A log-likelihood fit for extracting neutrino oscillation parameters

Andreas Vrikkis, 10 December 2021

***Abstract*—Optimised neutrino oscillation parameters were determined using simulated muon neutrino detections. By constructing a negative log-likehood fit between the observed and expected detection counts, estimators for the neutrino oscillation parameters θ_{23} and $\Delta\mathbf{m}^2_{23}$ were extracted from the minimum point of the fit. Three distinct numerical minimisation methods were tested and deployed: Univariate method, Newton method and Simulated Annealing. A subsequent update of the fit that included the free parameter α governing the rate of increase of the neutrino cross section with energy was minimised using the same methods. A discussion of the methods' accuracy, errors and reliability is presented. Simulated Annealing estimated the parameters with the lowest negative log-likelihood of the methods investigated. These were $\theta_{23} = \mathbf{0.7855^{+0.03654}_{-0.03674}rad}$, $\mathbf{\Delta m^2_{23} = 251.2^{+0.3912}_{-0.2576} KeV^2}$ and $\alpha = \mathbf{1.020^{+0.05271}_{-0.04485} eV^{-1}}$.**

I. Introduction

EXPERIMENTAL evidence of neutrino oscillations have commenced a vast new program of research seeking to answer questions regarding neutrino flavour masses and their role in the universe. It prompted the first major change to the standard model in the last twenty years allowing for the purely quantum mechanical description in which neutrinos oscillate between flavours and have a non-zero mass [1]. Due to the extremely long weak force interaction length, even the most sensitive neutrino detectors observe statistically limited numbers of neutrino interactions. To mitigate this limitation, statistical techniques have been largely employed which require computational minimisation algorithms to obtain results to a high degree of accuracy. The robustness and unique approaches of each method can significantly affect the results and their accuracies.

II. Theory

The probability that a muon neutrino ν_μ will be observed at its original flavour and will not have oscillated into a tau neutrino after it travelled a distance L(km) is given by

$$P\left(\upsilon_\mu \longrightarrow \upsilon_\mu\right) = 1 - \sin^2\left(2\theta_{23}\right)\sin^2\left(\frac{1.267\Delta m^2_{23}L}{E}\right) \quad (1)$$

where E is the neutrino energy in GeV, θ_{23} is the "mixing angle" and Δm^2_{23} is the difference between the squared masses of the neutrinos in each flavour [1]. Figure 1 illustrates how this survival probability's behaviour is governed by the values of the oscillation parameters.

Extraction of the oscillation parameters is possible through a statistical comparison between the counts observed at a known distance L and the expected number of detections after applying the survival probability. For this project, simulated data from the T2K experiment acted as the observation data at $L = 295$km. Figure 2 depicts the observed muon neutrino counts and the expected counts, also obtained through simulations, when the survival probability is neglected. The unoscillated expectation bins were then multiplied by the probability 1, taking the mid-point of the energy range in each bin as the energy E in the probability. To transform the extraction into a minimisation problem, the Negative Log-Likelihood NLL of the observation counts m_i and expected counts λ_i was constructed. It depended on the oscillation parameters $\vec{u} = (\theta_{23}, \Delta m^2_{23})$ and is given by

$$\mathrm{NLL}(\vec{u}) = \sum_{i=1}^{n}\left[\lambda_i(\vec{u}) - m_i + m_i \ln\left(\frac{m_i}{\lambda_i(\vec{u})}\right)\right]$$

using stirling's approximation and following a Poisson distribution for each bin [2]. The $\vec{u}_{min}$ that minimises the NLL represents the best fit for the data and expected counts, thus acting as estimators for the true parameter values.

For any negative log-likelihood distribution, it can be shown that the inherent experimental uncertainty is the range which

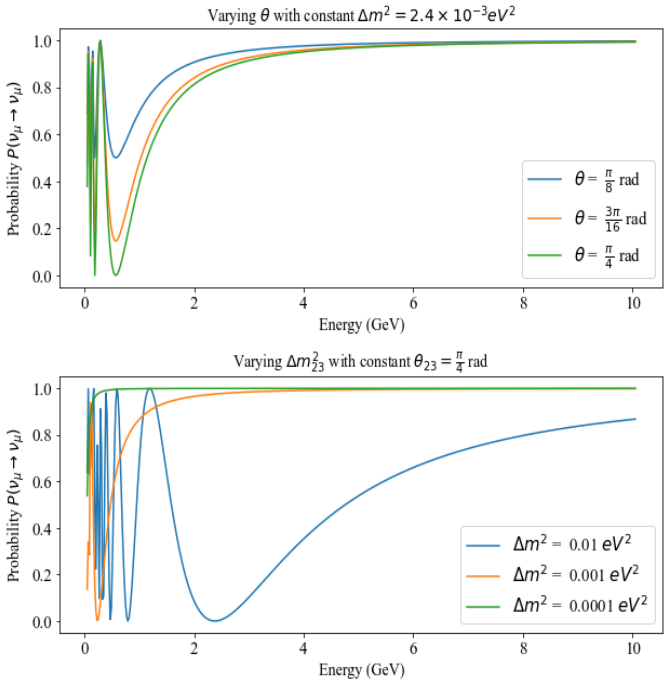

Fig. 1. The survival probability against muon neutrino energy. On the x-axis, the energy is depicted for the range 0 to 10 GeV, the y-axis shows the survival probability. The top plot depicts the case where Δm^2_{23} is kept constant, while the bottom plot depicts the case where θ_{23} is fixed. It can be observed that varying θ_{23} greatly affects the low-energy neutrino survival probability in the range of 0 to 4 GeV, as the depth of the widest minimum changes significantly. Varying Δm^2_{23} by an order of magnitude seems to affect almost all of the energy ranges.

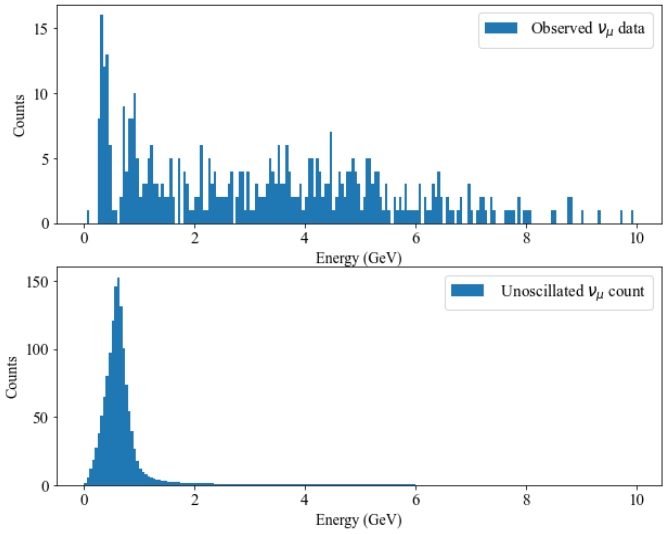

Fig. 2. The observed T2K muon neutrino detections on the top, expected number of unoscilatted detection on the bottom. Both distributions are plotted as histograms with 200 energy bins, each representing neutrinos with energies in successive ranges of 10 GeV. It is plain to deduce based on the disagreement of the plots that omitting the survival probability does not yield an accurate prediction, further fueling the evidence for the existence of neutrino oscillations.

makes the log-likelihood change by $+\frac{1}{2}$. This also holds for non-Gaussian likelihoods [3]. Generally, this can require a different shift in dependant variable when going up compared to down, yielding asymmetric uncertainties. However, this uncertainty does not reflect the numerical error occurring due to numerical methods utilised.

Neutrino Interaction cross-section

The expected rate of neutrino events is also linearly dependent on the neutrino interaction cross-section. Since the cross section increases approximately linearly with energy, the expected number of neutrino events were scaled according to

$$\lambda_i^{\text{new}}(\vec{u}) = \lambda_i^{\text{old}}(\vec{u}) \cdot \alpha \cdot E_\nu \tag{2}$$

where α is a new free parameter that governs the rate of increase of the cross-section with neutrino energy E_ν [2]. Thus, the NLL was also minimised by optimising all three parameters $[\theta_{23}, \Delta m^2_{23}, \alpha]$ in 3-dimensional minimisations.

III. Numerical Methods

To minimise the NLL, three methods were employed: Univariate, Newton and Simulated Annealing in both the 2-parameter(2D) and 3-parameter (3D) case.

Univariate Method

The univariate method incorporates a 1-dimensional minimiser deployed successively in each dimension until it spirals to the minimum [4]. In this investigation, this was a parabolic search function that approximated the NLL around the minimum as a parabola. In the case where θ_{23} was to be optimised, the parabolic search function was supplied with an NLL(θ) function that was programmed to keep the other parameter Δm^2_{23} constant. It was further supplied with two starting locations θ_1 and θ_2, with θ_3 being the midpoint of the two. The NLL values at each location were then fitted with a 2nd-order Langrange polynomial. The location of the mininimum of this fitted parabola was calculated using quadratic interpolation, along with its corresponding NLL value. These values were stored in a list and θ_{23} with the highest NLL was dropped, iterating this procedure until the change in minimum value was less than 1×10^{-7}. An equivalent process was deployed for the other parameters. The minimiser was tested on a negative gaussian

$$g(x) = -e^{-(x-5)^2} - 2e^{-(x-3)^2}$$

which has a local minimum at $x = 5$ and a global minimum at $x = 3$. It was established that if the starting point was sufficiently close to the true minimum, it correctly returned $x_{min} = 3$ at the accuracy of the threshold. Otherwise, it was highly susceptible to local minimima pitfalls.

The univariate algorithm cyclically minimised each parameter at each step while keeping the other parameters constant at their latest estimate - their value at the previous step. To automate the parabolic function supply process, a guess to the minimum was determined at each iteration by obtaining a list of NLL values for a range of the parameter currently mimised. The smallest value of the list acted as the first guess, with the second value being minimally displaced from the first. The univariate method was given a starting point and minimised successively until the change in estimates was less than a unique threshold in each dimension based on its parameter order of magnitude. In the 3D case, the minimum was obtained by minimising sequentially through all three paramaters $[\theta_{23}, \Delta m^2_{23}, \alpha]$ in a clockwise or anticlockwise manner.

Newton method

This method considers the local curvature at each step to simultaneously minimise both variables in $\vec{u} = (\theta_{23}, \Delta m^2_{23})$. By Taylor expanding around the function at each point it can be shown that the estimate of the minimum $\vec{u}$ follows the iterative relation

$$\vec{u}_{n+1} = \vec{u}_n - [\mathbf{H}(\vec{u}_n)]^{-1} \cdot \vec{\nabla} f(\vec{u}_n) \tag{3}$$

where $H_{ij}(\vec{u}) = \frac{\partial^2 f(\vec{u})}{\partial\theta_i \partial\Delta m_j}$ is the Hessian matrix [5]. To obtain the Hessian entries and the gradient vector, the derivatives were computed using the central difference scheme with a stepsize of 1×10^{-4}. The minimum value was then iterated according to equation 3 until changes in estimates were less than 1×10^{-9}. This convergence threshold was a compromise between iteration time and accuracy. The algorithm was ran for different starting points and was executed the same way in the 3D case.

Similarly to the previous method, validation was provided through the use of a testing function. The 2-dimensional function

$$t(x, y) = -sinc(100\sqrt{x^2 + y^2}) \tag{4}$$

was chosen as it features a global minimum at $x = y = 0$, encompassed by multiple minima and saddle points that the algorithm could drop into. It was tested for various starting points and finite difference step sizes. Analogously to the univariate method, it tended to fall into local minima, but was

less prone to do so if the starting point was sufficiently close to the true minimum. A contour plot for the NLL minimum also confirmed the accuracy of this method.

Simulated Annealing

This method introduces randomness in its search, enabling it to 'jump' out of local minima. The method takes the function f to be minimised and starts at given starting parameters [5]. A proposal function was supplied with the current parameter values and returned new random positions according to a normal distribution with mean at the current value. The normal's standard deviation was set to $starting\ value \times 1\%$ in each dimension, chosen by trial and error, ensuring sufficient variation in the proposed values while still keeping them statistically dependant. The proposed parameters were then accepted or rejected according to the acceptance probability

$$p_{\text{acc}} = \begin{cases} 1 & \text{if } \Delta f(\vec{u}) \leq 0 \\ \exp\left(-\Delta f(\vec{u})/T\right) & \text{if } \Delta f(\vec{u}) > 0 \end{cases}$$

where $\Delta f(\vec{u}) = |NLL(\vec{u}_{current}) - NLL(\vec{u}_{proposed})|$ and T is the "temperature" variable. As T gets smaller, the probability of accepting a proposal decreases, as does the overall range of values. The algorithm ran over 12 temperatures from $T = 3.00$ to $T = 0.25$, with 5×10^5 iterations in each temperature. The value and error of the minimum was taken to be the average and standard deviation respectively over the last 8 temperatures, so as to avoid biasing errors that may occur in the first stages of the algorithm, which are much more accepting that the last.

Validation was provided through the use of the testing function defined in (4). Multiple starting points, iteration number and temperatures were tested - most of which consistently returned the global minimum within the standard deviation.

NLL Error

To extract the log-likelihood error at any parameter position, a function was coded that took the parameter values and returned the errors in both directions by solving a root-finding problem. Specifically, for a minimum set of parameters $\vec{u}_{min}$ the equation

$$NLL(\vec{u}_{\pm}) - NLL(\vec{u}_{min}) - \frac{1}{2} = 0 \tag{5}$$

was defined in the method. By subsequently applying the bisection method, the roots $\vec{u}_{\pm}$ of equation 5 were the values at which the NLL was shifted by $\frac{1}{2}$. Subtracting the upper location $\vec{u}_{+}$ with the minimum $\vec{u}_{min}$ then gave the standard deviation σ_{+}, and similarly on the other side the lower deviation σ_{-}. The bisection method was given a threshold of 1×10^{-4} to act as the minimum amount of change of the last 2 estimates, defining convergence. The bisection method was always validated by a checking mechanism ensuring the change in NLL was indeed equal to $+\frac{1}{2}$ at the level of significance of the convergence threshold.

IV. Results

Univariate Method

Figure 3 depicts the NLL plot as a function of θ_{23} with a fixed trial $\Delta m^2_{23trial} = 2.4 \times 10^{-3} eV^2$. This plot and following discussion is representative of the algorithm execution at each minimisation. The value of the minimum was calculated to be $\theta_{trial} = 0.7664^{+0.009551}_{-0.01073} rad.$

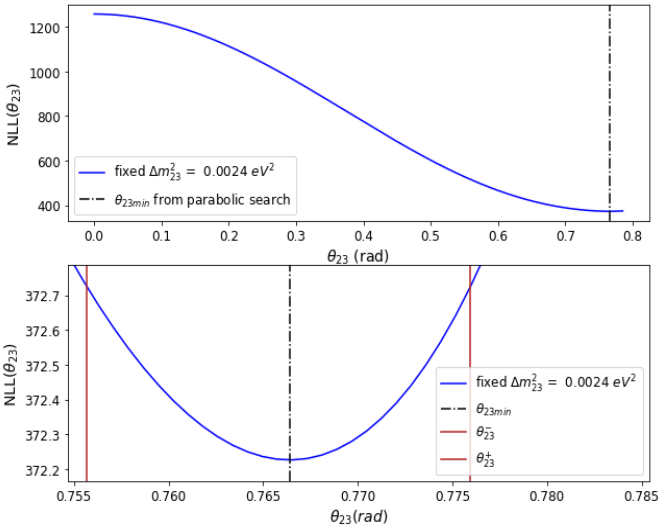

Fig. 3. Plot of NLL as a function of the parameter θ_{23} for fixed trial value of $\Delta m^2_{23} = 2.4 \times 10^{-3} eV^2$. The botton plot shows a magnified section of the NLL to highlight its parabolic nature around the minimum. θ_{trial} is also plotted as the black vertical line, bounded by the standard deviation locations at each side plotted as red vertical lines.

With a starting point at the trial values $\theta_0 = \frac{\pi}{4} rad$ and $\Delta m^2_0 = 2.4 \times 10^{-3} eV^2$, the univariate method estimated the minimum at

$$\theta_{23}^{univariate} = 0.7676^{+0.01129}_{-0.01125}\ rad$$
$$\Delta m_{23}^{univariate} = 2.389^{+0.003183}_{-0.002955} \times 10^{-3} eV^2$$

The corresponding value of NLL was

$$NLL^{univariate} = 372.18 \pm 0.5$$

after 7 iterations. This result was obtained by minimising θ_{23} first. Minimising Δm^2_{23} first gave an estimate with a difference that was negligible compared to the standard deviation. This could be because of the rather strict convergence threshold of 1×10^{-7} in the change of parameter values which forced the algorithm to iterate until it reached that point.

Univariate Error

In the case of the univariate method, the numerical error Σ associated with the parabolic search minimum value can be estimated using the curvature of the last parabolic estimate according to

$$\Sigma^2 = -\frac{1}{\frac{d^2 NLL(\theta_{23})}{d\theta^2_{23}}\Big|_{\theta_{min}}} \tag{6}$$

and similarly for Δm^2_{23} [3]. The analytic form of the 2nd-order Langrange polynomial with the 3 latest parabolic minimiser estimates was used to obtain the exact value of $\Sigma_{\theta_{23}} = 1.03 \times 10^{-5}$ for θ_{23} and $\Sigma_{\Delta m^2_{23}} = 1.11 \times 10^{-7}$ for Δm^2_{23}. These values are both 2 orders of magnitude smaller than the uncertainty inherent to NLL and are insignificant to the 4 significant figures used to quote the results. Therefore

they were omitted from the consideration. The narrowness of the numerical error is largely due to the strict convergence criterion set for the parabolic minimiser, which forced the latest estimates to be sufficiently close that their NLL values formed an approximately horizontal line. Consequently the curvature in equation 6 is greatly reduced as does the error.

The NLL uncertainty will become smaller on the positive side and larger on the negative side as θ_{23} tends to $\frac{\pi}{4}$, as observed in figure 4.

Newton Method

The same starting point $\theta_0 = \frac{\pi}{4} rad$ and $\Delta m^2_0 = 2.4 \times 10^{-3} eV^2$ was applied, with a stepsize of $h = 1 \times 10^{-4}$ for the finite difference approximation derivatives. Setting the convergence threshold to 1×10^{-9}, the method estimated the minimum as

$$\theta_{23}^{Newton} = (0.7853 \pm 0.03841)\ \ rad$$
$$\Delta m_{23}^{Newton} = 2.393^{+0.005350}_{-0.06310} \times 10^{-3} eV^2$$

with an NLL value of

$$NLL^{Newton} = 373.27 \pm 0.5$$

after 9 iterations. More testing suggested that enlarging the convergence threshold reduced the number of iterations without significantly affecting the result.

Newton Errors

The primary numerical error arises in the derivative calculations both in the Hessian entries and the gradient operator. Each finite difference scheme derivative carries a truncation error of $O(h^2)$ [4]. Writing

$$\tilde{\mathbf{H}} = \mathbf{H} + e_H$$

where e_H is the numerical error of the Hessian and is of $O(h^2)$ allows to calculate the total error by multiplying by the number of iterations. The small size of h generates an e_H much smaller than the significance of the quoted result and the inherent error of the NLL, thus did not merit further consideration.

The uncertainty range of θ_{23}^{Newton} is larger than the univariate method. This is because as Δm^2_{23} tends to its global minimum, the two θ_{23} local minima merge together to form a single wider minimum dip. The effect is illustrated in figure 4.

Simulated Annealing

For this method, the starting point was set to $\theta_0 = 0.5 rad$, $\Delta m^2_0 = 2.4 \times 10^{-3} eV^2$. These were deliberately chosen to be poor estimates so as to avoid starting in a local minimum, therefore allowing the algorithm to explore a wider range in parameter space. After running for 12 equally spaced temperatures the resulting estimates were

$$\theta_{23}^{SA} = (0.7850 \pm 0.04300)\ \ rad$$
$$\Delta m_{23}^{SA} = 2.393^{+0.003075}_{-0.006873} \times 10^{-3} eV^2$$

with an NLL value of

$$NLL^{SA} = 373.22 \pm 0.5$$

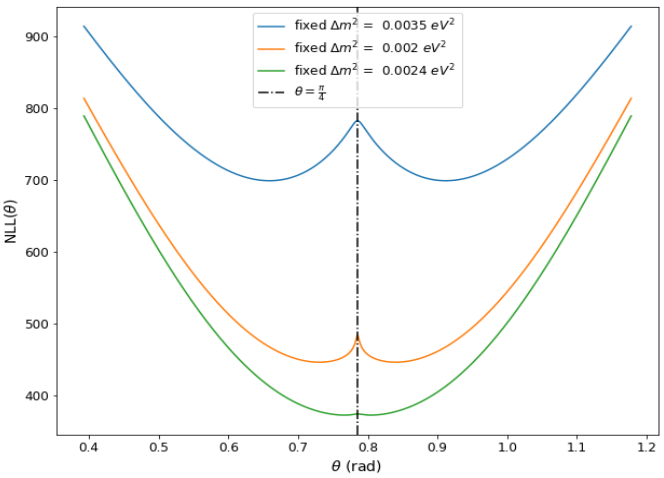

Fig. 4. Plot of NLL as a function of the parameter θ_{23} in rad for fixed value of $\Delta m^2_{23} eV^2$. It can be seen that as the value of Δm^2_{23} tends to 0.0024, the two θ_{23} minima fuse together until they form a larger dip. The NLL of the minimum in the merged dip is less than the NLL of the separate local minima, suggesting that the former is the closest to the global minimum of the paramaters. A black horizonal line indicates the location of $\theta = \frac{\pi}{4}$. As θ_{23} tends to $\theta = \frac{\pi}{4}$ the NLL needs to move less in the positive direction than the negative for a change of $+\frac{1}{2}$.

Simulated Annealing Errors

The standard deviation over the last 8 temperatures was $0.02038 rad$ and $3.024 \times 10^{-5} eV^2$ for θ_{23}^{SA} and Δm_{23}^{SA} respectively. Unlike the previous results, this is significant as it is comparabe to the NLL uncertainty. The final error was thus added in quadrature through

$$\sigma_{total} = \sqrt{\sigma^2_{numerical} + \sigma^2_{NLL}}$$

to give the error quoted above.

Comparison of 2-parameter results

The results for the 2D cases for each method are showcased on a contour plot in figure 5. Considering the figure, it can be deduced that while the univariate method's estimate returned the lowest NLL value, the parameters are stuck in a local minimum in θ_{23}. It is for this reason that it features the smallest error bars, as discussed in figure 4. Furthermore, both the Newton method and Simulated Annealing contain the highest 'peak' of the contour within their statistical uncertainty, rendering them the more accurate methods of the three investigated. Accounting for the fact that Simulated Annealing is rather agnostic of starting point location and consistently returns the lowest NLL result for bad starting estimates, it can be concluded that it is the more reliable method in this investigation, despite the larger error bars.

Adding Cross-section term

To account for the effect of the cross-section, each λ_i was updated according to equation 2. The free paramater α was plotted and a rough position of the minimum was estimated to be $\alpha_0 = 1$ which was used as the starting point in each method. Table I lists the resulting estimations from each method.

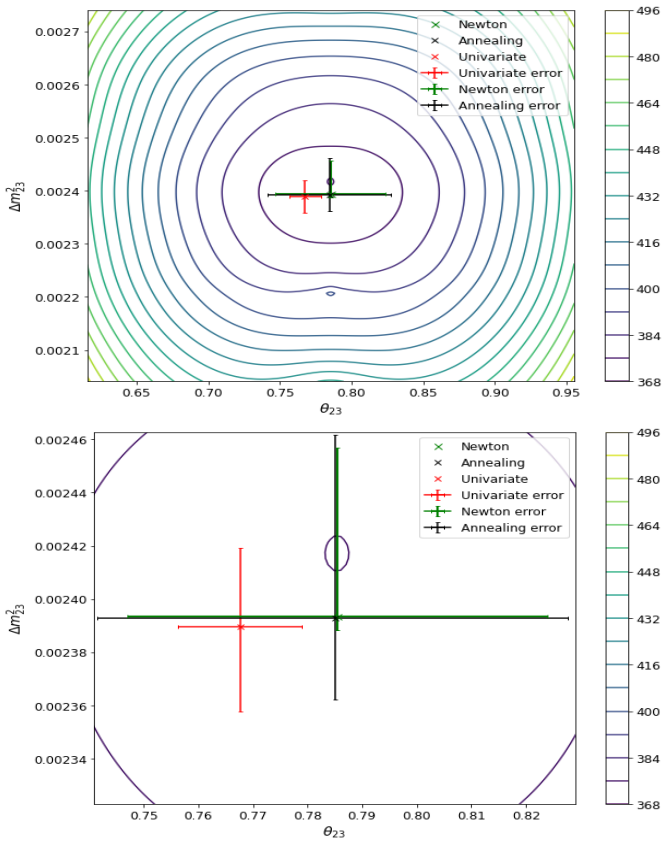

Fig. 5. Contour plots of the 2D case. The Δm^2_{23} and θ_{23} parameters are plotted on the y-axis and x-axis respectively. The contours indicate cuts on the NLL likelihood. The top plot features 50 cuts, while the bottom is a magnified view of the top plot in order to better compare the methods' results. The colour of the contour indicates their 'height' representing their NLL value, with the colour palette on the right matching colours with their NLL. Each value is fitted with its error bar.

Method	$\theta_{23min}(rad)$	Δm^2_{23min} $(10^{-3}eV^2)$	α (eV^{-1})
Univariate	0.7853 ± 0.03315	$2.536^{+0.001946}_{-0.005315}$	$1.026^{+0.04956}_{-0.04801}$
Newton	$0.7840^{+0.03534}_{-0.03258}$	$2.536^{+0.001964}_{-0.005277}$	$1.016^{+0.06039}_{-0.03915}$
SA	$0.7855^{+0.03654}_{-0.03674}$	$2.512^{+0.003912}_{-0.002576}$	$1.020^{+0.05271}_{-0.04485}$

TABLE I

The corresponding NLL values were:

$$NLL_{3D}^{Univariate} = 88.05 \pm 0.5$$
$$NLL_{3D}^{Newton} = 88.06 \pm 0.5$$
$$NLL_{3D}^{SA} = 87.92 \pm 0.5$$

Taking Simulated Annealing as the method with the lowest NLL as the best estimator, a final comparison plot of the optimised expected counts and observed counts is shown in figure 6. The peak's locations and values match, with the shape of the distribution being very similar, confirming the

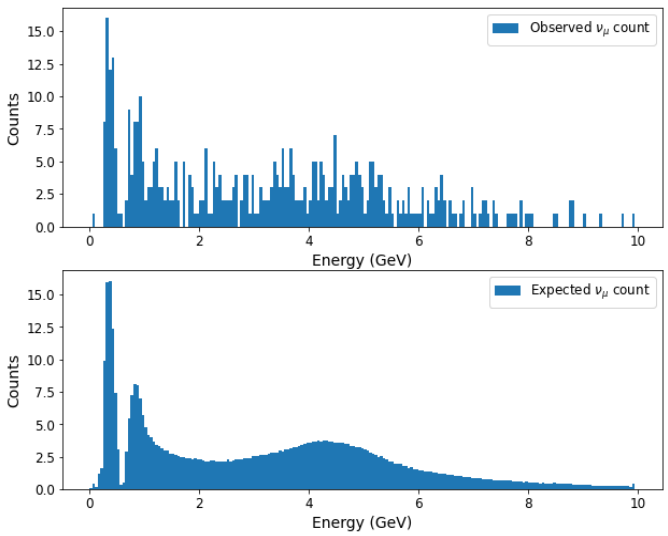

Fig. 6. Observed count number histogram on the top and expected count distribution histogram with parameters from Simulated Annealing minimisation on the bottom. The two plots match in shape, values and locations of peaks and troughs. Both distributions have a large peak with 16,12 and 15 counts in bins 7,8,9 respectively, before plowing to 0 and rising again to another peak. A steady rise can be seen in both plots, which is then followed by a decay that resembles negative exponentiation.

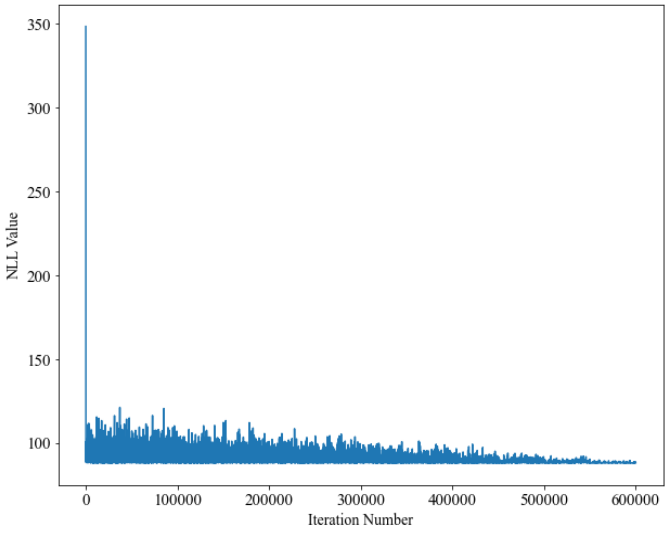

Fig. 7. Plot of NLL value against iteration number for 3D Simulated Annealing. The NLL rapidly drops from a high value of 348 and can be seen gradually converging with each iteration to the value closer to the final stated. This plot also acts as confirmation to the validity of this method, as the values seem correlated and not completely random due to the acceptance function.

accuracy of the result. Finally, a plot showcasing the evolution of the NLL value with each iteration in the annealing process is shown in figure 7.

V. Conclusion

In this investigation, simulated observation detections from the T2K experiment were used in conjunction with expected counts based on the survival probability of muon neutrinos to construct a negative log-likehood fit. The fit was subsequently minimised using three separate numerical methods

- the Univariate, Newton and Simulated Annealing. Each of them yielded estimators for three parameters θ_{23}, Δm^2_{23} and α, which were analysed and compared. The statistical uncertainty inherent to the NLL was the main source of error in the investigation, while numerical errors were also significant in the case of Simulated Annealing. A contour plot along with a comparison between observed and expected counts based on the optimised parameters validated the results, with Simulated Annealing yielding the lowest NLL while also marginally featuring the highest overall uncertainty.

REFERENCES

[1] Carlo Giunti and Chung W Kim. *Fundamentals of neutrino physics and astrophysics*. Oxford university press, 2007.
[2] Paul Dauncey Mark Scott. *3rd Year Computational Physics Project 1 Script*. Imperial College London, 2021-2022.
[3] Mark Richards. *First Year Statistics of Measurement L8 Notes*. Imperial College London, 2019-2020.
[4] Joe D Hoffman and Steven Frankel. *Numerical methods for engineers and scientists*. CRC press, 2018.
[5] Paul Dauncey Mark Scott. *3rd Year Computational Physics Course Notes*. Imperial College London, 2021-2022.