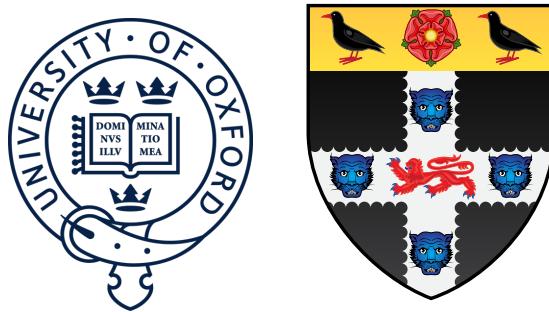


First Measurement of the Solar Neutrino Oscillation Parameters via Boron-8 Solar Neutrinos in SNO+



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

Draft - v0.1

Sunday 5th February, 2023 – 16:41

To my parents

Draft - v0.1

Sunday 5th February, 2023 – 16:41

Acknowledgements

And I would like to acknowledge ...

Draft - v0.1

Sunday 5th February, 2023 – 16:41

Abstract

This is where you write your abstract ...

Draft - v0.1

Sunday 5th February, 2023 – 16:41

Table of contents

List of figures	xi
List of tables	xv
1 The Theory of Neutrino Physics	1
2 The SNO+ Detector	3
3 Optical Scattering Theory	5
4 Simulating SMELLIE events	7
4.0.1 Previous attempts at SMELLIE event simulation	7
4.0.2 The new generator	10
4.0.3 Improving the beam profiles	15
4.0.4 Combining beam profile datasets	18
4.0.5 Results & Discussion	23
5 Solar Oscillation Analysis	29
5.1 Analysis Methodology	29
5.2 Analysis on Scintillator-Phase data	31
5.3 Sensitivity Projections	32
5.4 Conclusions	32

List of figures

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 [1].	9
4.2	Typical distribution of the number of attempts it takes for the existing 2D generator before the test direction gets accepted, per event.	11
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.	12

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.	14
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.	15
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.	22
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.	24
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.	25

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) . 27

Draft - v0.1

Sunday 5th February, 2023 – 16:41

List of tables

4.1 Water-phase runs used for new beam profiling.	17
---	----

Draft - v0.1

Sunday 5th February, 2023 – 16:41

Chapter 1

¹

The Theory of Neutrino Physics

²

Draft - v0.1

Sunday 5th February, 2023 – 16:41

Chapter 2

¹

The SNO+ Detector

²

Draft - v0.1

Sunday 5th February, 2023 – 16:41

Chapter 3

¹

Optical Scattering Theory

²

Draft - v0.1

Sunday 5th February, 2023 – 16:41

Chapter 4

Simulating SMELLIE events

4.0.1 Previous attempts at SMELLIE event simulation

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, the fibre emission positions were recorded during the installation of the fibres. The 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 is defined as the mean number of photons simulated per event; on an event-by-event basis we sample Poisson fluctuations about that mean value. Determination of the pulse magnitude must be done on a subrun-by-subrun basis. Unlike scintillation light, light from SMELLIE is

¹ not at all isotropic, and so we must specify some form of angular emission distribution.
² Determining and handling these angular emission distributions, also known as beam
³ profiles, is the focus of this chapter.

⁴ Before we can determine the beam profiles, we must first decide how to specify
⁵ them. Previous observations show that different fibres can have notably different beam
⁶ profiles [2], so we let each fibre's beam profiles be unique. We assume for now that a
⁷ given fibre's beam profile is stable over time, and independent of the wavelength of
⁸ light fired. A straightforward, naïve approach to parameterising a beam profile would
⁹ be as follows: specify some nominal fibre direction, corresponding to the direction light
¹⁰ takes travelling from the fibre to the centre of the “beamspot” observed on the other
¹¹ side of the detector. Then, specify a 1D beam profile, corresponding to the probability
¹² density of firing a photon at a given polar angle α relative to the nominal direction.
¹³ One might even assume this distribution is Gaussian. The distribution in azimuthal
¹⁴ direction, ϕ , is assumed to be uniform.

¹⁵ This 1D beam profile approach was used initially for SMELLIE, and remains in
¹⁶ use for the other ELLIE sub-systems within SNO+. However, when SMELLIE data
¹⁷ was taken in the water-phase of the experiment, simulations using these beam profiles
¹⁸ failed to match them well at all - see figure 4.1 for an example. Not only was the
¹⁹ distribution in α not Gaussian, a distinct speckle-pattern can be observed within the
²⁰ beamspot that is not uniform in ϕ . This fact led to colleague Esther Turner building a
²¹ SMELLIE generator that could handle 2D beam profiles: dependent on both α and ϕ .
²² The distribution was stored as a map from each inward-pointing PMT in the detector
²³ to a relative intensity value. This was chosen because the beam profile shapes were
²⁴ calibrated from existing SMELLIE data — more on this in section 4.0.3.

²⁵ This original 2D generator then sampled the beam profile via a rejection sampling
²⁶ approach, outlined as follows:

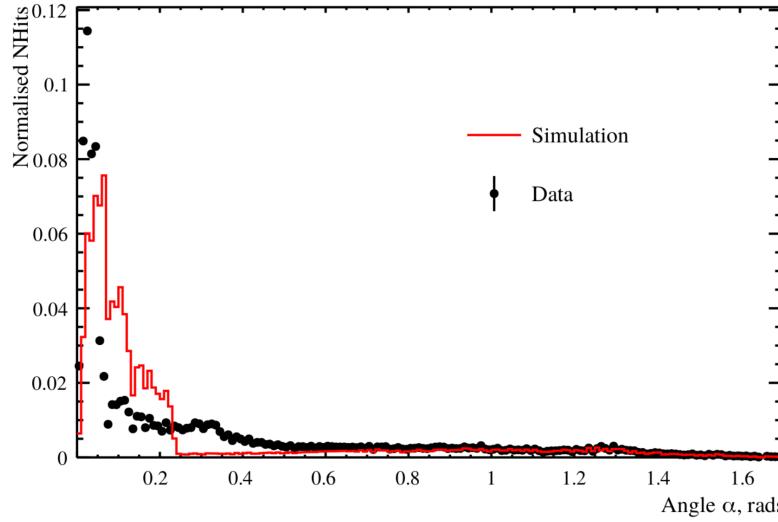


Fig. 4.1 Comparison between a simulation of one of the fibres, made from the 1D beam profile generator (red), with the associated data subrun that was used to create that beam profile (in black). For both MC and data, what is plotted is the PDF of observed PMT hits, as a function of the α angle. Poissonian 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 [1].

1. Propose a test direction (α, ϕ) , by generating ϕ uniformly in the interval $[0, 2\pi]$,
and α according to some pre-determined Gaussian distribution, known as the
Gaussian envelope.
2
2. Given this test direction, calculate where a line following this direction from the
fibre of interest will hit the PSUP on the other side of the detector. Find the
closest PMTs to that point.
3
3. From those PMTs, obtain their relative intensity values from the beam profile
mapping, and perform an interpolation based on how close each PMT is to the
PSUP intersection point. This gives an interpolated relative intensity value for
this test direction.
4
4. Because we are sampling using the angular coordinates (α, ϕ) , differential area
elements over this space of directions do not have the same size. We can correct
11
12

for this fact by multiplying our interpolated relative intensity by $\sin \alpha$, which corresponds to the Jacobian of the direction-space.

5. Calculate the value for the Gaussian envelope along this test direction.

6. Throw a random number uniformly between 0 and the Gaussian envelope value. If the random number is less than the interpolated intensity, then this test direction is accepted, and a photon is generated with that direction. Otherwise, we reject 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.

15 4.0.2 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

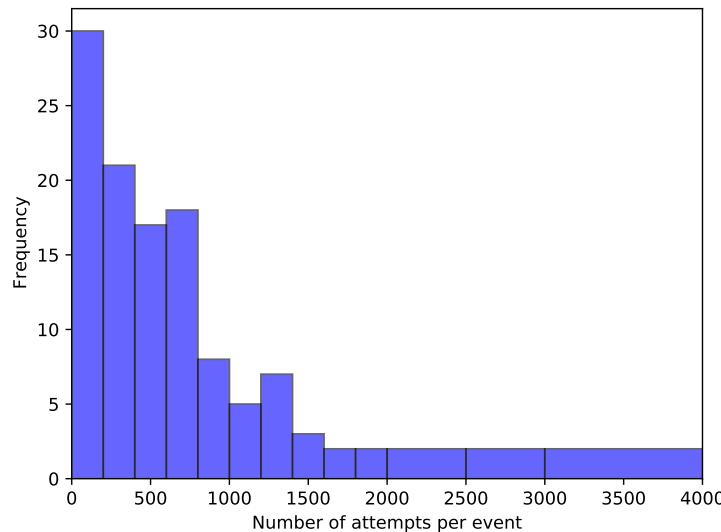
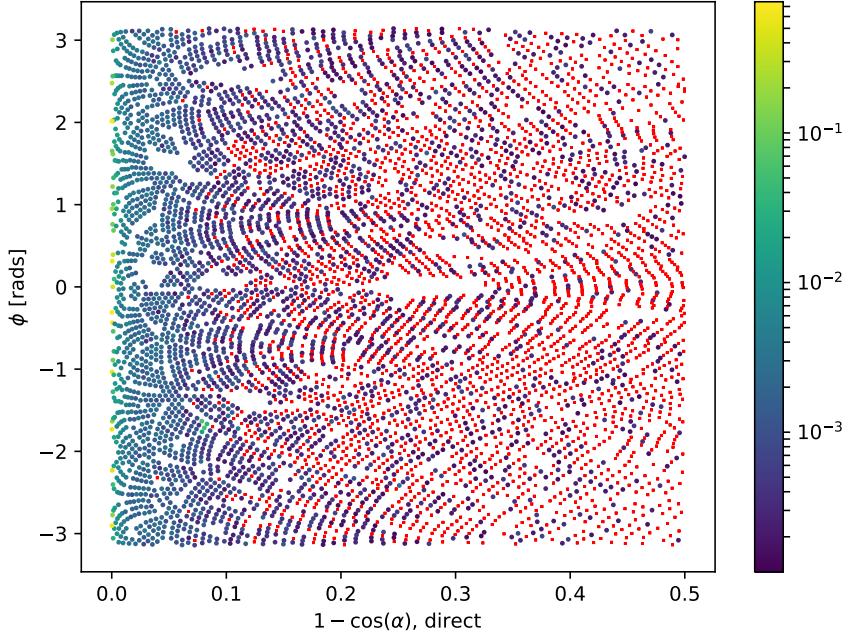


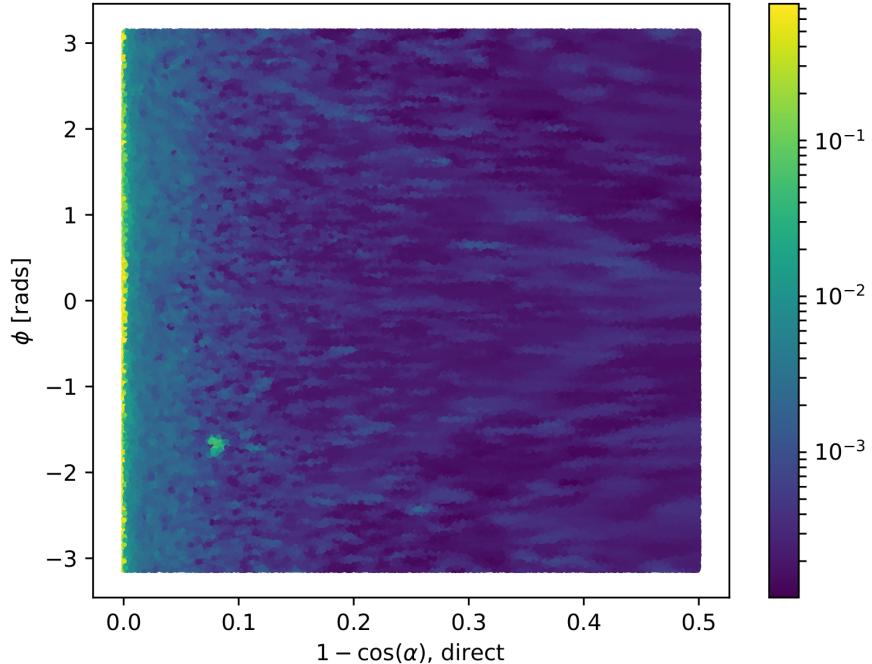
Fig. 4.2 Typical distribution of the number of attempts it takes for the existing 2D generator before the test direction gets accepted, per event.

try and “pre-calculate” as much as possible before run-time. Starting with the existing 1
PMT relative intensity maps, we plot these in the 2D direction-space $(1 - \cos \alpha, \phi)$: 2
see Figure 4.3a. In a toy-MC simulation, 500,000 directions are then thrown uniformly 3
in this 2D space per fibre. For each direction, the same method of obtaining an 4
interpolated intensity value from the nearest PMTs to the corresponding point on the 5
PSUP as from the original 2D generator was performed, the only difference being that 6
these calculations were done well before any actual SMELLIE simulation. Figure 4.3b 7
shows the interpolated intensities obtained for one fibre. 8

Following this, the sampled intensities were then binned into a 2D histogram, where 9
the bin value corresponds to the sum of all intensities for all directions found within this 10
bin. Choosing a sensible binning procedure is important: too few bins, and necessary 11
information about the shape of the beam is lost, whilst too many bins can oversample 12
the data and capture statistical artefacts in the sampling process instead of just the 13
beam profile. As a balance, 15 bins were chosen along the ϕ direction, and 60 in 14



(a) PMT relative intensity map



(b) Interpolated intensity map

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.

$r = 1 - \cos \alpha$. 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 matters greatly, and smooth out the very-low intensities seen at larger polar angles. One of these histograms can be seen in Figure 4.4: the large change in bin widths as a function of r is clear. One can also see that near the beamspot notable dependence on the intensity as a function of ϕ . The mysterious “spot” at $r = 0.08$, well out of the beamspot, is an indication that the underlying beam profile data being used requires improvement: more on this in section 4.0.3.

The Cumulative Density Function (CDF) of this intensity histogram as a function of bin was then produced, where the bins were ordered through a raster-scan: scanning first over ϕ , and then r . The CDF was then normalised to 1 so that it was well-defined. It is this CDF object that is then loaded in and sampled from during event generation. To do this, an “inverse-CDF” approach was used, which has the major benefit over rejection sampling of always producing a valid direction for every sample made. The algorithm works as follows:

1. Throw a random number uniformly in $[0, 1]$.
2. Perform a binary search to find the bin that has the largest CDF value below this random number.
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.

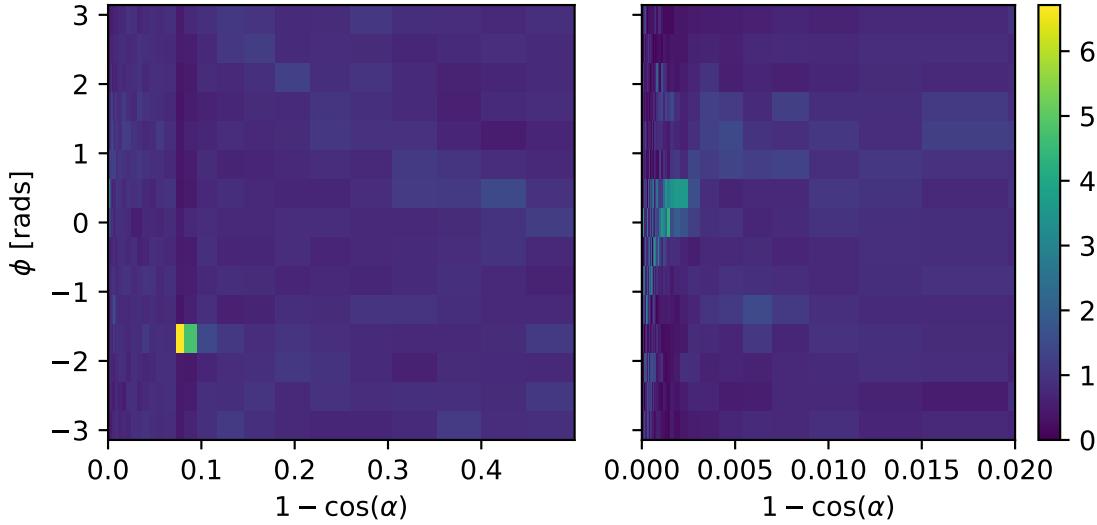


Fig. 4.4 Histogram of interpolated intensities within the 2D direction-space. The left view shows the full histogram; the right is a zoomed-in version near the beamspot. Unlike the binning in ϕ , the bin widths in r are not at all uniform. Instead, they have been determined such that the area summed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

- ¹ 4. Look at the selected bin’s r -bin edges, and select a value of r by throwing a
- ² second random number uniformly between the two edges. Convert this r into a
- ³ 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
- ⁶ generator, the speed improvement was very large: generation now took ~ 1 ms per
- ⁷ SMELLIE event, a speed improvement of nearly 50,000. Event generation became
- ⁸ as fast as it was when the 1D generator was being used. Furthermore, because of
- ⁹ the approach taken, this major speed improvement comes at no sacrifice in accuracy.
- ¹⁰ Figure 4.5 shows a comparison of the average number of photoelectrons (npe) per event
- ¹¹ per PMT between water-phase SMELLIE data and simulations with both the old and
- ¹² new 2D generator. One can see clearly that both generators are as accurate as one

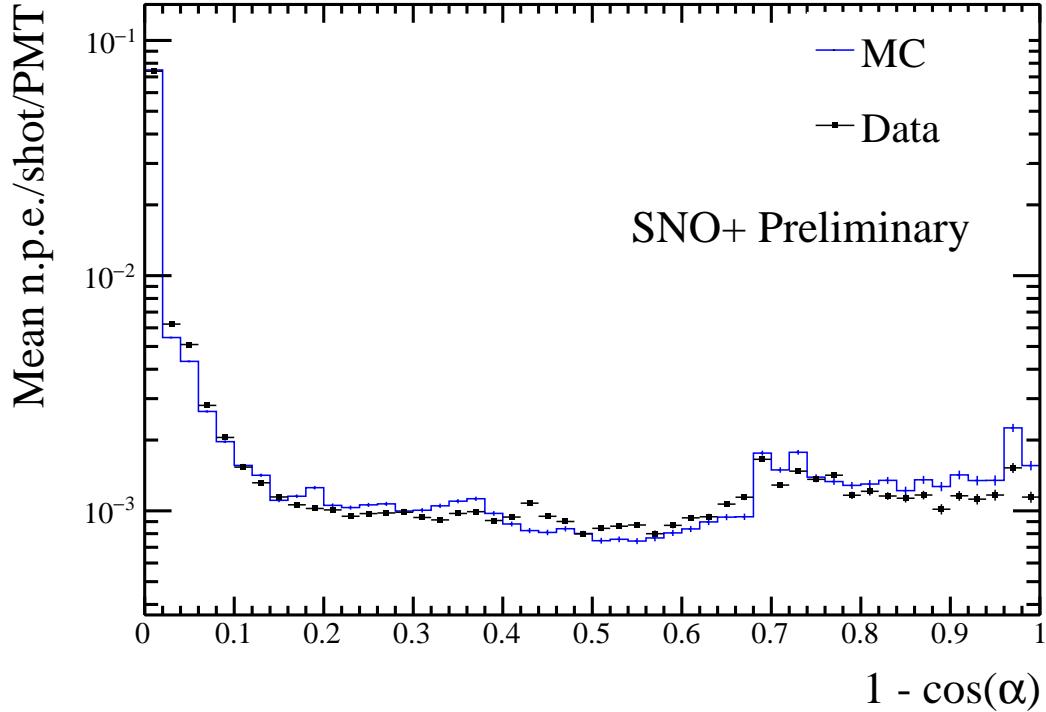


Fig. 4.5 Comparison of water-phase data to MC generated using both the old and new 2D beam profile generator approaches, with the updated beam profiles. Both versions of the generator are consistent with one another, but the new generator is many times faster.

another. Note that this plot uses the updated beam profiles as explained in the next section.

4.0.3 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 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 [1] that with the old

¹ 2D generator, the systematic uncertainty on the beam profiles was the dominant source
² of error in the main SMELLIE analysis. To help improve this situation, it was decided
³ to update the existing beam profiles.

⁴ These old beam profiles were originally determined by looking at SMELLIE data
⁵ taken during the water-phase. Specifically, a “medium”-intensity subrun with one
⁶ of the lasers firing at a wavelength of 495 nm, was chosen for each fibre. “Medium”-
⁷ intensity corresponds to firing the relevant laser at a set intensity determined during
⁸ an earlier commissioning process, for which the maximum occupancy of PMT hits at
⁹ that intensity, i.e. the proportion of hits per event, corresponded to roughly 80%. This
¹⁰ value was chosen as it allowed for high statistics in a relatively short run-time, but not
¹¹ so intense that the occupancy of any given PMT in the beamspot was 100%. Because
¹² Rayleigh scattering is strongly-dependent on wavelength, the long wavelength of light
¹³ was chosen so that impacts from this scattering were small in the data.

¹⁴ SNO+ PMTs are unable to distinguish the exact number of photoelectrons being
¹⁵ generated. One is typically only able to know if a PMT has been triggered at all, by
¹⁶ any number of photoelectrons. As a result, the occupancy of a PMT over a number
¹⁷ of SMELLIE events, o , is a biased estimator of the mean number of photoelectrons
¹⁸ generated, μ . Assuming the number of photoelectrons generated in a given event
¹⁹ follows Poisson statistics, the probability of generating k photoelectrons is:

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

²¹ The probability of observing a “hit” in a given PMT corresponds to generating at
²² least one photoelectron:

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

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.

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 4.0.2.

4.0.4 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 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 $\alpha < 3^\circ$. It was found that a “loose” time residual cut of $t_{res} \in [-10, +12]\text{ns}$ was

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.

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

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

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

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

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

¹ binomially-distributed:

$$P(m_{ij}|\mu_{ij}) = L(\mu_{ij}|m_{ij}) = \binom{N_j}{m_{ij}} p_{ij}^{m_{ij}} (1-p_{ij})^{N_j-m_{ij}} = \binom{N_j}{m_{ij}} (1-e^{-\mu_{ij}})^{m_{ij}} e^{-\mu_{ij}(N_j-m_{ij})}. \quad \blacksquare$$

²

(4.5)

³ Here we have used equation 4.2, and noted that this probability distribution in m can
⁴ be re-framed as a likelihood function for the parameter μ_{ij} . Considering only a single
⁵ subrun of data, the maximum likelihood estimate of the parameter μ_{ij} can be shown
⁶ to be:

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

⁷

(4.6)

⁸ where o_{ij} is just the occupancy of PMT i in subrun j . This is just the multi-hit
⁹ correction formula seen in equation 4.3, which makes sense.

¹⁰ When looking at multiple subruns for the same fibre, the total likelihood function
¹¹ for a given PMT when considering all the data for a given fibre will be the product of
¹² 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

(4.7)

¹⁴ This leads to a log-likelihood distribution of

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

¹⁵

(4.8)

¹⁶ Formally, one could combine the likelihoods of all the PMTs together, and by looking
¹⁷ at the maximum likelihood estimates for each of the parameters measure the parameter
¹⁸ values this way. However, the set of equations one obtains through this approach
¹⁹ quickly become analytically intractable, because the PMTs are coupled by the intensity
²⁰ values I_j . Even a direct numerical approach would be liable to fail: for a given fibre

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

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

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

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

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

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

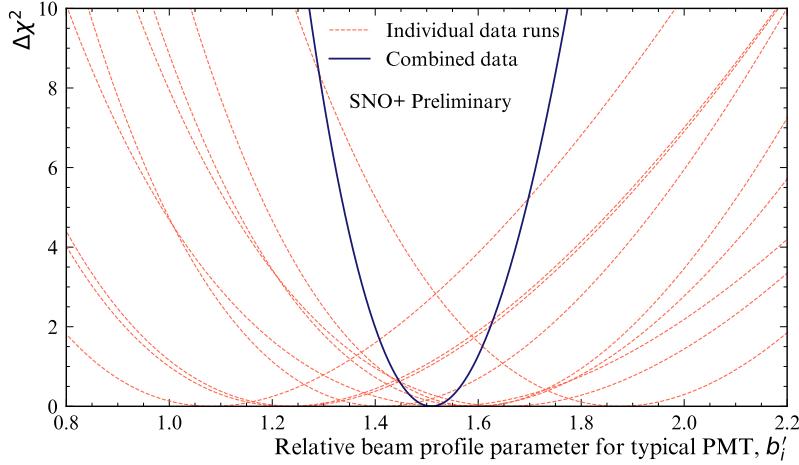


Fig. 4.6 Plot of $\Delta\chi^2 \simeq X_i$, twice the negative log-likelihood ratio, for both single subruns of a typical PMT, and when all relevant subruns are combined.

¹ equation 4.4, we find that:

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

³ As a result of this, given the set $\{S_j\}$ and k_i^{MC} , one can maximise the log-likelihood
⁴ \mathcal{L} with respect to $b'_i = acb_i$ numerically, to obtain the maximum likelihood estimate
⁵ of b'_i . Because a and c were global constants of proportionality, they would become
⁶ irrelevant as soon as the beam profile was normalised in the CDF-creation process
⁷ outlined in 4.0.2.

⁸ Figure 4.6 shows the shape of this log-likelihood distribution for a particular PMT
⁹ when considering fibre FS007's beam profile. One can see how individual subruns
¹⁰ provide much more information when combined than if one looked at a single subrun
¹¹ 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 [3] allows us to produce 1σ confidence intervals about the maximum likelihood

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

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

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

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

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

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

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

where $\text{erf}(x)$ is the error function.

4.0.5 Results & Discussion

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

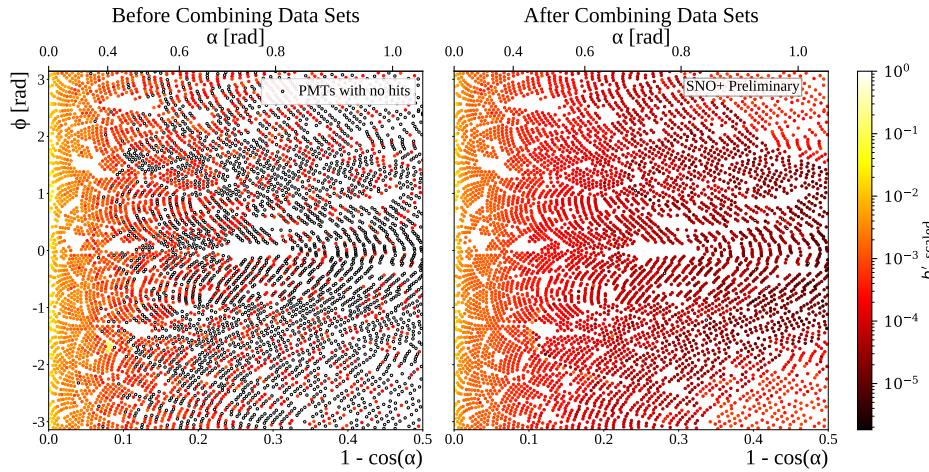


Fig. 4.7 Comparison between old and updated beam profiles for fibre FS055, after combining multiple data sets. Once again, the relative intensities (b'_i) for each PMT are given by the colour of each point, the position of each plotted in the 2D (r, ϕ) -space. The relative intensities have been both 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.

¹ in data. That many more data sets were included allowed for the major increase in
² dynamic range available for measuring these b'_i values. One can also note that by
³ including additional data the curious spot that was seen in the old beam profile our at
⁴ $r \approx 0.08$ has gone, further indicating that it was an artefact of that single data set.

⁵ Further details can be gathered from the interpolated intensity maps, one of which
⁶ can be seen in figure 4.8. There are two curious stand-out features that can be seen here:
⁷ firstly, there are multiple distinct parabolic arcs. These correspond to the shadows of
⁸ the ropes that hold up/down the AV. More precisely, they are the mismodelling of
⁹ those shadows — if the shadows were in the right place in the isotropic MC, then they
¹⁰ would correctly cancel out any decreased intensity seen in the data of shadowed PMTs.
¹¹ These shadows could be mismodelled either because the positions of the ropes in the
¹² MC are in the wrong place, or the fibre's emission position is wrong. Note that any
¹³ mismodelling of the fibre's nominal emission direction has no impact on this shadowing
¹⁴ problem, as changing that direction merely causes a change of basis in the (r, ϕ) -space.

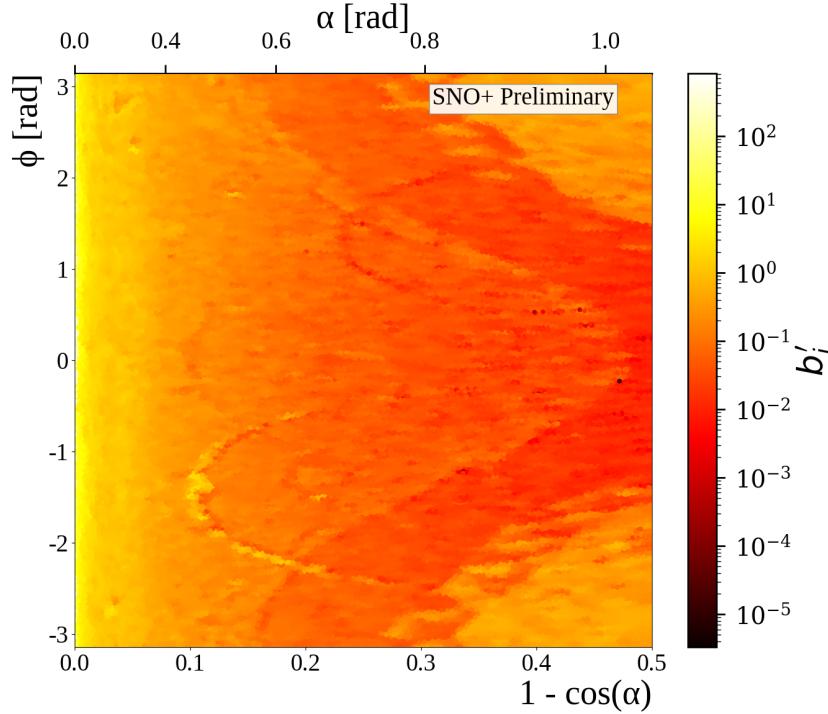


Fig. 4.8 Interpolated intensity map for the new updated beam profile of fibre FS055. The misalignment of rope shadows and AV effects, can both be seen.

The latter possibility of incorrect fibre positions are more likely, and in fact these arcs in the beam profiles could be used as an effective way to correct for this problem.

The second distinctive feature of this intensity map is the large band of lower intensity varying between $r \approx 0.2 - 0.5$, followed by larger intensity out at large r values. This feature comes from light reflecting off the AV surface, or internally-reflecting. The reason for this band's functional dependence on ϕ is that this particular fibre, FS055, has a nominal fibre direction $\sim 10^\circ$ from pointing radially-towards the detector's centre. This feature appears in the updated beam profiles of all fibres, but its shape depends on the particular fibre's direction — for fibres pointing directly towards the detector's centre, there is little ϕ -dependence observed. Like the ropes, this feature must come from some form of mismodelling of the optics of the AV. A de-facto shadowing of PMTs in line with tangents from the AV surface which intersect

1 the fibre position is to be expected. One also expects PMTs at polar angles larger than
2 this to have their observed intensities boosted from reflected light off the AV. However,
3 the discontinuities seen in the beam profiles indicate that for whatever reason this
4 effect has been over-emphasised in the simulation.

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

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

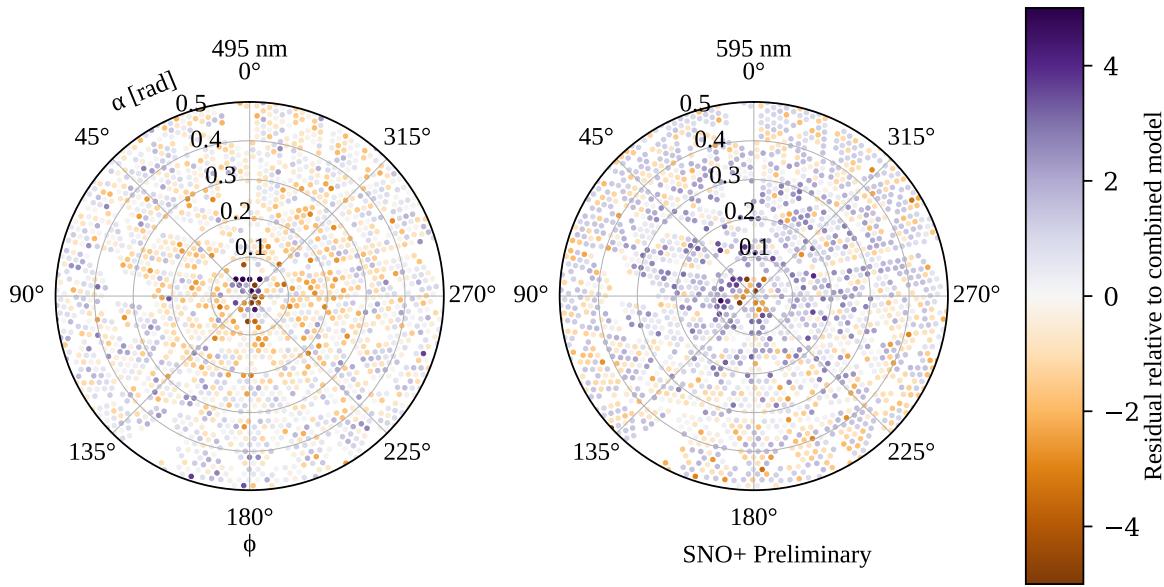


Fig. 4.9 Residuals from subruns at two different wavelengths, both compared to the combined beam profile model for fibre FS055. A negative sign, and hence bluer colours, indicate that the combined model underestimates the observed intensity for that particular subrun. Values with a magnitude beyond 5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

distribution. Certainly if one wanted to further improve the uncertainties in the beam profiles, tackling these challenges would be key.

Draft - v0.1

Sunday 5th February, 2023 – 16:41

Chapter 5

Solar Oscillation Analysis

5.1 Analysis Methodology

- Start with observational principle: changes in oscillation parameters lead to change in the energy spectrum of neutrinos elastically-scattering in the detector, and hence a change in the observed energy spectra of elastically-scattered electrons within the detector. Demonstrate basic impact on modifying Δm_{21}^2 and θ_{12} .
- Want to maximise the sensitivity to measuring these two parameters. Energy spectrum of signal is background-free above ~ 5 MeV, but rate is substantially larger at lower energies. If one can minimise backgrounds at the lower energies, then there should be hopes of obtaining a measurement with greater precision!
- Set up analysis approach: a Bayesian analysis using MCMC. Explain why this was chosen at a high level first: allows for us to perform a relatively complex analysis with multidimensional PDFs, numerous backgrounds, systematics, constraints on parameters, all whilst allowing us to obtain well-defined measures of uncertainty on our final results.

- 1 • Test statistic: binned extended log-likelihood. Explain why this is fundamentally
2 the “right” test statistic to use. Allows for handling of constraints and systematics.
- 3 • Give an overview of how an MCMC works. Key idea: exploring parameter space
4 in such a way as to reproduce posterior density distribution. Clever! Helps to
5 avoid “curse of dimensionality” that often arises in fits with numerous parameters.
6 Must also explain how we work in a fundamentally Bayesian, not frequentist,
7 statistical approach.
- 8 • Explain background processes we’ll be dealing with: can be distinguished by
9 event topology, energy spectrum, and position distribution.
- 10 • Explain cuts that I am using for the analysis. Demonstrate how these attempt to
11 maximise our signal sensitivity. Leads nicely into choice of 2D fit in both energy
12 and radius (cubed).
- 13 • Describe implementation of neutrino oscillations within the fitting procedure:
14 flat priors on the oscillation parameters, but a strong constraint from the solar
15 flux. Explain choices of constraint that are possible. MCMC varies oscillation
16 parameters and flux scaling factor, which then modifies the solar signal PDFs
17 through a de-facto systematic that is a function of true neutrino energy, a
18 3rd “bookkeeping” dimension in the signal’s PDFs only. Neutrino oscillations
19 simulated via PSelmaa, which accounts for MSW effect in both the Sun and the
20 Earth. For computational speed, at run-time we actually use a lookup table with
21 linear interpolation for the survival probability as a function of parameters.
- 22 • Implementation of systematics: we handle them generally as linear transforma-
23 tions acting on the vector of bin data. Clever! Which systematics do we expect
24 to be particularly important for this analysis? Well, mismodelling in detector
25 optics etc. can lead to changes in the measured energy spectrum of processes,

which can be decomposed into an energy scale term and an energy smearing term to first order. Systematics also possible in the radial dimension (expected to be less important?)

5.2 Analysis on Scintillator-Phase data

- Description of dataset chosen for analysis: full-scint data that satisfies the “gold” list of run selection requirements, between 1st June 2022 and March 2023. Starting date chosen to ensure radon levels have stabilised within the centre of the detector.
- Impact of cuts on data and MC. Show tables (the full details maybe in an appendix) indicating this.
- Describe the constraints chosen to apply to the fit, and why they can be justified.
- Running & validation of MCMC fitting. Show plots of parameter values versus step, to demonstrate that the step sizes have been tuned sufficiently. Show auto-correlation plots, to motivate a sensible “burn-in” size. Show posterior density plots for each nuisance parameter, to check that they all look sensible and have sufficient statistics. Show plot of correlation coefficients between parameters, and look at any correlations that are particularly interesting.
- Look at the data versus MC plot in energy, radius, and both. 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} .
- Show impact of modifying certain constraints on the final results of the measurement of θ_{12} .

1 5.3 Sensitivity Projections**2 5.4 Conclusions**

References

- [1] E. Turner, *A Measurement of Scattering Characteristics of the Detection Medium in the SNO+ Detector*, DPhil Thesis, University of Oxford, 2022. 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] 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. 6
7
8