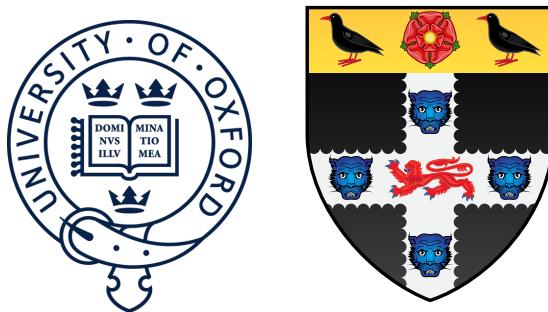


First Measurement of the Solar Neutrino Oscillation Parameters via Boron-8 Solar Neutrinos in SNO+



Daniel Cookman

Christ Church College

University of Oxford

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Trinity Term 2022

To my parents

and

To Oscar Jacobsson:

The best of us

Acknowledgements

And I would like to acknowledge ...

Abstract

Formal 1-page summary of the work completed in the thesis.

Table of contents

List of Figures	xiii
List of Tables	xv
Introduction	1
1 The Theory of Neutrino Physics	3
1.1 The Standard Model and Neutrinos	3
1.1.1 A Brief Introduction to the Standard Model	3
1.1.2 Neutrinos within the Standard Model	4
1.2 Neutrino Oscillations and Neutrino Masses	5
1.2.1 The Evidence for Neutrino Oscillations	5
1.2.2 The Phenomenology of Neutrino Oscillations	6
1.2.3 The Origins of Neutrino Mass	6
2 The SNO+ Detector	9
2.1 Detector Geometry and Design	9
2.2 Experimental Phases	11
2.3 Detecting and Recording an Event in SNO+: A Journey	12
2.3.1 Particle Interactions with Matter	13
2.3.2 Optical Processes	17

2.3.3	Detection by PMTs	21
2.3.4	Data Acquisition and Triggering	22
2.3.5	Operation of the Detector	23
2.4	Calibrations and Detector Modelling	23
2.4.1	Detector Monitoring	23
2.4.2	Electronic and PMT Calibrations	23
2.4.3	Energy and Optical Calibrations	24
2.4.4	Event Reconstruction	24
2.4.5	Event Simulation	25
3	The SMELLIE Calibration System	27
3.1	The SMELLIE Hardware	28
3.2	Software for SMELLIE Data-taking	28
3.3	Commissioning SMELLIE in the Scintillator Phase	29
4	Simulating SMELLIE Events	31
4.1	Improving the SMELLIE Generator Algorithm	32
4.1.1	Previous Attempts at SMELLIE Event Simulation	32
4.1.2	The new generator	35
4.2	Improving the beam profiles	39
4.2.1	Combining beam profile datasets	42
4.2.2	Results & Discussion	48
4.3	Comparisons between Data and Simulation	52
4.3.1	Forward Hemisphere Discrepancies	53
4.3.2	Emission Time Discrepancies	53
4.3.3	Backward Hemisphere Discrepancies	53

Table of contents	xi
5 Analysis of SMELLIE Data in the Scintillator Phase	55
5.1 Extinction Length Analysis	55
5.1.1 Motivation	55
5.1.2 Analysis Approach	56
5.1.3 Validation of the Analysis in Simulation	57
5.1.4 Results in Data	57
5.2 Scattering Analysis	58
5.2.1 Historical Approaches and the Problem of Systematics	58
5.2.2 New Methodology	59
5.2.3 Results	60
6 Solar Oscillation Analysis	61
6.1 Analysis Methodology	62
6.1.1 Observational Principle	62
6.1.2 Background Processes	64
6.1.3 The Log-likelihood Test Statistic	70
6.1.4 The Bayesian Statistical Approach	72
6.1.5 Markov Chain Monte Carlo	74
6.1.6 Choosing Priors	76
6.1.7 Including Systematics in the Fit	76
6.1.8 Including Oscillations in the Fit	78
6.2 Analysis on Scintillator-Phase data	80
6.2.1 Dataset and Livetime	80
6.2.2 Event Selection	81
6.2.3 Expected Rates and Constraints	81
6.2.4 Results	82
6.3 Sensitivity Projections	83

7 Conclusions and Suggestions for Future Work	85
References	89

List of Figures

2.1	3D model of the SNO+ detector	10
4.1	Comparison of SMELLIE between data and MC using a 1D generator .	33
4.2	Number of attempts taken for existing 2D SMELLIE generator to accept direction	36
4.3	Intensity map of existing beam profile, and interpolated sampling map	37
4.4	Histogram of interpolated intensities within the 2D direction-space . .	38
4.5	Comparison of water-phase data to MC generated using both the old and new 2D beam profile generator approaches, with the updated beam profiles	40
4.6	Plot of $\Delta\chi^2$ for both single subruns of a typical PMT, and when all relevant subruns are combined	47
4.7	Comparison between old and updated beam profiles for fibre FS055, after combining multiple data sets	49
4.8	Interpolated intensity map for the new updated beam profile of fibre FS055	50
4.9	Residuals from subruns at two different wavelengths, both compared to the combined beam profile model for fibre FS055	52

List of Tables

4.1 Water-phase runs used for new beam profiling. 43

Introduction

Couple of pages outlining document's structure and contents (this is what each of the chapters is here for). Less formal than the abstract, also explaining what the expected audience of this thesis is: who will find this document useful!

1

2

3

4

Chapter 1

The Theory of Neutrino Physics

Light

Light

The visible reminder of Invisible Light

1

2

3

The Rock

T. S. ELIOT

1.1 The Standard Model and Neutrinos

1.1.1 A Brief Introduction to the Standard Model

Covering how the SM works at the highest level, including:

- Quantum Field Theory and the Lagrangian dynamical framework

4

5

6

- The connection between symmetries of a QFT model and its gauge fields that describe the model's forces
- The SM's fundamental symmetries, and associated forces, but —

7

8

9

10

- 1 • Not (exactly) what we see “normally”! The electromagnetic and weak forces
2 appear distinct, and the weak gauge bosons have mass. To explain this, we need
3 a further component, the Brout-Englert-Higgs (BEH) Mechanism.

4 [2 pages total]

5 **1.1.2 Neutrinos within the Standard Model**

- 6 • Basic description of where neutrinos fit into SM: 3 kinds of neutral fermion, the
7 counterparts to the charged fermions. Interacts with the weak force only.
- 8 • Summary of the experimental evidence for this picture: mainly, the discovery of
9 electron anti-neutrinos by Cowan and Reines, the muon neutrino by Lederman,
10 Schwartz, and Steinberger, and the tau neutrino by the DONUT Collaboration.
11 Further critical experiments include the first measurement of a neutrino’s helicity
12 by Goldhaber et al. as well as Danby et al.’s demonstration that ν_μ are distinct
13 from ν_e .
- 14 • More detailed description, via Feynman diagrams, of the two fundamental modes
15 of interaction by neutrinos with the weak force: charged- and neutral-current
16 interactions. A brief mention of the quantitative theory that underlies description:
17 Gashow, Salam, and Weinberg’s Electroweak Theory. This explains not only the
18 V–A structure of charged-current interactions, but also predicted accurately the
19 nature of neutral-current interactions. (Given space constraints, I see no reason
20 to go into much of the details of the theory, or the many experimental tests of
21 its structure.)

22 [4 pages]

1.2 Neutrino Oscillations and Neutrino Masses

1

1.2.1 The Evidence for Neutrino Oscillations

2

- Describe status quo ante of massless nature of neutrinos: BEH mechanism as exists cannot allow for neutrinos to have mass as only left-handed neutrinos have been observed.
3
4
5
- Furthermore, strong experimental limits on neutrino masses, from e.g. tritium-decay endpoint measurements by the KATRIN experiment and cosmological inferences from the CMB by the Planck satellite.
6
7
8
- But — then neutrino oscillations are observed over a variety of experiments and contexts. Summarise critical bits of evidence:
9
10
- Electron neutrino disappearance in solar neutrino experiments, including Ray Davis' Homestake experiment, the SAGE/GALLEX experiments, and SNO. For the latter, the comparison of charged-current and neutral-current modes of interaction was clear evidence of neutrino oscillations over other types of process (e.g. neutrino decay).
11
12
13
14
15
- Include in the above a brief description of Bahcall's Standard Solar Model.
16
- Muon neutrino disappearance in atmospheric and long-baseline accelerator neutrino experiments, such as Super-Kamiokande, T2K, and No ν a.
17
18
- A few further observations to note are: reactor electron anti-neutrino disappearance from both KamLAND and Daya Bay; tau neutrino appearance at the OPERA experiment; short-baseline neutrino anomaly within LSND and MiniBooNE (with recent contrary evidence from MicroBooNE).
19
20
21
22

1.2.2 The Phenomenology of Neutrino Oscillations

- Describe the current phenomenological model of 3-flavour neutrino oscillations that can explain all of this evidence: the PMNS mixing matrix.
- Describe also the MSW effect, which is critical for explaining solar neutrino oscillations.
- Show the formula for solar neutrino oscillations, given this MSW effect in both the Sun and Earth. Note the dependence of solar neutrino oscillations on only the “solar” oscillation parameters. This is all particularly useful for the solar analysis chapter.

$$P_{ee} \left(\tan 2\theta_{12}^M, \sin \theta_{13}^M, \Delta m_{21,M}^2 \right) = BLAH \quad (1.1)$$

[3 pages]

1.2.3 The Origins of Neutrino Mass

- Observed neutrino oscillations require at least two neutrino mass states to be non-zero. Given constraints of the current SM, two main ways of adding neutrino masses: a Dirac mass term (i.e. allowing for sterile neutrinos), and a Majorana mass term.
- For latter, briefly describe what a Majorana particle is, and how with the Seesaw Mechanism (just the simple Type 1 described in-text) one can not only get neutrino masses but also explain their lightness relative to the other massive SM particles. Note that there exist more elaborate versions of this theory.
- Furthermore, with reference to the Sakharov conditions, describe qualitatively how the Seesaw Mechanism also allows for possible leptogenesis/baryogenesis in the early Universe, and hence could explain its matter-antimatter asymmetry.

1.2 Neutrino Oscillations and Neutrino Masses

7

- Describe briefly the nuclear physics behind double-beta decay (i.e. why it can happen at all over just normal beta decay), and then how Majorana neutrinos allow for neutrinoless double beta decay, $0\nu\beta\beta$.
1
2
3
- Describe the experimental signature of $0\nu\beta\beta$: a spike of events of observed energy equal to the Q-value of the decay.
4
5
- Note Schecter-Valle Theorem ensures that any observation of $0\nu\beta\beta$ must be the result of neutrinos being Majorana. I.e. the Universe cannot conspire against us and have $0\nu\beta\beta$ without Majorana neutrinos.
6
7
8
- Very briefly note the current status of the search for $0\nu\beta\beta$, describing the main varieties of experimental setup seen, along with a nice canonical example of such an experiment and their best limit. In particular, the Germanium-crystal detectors such as GERDA, Xenon-TPC detectors like EXO-200, and large-scale liquid scintillators such as KamLAND-Zen.
9
10
11
12
13

[3 pages]

14

[CHAPTER TOTAL: 17 pages]

15

Chapter 2

The SNO+ Detector

The light-soaked days are coming.

JOHN GREEN

2.1 Detector Geometry and Design

The SNO+ detector is a large, multi-purpose neutrino detector built in the SNOLAB underground laboratory near Sudbury, Canada. Its main detector structure is taken from the Nobel prize-winning Sudbury Neutrino Observatory (SNO) [], which can be seen in Fig. 2.1. The bulk of the detector is the main detector medium, which changes depending on the phase of the experiment — more on the specifics of this shortly. This medium is held within a 12 m diameter sphere known as the Acrylic Vessel (AV). The AV floats within a body of ultra-pure water (UPW), beyond which is a stainless steel support structure (PSUP) that holds ~ 9000 Photomultiplier Tubes (PMTs). It is these PMTs that detect the light generated from physics events that occur within the detector medium. The AV is kept in place relative to the PSUP through a series of ‘hold-up’ and ‘hold-down’ tensylon ropes. All of these components are suspended within a large cylindrical cavity also filled with UPW. Directly above the detector is



Fig. 2.1 3D model of the SNO+ detector [1].

1 the Deck, within which all the detector electronics are kept. Access within the AV for
2 calibration tools and filling is possible only through the acrylic ‘neck’ on top of the AV.

3 This choice of design is highly deliberate, with the details are discussed in [1]. The
4 location 2.2 km underground ensures that there is minimal impact from cosmic rays:
5 only 3 cosmic ray muon events are expected within the detector an hour [1]. By making
6 the detector spherical, the high degree of symmetry can easily be taken advantage of
7 in event reconstruction and analysis. Moreover, light produced throughout most of
8 the body of the AV will be minimally-impacted by refraction through the acrylic. The
9 only major exception to this is light emitted within ~ 50 cm of the AV, at which point
10 total internal reflection becomes possible. In order to make as much emitted light be
11 able to get detected as possible, all materials within the PSUP were chosen for their
12 optical transparency (excepting the ropes).

2.2 Experimental Phases**11**

Another major design consideration is that of radioactive backgrounds. Maintaining minimal levels of backgrounds is critical for effective particle physics, otherwise the signal one is searching for would become completely swamped. Materials within the detector were chosen to ensure these background levels would be low enough for the Collaboration’s physics goals to be achieved. Another benefit of the spherical design is its high volume-to-surface area ratio, which means that the relatively-high background levels of the PMT glass are kept far away from the detection medium.

2.2 Experimental Phases

As mentioned earlier, SNO+ was designed to fulfil a number of physics goals over multiple ‘phases’ of the detector’s lifetime. The phases are distinguished by the medium that fills the AV. The first main phase (after a brief ‘air-filled’ phase used only for detector commissioning) was that of the ‘water-fill’, with data taken between May 2017 and July 2019. This was used to perform fundamental optical calibrations of the detector [], measurements of the solar neutrino flux [], observation of neutrino oscillations in reactor anti-neutrinos [], and searches for nucleon decay [].

After this, the detector was filled with 800 tonnes of liquid scintillator known as linear alkylbenzene (LAB), mixed with the fluor PPO. More information on the physics of scintillators can be found in Section 2.3.1. Filling of the LABPPO cocktail had to be paused in March 2020 due to the COVID-19 pandemic, leading to the detector having its bottom half still filled with UPW, and the top half filled with LAB and PPO at 0.5 g/L. This impromptu phase became known as the ‘partial-fill’, and allowed for some creative analyses to be performed: an initial neutrino oscillation analysis from reactor anti-neutrinos [], as well as the first ever observation of directionality in a high light yield scintillator []. Eventually, filling of the detector with liquid scintillator was able to resume, being completed in May 2021. At that point, the concentration of PPO in the

1 detector was at 0.6 g/L, markedly below the target level of 2.0 g/L. A further ‘PPO
2 top-up’ campaign then proceeded, finishing in April 2022 with a final concentration of
3 2.2 g/L PPO. Thus began the ‘scintillator-fill’ of the experiment, which continues on
4 during the time of writing. The main goals for this phase include a number of solar
5 neutrino analyses (including the one described in Chapter 6), a precision measurement
6 of the neutrino oscillation parameter Δm_{21}^2 using reactor anti-neutrinos [], and further
7 calibrations of the detector and its backgrounds.

8 Finally, in the near future the detector will be loaded with Tellurium, allowing for
9 the flagship analysis of the experiment to begin: neutrinoless double-beta decay. The
10 details of this chemical loading process are described in []. Alongside the Te-loaded
11 liquid scintillator will be a number of other chemicals within the scintillator cocktail
12 that will help ensure optimal optical properties and chemical stability. These include:
13 the surfactant DDA to ensure the solubility and stability of the ‘Te-diol’ within the
14 LAB; the wavelength-shifter BisMSB to absorb light at short wavelengths and re-emit
15 closer to the optimal quantum efficiency of the detector’s PMTs; and the anti-oxidant
16 BHT to prevent any free-radicals within the liquid scintillator from ‘yellowing’ the
17 medium.

18 **2.3 Detecting and Recording an Event in SNO+:** 19 **A Journey**

20 To understand well the SNO+ detector, it is worth thinking about how the information
21 of a physics event, e.g. a solar neutrino interaction, gets observed. This section follows
22 the journey of such an event.

2.3.1 Particle Interactions with Matter

All observable physics events within the detector begin by the generation of some form of ionising radiation: α , β^\pm , γ , p or n . These can be created via numerous processes, both exciting (e.g. $0\nu\beta\beta$ or interactions of neutrinos) and banal (e.g. decay of background radioisotopes): see Section 6.1.2 for some of them. Regardless of their origin, these particles begin propagating through the detector, and interacting with the detector medium. A number of mechanisms then allow for the generation of optical-wavelength light as a result of these interactions.

Cherenkov Light Emission

Whenever a charged particle passes through a dielectric medium, nearby molecules polarise and attempt to align their dipoles in the direction of the charge. This results in a temporary polarisation of the medium as the charge passes though it. However, the molecules can only respond to the charge's movement at the speed of light in the medium, which is necessarily slower than the speed of light in a vacuum by a factor of the medium's refractive index. If the charged particle is able to move at a speed greater than the speed of light in the medium, then the dipoles formed within the medium struggle to respond to the motion. This superluminal motion results in a wake of polarisation in the medium, which propagates outwards from the direction of the charged particle: this is Cherenkov light. This process is much akin to the ‘sonic boom’ that occurs when an object travels at supersonic speeds.

Cherenkov light emanates outwards in a cone along the direction of the charge's travel; the angle of the cone θ_γ is purely a function of the speed of the charged particle relative to the speed of light, β , and the refractive index of the medium $n(\lambda)$: $\cos \theta_\gamma(\lambda) = \frac{1}{n(\lambda)\beta}$. This formula also demonstrates that there is a minimum speed necessary for Cherenkov light to be generated: $\beta = 1/n(\lambda)$.

1 All detection media are capable of allowing for Cherenkov light to be generated, as
2 long as sufficiently high energy particles can be produced. In the water-fill phase of
3 the detector, Cherenkov light was the only means by which light could be generated.
4 Light from Cherenkov emission can still be created in liquid scintillator, but it tends
5 to be swamped by another form of light generation: scintillation.

6 **Scintillation**

7 Whenever a high-energy particle passes through a medium, interactions between the
8 particle and the medium's molecules lead to an inexorable transfer of energy from the
9 former to the latter. These interactions are numerous, and are a function of the type
10 of particle and its kinetic energy. For example, a 5 MeV electron loses energy from
11 electromagnetic interactions with nearby molecules as it passes through, and scatters
12 after colliding with atomic electrons or nuclei. From the perspective of the medium,
13 its molecules use this new-found energy for a number of activities:

- 14 1. Exciting atomic electrons into various higher-energy states;
- 15 2. Ionising atomic electrons;
- 16 3. Increasing the overall kinetic energy of the molecule.

17 At the macroscopic level, process 3 is simply the conversion of the high energy particle's
18 energy into heating the medium. In fact, within most materials the other processes
19 also end up doing the same.

20 However, for certain special classes of material the excitation and ionisation of
21 electrons can lead to a unique process: scintillation (often generally referred to as
22 'luminescence' or 'fluorescence'). Here we choose to focus on the particular case of
23 organic liquid scintillators, given that this is the type of scintillator which currently fills
24 the SNO+ detector. Certain organic compounds (i.e. molecules built from a skeleton

2.3 Detecting and Recording an Event in SNO+: A Journey15

Fig. 2.2

Fig. 2.3

of carbon atoms) have ‘ π -bonds’ in addition to the single ‘ σ -bond’ that is needed to bond two carbon atoms together. A major example of these π -bonds is in the benzene rings of ‘aromatic’ hydrocarbons, such as LAB and PPO. Fig. 2.2 shows an image of the orbitals for these π -bonds in a benzene ring: note how the electrons that inhabit these particular orbitals become entirely delocalised.

Because of this delocalised structure, excited atomic π -electrons can stay in what is typically the first-excited state for somewhat longer than typical excited states: lifetimes of $\mathcal{O}(10^{-9}\text{ s})$ as opposed to $\mathcal{O}(10^{-12}\text{ s})$. Moreover, decays from this state can emit light typically in the optical-wavelength range. It is this light emission that is called ‘scintillation light’. Ionised electrons can also recombine — that is, re-enter atomic orbitals — into various excited states, and then decay back to the ground state, also allowing for the possibility of scintillation light to be generated.

The nature of atomic spin selection rules restrict singlet ground-state electrons from exciting into triplet orbitals. However, electrons that were ionised and then recombined have no such restriction, and so are able to enter various excited triplet states. The same selection rules that restrict excitation from singlet into triplet states also works in the other direction, restricting the ability of electrons in excited triplet states from decaying back into the ground state. This does eventually happen, but the lifetimes of these “phosphorescence” decays are typically order of magnitude slower than their singlet-singlet counterparts. As a result, scintillation light typically has, to first order, a ‘fast’ and ‘slow’ time component, which can be seen in Fig. 2.3.¹

¹In SNO+, we currently model emission of scintillation light from LAB with 3 time components, each roughly one order of magnitude slower than the other.

Fig. 2.4

When using just a single scintillating compound, the very same energy levels that can generate scintillation light are those that can absorb it. This can be a problem for large-scale detectors like SNO+, which depend on scintillation light being unobstructed in its path to the PMTs. Conveniently, this problem can be addressed with the addition of another scintillating component, known (somewhat confusingly) as the primary fluor. In SNO+, this is the PPO added to the LAB.

When an LAB molecule is excited, that energy can be transferred to a PPO molecule through what is known as a ‘non-radiative transfer’. In short, this transfer of energy occurs not through the emission and absorption of optical photons, but through the coupling of the molecules’ electric dipoles.² When the now-excited PPO molecule de-excites to emit scintillation light, the different molecular structure it has generates a different emission spectrum to that of LAB. These longer wavelengths of light are no longer able to be absorbed by the LAB, allowing for a scintillator with less optical absorption.

Adding in one additional component doesn’t have to be the end, either. In SNO+ we plan on adding in the compound BisMSB to the scintillator cocktail. This is a ‘wavelength-shifter’: scintillation light at short wavelengths is absorbed, and then re-emitted at longer wavelengths, where the detection efficiency of the PMTs is greater. More on the properties of the PMTs in SNO+ can be found in Section 2.3.3. The net effect of the three scintillating components within SNO+ can be seen in Fig. 2.4. Note how, as energy is transferred from one scintillation component to another, the wavelength of light emitted gets necessarily longer as energy is lost to heat.

²To be pedantic, photons are still transferred in this energy exchange, but they are virtual instead of real.

The light yield of a scintillator, i.e. the amount of optical photons generated per unit of energy deposited into the scintillator, is a function not just of the scintillator but also the incident particle. In particular, α particles are far more effective at exciting and ionising nearby atoms, and so can deposit far more of its energy into the scintillator per unit volume. However, the strength of this ionisation for α s can actually become at detriment to the generation of scintillation light. Empirically, scintillators follow to first order Birks' Law for their scintillation light yield []:

$$\frac{dL}{dx} = S \frac{\frac{dE}{dx}}{1 + k_{\text{Birks}} \frac{dE}{dx}}, \quad (2.1)$$

where $\frac{dL}{dx}$ is the number of photons emitted per unit track length, $\frac{dE}{dx}$ is the energy loss of the incident particle per unit track length, S is the scintillator's characteristic light yield constant, and k_{Birks} is the scintillator's “Birks' Constant”³. For minimum-ionising particles such as a 6 MeV electron, the denominator of this equation is close to 1, and so the amount of scintillation light generated is just $\frac{dL}{dx} = S \cdot \frac{dE}{dx}$. However, for α -particles generated in radioactive decays, this denominator can become substantial, and in the limiting case we have merely $\frac{dL}{dx} = S$. In the current phase of SNO+, S and k_{Birks} are measured to be 14,000 γ/MeV and $0.0798 \text{ mm MeV}^{-1}$,⁴ respectively [].

2.3.2 Optical Processes

Once optical-wavelength photons have been created within the detector, various processes can then occur that can hinder its path towards a PMT, and therefore modify the observed signal. This subsection covers the main optical processes, with a fo-

³Birks' Constant is often just written as k_B , but this is easily confused with the far-better known Boltzmann Constant, which is completely different!

⁴This is the Birks' constant used in simulation for α -particles. Other particles, such as electrons, are given slightly different values of k_{Birks} .

¹ cus on Rayleigh scattering, as an understanding of this phenomenon is critical for
² Chapters 3–5.

³ **Rayleigh Scattering**

⁴ Optical scattering is the general process of how light is scattered by particles within a
⁵ medium. This is fundamentally an electrodynamical process: an electromagnetic wave
⁶ is incident on the set of particles within the medium, which induces these particles to
⁷ oscillate within the field, and therefore generating their own electromagnetic radiation
⁸ in response. Usually, this ‘scattered’ radiation has the same frequency as that of the
⁹ incident radiation, and therefore the scattering is said to be *elastic*. It is possible under
¹⁰ certain circumstances for this scattered radiation to be of a longer wavelength than
¹¹ the incident radiation: in which case, energy was absorbed by the particles and so the
¹² scattering was *inelastic*. However, this latter type of scattering, also known as Raman
¹³ scattering, is not relevant for SNO+ [].

¹⁴ The general solution to elastic optical scattering was first described by Gustav
¹⁵ Mie [] and Ludvig Lorenz [] in what is now known as *Mie Theory*. In this theory, it is
¹⁶ assumed that a plane wave of wavelength λ is incident on a dielectric sphere of radius a .
¹⁷ While the general solution to the problem of Mie scattering is somewhat complicated
¹⁸ (if tractable), in certain regimes one can make further simplifying assumptions that
¹⁹ substantially reduce the complexity of the result. In particular, if one assumes that
²⁰ the size of the particle is much smaller than the wavelength of light, and that any
²¹ induced dipole moment can actually be established in the time window allowed by the
²² oscillation period of the electromagnetic field [], then one can obtain *Rayleigh scattering*.
²³ This simpler case is so-called because of its initial formulation by Lord Rayleigh [].

2.3 Detecting and Recording an Event in SNO+: A Journey

19

One can show that the differential cross-section associated with Rayleigh scattering of unpolarised light off of a single particle, $\frac{d^2\sigma_{\text{Ray}}}{d\theta d\phi}(\theta, \phi)$, is given by []:

$$\frac{d^2\sigma_{\text{Ray}}}{d\theta d\phi}(\theta, \phi) = \frac{8\pi a^6}{\lambda^4} \left(\frac{n_{\text{par}}^2 - 1}{n_{\text{par}}^2 + 2} \right)^2 (1 + \cos^2 \theta). \quad (2.2)$$

Here, θ and ϕ correspond respectively to the polar and azimuthal angles of the scattered waves, and n_{par} is the refractive index of the scattering particle. Most important to notice about this equation is that the cross-section follows a strong $1/\lambda^4$ dependence, meaning that short wavelengths of light will be scattered to far greater extents than that of longer wavelengths. Secondly, the light is not scattered isotropically, but according to a $1 + \cos^2 \theta$ dependence. This means that most light is either scattered directly forwards or backwards (known as a *backscattering*), and little gets scattered orthogonally to the direction of the incident light. This is useful when it comes to trying to measure scattering in the SNO+ detector, as it provides a handle upon which to distinguish scattered light from isotropically-emitted scintillation light.

Of course, we care about the scattering that occurs within an entire bulk medium, not just the scattering off of a single molecule. From a macroscopic perspective, the key quantity of interest is a material's *Rayleigh scattering length*, l_{Ray} : the mean distance a photon is expected to travel before Rayleigh scattering. One can show that, assuming the above differential scattering cross-section, the Rayleigh scattering length is given by []:

$$l_{\text{Ray}} = \left[\frac{16\pi}{3} R \right]^{-1}. \quad (2.3)$$

R is the *Rayleigh ratio*, $R = \frac{1}{V} \frac{d^2\sigma_{\text{Ray}}(90^\circ)}{d\theta d\phi}$, equivalent to the power of the scattered light per unit volume of the scattering medium per unit incident intensity at $\theta = 90^\circ$.

This can lead to a few changes to Rayleigh scattering that are worth noting. Firstly, unlike for a single particle, the electric polarisability of a material can be

¹ *anisotropic.* Anisotropic materials have a modified angular dependence on their
² differential cross-section, governed by the *depolarisation ratio*, δ . In particular, the
³ $(1 + \cos^2 \theta)$ dependence becomes $(1 + \frac{1-\delta}{1+\delta} \cos^2 \theta)$. For isotropic materials, $\delta = 0$, and
⁴ so the angular dependence reduces to the original form.

⁵ Secondly, the above model has been shown to be insufficient to describe liquids
⁶ or solids [], because of the non-negligible strength of their inter-molecular forces.
⁷ Fortunately, Einstein [], Smoluchowski [], and Cabannes [] developed a theory for
⁸ describing how photons can scatter off of the local charge density fluctuations that
⁹ naturally are present in a medium because of the thermal motion of molecules. The
¹⁰ theory shows that the Rayleigh ratio of a medium is related to the medium's dielectric
¹¹ constant, ε , by:

$$R = \frac{\pi^2}{2\lambda^4} \left[\rho \left(\frac{\partial \varepsilon}{\partial \rho} \right)_T \right]^2 k_B T \kappa_T \frac{6 + 6\delta}{6 - 7\delta}, \quad (2.4)$$

¹² where ρ is the density of the medium, $\left(\frac{\partial \varepsilon}{\partial \rho} \right)_T$ is the partial derivative of the dielectric
¹³ constant with respect to a changing density assuming a constant temperature T , k_B
¹⁴ is the Boltzmann Constant, and κ_T is the medium's isothermal compressibility. This
¹⁵ latter quantity is given by the rate of change of volume given a changing pressure of
¹⁶ the medium, all at a constant temperature.
¹⁷

¹⁸ Furthermore, the Eikman Equation [] has been shown to be

- ¹⁹ • Explain the electrodynamical model for scattering, giving rise to the expected
²⁰ $1/\lambda^4$ wavelength-dependence of the scattering length, as well as the $1 + \cos^2 \theta$
²¹ dependence of the scattering angle.
- ²² • Describe how the density-fluctuation theory gives rise to a possible modification
²³ to the scattering angle distribution due to the anisotropy of the optical media's
²⁴ polarisability vector. Studies by JUNO have shown that LABPPO has such a
²⁵ measurable anisotropy.

2.3 Detecting and Recording an Event in SNO+: A Journey

21

- Show the existing model for water and LABPPO used in RAT, and note that this anisotropy is currently not included in any simulations. I don't need to go in too much detail for this subsection as Krish did a nice job in his thesis and I can cite that, but I do need to write enough to cover the basics for my SMELLIE analysis chapter.

[3 pages]

6

Absorption and Re-emission

7

- State that light can get absorbed by materials, and if that medium is a scintillator then re-emission is possible. I think further details such as the specific shape of the absorption/re-emission of the scintillator and water can be shown in the SMELLIE analysis chapter, in which I have to explain about possible changes to the model anyway.

[1/2 page]

13

Surface reflection and refraction

14

- State that boundaries between media can induce reflections and refraction, as governed by the Fresnel transmission and reflection formulae. These formulae are worth mentioning because I use them in my SMELLIE extinction length analysis.

[1/2 page]

18

2.3.3 Detection by PMTs

19

- Light gets detected via the PMTs. Note the existence of the PMT concentrators to maximise coverage within the AV, but minimise it in the external water: show the calibrated angular response.

22

- 1 • Remind reader that a PMT converts photons to photoelectrons with a certain
- 2 quantum efficiency, dependent on the wavelength of light.
- 3 • The process of multiplying the signal induces a spread in the possible generated
- 4 time of the voltage signal, known as the PMT's transit time spread.
- 5 • The PMT response is also weakly dependent on the number of photoelectrons
- 6 generated; not enough to be able to confidently distinguish the npe under most
- 7 circumstances.

8 [2 pages]

9 **2.3.4 Data Acquisition and Triggering**

- 10 • Summarise how a PMT signal becomes digitised via the front-end electronics,
- 11 briefly.
- 12 • Summarise the triggering system, as I'll have to describe in a later chapter how
- 13 the SMELLIE hardware fits into this, especially as there have been ongoing
- 14 SMELLIE triggering issues worth mentioning there!
- 15 • Note the information stored by an event: importantly for this thesis, the TAC
- 16 and QHS per hit, the event's GTID as well as trigger time measured from the
- 17 50 MHz clock. The latter is worth mentioning as this is how I determine time
- 18 differences for my BiPo tagging in the solar analysis.
- 19 • Finally, note that these get written to file in the ZDAB format.

20 [3 pages]

2.3.5 Operation of the Detector

- Detector electronics operated through ORCA; allows for different running modes, such as calibrations.
- Mention that data gets split into run and subruns.

[1 page]

2.4 Calibrations and Detector Modelling**2.4.1 Detector Monitoring**

- Detector's state is continuously monitored via a number of systems for data quality purposes, including a human detector 'shifter'. This includes the alarm systems for the electronics and slow controls, and 'nearline' monitoring of the detector status.
- CHS and CSS ensure only "good" channels used in any analysis.

[1 page]

2.4.2 Electronic and PMT Calibrations

- First main set of calibrations are the ECAs and PCAs. These help us convert raw electronic signal information from PMTs into 'calibrated' hit times and charges. PCAs performed via TELLIE and the Laserball.
- These initial calibrations allow for the first two passes of data processing, resulting in a 'RATDS' file format used in optical calibrations.
- No need for many details here; this is not a thesis on this topic!

[1 page]

1 2.4.3 Energy and Optical Calibrations

- 2 • After ECAs and PCAs, further calibration is necessary to accurately model the
3 optical properties of the detector.
- 4 • Describe briefly the function of the Laserball, SMELLIE, and AMELLIE sources
5 in optical calibration (can reference the water-phase optical calibration paper).
6 Details of how each analysis works is obviously not needed here, especially not
7 for SMELLIE as we have 3 chapters to go over that!
- 8 • AmBe and N16 sources provide further information for calibration
- 9 • In-situ backgrounds used for more calibration sources, especially for determining
10 the light yield of the scintillator, and Birks' constant.

11 [2 pages]

12 2.4.4 Event Reconstruction

- 13 • Once detector is calibrated, event reconstruction becomes possible. Describe
14 main assumption of a SNO+ event: we assume a single-site electron event.
- 15 • Briefly describe the basics of how energy, position, and time reconstruction works,
16 as this is needed for the solar analysis chapter. Mention existence of direction
17 fitting in scintillator!
- 18 • ‘Physics’ runs have their events reconstructed in a third pass of data processing,
19 resulting in a ‘fully-processed’ RATDS file as well as a simplified NTUPLE file
20 used in high-level physics analyses (such as my solar analysis).

21 [2 pages]

2.4.5 Event Simulation

- Briefly describe overview of RAT software: not only provides the software by which the above data processing occurs, but allows for a GEANT4-based simulation of events in the detector. This simulation includes all parts of the physics described in Section 2.3, including particle and nuclear physics interactions, optical photon creation and propagation, signal generation and event building based on simulated triggers of the electronics. These simulated events then get processed in the same manner as actual data is.

[1 page] [24 PAGES TOTAL]

Chapter 3

The SMELLIE Calibration System

There's a certain Slant of light,

Winter Afternoons —

That oppresses, like the Heft

Of Cathedral Tunes —

EMILY DICKINSON

- Basic principle for how SMELLIE works: firing collimated laser light into detector to observe scattering events.
- Analysis will measure and monitor scattering in a detector with changing optics.
- One can try and measure some component of this: the cross-section/scattering length versus wavelength and time, and/or the relative scattering length versus wavelength and time.

[1 page]

3.1 The SMELLIE Hardware

- Describe the existing hardware, post-upgrade made in Summer 2022. For pre-upgrade hardware, can simply cite previous SMELLIE theses. This includes the path of light into the detector, as well as the path of the trigger signal.
- Make sure to mention explicitly these upgrades: Tony Zummo's fix to the TUBii trigger logic, as well as the addition of the VFA, updated MPU, and modified trigger window. Make sure to motivate why these updates were made.

[7 pages]

3.2 Software for SMELLIE Data-taking

- Can be brief here! Little has changed since previous theses, so can mostly just summarise and cite.
- Server running on SNODROP machine, which converts high-level commands into low-level ones that the hardware can interpret.
- Run plan files written in JSON handed to ORCA which then sends relevant commands to SNODROP which fires as appropriate.
- Operator interacts with ORCA to perform SMELLIE calibration runs.
- After SMELLIE data taken, run description file created, containing metadata about the run conditions, used in analysis.

[2 pages]

3.3 Commissioning SMELLIE in the

Scintillator Phase

- Explain why commissioning of SMELLIE is needed: Need to confirm that SMELLIE is working as expected; determine intensity "set-points" for different use cases.
- Commissioning originally performed by Esther and JeffL back in the water phase; explain why this needed to be re-done for both the scintillator phase and after the hardware upgrades.
- No need to describe the Tesseract in detail here - that can be in Jeff L's thesis. But, I do want to show the results of both commissioning campaigns in scintillator-fill, one before the new hardware was added, and one after.

[5 pages] [15 PAGES TOTAL]

Chapter 4

1

Simulating SMELLIE Events

2

Max Power : *Kids. From now on there are three ways of doing things: the right way, the wrong way, and the Max Power way.*

Bart Simpson : *Isn't that just the wrong way?*

Max Power : *Yes, but faster!*

THE SIMPSONS

Critical to extraction of scattering information from SMELLIE data is an accurate Monte Carlo (MC) simulation of the SMELLIE system. By modelling the laser light emission into the detector correctly, we can simulate how SMELLIE light will be impacted by changing scattering lengths in the detector. Because of the complexity of the optics of the optical fibres used to direct the laser light into the detector, a given SMELLIE event is simulated as a partially-collimated “flash” of visible photons emanating from the emission point of the fibre into the detector. This flash then requires a number of parameters to be correctly described. In particular:

4

5

6

7

8

9

10

11

- **Fibre emission positions** were recorded during the installation of the fibres.

12

- **Wavelength and emission timing distributions** of light pulses were taken from measurements of the laser heads by their manufacturers [], or by colleague Jeff Lidgard in the case of the SuperK wavelength distribution [].
- **The “pulse magnitude”**, defined as the mean number of photons simulated per event, is determined on a subrun-by-subrun basis, and is assumed to fluctuate as a Poisson distribution.
- **The beam profiles**, which describe the angular emission distributions of each fibre, is the focus of this chapter. These are necessary because unlike scintillation light, the light emitted from SMELLIE fibres is not isotropic.
- **Nominal fibre emission directions** attempt to define the centre of the beam for a given fibre.

This chapter is split into three sections. Improvements to the existing simulation algorithm for the beam profiles are first made, and then the beam profiles themselves are updated. Finally, comparisons between data and simulation are made after the upgrades to investigate any remaining discrepancies.

4.1 Improving the SMELLIE Generator Algorithm

4.1.1 Previous Attempts at SMELLIE Event Simulation

Before we can determine the beam profiles, we must first decide how to specify them. Previous observations show that different fibres can have notably different beam profiles [2], so we let each fibre’s beam profiles be unique. We assume for now that a given fibre’s beam profile is stable over time, and independent of the wavelength of light fired. A straightforward, naïve approach to parameterising a beam profile would be as follows: specify some nominal fibre direction, corresponding to the direction light

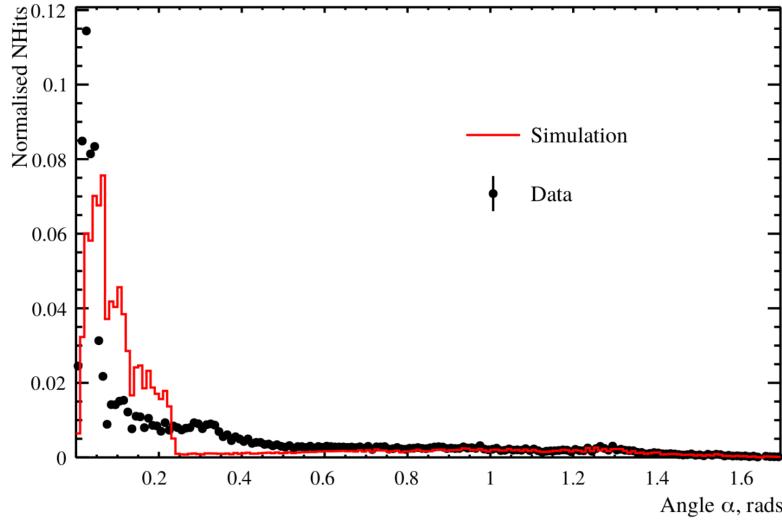


Fig. 4.1 Comparison between a simulation of one of the fibres, made from the 1D beam profile generator (red), with the associated data subrun that was used to create that beam profile (in black). For both MC and data, what is plotted is the PDF of observed PMT hits, as a function of the α angle. Poissonian errors have been added to the data points, but are too small to see. Clearly, this 1D generator does not replicate the observed beam profile correctly. Figure taken from [3].

takes travelling from the fibre to the centre of the “beamspot” observed on the other
1 side of the detector. Then, specify a 1D beam profile, corresponding to the probability
2 density of firing a photon at a given polar angle α relative to the nominal direction.
3 One might even assume this distribution is Gaussian. The distribution in azimuthal
4 direction, ϕ , is assumed to be uniform.
5

This 1D beam profile approach was used initially for SMELLIE, and remains in
6 use for the other ELLIE sub-systems within SNO+. However, when SMELLIE data
7 was taken in the water-phase of the experiment, simulations using these beam profiles
8 failed to match them well at all - see figure 4.1 for an example. Not only was the
9 distribution in α not Gaussian, a distinct speckle-pattern can be observed within the
10 beamspot that is not uniform in ϕ . This fact led to colleague Esther Turner building a
11 SMELLIE generator that could handle 2D beam profiles: dependent on both α and ϕ .
12 The distribution was stored as a map from each inward-pointing PMT in the detector
13

¹ to a relative intensity value. This was chosen because the beam profile shapes were
² calibrated from existing SMELLIE data — more on this in section 4.2.

³ This original 2D generator then sampled the beam profile via a rejection sampling
⁴ approach, outlined as follows:

- ⁵ 1. Propose a test direction (α, ϕ) , by generating ϕ uniformly in the interval $[0, 2\pi]$,
⁶ and α according to some pre-determined Gaussian distribution, known as the
⁷ Gaussian envelope.
- ⁸ 2. Given this test direction, calculate where a line following this direction from the
⁹ fibre of interest will hit the PSUP on the other side of the detector. Find the 3
¹⁰ closest PMTs to that point.
- ¹¹ 3. From those PMTs, obtain their relative intensity values from the beam profile
¹² mapping, and perform an interpolation based on how close each PMT is to the
¹³ PSUP intersection point. This gives an interpolated relative intensity value for
¹⁴ this test direction.
- ¹⁵ 4. Because we are sampling using the angular coordinates (α, ϕ) , differential area
¹⁶ elements over this space of directions do not have the same size. We can correct
¹⁷ for this fact by multiplying our interpolated relative intensity by $\sin \alpha$, which
¹⁸ corresponds to the Jacobian of the direction-space.
- ¹⁹ 5. Calculate the value for the Gaussian envelope along this test direction.
- ²⁰ 6. Throw a random number uniformly between 0 and the Gaussian envelope value. If
²¹ the random number is less than the interpolated intensity, then this test direction
²² is accepted, and a photon is generated with that direction. Otherwise, we reject
²³ the direction and try the whole process again.

This generator certainly works, but has a key problem: efficiency. The 1D generator was able to generate a SMELLIE event (that is, to fully specify the starting parameters of all the photons emitted from a fibre) at a speed of ~ 1 ms. However, the 2D generator specified here could take upwards of ~ 50 s *per event* to generate. Because a typical SMELLIE analysis requires simulating many millions of events, the CPU time taken to perform this quickly became unfeasible. Fixing this generator speed problem was a high priority for the SMELLIE analysis.

4.1.2 The new generator

On careful inspection of the existing 2D generator, the main reason for the slowness of the algorithm is the use of a rejection approach. Even with use of the Gaussian envelope, which was included to help with speed, the vast majority of proposed directions are never selected. Figure 4.2 shows a histogram of number of attempts per event it took for a valid direction to be chosen for a representative SMELLIE simulation. Moreover, the calculations needing to be done for every proposed direction are relatively complex, notably trying to find the 3 nearest PMTs to some point on the PSUP.

A new 2D generator was built with these thoughts in mind. Firstly, the rejection method would no longer be used, given its inefficiency. We would also endeavour to try and “pre-calculate” as much as possible before run-time. Starting with the existing PMT relative intensity maps, we plot these in the 2D direction-space ($1 - \cos \alpha, \phi$): see Figure 4.3a. In a toy-MC simulation, 500,000 directions are then thrown uniformly in this 2D space per fibre. For each direction, the same method of obtaining an interpolated intensity value from the nearest PMTs to the corresponding point on the PSUP as from the original 2D generator was performed, the only difference being that these calculations were done well before any actual SMELLIE simulation. Figure 4.3b shows the interpolated intensities obtained for one fibre.

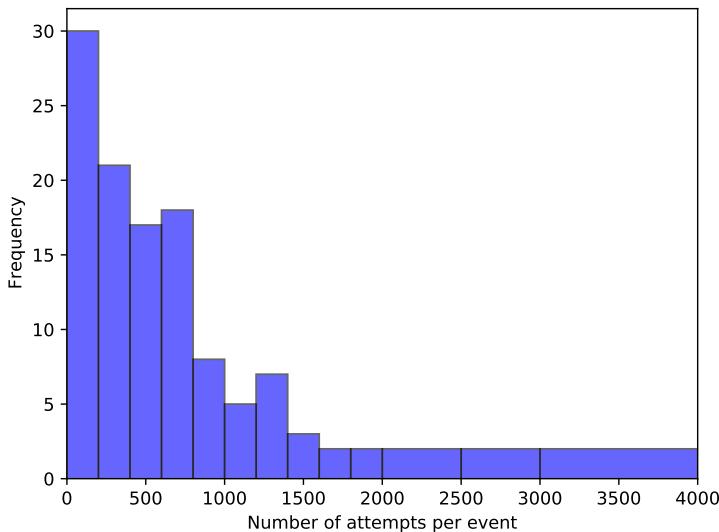


Fig. 4.2 Typical distribution of the number of attempts it takes for the existing 2D generator before the test direction gets accepted, per event.

Following this, the sampled intensities were then binned into a 2D histogram, where the bin value corresponds to the sum of all intensities for all directions found within this bin. Choosing a sensible binning procedure is important: too few bins, and necessary information about the shape of the beam is lost, whilst too many bins can oversample the data and capture statistical artefacts in the sampling process instead of just the beam profile. As a balance, 15 bins were chosen along the ϕ direction, and 60 in $r = 1 - \cos \alpha$. This was chosen to ensure that a reasonable number of PMTs were located within each bin, lessening the impact of any statistical fluctuations. Although the bins in ϕ were chosen to have uniform width, this was decided to be not the case for the other axis, as there is far more important information near $r = 0$ (the beamspot). Instead, the width of the bins in r were calculated so that roughly the same total probability was contained in each r -strip. By consequence, bins near the beamspot typically are of significantly smaller size than ones much further out. This allows us to both capture any rapid changes in intensity near the beamspot, where this

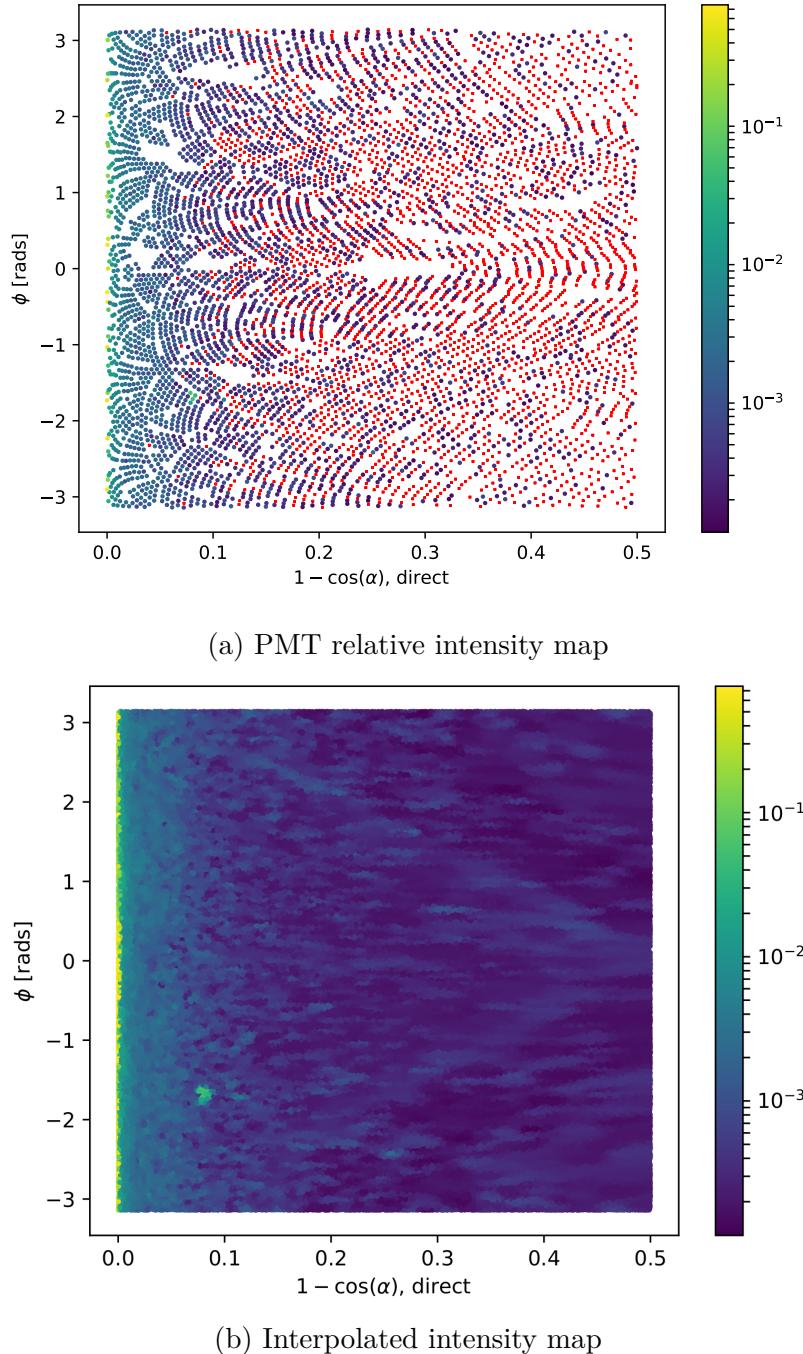


Fig. 4.3 The first step in the new method for preparing the new generator. In (a), the relative intensities used for the existing beam profile of fibre labelled FS055 are shown for each PMT, the position on the plot indicating the location of that PMT in the fibre coordinates. The colour indicates the relative intensity; PMTs marked red have an intensity of zero. Figure (b) shows the result of throwing 500,000 directions uniformly over this 2D space, the intensity of each point given by interpolating the intensities of nearby PMTs.

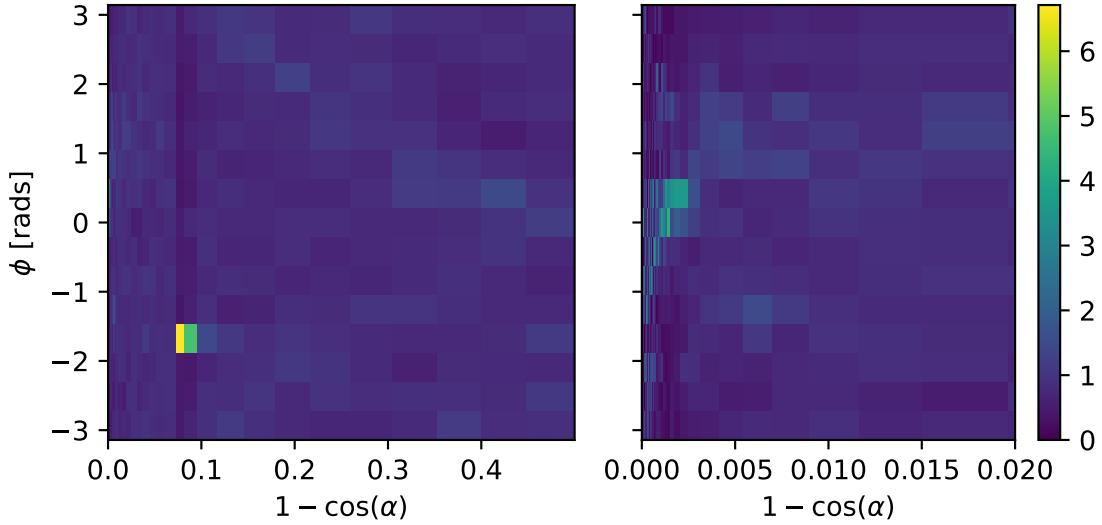


Fig. 4.4 Histogram of interpolated intensities within the 2D direction-space. The left view shows the full histogram; the right is a zoomed-in version near the beamspot. Unlike the binning in ϕ , the bin widths in r are not at all uniform. Instead, they have been determined such that the area summed over a given “strip” of bins of constant r will be the same.

¹ matters greatly, and smooth out the very-low intensities seen at larger polar angles.

² One of these histograms can be seen in Figure 4.4: the large change in bin widths as a
³ function of r is clear. One can also see that near the beamspot notable dependence on
⁴ the intensity as a function of ϕ . The mysterious “spot” at $r = 0.08$, well out of the
⁵ beamspot, is an indication that the underlying beam profile data being used requires
⁶ improvement: more on this in section 4.2.

⁷ The Cumulative Density Function (CDF) of this intensity histogram as a function
⁸ of bin was then produced, where the bins were ordered through a raster-scan: scanning
⁹ first over ϕ , and then r . The CDF was then normalised to 1 so that it was well-defined.

¹⁰ It is this CDF object that is then loaded in and sampled from during event generation.

¹¹ To do this, an “inverse-CDF” approach was used, which has the major benefit over
¹² rejection sampling of always producing a valid direction for every sample made. The
¹³ algorithm works as follows:

4.2 Improving the beam profiles

39

1. Throw a random number uniformly in $[0, 1]$. 1
2. Perform a binary search to find the bin that has the largest CDF value below this random number. 2
3. Look at the bin edges in ϕ of this selected bin: use linear interpolation of the random number to obtain a ϕ value located between these two ϕ -values. 3
4. Look at the selected bin's r -bin edges, and select a value of r by throwing a second random number uniformly between the two edges. Convert this r into a polar angle α . 4
5. The photon's direction is defined by the (α, ϕ) chosen by this process. 5

Because of the relative simplicity of this algorithm compared to the previous 2D generator, the speed improvement was very large: generation now took ~ 1 ms per SMELLIE event, a speed improvement of nearly 50,000. Event generation became as fast as it was when the 1D generator was being used. Furthermore, because of the approach taken, this major speed improvement comes at no sacrifice in accuracy. Figure 4.5 shows a comparison of the average number of photoelectrons (npe) per event per PMT between water-phase SMELLIE data and simulations with both the old and new 2D generator. One can see clearly that both generators are as accurate as one another. Note that this plot uses the updated beam profiles as explained in the next section.

4.2 Improving the beam profiles

Even with the new 2D profile generator, a problem remains: the simulation fails to reasonably recreate data, and much of this appears to be because of the poor beam profile data being used. The curious “spot” for one of the fibres was already noted in

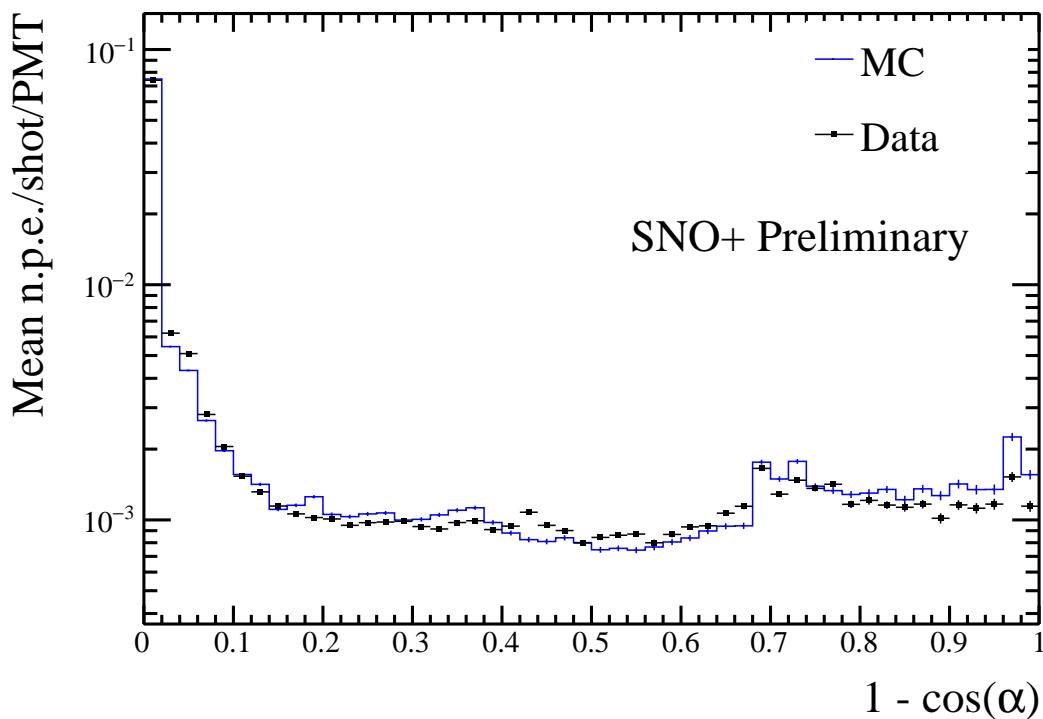


Fig. 4.5 Comparison of water-phase data to MC generated using both the old and new 2D beam profile generator approaches, with the updated beam profiles. Both versions of the generator are consistent with one another, but the new generator is many times faster.

4.2 Improving the beam profiles

41

the previous section that doesn't seem to be physical, and more broadly at large angles for all the fibres there are large swathes of PMTs with an intensity of zero, providing little useful information about the beam shape. It was shown in [3] that with the old 2D generator, the systematic uncertainty on the beam profiles was the dominant source of error in the main SMELLIE analysis. To help improve this situation, it was decided to update the existing beam profiles.

These old beam profiles were originally determined by looking at SMELLIE data taken during the water-phase. Specifically, a “medium”-intensity subrun with one of the lasers firing at a wavelength of 495 nm, was chosen for each fibre. “Medium”-intensity corresponds to firing the relevant laser at a set intensity determined during an earlier commissioning process, for which the maximum occupancy of PMT hits at that intensity, i.e. the proportion of hits per event, corresponded to roughly 80%. This value was chosen as it allowed for high statistics in a relatively short run-time, but not so intense that the occupancy of any given PMT in the beamspot was 100%. Because Rayleigh scattering is strongly-dependent on wavelength, the long wavelength of light was chosen so that impacts from this scattering were small in the data.

SNO+ PMTs are unable to distinguish the exact number of photoelectrons being generated. One is typically only able to know if a PMT has been triggered at all, by any number of photoelectrons. As a result, the occupancy of a PMT over a number of SMELLIE events, o , is a biased estimator of the mean number of photoelectrons generated, μ . Assuming the number of photoelectrons generated in a given event follows Poisson statistics, the probability of generating k photoelectrons is:

$$P(k|\mu) = \frac{\mu^k e^{-\mu}}{k!}. \quad (4.1)$$

The probability of observing a “hit” in a given PMT corresponds to generating at least one photoelectron:

$$P(\text{hit}|\mu) = P(k \geq 1|\mu) = 1 - P(k = 0|\mu) = 1 - e^{-\mu}, \quad (4.2)$$

which implies after rearrangement that one can determine the mean number of photoelectrons per event from the occupancy by:

$$\mu = \ln(1 - o). \quad (4.3)$$

This is the reason why we want to avoid PMTs with occupancies of 100%: they preclude one's ability to convert into a value for μ by looking at occupancy alone. We call this conversion from occupancy into npe the “multi-hit correction”. The impact of this correction is typically small for most PMTs, but can become very significant in a fibre's beamspot.

Once the npe mapping from data was obtained, a correction was then made for the detector's optics: even ignoring a fibre's beam profile, we still expect certain PMTs to be illuminated more than others because of e.g. reflections off the AV, or the solid angle subtended by the PMT bucket opening. For each fibre, a simulation was made where the beam profile was set as isotropic, and the corresponding npe mapping obtained: this map held information about the detector optics only. The beam profile mapping was then derived by simply dividing each fibre's npe mapping from data to its associated isotropic MC npe map. It is these maps that were first used in section 4.1.2.

4.2.1 Combining beam profile datasets

Fortunately, much more SMELLIE data was taken during the water-phase than was used for the original beam-profiling analysis. This additional data can be combined with that which was already used to far better constrain the beam profiles. In particular, given the existing assumption that scattering effects are minimal above wavelengths

4.2 Improving the beam profiles

43

Run Number	Run Type	Comments
114,018	All PQ lasers; SuperK laser in 400–500 nm range	Only PQ495 laser and SuperK at 495 nm is used
114,023	SuperK laser in 500–600 nm range	Part 1 of this wavelength range; crash occurred on last subrun, so that subrun is ignored
114,034	SuperK laser in 500–600 nm range	Part 2 of this wavelength range

Table 4.1 Water-phase runs used for new beam profiling.

of ~ 490 nm, all data taken with wavelengths above this can also be used. The specific runs (and associated comments about their specifics) are described in Table 4.1. Because high-intensity runs require a different analysis approach (PMTs with high occupancies must use charge, not occupancy, to estimate npe), for this analysis we only considered subruns that used low or medium intensity set-points.

For each subrun j of data per fibre, we look only at PMT hits for each PMT i that has been identified as “good” for that subrun¹, $i \in G_j$. G_j here represents the set of good PMTs in subrun j . In particular, a “good” PMT must have valid electronic and timing calibrations, be at high voltage and masked into the detector’s trigger system for that subrun. In addition, an angular cut of $\alpha < 60^\circ$ was made to remove PMTs that are well outside any reasonable beam direction. To isolates the hits arriving directly from the fibre without reflecting, scattering, or being noise, a time cut was also made. Because what matters is the time relative to emission from the fibre, and the expected time-of-flight from fibre to different PMTs varies, a quantity known as the time residual was used. Starting with the calibrated hit time of a given PMT relative to the event’s trigger time, t_{hit} , the expected time-of-flight t_{TOF} from the fibre to the PMT was subtracted, estimated with the collaboration’s “Light Path Calculator”. Then, the emission time was also subtracted, t_{emm} , estimated by looking at the second-earliest value of $t_{hit} - t_{TOF}$ within the fibre’s central beamspot, defined as the PMTs for which

¹Strictly speaking, a PMT’s “goodness” is only determined on a run-by-run, not a subrun-by-subrun level, but this has no impact on the analysis.

¹ $\alpha < 3^\circ$. It was found that a “loose” time residual cut of $t_{res} \in [-10, +12]\text{ns}$ was
² sufficient to remove the vast majority of non-direct light with little signal sacrifice. In
³ the situation where a subrun with intensity was very small, it would not regularly have
⁴ at least two hits in the beamspot, and so the time residuals calculated would not be
⁵ valid for many events. To avoid this situation, a cut was made on any subruns with
⁶ mean intensities below 9 within their beamspot. This value was chosen as it would
⁷ mean a 2σ fluctuation downwards of $2 \cdot \sqrt{9} = 2 \cdot 3 = 6 \text{ npe}$ would still have more than
⁸ the 2 hits necessary for timing reconstruction. One fibre, FS207, has no data subruns
⁹ that satisfy this condition, and as such will have to be dealt with separately. For the
¹⁰ time being, this fibre was ignored.

¹¹ Extracting the underlying beam profiles from these data required some careful
¹² thought, especially because different subruns could have different intensities. Consider-
¹³ ing a PMT i in subrun j , the mean number of photoelectrons generated per event in
¹⁴ that PMT for that subrun, μ_{ij} can be decomposed as follows:

$$\mu_{ij} = I_j k_i = I_j b_i f_i. \quad (4.4)$$

¹⁵ I_j is the intensity of the subrun, i.e. the mean number of photons generated from the
¹⁶ fibre in that subrun per event. k_i is the probability that a given photon generated at
¹⁷ the fibre source ends up generating a photoelectron in PMT i . This itself can be further
¹⁸ split into two components: b_i , the probability that a given photon at the fibre source
¹⁹ points in the direction of PMT i ; and f_i , the probability that a given correctly-pointed
²⁰ photon actually makes it to the PMT and successfully generates a photoelectron. It is
²¹ b_i that is the actual beam profile we would like to measure.
²²

²³ Letting p_{ij} be the probability of observing a hit for a given event on a given
²⁴ PMT, the probability of observing m_{ij} hits out of N_j events in the subrun will be

4.2 Improving the beam profiles**45**

binomially-distributed:

$$P(m_{ij}|\mu_{ij}) = L(\mu_{ij}|m_{ij}) = \binom{N_j}{m_{ij}} p_{ij}^{m_{ij}} (1-p_{ij})^{N_j-m_{ij}} = \binom{N_j}{m_{ij}} (1-e^{-\mu_{ij}})^{m_{ij}} e^{-\mu_{ij}(N_j-m_{ij})}. \quad \blacksquare$$
1
2
3
4
5
6
7

Here we have used equation 4.2, and noted that this probability distribution in m can be re-framed as a likelihood function for the parameter μ_{ij} . Considering only a single subrun of data, the maximum likelihood estimate of the parameter μ_{ij} can be shown to be:

$$\langle \mu_{ij} \rangle = -\ln \left(1 - \frac{m_{ij}}{N_j} \right) = \ln (1 - o_{ij}) \quad (m_{ij} \neq N_j), \quad \blacksquare$$
8
9
10
11
12

where o_{ij} is just the occupancy of PMT i in subrun j . This is just the multi-hit correction formula seen in equation 4.3, which makes sense.

When looking at multiple subruns for the same fibre, the total likelihood function for a given PMT when considering all the data for a given fibre will be the product of the likelihoods from each dataset,

$$L(\{I_j\}, k_i | \{m_{ij}\}) = \prod_j L(I_j, k_i | m_{ij}) = \prod_j \binom{N_j}{m_{ij}} (1 - e^{-I_j k_i})^{m_{ij}} e^{-I_j k_i (N_j - m_{ij})}. \quad \blacksquare$$
13

This leads to a log-likelihood distribution of

$$\mathcal{L}(\{I_j\}, k_i | \{m_{ij}\}) = \sum_j \left[\ln \left(\binom{N_j}{m_{ij}} C_{m_{ij}} \right) + m_{ij} \ln \left(1 - e^{-I_j k_i} \right) - I_j k_i (N_j - m_{ij}) \right]. \quad \blacksquare$$
14
15

Formally, one could combine the likelihoods of all the PMTs together, and by looking at the maximum likelihood estimates for each of the parameters measure the parameter values this way. However, the set of equations one obtains through this approach quickly become analytically intractable, because the PMTs are coupled by the intensity values I_j . Even a direct numerical approach would be liable to fail: for a given fibre

1

2

3

4

5

6

7

8

9

10

11

12

14

15

16

17

18

19

20

¹ there can be dozens of subruns, and many thousands of PMTs of relevance, so the
² dimensionality of the system of equations would be far too large.

³ Because of this, a different approach was taken. It is expected that in a subrun the
⁴ total npe, summed over all good PMTs, should be proportional to the intensity value
⁵ I_j . One must be careful about this construction — different subruns can have different
⁶ sets of good PMTs, so two subruns with identical I_j values could have a larger summed
⁷ npe merely because more PMTs were good in that subrun. To counter-act this effect,
⁸ only PMTs that were classified as good in *all* subruns being analysed for that fibre
⁹ would be used for the npe summation. In other words, we use data from PMT i for
¹⁰ summing only if:

$$\text{11} \quad i \in \mathcal{I} = \bigcap_j G_j. \quad (4.9)$$

¹² We can then define the summed npe for a given subrun as $S_j = \sum_{i \in \mathcal{I}} \text{npe}_{ij}$, and assert
¹³ that $I_j = cS_j$. By finding a value proportional to I_j , there is now enough information
¹⁴ to maximise the log-likelihood $\mathcal{L}(k_i | \{m_{ij}\}, \{I_j\})$ with respect to k_i for each PMT
¹⁵ independently, and hence obtain estimates for these k_i parameters.

¹⁶ Of course, what is actually wanted are the underlying b_i values, not k_i . This is
¹⁷ where isotropic simulations come in. For each run of data used, a matching isotropic
¹⁸ MC was produced. As an example, a simulation for run 114,023 contained 200,000
¹⁹ events for each fibre using an isotropic beam profile, over the full wavelength range
²⁰ considered in this run, 500–600 nm, using the same run conditions as in data (which
²¹ PMTs were at high voltage, etc.).

²² For each isotropic MC run, both I_j^{MC} and k_i^{MC} were calculated via the method
²³ described above. Because the simulations were isotropic, the underlying value for b_i
²⁴ was constant across all the PMTs, and so $ak_i^{MC} = f_i$. By doing some rearranging of

4.2 Improving the beam profiles

47

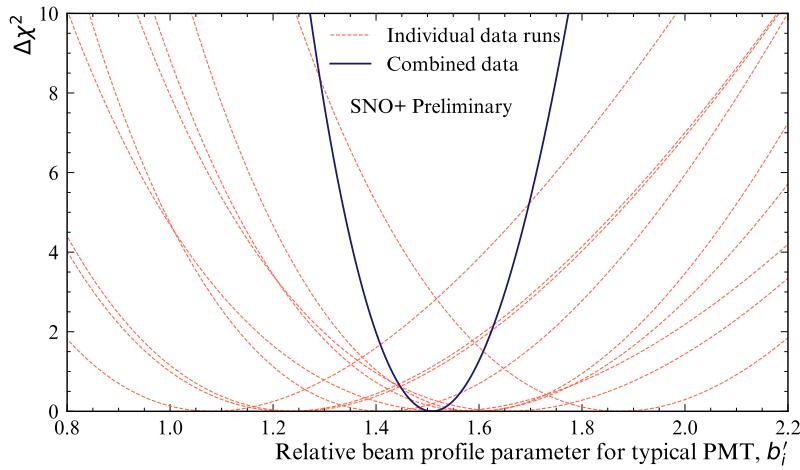


Fig. 4.6 Plot of $\Delta\chi^2 \simeq X_i$, twice the negative log-likelihood ratio, for both single subruns of a typical PMT, and when all relevant subruns are combined.

equation 4.4, we find that:

$$\mu_{ij} = I_j b_i f_i = c S_j b_i a k_i^{MC} = (acb_i) S_j k_i^{MC}. \quad (4.10)$$

As a result of this, given the set $\{S_j\}$ and k_i^{MC} , one can maximise the log-likelihood \mathcal{L} with respect to $b'_i = acb_i$ numerically, to obtain the maximum likelihood estimate of b'_i . Because a and c were global constants of proportionality, they would become irrelevant as soon as the beam profile was normalised in the CDF-creation process outlined in 4.1.2.

Figure 4.6 shows the shape of this log-likelihood distribution for a particular PMT when considering fibre FS007's beam profile. One can see how individual subruns provide much more information when combined than if one looked at a single subrun alone.

Another benefit of using this log-likelihood approach is that the resulting distribution's shape can be used for uncertainty estimation. In almost all cases, Wilks Theorem [4] allows us to produce 1σ confidence intervals about the maximum likelihood

estimate for b'_i , $\langle b'_i \rangle$, because

$$X(b'_i) = -2 [\mathcal{L}(b'_i) - \mathcal{L}(\langle b'_i \rangle)]$$

¹ approximates a χ^2 -distribution. As a result, the error bounds on our parameter estimate
² are given by when $X = 1$. The fact that the shape of X can be well-approximated by
³ a quadratic in the region near $X = 0$ indicates the validity of Wilks' Theorem being
⁴ used here.

⁵ Only a couple of exceptions to this approach of parameter estimation are possible.
⁶ In the case where $m_{ij} = N_j$, i.e. a PMT has 100% occupancy, no maximum likelihood
⁷ estimate exists: we need not worry about this, as subruns where this occurs have not
⁸ been used. On the other end, however, there are some PMTs for certain fibres where
⁹ after all subruns of data have been included, there remains no hits. In this scenario,
¹⁰ one can show that the log-likelihood becomes linear in the beam profile parameter:

$$\mathcal{L}(b'_i | \{m_{ij} = 0\}) = b'_i k_i^{MC} \cdot \sum_j [I_j N_j]. \quad (4.11)$$

¹² This scenario is very much reminiscent of rare-decay searches, and a similar approach
¹³ can be used. A 1σ upper limit on the possible value for b'_i can be analytically-calculated
¹⁴ to be:

$$b'_{i,ulim} = -\frac{k_i^{MC} \sum_j [I_j N_j]}{\ln [1 - \text{erf}(1/\sqrt{2})]}, \quad (4.12)$$

¹⁶ where $\text{erf}(x)$ is the error function. [18 pages for above two sections]

¹⁷ 4.2.2 Results & Discussion

¹⁸ WARNING: contents of this subsection will be gutted, focusing merely on impact of
¹⁹ combining data sets. Details about discrepancies will be covered in the next section.

4.2 Improving the beam profiles

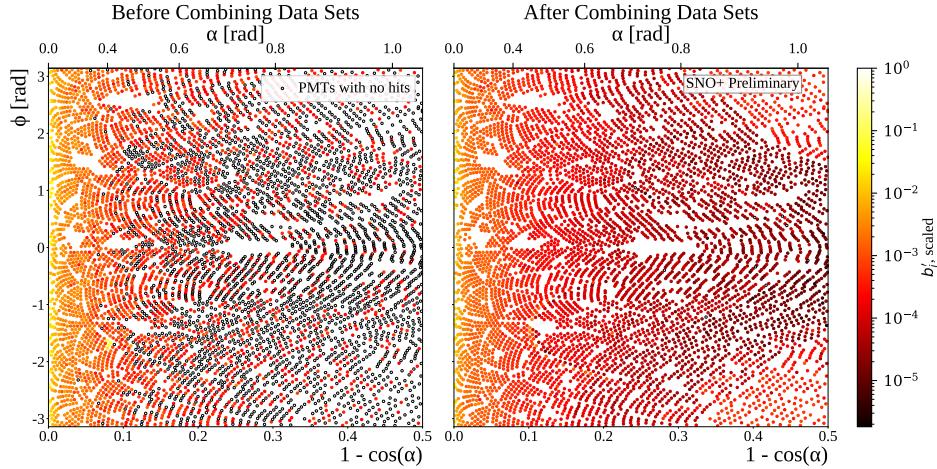


Fig. 4.7 Comparison between old and updated beam profiles for fibre FS055, after combining multiple data sets. Once again, the relative intensities (b'_i) for each PMT are given by the colour of each point, the position of each plotted in the 2D (r, ϕ) -space. The relative intensities have been both scaled here so that the largest value equals 1. Hollowed-out points are PMTs that, even after all relevant subruns have been combined, have no PMT hits.

[Will just be 1 page] Figure 4.7 shows the impact of using additional subruns of data on a typical beam profile. One can clearly see the great reduction in the number of PMTs with no hits in data. That many more data sets were included allowed for the major increase in dynamic range available for measuring these b'_i values. One can also note that by including additional data the curious spot that was seen in the old beam profile our at $r \approx 0.08$ has gone, further indicating that it was an artefact of that single data set.

Further details can be gathered from the interpolated intensity maps, one of which can be seen in figure 4.8. There are two curious stand-out features that can be seen here: firstly, there are multiple distinct parabolic arcs. These correspond to the shadows of the ropes that hold up/down the AV. More precisely, they are the mismodelling of those shadows — if the shadows were in the right place in the isotropic MC, then they would correctly cancel out any decreased intensity seen in the data of shadowed PMTs. These shadows could be mismodelled either because the positions of the ropes in the

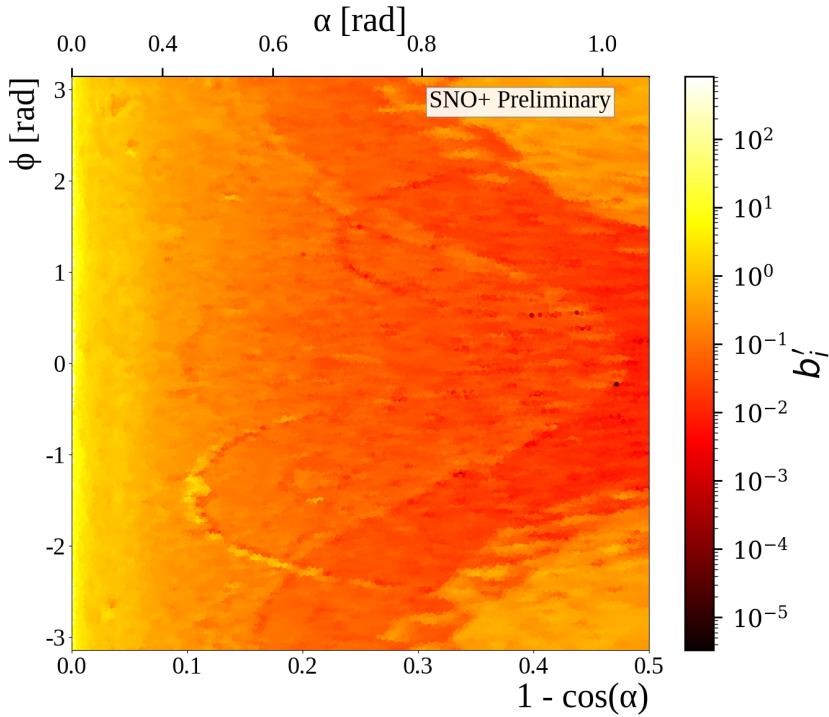


Fig. 4.8 Interpolated intensity map for the new updated beam profile of fibre FS055. The misalignment of rope shadows and AV effects, can both be seen.

¹ MC are in the wrong place, or the fibre's emission position is wrong. Note that any
² mismodelling of the fibre's nominal emission direction has no impact on this shadowing
³ problem, as changing that direction merely causes a change of basis in the (r, ϕ) -space.
⁴ The latter possibility of incorrect fibre positions are more likely, and in fact these arcs
⁵ in the beam profiles could be used as an effective way to correct for this problem.

⁶ The second distinctive feature of this intensity map is the large band of lower
⁷ intensity varying between $r \approx 0.2 - 0.5$, followed by larger intensity out at large
⁸ r values. This feature comes from light reflecting off the AV surface, or internally-
⁹ reflecting. The reason for this band's functional dependence on ϕ is that this particular
¹⁰ fibre, FS055, has a nominal fibre direction $\sim 10^\circ$ from pointing radially-towards the
¹¹ detector's centre. This feature appears in the updated beam profiles of all fibres, but
¹² its shape depends on the particular fibre's direction — for fibres pointing directly

4.2 Improving the beam profiles**51**

towards the detector’s centre, there is little ϕ -dependence observed. Like the ropes, 1
this feature must come from some form of mismodelling of the optics of the AV. A 2
de-facto shadowing of PMTs in line with tangents from the AV surface which intersect 3
the fibre position is to be expected. One also expects PMTs at polar angles larger than 4
this to have their observed intensities boosted from reflected light off the AV. However, 5
the discontinuities seen in the beam profiles indicate that for whatever reason this 6
effect has been over-emphasised in the simulation. 7

There is a further phenomenon that can be seen, by comparing beam profile values 8
obtained from a single subrun to the updated combined beam profile. This can be 9
done by calculating the residuals corresponding to the single subrun, relative to the 10
combined data set. The residual is negative if the combined data sets have a b'_i below 11
the equivalent for a given single subrun; that is, the combined model underestimates 12
this subrun for that PMT. 13

This information was plotted for two different subruns from the same fibre, seen in 14
figure 4.9. One subrun was the same one used by Esther Turner for the original 2D 15
beam profiling, with a wavelength of 495 nm; the latter was at the longer wavelength of 16
595 nm. For both subruns, most PMTs are seen to have intensities well-modelled by the 17
combined model. However, there appears to be a significant amount of mismodelling 18
within the beamspot. There also appears to be some systematic shift between data 19
and model at somewhat larger polar angles. Moreover, this mismodelling seems not to 20
be merely random, but a function of wavelength: at shorter wavelengths the beamspot 21
tends towards being overestimated and then underestimated at larger values of α . At 22
longer wavelengths, the beamspot becomes underestimated, with larger angles getting 23
overestimated. This indicates that there appears to be a wavelength-dependence on the 24
beam profiles, contradicting one of the main assumptions which we used to combine 25
the water-phase data in the first place! All three of these features — rope shadows, 26

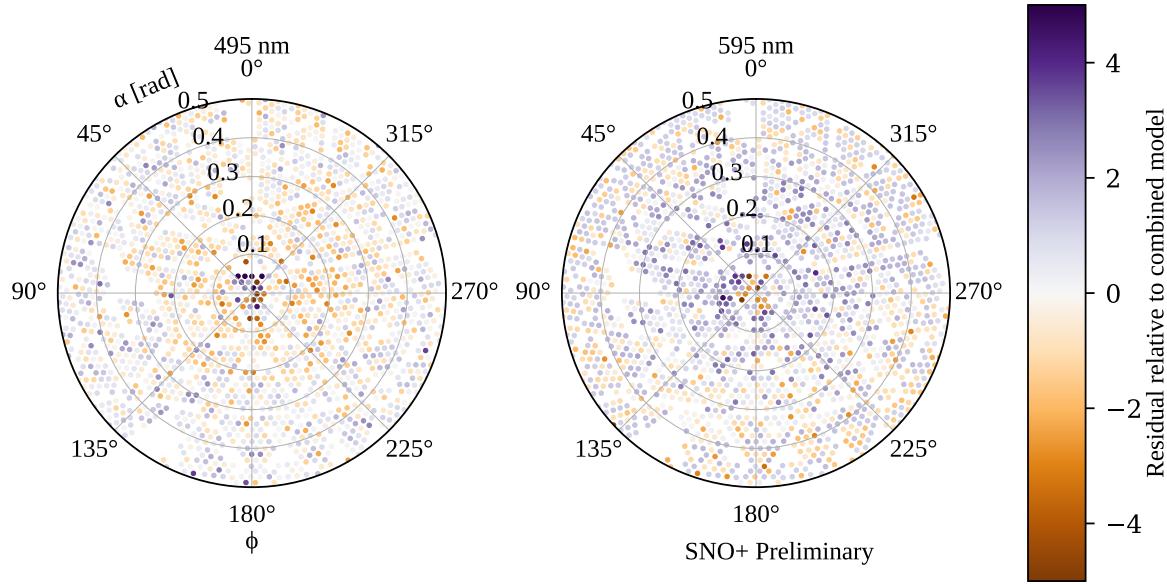


Fig. 4.9 Residuals from subruns at two different wavelengths, both compared to the combined beam profile model for fibre FS055. A negative sign, and hence bluer colours, indicate that the combined model underestimates the observed intensity for that particular subrun. Values with a magnitude beyond 5 are shown capped at this maximal value for the purposes of this plot. These PMTs are plotted in the polar fibre coordinates (α, ϕ) .

¹ AV reflections, and wavelength dependence — add systematic uncertainty to the beam
² profiles, beyond the statistical uncertainty as measured by the width of the likelihood
³ distribution. Certainly if one wanted to further improve the uncertainties in the beam
⁴ profiles, tackling these challenges would be key.

⁵ 4.3 Comparisons between Data and Simulation

- ⁶ • This focuses on disagreements noticed between data and MC even after the new
⁷ generator and beam profiles have been used.
⁸ • Important to mention that none of these are necessarily game-ending, they are
⁹ just systematics that may or may not be substantial in a given analysis with
¹⁰ SMELLIE..

4.3.1 Forward Hemisphere Discrepancies

- The continued disagreement between data and MC when it comes to measured npe in various parts of the “forward” hemisphere. This includes:
 - The central beamspot,
 - The TIR region,
 - Rope shadows,
 - A noticeable wavelength-dependence.
- This is pretty much most of the contents of Section 4.2.2.

[4 pages]

9

4.3.2 Emission Time Discrepancies

- For certain lasers, a strong mismatch in the observed hit time residuals for prompt light.
- A mysterious +18 ns bump seen for the PQ495 laser.
- A trigger jitter in the SuperK laser.

[3 pages]

15

4.3.3 Backward Hemisphere Discrepancies

- The observed distribution of hits vs time and angle in MC does not match data in a number of ways for PMTs near the fibre emission point.
- Includes the outer-water scattering length, rope reflections, and investigations into whether certain modifications to the optics could plausibly fix things (so far, no).

21

¹ [6 pages]

² [32 PAGES TOTAL]

Chapter 5

Analysis of SMELLIE Data in the Scintillator Phase

This chapter contains two sets of analyses: measurements of the extinction lengths of the scintillator as a function of wavelength and time, as well as monitoring the Rayleigh scattering length over time.

5.1 Extinction Length Analysis

5.1.1 Motivation

- Explain motivating observations for this analysis: a substantial discrepancy between MC and data seen in the radial profiles of nhitsCleaned for ^{210}Po after the PPO top-up campaign. Performed by Serena.
- Hypothesis of a shortening of the absorption/scattering length proposed, further strengthened by Ben Tam’s ex-situ absorption measurements with scintillator taken from the detector, as well as knowledge about a likely “cooking” of PPO during the PPO-fill.

- Describe the provisional optics model decided on based on these measurements, which includes an additional non-re-emitting component of the scintillator.
- As a further cross-check, SMELLIE should be sensitive to changes in the overall extinction length of the scintillator, especially for short extinction lengths relative to the size of the detector.
- More straightforward in measuring extinction length compared to scattering length — no need to distinguish between scintillator re-emission and scattering.
- Further uses: can be used to monitor the extinction length over time!

[4 pages]

5.1.2 Analysis Approach

The 2-PMT Case

- Outline theoretical approach for how one could measure the extinction length of scintillator through a comparison of SMELLIE data between the scintillator and water phases, in the simplified 1 dimensional case with only 2 PMTs.

[3 pages]

Combining Results Between PMTs

- Not doing analysis with just 2 PMTs, of course! Can combine results from multiple PMTs within a beamspot: I explain how here.

[2 pages]

5.1 Extinction Length Analysis**57****Corrections Between the Water and Scintillator Phases**

1

- Note the complications that we have to deal with. Namely, the differing refractive indices of the media bending the beamspot differently in the phases, as well as the method used to estimate t_{emm} .
2
3
4
- Explain how we deal with these, the former through MC simulation.
5

[2 pages]

6

5.1.3 Validation of the Analysis in Simulation

7

- Show results of this approach being used to measure the extinction length in simulation. How well does it do?
8
9

[3 pages]

10

5.1.4 Results in Data

11

- Describe the data used in this analysis, both water and scintillator, which can be shown in a table.
12
13
- Show examples of analysis of data in action for 375 nm data: typical t_{res} distributions of backscattered and beamspot PMTs; calculation of that particular extinction length measurement, followed by the graph for extinction length in 375 nm over all fibres and time periods.
14
15
16
17
- Discuss what results can be seen in this plot: consistency between fibres, the expected change as a function of PPO concentration, and stability of the extinction length during the main 2.2 g/L scintillator phase.
18
19
20

- Compare results to those made by Ben ex-situ: are they in agreement? If not, what possible systematics could there be? The main one for my analysis is likely to be uncertainties in the simulated beam profile that leak through into the refractive index correction of the beamspot. For the ex-situ analysis, the value of the extinction length obtained is achieved through background subtraction at some long wavelength, and the particular choice of this wavelength can lead to systematic changes in the obtained extinction length.
- Look at results at longer wavelengths: can anything reasonably be said at these longer wavelengths? Why/why not?
- Finally: describe any conclusions that can be reached, in particular whether we can affirm the optics model we use in RAT.

[8 pages]

5.2 Scattering Analysis

5.2.1 Historical Approaches and the Problem of Systematics

- Comparison to MC is necessary in scattering analysis, compared to merely being needed as a correction factor. This is because of the angular dependence of scattering. As a result, we can be far more susceptible to systematics from poor modelling!
- As a warning, show how Krish's/Esther's approach to the SMELLIE scattering analysis suffers majorly from these systematic effects. Requires describing their analysis approach briefly, and then explaining how the systematics described in Section 4.3 lead to major problems with this approach.

5.2 Scattering Analysis

59

- Motivates the need for either reduced systematics, or an alternative analysis approach that is more robust to them! 1
2

[2 pages] 3

5.2.2 New Methodology

4

Signal Region Selection

5

- Propose the new analysis approach: looking at light in the “bad light-path” PMT region. Define what this region is. 6
7
- Give qualitative argument for why we expect this region to be robust to the beam profile systematics: dominated by the scattered signal as no direct light can make it here, and changes to beam profile should get “smeared out” after scattering. 8
9
10
11
- Show how simulations indicate this should be a region with a very high purity of scattered light, and (assuming all else being equal) robust to beam profile uncertainties. 12
13
14
- Confirm robustness of selected PMT region to uncertainties in AV offset and fibre position. 15
16

[5 pages] 17

Measuring the Emission Intensity

18

- Remaining systematics is now in the calculation of an average absolute emission intensity. 19
20

60

Analysis of SMELLIE Data in the Scintillator Phase

- 1 • Show how various methods don't work particularly well: whole detector npe,
- 2 beamspot npe, backscattered light npe, "bad light-path" PMTs but at later times.
- 3 Explain why it goes wrong for each method.
- 4 • Look at "beamspot but excepting the central bit": if that works well, then we
- 5 can continue!
- 6 • Otherwise, we'll have to live with measuring relative scattering lengths instead
- 7 of absolute amounts, using the outer water back-scattering as a measure of the
- 8 relative emission intensity.

9 [4 pages]

10 **5.2.3 Results**

- 11 • Actually do the proposed analysis on data, versus time and wavelength. Do the
- 12 results seem consistent between fibres? Are they sensible values?

13 [5 pages] [33 PAGES TOTAL]

Chapter 6

Solar Oscillation Analysis

Driving out into the Sun

Let the ultraviolet cover me up

Looking for a Creation Myth

Ended up with a pair of black lips

This is the End

PHOEBE BRIDGERS

Measuring the “solar” neutrino oscillation parameters Δm_{21}^2 and θ_{12} is one of the principal aims of the SNO+ detector during the scintillator-phase. There are, in fact, two complementary methods of measuring these parameters: the oscillations of anti-neutrinos from terrestrial nuclear reactors, and the oscillations of neutrinos from the Sun.

This chapter focuses on the latter approach, using ${}^8\text{B}$ neutrinos coming from the Sun to measure the solar oscillation parameters. An initial background-free study was performed by Javi Caravaca [], which demonstrated that it was indeed possible to make such a measurement in the detector. The work in this chapter builds on substantially from that analysis. This chapter also draws on the associated reactor anti-neutrino

¹ analysis built by Iwan Morton-Blake []], and more broadly from the general techniques
² used in the $0\nu\beta\beta$ analysis of Tereza Kroupova [] and Jack Dunger [].

³ This chapter begins by explaining how it is possible to measure the solar oscillation
⁴ parameters via ^8B events. Then, the framework used to perform the analysis is then
⁵ explained: that of a *Bayesian Analysis using Markov Chain Monte Carlo techniques*.
⁶ After the method has been described, the dataset upon which the analysis is performed
⁷ is then introduced. The results and associated validation are then given. Given these
⁸ results, a projection is then made for the expected sensitivity to θ_{12} as a function of
⁹ livetime.

¹⁰ 6.1 Analysis Methodology

¹¹ 6.1.1 Observational Principle

¹² How can we measure neutrino oscillation parameters via solar neutrinos in the SNO+
¹³ detector? As discussed in Chapter 1, it is possible to detect all flavours of neutrino
¹⁴ through elastic scattering with electrons in the detector. If this interaction was purely
¹⁵ neutral-current, then there would be no way of telling the flavour-state of an interacting
¹⁶ neutrino. However, electron neutrinos are able to interact through an additional
¹⁷ charged-current mode. This modifies the cross-section for electron neutrinos, and
¹⁸ means that as the survival probability for electron neutrinos generated from the Sun,
¹⁹ P_{ee} , is modified, the interaction probability of neutrinos with the detector will also.

²⁰ Of course, we do not directly measure neutrino energies in the detector — only
²¹ the associated scattered electron. If there were no correlation between the observed
²² electron energy and its associated neutrino, then the only effect of neutrino oscillations
²³ would be to change the overall observed rate of events due to this process. There
²⁴ would be no change in the shape of the event’s energy spectrum, even though neutrino

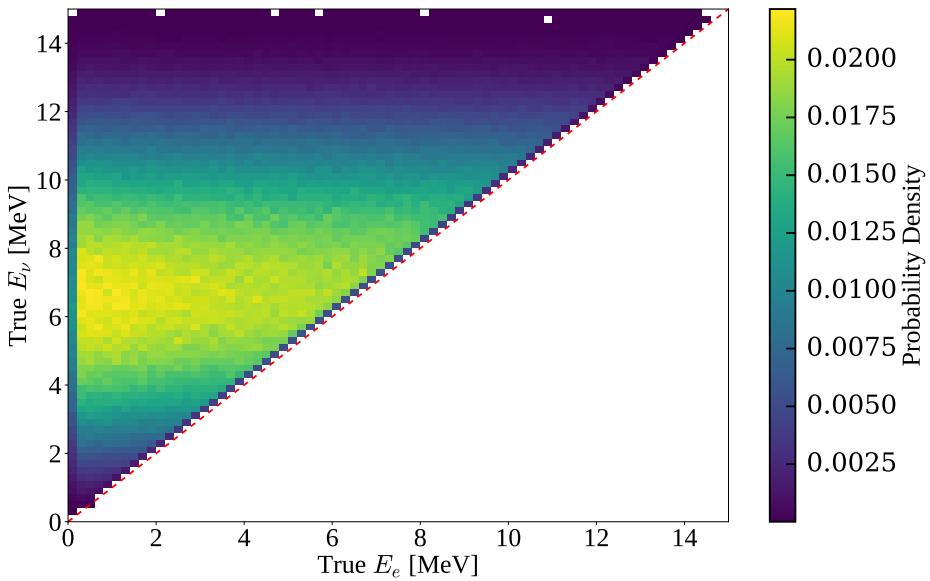


Fig. 6.1 2D probability distribution comparing the true neutrino energy from a ${}^8\text{B}$ ν_e to the true energy of the scattered electron. Also shown in red is the line $E_\nu = E_e$.

Fig. 6.2

oscillations are a function of neutrino energy. Fortunately, there is some dependence of the neutrino's energy, E_ν on that of the scattered electron, E_e . This dependence can be seen in Fig. 6.1 for ${}^8\text{B}$ electron neutrinos interacting in SNO+. As can be seen, the dependence is weak, and comes mostly from basic energy conservation: If one observes a 10 MeV electron event in the detector, it can't reasonably have come from a 5 MeV neutrino.

In Fig. 6.2 we can see the impact each physical process has on the energy spectrum that we eventually observe. We start with a broad energy distribution of ${}^8\text{B}$ electron neutrinos generated in the Sun. These neutrinos then oscillate their flavour state as they propagate to the detector, in an energy-dependent manner. When a (tiny) fraction of these neutrinos interact with the electrons in our detector, there is both an energy- and flavour-dependence on the cross-section. The scattered electrons gain a kinetic

Fig. 6.3

¹ energy with some mild dependence on the inciting neutrino's energy, which is then
² measured by the detector to within some energy resolution.

³ Let us now consider the dependence of P_{ee} on the individual neutrino oscillation parameters. Recall from Eq. 1.1 that, after considering matter-induced oscillations due to neutrinos passing through the Sun and possibly the Earth, $P_{ee} =$
⁶ $P_{ee}(\tan 2\theta_{12}^M, \sin \theta_{13}^M, \Delta m_{21,M}^2) = P_{ee}(\theta_{12}, \theta_{13}, \Delta m_{12}^2, \Delta m_{13}^2)$. Fig. 6.3 shows the dependence of each of these four oscillation parameters on $P_{ee}(E)$. We can see that in reality
⁸ only the two parameters Δm_{21}^2 and θ_{12} have a substantial impact on $P_{ee}(E)$ and hence
⁹ the observed electron energy spectrum. Because of this, for this analysis we will only
¹⁰ ever vary these two oscillation parameters, and keep θ_{13} and Δm_{13}^2 at their current
¹¹ global fit values¹ of $\sin^2 \theta_{13} = 0.0222$ and $\Delta m_{13}^2 = +2.515 \times 10^{-3} \text{ eV}^2$ [].

¹² 6.1.2 Background Processes

¹³ Sadly, elastically-scattered electrons from ${}^8\text{B}$ neutrinos are not the only events we
¹⁴ see in the SNO+ detector during the scintillator phase. There are a number of
¹⁵ background processes that our signal must compete against. Below a reconstructed
¹⁶ energy of $\sim 2 \text{ MeV}$, it is known that various backgrounds completely swamp any
¹⁷ possible ${}^8\text{B}$ signal, and so for this analysis we only consider processes that can generate
¹⁸ reconstructed energies of at least $E_{\min} = 2 \text{ MeV}$. The following subsections explain
¹⁹ each of these backgrounds, as well as methods that have been used to mitigate them
²⁰ as much as possible.

¹We use the global fit results excluding Super-Kamiokande's atmospheric data, and assuming normal ordering of the neutrino mass hierarchy. This choice has a tiny impact on the magnitudes of these two fixed parameters, the main impact being the sign of Δm_{13}^2 .

Fig. 6.4

Internal Uranium- and Thorium-Chain Backgrounds

Although every effort has been made to make the scintillator cocktail that fills SNO+ to be as radio-pure as possible, there inevitably remain trace amounts of the radioactive isotopes that derive from the decay chains of the ^{238}U and ^{232}Th isotopes. Fig. 6.4 shows these two decay chains. Fortunately, only a fraction of the radioactive isotopes in these chains actually are capable of generating events in the detector with energies above E_{\min} : these have been highlighted in Fig. 6.4 in gold.

Of particular note are the decays of ^{212}Bi and ^{214}Bi . Both are capable of either α - or β -decays to Tl or Po isotopes, respectively. For the former, it is the subsequent β -decay of the Tl that can have a reconstructed energy above E_{\min} . For the latter, the Bi decay is the part of the pair of decays that can lie above E_{\min} . Although the α -decays here certainly have Q-values well above 2 MeV, the liquid scintillator quenches the observed energy well below E_{\min} . The so-called “Bi–Po” decays are particularly special because the lifetimes of ^{212}Po and ^{214}Po are 300 ns and 164 μs , respectively, which are short enough to allow for highly-effective coincidence tagging.

There are two classes of Bi–Po event in the detector: “out-of-window” events for which the Bi and Po occur in separate event windows, and “in-window” events whereby the Bi and Po occur within the same event window. These lead to two distinct strategies for tagging these kinds of events. For out-of-window Bi–Pos, we look for a delayed coincidence of two events. Using the tagging algorithm suggested in [] as a starting point, the chosen procedure was as follows. There must be two events that trigger the detector within 4 μs of one another, and both have a valid `scintFit` position reconstruction within 2 m of one another. The delayed candidate event must also have at least 100 cleaned PMT hits. This very broad coincidence tagging procedure

¹ was designed to ensure that the cut was as *efficient* in tagging (and hence, rejecting)
² Bi–Pos as possible, whilst negligibly impacting the solar signal. This is in contrast to
³ the cuts chosen by Rafael Hunt-Stokes in [], which try and obtain a highly *pure* sample
⁴ of Bi–Po tags.

⁵ Of course, the above delayed coincidence procedure cannot catch any of the in-
⁶ window Bi–Po events. For these, we use a different approach. Because two decays
⁷ happened in the same event window, we expect to see two distinct peaks in the event’s
⁸ time residual spectrum. In order to look for this event topology, a likelihood-ratio
⁹ classifier was run over events, first developed by Eric Marzec [] and re-coordinated for
¹⁰ the 2.2 g l^{-1} LABPPO scintillator optics by Ziping Ye []. This classifier calculates the
¹¹ likelihood ratio between the null hypothesis of a $0\nu\beta\beta$ event (a proxy in this analysis for
¹² single-site events such as our ^8B signal) and the alternative hypothesis of an in-window
¹³ Bi–Po event. The more negative the value of the result, `alphabeta212`, the greater
¹⁴ the evidence there is for rejecting the null hypothesis of a single-site event. Events
¹⁵ with `alphabeta212 < 0` were then rejected.

¹⁶ Combining both out-of-window and in-window Bi–Po tagging, the impact on
¹⁷ ^{212}Bi –Po, ^{214}Bi –Po, and ^8B ν_e events can be seen in Fig. 6.5. We consider here only
¹⁸ events that pass all other cuts used in this analysis: see Section 6.2.2 for the specifics
¹⁹ of the cuts used. Because of the different lifetimes of the decays, ^{214}Bi –Po decays
²⁰ predominantly fall out-of-window whilst ^{212}Bi –Po events are typically in-window. This
²¹ explains why the out-of-window tagging is substantially better at cutting ^{214}Bi –Po
²² decays, whereas the in-window tagging far better tags ^{212}Bi –Po decays. Overall, within
²³ the analysis region of interest (ROI), the two combined cuts are able to tag TODO%
²⁴ of ^{214}Bi –Po triggered events, TODO% of ^{212}Bi –Po triggered events, whilst retaining
²⁵ TODO% of ^8B ν_e signal events.

Fig. 6.5

Fig. 6.6

(α, n) Reactions

The impact of ^{238}U - and ^{232}Th -chain isotopes does not simply end at their direct decays. It is possible for the α s generated during these decays to undergo their own interactions with nuclei in the detector. Within the organic scintillator of SNO+, the dominant interaction of this type is when an α collides with a ^{13}C nucleus, emitting a neutron: $\alpha + ^{13}\text{C} \longrightarrow ^{16}\text{O} + n$. This is known as an (α, n) reaction.

The topology of this reaction in the detector is a delayed coincidence, as shown in Fig. 6.6. For the prompt signal, there is the light emitted from the α just before, and the n just after the (α, n) . The neutron then thermalises and gets captured by another nucleus — usually hydrogen in SNO+ — which creates an excited state that then eventually decays, creating a γ that creates the delayed signal in the detector [].

As can be seen in Fig. 6.7, (α, n) interactions can lead to events reconstructed at a wide variety of energies, which could be an issue for this analysis. However, because they are delayed coincidence events with a typical decay time of ~ 100 ns, the aforementioned out-of-window Bi–Po tagging algorithm also efficiently tags (α, n) events. Looking again at Fig. 6.7, simply by using the out-of-window Bi–Po tagger without any further modifications TODO% of events in the ROI are cut.

External Backgrounds

All materials within the SNO+ detector are radioactive, not just the liquid scintillator cocktail. This includes the acrylic, ropes, external water, and PMTs. These components

Fig. 6.7

Fig. 6.8

1 have had their radiopurity “assayed” (that is, measured) throughout the detector’s
 2 lifetime, often back to the construction of the original SNO detector itself. The
 3 materials other than the liquid scintillator are known to have far higher background
 4 levels, especially in the important ^{238}U - and ^{232}Th -chain backgrounds []. To distinguish
 5 between the inherent backgrounds within the scintillator, and the backgrounds from
 6 materials at larger radii, we use the terminology “internal” and “external”, respectively.

7 Although there are numerous external backgrounds, with a suitably accurate and
 8 precise position reconstruction algorithm they can be suitably handled. The simplest
 9 approach is with a so-called “fiducial volume” (FV) cut: just throw out all events
 10 that reconstruct beyond some radius. The only external background events that will
 11 reach within the FV are those that have reconstructed very poorly, or have some
 12 long-distance radiation that manages to deposit radiation close to the centre of the
 13 AV. Because α and β radiation can only travel short distances through the detector,
 14 it is only γ radiation that can realistically travel far enough into the detector to
 15 be able to reconstruct anywhere near the centre. Moreover, the intensity of this γ
 16 radiation attenuates exponentially towards the centre of the detector, meaning only a
 17 tiny fraction of the total number of external events reconstruct within a 3.5 m FV, say.
 18 This strong radial-dependence can be seen in Fig. 6.8.

19 What this figure also demonstrates is that our solar signal has a completely different
 20 radial dependence to these backgrounds. As a result, if one considers not just the energy
 21 of events but also their reconstructed radius, then it is possible to get an additional
 22 handle on the external backgrounds. The FV cut can then be pushed further out to
 23 larger radii, allowing one to gain more signal statistics.

24 Work by Tereza Kroupova [] allows for additional means of distinguishing external
 25 backgrounds from the solar signal. The underlying assumption in the reconstruction of

Fig. 6.9

SNO+ events is that there was an electron at a single point, which is entirely valid for ${}^8\text{B}$ elastic scattering events. However, external backgrounds can fail this assumption in two ways. Firstly, these radioactive decays often generate γ radiation in addition to the main α/β particle, which creates a multi-site event. Because the `scintFit` position reconstruction algorithm is not prepared for a distribution of energy depositions in the scintillator, the t_{res} distribution will broaden. This allows an event classifier to be built that distinguishes between the t_{res} distributions of single-site events and externals, known as the “external background timing classifier”. Secondly, because external backgrounds that do reconstruct close to the centre of the detector typically have a γ that travelled a long distance towards the centre of the detector, we expect the earliest light that hits the PMTs to arrive most often along the direction of the reconstructed position vector. A distribution of PMT hits for a given event as a function of their angular distribution relative to the direction of position reconstruction can be built, and compared to the expected distributions for single-site and external background events. This is known as the “external background topological classifier”. Much like the classifier described in Section 6.1.2, the single-site events used for comparison were $0\nu\beta\beta$ events, but these have a similar single-site structure to the solar signal of interest in this analysis.

Fig. 6.9 shows the correlation between the two classifier results for both a typical external background, and ${}^8\text{B} \nu_e$ events.

Cosmogenic Isotopes

The final source of background events are radioactive isotopes that form via collisions of cosmic rays with atomic nuclei, known as cosmogenic isotopes. Most of these isotopes are short-lived [], with lifetimes $\mathcal{O}(1\text{ s})$. Fortunately, the depth of SNO+ means that

1 our rate of cosmic ray muons interacting with the detector is only 3 an hour []. Because
2 the rate is so low relative to other experiments, relatively straightforward approaches
3 to tagging and removing cosmic ray muons and their cosmogenic followers can be
4 utilised without substantial loss of livetime. Events are tagged as a cosmic ray muon if
5 they create a sufficient number of hits in the outward-looking PMTs above background
6 levels, as well as many hits within the detector itself. The details of this tagging
7 for the scintillator-fill were put in place by Lorna Nolan [], modifying the existing
8 algorithms used in the water-phase [] and in SNO []. After a tagged cosmic muon
9 event, all events for the next 20s are then vetoed as a means of rejecting followers.
10 This simple cut is enough to remove the vast majority of expected cosmogenic events
11 in the scintillator-phase. The expected impact on loss of livetime, and hence quantity
12 of signal events, is 3 lots of 20s vetoes an hour, that is to say 1/60th of the signal is
13 lost via this cut.

14 There is one cosmogenic isotope with a long-enough half-life that even the 20s
15 muon follower veto is not sufficient to remove all events. This is ¹¹C, which β^+ -decays
16 to stable ¹¹B with a half-life of 20.4 min. The maximum possible energy deposited in
17 this decay is 1.982 MeV [], just below E_{\min} , so only a small fraction of ¹¹C events end
18 up in the ROI: the ones with very high energies that get their energy reconstruction
19 falsely-inflated by some amount. As a result, this background is expected to be very
20 much sub-dominant to all other backgrounds in this analysis. Because this background
21 is important to consider for some other analyses, a triple-coincidence tagging algorithm
22 is being built by Katharine Dixon [], but not used for this analysis currently.

23 6.1.3 The Log-likelihood Test Statistic

24 At the highest level, this analysis involves taking the data observed in the scintillator-fill
25 after applying a certain set of cuts, along with simulated PDFs for all processes believed

to build up the observed data with those same cuts applied, and then attempting to fit the combined energy and radial distributions of the MC to that of the data. Given a set of PDFs, to try and match the distribution of observables in data we can modify a number of parameters. These consist of the normalisations of each PDF (i.e. the total number of events observed due to that process), and any systematic parameters that could modify the shapes of these distributions. For this analysis, the neutrino oscillation parameters act as *de facto* systematic parameters, as they modify the shape of the ⁸B PDFs. Of course, unlike usual systematics the oscillation parameters are what we are actively trying to measure, instead of being a nuisance.

In order to perform a fit to data in this way, we must first answer a set of questions:

1. Which signal and background processes must we consider?
2. In addition to their normalisations, are there any further parameters necessary to specify the distributions of the PDFs for each of the processes? Systematics and oscillation parameters are good examples.
3. What is our test statistic?
4. What algorithm do we use to try and find the best-fit result?
5. How do we measure uncertainties on these best-fit values for each parameter?

In section 6.1.2, question 1 was answered for this analysis. We now give the answer to question 3; all other questions on this list will be answered shortly.

The test statistic used for this analysis is the *binned extended log-likelihood*. Once the data and MC PDFs have been binned in both the observables of interest, it is assumed that the expected number of events in a given bin j is governed by a Poisson distribution:

$$P_j(n_j|\lambda_j) = \frac{\lambda_j^{n_j} e^{-\lambda_j}}{n_j!}, \quad (6.1)$$

1 where $P_j(n_j|\lambda_j)$ is the probability of observing n_j events in bin j , given an expectation
 2 of λ_j events in total from signal and background processes in that bin. This λ_j can be
 3 decomposed into each of the expected rates for each process, i :

$$4 \quad \lambda_j = \sum_{i=1}^{N_{\text{PDFs}}} \mathcal{N}_i P_{ij}(\boldsymbol{\theta}), \quad (6.2)$$

5 where N_{PDF} is the number of PDFs being considered in the analysis, \mathcal{N}_i is the nor-
 6 malisation parameter of the i^{th} PDF, and $P_{ij}(\boldsymbol{\theta})$ is the probability of observing an
 7 event of process type j in bin i , assuming a set of non-normalisation parameters $\boldsymbol{\theta}$. By
 8 combining the probabilities of all the bins together, the total probability for a given
 9 set of processes assuming these parameters to give rise to the data seen is:

$$10 \quad P(\mathbf{n}|\mathcal{N}, \boldsymbol{\theta}) = L(\mathcal{N}, \boldsymbol{\theta}|\mathbf{n}) = \prod_{j=1}^{N_{\text{bins}}} \frac{\left[\sum_{i=1}^{N_{\text{PDFs}}} \mathcal{N}_i P_{ij}(\boldsymbol{\theta}) \right]^{n_j} e^{-\sum_{i=1}^{N_{\text{PDFs}}} \mathcal{N}_i P_{ij}(\boldsymbol{\theta})}}{n_j!}, \quad (6.3)$$

12 where N_{bins} is the total number of bins being considered in the analysis. This probability
 13 can be re-framed as the likelihood of the vectors of parameters \mathcal{N} and $\boldsymbol{\theta}$ given
 14 the vector of number of events in each bin, \mathbf{n} : $L(\mathcal{N}, \boldsymbol{\theta}|\mathbf{n})$. It is rare to see the
 15 likelihood as-is, instead, for computational purposes the log-likelihood is used instead,
 16 $\mathcal{L}(\mathcal{N}, \boldsymbol{\theta}|\mathbf{n}) := \ln L(\mathcal{N}, \boldsymbol{\theta}|\mathbf{n})$. We can then get to the formula actually used for this
 17 analysis:

$$18 \quad \mathcal{L}(\mathcal{N}, \boldsymbol{\theta}|\mathbf{n}) = - \sum_{i=1}^{N_{\text{PDFs}}} \mathcal{N}_i + \sum_{j=1}^{N_{\text{bins}}} n_j \ln \left(\sum_{i=1}^{N_{\text{PDFs}}} \mathcal{N}_i P_{ij}(\boldsymbol{\theta}) \right). \quad (6.4)$$

20 6.1.4 The Bayesian Statistical Approach

21 There are two main schools of statistical inference, “Frequentist” and “Bayesian”. In
 22 the former, probabilities describe the fraction of times a situation can be found within
 23 the whole ensemble of possible worlds. For the latter, we care not about an ensemble

of worlds but instead our degree of belief in this current one. We update our beliefs as we acquire knowledge of the world through Bayes' Theorem:

$$P(\boldsymbol{\mu}|\mathbf{x}) = \frac{L(\boldsymbol{\mu}|\mathbf{x}) P(\boldsymbol{\mu})}{P(\mathbf{x})}. \quad (6.5)$$

Here, $\boldsymbol{\mu}$ is the set of parameters that model our system, $P(\boldsymbol{\mu})$ is our *prior* (pre-existing) distribution for those model parameters, and \mathbf{x} is the data taken in our experiment. The updated, *posterior* distribution $P(\boldsymbol{\mu}|\mathbf{x})$ is then the prior multiplied by the likelihood of parameters $\boldsymbol{\mu}$ given observations \mathbf{x} , $L(\boldsymbol{\mu}|\mathbf{x})$, and divided by the total probability $P(\mathbf{x})$ of observing \mathbf{x} under any circumstance.

Both approaches to statistics are widely-used in statistical analysis, in both particle physics and beyond. The Bayesian approach was used for this analysis, as it was believed that this helps keep transparent what assumptions are being made in the analysis.

Now, if one is able to determine the overall posterior distribution, then it is possible to derive best-fit values with uncertainties for all parameters in the fit. This is done by “marginalising” the posterior distribution, i.e. integrating over all parameters other than the one of interest. A sensible best-fit value is then the modal marginalised posterior density, the highest value in the marginalised distribution. The uncertainty on this value is derived from the spread of the marginalised posterior, by the calculation of the 1σ Credible Interval (CI): this is the set of values for a given parameter which has a total posterior probability of 68.3%, and contain the best-fit value. There are in fact an infinite number of CIs that satisfy this property; for this analysis, the values are chosen in decreasing order of marginalised posterior probability density.

6.1.5 Markov Chain Monte Carlo

Of course, all of this assumes that one can accurately determine the posterior density distribution. Whilst the likelihood and prior distribution are straightforward enough to calculate, often-times $P(\mathbf{x})$ (which acts as a normalisation) is very challenging to determine. This is because calculating this normalisation involves integrating the likelihood over all the parameter space, and if there are a large number of parameters this can become enormously numerically complex. An alternative approach is needed!

That alternative comes in the form of *Markov Chain Monte Carlo*, MCMC. A Markov Chain is any mathematical system for which the next state of the system is dependent only on its current state; the system is in some sense “memoryless”. For a large class of Markov Chains — those that are “ergodic” and “aperiodic” — one can prove that regardless of the initial position on the chain, the probability distribution converges to the same distribution π . MCMC uses such a Markov Chain which attempts to converge towards the posterior density distribution in particular. In MCMC, after choosing the initial position in the parameter space, successive states are chosen at random with a probability dependent only on the properties of the current position in parameter space and the proposed position. The convergence property of Markov Chains means that the set of steps made in the parameter space after some initial number of steps will have a distribution that converges to that of the posterior density distribution.

There are a number of MCMC algorithms, and the particular one used in this analysis is that of the *Random-Walk Metropolis-Hastings Algorithm*. In this algorithm, after the initial position in the parameter space μ , a new step is proposed some distance from the current one, μ' . This step is chosen at random from a multivariate Gaussian distribution centred on the current position, with widths in each dimension of the parameter space chosen beforehand as constants for tuning the MCMC process. This

choosing of a new proposed step at random is what gives the algorithm its Monte Carlo and Random Walk titles. Once a new step is proposed, it is accepted as the new step with a probability $S(\boldsymbol{\mu}'|\boldsymbol{\mu})$ according to the condition of *detailed balance*:

$$\begin{aligned} S(\boldsymbol{\mu}'|\boldsymbol{\mu}) &= \min\left(1, \frac{P(\boldsymbol{\mu}'|\mathbf{x})}{P(\boldsymbol{\mu}|\mathbf{x})} \frac{R(\boldsymbol{\mu}|\boldsymbol{\mu}')}{R(\boldsymbol{\mu}'|\boldsymbol{\mu})}\right) = \min\left(1, \frac{L(\boldsymbol{\mu}'|\mathbf{x}) P(\boldsymbol{\mu}')}{L(\boldsymbol{\mu}|\mathbf{x}) P(\boldsymbol{\mu})} \frac{R(\boldsymbol{\mu}|\boldsymbol{\mu}')}{R(\boldsymbol{\mu}'|\boldsymbol{\mu})}\right) \\ &= \min\left(1, \frac{R(\boldsymbol{\mu}|\boldsymbol{\mu}')}{R(\boldsymbol{\mu}'|\boldsymbol{\mu})} \exp\left[\mathcal{L}(\boldsymbol{\mu}'|\mathbf{x}) - \mathcal{L}(\boldsymbol{\mu}|\mathbf{x}) + \ln \frac{P(\boldsymbol{\mu}')}{P(\boldsymbol{\mu})}\right]\right). \end{aligned} \quad (6.6)$$

$R(\boldsymbol{\mu}'|\boldsymbol{\mu})$ is the probability density that position $\boldsymbol{\mu}'$ is proposed as a step from position $\boldsymbol{\mu}$, and vice versa for $R(\boldsymbol{\mu}|\boldsymbol{\mu}')$. In most cases, because of the use of the same multivariate Gaussian in choosing proposals, $\frac{R(\boldsymbol{\mu}|\boldsymbol{\mu}')}{R(\boldsymbol{\mu}'|\boldsymbol{\mu})} = 1$ simply. This component only becomes important at the edges of the parameter space, preventing the sampling probability being incorrectly impacted if a proposed step goes outside the allowed parameter space.

It is the detailed balance condition that ensures convergence of the MCMC algorithm to specifically the posterior density distribution. Crucially, because it is only dependent on the ratio of posterior densities, the hard-to-calculate normalisation $P(\mathbf{x})$ in both posterior density terms cancels out, meaning one only needs to calculate the likelihood and priors for each step, as well as $\frac{R(\boldsymbol{\mu}|\boldsymbol{\mu}')}{R(\boldsymbol{\mu}'|\boldsymbol{\mu})}$.

The specific implementation of MCMC used for this analysis is that of **OXO**, a C++ analysis framework first developed by Jack Dunger []. **OXO** is able to run the Metropolis-Hastings algorithm on multidimensional binned data, using the log-likelihood defined in 6.1.3. This framework also allows one to include systematic parameters that can float within the fit, and define non-uniform priors for normalisations and systematics that have constraints: the details of this will be discussed shortly.

6.1.6 Choosing Priors

For this analysis, the suggestions made by Biller & Oser in [] about choosing prior distributions are followed: for parameters that do not have some pre-existing constraint, a flat prior is used. A nice consequence of this choice is that $\ln \frac{P(\mu')}{P(\mu)} = 0$, so the actual value of the prior for these variables never needs to be calculated when running the MCMC algorithm. For the bulk of this analysis, uniform priors are assumed on the neutrino oscillation parameters Δm_{21}^2 and θ_{12} , as the magnitudes of these parameters are now well-established.

For parameters with existing asymmetric constraints $\beta_{-\sigma_-}^{+\sigma_+}$, this analysis uses an asymmetric Gaussian prior, equivalent to the logarithm of the prior being an asymmetric quadratic:

$$\ln P(\mu) = \mathcal{A} - \begin{cases} \frac{(\mu-\beta)^2}{2\sigma_+^2} & \text{if } \mu \geq \beta, \\ \frac{(\mu-\beta)^2}{2\sigma_-^2} & \text{if } \mu < \beta. \end{cases} \quad (6.7)$$

Here, \mathcal{A} is the logarithm of the prior's normalisation constant, and cancels out in the detailed balance condition. For parameters with symmetric constraints, $\sigma_+ = \sigma_-$, then $\ln P(\mu)$ reduces to a quadratic with maximum at $\mu = \beta$.

6.1.7 Including Systematics in the Fit

One important implementation detail is how systematics are applied within the MCMC fitting process. Once systematics are added to the fit, at every step the binned PDFs for all the processes considered in the fit must get modified appropriately, which can become extremely computationally-intensive if not approached carefully. The strategy used in the `OXO` framework starts by thinking of the contents of a binned PDF as a vector of bin probabilities, $\mathbf{p} = (p_1, p_2, \dots, p_{N_{\text{bins}}})^T$. Then, we can think of a systematic acting on the PDF as a linear transformation, and hence a matrix S acting

Fig. 6.10

on this vector: $\mathbf{p}' = S\mathbf{p}$. We only need to calculate this matrix once for a given set of systematic parameter values, and can then use the same matrix on all the PDFs in the fit. Furthermore, when multiple systematics are applied, the matrix for each systematic can then be combined via matrix multiplication into one single “detector response” matrix. `OXO` uses the `Armadillo` [] linear algebra package for efficient matrix manipulation.

There is a problem that can arise when considering the impact of systematics near the edge of the analysis ROI. Many systematics such as shifts, scalings, and convolutions use information about the contents of nearby bins to determine the contents of a particular bin. However, for bins near the edge some of that information does not exist — it has been lost to the cuts that define the ROI. This can lead to a bias in the generation of the modified PDFs, and therefore also the posterior distribution.

As an example, consider the impact of an energy scale systematic on the energy distribution of ^{234m}Pa events in the detector, shown in Fig. 6.10. Because the events seen for this process in the ROI are merely the high-energy tail, any systematic energy scaling $E'_{\text{reco}} := \beta E_{\text{reco}}$ should have a large impact on the number of events observed at the low end of the ROI. However, given that the information about data below E_{min} is lost to the ROI cuts, any energy scaling of $\beta > 1$ will not be applied correctly at all.

The solution to this problem is defining a “buffer region” of bins on either side of the ROI, which allow for tracking of events in and out of the ROI due to systematics, but aren’t considered when calculating the likelihood. This is also shown in Fig. 6.10. After the scaling systematic is applied, although incorrect bin values are found in the buffer region, this is fine because we are no longer calculating the likelihood with those bins. Note that because of this modification, the normalisation parameters we put into the model no longer represent the expected number of events in the ROI. Instead, they

¹ represent the number of events expected in both the ROI and buffer region, before any
² systematics have been applied.

³ 6.1.8 Including Oscillations in the Fit

⁴ Within the analysis MCMC code, the process of neutrino oscillations are thought of
⁵ as a *de facto* systematic that acts only on the ${}^8\text{B}$ ν_e and ν_x signal spectra. Three
⁶ parameters relevant to the signal are floated within the MCMC fit: Δm_{21}^2 , θ_{12} , and
⁷ $\Phi_{8\text{B}}$, the unoscillated ${}^8\text{B}$ flux relative to the expected rate. For the two signal PDFs, a
⁸ third “bookkeeping” dimension is added on top of reconstructed energy and radius:
⁹ the true neutrino energy, E_ν . This is necessary for correctly applying oscillations, as
¹⁰ the oscillation probability is a function of the neutrino’s energy, not the scattered
¹¹ electron’s. Before the fit, these 3D PDFs are given normalisations corresponding to
¹² the expectation of the number of events for each type, ν_e and ν_x , after cuts but before
¹³ oscillations have been applied. Strictly speaking there should be zero ν_x events before
¹⁴ neutrino oscillations: the pre-oscillation rate used here is the post-cut number of events
¹⁵ expected if 100% of the neutrinos oscillated to the ν_x type.

¹⁶ During the MCMC fit, for a given set of parameters $\boldsymbol{\theta} = (\Delta m_{21}^2, \theta_{12}, \Phi_{8\text{B}})$ the
¹⁷ following is performed to oscillate the signal PDFs. Firstly, the normalisations are scaled
¹⁸ by the factor $\Phi_{8\text{B}}$. Then, for each E_ν bin the survival probability $P_{ee}(E_\nu, \Delta m_{21}^2, \theta_{12})$ is
¹⁹ calculated. Each bin then has their probability scaled by either P_{ee} or $1 - P_{ee}$, for ν_e and
²⁰ ν_x respectively. Of course, within the structure of the **OXO** framework these bin-by-bin
²¹ scaling aren’t immediately applied, but instead a matrix describing the impact of
²² oscillations on each of the PDFs is made. Because the oscillation transformation is
²³ purely a bin-by-bin scaling, the resulting matrices are diagonal, with diagonal elements
²⁴ $\Phi_{8\text{B}} \cdot P_{ee}(E_\nu, \Delta m_{21}^2, \theta_{12})$ or $\Phi_{8\text{B}} \cdot (1 - P_{ee}(E_\nu, \Delta m_{21}^2, \theta_{12}))$ for ν_e and ν_x respectively.
²⁵ After the oscillation matrix along with all other systematic matrices are applied to the

signal PDFs, the PDFs are then marginalised over the E_ν dimension so that the signal PDFs match the dimensionality of all other processes.

Calculations of the survival probability are handled with `PSelmaa`, an algorithm written by Nuno Barros for the SNO 3-phase Analysis []. This considers not only the neutrino oscillations through the vacuum of space between the Sun and Earth, but also the impact of matter effects in both the Sun and Earth. This can usually be a very computationally-intensive process, but `PSelmaa` takes advantage of the assumption that the solar oscillation parameters are in the so-called “Large Mixing Angle” regime, making the calculation much faster. As seen in Section 1.2.1, previous solar oscillation experiments demonstrate that this assumption is reasonable. For this analysis, the standard MSW effect is assumed with neutrinos obeying the Normal Hierarchy, with the Sun following the `B16_GS98` metallicity model [] and the `PREM` model being used for the Earth [].

One final thing `PSelmaa` needs to know to calculate survival probabilities is the distribution of solar zenith angles during the data-taking. The solar zenith θ_z is the angle between the two following vectors: one going from the centre of the Earth through the centre of the SNO+ detector, and another starting from the detector’s centre and pointing towards the Sun. As an example, if the Sun were ever to be directly overhead the detector, both vectors would be along direction \hat{z} in detector coordinates, leading to a solar zenith angle of $\theta_z = 0$. The position of the SNO+ detector on Earth, as well as the times at which the detector was live, determine the solar zenith angle distribution. If not accounted for, this can lead to a bias in the result of the analysis, as a preponderance of livetime taken at night (say) would lead to a larger fraction of solar neutrinos having to pass through the bulk of the Earth to get to the detector, and hence the impact of the MSW effect would be greater.

Even after using the Large Mixing Angle approximation, having to call PSelmaa numerous times for every step in the MCMC algorithm would lead to exorbitant run times for the fitting. Therefore, a further approximation is made. Before running the MCMC fit, PSelmaa is used to calculate P_{ee} over the necessary 3D space of parameters. To get a fine scan of this space, 101 E_ν values from 1 MeV to 20 MeV, 101 Δm_{21}^2 values from 3×10^{-6} eV² to 1×10^{-3} eV², and 151 values for θ_{12} from 5° to 65° were looked over. This 3D grid of $101 \cdot 101 \cdot 151$ P_{ee} values is then written to disk, and loaded into memory for use during the fit as a lookup table. At run-time, as the Metropolis-Hastings algorithm samples this 3D space the survival probability is estimated through a trilinear interpolation of the 3D grid loaded in: a version of linear interpolation for three dimensions [].

[19 pages currently, without figures; probably +5 pages for figures.]

6.2 Analysis on Scintillator-Phase data

6.2.1 Dataset and Livetime

- Description of dataset chosen for analysis: 2.2 g/L scintillator-phase data that satisfies the “gold” list of run selection requirements, between May and November 2022.
- Explain requirements for run being selected into the Gold list.
- Note ‘raw’ livetime calculated for this dataset, and then calculate the impact that the muon and high-nhit vetos have on the livetime.

[2 pages]

6.2.2 Event Selection

- List final set of cuts chosen for analysis, along with any explanations for cuts that haven't already been motivated earlier (e.g. data cleaning). These are:
 - Event index cut (prevents MC events that don't trigger detector from being used) 1
 - Data cleaning cuts 2
 - High-nhit event timing veto cut 3
 - Valid scintFit reconstruction 4
 - Reconstructed energy $2 < E < 14$ 5
 - Radial fiducial volume cut 6
 - BiPo out-of-window tag 7
 - BiPo in-window classifier cut 8
 - Externals classifier cut 9
 - “Cleanliness” cut 10
 - Position FOM cut 11
 - Show impact of cuts on data and MC. Show tables (the full details maybe in an appendix) indicating this. 12
 - Describe the finalised choice of binning for PDFs and data. 13
- [4 pages] 14
- 6.2.3 Expected Rates and Constraints**
- 15
- Show expected rates calculation for both signal and background. 16

- ¹ • Describe the constraints chosen to apply to the fit, and why they can be justified.

² These are:

- ³ – B8 flux constraint
⁴ – U-238 and Th-232 constraints from BiPo tagging
⁵ – Alpha-n constraint from Po-210 tagging by Serena and Shengzhao
⁶ – External constraints from Tony Zummo’s water-phase externals analysis.

- ⁷ • Describe the systematics to be added to the fit (just energy scale for now, maybe
⁸ also energy smearing?). For other possible systematics, such as those in position,
⁹ my aim is to explain why they are sub-dominant and so don’t need to be added
¹⁰ to the fit. Will cover more about impact of systematics in the next subsection.

¹¹ [4 pages]

¹² 6.2.4 Results

¹³ Fit Validation

- ¹⁴ • Show plots of parameter values versus step, to demonstrate that the step sizes
¹⁵ have been tuned sufficiently.
- ¹⁶ • Show auto-correlation plots, to motivate a sensible “burn-in” size.
- ¹⁷ • Posterior density plots for each nuisance parameter, to check that they all look
¹⁸ sensible and have sufficient statistics.
- ¹⁹ • Show plot of correlation coefficients between parameters, and note any strongly-
²⁰ correlated parameters.

²¹ [6 pages]

6.3 Sensitivity Projections**83****Oscillation Fit Results**

- Look at the data versus ‘best-fit’ MC plot in energy, radius, and both. (Recall that in MCMC, the ‘best-fit’ is not the point of parameter scape reached at the end, but the point of highest posterior density). Is there a good fit to data? Any clear disagreements?
- Show 2D contour plot for oscillation parameter posterior density. Note salient features. Show 1D posterior densities for each oscillation parameter. Derive measurement result for θ_{12} , i.e. point of highest posterior density, with uncertainties given by the 1σ credible interval.

[5 pages]

10

Impact of Systematics

- Show impact of modifying certain constraints on the final results of the measurement of θ_{12} . In particular: fiducial volume, ${}^8\text{B}$ flux constraint.
- Discussion of systematics post-fit — Hopefully energy scaling parameter should be close to 1, given the Collaboration’s calibration of the optics (light yield in particular). Perform a scan over energy smearing, and check impact of possible radial scale systematic.

[8 pages]

18

6.3 Sensitivity Projections

19

- Using the same production MC, generate PDFs with the expected normalisations for longer periods of livetime: 1, 3, and 5 years. This still assumes a scintillator-fill

20

21

1 with identical detector conditions on average, so not considering the impact of
2 BisMSB loading at any point.

- 3 • Describe any further constraints to be assumed on top of the existing analysis
4 of data: expected improved constraints on various backgrounds, as well as a
5 possible energy-scale calibration constraint from the AmBe source or an internally-
6 deployed source, for example.
- 7 • Run MCMC fits to these Asimov PDFs for each livetime scenario. Describe
8 results in terms of the improvement to the sensitivity to θ_{12} as a function of
9 livetime.
- 10 • Also consider scenario of lower backgrounds! How does that impact the result?
- 11 • Could consider scenario of BisMSB deployment, leading to substantially greater
12 light yield and hence energy resolution. Given time it would take for Production
13 MC to be generated for this scenario, I would likely have to come up with
14 something clever and quick to actually do this. There quite possibly won't be
15 enough time.

16 [8 pages]

17 [61 PAGES TOTAL]

Chapter 7

Conclusions and Suggestions for Future Work

- Say what has been achieved in this thesis! In particular:
 - Substantial improvement to the SMELLIE generator in terms of speed and dynamic range
 - A much stronger understanding of the discrepancies between data and MC in SMELLIE
 - The creation of two analyses of SMELLIE data, designed explicitly around being robust to these systematics
 - A measurement of the extinction length of scintillator *in-situ* with SMELLIE at 375 nm, monitored over time
 - A first measurement of the scattering length of the scintillator *in-situ*, monitored over time
 - The creation of an analysis of ${}^8\text{B}$ solar neutrinos in the scintillator phase to measure the solar neutrino oscillation parameters

- ¹ – The first measurement of θ_{12} using ^8B neutrinos in SNO+
- ² – Projections of this solar analysis' precision at longer livetimes
- ³ • Give suggestions for further work that could be done on both SMELLIE and the
⁴ solar oscillation analysis:
 - ⁵ – Inclusion of LABPPO's polarisability anisotropy in the detector's optical
⁶ scattering simulation, and determination of its impact on both Physics and
⁷ SMELLIE measurements
 - ⁸ – Further investigation of SMELLIE's wavelength-dependence of the beam
⁹ profiles, looking both at possible origins for the phenomenon and correcting
¹⁰ for this in simulation
 - ¹¹ – Various computational improvements to allow for faster MCMC run-times,
¹² including the calculation of the systematic matrices, and possible use of
¹³ GPUs to parallelise some parts of the computation.
 - ¹⁴ – Inclusion of additional solar neutrino components at lower energy into the
¹⁵ fit, e.g. Be7. Maybe the addition of lower energies also helps to naturally
¹⁶ constrain backgrounds within the fit?
 - ¹⁷ – Looking at the impact of various advanced background-rejection procedures,
¹⁸ such as event directionality or topology. How much do they help with the
¹⁹ sensitivity?
 - ²⁰ – Looking at the impact of splitting data into day and night parts, to try
²¹ and provide further constraints on any matter effects. Not expected to be
²² significant, so was ignored for this analysis so far.
 - ²³ – Performing a combined fit with the reactor anti-neutrino oscillation analysis,
²⁴ which allows for the handling of correlated uncertainties, such as detector
²⁵ response systematics.

[3 PAGES TOTAL]

1

References

- [1] SNO+, V. Albanese *et al.*, The SNO+ experiment, Journal of Instrumentation **16**, P08059 (2021), Publisher: IOP Publishing. 1
2
3
- [2] K. Majumdar, *On the Measurement of Optical Scattering and Studies of Background Rejection in the SNO+ detector*, DPhil Thesis, University of Oxford, 2015. 4
5
- [3] E. Turner, *A Measurement of Scattering Characteristics of the Detection Medium in the SNO+ Detector*, DPhil Thesis, University of Oxford, 2022. 6
7
- [4] S. S. Wilks, The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses, The Annals of Mathematical Statistics **9**, 60 (1938), Publisher: Institute of Mathematical Statistics. 8
9
10