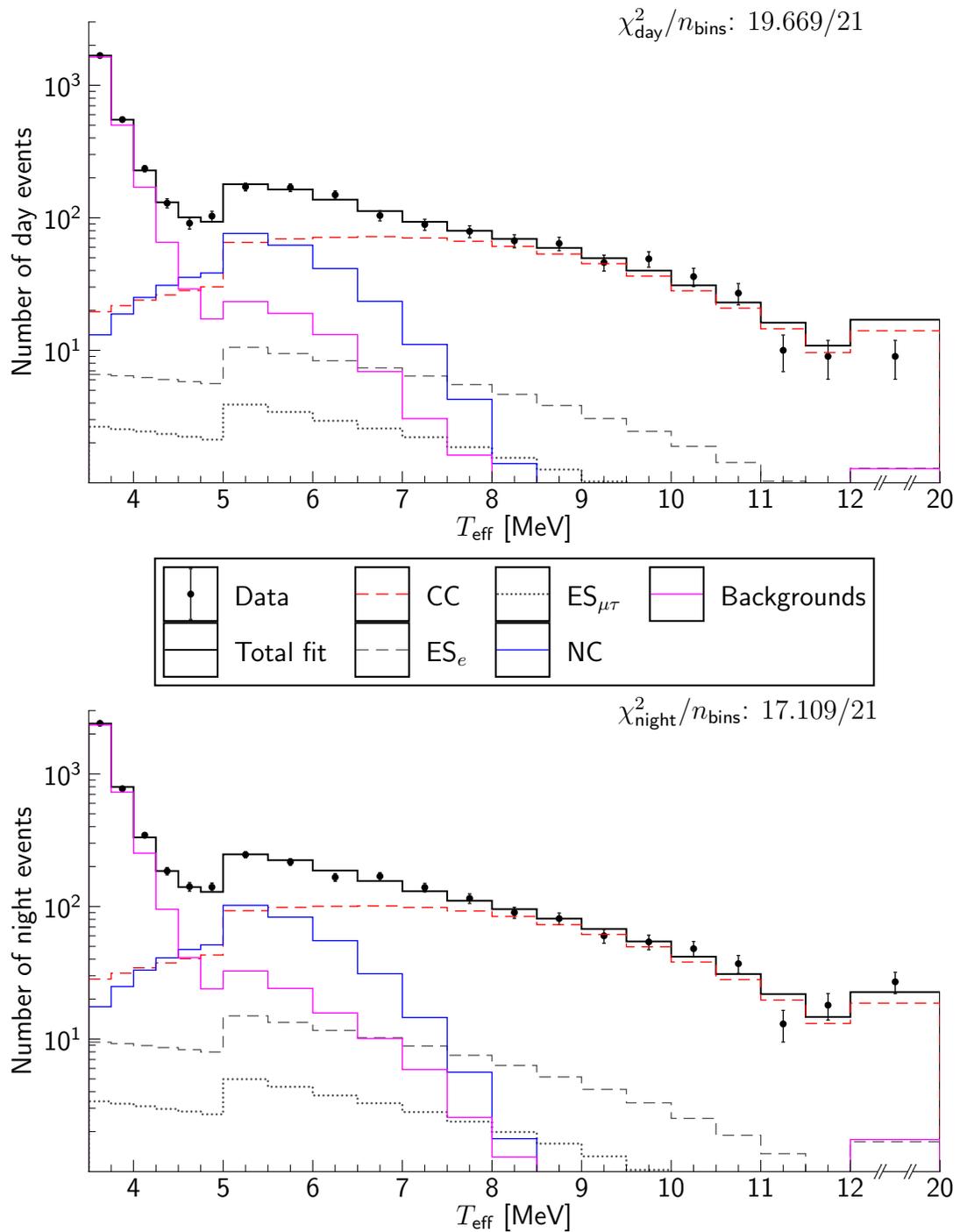
Figure 7.2: Projection of T_{eff} for the D$_2$O phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.
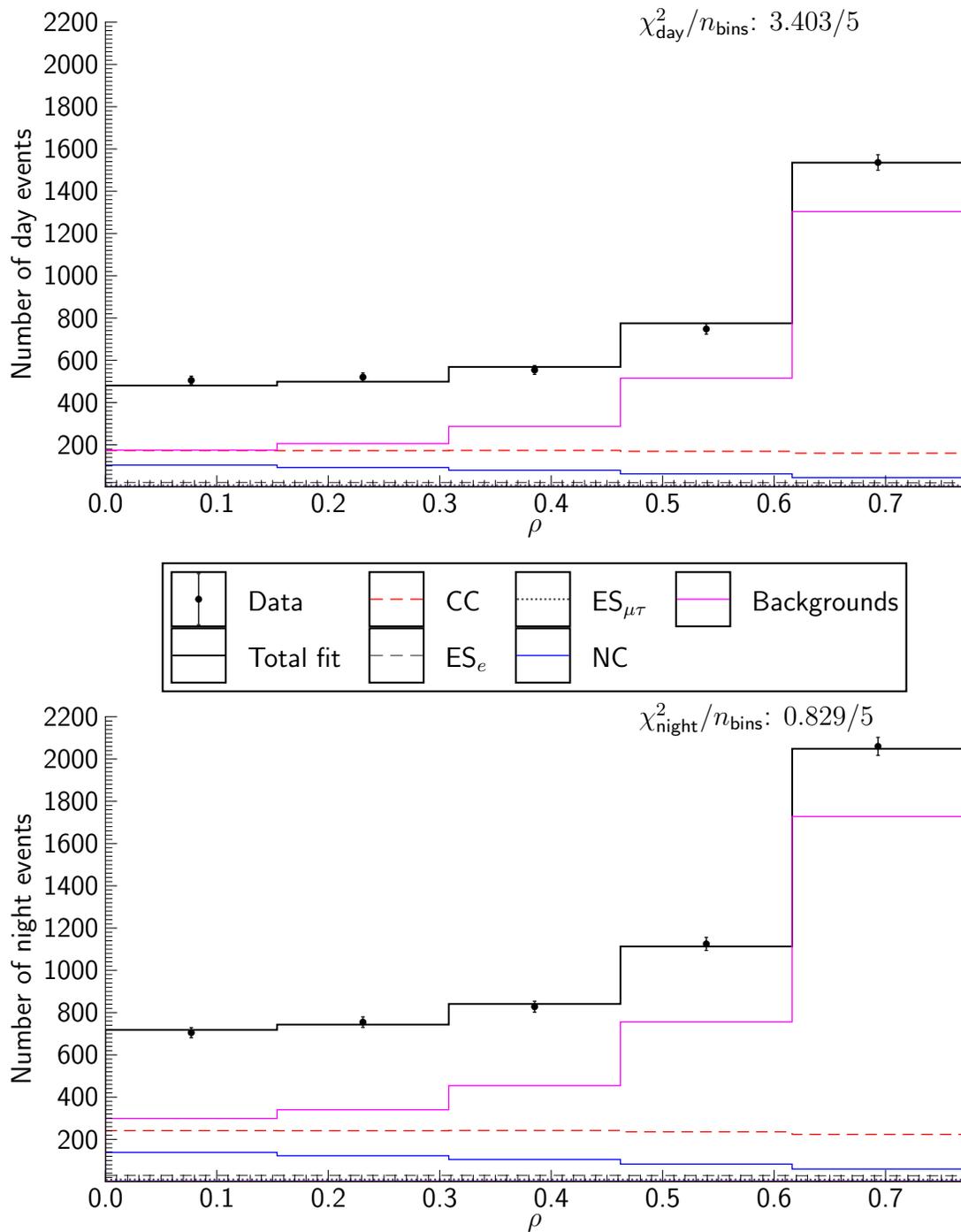
Figure 7.3: Projection of ρ for the D_2O phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.

Figure 7.4: Projection of $\cos\theta_{\odot}$ for the D$_2$O phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.
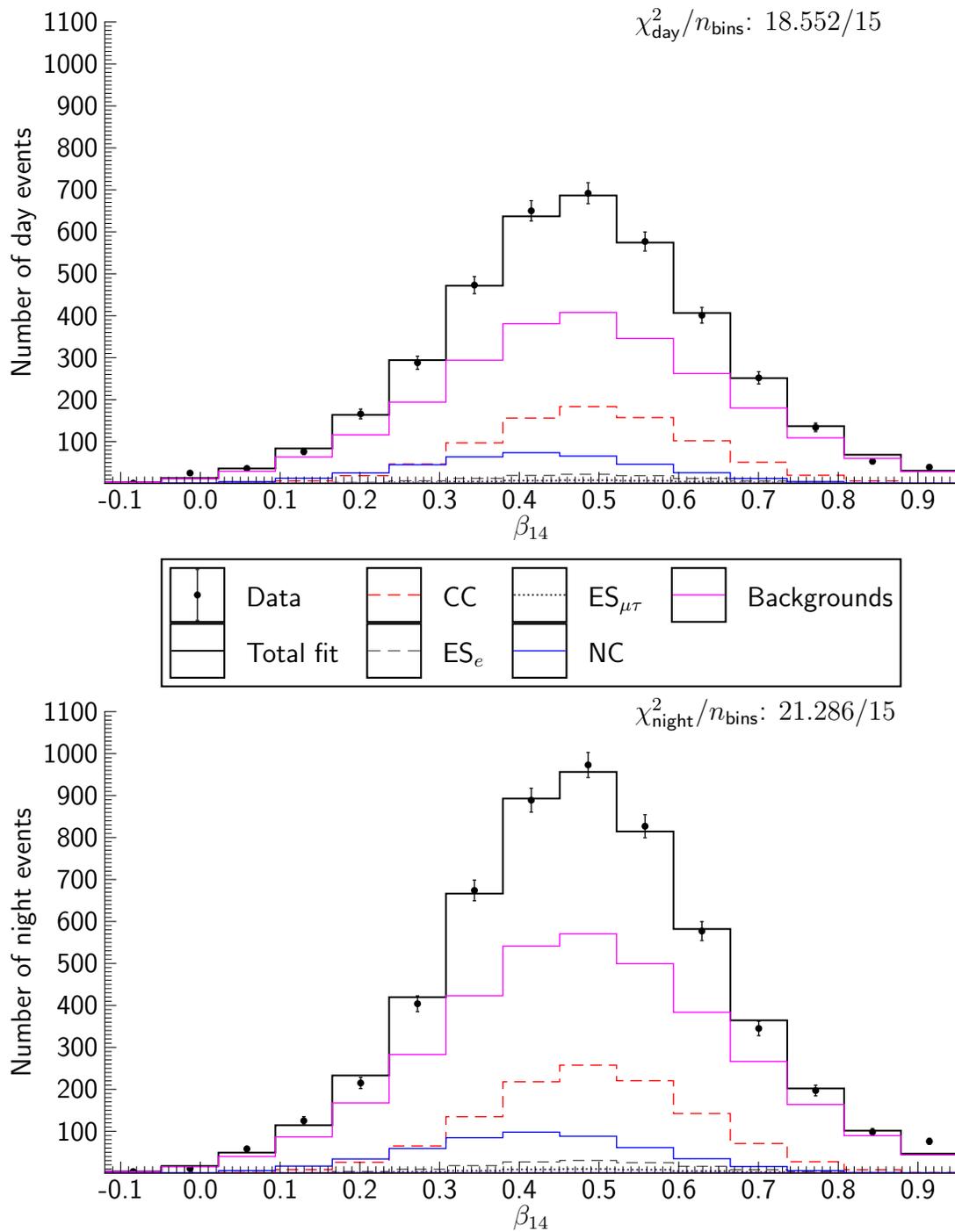
Figure 7.5: Projection of β_{14} for the D$_2$O phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.
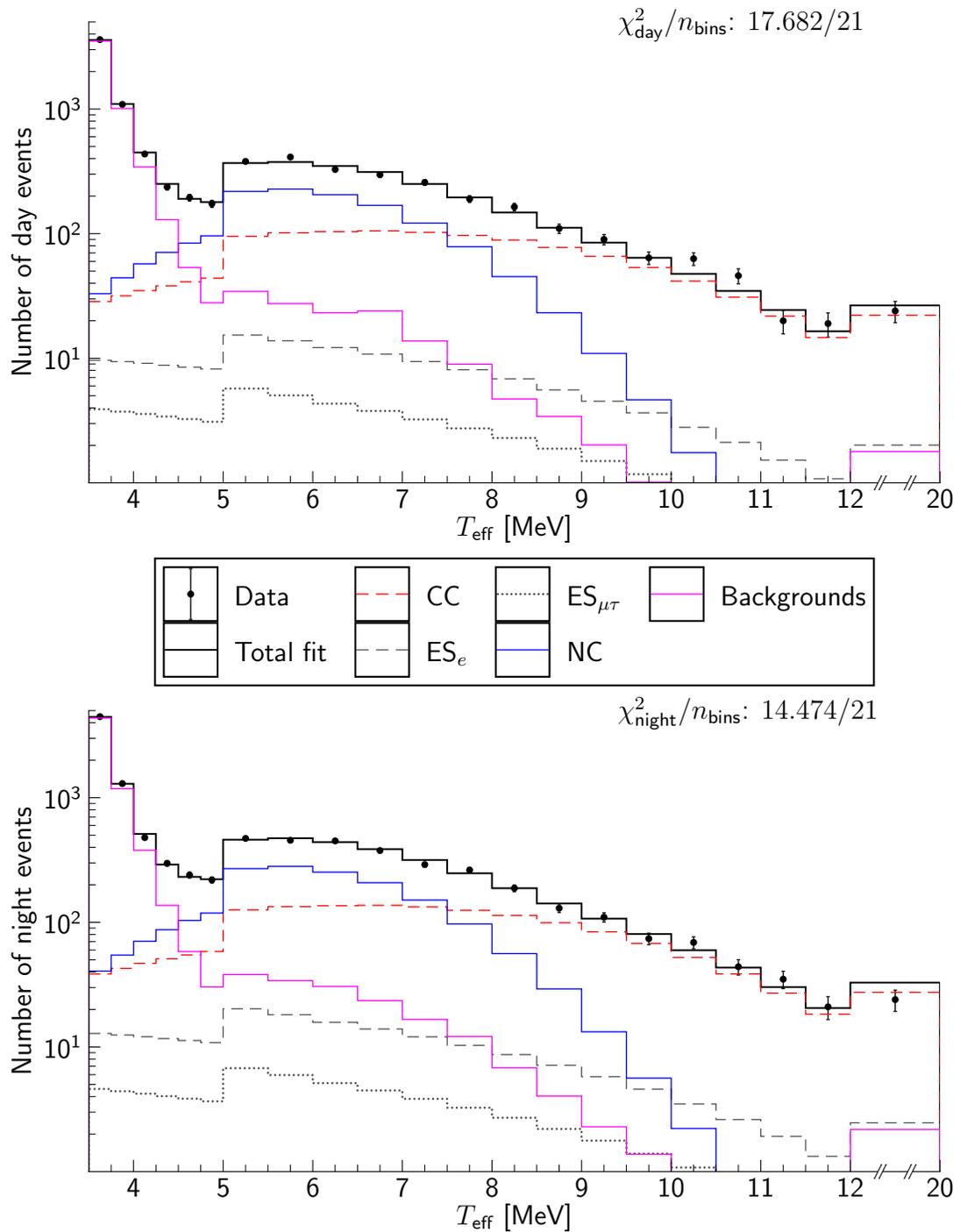
Figure 7.6: Projection of T_{eff} for the salt phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.
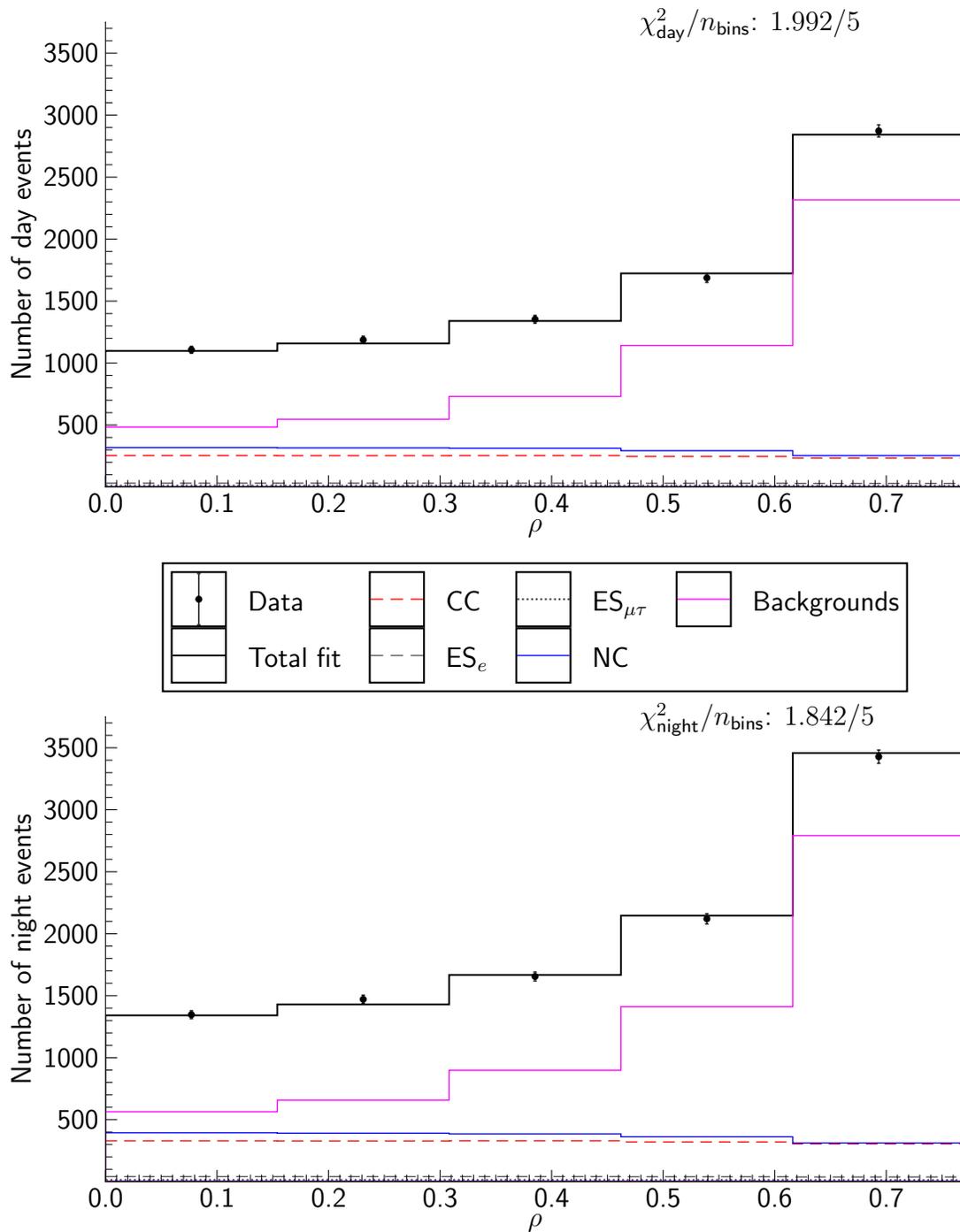
Figure 7.7: Projection of ρ for the salt phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.

Figure 7.8: Projection of $\cos\theta_\odot$ for the salt phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.
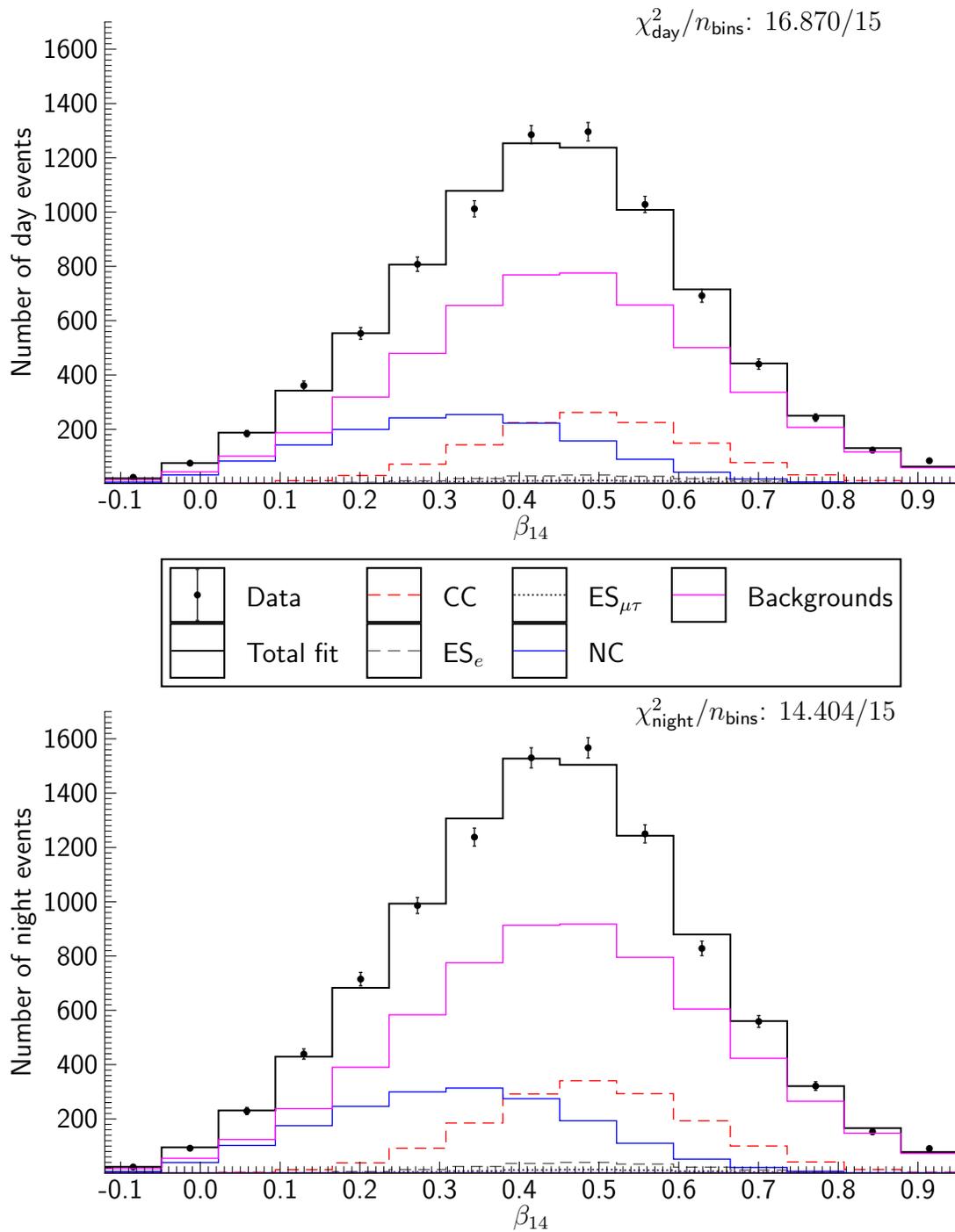
Figure 7.9: Projection of β_{14} for the salt phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.
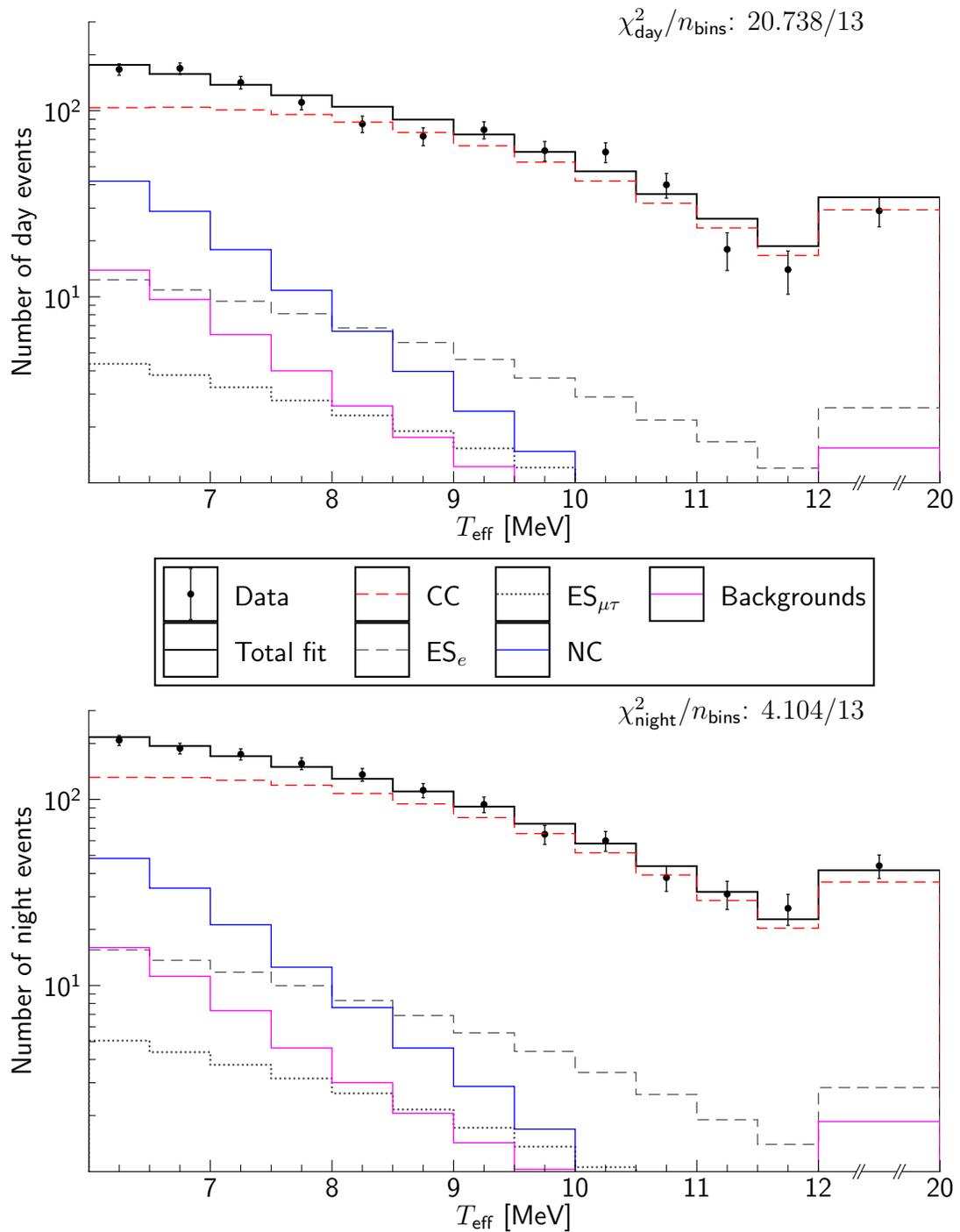
Figure 7.10: Projection of T_{eff} for the NCD phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.
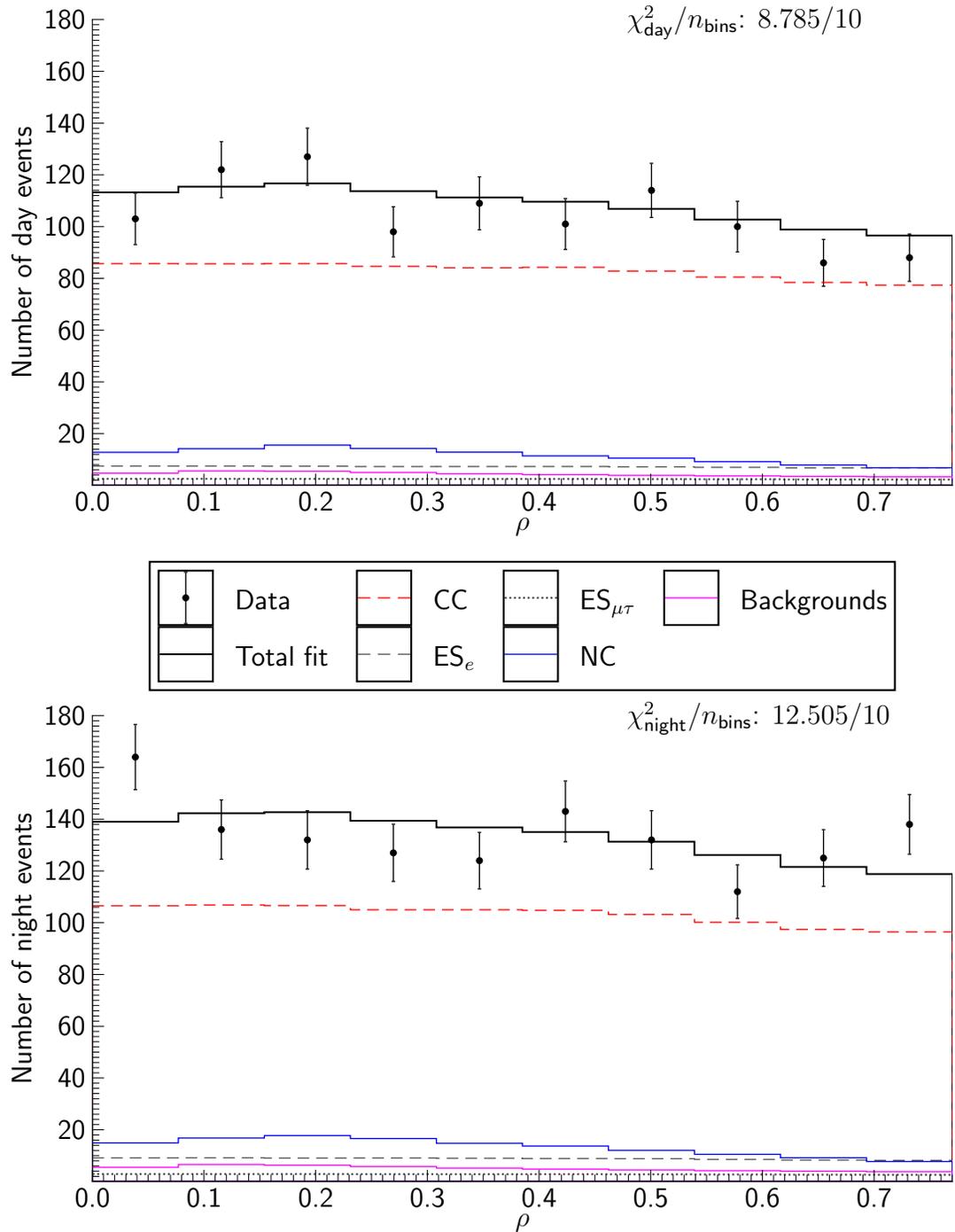
Figure 7.11: Projection of ρ for the NCD phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.

Figure 7.12: Projection of $\cos\theta_\odot$ for the NCD phase component of the P_{ee} fit. The top and bottom plots present the day and night events, respectively.

fluctuation in the data for two of the bins near $\cos\theta_\odot = 1$, as noted in [7].

7.2 $P_{ee} + P_{ea}$ Fit

As mentioned earlier, the usage of P_{ee} results is appropriate when analysing the SNO data in the context of a theoretical model where unitarity is assumed for the conversion of the electron-type solar neutrinos into one of the three weak flavour eigenstates. However, for theoretical models that consider oscillations to non-active neutrino flavours, i.e. for neutrinos that do not interact through the weak interaction, the P_{ee} method does not have the degrees of freedom that are required to allow the proper relative scaling of the expected number of CC, ES and NC events as a function of the different survival probability parameters, as well as the possible day/night asymmetry for the NC signal. The $P_{ee} + P_{ea}$ method introduces these additional degrees of freedom that are required for sterile neutrino analyses on the SNO data. As it can be seen through the equations of Section 4.4.1, for such a model, the ^{8}B flux cannot be measured by neutrino experiments such as SNO that only detect solar neutrinos through weak interactions. Such an experimental measurement could only be performed through the measurement of the solar neutrino rate via gravitational interactions, which is extremely difficult to realise. Since it is currently impossible to experimentally constrain the ^{8}B flux in the context of a sterile neutrino model, the $P_{ee} + P_{ea}$ method must thus rely on the theoretical value for the ^{8}B flux in order to provide measurements of the P_{ee} and P_{ea} parameters. However, as explained in Section 4.2.5, the extraction of the P_{ee} and P_{ea} parameters for SNO can be performed independently of the ^{8}B flux prediction, because the likelihood space only depends on the product between the ^{8}B scale and the P_{ee} and P_{ea} parameters, such that the best fit

parameters and their uncertainties for a given ^{8}B scale value can always be retrieved using

$$P'_{ee} = \frac{\alpha_B}{\alpha'_B} P_{ee} \qquad (4.46)$$

$$P'_{ea} = \frac{\alpha_B}{\alpha'_B} P_{ea}. \qquad (4.47)$$

This allows to combine the SNO $P_{ee} + P_{ea}$ results with other experimental results while using any chosen theoretical prediction for the ^{8}B flux.

A $P_{ee} + P_{ea}$ fit was thus performed on the SNO dataset, using a procedure identical to the one followed for the P_{ee} fit except for the change in the signal parameterisation. The ^{8}B scale was also fixed to a constant value, because of the aforementioned properties for the P_{ee} and P_{ea} parameters. It was arbitrarily set to the best value from the P_{ee} fit, only to facilitate the comparison of the results from both methods, before considering the effects from the uncertainty on the ^{8}B flux prediction, which has to be incorporated to provide a complete measurement in the case of the $P_{ee} + P_{ea}$ method. Tables 7.3 and 7.4 present the $P_{ee} + P_{ea}$ fit results, including the correlation matrix for the signal parameters. When compared to Table 7.1, Table 7.3 shows considerably smaller uncertainties for the P_{ee0} parameter, as anticipated when using a constant ^{8}B scale. The uncertainties for the A_{ee} parameters are considerably larger, on the other hand, since the NC signal rate is not constrained to be the same for day and night for a sterile model. The uncertainty on P_{ea0} is important, while still being bounded, because of the constraint provided by the scale of the NC signal in addition to the weak constraint from the number of ES$_{\mu\tau}$ events. The constraint provided through the NC signal scale is contingent on the accuracy of the P_{ee}

Parameter	Fit Value	Stat.	Basic Syst.	D/N Syst.	MC Effects	Total Syst.
^{8}B Scale	0.9235	N/A	N/A	N/A	N/A	N/A
P_{ee0}	0.316	$^{+0.012}_{-0.012}$	$^{+0.006}_{-0.007}$	$^{+0.004}_{-0.004}$	$^{+0.001}_{-0.001}$	$^{+0.007}_{-0.008}$
P_{ee1}	0.0076	$^{+0.0072}_{-0.0073}$	$^{+0.0045}_{-0.0031}$	$^{+0.0019}_{-0.0026}$	$^{+0.0006}_{-0.0010}$	$^{+0.0046}_{-0.0043}$
P_{ee2}	-0.0019	$^{+0.0029}_{-0.0029}$	$^{+0.0015}_{-0.0016}$	$^{+0.0003}_{-0.0005}$	$^{+0.0004}_{-0.0002}$	$^{+0.0015}_{-0.0017}$
A_{ee0}	0.067	$^{+0.039}_{-0.039}$	$^{+0.006}_{-0.009}$	$^{+0.022}_{-0.021}$	$^{+0.003}_{-0.004}$	$^{+0.022}_{-0.025}$
A_{ee1}	-0.031	$^{+0.030}_{-0.030}$	$^{+0.007}_{-0.005}$	$^{+0.016}_{-0.014}$	$^{+0.004}_{-0.002}$	$^{+0.019}_{-0.015}$
P_{ea0}	0.640	$^{+0.101}_{-0.096}$	$^{+0.026}_{-0.033}$	$^{+0.021}_{-0.020}$	$^{+0.010}_{-0.012}$	$^{+0.033}_{-0.041}$
P_{ea1}	-0.17	$^{+0.18}_{-0.17}$	$^{+0.03}_{-0.04}$	$^{+0.01}_{-0.01}$	$^{+0.02}_{-0.01}$	$^{+0.04}_{-0.04}$
A_{ea0}	-0.12	$^{+0.10}_{-0.10}$	$^{+0.05}_{-0.03}$	$^{+0.05}_{-0.06}$	$^{+0.02}_{-0.02}$	$^{+0.08}_{-0.06}$

Table 7.3: Fitted P_{ee} and P_{ea} parameters for the 3-phase $P_{ee} + P_{ea}$ fit. The ^{8}B scale was arbitrarily fixed to the value from the P_{ee} fit, since it cannot be measured experimentally in this case.

measurement though, because the NC scale is given by

$$\alpha_{\mathsf{NC}} = \alpha_{\mathsf{B}} \int_{-\infty}^{\infty} \left[P_{ee}(E_\nu) + P_{ea}(E_\nu) \right] f(E_\nu | \mathsf{NC}) dE_\nu. \tag{4.73}$$

The A_{ea} parameters do not benefit from a very good constraint through the NC scale, because the uncertainties on A_{ee} parameters are significant. The $\mathsf{ES}_{\mu\tau}$ signal provides an even weaker constraint on these parameters, such that their uncertainties are enormous. Nonetheless, these parameters are important since they contribute to generating the day/night uncertainty on NC, and thus have indirect effects on the P_{ee} parameter uncertainties. As a result of this interplay between the parameters of the $P_{ee} + P_{ea}$ fit, the correlation matrix shown in Table 7.4 differs significantly from the corresponding matrix for the P_{ee} fit shown in Table 7.2.

Figures 7.13 and 7.14 present the one-sigma uncertainty bands for the $P_{ee\mathsf{day}}$ and $P_{ea\mathsf{day}}$ probability polynomials, respectively, along with the associated bands for the day/night

	P_{ee0}	P_{ee1}	P_{ee2}	A_{ee0}	A_{ee1}	P_{ea0}	P_{ea1}	A_{ea0}
P_{ee0}	1.000	−0.237	−0.644	−0.617	0.332	−0.337	−0.136	0.454
P_{ee1}	−0.237	1.000	−0.213	0.366	−0.699	0.001	−0.319	−0.323
P_{ee2}	−0.644	−0.213	1.000	0.053	−0.054	0.166	0.287	−0.127
A_{ee0}	−0.617	0.366	0.053	1.000	−0.523	0.281	−0.008	−0.715
A_{ee1}	0.332	−0.699	−0.054	−0.523	1.000	−0.211	0.020	0.573
P_{ea0}	−0.337	0.001	0.166	0.281	−0.211	1.000	0.849	−0.382
P_{ea1}	−0.136	−0.319	0.287	−0.008	0.020	0.849	1.000	−0.036
A_{ea0}	0.454	−0.323	−0.127	−0.715	0.573	−0.382	−0.036	1.000

Table 7.4: Correlation matrix for the P_{ee} and P_{ea} parameters of the 3-phase $P_{ee} + P_{ea}$ fit, when the ^{8}B scale is arbitrarily fixed to the value from the P_{ee} fit.

asymmetry functions and the most likely curves, when the ^{8}B scale is arbitrarily fixed to the value obtained from the P_{ee} fit. The corresponding curves from the P_{ee} fit are also shown for comparison. Figure 7.13 shows that the additional degrees of freedom in the $P_{ee} + P_{ea}$ fit tend to pull down the $P_{ee\text{day}}$ curve slightly at low energy. This change is not significant though, particularly when considering the artificially narrower uncertainties for the $P_{ee} + P_{ea}$ fit due to the fixed ^{8}B scale. The same figure allows to appreciate the widening of the A_{ee} band with the $P_{ee} + P_{ea}$ fit, due to the NC signal rate which is no longer constrained to be the same for day and night. Finally, Figure 7.15 presents the one-sigma uncertainty bands for $P_{es\text{day}}$ and $P_{es\text{night}}$, the probabilities that electron-type neutrinos have oscillated to a sterile neutrino flavour when reaching the detector, given by $1 - P_{ee} - P_{ea}$. For the chosen fixed ^{8}B scale value, these bands are consistent with a model involving only three types of neutrinos, that is to say $P_{es} = 0$, as shown in the figure.

Since results from a $P_{ee} + P_{ea}$ fit are only meaningful after including the SSM uncertainty for the ^{8}B flux, the next step is thus to generate $P_{ee} + P_{ea}$ results by transforming the previous parameter values and correlation matrix based on a given SSM. As discussed
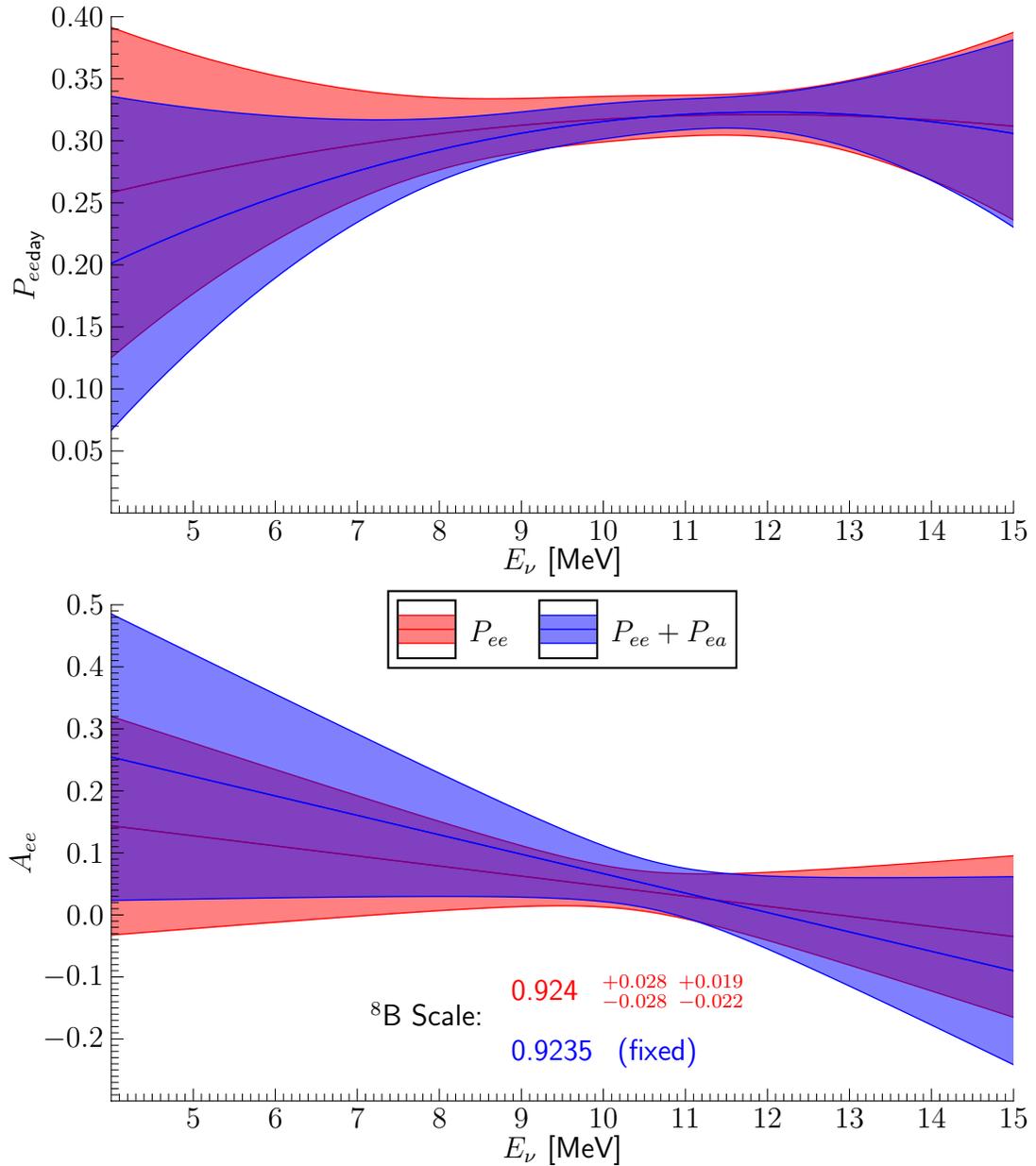
Figure 7.13: One-sigma (symmetrised) uncertainty bands of $P_{ee\text{day}}(E_\nu)$ and $A_{ee}(E_\nu)$ for the three-phase $P_{ee} + P_{ea}$ fit, which take into account the correlations between the parameters, but not the uncertainty on the ^{8}B flux whose value cannot be measured experimentally. The ^{8}B scale was arbitrarily fixed to the value from the P_{ee} fit. The bands from the P_{ee} fit are also shown.
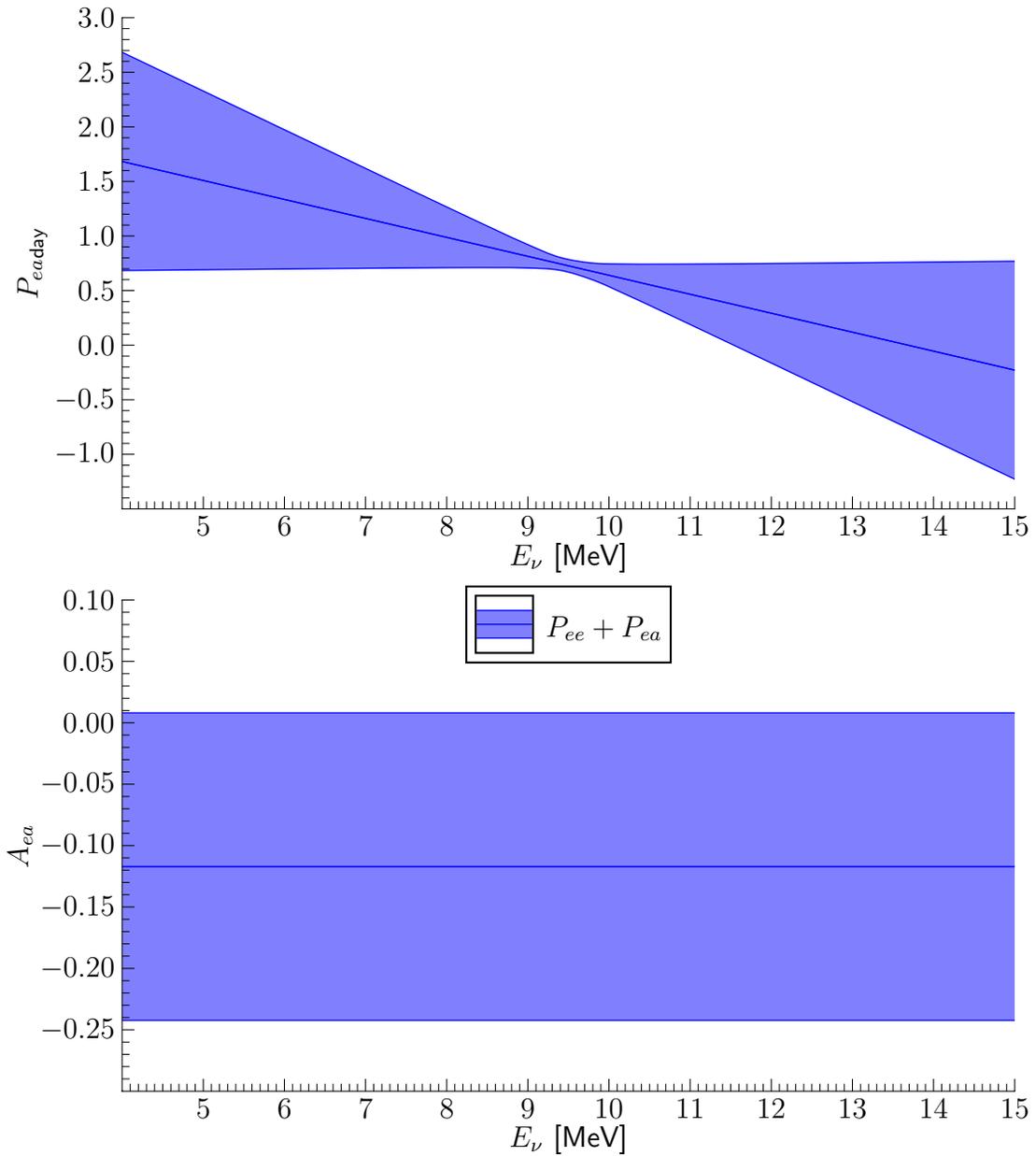
Figure 7.14: One-sigma (symmetrised) uncertainty bands of $P_{ea\,\mathrm{day}}(E_\nu)$ and $A_{ea}(E_\nu)$ for the three-phase $P_{ee} + P_{ea}$ fit, which take into account the correlations between the parameters, but not the uncertainty on the ^{8}B flux whose value cannot be measured experimentally. The ^{8}B scale was arbitrarily fixed to the value from the P_{ee} fit. Note that the error bands are smaller because there is no ^{8}B scale uncertainty.
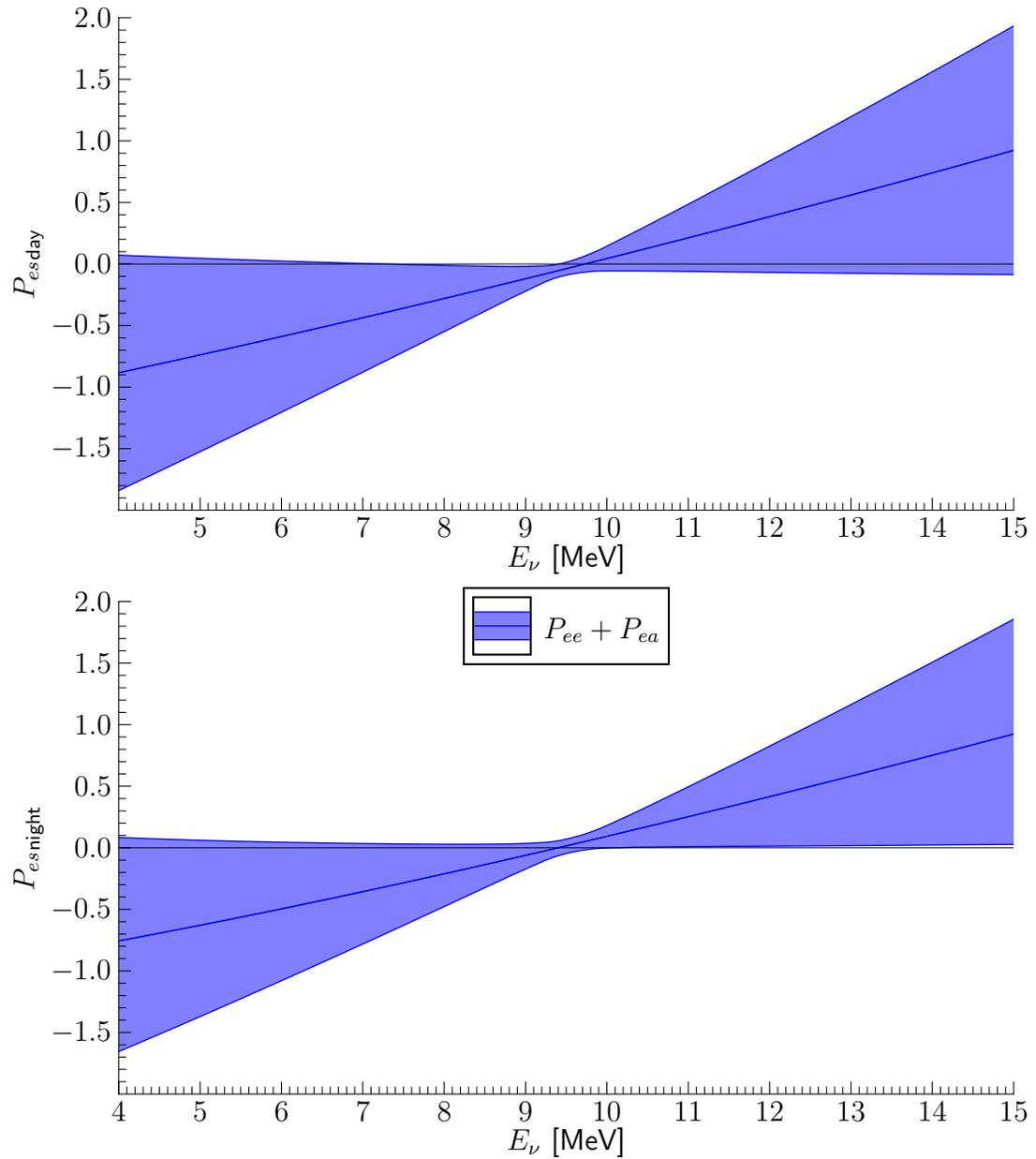
Figure 7.15: One-sigma (symmetrised) uncertainty bands of $P_{es\text{day}}(E_\nu)$ and $P_{es\text{night}}(E_\nu)$ for the three-phase $P_{ee} + P_{ea}$ fit, which take into account the correlations between the parameters, but not the uncertainty on the ^{8}B flux whose value cannot be measured experimentally. The ^{8}B scale was arbitrarily fixed to the value from the P_{ee} fit. Note that there is no ^{8}B scale uncertainty.

earlier, this transformation is exact and can be done easily. Table 7.5 shows the fitted SNO parameters for the $P_{ee} + P_{ea}$ fit, based on the BS05(OP) [18] model. The "Fit Value" column shows the most probable parameter values, which were simply obtained by scaling the P_{ee} and P_{ea} parameter values from Table 7.3 by 0.9235, as prescribed by Equations (4.46) and (4.47) with $\alpha'_{B} = 1$, since this value corresponds to the ^{8}B scale for the BS05(OP) model, as mentioned at the beginning of this chapter. The "Fixed" column presents the total parameter uncertainties, which were obtained after symmetrising the statistical and systematic uncertainties from Table 7.3 and after scaling the uncertainties for the P_{ee} and P_{ea} parameters by the same factor as for the best fit values. This column does not include the effects from the ^{8}B scale uncertainty however, but this can easily be done through a simple Monte Carlo simulation. To compute the uncertainties of the "Floated" column, as well as the correlation matrix shown in Table 7.6, 10^{6} sets of random P_{ee}, A_{ee}, P_{ea} and A_{ea} parameter values were generated using Tables 7.3 and 7.4, following the same procedure used to generate the one-sigma uncertainty bands of Figures 7.13 to 7.15 as described in the previous section. However, for each set of parameter values, a random ^{8}B scale value was generated using the constraint of the BS05(OP) model, which is 1.00 ± 0.16. The random P_{ee} and P_{ea} parameter values for a given set were then scaled using Equations (4.46) and (4.47), $\alpha_{B} = 0.9325$ and a value for α'_{B} given by the randomly generated ^{8}B scale value. The asymmetrical uncertainties of the "Floated" column in Table 7.5 were then generated from the resulting parameter distributions by computing RMS values on each side of the most likely parameter values separately. Similarly, the correlation matrix shown in Table 7.6 was generated by computing the covariances of the parameter distributions around the most likely parameter values. The fitted P_{ee} and P_{ea} parameters are inversely proportional to the ^{8}B scale of the BS05(OP) model,

Parameter	Fit Value	Fixed	Floated
^{8}B Scale	1.00	N/A	$+0.16$ -0.16
P_{ee0}	0.292	± 0.013	$+0.067$ -0.039
P_{ee1}	0.0070	± 0.0079	$+0.0086$ -0.0080
P_{ee2}	-0.0018	± 0.0031	$+0.0032$ -0.0033
A_{ee0}	0.067	± 0.045	$+0.045$ -0.045
A_{ee1}	-0.031	± 0.034	$+0.034$ -0.034
P_{ea0}	0.591	± 0.098	$+0.17$ -0.12
P_{ea1}	-0.16	± 0.17	$+0.17$ -0.18
A_{ea0}	-0.12	± 0.13	$+0.13$ -0.13
P_{es0}	0.117	± 0.094	$+0.14$ -0.23

Table 7.5: Fitted P_{ee}, P_{ea} and ^{8}B parameters for the 3-phase $P_{ee} + P_{ea}$ fit, along with the expression corresponding to a zeroth-order measurement of P_{es}. Shown are the parameter values and uncertainties when the ^{8}B flux is based on the BS05(OP) prediction. The first set of uncertainties was obtained after symmetrising the total uncertainties of a fit where the ^{8}B flux is fixed. The second set shows corresponding results after propagating the BS05(OP) flux uncertainties.

which has relatively large symmetrical uncertainties. Hence, the P_{ee} and P_{ea} parameter uncertainties become significantly larger and asymmetrical. Uncertainties for A_{ee} and A_{ea} parameters are unchanged, because these parameters are unaffected by the transformation. The computation of the asymmetrical uncertainties on $1 - P_{ee0} - P_{ea0}$, an expression that corresponds to a zeroth-order measurement of P_{es}, gives a value of $0.12^{+0.14}_{-0.23}$, which is not inconsistent with a model that only includes active neutrino flavours. Table 7.6 shows how the ^{8}B scale in a $P_{ee} + P_{ea}$ fit is typically very strongly anti-correlated with the P_{ee0} and P_{ea0} parameters. Except for the three correlations between the A_{ee} and A_{ea} parameters, the correlation factors are all affected when including the uncertainty on the ^{8}B flux from the SSM.

Figures 7.16 to 7.18 present the one-sigma uncertainty bands for the $P_{ee} + P_{ea}$ results

	^{8}B Scale	P_{ee0}	P_{ee1}	P_{ee2}	A_{ee0}	A_{ee1}	P_{ea0}	P_{ea1}	A_{ea0}
^{8}B Scale	1.000	−0.926	−0.147	0.094	0.001	0.000	−0.695	0.157	0.001
P_{ee0}	−0.926	1.000	0.091	−0.254	−0.152	0.081	0.647	−0.193	0.111
P_{ee1}	−0.147	0.091	1.000	−0.224	0.357	−0.680	0.113	−0.335	−0.315
P_{ee2}	0.094	−0.254	−0.224	1.000	0.053	−0.056	0.042	0.299	−0.126
A_{ee0}	0.001	−0.152	0.357	0.053	1.000	−0.522	0.190	−0.006	−0.716
A_{ee1}	0.000	0.081	−0.680	−0.056	−0.522	1.000	−0.143	0.018	0.573
P_{ea0}	−0.695	0.647	0.113	0.042	0.190	−0.143	1.000	0.455	−0.258
P_{ea1}	0.157	−0.193	−0.335	0.299	−0.006	0.018	0.455	1.000	−0.036
A_{ea0}	0.001	0.111	−0.315	−0.126	−0.716	0.573	−0.258	−0.036	1.000

Table 7.6: Correlation matrix for the P_{ee}, P_{ea} and ^{8}B parameters of the 3-phase $P_{ee} + P_{ea}$ fit, when the ^{8}B flux is floated according to the BS05(OP) constraint.

based on the BS05(OP) model, along with the bands from the P_{ee} fit, when applicable. The $P_{ee\mathrm{day}}$ band in Figure 7.16 shows a good overlap between the P_{ee} and $P_{ee} + P_{ea}$ results. The best fit curve for $P_{ee} + P_{ea}$ is lower on average, due to the fact that the BS05(OP) model predicts a higher ^{8}B flux than the best fit value obtained by the P_{ee} fit. Figure 7.18 presents the P_{es} bands for day and night. It shows how the SNO results alone are consistent with a model that only includes three types of neutrinos, with no significant day/night asymmetries for P_{es}. The agreement between the reconstructed distributions and the SNO dataset is very similar to what was obtained for the P_{ee} fit in Figures 7.2 to 7.12.
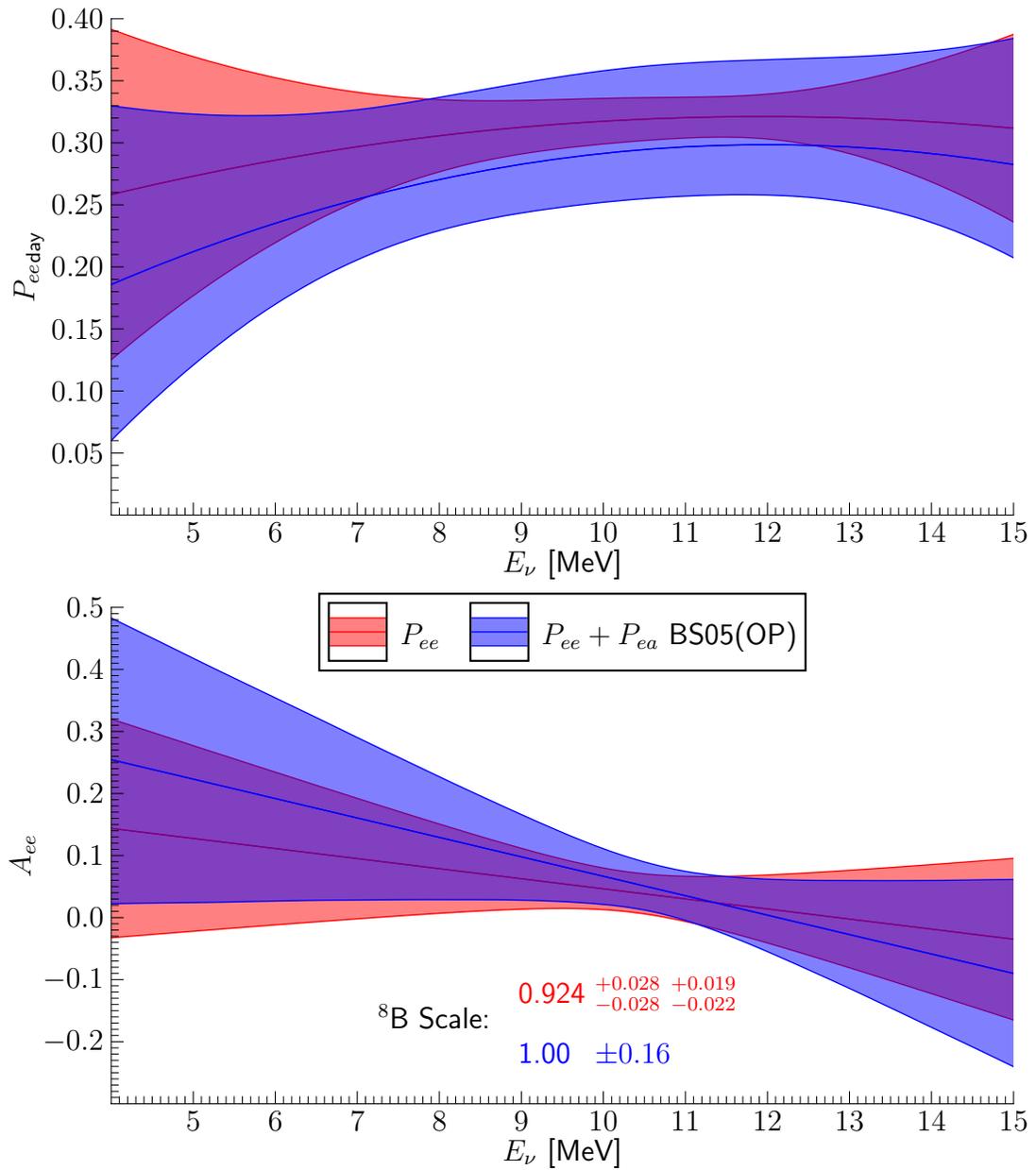
Figure 7.16: One-sigma (symmetrised) uncertainty bands of $P_{ee\mathsf{day}}(E_\nu)$ and $A_{ee}(E_\nu)$ for the three-phase $P_{ee} + P_{ea}$ fit, which take into account the correlations between the parameters. Results have been generated using the BS05(OP) ^{8}B flux and uncertainties. The bands from the P_{ee} fit are also shown.
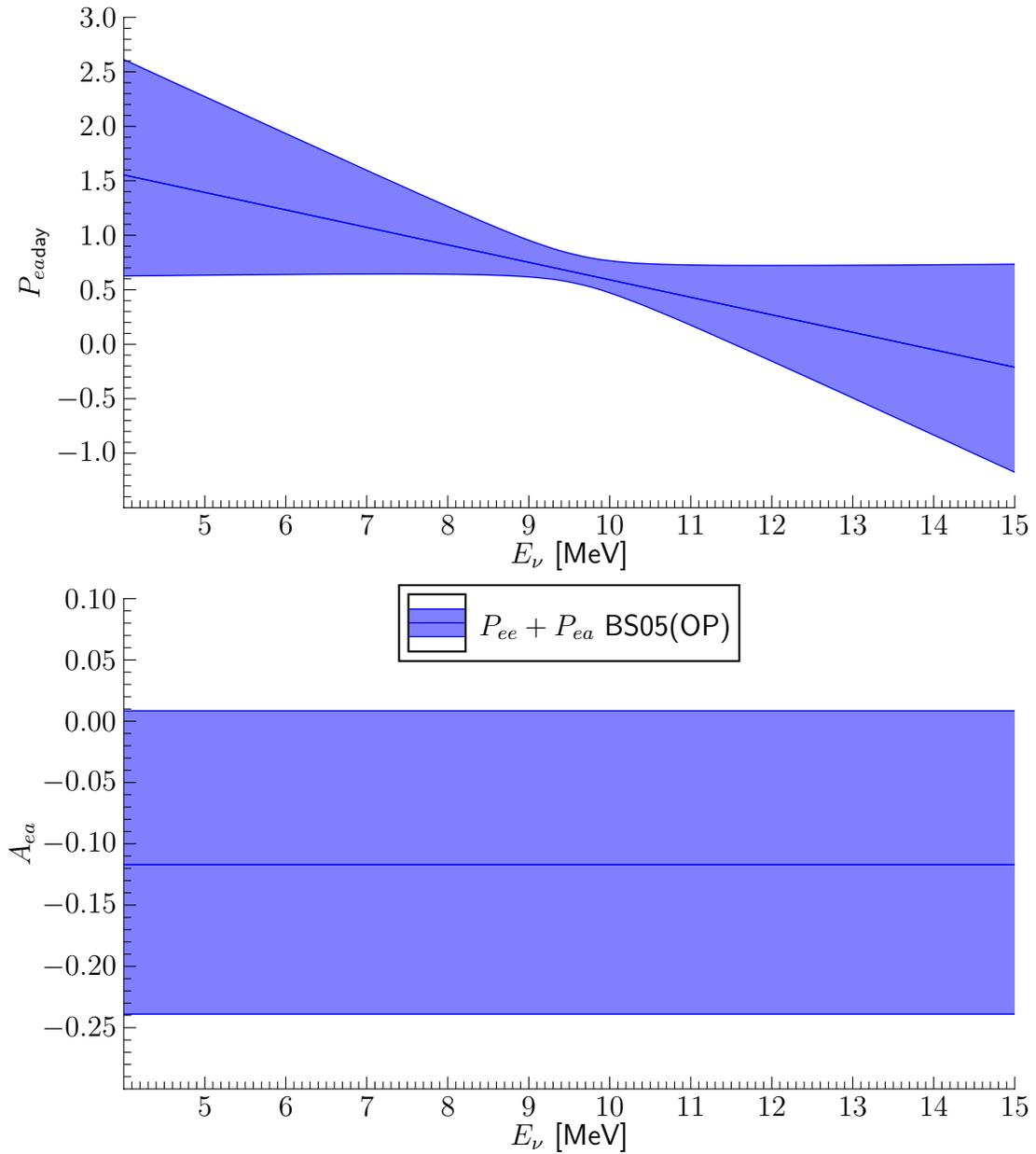
Figure 7.17: One-sigma (symmetrised) uncertainty bands of $P_{ea\,\text{day}}(E_\nu)$ and $A_{ea}(E_\nu)$ for the three-phase $P_{ee} + P_{ea}$ fit, which take into account the correlations between the parameters. Results have been generated using the BS05(OP) ^{8}B flux and uncertainties.
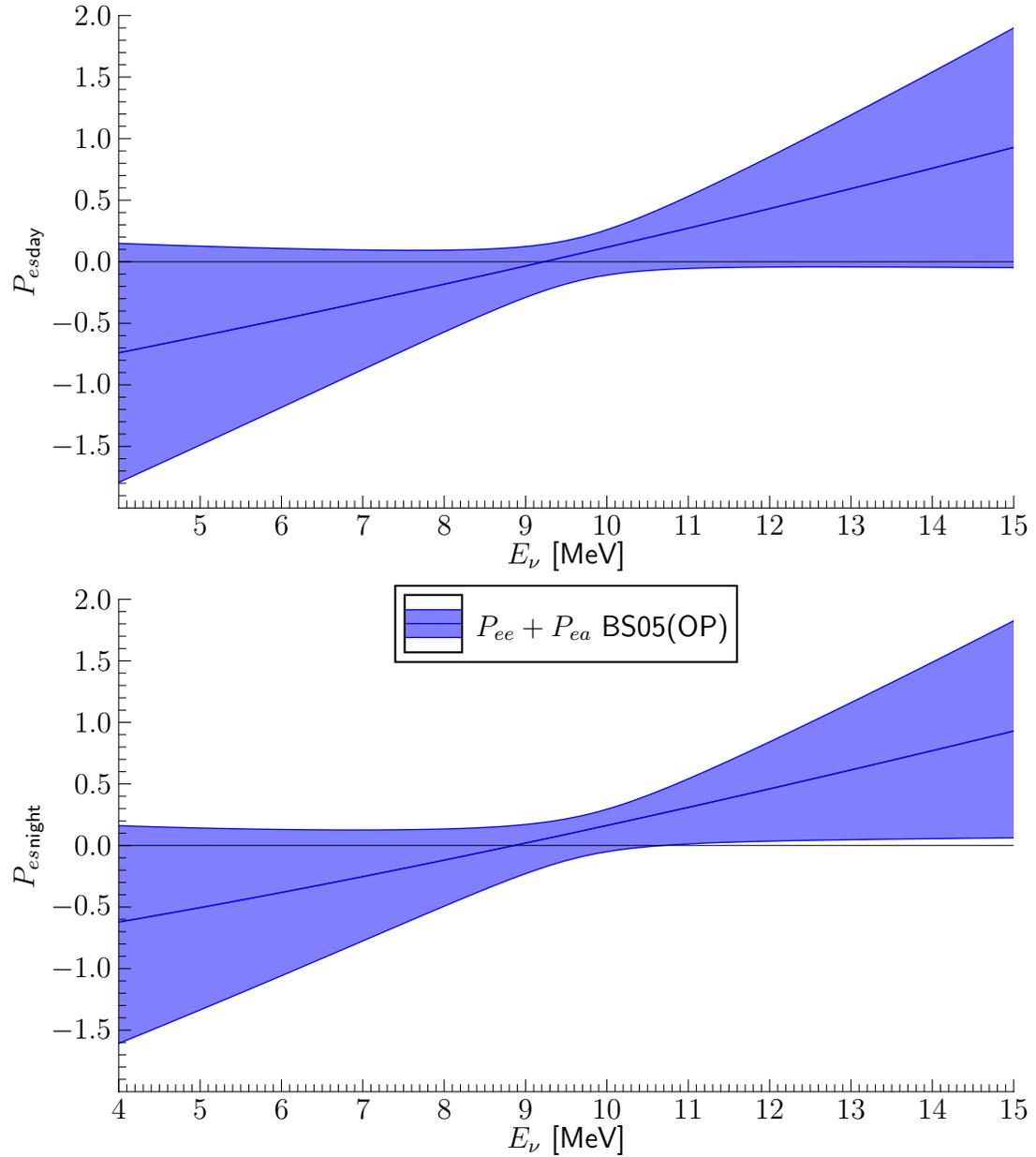
Figure 7.18: One-sigma (symmetrised) uncertainty bands of $P_{es\text{day}}(E_\nu)$ and $P_{es\text{night}}(E_\nu)$ for the three-phase $P_{ee} + P_{ea}$ fit, which take into account the correlations between the parameters. Results have been generated using the BS05(OP) ^{8}B flux and uncertainties.

Chapter 8

Conclusions

Designed to solve the solar neutrino problem, the SNO experiment has been very successful at measuring the total ^{8}B solar neutrino flux as well as the electron-type component of this flux. In its first published results in 2001 [26], SNO presented the first evidence for the oscillation of solar neutrinos. This result was obtained via the assumption that the CC energy spectrum was undistorted, since the D$_2$O phase only allowed to efficiently separate the CC and NC signals statistically using the reconstructed energy.

The data collected during the salt phase of the experiment allowed to better discriminate between the CC and NC signals, through the differentiation of their event isotropy. This allowed the SNO publication that analysed data from the salt phase [53] to successfully measure day and night CC energy spectra, which provided a model-independent measurement of the neutrino fluxes.

An analysis of the data from the NCD phase of the SNO experiment was completed in 2008 [7]. This phase provided an almost independent measurement of the ^{8}B flux, through the detection of neutrons from deployed proportional ^{3}He gas counters. Due to the breaking of the spherical geometry of the detector by the cylindrical NCDs, this phase necessitated the development of new pattern recognition algorithms for the detection of

light by the PMTs. The author of this thesis designed and implemented an efficient algorithm that allowed to recover approximately half of the loss in the reconstructed position uncertainty between the D_2O and NCD phases caused by the significant shadowing of the SNO PMTs by the NCDs.

In 2010, the collaboration published the results from a low-energy-threshold analysis of the D_2O and salt phases [4]. This analysis drastically reduced the measurement uncertainties due to the very significant increase of the signal statistics resulting from the lower energy threshold combined with careful studies of the SNO detector response, the backgrounds and systematic uncertainties. This publication also presented the first direct extraction of the survival probability function using SNO data. It was achieved using the binned P_{ee} fit method created by the author of this thesis, as described in Section 4.2.4.

In order to provide the best SNO measurement of the ^8B flux and the survival probability, it was then decided to perform a final SNO analysis which would consist of a three-phase extraction of these parameters. A pulse shape analysis of the data from the NCDs was used in order to reduce α and instrumental contamination of the neutron signal. To conduct such a combined analysis represented a considerable challenge, due to the number of floating systematic uncertainty parameters and the amount of Monte Carlo statistics involved. The author of this thesis thus designed QSigEx, a new signal extraction library which uses a variety of parallelisation, memory management and optimisation mechanisms to significantly reduce the time required to evaluate a likelihood function, while maintaining a very flexible structure that can support different types of analyses. In addition to these ameliorations at the programming level, significant work was performed to improve the efficiency and accuracy of the SNO likelihood function, as described in Chapters 3 and 4. Notably, an improved binned P_{ee} polynomial method was developed to

minimise biases in the extraction of the SNO signals, while being sufficiently efficient to allow the minimisation of the negative log-likelihood function in a reasonable amount of computer time.

A P_{ee} polynomial fit provides the best SNO constraint of the survival probability for electron-type neutrinos, for a theoretical model which does not allow non-active neutrino flavours, but it does not include the degrees of freedom which are necessary to correctly measure survival probabilities using other theoretical models, such as sterile neutrino models. In Chapter 4, $P_{ee} + P_{ea}$ methods are derived, which allow to introduce these necessary degrees of freedom to test these models.

For the three-phase analysis, improvements were also performed related to the Monte Carlo statistics. The significantly faster QSigEx library allowed to increase the sample size used to build the CC and ES_e PDFs by a factor of four and also to introduce the $ES_{\mu\tau}$ signal to the analysis, in addition to doubling the statistics for the simulation data used to generate some of the background PDFs. Although this increase in statistics contributed to reducing the signal extraction uncertainties due to statistical uncertainties of the PDFs, an algorithm was added to evaluate these uncertainties related to the finite Monte Carlo statistics. QSigEx also allowed to test the SNO signal extraction procedure for pulls and biases at an unprecedented level and to perform the final fits of the three-phase analysis using an unbinned P_{ee} polynomial method. The new signal extraction tool was also used to reproduce the results from the first analysis of the NCD phase data [7], as well as the results for the unconstrained and P_{ee} fits from LETA [4].

P_{ee} and $P_{ee} + P_{ea}$ combined fits of the three phases of the SNO experiment were then performed by the author of this thesis. As published in [39], the P_{ee} fit led to a

value of $5.25^{+0.16}_{-0.16}$(stat.)$^{+0.11}_{-0.13}$(syst.) $\times 10^6$ cm^{-2}s^{-1} for the total flux of neutrinos from ^{8}B decays in the Sun, assuming only active neutrino flavours and no constraint on the shape of the ^{8}B neutrino energy spectrum. The measurement of the P_{ee0} parameter resulted in a value of $0.317^{+0.016}_{-0.016}$(stat.)$^{+0.009}_{-0.009}$(syst.) that corresponds to direct evidence for flavour transformation of solar neutrinos. The precision of the P_{ee1} and P_{ee2} parameters does not allow SNO to reject the alternate hypothesis for the absence of spectral distortion. The fitted value for the A_{ee0} parameter is $0.046^{+0.031}_{-0.031}$(stat.)$^{+0.014}_{-0.014}$(syst.), which does not allow the rejection of a hypothesis that predicts no day/night asymmetry. The $P_{ee} + P_{ea}$ fit results, when applied to the BS05(OP) model, provide measurements of $0.292^{+0.067}_{-0.039}$ for P_{ee0} and $0.12^{+0.14}_{-0.23}$ for P_{es0} (defined as $1 - P_{ee0} - P_{ea0}$) that are not inconsistent with three active neutrino flavours.

The data of SNO show neither a large energy distortion nor a time-like asymmetry, but SNO clearly establishes in this analysis an undeniable signature of solar neutrino oscillations. The three-phase results can be seen as the legacy of the SNO experiment in the study of solar neutrinos.

Appendix A

Analytical PDF Event Generation

A.1 LETA PMT β-γ PDF/Events Generation

For LETA, an analytical multidimensional function is used to describe the PMT β-γ background events. The PDF is proportional to

$$f(T_{\text{eff}}, \beta_{14}, \rho, \cos\theta_\odot) \propto e^{\epsilon T_{\text{eff}}} \mathcal{N}(\beta_{14} \,|\, \bar{\beta}_{14} = \omega_0 + \omega_1(\rho - 0.61), \sigma = \beta_s)$$
$$(|b+1| - 1 + e^{\nu\rho}), \tag{A.1}$$

where $\mathcal{N}(x|\bar{x}, \sigma)$ is a Gaussian distribution with the given mean and width. To simplify the notation in the following discussion, B is defined as

$$B \equiv |b+1| - 1. \tag{A.2}$$

Since QSigEx uses binned PDFs for the other backgrounds and the signals, Equation (A.1) also has to be binned to avoid a biased fit (c.f. Section 3.3). Since the presence of the Gaussian function and the correlation with the event radius prevent the use of an analytical integration to generate this binned PDF, it is thus required to use a numerical method. One solution is to fill a histogram using values from random variables following

this distribution. It is also required to have such an event generation method to produce the fake datasets used for ensemble testing of the signal extraction code.

Luckily, this can be achieved quite efficiently using the transformation method. This method is based on the fact that given a random variable u with a known PDF $f_u(u)$, the probability of obtaining a value for a random variable $y = y(u)$ in the interval $[y(u), \, y(u) + dy(u)]$, $f_y(y(u))dy$, corresponds to the probability of u to be in the interval $[u, \, u + du]$, $f_u(u)du$. If the function $y(u)$ is chosen such that the minimal and maximal values for y and u (that is to say the minimal and maximal values of the random variables such that their respective PDFs are non-zero) are $y_{\text{min}} = y(u_{\text{min}})$ and $y_{\text{max}} = y(u_{\text{max}})$, respectively, the integration of the probabilities from u_{min} to some u value gives:

$$\int_{u_{\text{min}}}^{u} f_u(u')du' = \int_{y_{\text{min}}}^{y(u)} f_y(y')dy' = F_y(y(u)), \tag{A.3}$$

where $F_y(y)$ is the cumulative distribution for y. Solving for $y(u)$ gives

$$y(u) = F_y^{-1}\left(\int_{u_{\text{min}}}^{u} f_u(u')du'\right); \tag{A.4}$$

this simplifies to

$$y(u) = F_y^{-1}(u), \tag{A.5}$$

with y in the interval $]y_{\text{min}}, \, y_{\text{max}}]$, if u is a uniform deviate in the interval $]0, \, 1]$. Since histograms are usually defined such that a bin i has a range in an interval $[x_i, \, x_i + \Delta_x(i)[$, it is thus more convenient to generate y values in the interval $[y_{\text{min}}, \, y_{\text{max}}[$ using the transformation

$$y(u) = F_y^{-1}(1 - u), \tag{A.6}$$

with u being the same standard uniform deviate in the interval $]0,\ 1]$.

Using Equation (A.6) and the marginal PDF for the kinetic energy T that can be obtained from Equation (A.1), it is easy to show that proper random energies in the interval $[T_{\min},\ T_{\max}[$ can be generated using

$$T = \frac{1}{\epsilon} \log \left[(1 - u_0) \left(e^{\epsilon T_{\max}} - e^{\epsilon T_{\min}} \right) + e^{\epsilon T_{\min}} \right], \tag{A.7}$$

where u_0 is a standard uniform deviate.

Since the distribution for β_{14} in Equation (A.1) depends on ρ, it is easier to generate random values for the latter variable first and then use these values in the generation of random values for β_{14} (using a Gaussian random number generator). Although the expression for the distribution in ρ, $B + e^{\nu\rho}$, is very simple, random values for ρ cannot be generated directly using the transformation method because the function F_ρ cannot be inverted analytically. However, for a positive value of B, one can notice that the PDF for ρ can be written as

$$f(\rho) = P(\text{flat}) f(\rho|\text{flat}) + P(\text{exp}) f(\rho|\text{exp}), \tag{A.8}$$

where $P(\text{flat})$ is the probability for ρ to be located in the "flat" component, where $f(\rho|\text{flat})$ is the PDF for ρ given that it is located in this area, and where $P(\text{exp})$ and $f(\rho|\text{exp})$ are defined similarly for the exponential component. For the range $[\rho_{\min},\ \rho_{\max}[$ for ρ, it is straightforward to show, by integrating the areas of the two components, that the

probability $P(\text{flat})$ is given by

$$P(\text{flat}) = \frac{B(\rho_{\max} - \rho_{\min})}{B(\rho_{\max} - \rho_{\min}) + \frac{1}{\nu}\left(e^{\nu\rho_{\max}} - e^{\nu\rho_{\min}}\right)}. \tag{A.9}$$

For positive values of B, random ρ values can thus be generated using the following procedure:

1. Draw a pair of uniform deviates $(u_1, \ u_2)$.

2. If $u_1 \leq P(\text{flat})$, set ρ to

$$\rho = \rho_{\max} - u_2(\rho_{\max} - \rho_{\min}). \tag{A.10}$$

Otherwise, set ρ to

$$\rho = \frac{1}{\epsilon} \log\left[(1 - u_2)\left(e^{\epsilon\rho_{\max}} - e^{\epsilon\rho_{\min}}\right) + e^{\epsilon\rho_{\min}}\right]. \tag{A.11}$$

In the alternate case where $B < 0$, the technique described above does not work and F_ρ has to be inverted numerically. Since $F_\rho(\rho)$ is given by

$$F_\rho(\rho) = \frac{B\nu[\rho - \rho_{\min}] + e^{\nu\rho} - e^{\nu\rho_{\min}}}{B\nu(\rho_{\max} - \rho_{\min}) + e^{\nu\rho_{\max}} - e^{\nu\rho_{\min}}}, \tag{A.12}$$

solving for

$$F_\rho(\rho) = 1 - u, \tag{A.13}$$

where u is the value of a standard uniform deviate, is equivalent to finding the root of

$$G(\rho) \equiv B(\rho - \rho_{\text{min}}) + \frac{1}{\nu}\left(e^{\nu\rho} - e^{\nu\rho_{\text{min}}}\right)$$
$$+ (u - 1)\left[B(\rho_{\text{max}} - \rho_{\text{min}}) + \frac{1}{\nu}\left(e^{\nu\rho_{\text{max}}} - e^{\nu\rho_{\text{min}}}\right)\right]. \quad \text{(A.14)}$$

An efficient way to find this root is to use the Householder's method of order three, which leads to the following optimised algorithm:

1. Draw a pair of standard uniform deviates $(u_3,\ u_4)$.

2. Set

$$C_0 = e^{\nu\rho_{\text{min}}} \qquad\qquad\qquad\qquad\qquad\qquad \text{(A.15)}$$

$$C_1 = (u_3 - 1)\left[B(\rho_{\text{max}} - \rho_{\text{min}}) + \frac{1}{\nu}\left(e^{\nu\rho_{\text{max}}} - C_0\right)\right] \qquad \text{(A.16)}$$

$$\rho_0 = \rho_{\text{max}} - u_4(\rho_{\text{max}} - \rho_{\text{min}}). \qquad\qquad\qquad \text{(A.17)}$$

3. Compute

$$D_n = e^{\nu\rho_n} \qquad\qquad\qquad\qquad\qquad\qquad \text{(A.18)}$$

$$E_n = B + D_n \qquad\qquad\qquad\qquad\qquad \text{(A.19)}$$

$$G(\rho_n) = B(\rho_n - \rho_{\text{min}}) + \frac{1}{\nu}(D_n - C_0) + C_1 \qquad \text{(A.20)}$$

$$g(\rho_n) = \frac{E_n^2}{\nu D_n} \qquad\qquad\qquad\qquad\qquad \text{(A.21)}$$

$$\rho_{n+1} = \rho_n + \frac{3G(\rho_n)[2g(\rho_n) - G(\rho_n)]}{6E_n[G(\rho_n) - g(\rho_n)] - \nu G^2(\rho_n)}. \qquad \text{(A.22)}$$

4. Repeat Step 3 until the desired error on ρ, approximated by

$$\Delta \rho_n \approx \frac{G(\rho_n)}{E_n},\qquad\text{(A.23)}$$

is reached.

In the above procedure, the expression for $\Delta \rho$ is simply obtained by evaluating the first-order expansion of $G(\rho)$ at $\rho = \rho_\infty + \Delta \rho$, where ρ_∞ is the root of $G(\rho)$. For typical values of the PMT PDF parameters, approximately only 3.6 iterations on average are necessary to achieve an accuracy of 1×10^{-12}, which makes this algorithm very fast.

A.2 ^{16}N Source First Interaction Point

This section presents the details of a calculation that leads to an expression for the PDF of the ^{16}N interaction point along a Cartesian axis, under an isotropic point source approximation. Although the corresponding 3D PDF has a simple analytical expression, the projection on one of the axes cannot be expressed analytically. It is however possible to reduce it to a known integral form, for which numerical approximations already exist. The distribution can also be generated with a simple Monte Carlo simulation, using the transformation method.

From statistics, the distance between the centre of the ^{16}N source and the first interaction point of gamma rays follows an exponential distribution with a mean free path λ.

The marginal PDF $f(r)$ in spherical coordinates is thus given by

$$f(r) = \frac{1}{\lambda} e^{-\frac{r}{\lambda}}. \tag{A.24}$$

For an isotropic source, the random variables r, θ and ϕ are not correlated, such that the joint PDF $f(r, \theta, \phi)$ is given by

$$f(r, \theta, \phi) = f(r)f(\theta, \phi) = f(r)f(\theta)f(\phi). \tag{A.25}$$

The marginal PDF $f(\phi)$ is simply a uniform distribution defined between 0 and 2π:

$$f(\phi) = \frac{1}{2\pi}. \tag{A.26}$$

For an isotropic source in an isotropic medium, the probability of an event to occur in a given solid angle Ω must be proportional to that solid angle and must not depend on θ or ϕ, for any chosen solid angle. This can be translated to

$$\begin{aligned} P &= \iint_\Omega f(\theta, \phi) d\theta d\phi \\ &= \frac{1}{2\pi} \iint_\Omega f(\theta) d\theta d\phi \propto \Omega, \quad \forall \Omega. \end{aligned} \tag{A.27}$$

This must also hold at the infinitesimal level, where $d\Omega = \sin\theta d\theta d\phi$ and where $dP = \frac{1}{2\pi} f(\theta) d\theta d\phi$. This implies that $f(\theta) \propto \sin\theta$ and gives, after normalisation,

$$f(\theta) = \frac{1}{2} \sin\theta. \tag{A.28}$$

Therefore, from Equations (A.24), (A.25), (A.26) and (A.28), the joint PDF becomes

$$f(r, \theta, \phi) = \frac{1}{4\pi\lambda} e^{-\frac{r}{\lambda}} \sin\theta. \tag{A.29}$$

Since all marginal PDFs have been normalised, the expression for $f(r, \theta, \phi)$ is already normalised when integrating over the random variables:

$$\int_0^\infty \int_0^\pi \int_0^{2\pi} f(r, \theta, \phi) d\phi d\theta dr = \int_0^\infty f(r) dr \int_0^\pi f(\theta) d\theta \int_0^{2\pi} f(\phi) d\phi = 1. \tag{A.30}$$

The PDF $f(x, y, z)$ in Cartesian coordinates can be obtained from $f(r, \theta, \phi)$ using the Jacobian between the two coordinate systems:

$$
\begin{aligned}
f(x, y, z) &= \frac{f(r(x, y, z), \theta(x, y, z), \phi(x, y, z))}{|J(x, y, z)|} \\
&= \frac{f(r(x, y, z), \theta(x, y, z), \phi(x, y, z))}{r^2(x, y, z) \sin[\theta(x, y, z)]} \\
&= \frac{1}{4\pi\lambda} \frac{e^{-\frac{r(x,y,z)}{\lambda}}}{r^2(x, y, z)}.
\end{aligned}
\tag{A.31}
$$

The marginal PDF $f(x)$ is obtained by integrating $f(x, y, z)$ over y and z:

$$
\begin{aligned}
f(x) &= \int_{-\infty}^\infty \int_{-\infty}^\infty f(x, y, z) dy dz \\
&= \frac{1}{4\pi\lambda} \int_{-\infty}^\infty \int_{-\infty}^\infty \frac{e^{-\frac{\sqrt{x^2+y^2+z^2}}{\lambda}}}{x^2 + y^2 + z^2} dy dz.
\end{aligned}
\tag{A.32}
$$

There is no analytical solution to the integrals in Equation (A.32). However, the expression for $f(x)$ can be simplified to a known integral form. First, the integral is

expressed using spherical coordinates, by choosing the axes such that $x' = r \cos \theta$ and by using the property $\delta(au) = \frac{1}{|a|}\delta(u)$:

$$
\begin{aligned}
4\pi\lambda f(x) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{e^{\frac{-\sqrt{x^2+y^2+z^2}}{\lambda}}}{x^2 + y^2 + z^2} dydz \\
&= \int_{-\infty}^{+\infty} \delta(x' - x)dx \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{e^{\frac{-\sqrt{x^2+y^2+z^2}}{\lambda}}}{x^2 + y^2 + z^2} dydz \\
&= \int_{0}^{2\pi} \int_{-1}^{1} \int_{0}^{+\infty} \delta(r \cos\theta - x) \frac{e^{-r/\lambda}}{r^2} r^2 dr d\cos\theta d\phi \\
f(x) &= \frac{1}{2\lambda} \int_{-1}^{1} \int_{0}^{+\infty} \delta\left[\cos\theta \left(r - \frac{x}{\cos\theta}\right)\right] e^{-r/\lambda} dr d\cos\theta \\
&= \frac{1}{2\lambda} \int_{-1}^{1} \int_{0}^{+\infty} \frac{1}{|\cos\theta|} \delta\left(r - \frac{x}{\cos\theta}\right) e^{-r/\lambda} dr d\cos\theta \\
&= \frac{1}{2\lambda} \int_{-1}^{1} \frac{1}{|\cos\theta|} e^{\frac{-x}{\lambda\cos\theta}} d\cos\theta \begin{cases} 1 & \frac{x}{\cos\theta} \geq 0, \\ \\ 0 & \text{otherwise} \end{cases} \\
&= \frac{1}{2\lambda} \int_{0}^{1} \frac{1}{\cos\theta} e^{\frac{-|x|}{\lambda\cos\theta}} d\cos\theta.
\end{aligned}
\tag{A.33}
$$

Using the change of variable $\frac{|x|}{\lambda\cos\theta} \to t$, the last expression transforms to

$$
f(x) = -\frac{1}{2\lambda} \int_{+\infty}^{\frac{|x|}{\lambda}} \frac{e^{-t}}{|x|/(\lambda t)} \frac{|x|}{\lambda t^2} dt = \frac{1}{2\lambda} \int_{\frac{|x|}{\lambda}}^{+\infty} \frac{e^{-t}}{t} dt.
\tag{A.34}
$$

A random variable following the $f(x)$ distribution can also be generated using the transformation method [54] with Equations (A.24), (A.26) and (A.28). For a set of random variables u, u' and u'' uniformly distributed between 0 and 1, the transformation

method leads to

$$r = -\lambda \ln u$$

$$\phi = 2\pi u'$$

$$\cos \theta = 1 - 2u'', \tag{A.35}$$

where r, ϕ and θ are random variables distributed according to Equations (A.24), (A.26) and (A.28), respectively.

Since $x = r \sin \theta \cos \phi$, this gives

$$x = -2\lambda \ln u \sqrt{u''(1 - u'')} \cos(2\pi u'). \tag{A.36}$$

In this equation, x is a random variable that is distributed according to $f(x)$.

Appendix B

ES Direction Systematic Uncertainty

In Section 2.1.1.2.8, it was mentioned that the angular resolution function used to evaluate the SNO vertex reconstruction algorithms is given by the analytical expression

$$P(\cos\theta) = \alpha_\mathsf{M} \frac{\beta_\mathsf{M} e^{\beta_\mathsf{M}(\cos\theta - 1)}}{1 - e^{-2\beta_\mathsf{M}}} + (1 - \alpha_\mathsf{M}) \frac{\beta_\mathsf{S} e^{\beta_\mathsf{S}(\cos\theta - 1)}}{1 - e^{-2\beta_\mathsf{S}}}, \tag{2.9}$$

where θ is the angle between the true and the reconstructed event direction, β_S is the parameter for the exponential component associated to the main peak and β_M is the parameter associated to the multiple scattering component.

For ES events, the direction of electrons is strongly correlated with the incoming neutrino's direction, such that the $\cos\theta_\odot$ PDF itself has a shape that is well modelled by Equation (2.9). If the angular resolution systematic uncertainties are propagated in this PDF using the remapping equation

$$\cos\theta'_\odot = 1 + \frac{\cos\theta_\odot - 1}{1 \pm \Delta_\theta}, \tag{B.1}$$

where Δ_θ is the relative systematic uncertainty on both β_M and β_S (the relative systematic uncertainties on β_M and β_S have been found to be very similar and to have a strong,

positive correlation [55]), it transforms the PDF modelled by $P(\cos\theta)$ into

$$P'(\cos\theta'_{\odot}) = \alpha_{\mathsf{M}} \frac{\beta_{\mathsf{M}}(1 \pm \Delta_\theta)e^{\beta_{\mathsf{M}}(1\pm\Delta_\theta)(\cos\theta'_{\odot}-1)}}{1 - e^{-2\beta_{\mathsf{M}}}} +$$
$$(1 - \alpha_{\mathsf{M}})\frac{\beta_{\mathsf{S}}(1 \pm \Delta_\theta)e^{\beta_{\mathsf{S}}(1\pm\Delta_\theta)(\cos\theta'_{\odot}-1)}}{1 - e^{-2\beta_{\mathsf{S}}}}. \qquad (\text{B.2})$$

This follows from

$$g(y) = f(x(y))\left|\frac{dx}{dy}\right|, \qquad (\text{B.3})$$

which relates the PDF $g(y)$ to the PDF $f(x)$ through the usage of the Jacobian between x and y. From Equation (2.9), the true variation of the angular resolution function due to a relative systematic uncertainty Δ_θ on β_{M} and β_{S} is given by

$$P''(\cos\theta) = \alpha_{\mathsf{M}} \frac{\beta_{\mathsf{M}}(1 \pm \Delta_\theta)e^{\beta_{\mathsf{M}}(1\pm\Delta_\theta)(\cos\theta-1)}}{1 - e^{-2\beta_{\mathsf{M}}(1\pm\Delta_\theta)}} +$$
$$(1 - \alpha_{\mathsf{M}})\frac{\beta_{\mathsf{S}}(1 \pm \Delta_\theta)e^{\beta_{\mathsf{S}}(1\pm\Delta_\theta)(\cos\theta-1)}}{1 - e^{-2\beta_{\mathsf{S}}(1\pm\Delta_\theta)}}. \qquad (\text{B.4})$$

Comparing Equations (B.2) and (B.4), it can be seen that only the exponentials of the denominators are different. Equation (B.2) is normalised for $-2/(1\pm\Delta_\theta) < (\cos\theta'_{\odot}-1) < 0$ while Equation (B.4) is normalised for $-2 < (\cos\theta-1) < 0$. The integration of Equation (B.2) on the interval $-2 < (\cos\theta'_{\odot}-1) < 0$ gives

$$\int_{-1}^{1} P'(\cos\theta'_{\odot})d\cos\theta'_{\odot} = \alpha_{\mathsf{M}}\frac{1 - e^{-2\beta_{\mathsf{M}}(1\pm\Delta_\theta)}}{1 - e^{-2\beta_{\mathsf{M}}}} + (1 - \alpha_{\mathsf{M}})\frac{1 - e^{-2\beta_{\mathsf{S}}(1\pm\Delta_\theta)}}{1 - e^{-2\beta_{\mathsf{S}}}}. \qquad (\text{B.5})$$

Since β_{M} and β_{S} are typically greater than 3 and 15, respectively [55], the effect of Δ_θ in the previous equation is quite negligible, such that the relative weights of the exponential components are not significantly modified by the smearing.

The Equation (B.1) is thus a good approximation for the smearing of the ES $\cos\theta_\odot$ PDF due to the systematic uncertainties on β_{M} and β_{S}. Finally, at first-order approximation,

$$\cos\theta'_\odot = 1 + \frac{\cos\theta_\odot - 1}{1 \pm \Delta_\theta} \approx 1 + [1 \mp \Delta_\theta](\cos\theta_\odot - 1), \tag{B.6}$$

which is equivalent to Equation (2.11).

Appendix C

Configuration of Ensemble Test Fake Datasets

This appendix shows the configuration of the fake datasets and PDFs used for most ensemble tests. In each table, the second column gives the Poisson parameters used to generate the sets. These values exclude any P_{ee}/P_{ea} effect. The fourth column gives the number of "raw" PDF events, defined as the number of events in the skimmed Monte Carlo. The fifth column gives the number of "clean" PDF events, defined as the number of events in the analysis window. For the NCD phase, since the same events are used to generate NC, D$_2$O photodisintegration and atmospheric PDFs, the fake datasets have been generated by dividing about 50% of the events among the classes. For the NCD side, this also holds for NC, external neutrons, NCD, K2 and K5 photodisintegration backgrounds using the ^{24}Na source data.

	Fake Data		PDF Events	
Event Class	Poisson Param.	# Samples	# Raw	# Clean
D_2O ^8B CC	6257.910	992	4186597	3131937
D_2O ^8B ES_e	779.970	1985	1291739	779833
D_2O ^8B $ES_{\mu\tau}$ (P_{ee})	125.300	3000	3745027	2246251
D_2O ^8B $ES_{\mu\tau}$ (P_{ea})	125.300	3000	3744905	2245974
D_2O ^8B NC	908.850	250	268343	226689
D_2O ^8B NC $(P_{ee} + P_{ea})$	908.850	250	215538	182242
D_2O AV neutrons	41.490	2316	207629	95752
D_2O Bi D_2O	2145.700	64	2762655	100776
D_2O Tl D_2O	805.030	56	836317	44405
D_2O Bi AV bulk	42.800	19	284546	823
D_2O Tl AV bulk	66.600	15	257192	1027
D_2O Bi H_2O	85.070	14	759561	1238
D_2O Tl H_2O	106.500	16	556711	1713
D_2O hep CC	10.703	1000	762996	600770
D_2O hep ES	1.246	1000	156847	104702
D_2O hep NC	4.470	1000	5347	4504
D_2O PMT β-γ	2700.000	1000		
salt ^8B CC	8757.490	993	5839418	4377605
salt ^8B ES_e	1093.390	1986	1798262	1094917
salt ^8B $ES_{\mu\tau}$ (P_{ee})	175.800	3000	5221626	3151762
salt ^8B $ES_{\mu\tau}$ (P_{ea})	175.800	3000	5220337	3151427
salt ^8B NC	3369.410	250	1019554	840471
salt ^8B NC $(P_{ee} + P_{ea})$	3369.410	250	823874	679162
salt AV neutrons	150.000	2316	1104379	346879
salt Bi D_2O	3447.000	64	3801369	163874
salt Tl D_2O	1134.800	66	1150156	75708
salt ^{24}Na	390.300	347	1103113	134986
salt Bi AV bulk	81.100	18	377112	1511
salt Tl AV bulk	142.110	16	299586	2052
salt Bi H_2O	161.540	14	999577	2296
salt Tl H_2O	144.290	23	730405	3341
salt hep CC	15.061	1000	1074232	845508
salt hep ES	1.748	1000	218905	146946
salt hep NC	16.510	1000	19837	16305
salt PMT $\beta-\gamma$	5300.000	1000		

Table C.1: Poisson parameters in the LETA fake datasets and associated number of raw and clean PDF events for the different signals and backgrounds.

| Event Class | Fake Data | | PDF Events | |
	Poisson Param.	# Samples	# Raw	# Clean
D_2O ^8B CC	6257.910	992	4183691	3129633
D_2O ^8B ES_e	779.970	1985	1296830	782864
D_2O ^8B $ES_{\mu\tau}$ (P_{ee})	125.300	3000	3745684	2246160
D_2O ^8B $ES_{\mu\tau}$ (P_{ea})	125.300	3000	3745472	2245972
D_2O ^8B NC	908.850	250	268755	227201
D_2O ^8B NC ($P_{ee} + P_{ea}$)	908.850	250	217342	183607
D_2O AV neutrons	41.490	2316	208117	95930
D_2O Bi D_2O	2145.700	63	2840712	103541
D_2O Tl D_2O	805.030	55	851125	45276
D_2O Bi AV bulk	42.800	19	275737	763
D_2O Tl AV bulk	66.600	15	256566	1006
D_2O Bi H_2O	85.070	14	733985	1195
D_2O Tl H_2O	106.500	16	518561	1596
D_2O hep CC	10.703	1000	763213	601009
D_2O hep ES	1.246	1000	156858	104721
D_2O hep NC	4.470	1000	5322	4521
D_2O PMT β-γ	2700.000	1000		
salt ^8B CC	8757.490	992	5836635	4377206
salt ^8B ES_e	1093.390	1985	1798552	1095011
salt ^8B $ES_{\mu\tau}$ (P_{ee})	175.800	3000	5221045	3151373
salt ^8B $ES_{\mu\tau}$ (P_{ea})	175.800	3000	5224535	3153237
salt ^8B NC	3369.410	250	1020327	841558
salt ^8B NC ($P_{ee} + P_{ea}$)	3369.410	250	826103	681390
salt AV neutrons	150.000	2316	1106265	348260
salt Bi D_2O	3447.000	64	3805842	163924
salt Tl D_2O	1134.800	66	1145795	75772
salt ^{24}Na	390.300	347	1104300	135071
salt Bi AV bulk	81.100	18	360360	1478
salt Tl AV bulk	142.110	15	314336	2115
salt Bi H_2O	161.540	14	983038	2269
salt Tl H_2O	144.290	23	718818	3324
salt hep CC	15.061	1000	1074066	845386
salt hep ES	1.748	1000	218885	146941
salt hep NC	16.510	1000	20005	16497
salt PMT $\beta-\gamma$	5300.000	1000		

Table C.2: Poisson parameters in the LETA alternate fake datasets and associated number of raw and clean PDF events for the different signals and backgrounds.

	Fake Data		PDF Events	
Event Class	Poisson Param.	# Samples	# Raw	# Clean
NCD-PMT ^{8}B CC	5603.875	997	5816578	2816786
NCD-PMT ^{8}B ES$_e$	481.596	2001	1797864	485016
NCD-PMT ^{8}B ES$_{\mu\tau}$ (P_{ee})	74.833	3000	5252316	1353036
NCD-PMT ^{8}B ES$_{\mu\tau}$ (P_{ea})	74.833	3000	5255566	1354234
NCD-PMT ^{8}B NC	240.569	448	500125	122398
NCD-PMT ^{8}B NC ($P_{ee} + P_{ea}$)	237.636	370	500125	122398
NCD-PMT External n	20.754	2877	563920	59181
NCD-PMT Strings p.d.	5.938	15	157189	70092
NCD-PMT K2 p.d.	9.402	15	227498	112558
NCD-PMT K5 p.d.	8.378	15	288624	118326
NCD-PMT D$_2$O p.d.	8.305	448	500125	122398
NCD-PMT Atmos.	24.681	448	500125	122398
NCD-PMT hep CC	12.844	1000	1061931	724284
NCD-PMT hep ES	1.068	1000	218602	90581
NCD-PMT hep NC	1.156	1000	5417	1266
NCD-NCD ^{8}B NC	907.889	57	15632	14319
NCD-NCD External n	40.900	57	15632	14319
NCD-NCD Strings p.d.	35.600	57	15632	14319
NCD-NCD K2 p.d.	32.800	57	15632	14319
NCD-NCD K5 p.d.	31.600	57	15632	14319
NCD-NCD D$_2$O p.d.	31.000	57	15632	14319
NCD-NCD Atmos.	13.600	57	15632	14319
NCD-NCD hep NC	4.363	57	15632	14319

Table C.3: Poisson parameters in the NCD phase fake datasets and associated number of raw and clean PDF events for the different signals and backgrounds.

Event Class	Fake Data		PDF Events	
	Poisson Param.	# Samples	# Raw	# Clean
NCD-PMT ^{8}B CC	5603.875	997	5812480	2816837
NCD-PMT ^{8}B ES$_e$	481.596	2001	1792257	484284
NCD-PMT ^{8}B ES$_{\mu\tau}$ (P_{ee})	74.833	3000	5252635	1353535
NCD-PMT ^{8}B ES$_{\mu\tau}$ (P_{ea})	74.833	3000	5250438	1353066
NCD-PMT ^{8}B NC	240.569	448	500673	121940
NCD-PMT ^{8}B NC $(P_{ee} + P_{ea})$	237.622	370	500673	121940
NCD-PMT External n	20.754	2877	571079	60081
NCD-PMT Strings p.d.	5.938	15	157161	70075
NCD-PMT K2 p.d.	9.402	15	227545	112586
NCD-PMT K5 p.d.	8.378	15	288648	118340
NCD-PMT D$_2$O p.d.	8.305	448	500673	121940
NCD-PMT Atmos.	24.681	448	500673	121940
NCD-PMT hep CC	12.844	1000	1061587	724206
NCD-PMT hep ES	1.068	1000	218503	90540
NCD-PMT hep NC	1.156	1000	5173	1232
NCD-NCD ^{8}B NC	907.889	57	15496	14325
NCD-NCD External n	40.900	57	15496	14325
NCD-NCD Strings p.d.	35.600	57	15496	14325
NCD-NCD K2 p.d.	32.800	57	15496	14325
NCD-NCD K5 p.d.	31.600	57	15496	14325
NCD-NCD D$_2$O p.d.	31.000	57	15496	14325
NCD-NCD Atmos.	13.600	57	15496	14325
NCD-NCD hep NC	4.363	57	15496	14325

Table C.4: Poisson parameters in the NCD alternate phase fake datasets and associated number of raw and clean PDF events for the different signals and backgrounds.

Appendix D

Final Configuration of Three-Phase Fits

This appendix provides the numerical values along with some details concerning the configuration of the three-phase fits [39].

D.1 Nominal number of events

Tables D.1 to D.3 give the nominal number of events ("ν_{nom}", as described in Chapter 3) that are used in the final three-phase fits. For solar neutrino classes, the provided numbers are based on BS05(OP) [18] predictions and they are not modified by the survival probability functions. In the case of hep CC and hep ES, constant survival probabilities values of 0.35 and 0.47 are used, respectively.

Event Class	# Day Events	# Night Events
D_2O ^8B CC	2939.54	3901.26
D_2O ^8B ES_e	368.57	483.72
D_2O ^8B $ES_{\mu\tau}$	59.16	77.75
D_2O ^8B NC	415.61	553.49
D_2O AV neutrons	17.46	22.98
D_2O Bi D_2O	905.91	1183.65
D_2O Tl D_2O	339.42	444.58
D_2O Bi AV bulk	18.61	23.08
D_2O Tl AV bulk	30.58	41.49
D_2O Bi H_2O	85.13	107.53
D_2O Tl H_2O	44.75	58.97
D_2O Other n	1.33	1.77
D_2O hep CC	11.09	14.73
D_2O hep ES	0.97	1.26
D_2O hep NC	1.59	2.08
D_2O Atmos.	8.85	11.70
D_2O Leslie	13.32	10.27
D_2O PMT β-γ	1167.53	1532.47

Table D.1: Nominal number of events used by the final three-phase fit for the D_2O phase. The values correspond to the expected number of events in the analysis window. The numbers of events are not modified by survival probabilities.

Event Class	# Day Events	# Night Events
salt ^{8}B CC	4307.56	5302.95
salt ^{8}B ES$_e$	540.58	658.61
salt ^{8}B ES$_{\mu\tau}$	86.99	105.81
salt ^{8}B NC	1619.71	1996.45
salt AV neutrons	66.18	80.98
salt Bi D$_2$O	1522.03	1847.47
salt Tl D$_2$O	500.03	609.59
salt ^{24}Na	172.55	209.00
salt Bi AV bulk	35.07	44.23
salt Tl AV bulk	69.32	85.16
salt Bi H$_2$O	111.29	131.64
salt Tl H$_2$O	63.81	77.29
salt Other n	5.18	6.39
salt hep CC	16.35	20.12
salt hep ES	1.42	1.73
salt hep NC	6.10	7.56
salt Atmos.	13.14	15.79
salt Leslie	16.14	18.94
salt PMT $\beta-\gamma$	2390.29	2909.71

Table D.2: Nominal number of events used by the final three-phase fit for the salt phase. The values correspond to the expected number of events in the analysis window. The numbers of events are not modified by survival probabilities.

Event Class	# Day Events	# Night Events
NCD-PMT ^{8}B CC	2821.95	3375.53
NCD-PMT ^{8}B ES$_e$	247.05	285.56
NCD-PMT ^{8}B ES$_{\mu\tau}$	38.35	44.41
NCD-PMT ^{8}B NC	123.29	142.76
NCD-PMT External n	9.64	11.11
NCD-PMT Strings p.d.	2.72	3.22
NCD-PMT K2 p.d.	4.32	5.08
NCD-PMT K5 p.d.	5.55	6.51
NCD-PMT D$_2$O p.d.	3.85	4.46
NCD-PMT Atmos.	11.44	13.24
NCD-PMT hep CC	14.24	17.05
NCD-PMT hep ES	0.90	1.04
NCD-PMT hep NC	0.45	0.54

Event Class	# Events
NCD-NCD ^{8}B NC	1004.06
NCD-NCD External n	40.90
NCD-NCD Strings p.d.	35.60
NCD-NCD K2 p.d.	32.80
NCD-NCD K5 p.d.	45.50
NCD-NCD D$_2$O p.d.	31.00
NCD-NCD Atmos.	13.60
NCD-NCD hep NC	3.72

Table D.3: Nominal number of events used by the final three-phase fit for the NCD phase. The values correspond to the expected number of events in the analysis window. The numbers of events are not modified by survival probabilities.

Constrained Parameter	σ_{r+} [%]	σ_{r-} [%]
$\alpha_{D_2O \; Bi \; D_2O}$	33.33	20.00
$\alpha_{D_2O \; Tl \; D_2O}$	46.66	44.44
$\alpha_{D_2O \; Bi \; H_2O}$	17.30	17.30
$\alpha_{D_2O \; Tl \; H_2O}$	33.43	28.35
$\alpha_{salt \; Tl \; D_2O}$	46.05	59.21
$\alpha_{salt \; ^{24}Na}$	28.57	28.57
$\alpha_{salt \; Bi \; H_2O}$	24.22	24.22
$\alpha_{salt \; Tl \; H_2O}$	30.69	30.69

Table D.4: Relative constraints on day/night averaged background production rates used by the final three-phase fit for the D_2O and salt phases. The constraints are applied as described in Section 3.5.4, with the sum over the day and night components and $\mu_r = 1$. Day/night background rates are free to float separately only for D_2O, H_2O, salt ^{24}Na backgrounds and PMT β-γ events.

D.2 Background Constraints

In this section, constraints on the background production rates are presented. Although these constraints always affect the α parameters (see Equations (3.15) and (3.17) for the parameterisation of the generic form of the likelihood function) rather than the number of detected events, they are applied differently depending on the SNO phase. In the case of the D_2O and the salt phases, day and night rates are floated separately and the constraints are applied to the averaged day and night rates, as described in Section 3.5.4. For the NCD phase, the background rates are parameterised using day/night asymmetries, as presented in Section 3.7, so simple univariate penalty terms are used to constrain the resulting parameters. Table D.4 shows the constraints used for LETA and Tables D.5 and D.6 present the constraints used for the NCD phase.

Constrained Parameter	σ_{r+} [%]	σ_{r-} [%]
$\alpha_{\text{NCD-PMT External } n}$	50.37	50.37
$\alpha_{\text{NCD-PMT Strings p.d.}}$	34.27	34.27
$\alpha_{\text{NCD-PMT K2 p.d.}}$	15.85	15.85
$\alpha_{\text{NCD-PMT K5 p.d.}}$	16.48	18.46
$\alpha_{\text{NCD-PMT D}_2\text{O p.d.}}$	15.48	15.48
$\alpha_{\text{NCD-PMT Atmos.}}$	19.85	19.85

Table D.5: Relative constraints on day/night central background production rates used by the final three-phase fit for the NCD phase. The constraints are applied as simple univariate penalty terms (Section 3.5.1) with a most probable value of $\mu = 1$.

Constrained Parameter	μ	σ
$A_{\text{NCD-PMT External } n}$	-0.0195	0.0112
$A_{\text{NCD-PMT D}_2\text{O p.d.}}$	-0.034	0.112

Table D.6: Constraints on background day/night asymmetries used by the final three-phase fit for the NCD phase. The constraints are applied as simple univariate penalty terms (Section 3.5.1). Asymmetries for the other backgrounds of third phase are fixed to 0.

D.3   Systematic Uncertainties

This section presents the systematic uncertainty constraints that are used for the final fits. Since some of these uncertainties are correlated between phases, event classes or particle types, some uncertainties are expressed as a product between a fixed scale and a constrained variable. Variables that are represented using the symbol $\aleph$ are constrained using $N(0,1)$ Gaussian penalties.

D.3.1   Reconstructed Kinetic Energy

D.3.1.1   Scale

Energy scale systematic uncertainties are propagated as

$$
T_{\text{eff}'}^{(\text{D}_2\text{O,salt})} = \left\{ 1 + E_{\text{scale}}^{\text{corr}} + E_{\text{scale}}^{(\text{D}_2\text{O,salt})} \pm \frac{1}{2} \left[A_{\text{diu}} \left(E_{\text{scale}}^{(\text{D}_2\text{O,salt})} \right) + \right. \right.
$$
$$
\left. \left. \aleph_{E_{\text{scale}}}^{(\text{D}_2\text{O,salt})} A_{\text{dir}}^{(\text{CC,ES})} \left(E_{\text{scale}}^{(\text{D}_2\text{O,salt})} \right) \right] \right\} T_{\text{eff}}^{(\text{D}_2\text{O,salt})} \tag{D.1}
$$

$$
T_{\text{eff}'}^{\text{NCD}-\text{PMT}} = \left(1 + E_{\text{scale}}^{\text{corr}} + E_{\text{scale}}^{\text{NCD}-\text{PMT}} \right) \left\{ 1 \pm \frac{1}{2} \left[A_{\text{diu}} \left(E_{\text{scale}}^{\text{NCD}-\text{PMT}} \right) + \right. \right.
$$
$$
\left. \left. \aleph_{E_{\text{scale}}}^{\text{NCD}-\text{PMT}} A_{\text{dir}}^{\text{ES}} \left(E_{\text{scale}}^{\text{NCD}-\text{PMT}} \right) \right] \right\} T_{\text{eff}}^{\text{NCD}-\text{PMT}}, \tag{D.2}
$$

where $A_{\text{dir}} \left(E_{\text{scale}}^{(\text{D}_2\text{O,salt})} \right)$ only applies for the indicated event classes and where the positive shift is used for night events and the negative shift for day events.

Parameter	Nominal	σ_+	σ_-
$E_{\text{scale}}^{\text{corr}}$	0.0000	0.0041	0.0041
$E_{\text{scale}}^{\text{D}_2\text{O}}$	0.0000	0.0039	0.0047
$E_{\text{scale}}^{\text{salt}}$	0.0000	0.0034	0.0032
$E_{\text{scale}}^{\text{NCD}-\text{PMT}}$	0.0000	0.0081	0.0081
$A_{\text{diu}}\left(E_{\text{scale}}^{\text{D}_2\text{O}}\right)$	0.00%	0.32%	0.32%
$A_{\text{diu}}\left(E_{\text{scale}}^{\text{salt}}\right)$	0.0%	0.4%	0.4%
$A_{\text{diu}}\left(E_{\text{scale}}^{\text{NCD}-\text{PMT}}\right)$	0.00%	0.38%	0.38%
$A_{\text{dir}}^{\text{CC}}\left(E_{\text{scale}}^{\text{D}_2\text{O}}\right)$	0.09%		
$A_{\text{dir}}^{\text{ES}}\left(E_{\text{scale}}^{\text{D}_2\text{O}}\right)$	-0.92%		
$A_{\text{dir}}^{\text{CC}}\left(E_{\text{scale}}^{\text{salt}}\right)$	0.09%		
$A_{\text{dir}}^{\text{ES}}\left(E_{\text{scale}}^{\text{salt}}\right)$	-0.79%		
$A_{\text{dir}}^{\text{ES}}\left(E_{\text{scale}}^{\text{NCD}-\text{PMT}}\right)$	-0.99%		

Table D.7: Energy scale systematic uncertainty parameters.

D.3.1.2 Non-linearity

The energy non-linearity systematic uncertainty is applied as

$$T_{\text{eff}''}^{e^-} = T_{\text{eff}'}^{e^-}\left(1 + E_{\text{non}-\text{lin.}}\frac{T_{\text{eff}'}^{e^-} - 5.05\text{ MeV}}{19.0\text{ MeV} - 5.05\text{ MeV}}\right). \tag{D.3}$$

The same transformation is used for the three phases and does not apply to neutron events $(T_{\text{eff}''}^n = T_{\text{eff}'}^n)$.

Parameter	Nominal	σ
$E_{\text{non}-\text{lin.}}$	0.0000	0.0069

Table D.8: Energy non-linearity systematic uncertainty parameter.

D.3.1.3 Resolution

Energy resolution systematic uncertainties are propagated as

$$
\begin{aligned}
T_{\mathsf{eff}'''}^{\mathsf{D_2O}} =&\, T_{\mathsf{eff}''}^{\mathsf{D_2O}} + \mathsf{Gauss}\bigg\{ 0, \left[E_{\mathsf{res}}^{\mathsf{D_2O}}\right]^2 + \\
&\left[-0.516\ \mathsf{MeV} + 0.612\ \mathsf{MeV}^{1/2}\sqrt{T_{\mathsf{eff}''}^{\mathsf{D_2O}}}\right]^2 \\
&\left[H\left(\pm A_{\mathsf{diu}}\left(E_{\mathsf{res}}^{\mathsf{D_2O}}\right)\right) + H\left(\pm \aleph_{E_{\mathsf{res}}}^{\mathsf{D_2O}} A_{\mathsf{dir}}^{(\mathsf{CC,ES})}\left(E_{\mathsf{res}}^{\mathsf{D_2O}}\right)\right)\right] \\
&\left[2 + H\left(\pm A_{\mathsf{diu}}\left(E_{\mathsf{res}}^{\mathsf{D_2O}}\right)\right) + H\left(\pm \aleph_{E_{\mathsf{res}}}^{\mathsf{D_2O}} A_{\mathsf{dir}}^{(\mathsf{CC,ES})}\left(E_{\mathsf{res}}^{\mathsf{D_2O}}\right)\right)\right]\bigg\} \qquad \text{(D.4)}
\end{aligned}
$$

$$
\begin{aligned}
T_{\mathsf{eff}'''}^{\mathsf{salt}(e^-,n)} =&\, T_{\mathsf{eff}''}^{\mathsf{salt}(e^-,n)} + \mathsf{Gauss}\bigg\{ 0, \left[E_{\mathsf{res}}^{\mathsf{salt}(e^-,n)}\right]^2 + \\
&\left[-0.570\ \mathsf{MeV} + 0.651\ \mathsf{MeV}^{1/2}\sqrt{T_{\mathsf{eff}''}^{\mathsf{salt}(e^-,n)}}\right]^2 \\
&\left[H\left(\pm A_{\mathsf{diu}}\left(E_{\mathsf{res}}^{\mathsf{salt}}\right)\right) + H\left(\pm \aleph_{E_{\mathsf{res}}}^{\mathsf{salt}} A_{\mathsf{dir}}^{(\mathsf{CC,ES})}\left(E_{\mathsf{res}}^{\mathsf{salt}}\right)\right)\right] \\
&\left[2 + H\left(\pm A_{\mathsf{diu}}\left(E_{\mathsf{res}}^{\mathsf{salt}}\right)\right) + H\left(\pm \aleph_{E_{\mathsf{res}}}^{\mathsf{salt}} A_{\mathsf{dir}}^{(\mathsf{CC,ES})}\left(E_{\mathsf{res}}^{\mathsf{salt}}\right)\right)\right]\bigg\} \qquad \text{(D.5)}
\end{aligned}
$$

$$
\begin{aligned}
T_{\mathsf{eff}'''}^{\mathsf{NCD-PMT}(e^-)} =&\, T_{\mathsf{eff}''}^{\mathsf{NCD-PMT}(e^-)} + E_{\mathsf{rel\ res}}^{\mathsf{NCD-PMT}} E_{\mathsf{res\ scale}}^{\mathsf{NCD-PMT}(e^-)} \\
&\left[T_{\mathsf{eff}''}^{\mathsf{NCD-PMT}(e^-)} - T_{\mathsf{gen}}^{\mathsf{NCD-PMT}(e^-)}\right] \\
&\left\{1 \pm \frac{1}{2}\left[\aleph_{E_{\mathsf{res}}}^{\mathsf{NCD-PMT}} A_{\mathsf{dir}}^{\mathsf{ES}}\left(E_{\mathsf{res}}^{\mathsf{NCD-PMT}}\right)\right]\right\} \qquad \text{(D.6)}
\end{aligned}
$$

$$
\begin{aligned}
T_{\mathsf{eff}'''}^{\mathsf{NCD-PMT}(n)} =&\, T_{\mathsf{eff}''}^{\mathsf{NCD-PMT}(n)} + E_{\mathsf{rel\ res}}^{\mathsf{NCD-PMT}} E_{\mathsf{res\ scale}}^{\mathsf{NCD-PMT}(n)} \\
&\left[T_{\mathsf{eff}''}^{\mathsf{NCD-PMT}(n)} - 5.646\ \mathsf{MeV}\right] \\
&\left\{1 \pm \frac{1}{2}\left[\aleph_{E_{\mathsf{res}}}^{\mathsf{NCD-PMT}} A_{\mathsf{dir}}^{\mathsf{ES}}\left(E_{\mathsf{res}}^{\mathsf{NCD-PMT}}\right)\right]\right\}, \qquad \text{(D.7)}
\end{aligned}
$$

where $\text{Gauss}(\mu, \sigma^2)$ is a function that generates a random variable distributed according to $N(\mu, \sigma^2)$, where $H(x)$ is the Heaviside step function and where the positive signs are used for night events. T_{gen} is the Monte Carlo generated kinetic energy.

Parameter	Nominal	σ_+	σ_-
$E_{\text{res}}^{\text{D}_2\text{O}}$ [MeV]	0.155	0.041	0.080
$E_{\text{res}}^{\text{salt}(e^-)}$ [MeV]	0.168	0.041	0.080
$E_{\text{res}}^{\text{salt}(n)}$ [MeV]	0.154	0.018	0.018
$E_{\text{rel res}}^{\text{NCD}-\text{PMT}}$	1.0000	0.8739	0.8739
$E_{\text{res scale}}^{\text{NCD}-\text{PMT}(e^-)}$ [MeV]	0.016184		
$E_{\text{res scale}}^{\text{NCD}-\text{PMT}(n)}$ [MeV]	0.0119		
$A_{\text{diu}}\left(E_{\text{res}}^{\text{D}_2\text{O}}\right)$	0.0%	0.3%	0.3%
$A_{\text{diu}}\left(E_{\text{res}}^{\text{salt}}\right)$	0.0%	0.5%	0.5%
$A_{\text{dir}}^{\text{CC}}\left(E_{\text{res}}^{\text{D}_2\text{O}}\right)$	0.14%		
$A_{\text{dir}}^{\text{ES}}\left(E_{\text{res}}^{\text{D}_2\text{O}}\right)$	-0.64%		
$A_{\text{dir}}^{\text{CC}}\left(E_{\text{res}}^{\text{salt}}\right)$	0.13%		
$A_{\text{dir}}^{\text{ES}}\left(E_{\text{res}}^{\text{salt}}\right)$	-1.3%		
$A_{\text{dir}}^{\text{ES}}\left(E_{\text{rel res}}^{\text{NCD}-\text{PMT}}\right)$	-1.2%		

Table D.9: Energy resolution systematic uncertainty parameters.

D.3.2   Position

D.3.2.1   Shift

For all phases, position shift systematic uncertainties are applied as

$$(x', y', z') = (x, y, z) + (x, y, z)_{\text{shift}}, \tag{D.8}$$

where the uncertainties $(x, y, z)_{\text{shift}}$ are specific to each phase.

Parameter	Nominal [cm]	σ_+ [cm]	σ_- [cm]
$x_{\mathrm{shift}}^{\mathrm{D_2O}}$	0.00	1.15	0.13
$y_{\mathrm{shift}}^{\mathrm{D_2O}}$	0.00	2.87	0.17
$z_{\mathrm{shift}}^{\mathrm{D_2O}}$	5.00	2.58	0.15
$x_{\mathrm{shift}}^{\mathrm{salt}}$	0.00	0.62	0.07
$y_{\mathrm{shift}}^{\mathrm{salt}}$	0.00	2.29	0.09
$z_{\mathrm{shift}}^{\mathrm{salt}}$	5.00	3.11	0.16
$x_{\mathrm{shift}}^{\mathrm{NCD-PMT}}$	0.0	4.0	4.0
$y_{\mathrm{shift}}^{\mathrm{NCD-PMT}}$	0.0	4.0	4.0
$z_{\mathrm{shift}}^{\mathrm{NCD-PMT}}$	5.0	4.0	4.0

Table D.10: Position shift systematic uncertainty parameters.

D.3.2.2 Scale

Position scale systematic uncertainties are applied as

$$
(x'', y'')^{(\mathrm{D_2O,salt})} = \left\{ 1 + V_{r\ \mathrm{scale}}^{(\mathrm{D_2O,salt})} \pm \frac{1}{2} \Big[A_{\mathrm{diu}} \left(V_{\mathrm{scale}}^{(\mathrm{D_2O,salt})} \right) + \right.
$$
$$
\left. \aleph_{V_{r\ \mathrm{scale}}}^{(\mathrm{D_2O,salt})} A_{\mathrm{dir}}^{(\mathrm{CC,ES})} \left(V_{\mathrm{scale}}^{(\mathrm{D_2O,salt})} \right) \Big] \right\} (x', y')^{(\mathrm{D_2O,salt})} \tag{D.9}
$$

$$
z''^{(\mathrm{D_2O,salt})} = \left\{ 1 + V_{r\ \mathrm{scale}}^{(\mathrm{D_2O,salt})} + V_{z\ \mathrm{scale}}^{(\mathrm{D_2O,salt})} \pm \frac{1}{2} \Big[A_{\mathrm{diu}} \left(V_{\mathrm{scale}}^{(\mathrm{D_2O,salt})} \right) + \right.
$$
$$
\left. \aleph_{V_{r\ \mathrm{scale}}}^{(\mathrm{D_2O,salt})} A_{\mathrm{dir}}^{(\mathrm{CC,ES})} \left(V_{\mathrm{scale}}^{(\mathrm{D_2O,salt})} \right) \Big] \right\} z'^{(\mathrm{D_2O,salt})} \tag{D.10}
$$

$$
(x'', y'')^{\mathrm{NCD-PMT}} = \left(1 + V_{r\ \mathrm{scale}}^{\mathrm{NCD-PMT}} \right) \left\{ 1 \pm \frac{1}{2} \Big[A_{\mathrm{diu}} \left(V_{\mathrm{scale}}^{\mathrm{NCD-PMT}} \right) + \right.
$$
$$
\left. \aleph_{V_{r\ \mathrm{scale}}}^{\mathrm{NCD-PMT}} A_{\mathrm{dir}}^{\mathrm{ES}} \left(V_{\mathrm{scale}}^{\mathrm{NCD-PMT}} \right) \Big] \right\} (x', y')^{\mathrm{NCD-PMT}} \tag{D.11}
$$

$$
z''^{\mathrm{NCD-PMT}} = \left(1 + V_{r\ \mathrm{scale}}^{\mathrm{NCD-PMT}} + V_{z\ \mathrm{scale}}^{\mathrm{NCD-PMT}} \right) \left\{ 1 \pm \frac{1}{2} \Big[A_{\mathrm{diu}} \left(V_{\mathrm{scale}}^{\mathrm{NCD-PMT}} \right) + \right.
$$
$$
\left. \aleph_{V_{r\ \mathrm{scale}}}^{\mathrm{NCD-PMT}} A_{\mathrm{dir}}^{\mathrm{ES}} \left(V_{\mathrm{scale}}^{\mathrm{NCD-PMT}} \right) \Big] \right\} z'^{\mathrm{NCD-PMT}}, \tag{D.12}
$$

where the positive signs are used for night events.

Parameter	Nominal	σ_+	σ_-
$V_{r\ \text{scale}}^{\text{D}_2\text{O}}$	0.0000	0.0010	0.0057
$V_{z\ \text{scale}}^{\text{D}_2\text{O}}$	0.0000	0.0040	0.0000
$V_{r\ \text{scale}}^{\text{salt}}$	0.0000	0.0004	0.0034
$V_{z\ \text{scale}}^{\text{salt}}$	0.0000	0.0003	0.0025
$V_{r\ \text{scale}}^{\text{NCD}-\text{PMT}}$	0.0000	0.0029	0.0077
$V_{z\ \text{scale}}^{\text{NCD}-\text{PMT}}$	0.0000	0.0015	0.0012
$A_{\text{diu}}\left(V_{\text{scale}}^{\text{D}_2\text{O}}\right)$	0.0%	0.2%	0.2%
$A_{\text{diu}}\left(V_{\text{scale}}^{\text{salt}}\right)$	0.0%	0.3%	0.3%
$A_{\text{diu}}\left(V_{\text{scale}}^{\text{NCD}-\text{PMT}}\right)$	0.0%	0.1%	0.1%
$A_{\text{dir}}^{\text{CC}}\left(V_{\text{scale}}^{\text{D}_2\text{O}}\right)$	0.04%		
$A_{\text{dir}}^{\text{ES}}\left(V_{\text{scale}}^{\text{D}_2\text{O}}\right)$	-0.5 %		
$A_{\text{dir}}^{\text{CC}}\left(V_{\text{scale}}^{\text{salt}}\right)$	0.02%		
$A_{\text{dir}}^{\text{ES}}\left(V_{\text{scale}}^{\text{salt}}\right)$	-0.15%		
$A_{\text{dir}}^{\text{ES}}\left(V_{\text{scale}}^{\text{NCD}-\text{PMT}}\right)$	-0.18%		

Table D.11: Position scale systematic uncertainty parameters.

D.3.2.3 Resolution

Position resolution systematic uncertainties are propagated as

$$
\begin{aligned}
(x''', y''', z''')^{\mathsf{D_2O}} =& (x'', y'', z'')^{\mathsf{D_2O}} + \mathsf{Gauss}\Bigg\{0, \left[V^{\mathsf{D_2O}}_{\mathrm{res}\ (\mathrm{x,y})}\right]^2 + \\
& H\left(\pm A_{\mathrm{diu}}\left(V^{\mathsf{D_2O}}_{\mathrm{res}}\right)\right)^2\Bigg\}
\end{aligned} \tag{D.13}
$$

$$
\begin{aligned}
(x''', y''', z''')^{\mathsf{salt}} =& (x'', y'', z'')^{\mathsf{salt}} + \mathsf{Gauss}\Bigg\{0, \left[V^{\mathsf{salt}}_{\mathrm{res}\ (\mathrm{x,y})}\right]^2 + \\
& \left[H\left(\pm A_{\mathrm{diu}}\left(V^{\mathsf{salt}}_{\mathrm{res}}\right)\right) + H\left(\pm \aleph^{\mathsf{salt}}_{V_{\mathrm{res}}} A^{(\mathsf{CC,ES})}_{\mathrm{dir}}\left(V^{\mathsf{salt}}_{\mathrm{res}}\right)\right)\right]^2\Bigg\}
\end{aligned} \tag{D.14}
$$

$$
\begin{aligned}
(x''', y''')^{\mathsf{NCD-PMT}} =& (x'', y'')^{\mathsf{NCD-PMT}} + \left[V^{\mathsf{NCD-PMT}}_{\mathrm{res}\ xy\ 0} + V^{\mathsf{NCD-PMT}}_{\mathrm{res}\ xy\ 1}z + V^{\mathsf{NCD-PMT}}_{\mathrm{res}\ xy\ 2}z^2\right] \\
& \left[(x'', y'')^{\mathsf{NCD-PMT}} - (x_{\mathrm{gen}}, y_{\mathrm{gen}})^{\mathsf{NCD-PMT}}\right]
\end{aligned} \tag{D.15}
$$

$$
\begin{aligned}
z'''^{\mathsf{NCD-PMT}} =& z''^{\mathsf{NCD-PMT}} + \left[V^{\mathsf{NCD-PMT}}_{\mathrm{res}\ z\ 0} + V^{\mathsf{NCD-PMT}}_{\mathrm{res}\ z\ 1}z\right] \\
& \left[z''^{\mathsf{NCD-PMT}} - z^{\mathsf{NCD-PMT}}_{\mathrm{gen}}\right],
\end{aligned} \tag{D.16}
$$

where $\mathsf{Gauss}(\mu, \sigma^2)$ and $H(x)$ are defined as mentioned in Section D.3.1.3 and where the Gaussian random values are uncorrelated across phases.

D.3.2.4 Energy Dependence

To propagate the systematic uncertainties for the energy dependence of the fiducial volume, each event is weighted using the equation

$$
W = 1 + V_{E\ \mathrm{dep}}(T_{\mathrm{eff}'''} - 5.05\ \mathsf{MeV}), \tag{D.17}
$$

where the systematic uncertainties $V_{E\ \mathrm{dep}}$ are specific to each phase.

Parameter	Nominal [cm]	σ [cm]
$V_{\text{res } x}^{\text{D}_2\text{O}}$	0.	3.3
$V_{\text{res } y}^{\text{D}_2\text{O}}$	0.	2.2
$V_{\text{res } z}^{\text{D}_2\text{O}}$	0.	1.5
$V_{\text{res } x}^{\text{salt}}$	0.	3.1
$V_{\text{res } y}^{\text{salt}}$	0.	3.4
$V_{\text{res } z}^{\text{salt}}$	0.	5.3
$V_{\text{res } xy\ 0}^{\text{NCD}-\text{PMT}}$	0.06546	see Table D.13
$V_{\text{res } xy\ 1}^{\text{NCD}-\text{PMT}}$	-5.501×10^{-5}	see Table D.13
$V_{\text{res } xy\ 2}^{\text{NCD}-\text{PMT}}$	3.9×10^{-7}	see Table D.13
$V_{\text{res } z\ 0}^{\text{NCD}-\text{PMT}}$	0.07096	see Table D.14
$V_{\text{res } z\ 1}^{\text{NCD}-\text{PMT}}$	1.155×10^{-4}	see Table D.14
$A_{\text{diu}}\left(V_{\text{res}}^{\text{D}_2\text{O}}\right)$	0.	6.82
$A_{\text{diu}}\left(V_{\text{res}}^{\text{salt}}\right)$	0.	7.21
$A_{\text{dir}}^{\text{CC}}\left(V_{\text{res}}^{\text{salt}}\right)$	1.02	
$A_{\text{dir}}^{\text{ES}}\left(V_{\text{res}}^{\text{salt}}\right)$	3.36	

Table D.12: Position resolution systematic uncertainty parameters.

	$V_{\text{res } xy\ 0}^{\text{NCD}-\text{PMT}}$	$V_{\text{res } xy\ 1}^{\text{NCD}-\text{PMT}}$	$V_{\text{res } xy\ 2}^{\text{NCD}-\text{PMT}}$
$V_{\text{res } xy\ 0}^{\text{NCD}-\text{PMT}}$	0.000818124	-2.24984×10^{-7}	-4.19131×10^{-9}
$V_{\text{res } xy\ 1}^{\text{NCD}-\text{PMT}}$	-2.24984×10^{-7}	3.66098×10^{-9}	3.71423×10^{-12}
$V_{\text{res } xy\ 2}^{\text{NCD}-\text{PMT}}$	-4.19131×10^{-9}	3.71423×10^{-12}	3.92118×10^{-14}

Table D.13: Covariance matrix used to apply the position resolution uncertainty in the x and y directions.

	$V_{\text{res } z\ 0}^{\text{NCD}-\text{PMT}}$	$V_{\text{res } z\ 1}^{\text{NCD}-\text{PMT}}$
$V_{\text{res } z\ 0}^{\text{NCD}-\text{PMT}}$	0.00078696	3.47188×10^{-7}
$V_{\text{res } z\ 1}^{\text{NCD}-\text{PMT}}$	3.47188×10^{-7}	6.80761×10^{-9}

Table D.14: Covariance matrix used to apply the position resolution uncertainty in the z direction.

Parameter	Nominal $[\text{MeV}^{-1}]$	σ_+ $[\text{MeV}^{-1}]$	σ_- $[\text{MeV}^{-1}]$
$V_{E\ \text{dep}}^{\text{D}_2\text{O}}$	0	0.0085	0.0049
$V_{E\ \text{dep}}^{\text{salt}}$	0	0.0041	0.0048
$V_{E\ \text{dep}}^{\text{NCD−PMT}}$	0	0.0088	0.0067

Table D.15: Fiducial volume energy dependence systematic uncertainty parameters.

D.3.3   Direction

Systematic uncertainties on the $\cos\theta_\odot$ observable are applied as

$$
\cos\theta_\odot^{\prime(\text{D}_2\text{O,salt})} = 1 + \left[1 + \Delta_\theta^{(\text{D}_2\text{O,salt})} \pm \frac{1}{2}\aleph_{\Delta_\theta}^{(\text{D}_2\text{O,salt})} A_{\text{dir}}^{\text{ES}}\left(\Delta_\theta^{(\text{D}_2\text{O,salt})}\right)\right]
$$
$$
\left(\cos\theta_\odot^{(\text{D}_2\text{O,salt})} - 1\right) \tag{D.18}
$$
$$
\cos\theta_\odot^{\prime\text{NCD−PMT}} = 1 + \left[1 + \Delta_\theta^{\text{NCD−PMT}}\right]\left[1 \pm \frac{1}{2}\aleph_{\Delta_\theta}^{\text{NCD−PMT}} A_{\text{dir}}^{\text{ES}}\left(\Delta_\theta^{\text{NCD−PMT}}\right)\right]
$$
$$
\left(\cos\theta_\odot^{\text{NCD−PMT}} - 1\right), \tag{D.19}
$$

using the same sign convention as the previous uncertainties.

Parameter	Nominal	σ
$\Delta_\theta^{\text{D}_2\text{O}}$	0.	0.11
$\Delta_\theta^{\text{salt}}$	0.	0.11
$\Delta_\theta^{\text{NCD−PMT}}$	0.	0.12
$A_{\text{dir}}^{\text{ES}}\left(\Delta_\theta^{\text{D}_2\text{O}}\right)$	-2.18%	
$A_{\text{dir}}^{\text{ES}}\left(\Delta_\theta^{\text{salt}}\right)$	-5.2%	
$A_{\text{dir}}^{\text{ES}}\left(\Delta_\theta^{\text{NCD−PMT}}\right)$	-6.9%	

Table D.16: $\cos\theta_\odot$ systematic uncertainty parameters.

D.3.4 Isotropy

D.3.4.1 Scale

β_{14} scale systematic uncertainties are propagated as

$$\beta_{14}^{\prime(D_2O,\text{salt } e^-)} = \beta_{14}^{(D_2O,\text{salt } e^-)} \Big[1 + B_{\text{scale shift } 0}^{(D_2O,\text{salt } e^-)} + B_{\text{scale } 0}^{D_2O \text{ \& salt } e^-} +$$

$$B_{\text{scale } 1} \left(T_{\text{eff}}^{(D_2O,\text{salt } e^-)} - 5.589 \text{ MeV} \right) \Big] \tag{D.20}$$

$$\beta_{14}^{\prime\text{salt } n} = \beta_{14}^{\text{salt } n} \left[1 + B_{\text{scale } 0}^{\text{salt } n} + B_{\text{scale } 1} \left(T_{\text{eff}}^{\text{salt } n} - 5.589 \text{ MeV} \right) \right]. \tag{D.21}$$

Parameter	Nominal	σ_+	σ_-
$B_{\text{scale shift } 0}^{D_2O}$	-0.0081		
$B_{\text{scale shift } 0}^{\text{salt } e^-}$	$0.$		
$B_{\text{scale } 0}^{D_2O \text{ \& salt } e^-}$	$0.$	0.002084	0.002084
$B_{\text{scale } 0}^{\text{salt } n}$	-0.0144	0.0038	0.0020
$B_{\text{scale } 1} \text{ [MeV}^{-1}\text{]}$	0.00275597	0.00048790	0.00048790

Table D.17: β_{14} scale systematic uncertainty parameters.

D.3.4.2 Shift

The β_{14} shift systematic uncertainty is propagated as

$$\beta_{14}^{\prime\prime(D_2O,\text{salt})} = \beta_{14}^{\prime(D_2O,\text{salt})} \pm \frac{1}{2} \Big[A_{\text{diu}} \left(B_{\text{shift}}^{(D_2O,\text{salt})} \right) +$$

$$\aleph_{B_{\text{shift}}}^{(D_2O,\text{salt})} A_{\text{dir}}^{(CC,ES)} \left(B_{\text{shift}}^{(D_2O,\text{salt})} \right) \Big], \tag{D.22}$$

using the usual conventions.

Parameter	Nominal	σ
$A_{\text{diu}}\left(B_{\text{shift}}^{\text{D}_2\text{O}}\right)$	0.	0.0043
$A_{\text{diu}}\left(B_{\text{shift}}^{\text{salt}}\right)$	0.	0.0043
$A_{\text{dir}}^{\text{CC}}\left(B_{\text{shift}}^{\text{D}_2\text{O}}\right)$	0.00038	
$A_{\text{dir}}^{\text{ES}}\left(B_{\text{shift}}^{\text{D}_2\text{O}}\right)$	-0.0034	
$A_{\text{dir}}^{\text{CC}}\left(B_{\text{shift}}^{\text{salt}}\right)$	0.00038	
$A_{\text{dir}}^{\text{ES}}\left(B_{\text{shift}}^{\text{salt}}\right)$	-0.0034	

Table D.18: β_{14} shift systematic uncertainty parameters.

D.3.4.3 Width

β_{14} width systematic uncertainties are applied as

$$\beta_{14}'''^{\text{D}_2\text{O}(e^-,n)} = \beta_{14}''^{\text{D}_2\text{O}(e^-,n)} + B_{\text{width}}^{\text{D}_2\text{O \& salt } e^-}$$

$$\left[\beta_{14}''^{\text{D}_2\text{O}(e^-,n)} - \bar{\beta}_{14_j}^{\text{D}_2\text{O}(e^-,n)}\left(T_{\text{eff}}^{\text{D}_2\text{O}(e^-,n)}\right)\right] \tag{D.23}$$

$$\beta_{14}'''^{\text{salt } e^-} = \beta_{14}''^{\text{salt } e^-} + B_{\text{width}}^{\text{D}_2\text{O \& salt } e^-}$$

$$\left[\beta_{14}''^{\text{salt } e^-} - \bar{\beta}_{14_j}^{\text{salt } e^-}\left(T_{\text{eff}}^{\text{salt } e^-}\right)\right] \tag{D.24}$$

$$\beta_{14}'''^{\text{salt } n} = \beta_{14}''^{\text{salt } n} + \text{Gauss}\left(0, \left[B_{\text{width}}^{\text{salt } n}\right]^2\right), \tag{D.25}$$

where the functions $\bar{\beta}_{14_j}\left(T_{\text{eff}}\right)$ give the β_{14} mean in the simulation data as a function of the kinetic energy for a given type of particle (electron-like or neutrons) and a given class of events, and where $\text{Gauss}(\mu, \sigma^2)$ is defined as usual.

Parameter	Nominal	σ
$B_{\text{width}}^{\text{D}_2\text{O \& salt } e^-}$	0.	0.003315
$B_{\text{width}}^{\text{salt } n}$	0.0150	0.0045

Table D.19: β_{14} width systematic uncertainty parameter.

D.3.5 PMT Photodisintegration Events

PMT $\beta - \gamma$ background events for the D_2O and salt phases are modelled using the analytical PDF

$$f(T_{\text{eff}}, \beta_{14}, \rho, \cos\theta_\odot) \propto e^{\epsilon\,T_{\text{eff}}} \mathcal{N}(\beta_{14} \,|\, \bar{\beta}_{14} = \omega_0 + \omega_1(\rho - 0.61), \sigma = \beta_s)$$

$$(|b + 1| - 1 + e^{\nu\rho}), \tag{A.1}$$

where $\mathcal{N}(x|\bar{x}, \sigma)$ is a Gaussian distribution with the given mean and width. All parameters are phase and day/night specific. Parameters b and ν are expressed as

$$b = b_0 + b_u \aleph_1 \tag{D.26}$$

$$\nu = \nu_0 + \nu_u \left(\rho[b, \nu]\aleph_1 + \sqrt{1 - \rho[b, \nu]^2}\aleph_2 \right), \tag{D.27}$$

where $\aleph_1$ and $\aleph_2$ are also specific (independent) for each phase and day/night.

Parameter	D_2O Phase		salt Phase	
	Day	Night	Day	Night
ϵ	-6.73 ± 1.29	-5.64 ± 1.02	-6.26 ± 0.91	-6.98 ± 0.91
ω_0	0.533 ± 0.014	0.533 ± 0.014	0.511 ± 0.007	0.511 ± 0.007
ω_1	0.182 ± 0.095	0.182 ± 0.095	0.237 ± 0.051	0.237 ± 0.051
β_s	0.182 ± 0.011	0.182 ± 0.011	0.195 ± 0.007	0.195 ± 0.007
$b_0 \pm b_u$	-1.00 ± 1.29	3.27 ± 12.04	-0.33 ± 2.08	0.49 ± 3.02
$\nu_0 \pm \nu_u$	6.63 ± 0.93	6.78 ± 1.52	5.32 ± 1.01	5.66 ± 1.07
$\rho[b, \nu]$	0.60	0.96	0.91	0.94

Table D.20: Systematic uncertainty parameters for the PMT β-γ analytical PDFs.

D.3.6   Other

D.3.6.1   Winter et al. ^{8}B Spectral Shape Uncertainty

For the three phases, the Winter *et al.* [56] ^{8}B spectral shape uncertainty is propagated
by reweighting the CC, ES and NC events using the function

$$W^{(\text{CC,ES,NC})} = 1 + \frac{\aleph_{^8\text{B}}^{(\text{CC,ES,NC})}}{3}[0.018 - 0.001999 \text{ MeV}^{-1} \times E_\nu -$$
$$8.8769 \times 10^{-5} \text{ MeV}^{-2} \times E_\nu^2], \tag{D.28}$$

where parameters $\aleph_{^8\text{B}}$ are phase-correlated.

D.3.6.2   Salt Phase ^{24}Na Distribution

For the salt phase, the uncertainty associated to the distribution of ^{24}Na background
events is evaluated by reweighting these events using

$$W = \frac{(z''' \pm 600)^2}{432000}. \tag{D.29}$$

D.3.6.3 Efficiency Systematic Uncertainties

Efficiency systematic uncertainties are applied as

$$W^{(\text{D}_2\text{O NC,salt NC})} = 1 + \Delta\epsilon_n^{\text{corr}} + \Delta\epsilon_n^{(\text{D}_2\text{O,salt})} \tag{D.30}$$

$$W^{(\text{D}_2\text{O } n \text{ bg,salt } n \text{ bg})} = \left[1 + \Delta\epsilon_n^{\text{corr}} + \Delta\epsilon_n^{(\text{D}_2\text{O,salt})}\right]\left[1 + \Delta\epsilon_{\text{photo}}\right] \tag{D.31}$$

$$W^{(\text{NCD}-\text{PMT NC,NCD}-\text{NCD NC})} = \epsilon_{\text{NC}}^{(\text{NCD}-\text{PMT,NCD}-\text{NCD})}, \tag{D.32}$$

where $\Delta\epsilon_n^{\text{corr}}$ is the correlated neutron capture efficiency uncertainty, where $\Delta\epsilon_n^{\text{D}_2\text{O}}$ and $\Delta\epsilon_n^{\text{salt}}$ are the neutron capture efficiency uncertainties for the D_2O and salt phases respectively, where $\Delta\epsilon_{\text{photo}}$ is the efficiency uncertainty for photodisintegration events and where $\epsilon_{\text{NC}}^{\text{NCD}-\text{PMT}}$ and $\epsilon_{\text{NC}}^{\text{NCD}-\text{NCD}}$ are respectively the NC capture efficiency uncertainties for the PMT and NCD side of the NCD phase.

Parameter	Nominal	σ
$\Delta\epsilon_n^{\text{corr}}$	0	0.007
$\Delta\epsilon_n^{\text{D}_2\text{O}}$	0	0.0187
$\Delta\epsilon_n^{\text{salt}}$	0	0.0124
$\epsilon_{\text{NC}}^{\text{NCD}-\text{PMT}}$	1	0.0279
$\epsilon_{\text{NC}}^{\text{NCD}-\text{NCD}}$	1	0.0237
$\Delta\epsilon_{\text{photo}}$	0	0.02

Table D.21: Efficiency systematic uncertainty parameters.

D.3.6.4 Sacrifice Uncertainties

In addition to the systematic uncertainties that were presented in the previous section, other parameters are added to consider the uncertainties on the effects of the analysis

cuts [57]. These uncertainties have been evaluated separately for each signal and background in an energy-dependent fashion and they are applied as such during the signal extraction process.

D.4 PSA Constraint

As shown in Section 3.6, the PSA constraint is applied as

$$LL_{\mathsf{PSA}} = -\frac{1}{2} \left(\frac{\left[\sum_{j \in \{n \text{ classes}\}} \alpha_j \nu_{j\,\mathsf{exp}}(\vec{\eta}) \right] - n_n}{n_n \delta_{nn}} \right)^2.$$ (3.33)

The constraint mean and uncertainty are shown on Table D.22.

Parameter	Nominal	σ
$n_n \pm n_n \delta_{nn}$	1114.6	78.3

Table D.22: PSA constraint used by the final three-phase fits.

References

[1] Q. R. Ahmad et al. (SNO Collaboration). *Phys. Rev. Lett.*, **89**, 011301 (2002).

[2] S. Ahmed et al. (SNO Collaboration). *Phys. Rev. Lett.*, **92**, 181301 (2004).

[3] R. Brun, F. Rademakers et al. "ROOT — A Data Analysis Framework",
`http://root.cern.ch` (2011).

[4] B. Aharmim et al. (SNO Collaboration). *Phys. Rev. C*, **81**, 055504 (2010).

[5] GNU Project. "The GNU Operating System", `http://www.gnu.org` (2011).

[6] FreeBSD Project. "The FreeBSD Project", `http://www.freebsd.org` (2011).

[7] B. Aharmin et al. (SNO Collaboration). *Phys. Rev. Lett.*, **101**, 111301 (2008).

[8] L. Lamport et al. "LaTeX – A document preparation system",
`http://www.latex-project.org` (2011).

[9] B. Moolenaar et al. "Vim the editor", `http://www.vim.org` (2011).

[10] T. Ohl. "FeynMP MetaPost Pacakge",
`http://www.ctan.org/tex-archive/macros/latex/contrib/feynmf` (2011).

[11] D. Knuth. "TeX Users Group home page", `http://www.tug.org` (2011).

[12] T. Van Zandt et al. "PSTricks web site", `http://tug.org/PSTricks` (2011).

[13] W. Pauli. "Letter to L. Meitner and her colleagues dated 4 December 1930".

[14] J. Chadwick. *Verh. d. deutschen phys. Ges.*, **16**, 383 (1914).

[15] C.L. Cowan Jr., F. Reines, F.B. Harrison, H.W. Kruse and A.D. McGuire. *Science*, **124**, 103 (1956).

[16] G. Danby, J.-M. Gaillard, K. Goulianos, L.M. Lederman, N. Mistry, M. Schwartz and J. Steinberger. *Phys. Rev. Lett.*, **9**, 36 (1962).

[17] K. Kodama et al. (DONUT Collaboration). *Phys. Lett. B*, **504**, 3 (2001) 218–224.

[18] J. N. Bahcall, A. Serenelli and S. Basu. *ApJ*, **621**, L85 (2000).

[19] J. N. Bahcall. *Nucl. Phys. Proc. Suppl.*, **118** (2003) 77–86.

[20] B. T. Cleveland et al. *Astrophys. J.*, **496** (1998) 505–526.

[21] B. Pontecorvo. *Sov. Phys. JETP*, **6** (1957) 429.

[22] Z. Maki, M. Nakagawa and S. Sakata. *Progress of Theoretical Physics*, **28**, 5 (1962) 870–880.

[23] L. Wolfenstein. *Phys. Rev.*, **D17**, 2369 (1978).

[24] S. P. Mikheev and A. Y. Smirnov. *Sov. J. Nucl. Phys.*, **42**, 913 (1985).

[25] K. Nakamura et al. (Particle Data Group). *J. Phys. G*, **37**, 075021 (2010).

[26] Q. R. Ahmad et al. (SNO Collaboration). *Phys. Rev. Lett.*, **87**, 071301 (2001).

[27] B. A. Moffat. "The Optical Calibration of the Sudbury Neutrino Observatory". Ph.D. thesis, Queen's University (2001).

[28] B. Aharmim et al. (SNO Collaboration). *Nucl. Instrum. Meth.*, **A604** (2009) 531–535.

[29] M. R. Dragowsky et al. *Nucl. Instrum. Meth.*, **A481**, 284 (2002).

[30] N. Tagg et al. *Nucl. Instrum. Meth.*, **A489**, 178 (2002).

[31] M. S. Kos. "Low Energy Solar Neutrino Analysis of the Salt Phase of the Sudbury Neutrino Observatory". Ph.D. thesis, Queen's University (2007).

[32] C. Howard. "Injection of a Uniform ^{24}Na Radioactive Source into the Sudbury Neutrino Observatory". Master's thesis, Queen's University (2006).

[33] M. G. Boulay. "Direct Evidence for Weak Flavour Mixing with the Sudbury Neutrino Observatory". Ph.D. thesis, Queen's University (2001).

[34] J. Boger et al. (SNO Collaboration). *Nucl. Instrum. Meth.*, **A449** (2000) 172–207.

[35] H. H. Chen. *Phys. Rev. Lett.*, **55** (1985) 1534–1536.

[36] J. C. Loach. "Measurement of the Flux of ^{8}B Solar Neutrinos at the Sudbury Neutrino Observatory". Ph.D. thesis, University of Oxford (2008).

[37] J.F. Amsbaugh et al. *Nucl. Instrum. Meth.*, **A579**, 3 (2007) 1054–1080.

[38] B. Aharmim et al. (SNO Collaboration). arXiv:1107.2901 (Submitted to Phys. Rev. C) (2011).

[39] B. Aharmim et al. (SNO Collaboration). arXiv:1109.0763 (Submitted to Phys. Rev. C) (2011).

[40] K. Levenberg. *The Quarterly of Applied Mathematics*, **2** (1944) 164–168.

[41] D. Marquardt. *SIAM Journal on Applied Mathematics*, **11**, 2 (1963) 431–441.

[42] M. S. Neubauer. "Evidence for Electron Neutrino Flavor Change Through Measurement of the ^{8}B Solar Neutrino Flux at the Sudbury Neutrino Observatory". Ph.D. thesis, University of Pennsylvania (2001).

[43] B. Aharmim et al. (SNO Collaboration). *Phys. Rev. C*, **75**, 4 (2007).

[44] C. J. Jillings. "The Electron-Scattering Reaction in the Sudbury Neutrino Observatory". Ph.D. thesis, Queen's University (1999).

[45] S. M. Oser. "A Comment on Implementing the Angular Resolution Uncertainty in Sigex", (SNO Internal Note) (2008).

[46] M. Dunford. "Measurement of the ^{8}B Solar Neutrino Energy Spectrum at the Sudbury Neutrino Observatory". Ph.D. thesis, University of Pennsylvania (2006).

[47] R. F. MacLellan. "The Energy Calibration for the Solar Neutrino Analysis of All Three Phases of the Sudbury Neutrino Observatory". Ph.D. thesis, Queen's University (2009).

[48] A. Wright. "Robust Signal Extraction Methods and Monte Carlo Sensitivity Studies for the Sudbury Neutrino Observatory and SNO+ Experiments". Ph.D. thesis, Univeristy of Washington (2009).

[49] G. D. Orebi Gann. "An Improved Measurement of the ^{8}B Solar Neutrino Energy Spectrum at the Sudbury Neutrino Observatory". Ph.D. thesis, Merton College, Oxford (2008).

[50] S. R. Seibert. "A Low Energy Measurement of the ^{8}B Solar Neutrino Spectrum at the Sudbury Neutrino Observatory". Ph.D. thesis, University of Texas at Austin (2008).

[51] J. N. Bahcall et al. arXiv:astro-ph/0010346v1 (2000).

[52] A. M. Serenelli, S. Basu, J. W. Ferguson and M. Asplund. *ApJ*, **705**, L123 (2009).

[53] B. Aharmim et al. (SNO Collaboration). *Phys. Rev. C*, **72**, 055502 (2005).

[54] G. Cowan. "Statistical Data Analysis". Oxford Science Publications, Clarendon Press, Oxford (1998).

[55] P.-L. Drouin. "FTN: A Time Fitter for the NCD Phase", (SNO Internal Note) (2008).

[56] W. T. Winter, S. J. Freedman, K. E. Rehm and J. P. Schiffer. *Phys. Rev. C*, **73**, 025503 (2006).

[57] G. D. Orebi Gann. "Private Communication".