

Measuring charged current neutrino interactions in the Electromagnetic Calorimeters in near detector of T2K

Dominic Brailsford
TODO

A thesis submitted to Imperial College London
for the degree of Doctor of Philosophy

Abstract

This thesis presents a study of using the Electromagnetic Calorimeters (ECals) of the Tokai-to-Kamioka (T2K) off-axis near detector (ND280) as a target to study neutrino interactions using data collected during T2K run 3C.

The analysis presented shows the development of a new set of reconstruction algorithms which are able to reconstruct multiple tracks which originate from the same neutrino interaction. The output of this reconstruction is the used as the basis for a ν_μ charged current inclusive selection in the ND280 ECals. The selected events are then used in a simple χ^2 fit to extract the T2K flux-averaged $\nu_\mu u$ charged current inclusive cross-section on lead, which is measured as $\langle\sigma_{\text{Pb}}^{\text{CC}}\rangle_\phi = 8.13_{-1.26}^{+1.33} \times 10^{-39} \text{ cm}^2 \text{ nucleon}^{-1}$.

Declaration

This dissertation is the result of my own work, except where explicit reference is made to the work of others, and has not been submitted for another qualification to this or any other university. This dissertation does not exceed the word limit for the respective Degree Committee.

Dominic Brailsford

Acknowledgements

Something about my supervisor ...

Preface

This thesis describes my analysis of the ν_μ charged-current cross-section on lead using the T2K near detector electromagnetic calorimeters.

Contents

1	Measurement of the ν_μ charged current inclusive cross-section on Pb	3
1.1	Measurement method	3
1.2	Input samples to the fit	4
1.3	The ECal rate fit	6
1.4	Systematic uncertainties	8
1.4.1	Flux systematic evaluation	9
1.4.2	Cross-section systematic evaluation	13
1.4.3	ECal detector systematic evaluation	19
1.4.4	FGD detector uncertainty evaluation	34
1.4.5	The measured cross-section uncertainty	39
1.5	Validation of method	40
1.5.1	Validation of the fit machinery	40
1.5.2	Physics validation of the fitter	45
1.6	Applying the fit to ND280 data	49
1.7	The ν_μ charged current inclusive cross-section on lead	62
2	Discussion and conclusions	65
2.1	Discussion	65
2.2	Conclusions	66
2.3	The future	67
	Bibliography	71
	List of Figures	73
	List of Tables	79

“These chickens jackin’ my style. ”

— Fergie

Chapter 1

Measurement of the ν_μ charged current inclusive cross-section on Pb

This chapter presents the measurement of the ν_μ CC inclusive cross-section on Pb using the ND280 ECals. This chapter details the measurement method, the samples used in the measurement, the assessment of the systematic uncertainties, validation of the measurement method and finally the measurement itself.

1.1 Measurement method

The chosen method fits a prediction to measured data using multiple data samples [1, 2]¹. Here, a “sample” refers to, for example, selected events in a particular ECal module.² The core of the analysis method is a χ^2 fit which tries to minimise the difference between the prediction and the data. The χ^2 is defined as

$$\chi^2 = \Delta\vec{N}^T (\underline{\underline{V}}^{\text{syst}} + \underline{\underline{V}}^{\text{stat}})^{-1} \Delta\vec{N}, \quad (1.1)$$

where $\underline{\underline{V}}^{\text{syst}}$ and $\underline{\underline{V}}^{\text{stat}}$ are the systematic and statistical covariance matrices for the sample and $\Delta\vec{N}$ contains the difference between the data and the prediction for each

¹I put the references at the end of the sentence because the text doesn’t flow well if the first sentence in the chapter is cut off after just “The chosen method”.

²because “sample” could mean anything (e.g., different run periods or something), and without some idea of what it refers to, it is hard to follow the rest.

sample. If the number of samples used is M , $\Delta\vec{N}$ is defined as

$$\Delta\vec{N} = \begin{pmatrix} N_1^{\text{data}} - N_1^{\text{pred}} \\ \vdots \\ N_M^{\text{data}} - N_M^{\text{pred}} \end{pmatrix}. \quad (1.2)$$

For sample j of the sample set M , N_j^{data} and N_j^{pred} are the number of measured data events and number of predicted events. The extractable information from the fit is located in N_j^{pred} which can be subdivided into a set of templates, each of which are assigned a normalisation parameter, namely

$$N_j^{\text{pred}} = \sum_i R_i n_{ij}, \quad (1.3)$$

where n_{ij} is the number of events in template i of sample j and R_i is the normalisation assigned to that template. The normalisation parameters are free to float in the fit and so it is in the normalisation parameters that the desired physical information is located.

1.2 Input samples to the fit

The DS-ECal provides the main signal sample which will be used to extract the cross-section. As shown in table ??, the selected sample in the DS-ECal has the highest purity and efficiency in the whole selection making the sample the natural choice to extract the cross-section.

The method outlined above allows for the simultaneous constraint of physical parameters and the background contamination of a set of input samples. The method works particularly well when the input samples do not share the same sensitivity to a particular background type. For example, in ND280, each ECal module is exposed to a different beam intensity and energy spectrum and so is exposed to a unique amount of beam induced background. Each ECal module can be thought of as a separate input sample which fits well with the method outlined above. So, while the DS-ECal will provide the target in which the cross-section will be extracted, the barrel ECals are to also be included in the fit as a background constraint. To further allow the barrel ECals

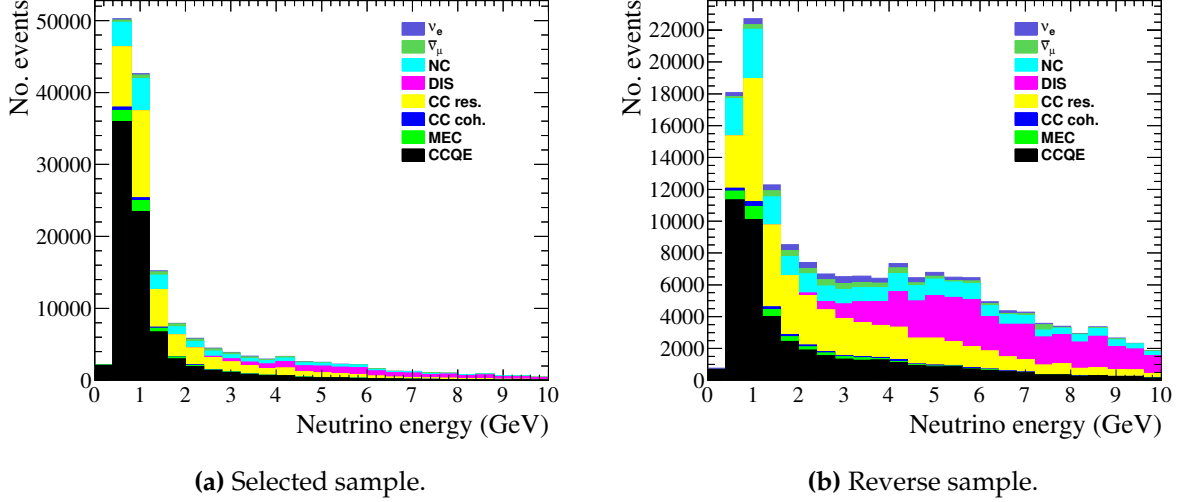


Figure 1.1: The neutrino energy of events seen in the bottom-left ECal module broken down by the neutrino interaction mode.

to achieve this, an additional sample set of barrel ECal events are to be included. This additional sample set, called the ‘reverse’ sample set, comes from the same data set as the selected sample described in chapter ?? . However, the events in the reverse sample set are events which pass the fiducial volume cuts but fail any other cut. An example of a selected sample compared with a reverse sample is shown in Fig. 1.1. There are clear differences in the shape of the energy spectrum between Fig. 1.1a and Fig. 1.1b which suggests that the selection is biased towards lower energy neutrino interactions, despite the flat efficiency curves shown in Fig. ?? . The apparent bias is actually an effect of higher energy interactions creating a surplus of reconstructed objects, most of which are cut away by the selection. The shape of the true neutrino energy spectrum (see Fig. ?? for an example) is actually more like the selected sample shown in Fig. 1.1a.

As described in chapter ?? , the developed selection does not distinguish between lead and carbon interactions, partly because such events are indistinguishable from one another. It is however desirable to separate the two categories out as the presented measurement purely involves lead. As the ECal is not capable of simultaneously constraining the lead and carbon events by itself, an additional constraint is needed. So, a sample of CC interactions on carbon occurring in the FGD are also included, which are taken from the official ND280 oscillation input analysis [3].

To summarise, there are many inputs samples to the measurement. Specifically, there are 12 barrel ECal samples (6 selected and 6 reverse), the DS-ECal sample and the FGD sample which totals to 12 input samples.

In terms of statistics, 3.949×10^{20} POT worth of NEUT MC events are used for this measurement.

1.3 The ECal rate fit

Now that the general method and the input samples have been introduced, the fit used by the analysis can be introduced. The fit itself is a data-driven constraint of the event rate in each ECal module where the Monte Carlo prediction is separated into templates whose normalisation is allowed to vary. The machinery outlined in section 1.1 almost completely describes what is required to understand the fit, however two specific definitions are needed. The first is that there are 12 input samples to the fit. The second is the definition of N_i^{pred} . To constrain the lead event rate, three Monte Carlo templates are needed: a lead template which contains any reconstructed event associated with a ν_μ CC interaction on lead, a carbon template which contains any reconstructed event associated with a ν_μ CC interaction on carbon and an 'other' template which contains any reconstructed event which does not fall into the above two categories. N_i^{pred} is then defined as

$$N_i^{\text{pred}} = R^{\text{Pb}} n_i^{\text{Pb}} + R^{\text{C}} n_i^{\text{C}} + R^{\text{other}} n_i^{\text{other}}, \quad (1.4)$$

where n_i^{X} is the number of events in template X of sample i and R_i^{X} is the variable normalisation of that template.

The rate fit allows R_i^{Pb} , R_i^{C} and R_i^{other} to vary with no prior constraint in an attempt to minimise the χ^2 as defined in equation 1.1. Once the minimum has been found, the fitted R_i^{Pb} can then be used to extract the cross-section using the selected events in the DS-ECal.

The number of events in each input sample, broken down by the template contributions, is shown in Fig. 1.2. As should be expected, the relative background contribution varies between each ECal module.

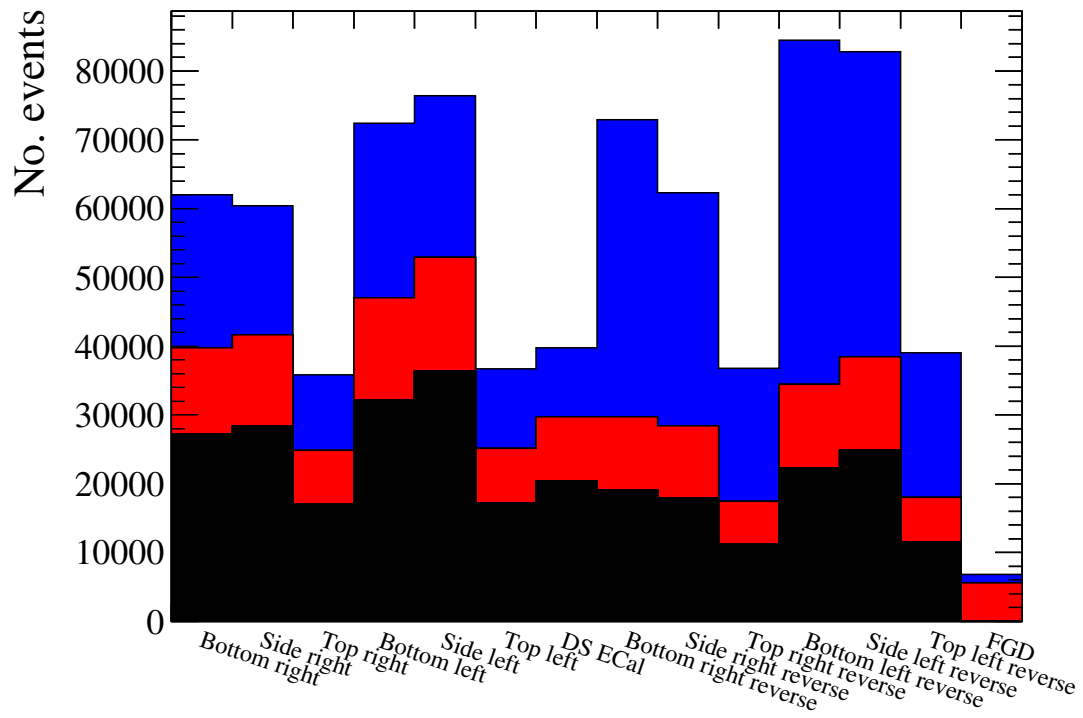


Figure 1.2: The number of events in each input sample separated into the lead (black), carbon (red) and other (blue) templates.

The underlying fit machinery uses the Minuit2 algorithm provided by the ROOT TFitter framework, chosen primarily because of its ease of use.

1.4 Systematic uncertainties

The χ^2 definition shown in equation 1.1 allows for systematic uncertainties to be directly included in the fit. As the systematic uncertainty implementation takes the form of a covariance matrix for the input samples, a good understanding of not only the systematic uncertainties, but how the samples correlate with one another is needed.

As stated above, there are 12 input samples so the covariance matrix will be a 12×12 symmetric matrix. While it may be obvious that the contents of the matrix will be the covariances of each input sample, it may not be immediately clear what type of covariances are needed. As described in section 1.3, the algorithm attempts to fit a set of predictions, separated into templates, to a set of measured data samples simultaneously. The variation comes from the normalisations assigned to each template. This means that the overall normalisation of the Monte Carlo prediction is free to vary without prior constraint and so it is the shape information between the prediction and data that constrains the parameters. Therefore the covariance matrix used in the fit should only contain uncertainties which account for the differences in the shape of the samples which the systematic sources introduce.

The ND280 event simulation is essentially broken down into three areas: Simulation of the neutrino flux, simulation of the neutrino interactions and simulation of the detector response. No matter how complicated a simulation can become, it is unlikely to ever simulate reality with 100% accuracy. The differences between what is simulated (be it flux, interaction or detector) and what happens in nature causes a systematic difference to be seen between collected data and Monte Carlo which must be accounted for. As the simulation can be broken down into three key areas, the systematic assessment can be broken down into the same three key areas, each of which are presented.

The actual assessment of a systematic uncertainty can be broken down into two areas: identification of a systematic error source and the propagation of that systematic error source through the analysis. The identification stage is typically handled by

a respective working group or by an analyser with external sources (e.g. control samples). The propagation of the systematic error source has no strict recipe but it typically involves variation of the identified systematic uncertainty and then either a reweighting of events is applied or the analysis chain is re-run. The systematic uncertainty treatment used in this analysis uses a mixture of the above.

To propagate the systematic uncertainty effect, a Monte Carlo sample akin to the prediction sample set used in the fit is required. So, a 2.5×10^{19} POT sub sample of beam and sand Monte Carlo is used for this purpose. All sample systematic covariance matrices are presented as fractional covariance matrices between each of the different detector samples. For the barrel samples, the selected samples are labelled by their respective module name and the reverse barrel samples (background enriched samples) are labelled with their module name and 'reverse'. The assessment of each systematic source is presented individually, resulting in a fractional covariance matrix for that source. The final systematic covariance matrix is then found by summing all of the individual covariance matrices.

1.4.1 Flux systematic evaluation

The neutrino flux is one of the core components of the ND280 event simulation and any uncertainties associated with the flux have a direct impact on the measured ND280 event rates. The associated flux uncertainties can be broken down into four categories:

- Properties of the proton beam such as profile and alignment
- Alignment of the target and focusing horns
- The current in the focusing horns and the magnetic field it generates
- Hadron production induced by beam interactions with the target

As the neutrino flux prediction is so important for all T2K analyses, the flux uncertainties are constrained using a range of measurements. These include external measurements from the NA61/SHINE collaboration, from the proton beam monitors, measurements of a spare magnetic horn and from the INGRID detector. Measured uncertainties from each source are used to vary the flux in JNUBEAM which produces a covariance matrix for each source. The covariance matrix to be used by the final analyser is the sum of each covariance matrix which is shown in Fig. 1.3. The

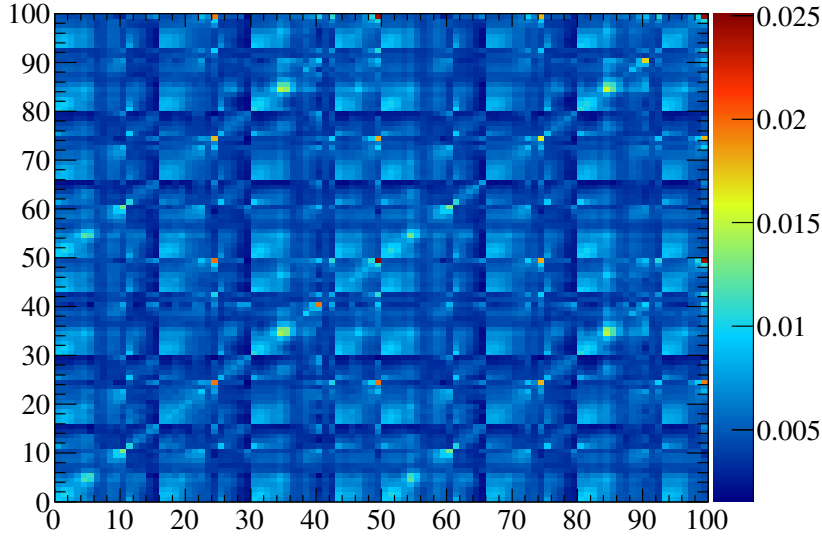


Figure 1.3: The flux prediction covariance matrix. The matrix is binned in detector (SK and ND280), neutrino flavour and energy.

covariance matrix is split into two detector sections: ND280 and SK. Each detector section is then separated into two beam running modes (ν and $\bar{\nu}$ runnings) which are also separated into four neutrino flavours (ν_μ , $\bar{\nu}_\mu$, ν_e and $\bar{\nu}_e$). Finally, each neutrino flavour is separated into a set of energy bins. The presented analysis is a ν_μ cross-section measurement using ND280, so only the ND280, ν -running mode section of the covariance matrix needs consideration. The provided flux covariance matrix is a fractional covariance matrix, binned in neutrino energy and flavour. As the neutrino energy and flavour information of events seen in this analysis is readily available, the flux covariance matrix can be used to reweight the events in the Monte Carlo sample to propagate the effect of the systematic uncertainty through the analysis. To do this, the flux covariance matrix is Cholesky decomposed and the resulting matrix multiplied by a vector filled with random throws from a Gaussian of mean 0 and width 1. Each element in this multiplied vector constitutes a fractional change in the number of a specific neutrino flavour and energy seen in the analysis. So, the event weightings are generated by adding 1 to each element of this vector. Every event selected in the sample is then weighted by the correct event weight (in this case defined by the neutrino flavour and energy) and this number of reweighted events is then recorded. The above description describes a single throw of the flux systematic propagation. This process is repeated 1000 times to build up the covariance matrix for the sample.

The sample covariance matrix elements are calculated as

$$V_{ab} = \frac{1}{N^{\text{throws}}} \sum_{i=1}^{N^{\text{throws}}} \frac{(N_a^i - N_a^{\text{nom}})(N_b^i - N_b^{\text{nom}})}{N_a^{\text{nom}} N_b^{\text{nom}}}, \quad (1.5)$$

where N_a^{nom} is the number of events seen in the nominal Monte Carlo for sample 'a' and N^{throws} is the number of systematics throws. N_a^i is generically the number of events seen in sample 'a' for systematic throw i , but its definition depends on which kind of covariance matrix is being calculated. For a shape+normalisation covariance matrix, N_a^i is defined as

$$N_a^i = N_a^{i'}, \quad (1.6)$$

where $N_a^{i'}$ is simply the number of events in sample 'a' for throw i after applying the systematic variation. When considering a shape-only covariance matrix, N_a^i is

$$N_a^i = N_a^{i'} \frac{\sum_{j=1}^{12} N_j^{\text{nom}}}{\sum_{j=1}^{12} N_j^{i'}}. \quad (1.7)$$

The extra factor on the RHS of equation 1.7 conserves the total number of events seen such that there are an equal total number of events before and after applying the systematic variation. The sample covariance matrices found by applying the above process are shown in Fig. 1.4.

The effect of the flux systematic on the selection efficiency of lead interactions was also evaluated. As the main target used in this analysis is the DS-ECal, the efficiency variation was evaluated for this detector only. For each systematic throw described above, the selection efficiency was calculated and recorded. By plotting the efficiency for each throw, the width of the efficiency distribution can be found which gives the efficiency uncertainty induced by the flux systematic. This efficiency uncertainty was found to be 0.00097. The lead selection efficiency in the DS-ECal is nominally 0.539, which implies a fractional change of 0.18%, which is essentially negligible.

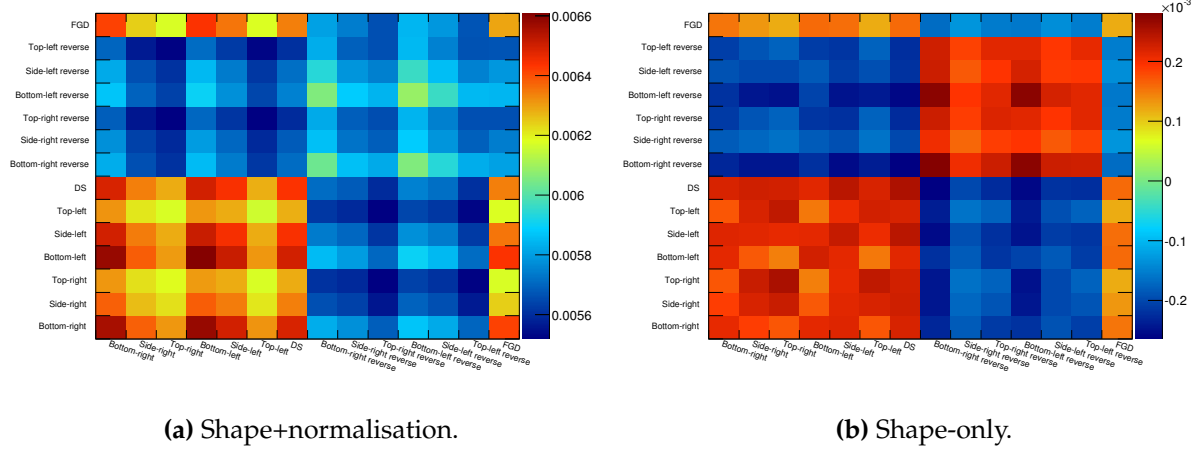


Figure 1.4: The sample covariance induced by the flux uncertainties.

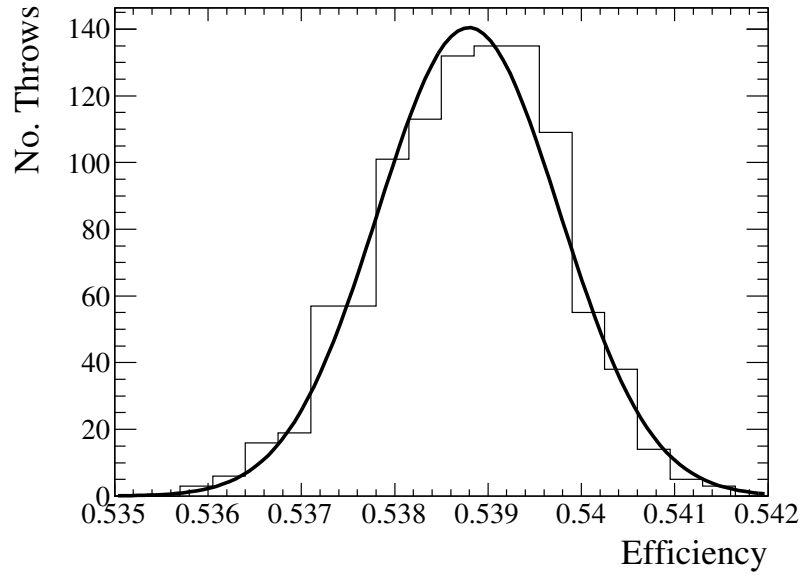


Figure 1.5: The variation in the selection efficiency of lead interactions in the DS-ECal caused by the flux systematic.

1.4.2 Cross-section systematic evaluation

The neutrino cross-section picture is still a very much open area of research. The interaction model used by current neutrino interaction generators use a set of empirical parameters which are tuned on experimental results. Uncertainties on any one of these empirical parameters will alter the simulated cross-section and thus the event rate seen in the ND280 simulation. A dedicated working group handles the assessment of the empirical parameters and their corresponding uncertainties. The parameter uncertainties are given as a covariance matrix along with the known correlations which is shown in Fig. 1.6. The interaction parameter assessment is chiefly undertaken with the T2K oscillation analyses in mind which means that the nuclear model parameters are only considered for carbon and oxygen. In the presented analysis, iron and aluminium are two major sources of background interactions which must also be treated in the systematic assessment. The chosen treatment is to assign a data-motivated, normalisation uncertainty to such backgrounds. The iron uncertainty is taken from the iron cross-section measurement using INGRID [4]. The aluminium uncertainty is taken from the total ν_μ cross-section on aluminium measurement using the IHEP-JINR detector [5]. The IHEP-JINR detector was exposed to a neutrino beam with a 3 GeV to 30 GeV energy range which is samples a higher energy than the T2K neutrino beam. Because of this fact, it was decided to double the quoted aluminium cross-section uncertainty. As element specific parameters for lead do not appear in Fig. 1.6, such parameters are additionally assessed separately using NEUT.

To propagate the effect of the cross-section systematic through the analysis, the covariance matrix shown in Fig. 1.6 was Cholesky decomposed and the resultant matrix multiplied by a vector of random numbers thrown from a gaussian (the same approach as described in section 1.4.1). However, the re-weighting in this case is more complicated. The T2KReWeight [6] software takes the parameters errors from the throw along with the simulated NEUT vertex which is matched to the event being varied. T2KReWeight then calculates a weight based on the inputted errors and the NEUT vertex in question. Using the T2KReWeight machinery, 1000 systematic throws from the covariance matrix shown in Fig. 1.6 were made and sample covariance matrices were calculated, which are shown in Fig. 1.7.

As with the flux systematic, the uncertainty in the selection efficiency for lead interactions in the DS-ECal due to the cross-section model systematic was found by

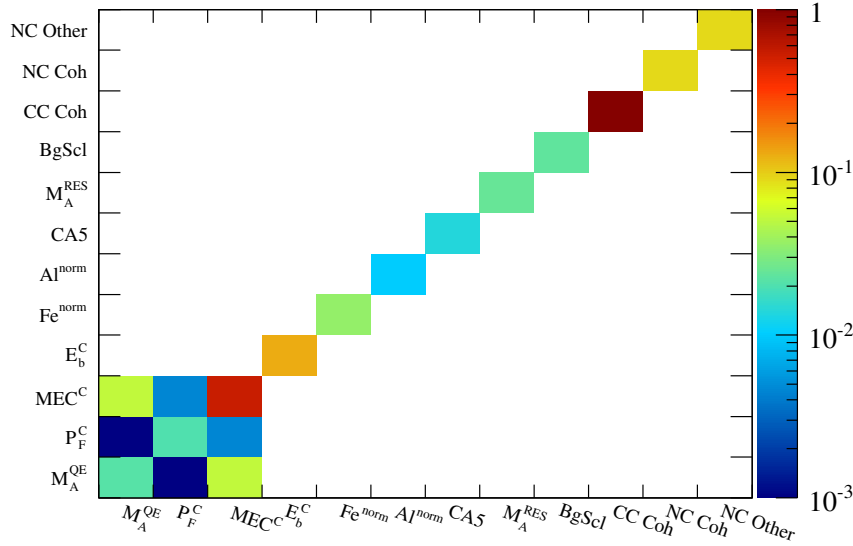
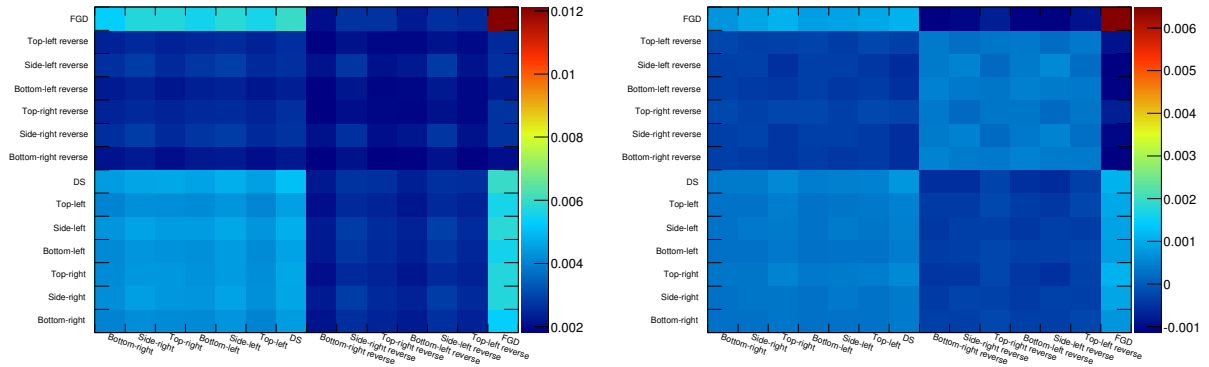


Figure 1.6: The neutrino cross-section model covariance matrix.



(a) Shape+normalisation.

(b) Shape-only.

Figure 1.7: The sample covariance induced by the cross-section model uncertainties.

recording the efficiency for each throw. The variation in the efficiency is shown in Fig. 1.8. The efficiency uncertainty was found to be 0.001 which corresponds to a 0.2% error which is, again, negligible.

As discussed above, the re-weighting of element specific parameters for events matched to lead interactions was not possible using T2KReWeight. As meson-exchange current interactions for lead interactions are not implemented in NEUT, the only parameters necessary to vary are P_F^{Pb} and E_b^{Pb} which are the Fermi-momentum and binding energy

for lead respectively. So, NEUT was used to produce 10,000 neutrino interactions on lead with uniform energy between 0 GeV and 10 GeV for the nominal parameters and $\pm \sigma$ variations of the parameters. The associated uncertainties were not readily available in the literature so 1.5 times the carbon parameter uncertainties were used as a conservative estimate. The NEUT events were used to plot the ratio of the varied cross-section to the nominal cross-section as a function of energy. The ratios for P_F^{Pb} and E_b^{Pb} are shown in Fig. 1.9a and Fig. 1.9b respectively. The variation in the cross-section caused by variation in E_b^{Pb} is sub-percent so there should be a negligible variation in the event rate which means this can be ignored. However, there is significant variation in the cross-section at low energy due to variations in P_F^{Pb} so this effect must be propagated through the analysis. To do this, the ratio distributions shown in Fig. 1.9a were treated as a set of events weights. The $+\sigma$ and $-\sigma$ weights were applied to the sample separately, creating two systematics throws. The results of the systematic throws were used to generate the sample covariances, which are shown in Fig. 1.10.

Final-state interactions (FSI) within the nucleus after a neutrino interaction has occurred could also cause a variation in the event rate. While the selection used in this analysis is CC inclusive, the method used separates events out into prong topologies. Any uncertainty in the FSI could cause migration of events from one prong topology to another. The Neutrino Interaction Working Group (NIWG) assesses these uncertainties

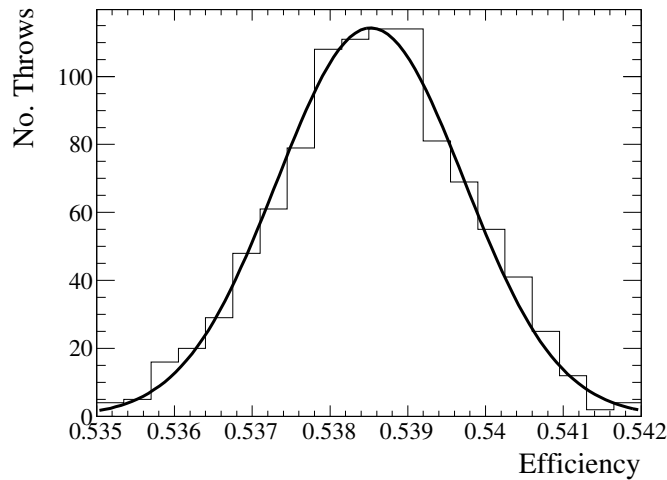


Figure 1.8: The variation in the selection efficiency of lead interactions in the DS-ECal caused by the cross-section model systematic.

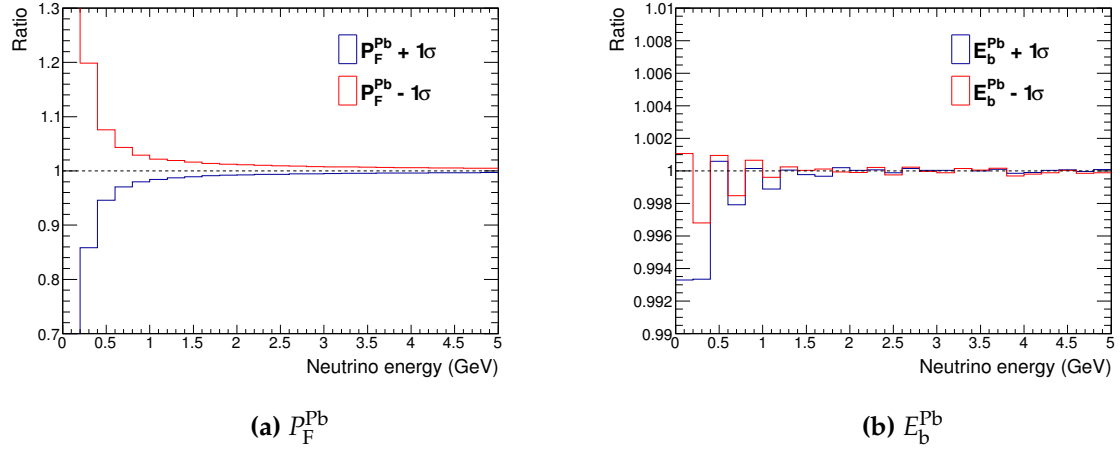


Figure 1.9: The ratio of the $\pm \sigma$ cross-section to the nominal cross-section for lead as a function of neutrino energy. Distributions produced by W. Ma.

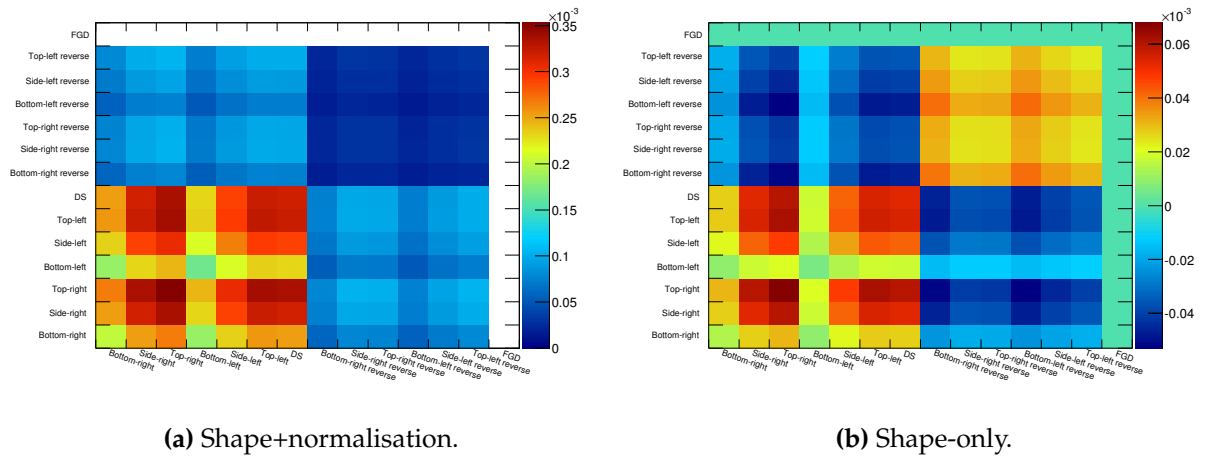


Figure 1.10: The sample covariance induced by variation of P_F^{Pb} .

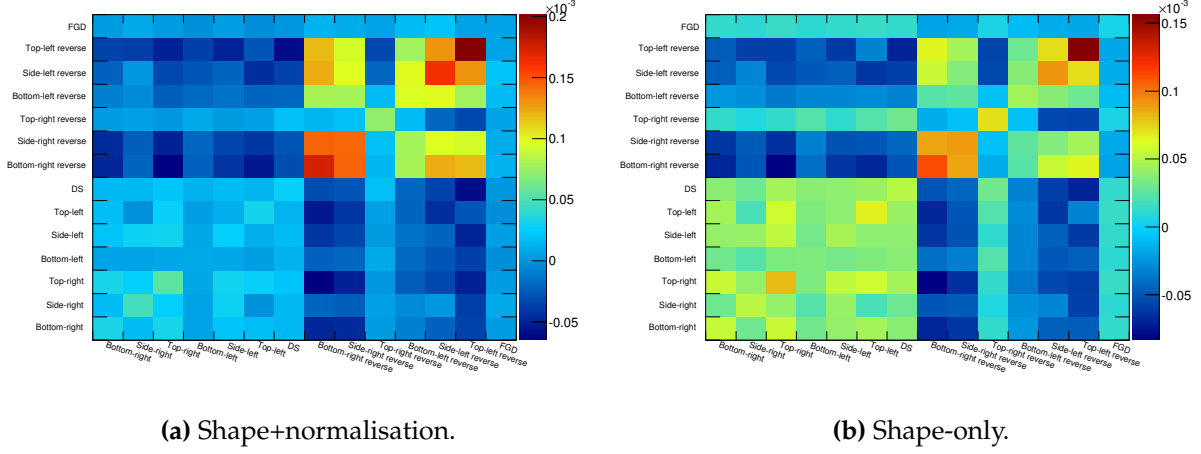


Figure 1.11: The sample covariance induced by variation of of the FSI parameters outlined in table 1.1.

and provides a set of parameters for an analyser to vary, all of which are associated with π^\pm interactions. There are six parameters associated with FSI:

- Low energy quasi-elastic scattering (FSIQEL)
- High energy quasi-elastic scattering (FSIQEH)
- Pion production (FSIINEL)
- Pion absorption (FSIABS)
- Low energy single charge exchange (FSICXL)
- High energy single charge exchange (FSICXH)

The NIWG provides combinations of the 1σ errors for the parameters highlighted above which span the whole FSI parameter space. The errors are provided as weights in 16 parameters sets which represent the 1σ contour for the FSI parameter space and are shown in table 1.1. Each parameter set can be used to generate one systematic throw, by passing the parameters to T2KReWeight and subsequently re-weighting the events. As with the cross-section model systematics described above, the parameter variations only affect carbon and oxygen. The FSI systematic throws were used to generate the sample covariance matrices which are shown in Fig. 1.11.

Parameter set	FSIQEL	FSIQEH	FSIINEL	FSIABS	FSICXL	FSICXH
Nominal	1.0	1.8	1.0	1.1	1.0	1.8
15	0.6	1.1	1.5	0.7	0.5	2.3
16	0.6	1.1	1.5	0.7	1.6	2.3
17	0.7	1.1	1.5	1.6	0.4	2.3
18	0.7	1.1	1.5	1.6	1.6	2.3
19	1.4	1.1	1.5	0.6	0.6	2.3
20	1.3	1.1	1.5	0.7	1.6	2.3
21	1.5	1.1	1.5	1.5	0.4	2.3
22	1.6	1.1	1.5	1.6	1.6	2.3
23	0.6	2.3	0.5	0.7	0.5	1.3
24	0.6	2.3	0.5	0.7	1.6	1.3
25	0.7	2.3	0.5	1.6	0.4	1.3
26	0.7	2.3	0.5	1.6	1.6	1.3
27	1.4	2.3	0.5	0.6	0.6	1.3
28	1.3	2.3	0.5	0.7	1.6	1.3
29	1.5	2.3	0.5	1.5	0.4	1.3
30	1.6	2.3	0.5	1.6	1.6	1.3

Table 1.1: FSI parameters, provided by the NIWG, representing the 1σ contour in the FSI parameter space [7].

1.4.3 ECal detector systematic evaluation

The analysis presented makes heavy use of a computational model of the ND280 ECal. Any differences between this model and the real ND280 ECal could cause a systematic difference between the analysed Monte Carlo and the collected data used in this study. So, several key areas have been identified which could cause such a systematic discrepancy, under the assumption that the differences exist.

The implemented selections uses the enhanced reconstruction, outlined in chapter ?? . The enhanced reconstruction depends on a set of clustering algorithms which are outlined in section ?? and those clustering algorithms depend on the hit scintillator bars themselves. Following this trail back, it can be argued that the vast majority of the ECal detector systematics can be covered by assessing the ability of the ECal to reconstruct individual hits. Specifically, the vast majority of the ECal's systematic uncertainties can be assessed by investigating the hit efficiency of the scintillator bars.

The aim of this study is to quantify the relative ECal hit inefficiency of collected data when compared to MC and use that information to invoke a similar ECal hit inefficiency in the MC. The effect of the hit inefficiency can then be propagated through the reconstruction and analysis to study its effect on the number of selected events. Fig. 1.12 shows the layer-by-layer efficiency of the ND280 ECals, measured using cosmic data [8]. Cosmic rays, which are largely composed of MIPs, are typically through going and only hit one bar per layer. So, the efficiencies shown in Fig. 1.12, can be used as a proxy for the hit efficiency of the scintillator bars. Fig. 1.12 contains almost all of the necessary information required to study this systematic uncertainty with the only missing information being the hit efficiency of the MC ECal. Due to time constraints, it was necessary to make a couple of conservative assumptions:

- The MC hit efficiency is 100%.
- The data hit efficiency for all bars is the lowest efficiency shown in Fig. 1.12, which is 96.2%.

Bearing the assumptions in mind, the hit inefficiency which must be applied to the MC is 3.8%. Care must be taken when attempting to apply the hit inefficiency to the MC. Naively, one might attempt to randomly remove 3.8% of hits from the reconstruction, however this would not best represent the hit inefficiency. There are many reasons why a scintillator bar would not be 100% efficient, however the efficiency should

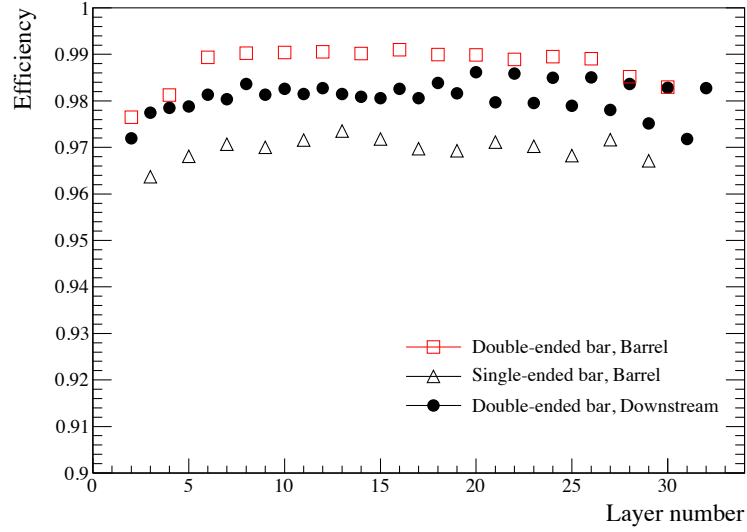


Figure 1.12: The layer-by-layer efficiency for the ND280 ECals, measured using cosmic data [8].

correlate with the deposited charge. So, to best model this systematic effect, the lowest 3.8% of hits should be removed from the MIP peak in charge space. Fig. 1.13 shows the deposited charge for each hit created by MC cosmic rays. The distribution is cut off at 5.5 PEU as this is a minimum threshold for output hits in the calibration stage of the ND280 software. The clearly visible MIP peak is very close to this threshold and is actually partially cut off by this limit. So, to apply the hit inefficiency, a new threshold needs to be calculated which removes an extra 3.8% of the lowest charge hits. By applying this logic to the hits used in Fig 1.13, the new threshold was calculated to be 7.55 PEU. This new threshold was applied as an initial stage of the reconstruction. The

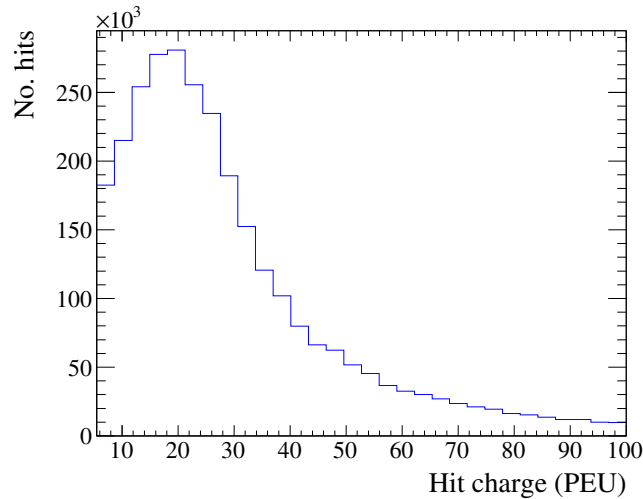


Figure 1.13: The charge distribution of scintillator bar hits created by MC cosmic rays.

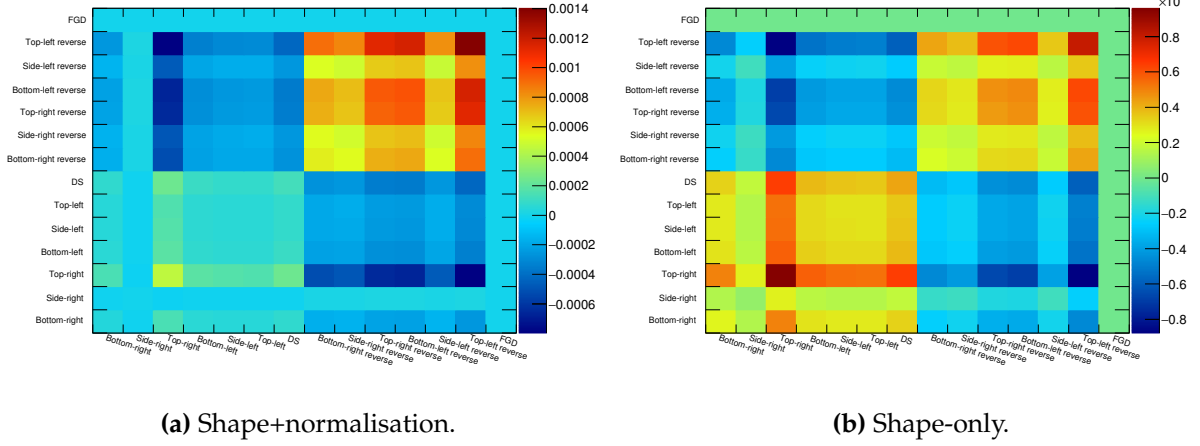


Figure 1.14: The sample covariance induced by the ECal hit inefficiency.

reconstruction and selection was then reapplied to the sample. This new sample was then compared to the nominal sample to construct the covariance matrices, which are shown in Fig. 1.15. The largest uncertainty for the samples was found to be 3.74%.

The hit efficiency systematic uncertainty actually covers several areas of uncertainty. Firstly, it addresses how a change in the core input to the reconstruction changes its output and so, by extension, addresses the efficiency of the reconstruction. However, because the inefficiency is invoked by removing low charge hits, it also addresses how low charge hits, which could migrate below the minimum reconstruction threshold, could affect the analysis. Because treatment of the charge by the hit efficiency systematic is partially covered, it is only necessary to additionally address how variation in the charge of the already reconstructed objects could alter the number of objects which pass the selection cuts. To assess this, a well understood sideband sample is needed such that the hit charge in MC and data can be compared. Any shift or smearing of the charge in data, relative to Monte Carlo must then be added to the reconstructed objects, which can then be passed through the selection, allowing the covariance matrices to be constructed. To study the variation in the hit charge, a sample of cosmic MC and data was used. The two samples were passed through the enhanced reconstruction and the charge of the hits associated to the reconstructed events were then analysed. An example of the reconstructed hit charges is shown in Fig. 1.15a. There are clear differences in the shape of the two distributions; the data charge peak is wider than the MC peak. To quantify this difference in the peak of the distributions, a gaussian was fit to the top 66% of the two peaks which were then compared. The relevant parameters

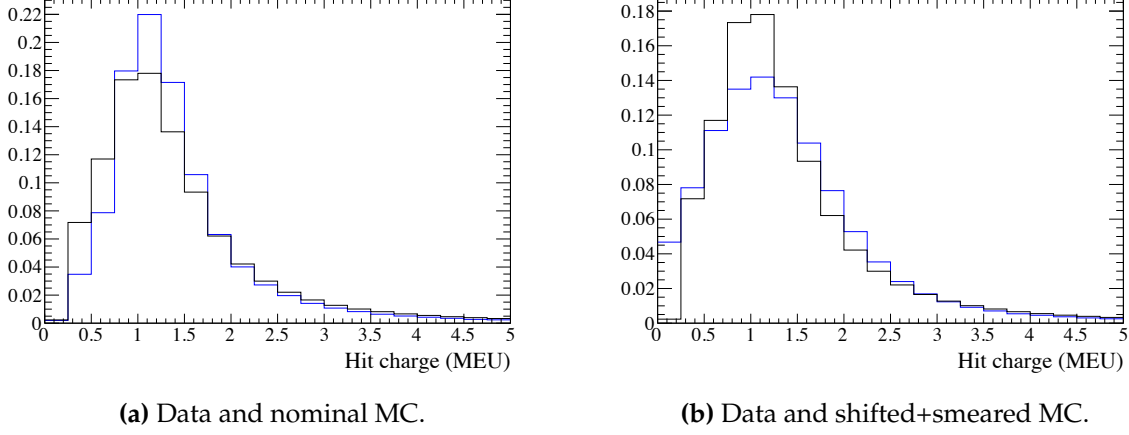


Figure 1.15: The charge of ECal hits after applying the enhanced reconstruction to run 3C cosmic events. The distributions only show information for reconstructed objects in the top-left barrel ECal module. The blue and black histograms are Monte Carlo and data respectively.

for the comparison are the width of the data and MC fits, σ_{Data}^Q and σ_{MC}^Q respectively, and

$$\Delta^Q = \mu_{\text{Data}}^Q - \mu_{\text{MC}}^Q, \quad (1.8)$$

where μ_{Data}^Q and μ_{MC}^Q are the mean of the data and MC fits respectively. Δ^Q measures the shift of the MC peak relative to data, while σ_{Data}^Q and σ_{MC}^Q can be used to quantify the relative width. These values are shown in table 1.2. To truly quantify the difference in the charge distributions, not only is a well understood sample required, but also a clean set of reconstructed events to remove any additional effects which may mask the charge difference. As the cosmic events will generally be almost parallel to the vertical axis, only the top and bottom ECal modules will be able to cleanly reconstruct the cosmic events, so only those events should be considered. Ideally, the values shown in table 1.2, would be used to correct and smear the MC peak to match data. However,

due to time constraints, it was only possible to use said values to over-correct and over-smear the MC. In terms of the individual hits, this meant using Δ^Q and σ_{Data}^Q for the top-left ECal to apply a correction and smearing to the hit charges, regardless of which module the hit occurred in. The top-left ECal module was chosen because it has the largest value of Δ^Q . The adjusted hit charge is

$$Q' = Q + X \quad (1.9)$$

ECal module	Δ^Q (MEU)	σ_{Data}^Q (MEU)	σ_{MC}^Q (MEU)
Bottom-right	-0.095	0.37	0.51
Bottom-left	+0.045	0.37	0.61
Top-right	-0.003	0.38	0.58
Top-left	-0.099	0.38	0.50
Side-right	-0.225	0.83	0.64
Side-left	-0.315	0.85	0.55
DS	-0.472	0.87	0.46

Table 1.2: Summary of the lead absorbers for the ND280 tracker ECals [8].

where Q is the nominal hit charge and X is a random variable which is defined as

$$X \sim N(\Delta^Q, \sigma_{\text{Data}}^Q). \quad (1.10)$$

It is important to note here that the width of the Gaussian distribution in which the adjustment is drawn from is σ_{Data}^Q . By using this width, the MC hit charge is over-smearred relative to data. An example application of equation 1.9 is shown in Fig. 1.15b which uses the same hit information as that in Fig. 1.15a, but with the correction applied to the MC hit charge. As can be seen by comparing the two figures, the application of the over-smearing has caused the MC charge peak to be wider than the data peak. To propagate the effect of this systematic, this smearing and correction needs to be applied to the reconstructed prongs before any selection takes place. Unfortunately, hit level information is not available at the level where the selection is applied. So, a slightly modified charge adjustment is needed. If the total charge contained on a prong is Q_{prong} , then the total adjusted charge is

$$Q'_{\text{prong}} = Q_{\text{prong}} + X_{\text{prong}} \quad (1.11)$$

where X_{prong} is a random variable and is defined as

$$X \sim N(\Delta^Q N_{\text{prong}}, \sigma_{\text{Data}}^Q), \quad (1.12)$$

where N_{prong} is the number of hits contained on the prong. Following the above description, the prong charge variation was applied to every reconstructed prong in the systematic sample and the selection was then applied. This process was repeated 1000 times to build up the covariance matrices, which are shown in Fig. 1.16.

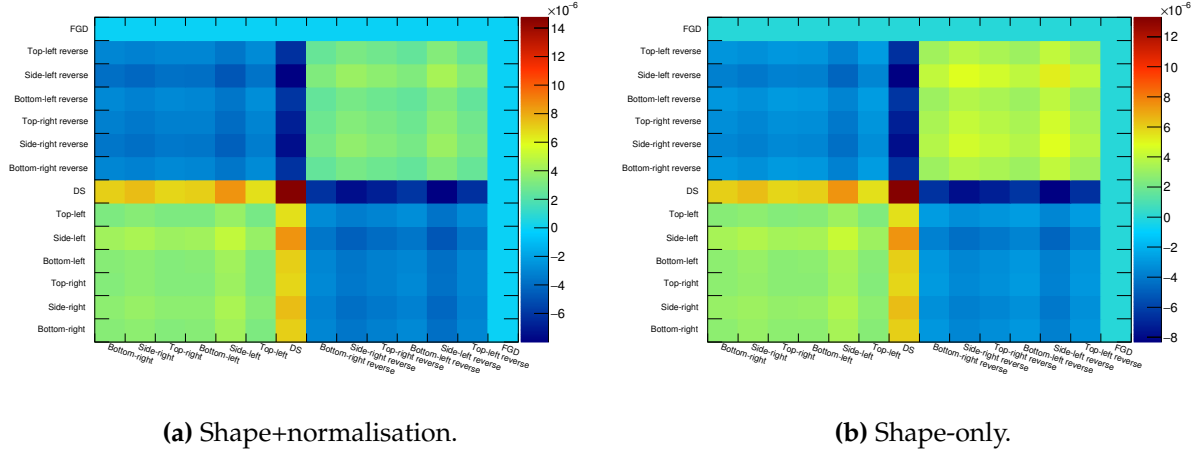
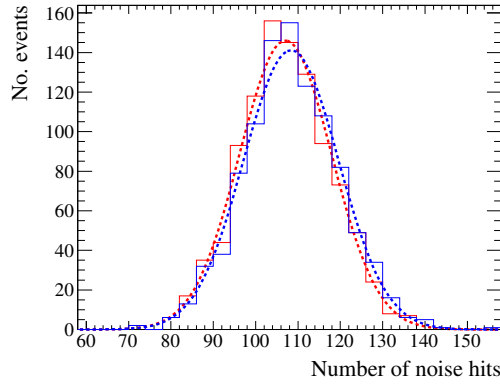


Figure 1.16: The sample covariance induced by the ECal charge uncertainties.

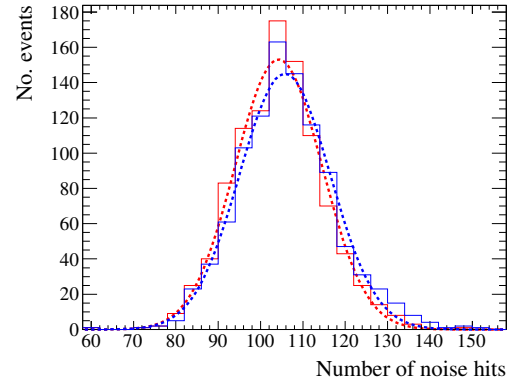
While the hit efficiency and charge systematic assessments are fairly all-encompassing in terms of ECal detector uncertainties, they can not address how the inherent noise in the ECal can affect the reconstruction and, by extension, the selected number of events in this analysis. There is only a need to address a noise systematic uncertainty if the simulated noise rate in the Monte Carlo is different to what actually happens in the ECal.

To measure the noise rate, a control sample of cosmic rays (for the barrel) and through-going muons (for the DS-ECal) were passed through the reconstruction. Before the clustering stages were initiated, the number of hits in the relevant ECal module were recorded. Then, after the reconstruction chain was completed, the number of hits which formed the final reconstructed objects were also counted. The difference between these two numbers forms the noise hit estimate. To test this estimator, the true number of noise hits in the Monte Carlo were also counted. The results of this test are shown in Fig. 1.17. The Gaussian fits to the distribution are used to quantify the noise levels per event (by using the mean and width of the fits). These values are shown in table 1.3. The estimate is clearly in good agreement with the true noise levels in the Monte Carlo. So, it was decided to apply this method to estimate the noise levels in the data samples, the results of which are shown in Fig. 1.18 and also shown in table 1.3. It is clear that there are differences between the noise levels in data and

Monte Carlo. So, to propagate the effect of this systematic uncertainty, elecSim was retuned to produce the data-like noise levels in both the barrel and DS-ECal. The

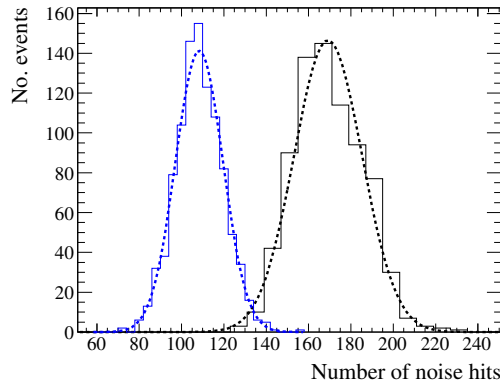


(a) Cosmic control sample in the barrel ECals.

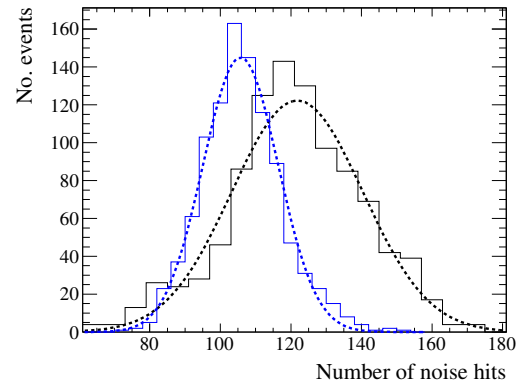


(b) Through-going muon control sample in the DS-ECal.

Figure 1.17: The estimation of the number of noise hits (solid blue line) and the number of true noise hits (solid red line) per event in Monte Carlo. The coloured dashed lines are the Gaussian fits to the respective distributions.



(a) Cosmic control sample in the barrel ECals.



(b) Through-going muon control sample in the DS-ECal.

Figure 1.18: The estimation of the number of noise hits in Monte Carlo (solid blue line) and data (solid black line) per event. The coloured dashed lines are the Gaussian fits to the respective distributions.

ECal module(s)	True (MC)	Estimated (MC)	Estimated (data)
Barrel	107 ± 10	108 ± 11	169 ± 16
DS	104 ± 10	106 ± 11	122 ± 18

Table 1.3: Summary of the study to estimate the number of noise hits per event in the ECals.

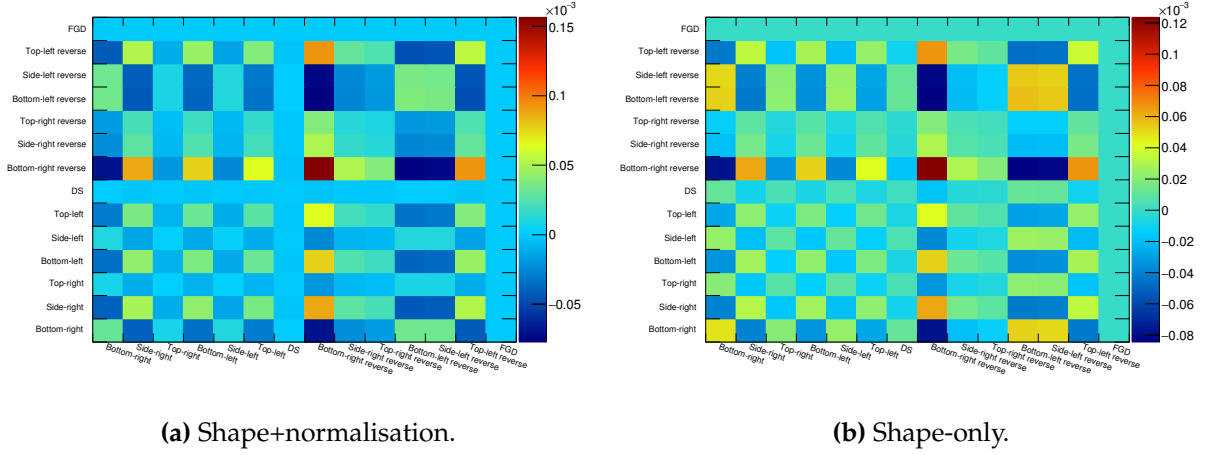


Figure 1.19: The sample covariance induced by the ECal noise uncertainties.

systematic sample was then passed through the modified elecSim and all pieces of the software chain downstream. After this, the selection was applied to the modified sample. This sample was then compared to nominal and the difference was used to construct the covariance matrices, which are shown in Fig. 1.19.

The ECal model used in the simulation was based on the as built drawings and so should be an accurate model. However, the components of each ECal module do have associated uncertainties which are not, and can not easily be, taken into account in the simulation, but could easily cause a systematic difference between data and Monte Carlo. For example, if the DS-ECal lead absorbers were thicker in the simulated model, one should expect to see a relative deficit of beam triggered events in the DS-ECal collected data. Provided that the component dimensions, compositions and uncertainties are known, toy Monte Carlo can be used to construct a set of toy ECal modules which model the variation in the module mass. This can be expanded further, to construct a covariance matrix for the total mass of each contributing element. This covariance matrix can then be used to re-weight the events, based on target element, in the Monte Carlo sample to model the effect of the ECal mass systematic.

The ECal module designs are split into three types: side, top/bottom and DS. It is only necessary to generate a mass covariance matrix for each of the three types. All values used for the component dimensions/uncertainties are taken from [8]. Precise component dimensions are only quoted for the components in the active regions of each module, namely the lead absorbers, the scintillator bars and the wavelength-

	Top/Bottom	Side	DS
No. per layer	2	4	2
Length (mm)	3858 ± 4	964.5 ± 4	2016 ± 1
Width (mm)	765 ± 4	2330 ± 4	1008 ± 4
Height (mm)	1.75 ± 0.1	1.75 ± 0.1	1.75 ± 0.1
Sb doping	$2.0 \pm 0.2\%$	$2.0 \pm 0.2\%$	$2.0 \pm 0.2\%$

Table 1.4: Summary of the lead absorbers for the ND280 tracker ECals [8].

shifting fibres. So, only these components are considered in the toy Monte Carlo study.

1.4.3.1 The lead absorbers

Reference [8] provided a clear description of the composition of the lead absorbers for the ECal module designs, which are summarised in table 1.4. All component uncertainties (both dimension and composition) were given as symmetric errors. To construct an ECal lead layer, each layer dimension was randomly taken from a gaussian with a mean equal to the quoted dimension and a width equal to the dimension uncertainty. After the layer volume was defined, the Sb contamination was calculated, also using a gaussian throw, but using the quoted Sb contamination and its uncertainty as the Gaussian's mean and width. The described information could then be used to retrieve the mass of the Pb and Sb in an ECal lead layer.

1.4.3.2 The scintillator bars

The information used to construct the toy scintillator bars is summarised in table 1.5, most of which was found in reference [8]. The length of the bars were quoted with symmetric uncertainties (like the lead absorbers), so the length was treated in the same way as the lead absorber dimensions. The hole and coating dimensions were also treated in the same manner. However, the width and height were quoted with asymmetric errors and so could not be treated in the same way. Only a negative uncertainty was quoted, which meant the width and height uncertainties could be treated in a somewhat similar manner, but with random numbers being drawn from a

	Top/Bottom	Side	DS
No. per layer (para)	38	57	-
No. per layer (perp)	96	96	50
Length (para) (mm)	3840 ± 0.1	3840 ± 0.1	-
Length (perp) (mm)	1520 ± 0.1	2280 ± 0.1	2000 ± 0.1
Width (mm)	$40^{+0.0}_{-0.4}$	$40^{+0.0}_{-0.4}$	$40^{+0.0}_{-0.4}$
Height (mm)	$10^{+0.0}_{-0.4}$	$10^{+0.0}_{-0.4}$	$10^{+0.0}_{-0.4}$
Hole diameter (mm)	1.75 ± 0.1	1.75 ± 0.1	1.75 ± 0.1
Composition (CH:PPO)	99:1	99:1	99:1
Coating thickness (mm)	0.25 ± 0.13	0.25 ± 0.13	0.25 ± 0.13
Coating composition (CH:TiO ₂)	85:15	85:15	85:15

Table 1.5: Summary of the scintillator bars for the ND280 tracker ECals [8].

half-gaussian. The bar composition was reported as polystyrene doped with 1% PPO (C₁₅H₁₁NO) and 0.03% POPOP. For the toy bar construction, the POPOP was considered negligible and so was not included. The result was that the toy bar composition (excluding the coating) was 99% CH and 1% C₁₅H₁₁NO by mass. Unfortunately, the full composition of the TiO₂ coating was not extensively described. However, a similar systematic study has already been undertaken for the FGD [9] which included an investigation of the coating composition. A representative for the Fermilab facility which extruded the scintillator bars reported that the final composition of the coating was 15% rutile-form TiO₂ and 85% CH with other contributing compounds considered negligible. This information was also included in the modelling. The collected information could then be used to retrieve the contributing mass of each element in a scintillator bar.

1.4.3.3 The WLS fibres

The information used to construct the WLS fibres is shown in table 1.6. As with the other toy components, reference [8] provided most of the necessary information for the construction and could be largely treated using the same methods as above. The major difference is that the WLS fibre diameter had an asymmetric error in which both components were not zero. So, two toy gaussians were constructed to draw from (one representing each uncertainty) and a random number was first drawn from a

	Top/Bottom	Side	DS
No. per layer (para)	38	57	-
No. per layer (perp)	96	96	50
Length (para) (mm)	3840 ± 0.05	3840 ± 0.05	-
Length (perp) (mm)	1520 ± 0.05	2280 ± 0.05	2000 ± 0.05
Diameter (mm)	$1^{+0.02}_{-0.03}$	$1^{+0.02}_{-0.03}$	$1^{+0.02}_{-0.03}$
Composition	CH	CH	CH

Table 1.6: Summary of the WLSfibres for the ND280 tracker ECals [8].

uniform distribution to decide which gaussian would be used. Very little information was provided about the composition of the WLS fibre, however the mass study in the FGD [9] assumed the WLS fibre was composed entirely of scintillator, so the same assumption was used for this study. As with the other components, the provided information could then be used to retrieve the contributing mass of the elements in the WLS fibres.

1.4.3.4 ECal mass covariance

Toy simulation was used to construct each component described in the above sections. Each component was repeatedly built and was subsequently used to construct an ECal module. The contributing mass of each element to the toy ECal module was then recorded. 10,000 toy ECal modules were constructed for each type, totalling 30,000 constructed modules. The recorded contributing elemental mass for each construction was used to build up a mass covariance matrix using eqn. 1.5 and eqn. 1.6, an example of which is shown in Fig. 1.20 for the DS-ECal. Unfortunately, it was found that the covariance matrices for all three module types were singular, which is of immediate concern as the matrices need to be Cholesky decomposed as part of the systematic uncertainty evaluation. It was found that the singular nature of the covariance matrices was caused by the contributing elements which have small covariance values (both on and off diagonal) and the problem could be solved by not considering some of them in the covariance matrix. The three problematic elements are carbon, nitrogen and hydrogen. Both nitrogen and hydrogen only constitute a relatively small amount of the mass and an example of this is shown in Fig. 1.21 which shows the mass of the 10,000 toy DS-ECal modules, with and without nitrogen and hydrogen. There is

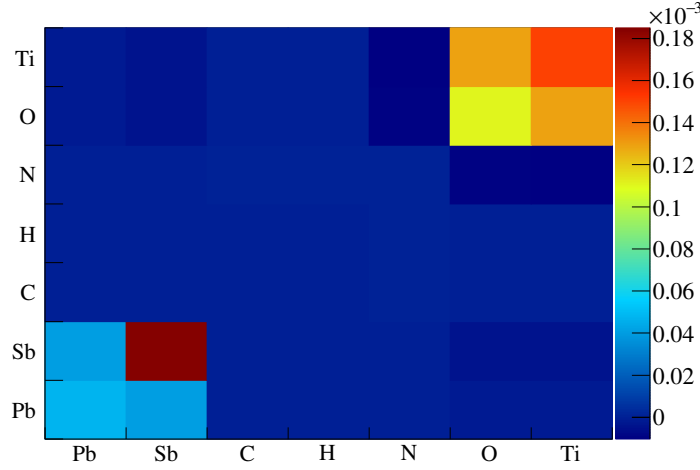


Figure 1.20: The covariance matrix for the mass of the contributing elements in the DS-ECal, with all contributing elements considered.

a 2.45% difference between the two constructions which is small but not necessarily negligible. While ideally the hydrogen and nitrogen terms should be included in the systematic evaluation, due to time constraints and that all covariance terms involving them were very small, it was decided that they would be excluded from the covariance matrix. With the extra elements excluded, a non-singular covariance matrix for each module type could be constructed, all of which are shown in Fig. 1.22.

1.4.3.5 Propagation of the mass systematic uncertainty

The covariance matrices shown in Fig. 1.22 were then used to generate systematic throws, in the same manner as the flux systematic uncertainty evaluation. For the input into each systematic throw i.e. one run through of the selected Monte Carlo events, a set of event weights were generated for each of the seven ECal modules using the above mass covariance matrices. Every neutrino interaction in the ECal active volume was weighted according to its target element. As with the other systematic uncertainty evaluations, 1,000 systematic throws were generated and the subsequent covariance matrices were constructed which are shown in Fig. 1.23. The maximum uncertainty, which was for the DS-ECal events, was found to be 0.39%.

An additional source of systematic uncertainty could come from the ECal's ability to reject OOFV backgrounds. As fig. ?? shows, the dominant background selected, particularly for the barrel ECals, is the OOFV background. If the simulation of the

ECal had a different OOFV background rejection capability to that of the actual ECal, the selected event composition between data and MC would differ, resulting in a systematic difference. The only feasible way to assess this systematic uncertainty is to use control samples. Unfortunately, only μ based control samples were available during the analysis. However, fig. ?? shows that the one prong topology, which contains the largest number of selected events, has the biggest OOFV background contamination and as Fig. ?? and Fig. ?? show, μ s are the dominant particle species in the surviving OOFV background. To assess the systematic uncertainty, the μ based control samples, both data and MC, were passed through the selection and the number of reconstructed events which pass the fiducial volume cut were recorded, N_{CS}^{pass} . The efficiency of the fiducial volume cut was then defined as

$$\epsilon_{CS} = \frac{N_{CS}^{\text{pass}}}{N_{CS}}, \quad (1.13)$$

where N_{CS} is the number of reconstructed control sample events which enter the selection. It is important to note why N_{CS}^{pass} counts the number of events which pass the fiducial volume cut, rather than as the number of events which pass all of the selection cuts. The reason for this is the inclusion of the reverse sample (background enriched sample) in the rate fit which includes any event which passes the fiducial

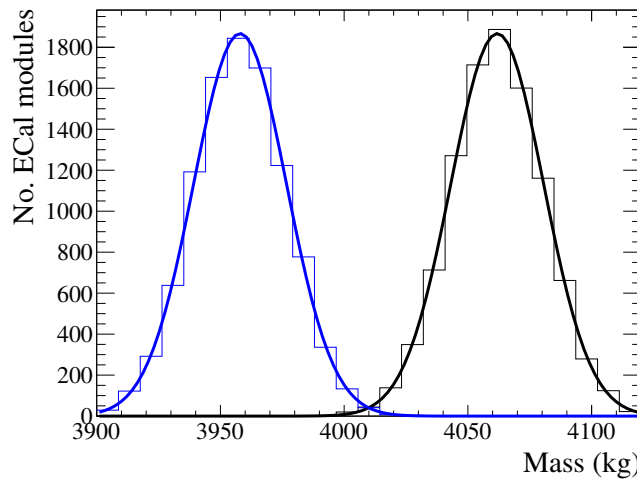


Figure 1.21: The total mass of the DS-ECal active volume with (black) and without (blue) nitrogen and hydrogen as contributing elements. The smooth lines are the Gaussian fits to the distributions. Using the Gaussian mean and width as the mass central value and uncertainty respectively, the DS-ECal active mass with (without) hydrogen and nitrogen is 4060 ± 20 kg (3960 ± 20 kg).

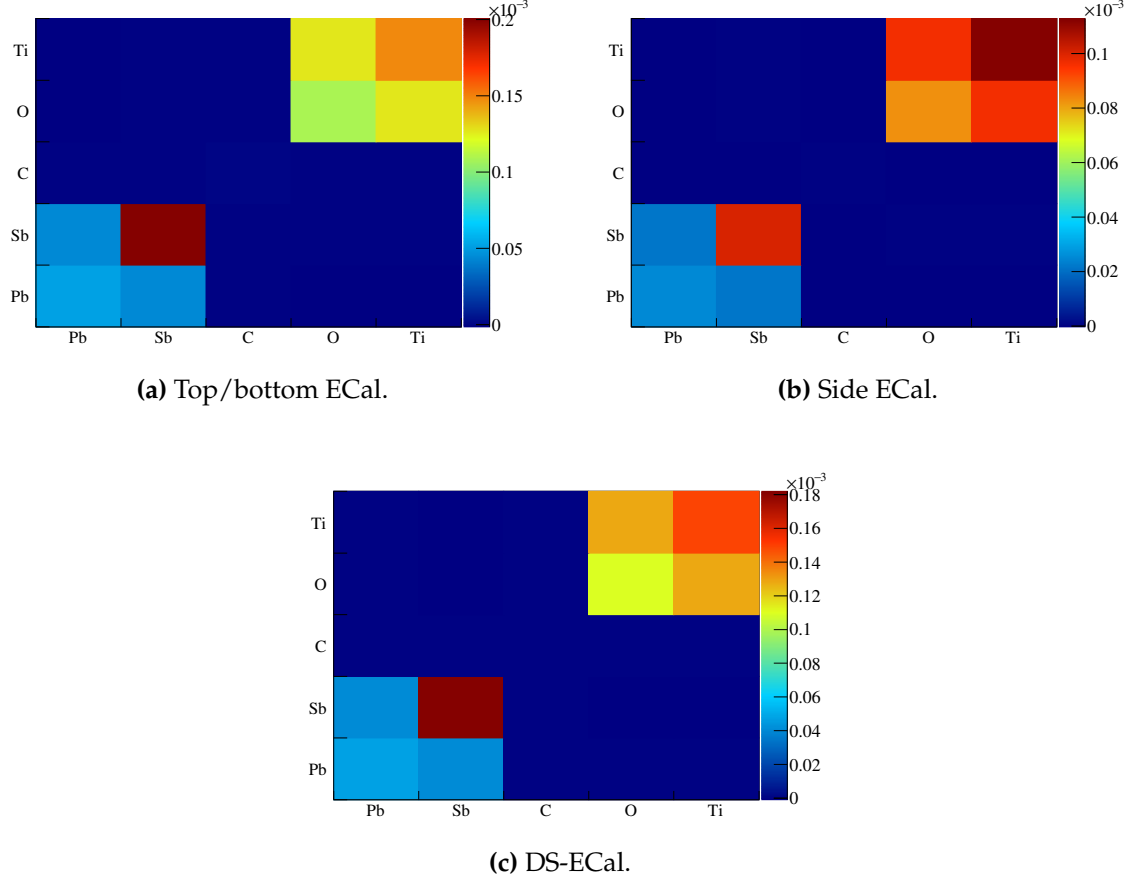


Figure 1.22: The fractional covariance matrix for the mass of each contributing element to an ECal module (hydrogen and nitrogen omitted).

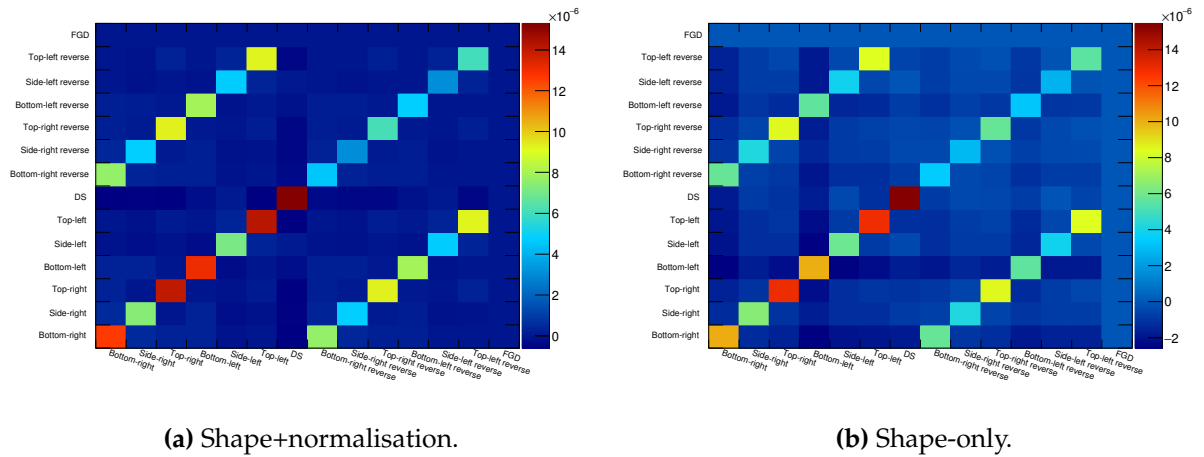


Figure 1.23: The sample covariance induced by the ECal mass uncertainties.

ECal module(s)	N_{CS} (MC)	N_{CS} (data)	ϵ_{CS} (MC)	ϵ_{CS} (data)	α_{OOFV}
Barrel	26560	42961	2.12%	5.92%	3.80%
DS	6225	22555	2.27%	5.97%	3.70%

Table 1.7: Summary of the ECal OOFV systematic study when using the FGD cosmic control sample.

volume cut, but fails the most upstream cut. By calculating ϵ_{CS} separately for data and MC, the systematic uncertainty can then be defined as

$$\alpha_{OOFV} = \epsilon_{CS}^{\text{data}} - \epsilon_{CS}^{\text{MC}}. \quad (1.14)$$

The FGD cosmic control samples were used for this study, which are events with coincident reconstructed tracks in both FGDs and which occur outside of a beam spill window. A summary of the information relevant for systematic uncertainty is located in table 1.7.

To propagate the effect of the systematic uncertainty, a similar approach to the flux systematic uncertainty propagation. However, as the OOFV systematic uncertainty is a pair of single numbers rather than a covariance matrix, only two event weights (one for the barrel ECal and one for the DS-ECal) were generated per systematic throw. The event weights were used to reweight only the OOFV events in the sample and the subsequent sample covariance matrices were constructed, which are shown in Fig. 1.24. The largest uncertainty is in the bottom-left reverse sample and is 1.82%.

As described in section ??, a significant data/MC difference was found in the DS-ECal reconstructed events when looking at the through-going muon control sample. The 2D reconstructed track quality check required that tracks contained no layer gaps. There was an unforeseen hit inefficiency in the DS-ECal which was not modelled in MC, which would cause tracks to be rejected because of the quality check. This effect is shown in Fig [INSERT FIGURE].

$$INSERTFIGURE \quad (1.15)$$

Because this problem was found after significant progress had been made on the MC selection, it was decided that this bug would be treated as a systematic uncertainty

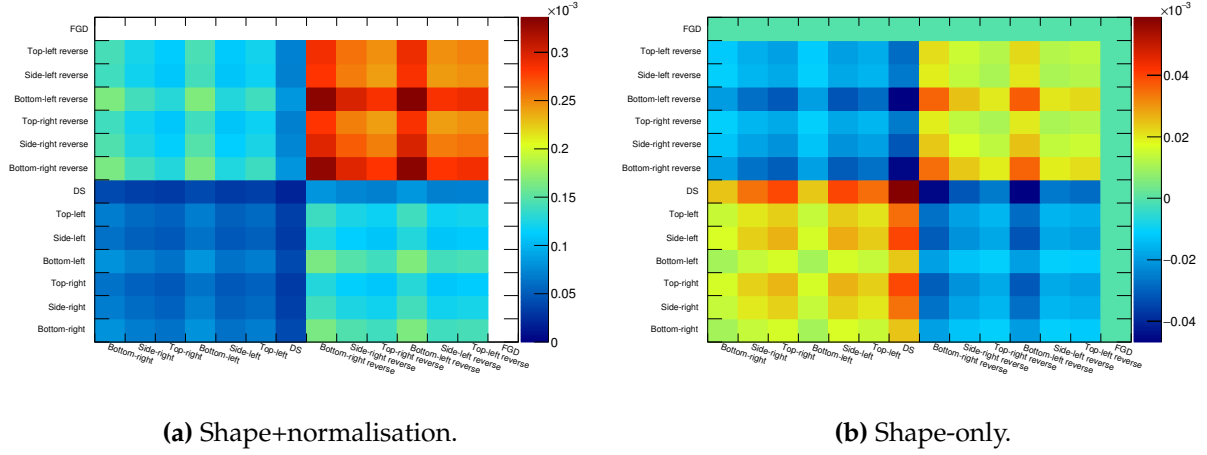


Figure 1.24: The sample covariance induced by the uncertainty in the ECal OOFV events.

in the analysis. To assess this ad hoc systematic uncertainty, the quality check was relaxed to allow 2D tracks to skip one layer in a given view. The reconstruction and selection was then applied to the sample and was compared to the nominal sample to construct the covariance matrices, which are shown in Fig. 1.25.

1.4.4 FGD detector uncertainty evaluation

The systematic uncertainties associated with the FGD are described extensively in [3]. In recent times, there has been a significant push to develop a framework in which to analyse and propagate systematic uncertainties associated with ND280 analyses,

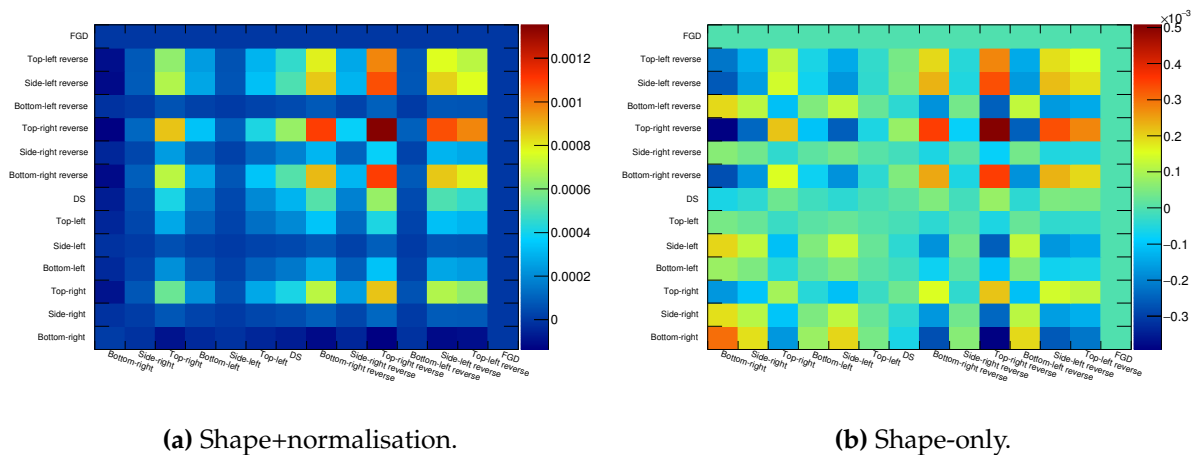


Figure 1.25: The sample covariance induced by the bug fix to the reconstruction.

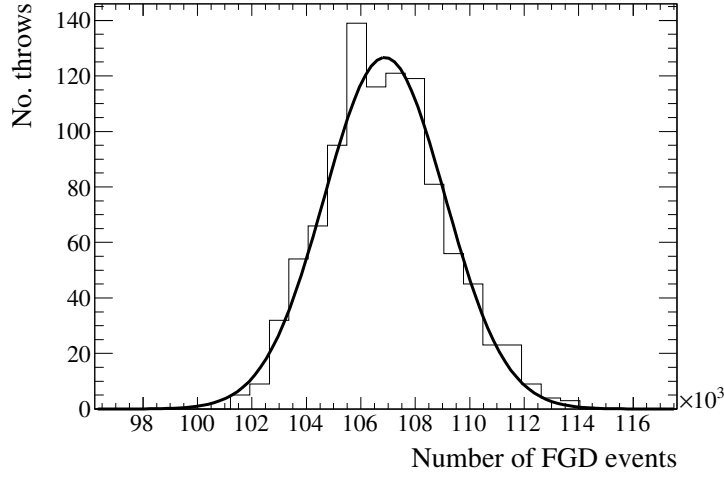


Figure 1.26: The number of selected FGD events for each systematic throw using PSYCHE. The smooth line is a Gaussian fit to the number of events.

which is called the PSYCHE framework. Even though outer detector analyses have not been incorporated into PSYCHE, this is not true for tracker analyses and so PSYCHE could be partly utilised in this analysis. The design of PSYCHE allows the user to enable/disable relevant systematic uncertainties and then simultaneously propagate them through an analysis. PSYCHE was set up to apply the FGD detector systematics and produce 1000 throws, in which every systematic uncertainty was simultaneously applied. The output, in the form of a distribution of a varied number of selected CC events in the FGD, was then used to define the systematic uncertainty. This distribution is shown in Fig. 1.26. The Gaussian fit to the thrown number of events implies that the number of FGD events is $(1.069 \pm 0.022) \times 10^5$ events i.e. a 2.01% uncertainty. However, the number of nominal selected events is 1.054×10^5 and so the shift in the number of events suggests an additional 1.41% error. Combining these in quadrature, the FGD detector systematic uncertainty was found to be 2.51%. This error purely relates to the number of CC events in the FGD and so does not correlate with any of the events in the ECal.

The total detector covariance matrices are shown in Fig. 1.27. These matrices were found by summing all of the detector related covariance matrices. The exception to this is the FGD detector uncertainty in which the single variance was added to the FGD on-diagonal bin. The largest uncertainty was found to be 5.04%. The total covariance matrices, combining all of the covariance matrices discussed above, are shown in Fig. 1.28. The largest uncertainty is 13.8%.

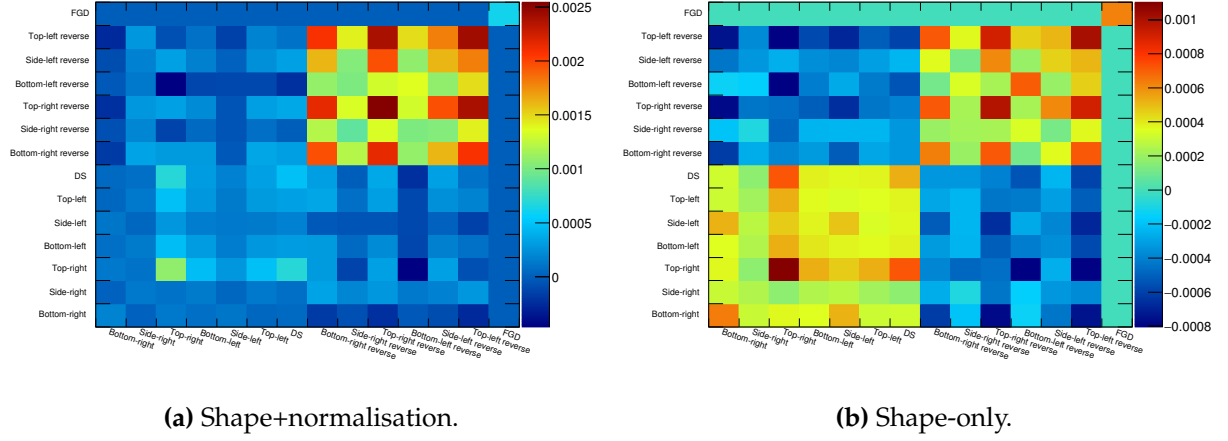


Figure 1.27: The sample covariance induced by all of the detector related uncertainties.

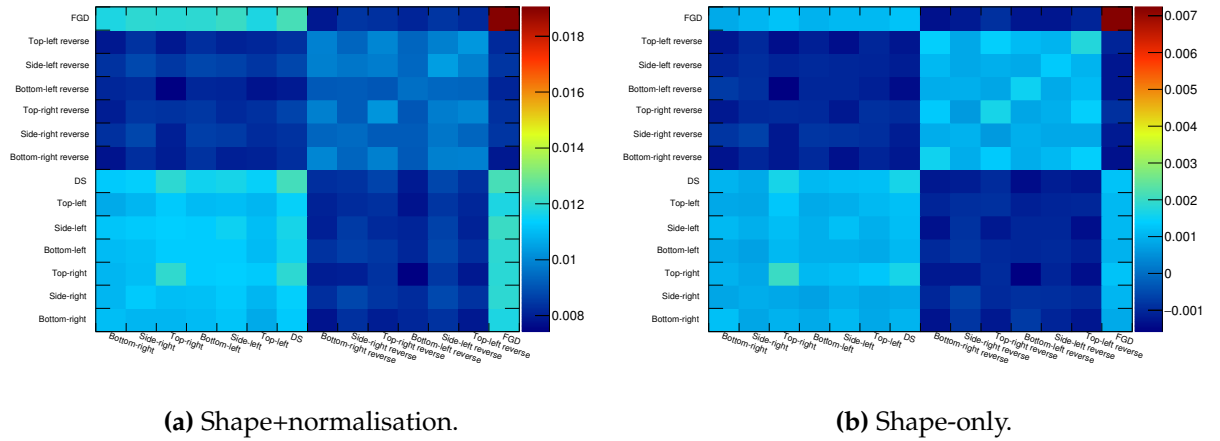


Figure 1.28: The sample covariance induced by all of the systematic uncertainties.

ECal module	1 prong topology efficiency (%)	2 prong topology efficiency (%)	3 prong topology efficiency (%)	4+ prong topology efficiency (%)
Barrel	46.3	54.2	66.0	65.4
DS	55.9	58.3	59.5	59.9

Table 1.8: The selection efficiencies from the GENIE-based Monte Carlo sample for each prong topology and ECal module.

The selection presented in chapter ?? heavily used Monte Carlo simulation of neutrino interactions to estimate the efficiency of the selection. At the core of the simulation, NEUT was used to initially generate the simulated neutrino interactions. Ideally, the selection efficiency should be a property of the selection itself. However, because of the use of a neutrino generator in the simulation, it is possible that the selection could contain an element of generator dependency. It is non-trivial to use a data control sample to assess a systematic uncertainty on the selection efficiency. In fact, the only feasible way to address this analysis issue would be to use multiple neutrino generators as inputs to the same ND280 simulation and compare the differences. Unfortunately, a large set of generators were not available for use in the time frame of the analysis. In ND280 software productions, a large amount of neutrino events based on the GENIE neutrino generator [10] are produced, separately to the NEUT events. So 2.5×10^{19} POT worth of GENIE events were passed through the ND280 software chain (including the enhanced reconstruction) and the selection presented in chapter ?. No GENIE-based sand MC has ever been produced in ND280 software productions. However, as this check only concerns changes to the selection efficiency, an absence of sand MC should not be a problem.

The selected events distributions, separated into the prong topologies, are shown in Fig. 1.29. The selection efficiencies for each prong topology are shown in Fig. 1.8. As described in section ??, the selection efficiencies do not include neutrino events which were not reconstructed. So, the topology combined efficiencies which do include said events are shown in table 1.9.

Overall, there are differences in the selection efficiencies between NEUT and GENIE, particularly for the DS-ECal. However, when combining the topologies and including

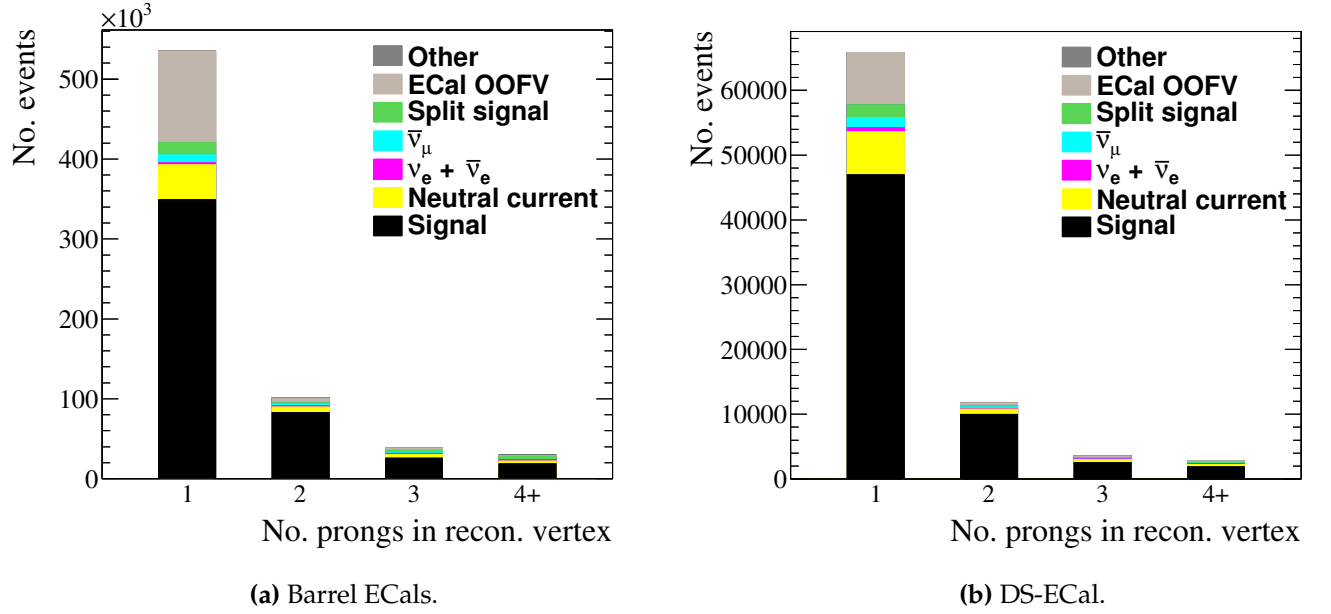


Figure 1.29: The number of selected events in the GENIE-based Monte Carlo sample, separated out into the prong topologies. Each event is categorised by the associated truth information from the simulation.

ECal module	Efficiency (%)
Barrel	40.7
DS	49.9

Table 1.9: The topology combined efficiency for the GENIE-based Monte Carlo sample. The efficiency values include simulated events which were not reconstructed.

the events below reconstruction threshold, the efficiency differences are somewhat minor. As described at the start of this chapter, the DS-ECal will form the primary target and so it is the DS-ECal efficiency which is the most important. The DS-ECal efficiency uncertainty is simply defined as the difference between the NEUT and GENIE efficiencies, which is $53.0\% - 49.9\% = 3.1\%$. This error as a fraction of the nominal uncertainty is 5.85%.

1.4.5 The measured cross-section uncertainty

Using all of the information described in chapter ?? and chapter 1, the ν_μ CC-inclusive cross-section on lead will be calculated as

$$\sigma_{\text{Pb}}^{\text{CC}} = \frac{R_{\text{Pb}} N_{\text{Pb}}^{\text{MC}} \eta_{\text{Pb}}}{T_{\text{Pb}} \Phi^{\text{MC}} \epsilon_{\text{Pb}}}, \quad (1.16)$$

where R_{Pb} is the Pb normalisation returned from the rate fit, $N_{\text{Pb}}^{\text{MC}}$ is the number of selected events in the DS-ECal, T_{Pb} is the number of Pb nuclei present in the DS-ECal, Φ^{MC} is the total flux used to create the MC and η_{Pb} and ϵ_{Pb} are the purity and efficiency of the DS-ECal selection. The shape covariance matrix (shown in Fig. 1.28b) will be used as an input to the fit. This covariance matrix includes uncertainties on the flux and (partly) on the number of Pb nuclei. Therefore, the error on R_{Pb} will provide the error from the shape systematic uncertainties which means it is only necessary to include a couple of extra uncertainties on the actual measurement of the cross-section. It is necessary to include the normalisation uncertainty on the number of selected DS-ECal events. This value, taken from Fig. 1.28a is 11.1%. The error on the efficiency, as discussed above, is 5.85%. Despite having investigated systematic differences in the ECal mass, the simulation does not include antimony in the absorber sheets. The absorber sheets are quoted as having a 0.2% uncertainty on the antimony doping [8] which also needs to be included in the final uncertainty. Combining all of this information, the final uncertainty on the measured cross-section will be

$$\sigma_{\sigma_{\text{Pb}}^{\text{CC}}} = \sigma_{R_{\text{Pb}}} \pm 12.5\%, \quad (1.17)$$

where $\sigma_{R_{\text{Pb}}}$ is the error on R_{Pb} returned from the rate fit.

Normalisation set	R^{Pb}	R^{C}	R^{Other}
1	1.0	1.0	1.0
2	3.0	1.0	1.0
3	0.7	1.5	0.8
4	1.0	0	1.0
5	-0.3	1.0	1.0

Table 1.10: The normalisation sets used in the machinery validation of the ECal rate fit.

1.5 Validation of method

The ECal rate fit described in section 1.3 requires validation to ensure that not only the machinery works as intended, but also to check that the fit is suitable for such an analysis. The first set of validation checks performed deals with the former case.

1.5.1 Validation of the fit machinery

The initial validation method used the same Monte Carlo described in section 1.2 to create toy datasets which were passed through the fitter. To create said datasets, new normalisation parameters, R^{Val} , were defined which are shown in table 1.10. These new normalisation parameters scale the relevant Monte Carlo templates.

For the MC population X in each sub-detector sample i , the normalisation parameter is applied to create the new population as

$$n_i^{X \text{ Val}} = R_i^{X \text{ Val}} n_i^{X \text{ Nom}}, \quad (1.18)$$

where $n_i^{X \text{ Nom}}$ is the number of events in population X of sub-detector sample i for the nominal MC. The total number of events in each sub-detector sample is then constructed as

$$N_i^{\text{Val}} = \sum_X n_i^{X \text{ Val}}. \quad (1.19)$$

The final step in creation of the toy datasets was to vary the total number of events in the sub-detector samples. The method chosen was the same as that used to generate

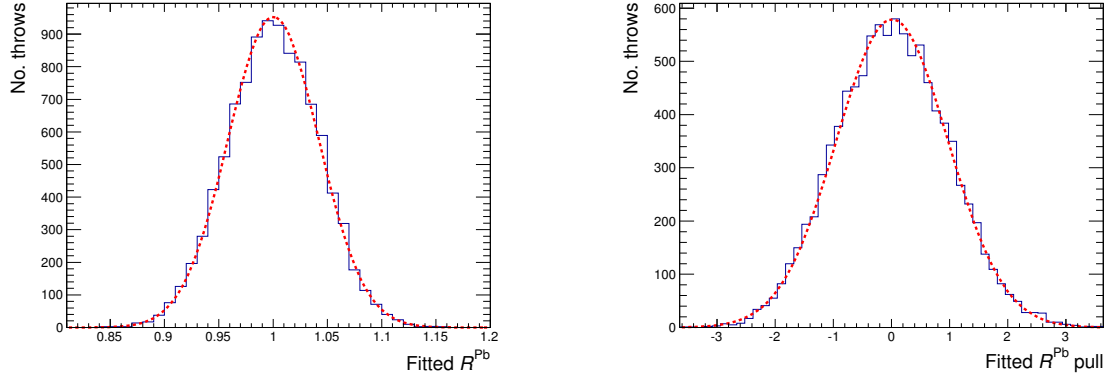
event weights for the systematic uncertainty studies described in section 1.4. The covariance matrix shown in Fig. 1.28b was cholesky decomposed and multiplied by a vector of random numbers, all drawn from a Gaussian of mean 0 and width 1. The sub-detector sample weights were then created by adding 1 to each element of the multiplied vector. Each scaled sub-detector sample, N_i^{Val} , was then reweighted by the relevant sample weight. This toy dataset was then passed to the fitter and its output recorded. The steps outlined above describe one pass of the fitter. This process was repeated 10,000 times for each validation normalisation set.

The value of R^{Pb} returned for each throw in parameter set 1 is shown in Fig. 1.30a. The Gaussian fit quantifies the performance of the fit: the mean specifies the final value and the width specifies a pseudo-error. To further quantify how well the machinery works, the parameter pull, P_i^X , is defined as

$$P_i^X = \frac{R_i^{X \text{ Fit}} - R_i^{X \text{ Val}}}{\sigma_i^{X \text{ Fit}}}, \quad (1.20)$$

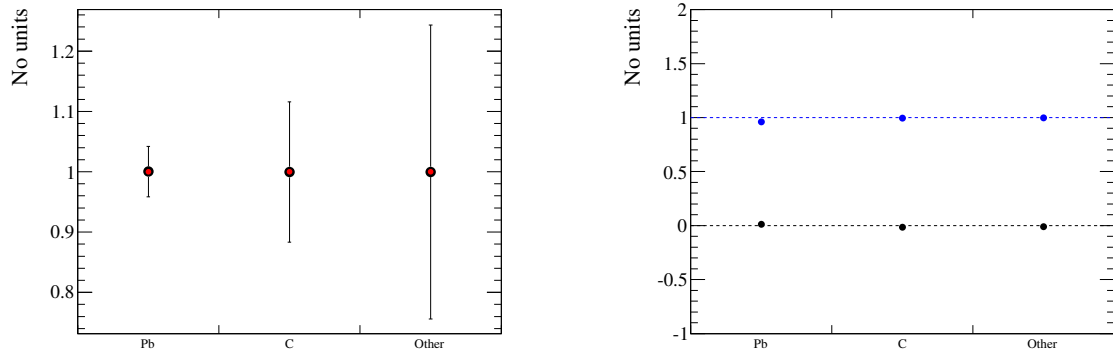
where $R_i^{X \text{ Fit}}$ and $\sigma_i^{X \text{ Fit}}$ are the value and error of parameter X returned from the fit for throw i respectively. The MINOS routine used in the fitter returns asymmetric errors and so care must be taken when deciding which error to use in equation 1.20. It was decided that the positive error would be used when the fitter returned a value higher than the input. P^{Pb} for parameter set 1 is shown in Fig. 1.30b. The Gaussian fit to this distribution has a width close to, but slightly below, 1. This suggests that the errors returned by the MINOS routine very slightly overestimate the error, but are still a very good representation of the uncertainty. The performance of the fit for all of the normalisations in normalisation set 1 are shown in Fig. 1.31. In each case, the fitter returns the correct normalisation and the pull characteristics suggest the fit returns the correct error and is unbiased. The performance of the fit for normalisation sets 2, 3, 4 and 5 are shown in Fig. 1.32, Fig. 1.33, Fig. 1.34 and Fig. 1.35 respectively. Generally speaking, the fitter performs well for all of the situations. It is clear that the output is sensible even when extreme situations are presented (normalisation set 2) as well as when minor shifts are applied to all three parameters (normalisation set 3). Even in un-physical situations the fitter returns the input correctly (normalisation set 5). It is only when a template is removed (normalisation set 4) that the fitter becomes biased. However, as Fig. 1.34b shows, the bias is minor.

As discussed in section 1.2, a set of barrel ECal samples which contain events that



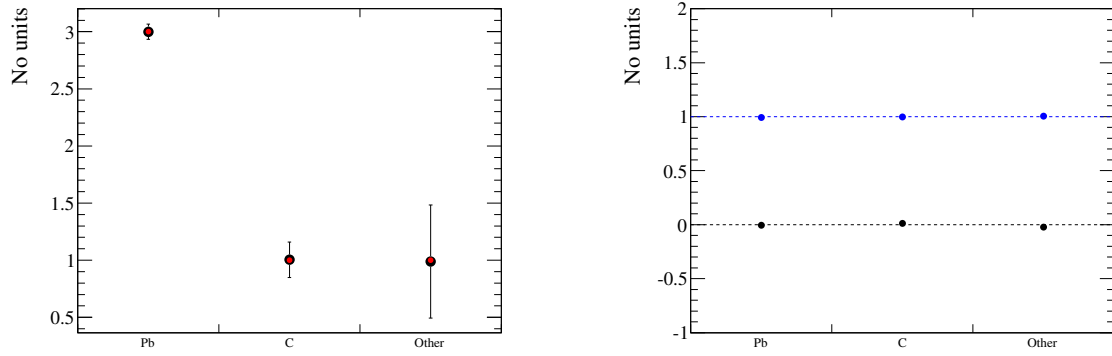
(a) The returned parameter for each throw. The fitted Gaussian has a mean of 1.00 and a width of 0.04. (b) The parameter pull for each throw. The fitted Gaussian has a mean of 0.01 and a width of 0.96.

Figure 1.30: Distributions of R^{Pb} related information returned from the fitter for the 10,000 fake data throws. The red dashed lines are Gaussian fits to the distributions.



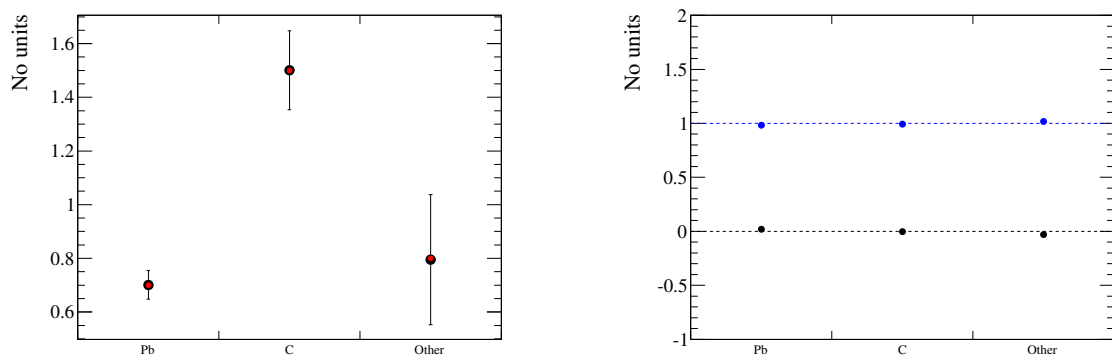
(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red). (b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

Figure 1.31: Performance of the fit for the 10,000 fake data sets when using normalisation set 1.



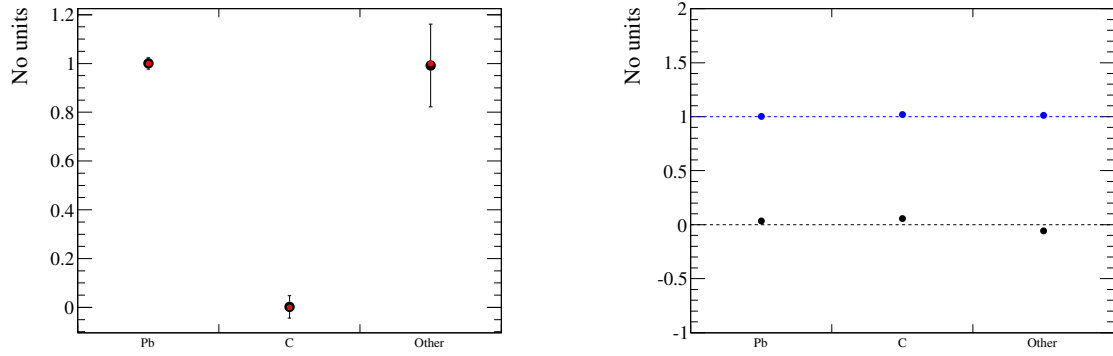
(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red).
 (b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

Figure 1.32: Performance of the fit for the 10,000 fake data sets when using normalisation set 2.



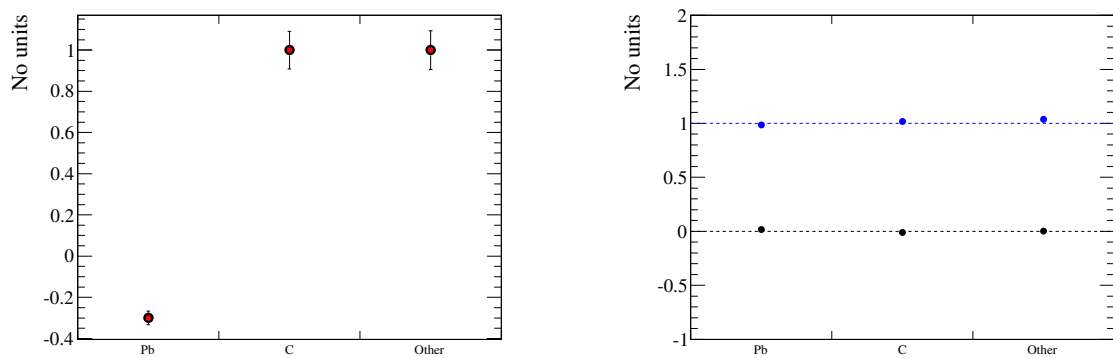
(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red).
 (b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

Figure 1.33: Performance of the fit for the 10,000 fake data sets when using normalisation set 3.



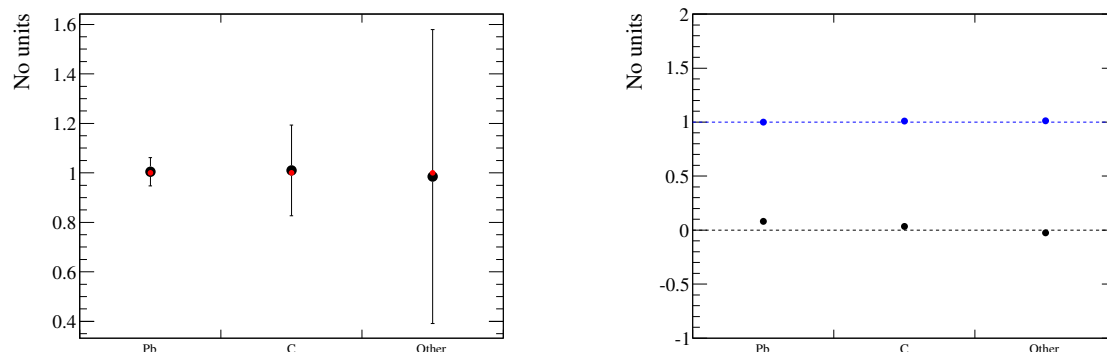
(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red).
 (b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

Figure 1.34: Performance of the fit for the 10,000 fake data sets when using normalisation set 4.



(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red).
 (b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

Figure 1.35: Performance of the fit for the 10,000 fake data sets when using normalisation set 5.



(a) The mean fitted values for each normalisation parameter (black) compared with the input normalisation values (red). (b) The means (black) and widths (blue) of the pull distributions after fitting with a Gaussian.

Figure 1.36: Performance of the fit with the reverse ECal samples removed for the 10,000 fake data sets when using normalisation set 1.

failed the selection cuts are included. To test that the inclusion of this sample is not degrading the performance, an extra set of 10,000 fake data throws were performed for normalisation set 1. However, the reverse ECal samples were not included in the fitter. The results of this test are shown in Fig. 1.36. Comparing this with the performance of the fitter with the reverse sample included (Fig. 1.31), it is clear that the absence of the reverse sample causes a wider range of normalisation values to be returned from the fitter. Additionally, a small bias is introduced to all three normalisations. While the issues present in the reverse-absent fitter are minor, It should be clear that the inclusion of the reverse ECal samples leads to more accurate and precise results.

1.5.2 Physics validation of the fitter

If the fitter truly works as intended, then it should be able to successfully fit one cross-section model to another. As described in section 1.4, T2K produces a set of GENIE-based ND280 events during software production in addition to the NEUT-based events. As the NEUT events are being used for the MC templates in the fitter, the GENIE samples make an excellent candidate for a data data test.

The same GENIE sampled used in section 1.4 was passed through the selection and binned into the ECal samples. Unfortunately, no GENIE-based sand Monte Carlo or FGD sample were available for this study. So, the same NEUT-based sand MC and

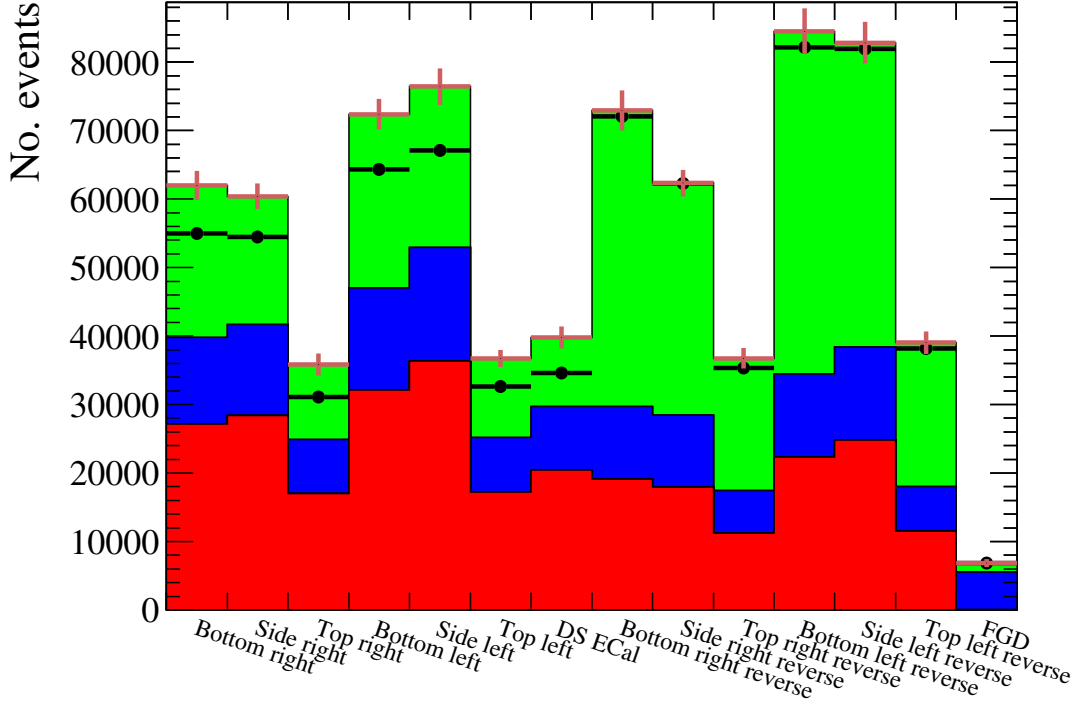


Figure 1.37: The prefit number of NEUT Monte Carlo events compared with GENIE fake data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the GENIE fake data.

FGD sample used in the MC template generation were used to fill this gap. The MC templates were normalised to the same POT as the fake data. The pre-fit samples are shown in Fig. 1.37.

Before the NEUT MC templates were actually fit to the genie fake data, it was necessary to calculate what the fit should be predicting. To calculate what the true GENIE to NEUT cross-section ratios are, the truth information provided by the MC simulation was used. For both the GENIE and NEUT samples, the number of true interactions in the DS-ECal were recorded, including the target element. These numbers were weighted to the same POT and the ratio taken. These values are shown in table 1.11.

The MC templates and fake data shown in Fig. 1.37 were then passed to the fitter, with the fitted normalisations shown in table 1.12 and a comparison of the post-fit MC templates with the GENIE fake data shown in Fig. 1.38. The reduced χ^2 of the fit was

$\frac{\sigma_{\text{GENIE}}^{\text{Pb}}}{\sigma_{\text{NEUT}}^{\text{Pb}}}$	$\frac{\sigma_{\text{GENIE}}^{\text{C}}}{\sigma_{\text{NEUT}}^{\text{C}}}$
0.908	0.926

Table 1.11: The GENIE to NEUT cross-section ratios, calculated from the truth information provided by the ND280 simulation.

found to be 1.71.

The values returned from the fit do not agree with the true values when considering only the errors returned from the fit. If the extra 12.5% uncertainty defined in equation 1.17 is added in quadrature with the positive error on R^{Pb} , the fitted normalisation is in agreement with its corresponding true value. It should be noted at this point that both the sand muon component and the FGD sample in the fake data are in fact NEUT, rather than GENIE. So, this test is somewhat unrealistic. While it is not trivial to produce a GENIE-based sand muon Monte Carlo, the FGD sample in the genie fake data can be re-weighted to better represent what the FGD sample should be. Using the carbon cross-section ratio shown in table 1.11, the carbon component of the FGD sample in the GENIE fake data was scaled. This modified set of fake data was then passed to the fitter, along with the NEUT MC templates. The returned normalisations are shown in table 1.13 and the comparison of the fitted MC templates with the fake data is shown in Fig. 1.39. It is clear that the situation is much improved as both the lead and carbon normalisation have moved towards their true values. The carbon normalisation still disagrees with the simulated value, however it has already been mentioned that the sand muon component of the GENIE fake data is actually NEUT. Ideally, a GENIE-based set of sand muon MC events would be used in the fake data but due to time constraints this was not possible. As the lead normalisation agrees with the generator prediction, this test can be considered a success.

Normalisation parameter	Fitted value
R^{Pb}	$0.79^{+0.08}_{-0.08}$
R^{C}	$1.19^{+0.15}_{-0.13}$
R^{Other}	$0.94^{+0.08}_{-0.08}$

Table 1.12: The values returned from the fitter when passing the NEUT MC templates and GENIE fake data shown in Fig. 1.37.

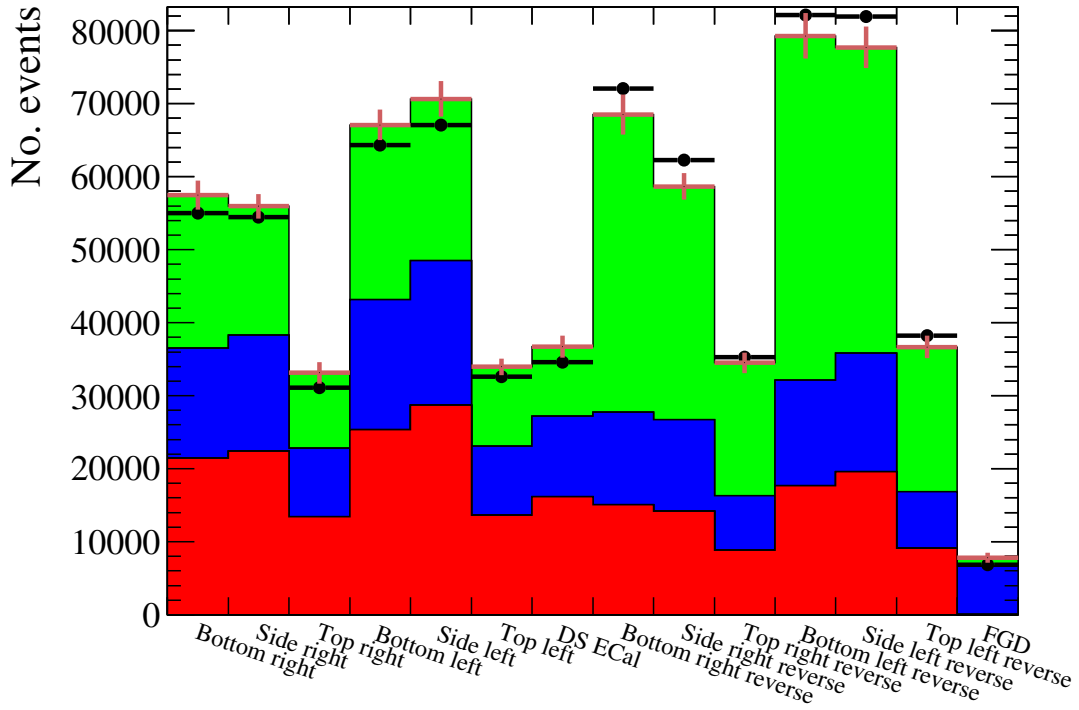


Figure 1.38: The post-fit number of NEUT Monte Carlo events compared with GENIE fake data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the GENIE fake data.

Normalisation parameter	Fitted value
R^{Pb}	$0.83^{+0.08}_{-0.08}$
R^{C}	$1.09^{+0.14}_{-0.12}$
R^{Other}	$0.95^{+0.08}_{-0.09}$

Table 1.13: The values returned from the fitter when passing the NEUT MC templates and GENIE fake data which contained a re-weighted FGD sample.

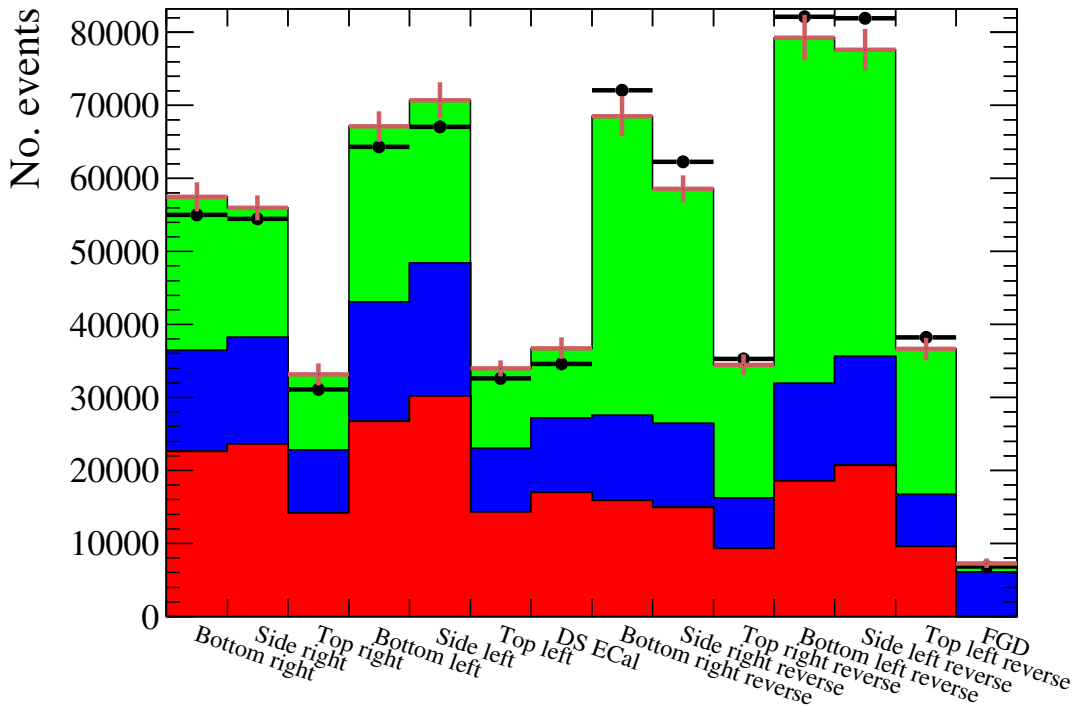


Figure 1.39: The post-fit number of NEUT Monte Carlo events compared with GENIE fake data which has a re-weighted FGD sample. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the GENIE fake data.

1.6 Applying the fit to ND280 data

To measure the ν_μ CC inclusive cross-section on lead, a sample of ECal data events from the neutrino beam must be passed through the selection presented in chapter ?? . However, an extra data quality check must be applied to the data events before the selection takes place. These data quality checks are separated into two parts. The first check ensures that the components of the beamline were nominally operating when the

Sample	No. Selected (MC)	No. Selected (data)
Bottom right ECal	55749	54396
Side right ECal	54011	56583
Top right ECal	32006	32233
Bottom left ECal	65169	65448
Side left ECal	68031	66095
Top left ECal	32760	33470
DS-ECal	35420	34372
Bottom right ECal (Reverse)	72112	70885
Side right ECal (Reverse)	61159	64445
Top right ECal (Reverse)	35997	36191
Bottom left ECal (Reverse)	83059	80967
Side left ECal (Reverse)	81240	84145
Top left ECal (Reverse)	38282	40293
FGD	5857	5992

Table 1.14: The number of selected events in each sample.

data was collected. This involves basic hardware checks, that the horn currents were operating within 5 kA of their mean value, that the beam angle was within 1 mrad of its mean value and that the muon rate was within 5% of its mean value. The second check ensures that ND280 itself was operating optimally during data collection. A dedicated group studies the sub-detector outputs on a per spill basis, ensuring that the sub-detectors (including the ECals) were operating correctly. The output of these two checks are a pair of boolean flags which state whether the data that was provided by the beam and collected by ND280 were sound. So, before data events are passed to the selection, these flags are queried and if either flag reports bad data, the event is rejected.

The data quality checks and the event selection described in chapter ?? were applied to 1.31×10^{20} POT of neutrino beam data collected from T2K run 3C. The FGD sample was also taken from run 3C, corresponding to 1.35×10^{20} POT. The number of selected events in each sample is summarised in table 1.14.

The samples were then passed to the fitter described in section 1.3 and the parameters found are shown in table 1.15. The minimum χ^2 per degree of freedom was found

Normalisation parameter	Fitted value
R^{Pb}	$1.58^{+0.10}_{-0.11}$
R^{C}	$1.87^{+0.21}_{-0.18}$
R^{Other}	$0.26^{+0.06}_{-0.05}$

Table 1.15: The values returned from the fitter when processing the NEUT MC templates with T2K data.

to be 13.4. The MC template comparison with the data before and after the χ^2 fit are shown in Fig. 1.40 and Fig. 1.41 respectively. After the normalisation parameters had been returned from the fit, the χ^2/NDOF as a function of the normalisation parameters could be found. This was achieved by varying a single normalisation parameter whilst keeping all other normalisation parameters fixed at their χ^2 minimum value. At each step in the single normalisation parameter variation, the χ^2/NDOF was calculated. These results are shown in Fig. 1.42.

It should be clear from Fig. 1.41 that the Monte Carlo is not a good fit to the data. Before the fitting, it is clear that there are shape differences between the data and Monte Carlo templates. For example, there is a minor data deficit in the bottom-right ECal module and a larger data excess in the side-right ECal. A similar situation occurs between the bottom-left reverse and side-left reverse samples. Fig. 1.43 shows the shape-only correlation matrix for the samples used in the fit. The correlation matrix shows that all of the barrel samples are highly correlated with each other, as are all of the barrel-reverse samples. This suggests that the fit has little freedom to easily resolve the shape differences. This fact is enforced by the pre-fit information shown in Fig. 1.41, where the only way to resolve such shape differences with highly correlated constraints is to dramatically alter the normalisation of the templates. The pre-fit information shown in Fig. 1.40 can help to illuminate the underlying issue. An area normalised comparison of a sub-set of the MC templates with the run 3C data is shown in Fig. 1.44 for the barrel samples and Fig. 1.45 for the barrel-reverse samples. As Fig. 1.44a and Fig. 1.45a show, the shape of the ECal interaction templates in both the barrel and barrel-reverse samples generally follow the data shape. This is not the case for the shape of the other template (Fig. 1.44b and Fig. 1.45b). This finding suggests that it is the composition of the other template that is the cause of the fit failure. The other template contains all events which are not CC-carbon or CC-lead interactions.

As was found in chapter ?? and shown in Fig. ??, the dominant background is the ECal OOFV background which is the primary event category which falls into the other template. The number of events in the other template as a function of true neutrino energy is shown in Fig. 1.46 and Fig. 1.47 for the barrel and barrel-reverse samples respectively where the distributions have been separated into the true neutrino interaction location. In all cases, entering backgrounds which originate from the magnet and sand are the dominant contributor. These events sit in a region far away from the ND280 tracker which is the only place where the neutrino flux has been studied in great detail.

Comparisons of data and MC can help to illuminate what the underlying cause is. Fig. 1.48 shows the number of events which pass the fiducial volume cut as a function of the number of hits in the reconstructed ECal cluster for both data and MC. Firstly, it is important to state why this particular set of events are being shown; the definition of the reverse selection is any event which passes the fiducial volume cut i.e. all events

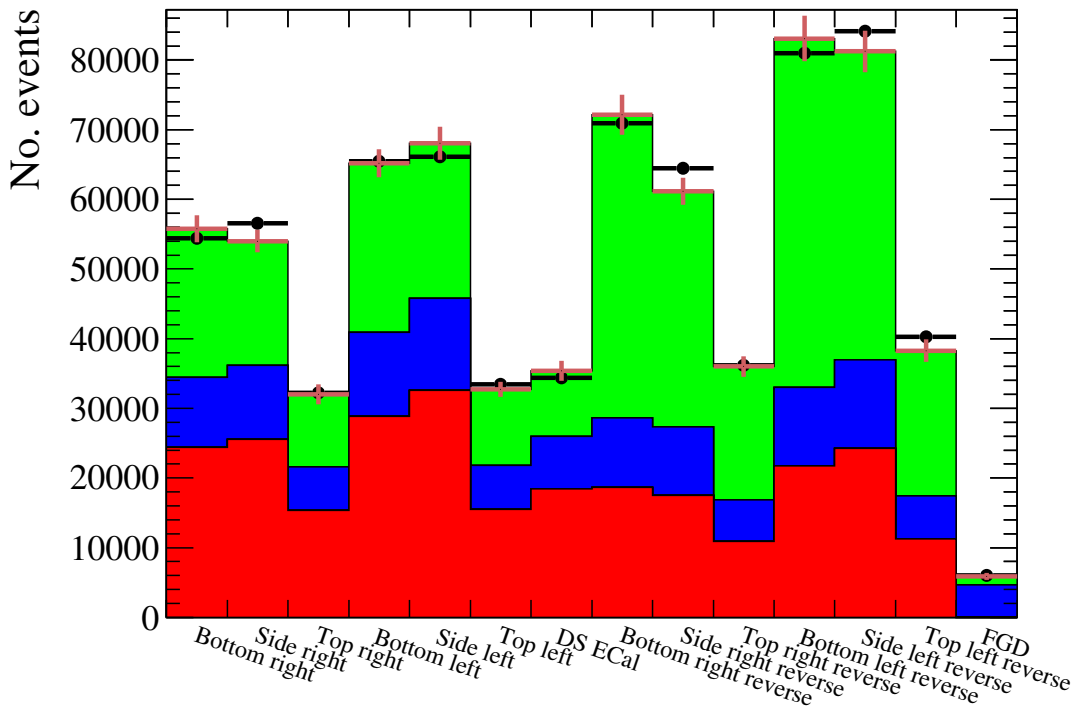


Figure 1.40: The pre-fit number of NEUT Monte Carlo events compared with T2K data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the T2K data.

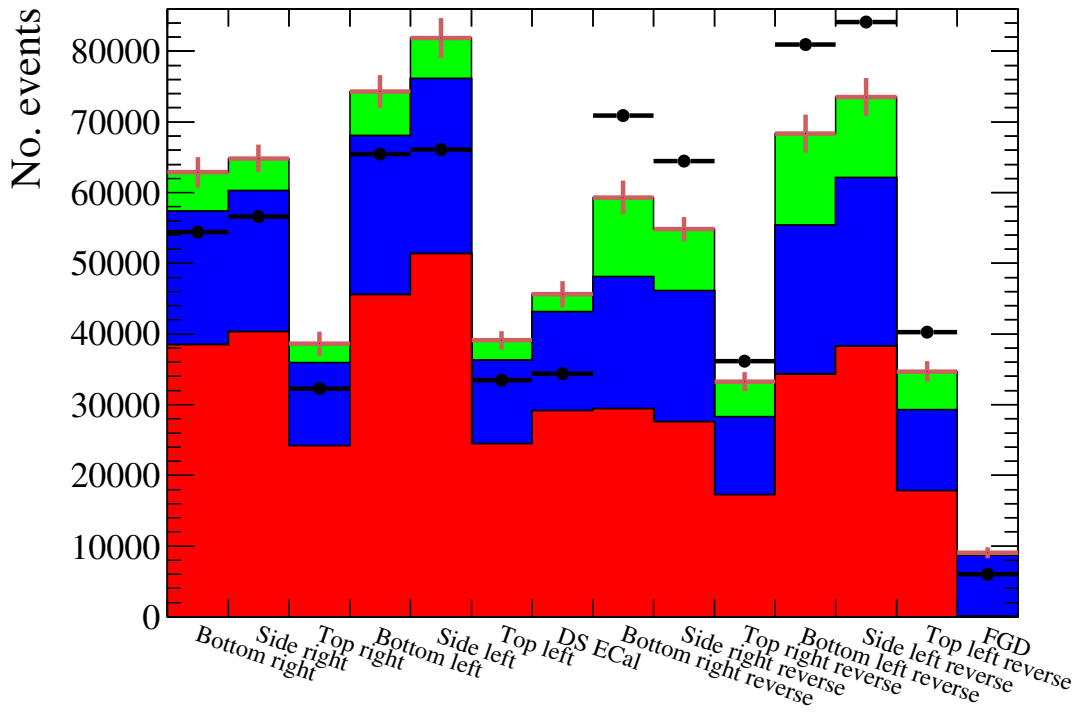


Figure 1.41: The post-fit number of NEUT Monte Carlo events compared with T2K data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the T2K data.

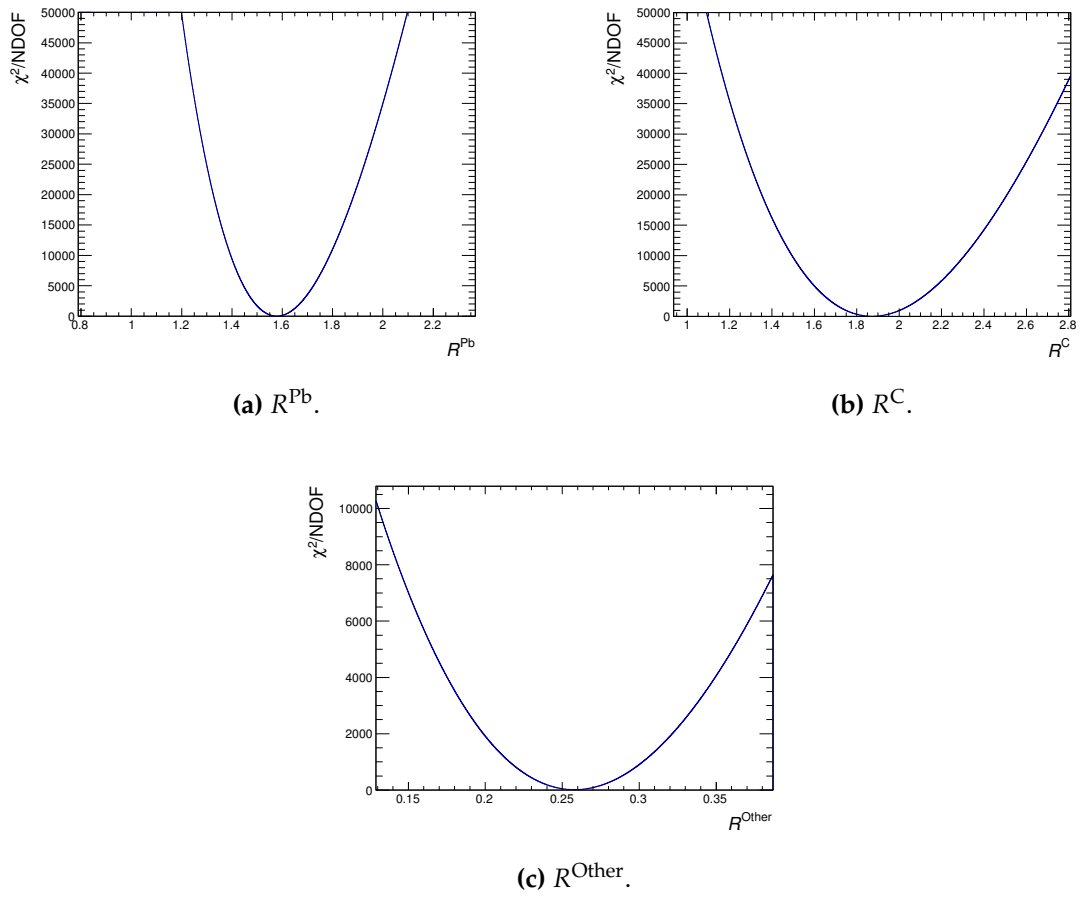


Figure 1.42: The χ^2/NDOF as a function of the normalisation parameters.

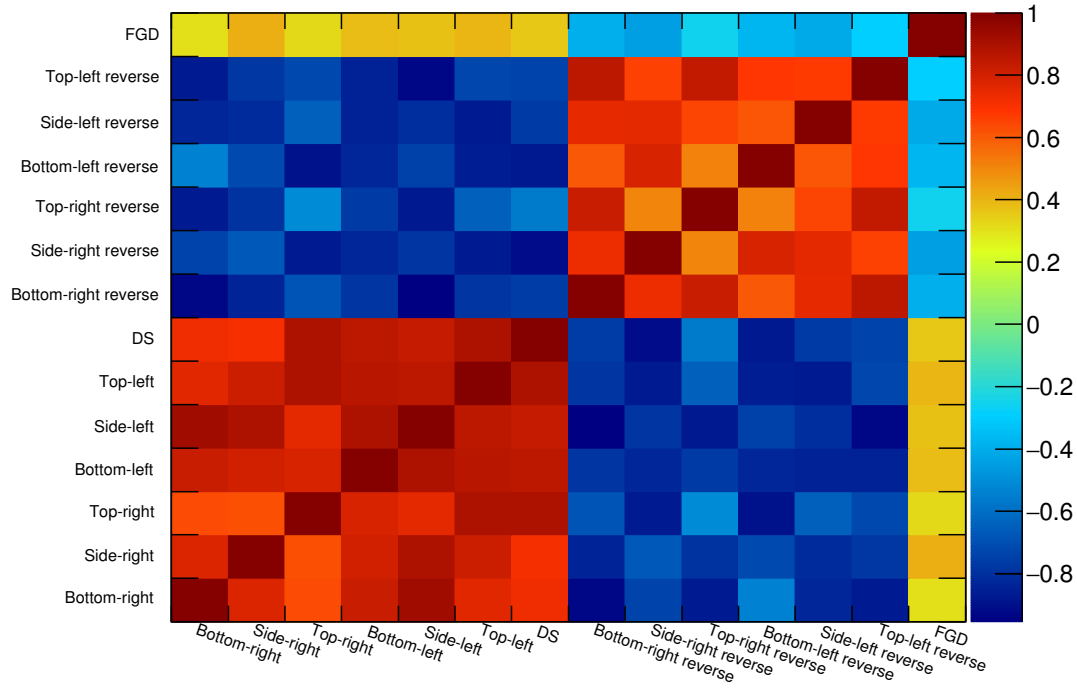
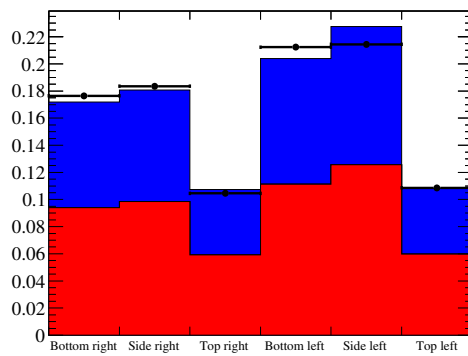
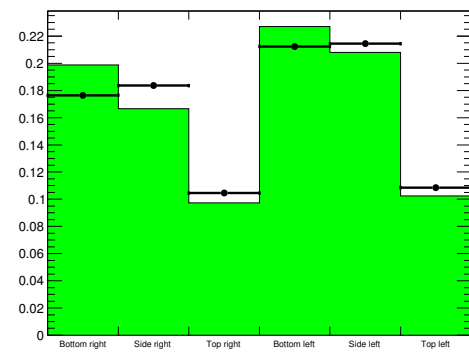


Figure 1.43: The shape-only correlation matrix for the samples used in the fit.



(a) The ECal interaction templates.



(b) The other template.

Figure 1.44: An area-normalised comparison of a sub-set of the MC templates with data for the barrel ECal samples. The coloured histograms and data points are the MC Templates and data respectively. The red, blue and green histograms are the lead, carbon and other templates respectively.

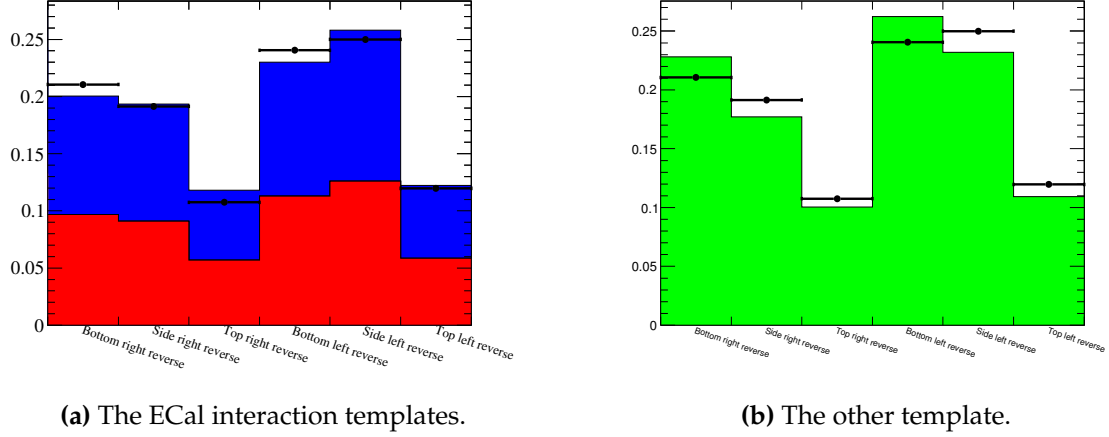
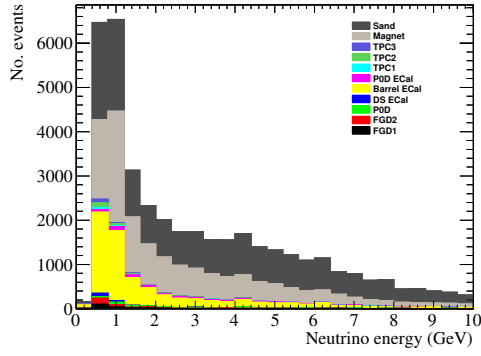
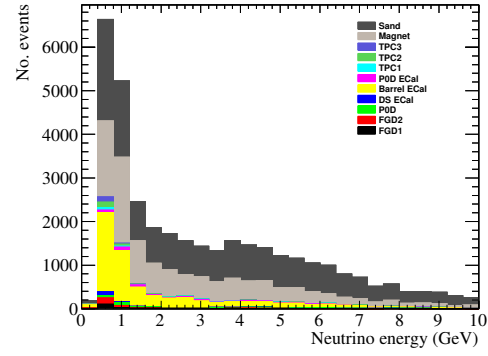


Figure 1.45: An area-normalised comparison of a sub-set of the MC templates with data for the barrel-reverse ECal samples. The coloured histograms and data points are the MC Templates and data respectively. The red, blue and green histograms are the lead, carbon and other templates respectively.

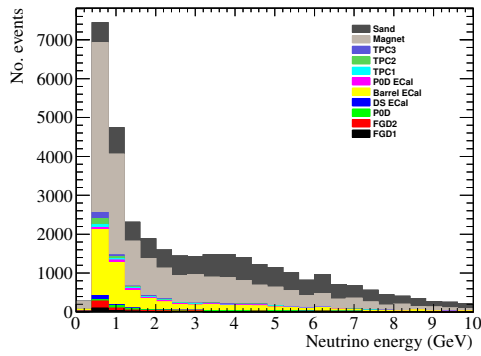
in Fig. 1.48 have to enter the ECal rate fit in one of the samples (selected or reverse). What is particularly interesting about these distributions is that there appears to be two significant areas of discrepancy between data and MC. The first discrepancy is an excess of data events with a low number of hits in the reconstructed clusters and is common to all barrel-ECal modules. A handscan of these events in the event display showed well reconstructed, 1-prong like events in each ECal module which appear to aim forwards, towards the magnet region. While an indepth study of these events is needed, it is possible that a mismodelling of the ν_μ cross-section on iron has caused such events to not appear in the MC. The second significant area of discrepancy, also shown in Fig. 1.48, is where there the number of hits in the reconstructed clusters is between 30 and 40. Unlike the first discrepancy, this problem is only an issue for all left ECals and the bottom-right ECal. What is notable about this discrepancy is that it sits in a region where the sand MC reaches its maximum. It has already been discussed that the sand MC is not well modelled and so it is not surprising that a discrepancy between data and MC would occur here. It is, however, troubling that these discrepancies exist after the fiducial volume cut has been applied which means that this undefined behaviour enters the analysis in some way. The efficiency of these events under the cut streams seem to have a module dependence and, as they have to appear in the analysis somewhere, this causes a non-trivial migration of such events between the selected samples and the reverse samples.



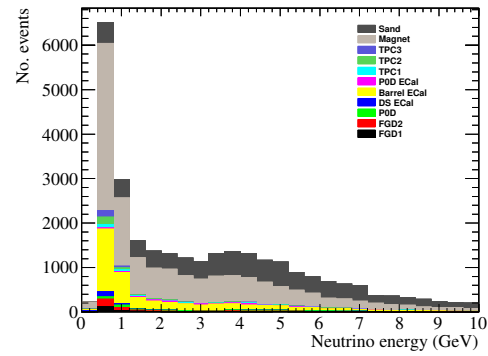
(a) Bottom left ECal.



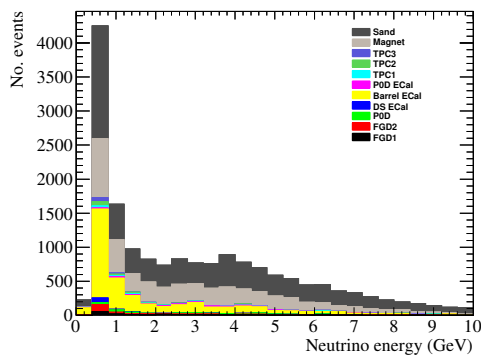
(b) Bottom right ECal.



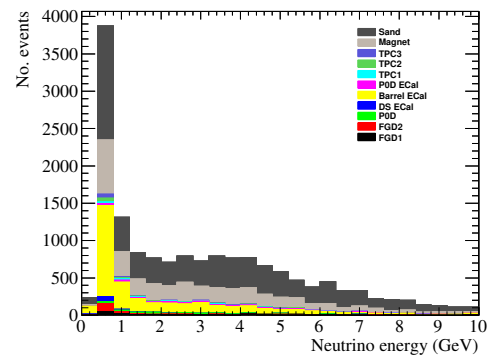
(c) Side left ECal.



(d) Side right ECal.

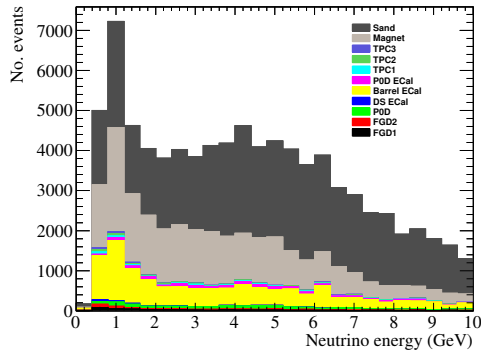


(e) Top left ECal.

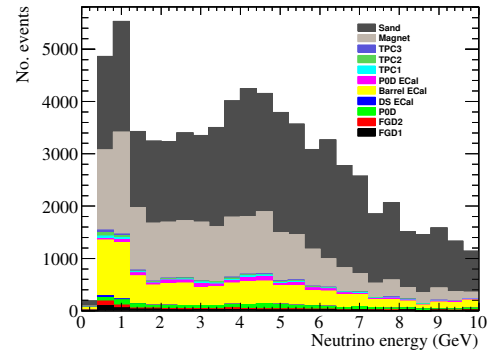


(f) Top right ECal.

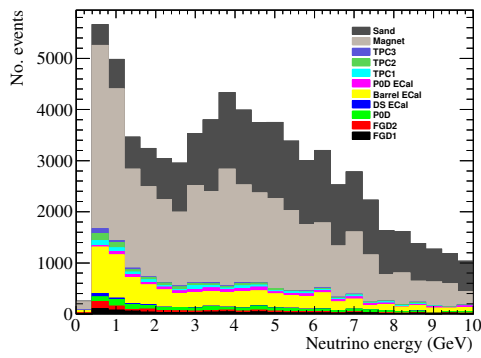
Figure 1.46: The number of events in the other template as a function of true neutrino energy for the barrel ECal samples. The distributions are broken down by true neutrino interaction location.



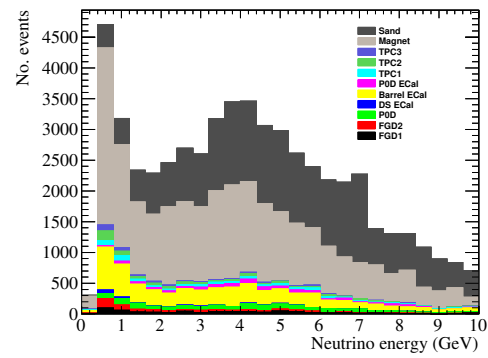
(a) Bottom left ECal.



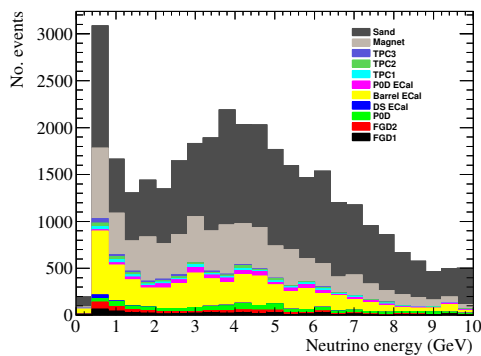
(b) Bottom right ECal.



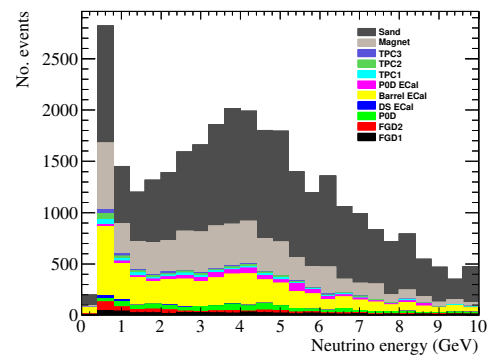
(c) Side left ECal.



(d) Side right ECal.



(e) Top left ECal.



(f) Top right ECal.

Figure 1.47: The number of events in the other template as a function of true neutrino energy for the barrel reverse ECal samples. The distributions are broken down by true neutrino interaction location.

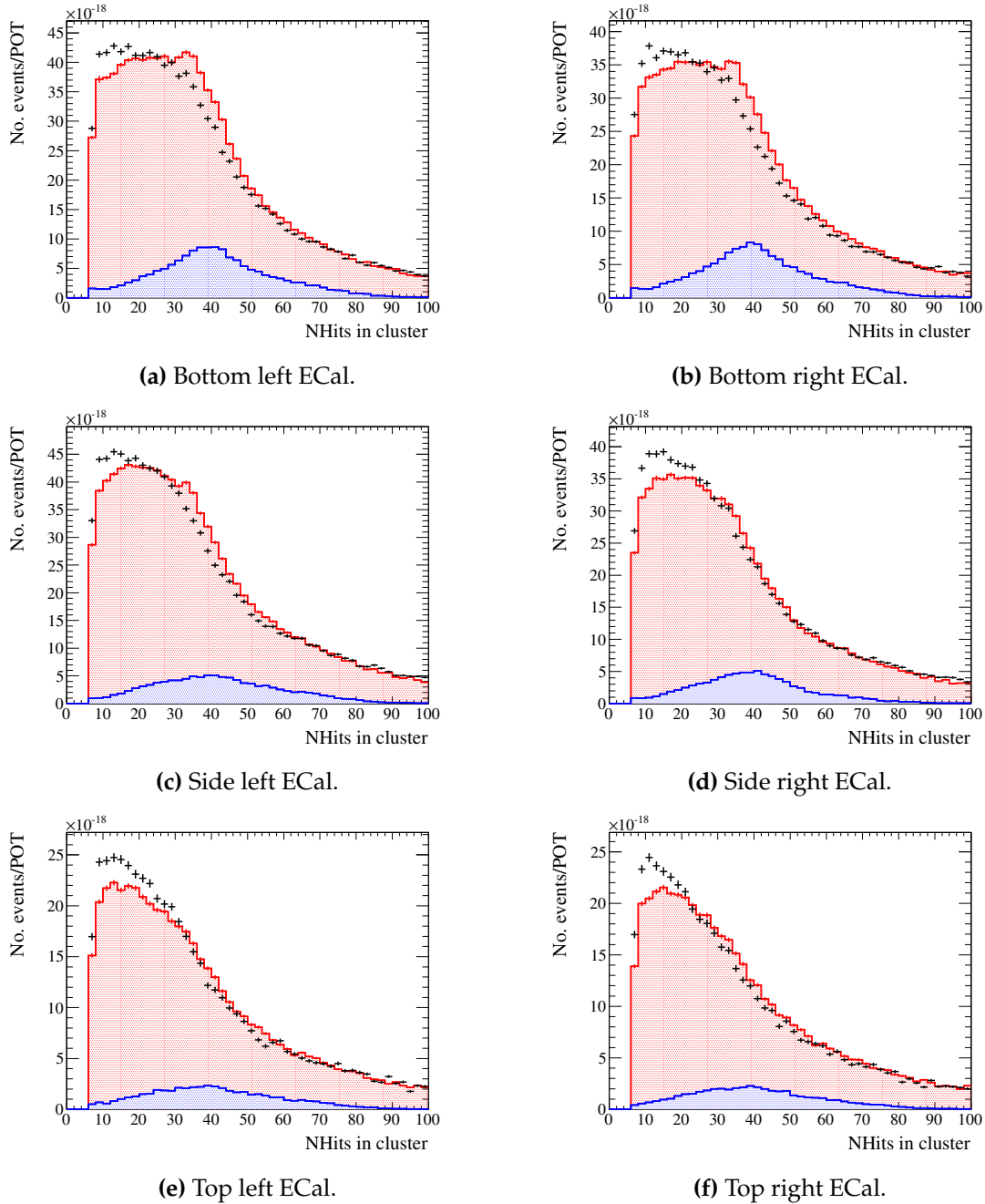


Figure 1.48: The number of events which pass the fiducial volume cut as a function of the number of hits in the reconstructed cluster. The red and blue histograms show the number of beam MC events and sand MC events respectively. The black points show the run 3C data.

Under normal circumstances, the background modelling needs extra study. However, due to time constraints, it is sufficient to introduce an adhoc uncertainty which encompasses the lack of background understanding. The background modelling has already been discussed as a possible cause of the fit failure and the barrel-reverse samples, which contain the largest background, can be used to quantify this lack of understanding. Table 1.16 shows the ratio of the number of events in each barrel sample to the number of events in each barrel-reverse sample for both data and MC. Also included in table 1.16 is the percentage difference between the data and MC ratio and it should be noted that this percentage difference has a geometric dependence. This strong dependence provides an extra hint that this background uncertainty is perhaps correlated with the neutrino beam flux. Regardless of the provenance of the underlying issue, the ratios presented in table 1.16 provide a good candidate to describe this adhoc uncertainty. The ratio itself essentially describes the size of the signal enhanced regions to the background enhanced regions. By quantifying this data and Monte Carlo difference, a crude uncertainty on the migration of events between the signal and background regions is crudely assessed. To include this extra information in the fit, the percentage difference shown in the fourth column of table 1.16 needs to be included in the shape-only covariance matrix. Specifically, for each barrel sample and its reverse counterpart, the corresponding percentage difference is added to the diagonal element of the covariance matrix. Using the modified covariance matrix as an input, the fit was then re-run using the NEUT MC templates and run 3C data. The number of events (post-fit) for each sample are shown in Fig. 1.49 and the normalisation parameters found by the fit are shown in table 1.17. The minimum χ^2/NDOF returned by the fit was 2.54.

The output of the fitter is more reassuring. The reduced χ^2 , while still a little high, suggests that the Monte Carlo is a much better fit to the data. Notably, all of the fitted normalisation parameters agree with the NEUT prediction within their shape-only error.

Barrel module	NSel/NRev (MC)	NSel/NRev (data)	% diff
Bottom right	0.773	0.767	0.7%
Side right	0.883	0.878	0.6%
Top right	0.889	0.891	0.2%
Bottom left	0.785	0.808	2.9%
Side left	0.837	0.785	6.2%
Top left	0.856	0.831	2.9%

Table 1.16: Comparison of the number of events between the barrel and barrel-reverse samples. The second and third column show the ratio of the number of events in the barrel samples to the number of events in the barrel-reverse samples for Monte Carlo and data respectively. The fourth column shows the percentage difference between the Monte Carlo and data ratios.

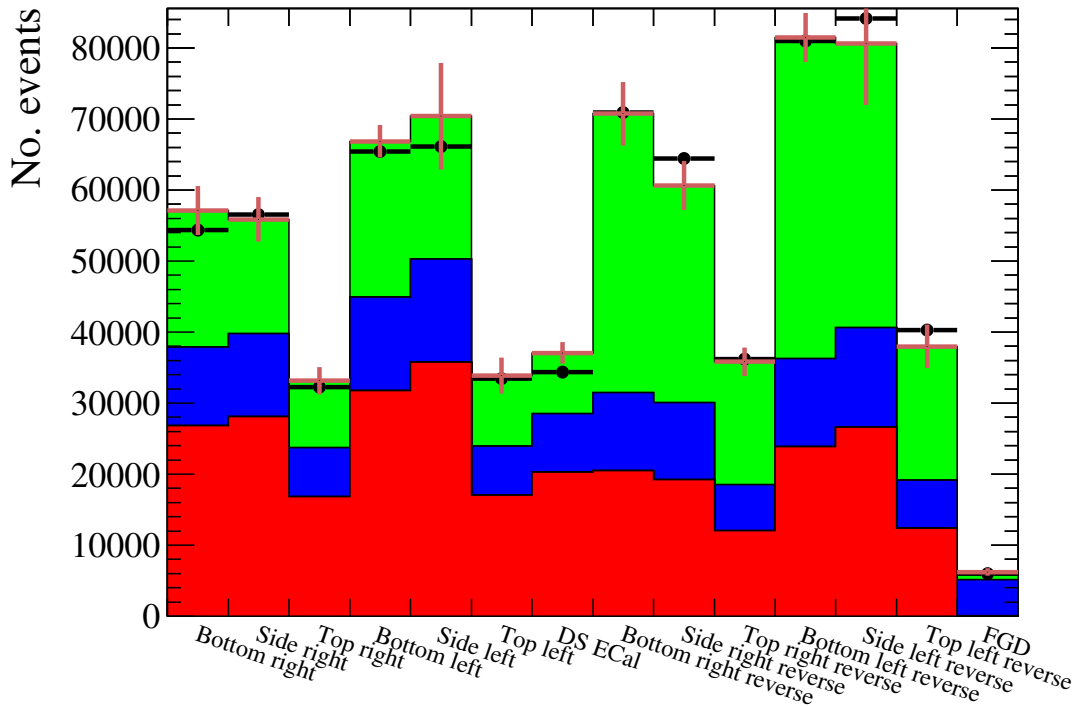


Figure 1.49: The post-fit number of NEUT Monte Carlo events compared with T2K data when including the adhoc uncertainty shown in table 1.16. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic and adhoc uncertainties on the number of NEUT MC events. The black points are the T2K data.

Normalisation parameter	Fitted value
R^{Pb}	$1.10^{+0.11}_{-0.10}$
R^{C}	$1.10^{+0.14}_{-0.12}$
R^{Other}	$0.90^{+0.10}_{-0.11}$

Table 1.17: The values returned from the fitter modified with the adhoc uncertainty when processing the NEUT MC templates with T2K data.

1.7 The ν_μ charged current inclusive cross-section on lead

Using the fitted value of R^{Pb} , the cross-section can be calculated using equation 1.16 and event information from the DS-ECal. As described in section 1.4, the errors on R^{Pb} returned from the fitter provides the shape systematic uncertainties. To calculate the full error, the extra error shown in equation 1.17 (12.5%) can be added in quadrature with the shape-only errors for R^{Pb} . Following this route, the value of R^{Pb} and its percentage uncertainty, which is the same percentage uncertainty value for the CC inclusive cross-section, was found to be

$$R^{\text{Pb}} = 1.10^{+16.4\%}_{-15.5\%}. \quad (1.21)$$

The dimensions and quantity of the lead absorbers in the DS-ECal are described in chapter ?? and section 1.4. Taking the density of lead to be 11.34 g/cm^3 , the total mass of lead in the DS-ECal is 2661 kg. The atomic mass of lead is 207.2 ± 0.1 atomic units [11], which infers there are 7.74×10^{27} lead nuclei present in the Monte Carlo simulation of the DS-ECal. The lack of antimony in the simulation has already been quantified in the final uncertainty above.

The selection presented in chapter ?? had a 53.0% efficiency and 72.4% purity for selecting charged-current neutrino interactions in the active volume of the DS-ECal. Considering interactions on lead alone, the efficiency and purity are 53.8% and 49.7% respectively with 32427 reconstructed events selected when scaled to the collected data POT.

Fig. 1.50 shows the simulated neutrino flux (per 10^{21} POT) for T2K run 3C. The POT-scaled integral of this distribution provides the total neutrino flux used in this

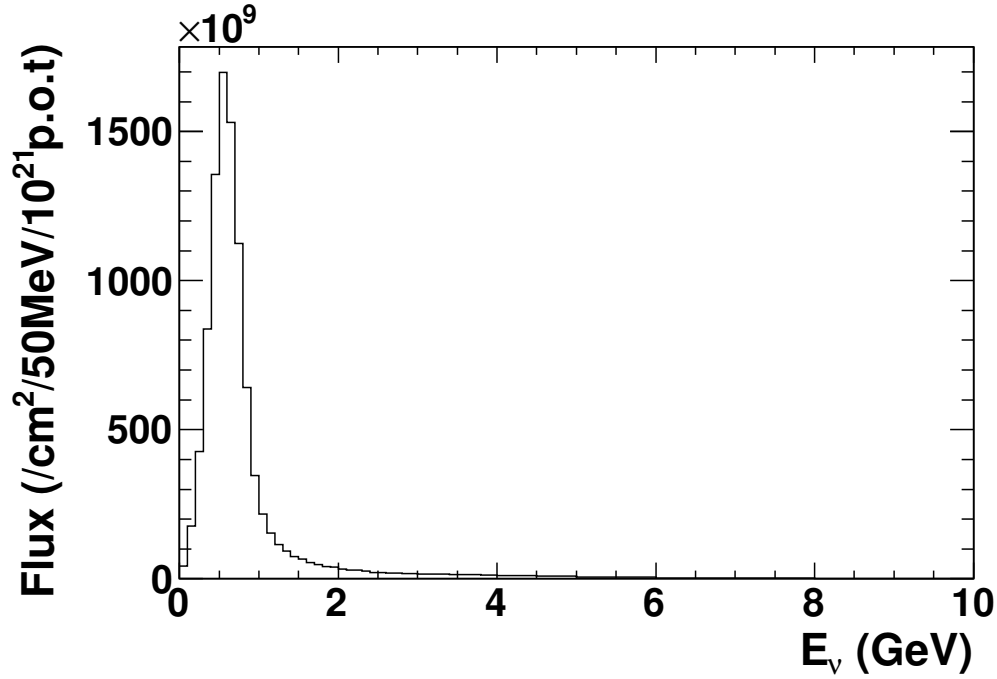


Figure 1.50: The simulated flux for T2K run 3C as a function of neutrino energy.

study. As the analysis ran over 1.31×10^{20} POT, the integral of Fig. 1.50 multiplied by the collected POT gives a total neutrino flux of $2.53 \times 10^{12} \text{ cm}^{-2}$.

Collating all of this information together and applying equation 1.16, the T2K flux-averaged ν_μ charged current cross-section on lead is measured as

$$\langle \sigma_{\text{Pb}}^{\text{CC}} \rangle_\phi = 8.13_{-1.26}^{+1.33} \times 10^{-39} \text{ cm}^2 \text{ nucleon}^{-1}. \quad (1.22)$$

Chapter 2

Discussion and conclusions

2.1 Discussion

The results presented in chapter 1 have potentially highlighted areas for improvement in the ND280 simulation. Fig. 1.40 showed discrepancies between data and MC for each ECal sample which ultimately led to the failure of the ECal rate fit when applied to T2K data. Upon further inspection of the fit, it was noted that there were two significant areas of discrepancy between data and MC (shown in Fig. 1.48).

As already discussed in chapter 1, one of the discrepancies appears to be correlated with the sand MC. The sand MC is a relatively new addition to the ND280 simulation and has been tuned to tracker region data. In most tracker analyses, the number of sand events which make it into the final selection generally number below 10 and so it is an area which has not required in depth study. This is obviously not the case for ECal based analyses which see a much higher number of events originating from the surrounding pit. The second discrepancy identified showed an excess of data events with a low number of hits in the reconstructed ECal clusters. A brief study of such events suggests that they originate from the surrounding magnet region. What is most troubling about these two issues is that they appear in the analysis at all. Both topologies appear to be an entering background which strongly suggests that further study of the ECal fiducial volume is needed.

As a separate problem, the reverse selection itself should be revisited. This selection is currently defined as any event which passes the fiducial volume cut but fails the final (most upstream) cut. While this allowed a large number of events to be included

in the analysis, it also opens up the possibility of data and MC mismodelling which either increases the systematic uncertainty or, like in this analysis, causes major issues in the final measurement.

However, this analysis has provided several important contributions to the T2K experiment. The first is that the ECal is capable of reconstructing topologies outside of the single track or single shower hypotheses. This is a potentially important milestone for the ECal as a neutrino detector as it paves the way for reconstructing neutrino events where the ECal is the active target. The second contribution is that it is possible to apply a selection to the enhanced reconstruction to actually select neutrino interactions in the ECal. While these developments have been used in a cross-section measurement, they were in no way developed solely with this analysis in mind. So, the cross-section measurement itself can essentially be thought of as a by-product of the development of the reconstruction and selection.

2.2 Conclusions

As highlighted in chapter ??, the field of neutrino physics has rapidly evolved since the neutrino's discovery. At time of writing, current and future generation neutrino oscillation experiments are aiming towards a measurements of the mass-hierarchy and the CP violating phase δ . To do this, a solid understanding of neutrino cross-section physics with atomic nuclei is vital.

The overall aim of this analysis was to study neutrino interactions in the ND280 ECals, with a measurement of the ν_μ CC inclusive cross-section on lead as the final goal. To realise this, a new set of reconstruction algorithms were developed to allow the reconstruction of the charged final-states of the neutrino interactions in the ECal. The output of this reconstruction was then used as the foundation of a ν_μ CC inclusive selection which achieved a good efficiency and purity for both the barrel ECals and DS-ECal. The final stage of the analysis was to use the selected events in a χ^2 fit to extract the data-constrained rate of neutrino interactions on lead. An important input to the fit itself was the evaluation of systematic uncertainties. This involved a set of ECal systematic uncertainties which were evaluated for the first time and should be extremely useful to future analysers.

Unfortunately, there were unforeseen discrepancies between data and MC which caused major problems in the final fit. These were overcome by introducing an adhoc uncertainty which encompassed the migration of problematic background events between the signal enhanced and background enhanced samples. With the inclusion of this extra uncertainty, the T2K flux-averaged ν_μ charged current cross-section on lead was measured to be

$$\langle \sigma_{\text{Pb}}^{\text{CC}} \rangle_\phi = 8.13_{-1.26}^{+1.33} \times 10^{-39} \text{ cm}^2 \text{ nucleon}^{-1}. \quad (2.1)$$

2.3 The future

The highest priority for future iterations of this analysis should be an in depth analysis of the problematic background events shown in Fig. 1.48. There are a few routes to take with such events but the general aim should be to either cut the events away in the selection and/or make improvements to the simulation itself.

There is an immense amount of scope for the reconstruction algorithms. While the Hough transform has shown itself to be a very powerful method of reconstruction, analysis of the generated parameter space is actually very rudimentary. As a reminder, the current method looks for the maxima in the parameter space which actually represent the 2D lines which pass through the most hits. This analysis method needs improvement. For example, the hits which are selected as constituents of the track are removed from the parameter space which means any subsequent track does not have that information available. A more advanced method of hit masking would allow hit sharing between tracks to take place.

Improvements can also be made to the selection. As it stands, the selection does not fully harness the power of the enhanced reconstruction. For example, particle identification could be applied to the individual tracks which would potentially allow removal of intrinsic background.

Most importantly, the reconstruction and selection provide much scope for expanding this analysis far beyond what is presented. At the time of writing, a dedicated group of analyser is being set up with a common aim of studying neutrino interactions in the ECal and its potential for physics contributions. For example, using the ν_μ interac-

tion rates in the barrel ECals to help constrain the flux in the ND280 tracker region and matching reconstructed ECal tracks to TPC tracks to make a ν_μ CC differential cross-section measurement on lead.

Despite the large issues raised here, the main contribution of this analysis is that the many tools developed here pave the way for a wide range of neutrino interaction studies in which the ECal is the target.

Bibliography

- [1] A. Rodriguez *et al.*, Phys. Rev. D **78**, 032003 (2008).
- [2] Y. Nakajima *et al.*, Phys. Rev. D **83**, 012005 (2011).
- [3] M. Ravonel, J. Lagoda, and A. Hillairet, T2K Technical note **212** (2015).
- [4] (T2K Collaboration), K. Abe *et al.*, Phys. Rev. D **90**, 052010 (2014).
- [5] V. B. Anikeev *et al.*, Z. Phys. **C70**, 39 (1996).
- [6] C. Andreopoulos, J. Dobson, and S. Dytman, T2K Technical note **7** (2009).
- [7] P. de Perio, K. Mahn, and P. Sinclair, T2K Technical note **113** (2012).
- [8] D. Allan *et al.*, Journal of Instrumentation **8**, P10019 (2013).
- [9] S. Oser, T2K Technical note **91** (2011).
- [10] C. Andreopoulos *et al.*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **614**, 87 (2010).
- [11] J. R. de Laeter *et al.*, Pure and Applied Chemistry **75**, 683 (2003).

List of Figures

1.1	The neutrino energy of events seen in the bottom-left ECal module broken down by the neutrino interaction mode.	5
1.2	The number of events in each input sample separated into the lead (black), carbon (red) and other (blue) templates.	7
1.3	The flux prediction covariance matrix. The matrix is binned in detector (SK and ND280), neutrino flavour and energy.	10
1.4	The sample covariance induced by the flux uncertainties.	12
1.5	The variation in the selection efficiency of lead interactions in the DS-ECal caused by the flux systematic.	12
1.6	The neutrino cross-section model covariance matrix.	14
1.7	The sample covariance induced by the cross-section model uncertainties.	14
1.8	The variation in the selection efficiency of lead interactions in the DS-ECal caused by the cross-section model systematic.	15
1.9	The ratio of the $\pm \sigma$ cross-section to the nominal cross-section for lead as a function of neutrino energy. Distributions produced by W. Ma. . .	16
1.10	The sample covariance induced by variation of P_F^{Pb}	16
1.11	The sample covariance induced by variation of of the FSI parameters outlined in table 1.1.	17
1.12	The layer-by-layer efficiency for the ND280 ECals, measured using cosmic data [8].	20
1.13	The charge distribution of scintillator bar hits created by MC cosmics rays.	20

1.14	The sample covariance induced by the ECal hit inefficiency.	21
1.15	The charge of ECal hits after applying the enhanced reconstruction to run 3C cosmic events. The distributions only show information for reconstructed objects in the top-left barrel ECal module. The blue and black histograms are Monte Carlo and data respectively.	22
1.16	The sample covariance induced by the ECal charge uncertainties. . . .	24
1.17	The estimation of the number of noise hits (solid blue line) and the number of true noise hits (solid red line) per event in Monte Carlo. The coloured dashed lines are the Gaussian fits to the respective distributions.	25
1.18	The estimation of the number of noise hits in Monte Carlo (solid blue line) and data (solid black line) per event. The coloured dashed lines are the Gaussian fits to the respective distributions.	25
1.19	The sample covariance induced by the ECal noise uncertainties.	26
1.20	The covariance matrix for the mass of the contributing elements in the DS-ECal, with all contributing elements considered.	30
1.21	The total mass of the DS-ECal active volume with (black) and without (blue) nitrogen and hydrogen as contributing elements. The smooth lines are the Gaussian fits to the distributions. Using the Gaussian mean and width as the mass central value and uncertainty respectively, the DS-ECal active mass with (without) hydrogen and nitrogen is 4060 ± 20 kg (3960 ± 20 kg).	30
1.22	The fractional covariance matrix for the mass of each contributing element to an ECal module (hydrogen and nitrogen omitted).	31
1.23	The sample covariance induced by the ECal mass uncertainties.	32
1.24	The sample covariance induced by the uncertainty in the ECal OOFV events.	34
1.25	The sample covariance induced by the bug fix to the reconstruction. . . .	35
1.26	The number of selected FGD events for each systematic throw using PSYCHE. The smooth line is a Gaussian fit to the number of events. . . .	36
1.27	The sample covariance induced by all of the detector related uncertainties. . . .	36

1.28	The sample covariance induced by all of the systematic uncertainties. .	37
1.29	The number of selected events in the GENIE-based Monte Carlo sample, separated out into the prong topologies. Each event is categorised by the associated truth information from the simulation.	38
1.30	Distributions of R^{Pb} related information returned from the fitter for the 10,000 fake data throws. The red dashed lines are Gaussian fits to the distributions.	42
1.31	Performance of the fit for the 10,000 fake data sets when using normalisation set 1.	42
1.32	Performance of the fit for the 10,000 fake data sets when using normalisation set 2.	43
1.33	Performance of the fit for the 10,000 fake data sets when using normalisation set 3.	43
1.34	Performance of the fit for the 10,000 fake data sets when using normalisation set 4.	44
1.35	Performance of the fit for the 10,000 fake data sets when using normalisation set 5.	44
1.36	Performance of the fit with the reverse ECal samples removed for the 10,000 fake data sets when using normalisation set 1.	45
1.37	The prefit number of NEUT Monte Carlo events compared with GENIE fake data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the GENIE fake data.	46
1.38	The post-fit number of NEUT Monte Carlo events compared with GENIE fake data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the GENIE fake data.	48

1.39	The post-fit number of NEUT Monte Carlo events compared with GENIE fake data which has a re-weighted FGD sample. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the GENIE fake data.	49
1.40	The pre-fit number of NEUT Monte Carlo events compared with T2K data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the T2K data.	52
1.41	The post-fit number of NEUT Monte Carlo events compared with T2K data. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic uncertainty on the number of NEUT MC events. The black points are the T2K data.	53
1.42	The χ^2/NDOF as a function of the normalisation parameters.	54
1.43	The shape-only correlation matrix for the samples used in the fit. . . .	55
1.44	An area-normalised comparison of a sub-set of the MC templates with data for the barrel ECal samples. The coloured histograms and data points are the MC Templates and data respectively. The red, blue and green histograms are the lead, carbon and other templates respectively.	55
1.45	An area-normalised comparison of a sub-set of the MC templates with data for the barrel-reverse ECal samples. The coloured histograms and data points are the MC Templates and data respectively. The red, blue and green histograms are the lead, carbon and other templates respectively.	56
1.46	The number of events in the other template as a function of true neutrino energy for the barrel ECal samples. The distributions are broken down by true neutrino interaction location.	57

1.47	The number of events in the other template as a function of true neutrino energy for the barrel reverse ECal samples. The distributions are broken down by true neutrino interaction location.	58
1.48	The number of events which pass the fiducial volume cut as a function of the number of hits in the reconstructed cluster. The red and blue histograms show the number of beam MC events and sand MC events respectively. The black points show the run 3C data.	59
1.49	The post-fit number of NEUT Monte Carlo events compared with T2K data when including the adhoc uncertainty shown in table 1.16. The red, blue and green histograms are the lead, carbon and other NEUT MC templates. The brown error bars represent the shape-only systematic and adhoc uncertainties on the number of NEUT MC events. The black points are the T2K data.	61
1.50	The simulated flux for T2K run 3C as a function of neutrino energy. . .	63

List of Tables

1.1	FSI parameters, provided by the NIWG, representing the 1σ contour in the FSI parameter space [7].	18
1.2	Summary of the lead absorbers for the ND280 tracker ECals [8].	23
1.3	Summary of the study to estimate the number of noise hits per event in the ECals.	25
1.4	Summary of the lead absorbers for the ND280 tracker ECals [8].	27
1.5	Summary of the scintillator bars for the ND280 tracker ECals [8].	28
1.6	Summary of the WLSfibres for the ND280 tracker ECals [8].	29
1.7	Summary of the ECal OOFV systematic study when using the FGD cosmic control sample.	33
1.8	The selection efficiencies from the GENIE-based Monte Carlo sample for each prong topology and ECal module.	38
1.9	The topology combined efficiency for the GENIE-based Monte Carlo sample. The efficiency values include simulated events which were not reconstructed.	38
1.10	The normalisation sets used in the machinery validation of the ECal rate fit.	40
1.11	The GENIE to NEUT cross-section ratios, calculated from the truth information provided by the ND280 simulation.	47
1.12	The values returned from the fitter when passing the NEUT MC templates and GENIE fake data shown in Fig. 1.37.	47

1.13	The values returned from the fitter when passing the NEUT MC templates and GENIE fake data which contained a re-weighted FGD sample.	49
1.14	The number of selected events in each sample.	50
1.15	The values returned from the fitter when processing the NEUT MC templates with T2K data.	51
1.16	Comparison of the number of events between the barrel and barrel-reverse samples. The second and thrid column show the ratio of the number of events in the barrel samples to the number of events in the barrel-reverse samples for Monte Carlo and data respectively. The fourth column shows the percentage difference between the Monte Carlo and data ratios.	61
1.17	The values returned from the fitter modified with the adhoc uncertainty when processing the NEUT MC templates with T2K data.	62