Measurement of the Electron Anti-neutrino Cross-section on Carbon at the T2K Near Detector

Fady Shaker

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Physics and Astronomy

University of Manitoba

Winnipeg

To my mother, and in memory of my father.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Fady Shaker

Acknowledgements

Firstly I would like to thank my supervisor, Dr. Blair Jamieson, for encouraging my research, providing the support and advice, and for giving me many opportunities to attend and contribute to different conferences and summer schools. I would also like to thank my committee, Dr. Peter Blunden, Dr. Michael Gericke and Dr. Sherif Sherif for their comments, suggestions and questions which helped me to understand my own work at a deeper level. I would especially like to thank Dr. Kendall Mahen and Dr. Dean Karlen who led some of our meetings and provided valuable feedback and comments on my work.

I thank my fellows at T2K collaboration who made the stay in Japan culturally enriching and fun. I thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support and the University of Manitoba for its generous scholarship and graduate fellowship.

Finally, I would like to thank my friends and family for their support during this long journey.

Abstract

The intrinsic electron anti-neutrino ($\overline{\nu}_e$) component in the muon anti-neutrino ($\overline{\nu}_\mu$) Tokai to Kamiokande (T2K) beam represents the largest irreducible background for the ($\overline{\nu}_\mu \rightarrow \overline{\nu}_e$) appearance measurement at Super-Kamiokande.

In this thesis, the $\overline{\nu}_e$ charged-current inclusive cross-section on carbon is measured using the T2K near detector (ND280). The measurement is performed in two steps: the selection of the $\overline{\nu}_e$ events, and the correction of the detector inefficiency and mis-reconstruction (unfolding).

In the selection step, a new particle identification based on the Boosted Decision Tree (BDT) was developed. The Particle IDentifications (PIDs) from the Fine-Grained Detectors (FGDs), Time Projection Chambers (TPCs) and Electromagnetic Calorimeters (ECALs) are used as input for the BDT to select the positron track; a characteristic feature of the charged-current interaction of the $\overline{\nu}_e$.

The BDT parameters were tuned to maximize the selection efficiency, purity and Kolmogorov-Smirnov (KS) test. The uncertainty on the BDT output test statistic was calculated by propagating the input variables systematics using Monte Carlo (MC) simulation.

Moreover, we have developed an improved unfolding algorithm based on the D'Agostini 2010 method [1] as well as a framework to test the unfolding performance.

In order to to constrain the gamma background, we have also constructed a control sample to select the gamma events and to use it in the simultaneous unfolding framework.

Finally, we measured the $\overline{\nu}_e$ charged-current inclusive cross-section on carbon with and without using the control sample. The measured values are 0.279 $\pm$ 0.054 (19.4% stat) $\pm$ 0.027 (9.7% syst) $\times 10^{-38}$ cm^2/nucleon and 0.332 $\pm$ 0.052 (15.8% stat) $\pm$ 0.025 (7.5% syst) $\times 10^{-38}$ cm^2/nucleon, respectively, which are in agreement with the MC simulation (0.290×10^{-38} cm^2/nucleon) as well as the Gargamelle [2] measurement (0.25 $\pm$ 0.07 $\times 10^{-38}$ cm^2/nucleon).

Table of contents

List of tables

List of figures

List of abbreviations

2p-2h Two particles-two holes

AdaBoost Adaptive Boosting

ASIC Application Specific Integrated Circuit

Bagging Bootstrap Aggregation

BDT Boosted Decision Tree

CCQE Charged-Current Quasi-Elastic

CCRes Charged-Current Resonance

CL Confidence Level

CM Central Mass

CP Charge Parity

CVC Conserved Vector Current

DAQ Data Acquisition System

DCC Data Concentration Card

DIS Deep Inelastic Scattering

dLips differential Lorentz invariant phase space

ECAL Electromagnetic Calorimeter

FEC Front-end Electronics

FEM Front-end Mezzanine

FF Fragmentation Function

FFT Fast Fourier Transform

FGD Fine Grained Detectors

FoM Figure of Merit

FV Fiducial Volume

GPS Global Position System

GWS Glashow-Weinberg-Salam

HEP High Energy Physics

IH Inverted Hierarchy

INGRID Interactive Neutrino Grid

J-PARC Japan Proton Accelerator Complex

KDE Kernel Density Estimation

KS Test Kolmogorov-Smirnov Test

LFG Local Fermi Gas

LINAC Linear Accelerator

MC Monte Carlo

MEC Meson Exchange Current

Micromegas (MM) Micro Mesh Gaseous Structure

MIP Minimum Ionizing Particle

MISE Mean Integrated Square Error

MPPC Multi-Pixel Photon Counter

MR Main Ring

ND280 Near Detector at 280 m from beam production

NH Normal Hierarchy

NIWG Neutrino Interaction Working Group

OOFGD Out Of Fine Grained Detector

OOFV Out Of fiducial Volume

PØD Pi-Zero Detector

PCA Principal Component Analysis

PCAC Partially Conserved Axial Current

PDF Probability Density Function

PID Particle Identification

PSYCHE Propagation of SYstematics and CHaracterization of Events

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

QH Quark Hadron

RCS Rapid-cycling Synchrotron

RES Resonance

RFG Relativistic Fermi Gas

RPA Random Phase Approximation

RS Rein Sehgal

SCA Switched Capacitor Array

SIS Shallow Inelastic Scattering

SK Super-Kamiokande

SMRD Side Muon Range Detector

SRC Short Range Correlation

SVD Singular Value Decomposition

T2K Tokai to Kamiokande experiment

TMVA Toolkit for MultiVariate Analysis

ToF Time of Flight

TPC Time Projection Chamber

WLS Wavelength Shifting

Part I

Introduction

Chapter 1

The Standard Model

The Standard Model (SM) of particle physics tries to answer the question: what is matter made of, at the most fundamental level? It incorporates quantum mechanics and special relativity to describe the elementary particles, along with three of the four known fundamental forces in the universe: the electromagnetic, weak, and strong forces [3].

The SM classifies the elementary particles into two groups called leptons and quarks. Each group is made up of three generations of paired particles. All stable matter in the universe is made from particles of the first (lightest) generation. Heavier particles decay to lighter, more stable, particles.

In the lepton sector, the “electron” and the “electron neutrino” form the first generation, followed by the “muon” and the “muon neutrino”, then finally the “tau” and the “tau neutrino”. The electron, the muon and the tau all have an electric charge and a considerable mass, whereas the neutrinos are electrically neutral and have very little mass.

In the quark sector, the “up” and the“down” quarks form the first generation, followed by the “charm” and “strange” annd the “top” and “bottom” quarks. In addition to the electrical charge, quarks also possess a different kind of charge called “colour” with three different colours: red, green and blue. Quarks are always found in bound state colourless hadrons. It

is believed, although not yet proven, that all observed free particles are colourless, which is the Colour Confinement Hypothesis [3].

For each particle, there is an associated antiparticle with the same mass and opposite charge. Quantum mechanically, the particle and antiparticle states are related through charge conjugation (C)[1]. The charge conjugation operator flips the particle's charge as well as all of its internal quantum numbers like baryon number, lepton number, strangeness, charm, beauty and truth, but keeps the particle's mass, energy, momentum and spin [3, Chapter 4].

The interaction between different particles is facilitated through the exchange of a force carrier (gauge boson). The photon (γ) carries the electromagnetic force, the gluon (g) carries the strong force, while the weak force is carried by the "$W^{\pm}$" and "Z" bosons. In the SM, the Higgs boson is responsible for giving masses to all particles.

The elementary particles in the SM, along with the gauge and Higgs bosons are shown in Fig. 1.1.

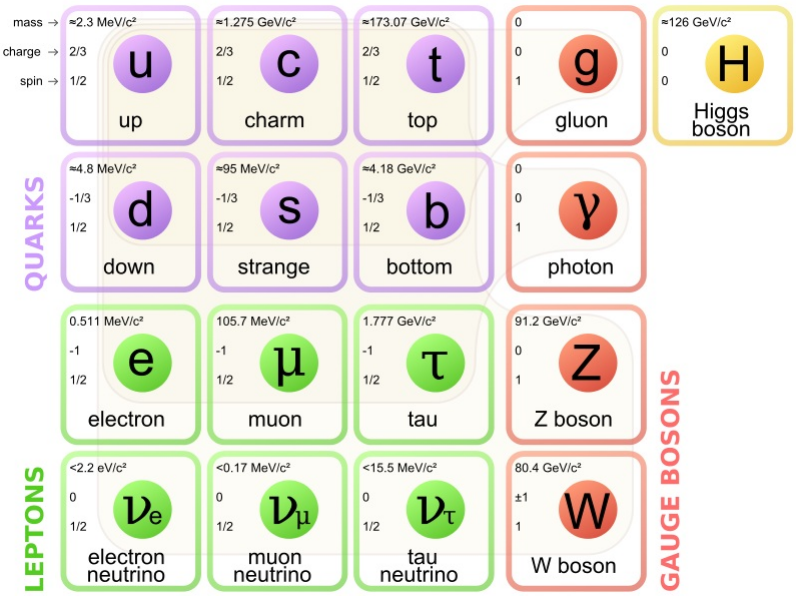


Fig. 1.1 The SM of elementary particles, with the three generations of matter, gauge bosons in the fourth column, and the Higgs boson in the fifth. The figure is taken from [4].

[1]Some particles are the eigenstates of the charge conjugation operator. Therefore, they are their own antiparticles, e.g. photons and the neutral pion [3, Chapter 4]

In the SM, the Lagrangian approach for quantum field theory is commonly used. This approach is ideal to discuss symmetries and conservation laws. Noether's theory (cf. [5, Chapter 6], [3, Chapter 4]) shows that each continuous symmetry transformation under which the Lagrangian is invariant leads to a conservation law. For example, invariance in time, space translation, and rotation lead to conservation of energy, conservation of momentum, and conservation of angular momentum respectively. Invariance under internal operations, e.g. the rotation of the complex phase of wave functions, leads to a conservation of charge.

The SM is not a complete theory of physics though, as it does not include a theory of gravity, baryon asymmetry, dark energy, dark matter or neutrino oscillations. Nevertheless, the SM is successful in describing the fundamental elementary interactions of the quarks and leptons based on gauge symmetry principles, or more precisely local gauge invariance theory.

1.1 Dirac Equation

The relativistic formulation of the quantum mechanics of the electron was introduced by Dirac in 1928 [3]. The spin, an intrinsic form of angular momentum carried by elementary particles with no classical analogy, appears as a natural consequence of this relativistic treatment. The Dirac equation describes the equation of motion for any spin $\frac{1}{2}$ particle

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi = 0, \tag{1.1}$$

where m is the particle mass, γ^{μ} are the 4x4 Dirac gamma matrices and ψ is a four component spinor. The derivation of the above equation can be found in any quantum field theory book (cf. [3, Chapter 7], [6, Chapter 4]).

In the Dirac representation the γ^μ are connected to the Pauli matrices as follows

$$\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}; \gamma^a = \begin{pmatrix} 0 & \sigma_a \\ -\sigma_a & 0 \end{pmatrix}, a = 1,2,3. \tag{1.2}$$

Where I is a 2x2 identity matrix, $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ and $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.
The γ matrices haves interesting properties

- All γ matrices are traceless and Hermitian[2].
- $(\gamma)^2 = I$.
- $\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}$, where $g^{\mu\nu}$ is the metric tensor.
- $\{\gamma^0, \gamma^\mu\} = 0$, $\{\gamma^0, \gamma^5\} = 0$, with $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$.

The Dirac Lagrangian for one free spin $\frac{1}{2}$ particle can be expressed as

$$\mathscr{L}_{free} = \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi, \tag{1.3}$$

with

$$\overline{\psi} = \psi^\dagger\gamma^0. \tag{1.4}$$

Applying Euler-Lagrange equation[3] to Eq. (1.3), we obtain the corresponding equation of motion.

[2]A Hermitian matrix, a.k.a self-adjoint, H satisfies $H = H^\dagger$. Where $H^\dagger$ is the transpose complex conjugate of H.

[3]Euler-Lagrange equation is, in general, a second order differential equation, when applied to the Lagrangian we get the equation of motion. In relativistic form
$\partial_\mu(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\phi)}) - \frac{\partial\mathscr{L}}{\partial\phi} = 0$.

1.2 Local Gauge Invariance

It is known that classical electromagnetics is gauge invariant. Modifying the vector potential $\vec{A}$ to $\vec{A}+\nabla f$, where f is any continuously differentiable scalar function does not change the underlying physics. For example, the magnetic field $\vec{B}$ expressed as the curl of some vector potential does not change with the vector potential transformation

$$\vec{B}' = \nabla \times \vec{A}' = \nabla \times (\vec{A} + \nabla f) = \nabla \times \vec{A} + \nabla \times \nabla f = \vec{B}, \tag{1.5}$$

as the curl of a gradient equals to zero. This non-uniqueness leads to a degree of freedom in the formulation of electrodynamics, or gauge freedom.

In a relativistic quantum theory, for the simplest case of a non-interacting scalar field, ϕ, the Lagrangian normally contains kinetic terms like $\partial_\mu \phi^\dagger \partial^\mu \phi$ and a mass term that goes like $m^2\phi^\dagger\phi$. The question that arose was: what happens to the Lagrangian if we transform the field ϕ by a unitary transformation ($\phi \to U\phi$)? The mass term will not change because $m^2(U\phi)^\dagger U\phi = m^2\phi^\dagger U^\dagger U\phi = m^2\phi^\dagger\phi$.

A problem may arise in the derivative term. Expressing the unitary transformation as $U = e^{iq\theta(x)}$, where q is a real number representing the particle charge, and $\theta(x)$ is an arbitrary function of spacetime. If $\theta(x)$ was constant, this would correspond to a global phase change in the wave function which is not detectable. If we insist on local gauge invariance, we will have to introduce a gauge field to the Lagrangian, with certain transformation properties, such that it cancels out the extra derivative terms that arose from the field transformation. We can achieve that, for the above simple example, by introducing a gauge field A_μ that transforms like $A_\mu \to A_\mu + \partial_\mu\theta(x)$ and replace the derivative in the Lagrangian by a covariant derivative

$$\mathscr{D}_\mu \equiv \partial_\mu + iqA_\mu, \tag{1.6}$$

$$\mathscr{D}^\mu \equiv \partial^\mu - iqA^\mu. \tag{1.7}$$

This will ensure that the kinetic term is invariant under such transformation

$$\mathscr{D}'_{\mu} e^{-iq\theta(x)}(\phi^*)\mathscr{D}'^{\mu} e^{+iq\theta(x)}(\phi) = \mathscr{D}_{\mu}(\phi^*)\mathscr{D}^{\mu}(\phi). \tag{1.8}$$

The new Lagrangian in such case would look like

$$\mathscr{L} = \mathscr{D}_{\mu}\phi^{\dagger}\mathscr{D}^{\mu}\phi - V(\phi^{\dagger}\phi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{1.9}$$

where $F_{\mu\nu}$ represents the kinetic term of the gauge field

$$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}. \tag{1.10}$$

1.3 Electroweak Theory

Electroweak theory is a unification of the electromagnetic, U(1) gauge theory, with the weak interaction, SU(2) gauge theory. It is based on the Glashow-Weinberg-Salam (GWS) model [7], who were awarded the Nobel prize for it in 1979.

For the local U(1) gauge invariance, the Dirac Lagrangian Eq. (1.3) is modified to

$$\mathscr{L}^{U(1)}_{free} = \overline{\psi}(i\gamma^{\mu}\mathscr{D}_{\mu} - m)\psi = \overline{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi - q\overline{\psi}\gamma^{\mu}\psi A_{\mu}. \tag{1.11}$$

Here we can also define the electromagnetic current $J^{\mu} \equiv q\overline{\psi}\gamma^{\mu}\psi$. This current satisfies the continuity equation ($\partial_{\mu}J^{\mu} = 0$), hence, the current is a conserved quantity.

The Dirac Lagrangian satisfying a local gauge invariance SU(2) is expressed as

$$\mathscr{L}^{SU(2)}_{free} = \overline{\psi}(i\gamma^{\mu}\mathscr{D}_{\mu} - m)\psi = \overline{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi - \frac{g}{2}\overline{\psi}\gamma^{\mu}\vec{\tau}\psi\vec{b_{\mu}}. \tag{1.12}$$

Where we introduce a new covariant derivative

$$\mathscr{D}_\mu = \partial_\mu + igB_\mu, \tag{1.13}$$

where $B_\mu = \frac{1}{2}\vec{\tau}\cdot\vec{b_\mu}$, g represents the weak coupling strength, $\vec{\tau}$ is the isospin generator vector (Pauli matrices) and $\vec{b_\mu}$ is the gauge field vector.

These three massless vector gauge fields (b_1, b_2, b_3) will acquire mass through spontaneous symmetry breaking when interacting with the Higgs field, to have the known massive vector bosons (W^+, W^-, Z^0) acting as carriers of the weak force (cf. [7, Chapter 19]).

Again, we can define a weak current vector to be $\vec{j^\mu} \equiv \frac{g}{2}\overline{\psi}\gamma^\mu\vec{\tau}\psi$.

The SU(2) weak isospin symmetry, basically says that neglecting the electric charge, neutrinos and their corresponding charged leptons fields would describe the same field. Therefore, we can arrange them into a doublet, and express this symmetry as rotation in the weak-isospin space

$$\psi = \begin{pmatrix} \nu_l \\ l \end{pmatrix}, \tag{1.14}$$

where l here represents the flavour of the charged lepton: e, μ or τ.

It has been discovered that weak decay violates parity in charged weak interaction and that only the left-handed component of the fields contribute to the weak interaction a.k.a " maximal violation of parity"(cf. [3, Chapter 4 and 9]). Consequently, we arbitrarily split the weak interaction, by decomposing the field into its chiral representation, by defining a left-handed field ψ_L and a right-handed field ψ_R as

$$\psi_L = \frac{1}{2}(I - \gamma_5)\psi, \tag{1.15}$$

and

$$\psi_R = \frac{1}{2}(I + \gamma_5)\psi. \tag{1.16}$$

In this chiral representation $\gamma_5 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix}$ and $\gamma_0 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$.

Now if we combined the neutrino field and the lepton one into a doublet as in Eq. (1.14), the result is

$$\psi_L = \begin{pmatrix} \nu_{l_L} \\ l_L \end{pmatrix}, \tag{1.17}$$

and

$$\psi_R = \begin{pmatrix} \nu_{l_R} \\ l_R \end{pmatrix}. \tag{1.18}$$

1.4 Charged-current Interaction

The b_1 and b_2 fields of Eq. (1.12) are responsible for the charged-current in weak interactions. Hence we can redefine the known W fields as follows

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(b_\mu^1 \mp b_\mu^2). \tag{1.19}$$

Furthermore, we can redefine the currents in terms of the raising and lowering operators $\tau^+ \equiv \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\tau^- \equiv \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. Consequently, the charged-current interaction Lagrangian can be expressed as

$$\begin{aligned} \mathscr{L}_{CC} &= -b_\mu^1 J_\mu^1 - b_\mu^2 J_\mu^2 \\ &= -W_\mu^+ J_\mu^+ - W_\mu^- J_\mu^-, \end{aligned} \tag{1.20}$$

with the leptonic charge raising current defined as

$$\begin{aligned} J^{+\mu} &= \frac{g}{\sqrt{2}}\overline{\nu_L}\gamma^\mu l_L \\ &= \frac{g}{\sqrt{2}}\overline{\nu}\frac{1}{2}(1+\gamma_5)\gamma^\mu\frac{1}{2}(1-\gamma_5)l \\ &= \frac{g}{2\sqrt{2}}\overline{\nu}\gamma^\mu(1-\gamma_5)l. \end{aligned} \tag{1.21}$$

The same holds for the leptonic charge lowering current $J^{-\mu}$

$$J^{-\mu} = \frac{g}{2\sqrt{2}}\bar{l}\gamma^\mu(1-\gamma_5)\nu. \tag{1.22}$$

1.5 Neutral Current Interaction

For the neutral current, experiments have shown that Z boson coupling is not totally left handed (cf. [3, Chapter 9]) but

$$\begin{aligned} J^{\mu f}_{NC} &= \frac{g}{2}\overline{\psi^f}\gamma^\mu(C^f_V - C^f_A\gamma_5)\psi^f, \\ &= \frac{g}{2}\overline{\nu}\gamma^\mu(C^\nu_V - C^\nu_A\gamma_5)\nu + \frac{g}{2}\bar{l}\gamma^\mu(C^l_V - C^l_A\gamma_5)l. \end{aligned} \tag{1.23}$$

Where the vector and axial vector coupling strength constants, C_V and C_A, respectively, are no longer equal to 1, but depend on the fermion type f. In the GWS model, C_V and C_A are fully determined by a single parameter, θ_W, called "the weak mixing angle" a.k.a "Weinberg angle" [7] as shown in Table 1.1.

Flavour (f)	C_V	C_A
ν_e,ν_μ,ν_τ	$\frac{1}{2}$	$\frac{1}{2}$
e,μ,τ	$-\frac{1}{2}+2\sin^2\theta_W$	$-\frac{1}{2}$
u,c,t	$\frac{1}{2}-\frac{4}{3}\sin^2\theta_W$	$\frac{1}{2}$
d,s,b	$-\frac{1}{2}+\frac{2}{3}\sin^2\theta_W$	$-\frac{1}{2}$

Table 1.1 Vector and axial coupling strength constants for the neutral current interaction as a function of the flavour (f) (cf. [3, Chapter 9]).

Experimentally, it has been found that $\theta_W \approx 28.7°$ [8].

1.6 Cross-section

We are interested in calculating the neutrino scattering cross-section, which is a measure of the probability that the incident neutrino scatters off the target. A cross-section (σ) is given by

$$\sigma = \frac{R_T}{N_{target}\phi}, \tag{1.24}$$

where R_T is the interaction rate, N_{target} is the number of targets and ϕ is the incident flux, i.e. number of incident particles per unit area per unit time.

Cross-section has the dimension of area and is normally measured in barn.[4] Its numerical value is chosen such that the specific reaction occurs if and only if the incident particle hits a circular area of this size, perpendicular to its path, and centred on the target.

The inclusive cross-section is simply the sum of all cross-sections coming from all possible interaction channels

$$\sigma_{inclusive} = \sum_i \sigma_i. \tag{1.25}$$

[4] A barn (symbol: b) is a unit of area equals to 10^{-28}m^2 (100fm^2).

A differential cross-section, $\frac{d\sigma}{d\Omega}$, is a measure of the probability that the particles are scattered by a solid angle $d\Omega$.

The cross-section can be calculated theoretically using Fermi's golden rule [3], which relates the differential cross-section, $d\sigma$, to the quantum dynamics, as well as the available phase space to the interaction, as follows (cf. [3, Chapter 6])

$$d\sigma \sim |\mathscr{M}|^2 \times \text{dLips}. \tag{1.26}$$

Here dLips is the differential Lorentz invariant phase space and $\mathscr{M}$ represents the Lorentz invariant amplitude for the transition of the field from the initial state to the final state ($|\psi_i> \rightarrow |\psi_f>$). Generally, $|\mathscr{M}|^2$ is the product of a coupling term and the contraction of two tensors, formed from the interaction currents. For example, the scattering cross-section (σ) of two particles (1&2) producing particles (3,4,..,n) in final state ($1+2 \rightarrow 3+4+..+n$), is given by

$$\begin{aligned}\sigma = \frac{S}{\sqrt{(p_1 p_2)^2 - (m_1 m_2)^2}} \int |\mathscr{M}|^2 (2\pi)^4 \delta^4(p_1 + p_2 - P_3 - ... - P_n) \\ \times \prod_{j=3}^{n} 2\pi\delta(p_j^2 - m_j^2)\Theta(p_j^0)\frac{d^4 p_j}{(2\pi)^4},\end{aligned} \tag{1.27}$$

where S is a statistical factor that corrects for double counting when there are identical particles in the final state. For each group of identical n particles, S gets a factor of $(\frac{1}{n!})$. The terms p and m are the particle four-momentum and mass, respectively, $\delta^4(p_1 + p_2 - P_3 - ... - P_n)$ ensures energy/momentum conservation of the reaction, $\delta(p_j^2 - m_j^2)$ forces the outgoing particles to lie on their mass shell, and the step function $\Theta(p_j^0)$ ensures positive energies.

A detailed description of the cross-section for different interaction channels is presented in the subsequent sections.

$\mathscr{M}$ is normally calculated from Feynman diagrams, which are pictorial representations of the interaction process. The full description of Feynman rules can be found in (cf. [3, Chapter 6 & Appendix D] or [6, Appendix L] for QED and [7, Appendix Q] for QCD and EW). These references also show that the Feynman propagator and vertex factors are related to the free and interaction Lagrangian respectively.

Chapter 2

Neutrino–Nucleus Cross-section

Neutrino interactions in the SM are defined for point-like particles. Practically, what we can measure is folded twice: once to account for the fact that the neutrino interacts with a distribution of quarks inside the nucleon by introducing form factors, another is due to nucleon-nucleon interaction inside the nucleus. A theoretical, as well as phenomenological approach, is needed in such nuclear models.

Most of the neutrino–nucleus cross-section models are based on electron–nucleus scattering in one way or another. Unlike the monochromatic electron beam, a neutrino beam is broad in energy and, consequently, has a higher systematic uncertainty in the reconstructed true neutrino energy.

There are two kinds of interaction models: inclusive and exclusive. In the inclusive model, only a lepton of the same flavour as the neutrino is detected at the final state. For the exclusive model, we categorize the interaction channels by the outgoing hadrons; for example, we may have Charged-Current Quasi-Elastic (CCQE), Charged-Current Resonance (CCRes), or Deep Inelastic Scattering (DIS).

Figure 2.1 shows the different interaction channels of the neutrino–nucleus cross-section, modelled in the neutrino interaction simulator NEUT [9]. Each of these interaction channels will be explained in detail in the next sections.

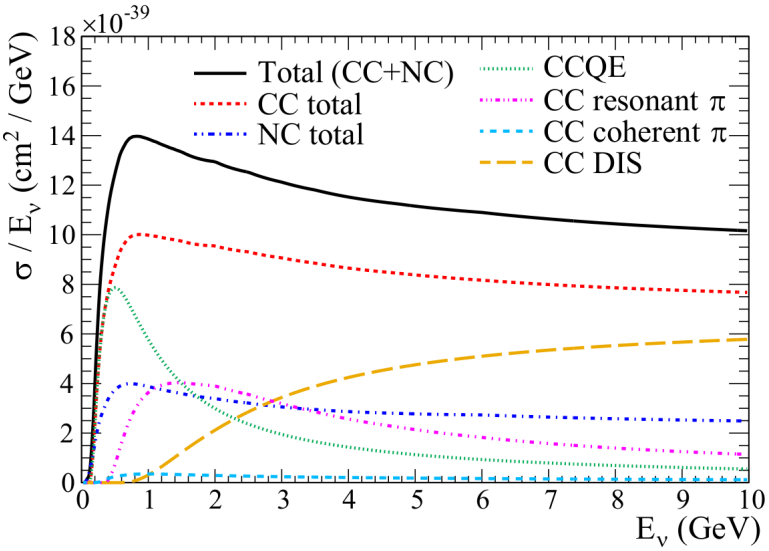


Fig. 2.1 The neutrino–nucleus cross-section per nucleon divided by the neutrino energy, modelled in NEUT. Figure courtesy of [10].

2.1 Inclusive Cross-section

In the inclusive cross-section the neutrino interacts with the nucleon, generating a lepton of the same flavour and any number of other hadronic products in the final state. The inclusive processes cover a broad range of neutrino energies.

In the electron anti-neutrino case, the inclusive cross-section is described by the following reaction

$$\overline{\nu}_e + N \rightarrow e^+ + X, \tag{2.1}$$

where N is the nucleon and X is any hadronic state. The inclusive cross-section is described as the coherent sum of cross-sections from all possible interaction channels. First principle (ab-initio) techniques only work for inclusive cross-section models. In order to apply the ab-initio techniques to the semi-exclusive cases, one must make some assumptions, such as modelling the energy/momentum distribution of the nucleus using a spectral function, or as a relativistic Fermi gas.

2.2 Charged-Current Quasi-Elastic Scattering

Quasi-Elastic scattering represents the main interaction up to an energy regime of 1 GeV. It defines the interaction channel in which the neutrino is scattered off one bound nucleon, knocking it out but leaving the nucleon unexcited, i.e. there is no pion in the final state.

For the electron anti-neutrino ($\overline{\nu}_e$) case, the reaction is given by

$$\overline{\nu}_e + p \rightarrow e^+ + n. \tag{2.2}$$

Figure 2.2 shows the Feynman diagram representing Eq. (2.2).

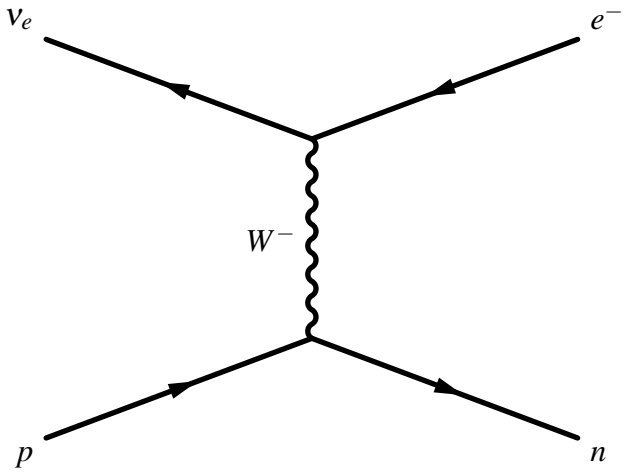


Fig. 2.2 CCQE interaction Feynman diagram. In this diagram, time flows from left to right.

Practically, it is not possible to measure this channel precisely due to nuclear effects. What is measured in the lab is a positron appearance with no pions in the final state (CC0π). For instance, a final state interaction (FSI) can transform a CC1π interaction into CC0π if the pion is absorbed inside the nucleus.

The derivation of the cross-section of this interaction channel follows the Llewellyn Smith procedure [11], which is implemented in NEUT [9], and will be outlined here.

The interaction Lagrangian can be expressed as the contraction of the leptonic and hadronic currents as follows

$$\mathscr{L}_{int} = \frac{G_F}{\sqrt{2}} J_{\alpha}^{lep} J_{had}^{\alpha}, \tag{2.3}$$

where G_F is the Fermi constant, J_{α}^{lep} is the leptonic current and J_{had}^{α} is the hadronic current. The matrix elements of J_{α}^{lep} connect the neutrino initial state to the corresponding final leptonic state. This current can be calculated from the standard model as follows

$$\left\langle l(k')|J_{\alpha}^{lep}|\nu(k)\right\rangle = \overline{u}_l(k')\gamma_{\alpha}\left(1-\gamma_5\right)u_{\nu}(k). \tag{2.4}$$

The matrix elements for J_{had}^{α} are more complicated as the exact initial and final states are unknown, but it may be writen in the most general form as

$$J_{had}^{\alpha} = \cos\theta_c\left(V^{\alpha} - A^{\alpha}\right), \tag{2.5}$$

where θ_c is the Cabibbo angle,cf. [3, Chapter 9], to account for the quark flavour mixing, and $\left(V^{\alpha} - A^{\alpha}\right)$ is an operator accounting for the V-A structure characterizing the neutrino interaction.

Using Lorentz invariance and knowing that the expectation value of a vector is a vector, while the expectation value of an axial vector is a pseudo-vector [12, 11], we can write

$$\left\langle n(p_f)|V^{\alpha}|p(p_i)\right\rangle = \overline{u}_n(p_f)\Gamma_V^{\alpha}u_p(p_i), \tag{2.6}$$

where p and n denote proton and neutron states, respectively, and $u(p_i)$ is the corresponding Dirac spinor of momentum p_i. The most general form of the Γ_V^{α} can be expressed in terms of the available vectors and Dirac matrices combination, cf. [13, Chapter 4], as

$$\Gamma_V^{\alpha} = \gamma^{\alpha}F_V(Q^2) + i\sigma^{\alpha\beta}q_{\beta}\frac{F_M(Q^2)}{2M} + q^{\alpha}F_s(Q^2), \tag{2.7}$$

where F_V, F_M and F_s are form factors, q is the four-momentum transfer and $\sigma^{\alpha\beta} = \frac{i}{2}\left[\gamma^\alpha, \gamma^\beta\right]$. Similarly we may construct the most general axial vector as

$$\Gamma_A^\alpha = \gamma^\alpha \gamma_5 F_A(Q^2) + i\sigma^{\alpha\beta}\gamma_5 q_\beta \frac{F_T(Q^2)}{2M} + q^\alpha \gamma_5 F_p(Q^2). \tag{2.8}$$

Where F_A, F_T and F_p are form factors. With six form factors, some symmetries may be applied to constrain them. Requiring "time invariance"[1] will result in all form factors having real values. Applying Conserved Vector Current (CVC), i.e. $\partial_\alpha V^\alpha = 0$ will result in $F_s = 0$, similar to what is observed for the electromagnetic current. Charge invariance requires $F_T = 0$. Therefore, the hadronic current can be parametrized with only four real form factors as

$$\left\langle n(p_f)|J^\alpha|p(p_i)\right\rangle = \cos\theta_c \overline{u}_n(p_f)\Gamma^\alpha u_p(p_i), \tag{2.9}$$

with

$$\Gamma^\alpha = \gamma^\alpha F_V(Q^2) + i\sigma^{\alpha\beta} q_\beta \frac{F_M(Q^2)}{2M} + \gamma^\alpha \gamma_5 F_A(Q^2) + q^\alpha \gamma_5 F_p(Q^2). \tag{2.10}$$

Using the isotriplet vector current hypothesis[2]

$$F_V(Q^2) = F_1^p(Q^2) - F_1^n(Q^2), \tag{2.11}$$

$$F_M(Q^2) = F_2^p(Q^2) - F_2^n(Q^2). \tag{2.12}$$

Accordingly, we can divide the hadronic current matrix element into two parts: the weak matrix elements (axial) and the electromagnetic matrix elements (vector). The Electromagnetic

[1] It is believed that neutrinos may violate CP. If we believe in CPT invariance, T invariance must also be violated. It is also expected that this violation is very small in the amplitude.

[2] The isotriplet vector hypothesis relates the matrix elements and the form factors of the weak vector current to that of the electromagnetic current. It states that the vector part of the hadron current, its adjoint and isovector part of the electromagnetic current form a triplet with I_z = -1,0 and +1 respectively.

(vector) form factors are well understood from electron scattering experiments

$$\Gamma_V^\alpha = \gamma^\alpha F_1^{p,n}(Q^2) + i\sigma^{\alpha\beta} q_\beta \frac{F_2^{p,n}(Q^2)}{2M}. \tag{2.13}$$

Applying Partially Conserved Axial Current (PCAC) and pion pole dominance [14] relates F_A and F_p together as follows

$$F_p(Q^2) = \frac{2M^2 F_A(Q^2)}{m_\pi^2 + Q^2}. \tag{2.14}$$

From beta decay experiments it was found that $F_A(0) \approx 1.26$ [14]. The exact parameterization of the form factor is unknown. Nevertheless, the assumption of a dipole form seems to fit the data to a good extent with only one free parameter, the axial mass M_A, as follows

$$F_A(Q^2) = \frac{F_A(0)}{1 + \frac{q^2}{M_a^2}}. \tag{2.15}$$

Other parameterizations like the Z-expansion, which are model independent descriptions of the axial form factor, have been recently developed to better fit the data, and to try to fix the large M_A anomaly [14, Sec.2.1].

2.3 Many-particle Many-hole (np-nh)

The MiniBoone experiment [15] has observed an excess of data above the predicted theoretical model. Another interaction channel has been proposed to fill this gap, assuming nucleon-nucleon correlation, and np-nh interaction. In this channel, the neutrino interacts with not only one nucleon but n correlated particles. The two particles-two holes (2p–2h) interaction is composed of contributions from the meson exchange current (MEC) and nucleon-nucleon (NN) correlation [16]. Figure 2.3 shows the Feynman diagrams representing the NN correlations between two nucleons (n1 and n2).

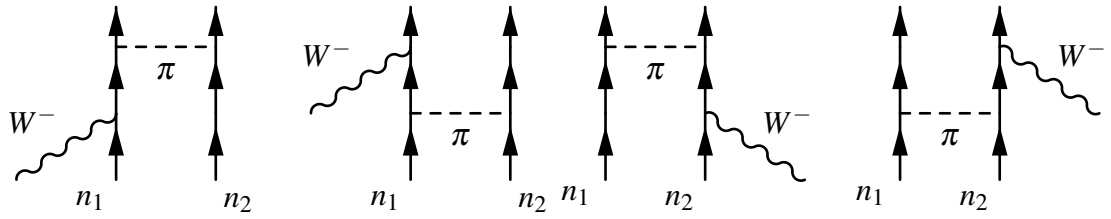


Fig. 2.3 NN interaction Feynman diagram. In (a), n_1 absorbs a W-boson, then exchanges a π meson with n_2. In (b), n_1 exchanges a π meson with n_2 first, then absorbs a W-boson. (c) and (d) describe the same interactions as in (a) and (b) but for n_2, since n_1 and n_2 could be distinguishable particles. In these diagrams, time flows from bottom to top.

The MEC is a two-body current, where the weak boson from the leptonic current is exchanged with a pair of nucleons, leading to two-nucleon emission from the primary vertex, instead of a single nucleon emission from the CCQE. The MEC is also responsible for the so-called “dip region”, i.e. the region between the quasi-elastic and resonance peaks. Taking this component into consideration successfully explained the electron scattering data. Figure 2.4 shows the different components of the MEC.

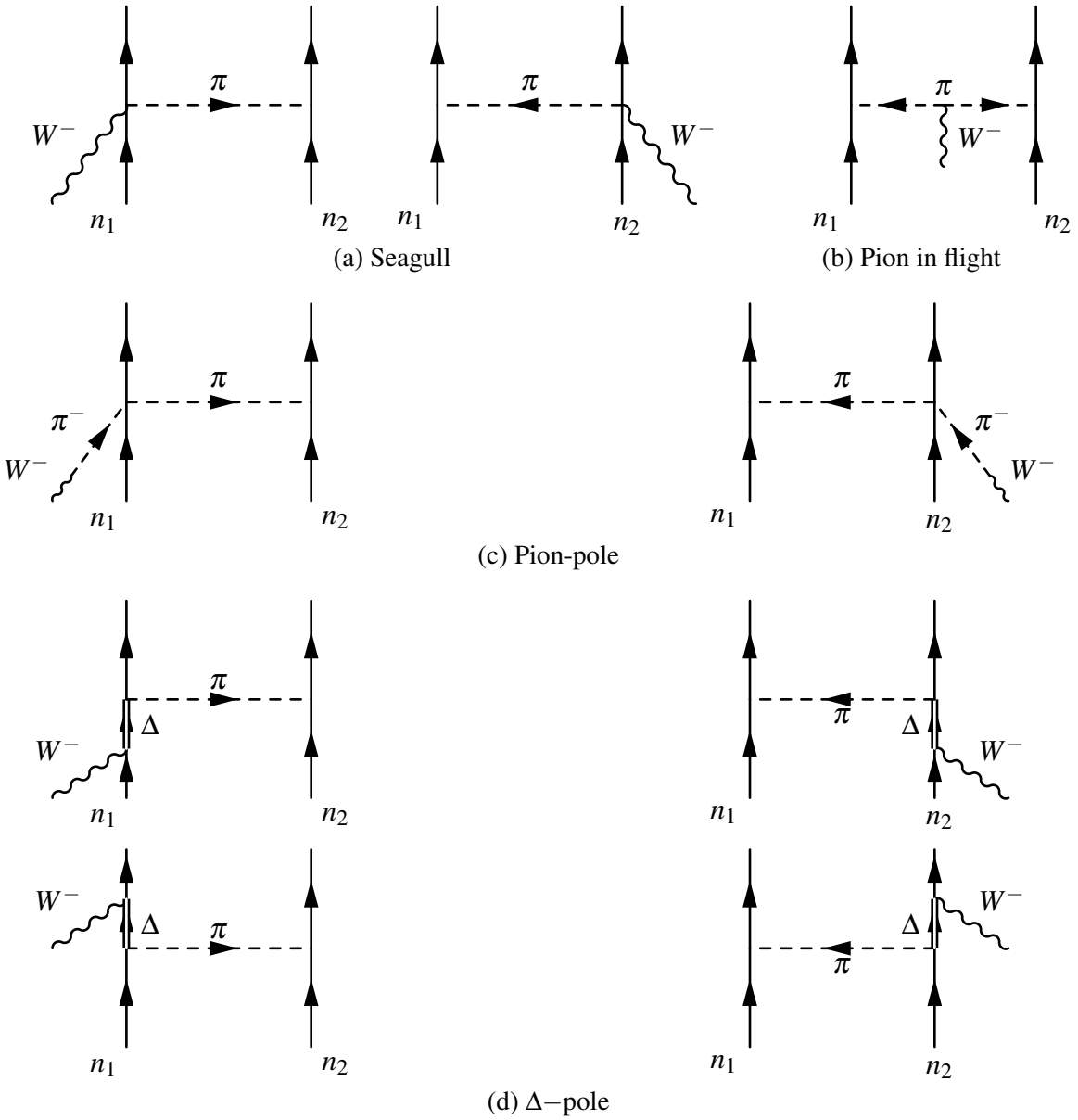


Fig. 2.4 MEC interaction Feynman diagrams. In these diagrams, time flows from bottom to top.

The two-body MEC current is the sum of seagull (Fig. 2.4a), pion-in-flight (Fig. 2.4b), pion-pole (Fig. 2.4c) and Δ-pole (Fig. 2.4d) operators. A full derivation of the MEC model for neutrino-nucleus interaction can be found in [16].

For the neutrino scattering generators, the computation of such processes is very challenging and requires approximations to select the dominant contribution from all possible

Feynman diagrams, hence its implementation is model dependent. In the nuclear model used by T2K (Nieves model) [17], the 2p2h process is the only source of multi-nucleon correlations.

2.4 Resonance

If the interacting neutrino has high enough energy (> 200 MeV), meson production (e.g. π, K, η,etc.) is possible via baryon resonances. The resonance interaction channel (RES) is characterized by an invariant mass < 2 GeV. In this energy region, the effective degrees of freedom are the mesons and baryons.

The dominant resonance is that of the Δ_{33} ($\approx$ 1.232 GeV). Higher resonances like P_{11} ($\approx$ 1.44 GeV), D_{13} ($\approx$ 1.52 GeV), and S_{11} ($\approx$ 1.535 GeV) are also considered in some models. These resonances are mostly spin $\frac{1}{2}$ or spin $\frac{3}{2}$. The nucleon, excited by electromagnetic, strong or weak forces, decays most of the time into a single pion and a nucleon, but can also decay into multiple pions, kaons or even photons.

In the neutrino case, pions can be produced through

$$\nu_l + p \rightarrow l^- + p + \pi^+, \tag{2.16}$$

$$\nu_l + n \rightarrow l^- + n + \pi^+, \text{ and} \tag{2.17}$$

$$\nu_l + n \rightarrow l^- + p + \pi^0. \tag{2.18}$$

For the anti-neutrino case, the interactions are

$$\overline{\nu}_l + p \rightarrow l^+ + n + \pi^0, \tag{2.19}$$

$$\overline{\nu}_l + p \rightarrow l^+ + p + \pi^-, \text{and} \tag{2.20}$$

$$\overline{\nu}_l + n \rightarrow l^+ + n + \pi^-. \tag{2.21}$$

A resonance interaction can be depicted using Feynman diagrams, for instance the electron anti-neutrino interaction of Eq. (2.21), would look like

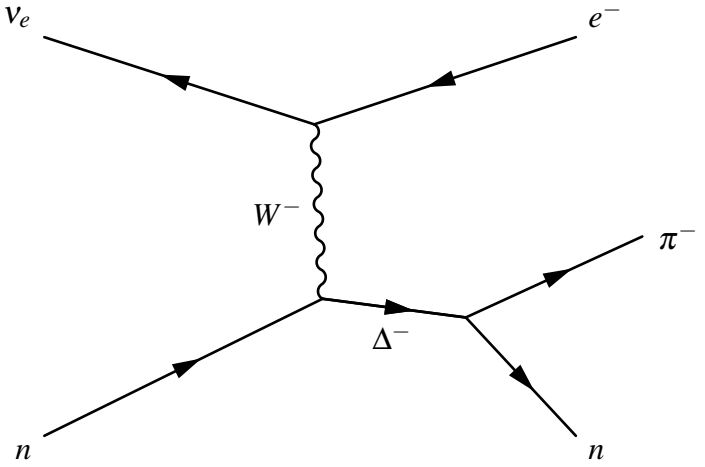


Fig. 2.5 Charged-current resonance interaction Feynman diagram, in which the W^- boson couples to a neutron, producing a Δ^-, that will decay into a neutron and a π^-. In this diagram, time flows from left to right.

The Rein and Sehgal (RS) [18] model is used in T2K, which uses a diagrammatic approach to calculate the corresponding cross-section. RS assumes the interactions

$$\begin{aligned} \nu + N &\rightarrow l + N^*, \\ N^* &\rightarrow N' + m, \end{aligned} \tag{2.22}$$

where ν is the incoming neutrino, l is the outgoing charged lepton, N^* is the intermediate baryon resonance and N' is the final nucleon state.

As in CCQE, Eqs. (2.3) to (2.5), the amplitude of the resonance interaction is given by [11]

$$M_{CCRES} = \frac{G_F \cos\theta_c}{\sqrt{2}} \overline{u}_l \gamma^\mu \left(1-\gamma_5\right) u_\nu \left\langle N' | J_\mu^{CCRES} | N \right\rangle, \tag{2.23}$$

where J_μ^{CCRES} is the hadronic current operator for the resonant single meson production. As in Eq. (2.5), J_μ^{CCRES} contains a vector and an axial-vector part. Both the vector and axial-vector form factors in J_μ^{CCRES} are assumed to be dipoles, whereas its structure is more complicated since it has to account for the spin transition of the resonance state (cf. [19]). The axial current for the spin-$\frac{3}{2}$ resonance is given by

$$A_{3/2}^{\alpha\beta} = -\Big(\frac{C_3^A}{M}(g^{\alpha\beta}\not{q} - q^\alpha\gamma^\beta) + \frac{C_4^A}{M^2}(g^{\alpha\beta} q\cdot p^{'} - q^\alpha p^{'\beta}) + C_5^A g^{\alpha\beta} + \frac{C_6^A}{M^2} q^\alpha q^\beta\Big), \tag{2.24}$$

where C_i^A, with i = 3,4,5 and 6, are the form factors for the nucleon-Δ transition, M is the nucleon mass, g is the metric tensor, $p^{'}$ is the nucleon four-momentum after pion production, q is the transferred four-momentum, and $\not{q}$ is the contracted transferred four-momentum with the gamma matrices, $\not{q} = q_\lambda \gamma^\lambda$. Llewellyn and Smith related all the form factors together, resulting in only two independent parameters to be measured: the axial form factor C_5^A and the resonance axial mass M_{RES}^A, cf. [11]. The axial form factor C_5^A is assumed to have a dipole form given by

$$C_5^A(Q^2) = C_5^A(0)\Big(1+\frac{Q^2}{M_{RES}^{A^2}}\Big)^{-2}\Big(1+\frac{Q^2}{M_A^2}\Big)^{-1}. \tag{2.25}$$

The double differential cross section of single meson production is then given by [10]

$$\frac{d^2\sigma}{dQ^2 d\mathscr{E}} = \frac{1}{32\pi m_N E_\nu^2}\cdot\frac{1}{2}\sum_{spins}|\mathscr{M}|^2\delta(W^2 - M_{N^*}^2), \tag{2.26}$$

where $\mathscr{E} = E_\nu - E_l$ is the energy transfer to the nucleon, E_ν is the energy of the incoming neutrino, E_l is the energy of the outgoing charged lepton, W is the hadronic invariant mass

and M_{N^*} is the resonance mass. The above formula neglects the resonance decay width. To take it into consideration, the δ-function in Eq. (2.26) is replaced with a Breit-Wigner formula

$$\frac{d^2\sigma}{dQ^2 d\mathscr{E}} = \frac{1}{32\pi m_N E_\nu^2} \cdot \frac{1}{2} \sum_{spins} |\mathscr{M}|^2 \frac{\Gamma}{2\pi((W-M)^2 + \frac{\Gamma^2}{4})}, \tag{2.27}$$

where Γ is the decay width of N^*. The RS model includes 18 possible baryon resonances with mass less than 2 GeV. The model parameters are then updated from the bubble chamber experiments, BNL [20] and ANL [21], to best fit the data.

In RS model, pion production is affected by the initial state of the nucleon, as well as the final state interaction, hence is model dependent. For example, $NC1\pi^0$ is the same as $NC1\pi^+$ if the π^+ had a charge exchange inside the nucleus and emitted a π^0. Also, the decay width of the Δ depends on the nuclear medium model. Generally, it is bigger than that of the free case, since we do have more decay channels. For example, in nuclear medium we may have $\Delta \to N\pi$, $\Delta N \to NN\pi$, $\Delta NN \to NN\pi$, etc. while in the free case we only have $\Delta \to N\pi$.

2.5 Shallow and Deep Inelastic Scattering

There is no precise way to distinguish between the Deep Inelastic Scattering (DIS) and the Resonance (RES). However, the DIS is usually characterized by an invariant mass W $\geq$ 2.0 GeV and a transferred four-momentum $Q^2 \geq 1.0$ GeV. At this energy level, the neutrino can probe the structure of the nucleon and interact with a quark inside the bound nucleon [22].

The general form of the cross-section is then given by [22]

$$\frac{d^2\sigma}{d\Omega' dE'} = \frac{G_F^2}{(2\pi)^2} \frac{|\vec{k'}|}{|\vec{k}|} \frac{m_W^2}{q^2 - m_W^2} L^{\alpha\beta} W_{\alpha\beta}, \tag{2.28}$$

where $L^{\alpha\beta}$ is the leptonic tensor. In the limit of $m_l \to 0$,

$$L^{\alpha\beta} = k^\alpha k'^\beta + k^\beta k^\alpha - k.k' g^{\alpha\beta} \pm i\varepsilon^{\alpha\beta\rho\sigma} k_\rho k'_\sigma, \tag{2.29}$$

where the "+" is used for neutrino and the "-" for the anti-neutrino case, k is the four-momentum and ε is the Levi-Civita tensor.

Moreover, the hadronic tensor $W_{\alpha\beta}$ is given by

$$W_{\alpha\beta} = \left(\frac{q_\alpha q_\beta}{q^2} - g_{\alpha\beta}\right)W_1 + \frac{1}{M^2}\left(p_\alpha - \frac{p.q}{q^2}q_\alpha\right)\left(p_\beta - \frac{p.q}{q^2}q_\beta\right)W_2 - \frac{i}{2MA}\varepsilon_{\alpha\beta\rho\sigma}p^\rho q^\sigma W_3, \tag{2.30}$$

where the W_i's are the structure functions, expressed in form factors as a function of the fraction Bjorken $x = \frac{Q^2}{p.q}$, and the four momentum transfer $q^2 = -Q^2$. DIS measures the structure function in the leading order approximation of perturbative QCD, taking only the four light quarks into consideration (u,d,c,s).

The Shallow Inelastic Scattering (SIS) forms the region between RES and DIS, and is poorly understood. Some generators, like NEUT [9] and GENIE [23] used by T2K, have a discontinuity between the RES and DIS regions, while others, e.g. NUWRO [24], adapt the idea of Quark-Hadron (QH) [22] duality to have a smooth transition covering the SIS.

The QH duality assumes that the RES structure function at low Q^2 averaged over the energy range of the RES equals the DIS structure function at high Q^2 for the same energy range. This method connects the nucleon and meson degrees of freedom with those of the quark and gluon.

DIS is also used by all generators to account for the non-resonant pion production. At lower values of Q^2, e.g. $Q^2 \leq 10$ GeV, QCD is non-perturbative [22], so most generators use an empirical implementation based on what is called "higher twist contribution" to qualitatively describe the process. In this method, the form factors are represented as quark plus some correction terms that arise from partons correlations.

The hadronization process in DIS is characterized by the Fragmentation Function (FF), which is the probability distribution to produce a hadron with a certain fraction of the longitudinal momentum of the scattered parton. The fragmentation function is determined

phenomenologically from empirical models fit to data, and implemented in PYTHIA [25], by fitting the free parameters to the HERMES [26] experiment data.

In the Parton model, each quark contribution is weighted by the parton distribution. Therefore, the cross-section is then given by [27]

$$\frac{d^2\sigma}{dxdy} = \frac{G_F^2 m_N E_\nu}{\pi\left(1+\frac{Q^2}{M_W^2}\right)^2}\left[(1-y-\frac{M_N xy}{2E_\nu})F_2 + y^2 xF_1 \pm y(1-\frac{y}{2})xF_3\right], \tag{2.31}$$

with

$$x = \frac{Q^2}{2m_N(E_\nu - E_l)}, \text{ and} \tag{2.32}$$

$$y = \frac{E_\nu - E_l}{E_\nu}. \tag{2.33}$$

Here, F_1, F_2 and F_3 are structure functions related to the parton distribution functions. They are functions of Q^2 and the Bjorken x. The function F_2 measures the density distribution of all quarks and anti-quarks in the nucleon. The function F_3 accounts for parity violating process, changes sign for ν and $\overline{\nu}$, and measures the valence quark distribution.

2.6 Final State Interactions

Final State Interactions (FSI) account for the scattering of the original hadron from the primary vertex inside the nucleus. For instance, the generated pion may suffer from:

- elastic scattering, changing the pion final momentum,
- absorption, e.g. $\pi^+ + d \rightarrow p + p$,
- charge exchange , e.g. $\pi^+ + n \rightarrow \pi^0 + p$,
- production of other hadrons, e.g. $\pi^+ + n \rightarrow \pi^+ + \pi^0$,

or it can knock out other nucleons. Consequently, FSI can mask the neutrino interaction, as a CC1π may look like a CC0π due to pion absorption, or smear the visible hadron multiplicity, making the reconstruction of the neutrino energy very challenging and model dependent. Figure 2.6 shows the different pion interaction effects inside the nucleus.

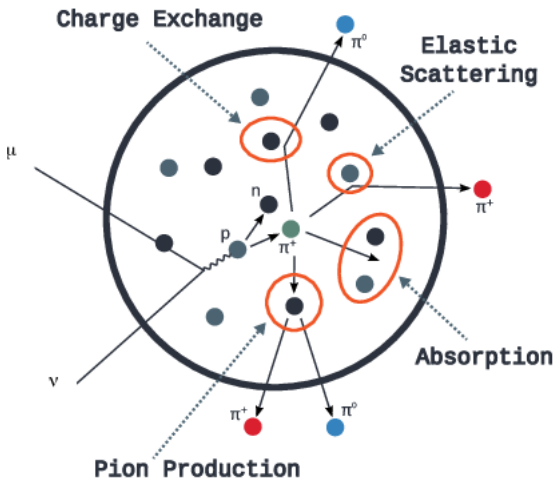


Fig. 2.6 A diagram showing different pion interactions inside the nucleus, such as absorption, elastic scattering, charge exchange and pions production. Figure courtesy of [28]

NEUT uses a semi-classical cascade model to implement FSI. In this model, the original hadron is stepped inside the nuclei, and at each step the mean free path $\lambda(r)$ is calculated as

$$\lambda(r) = \left(\sigma_{\pi_{interaction}} \rho(r)\right)^{-1}, \tag{2.34}$$

where $\sigma_{\pi_{intearction}}$ is the pion cross-section. NEUT uses the Oset model [29], a many body calculation in an infinite nuclear matter plus a local density approximation, to calculate the

pion cross-section. The local nuclear matter density, $\rho(r)$, depends on the nuclear mass number. An empirical equation with free parameters (α, ω and c) is used to model $\rho(r)$, and is taken to be of the form

$$\frac{\rho(r)}{\rho(0)} = \frac{1+\omega\frac{r^2}{c^2}}{1+e^{\frac{r-c}{\alpha}}}. \tag{2.35}$$

Monte-Carlo techniques are used to determine if an interaction took place or move to the next step.

2.7 Coherent and Diffractive Scattering

Coherent scattering is the interaction channel in which the neutrino interacts with the nucleus as a whole leaving it at the ground state, where no nucleon is knocked out. On the other hand, in the diffractive channel, a neutrino interacts with the nucleon itself and leaves the nucleon in the ground state without knocking it out of the nucleus. In these channels, the transferred energy to the nucleon is close to zero.

Coherent elastic neutrino-nucleus scattering does not affect the neutrino oscillation analysis. However, it represents an irreducible background for the dark matter search. Nevertheless, coherent meson and photon production can mimic the signal definition of the neutrino–nucleus cross-section and hence affect the oscillation analysis.

In the charged-current channel, the neutrino interaction can be represented as follows

$$\nu_l + A \rightarrow l^- + m^+ + A, \tag{2.36}$$

$$\overline{\nu}_l + A \rightarrow l^+ + m^- + A, \tag{2.37}$$

where $m^\pm = \pi^\pm, K^\pm, etc.$ is the produced charged meson.

In the neutral current channel,

$$\nu_l + A \rightarrow \nu_l + m^0 + A, \tag{2.38}$$

$$\overline{\nu}_l + A \rightarrow \overline{\nu}_l + m^0 + A, \tag{2.39}$$

where $m^0 = \pi^0, k^0$, etc. is the produced neutral meson.

The electron anti-neutrino coherent and diffractive pion production can be depicted in Feynman diagrams, as shown in Fig. 2.7.

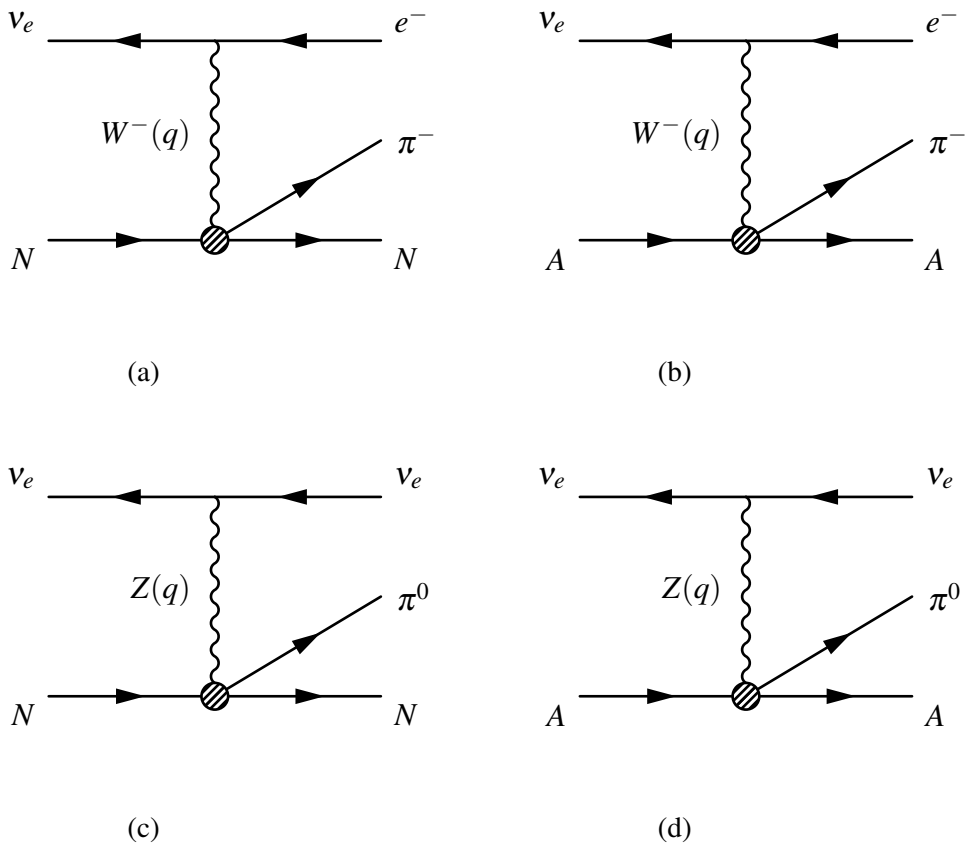


Fig. 2.7 (a) π^- production from $\overline{\nu}_e$ charged-current diffractive scattering. (b) π^- production from $\overline{\nu}_e$ charged-current coherent scattering. (c) π^0 production from $\overline{\nu}_e$ neutral current diffractive scattering. (d) π^0 production from $\overline{\nu}_e$ neutral current coherent scattering. In these diagrams, time flows from left to right.

In Fig. 2.7, the square of the four-momentum transferred to the nucleon/nucleus t is given by

$$t = \left(q - p_m\right)^2, \tag{2.40}$$

where q and p_m are the total transferred four-momentum and the meson four-momentum respectively.

For a small energy transfer to the nucleus/nucleon, most of the energy of the W is transferred to the pion, leading to

$$q^0 \approx p_m^0. \tag{2.41}$$

In that case we can write

$$t \approx -\left(\vec{q} - \vec{p}_m\right)^2, \tag{2.42}$$

with

$$|\vec{p}_m| \approx \sqrt{q_0^2 - m^2}. \tag{2.43}$$

In order to achieve a small momentum transfer to the nucleus, the following three conditions must be fulfilled:

1. $\vec{q}$ and $\vec{p}_m$ are parallel,
2. $q_0 \approx |\vec{q}|$, as $q^2 = q_0^2 - \vec{q}^2$, and
3. the produced particle mass m is small.

In the Diffraction process, even with a small energy transfer the nucleon may be ejected out of the nucleus and detected, mimicking a CCQE interaction channel. This phenomenon can be understood as follows, since for the nucleus

$$t = (p' - p)^2 = -2m_A T_A, \tag{2.44}$$

where m_A is the mass of the nucleus, T_A is the kinetic energy of the nucleus in the lab frame, and p and p' are the nucleus' four-momentum before and after the interaction respectively. For a nucleon

$$t = (p' - p)^2 = -2m_N T_N, \tag{2.45}$$

where M_N is the mass of the nucleon, T_A is the kinetic energy of the nucleon in the lab frame, p and p' are the nucleon four-momentum before and after the interaction respectively. Since $m_A \gg m_N$, $T_N \gg T_A$ and the nucleon can escape the nucleus and be experimentally detected.

The coherent and diffraction interaction channels have a non-negligible effect on the oscillation analysis. The π^0 produced from the neutral current channel can represent a background for the $\nu_\mu \to \nu_e$ oscillation, particularly when the resultant pair of photons ($\pi^0 \to \gamma\gamma$) are collinear or one is not detected. Another source of systematic error comes from the misclassification of the produced π^+ as a proton.

Using the partially conserved axial current (PCAC) hypothesis, at low Q^2, the coherent pion cross-section is related to pion-nucleus elastic scattering by the soft pion theorem [30, 22] as follows

$$\frac{d\sigma}{dq^2 dy dt}|_{q^2=0} = r\frac{G_F^2 f_\pi^2}{2\pi^2}\frac{1-y}{y}\frac{d\sigma}{dt}(\pi A \to \pi A)|_{q^2=0,\omega_{pi}=q^0}, \tag{2.46}$$

where $y = \frac{q^0}{E_\nu}$, $r = 1$ for NC and = $2|V_{ud}|^2$ for CC , and f_π = 92.4 MeV.

Rein and Sehgal built an empirical model [31] for the coherent pion production, using the above theory. Their model only works at $q^2 \approx 0$ and does not simulate other meson production. The heavier meson production, e.g. K production, will be suppressed twice: once from Cabbibo suppression due to the strange quark, and then kinematically, since $m_k > m_\pi$.

2.8 Short-Range Correlation

Short-Range Correlation (SRC) models the nucleon pairs that are very close to each other in space, such that their wave functions overlap. In momentum space, they will be characterized by high relative momentum ($\gg p_F$) and low centre of mass (CM) momentum ($< p_F$), where p_F is the Fermi momentum.

When a neutrino interacts with one of these correlated pairs, two nucleons are ejected back to back in their CM frame. SRC accounts for the long tail in the nucleon momentum distribution.

Chapter 3

Neutrino Interaction Models

The charged-current inclusive data at T2K is successfully modelled by adding the following interaction channels and nuclear effects: Quasi-Elastic, Relativistic Fermi Gas (RFG) model, Random Phase Approximation (RPA), np–nh, and single coherent and incoherent pion production. The next sections present a brief overview of these models along with their main limitations.

3.1 Relativistic Fermi Gas (RFG)

Relativistic Fermi Gas (RFG) is the simplest possible model for nuclear effects. It is based on the Smith-Moniz model [32]. This model treats nucleons as a non-interacting Fermi gas, in the limit of infinite nucleus radius. It has two parameters only: a global Fermi momentum p_F and a constant nucleon binding energy E_B. Figure 3.1 sketches the main properties of this model.

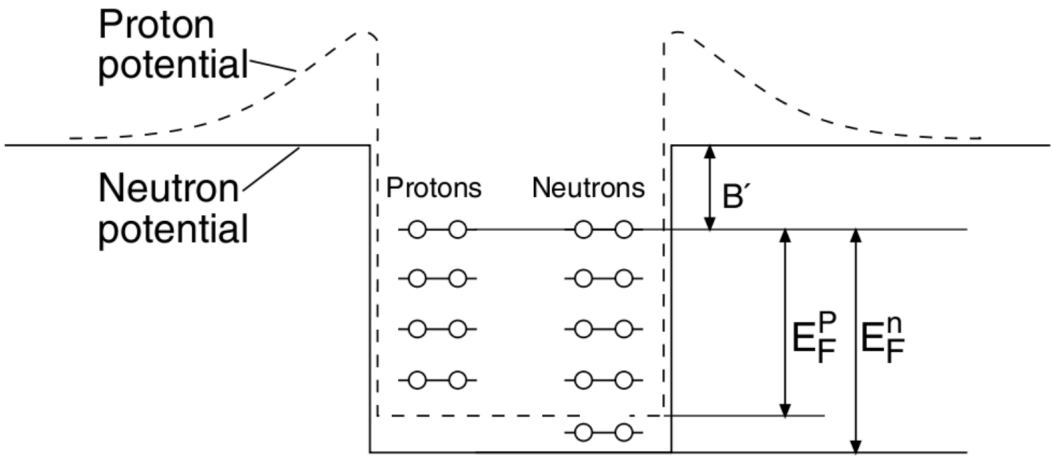


Fig. 3.1 Fermi Gas model for proton and neutrons. Figure courtesy of [33].

Since we can distinguish between protons and neutrons, they are positioned in two separate potential wells. The nuclear potential wells are approximated to be constant inside the nucleus and stop sharply at the end.

Each energy state can be occupied by only two nucleons with different spin projections to satisfy the Pauli exclusion principle. At zero absolute temperature T = 0, all available energy states are filled up to E_F. Mathematically, this is modelled using a simple step function $\Theta(p_F - p)$.

The difference, B', between the top of the well and the Fermi level is constant for most nuclei and represents the average binding energy per nucleon.

In a stable nucleus, the Fermi level of the protons and neutrons have to be equal, otherwise the nucleon will move to a more energetically favoured state through beta-decay. For heavier nuclei having more neutrons than protons, the depth of the neutron potential well has to be larger than that of the proton, such that they have the same final E_F. This implies that protons are, on average, less bounded in nuclei than neutrons, which can be understood as a consequence of the Coulomb repulsion of the charged protons.

The A dependence of the finite nuclear size is implemented by choosing a suitable E_B and p_F.

In this model, a neutrino knocks out one nucleon, producing one particle and one hole (1p–1h). The knocked-out nucleon is required to have a momentum larger than p_F due to the Pauli blocking effect. This model neglects any nucleon-nucleon correlation and does not include two-body currents like two particles two holes (2p–2h). Using this model alone, failed to reproduce the inclusive cross-section measurements.

A more sophisticated Fermi Gas model is also implemented in T2K, called Local Fermi Gas (LFG) (cf. [17]), in which the momentum of the initial nucleon depends on its radial position inside the nucleus. LFG is planned to be used in the future T2K analysis.

3.2 Random Phase Approximation (RPA)

Random Phase Approximation (RPA) was first introduced by Bohm and Pines [34] in the 1950s to describe plasma oscillations. Later it was also adopted in nuclear physics as a phenomenological nuclear potential.

RPA is used to describe the collective vibrations that represent a coherent motion of many nucleons. At low energies, a few MeV, these vibrations result in nuclear surface vibration, while at higher energies, 10–30 MeV, it is responsible for the density oscillation a.k.a Giant Resonance.

RPA assumes that the excited states of the many-body system can be described as linear combinations of one-particle on-hole (1p–1h) excitations. If one uses a mean-field model, like Hartree-Fock (HF) [35], in the ground state the effect of a 1p–1h excitation can be represented as follows

$$|1p-1h\rangle = \hat{c}_m^{\dagger} c_i |\mathrm{HF}\rangle, \tag{3.1}$$

where $\hat{c}_m^\dagger$ is the one particle creation operator with energy state m, and c_i is the one particle annihilation operator with energy state i. In this framework, the collective vibration, modelled as a superposition of large number of particle-hole excitations, can be represented as

$$|\mathrm{RPA}\rangle = \sum_{m,i} A_{m,i} \hat{c}_m^\dagger c_i |\mathrm{HF}\rangle. \tag{3.2}$$

The goal of the theory is to find the coefficients of the linear combinations of $A_{m,i}$ for a given interaction between particles and holes.

RPA accounts for the long-range correlation between nucleons within a nucleus. This long-range correlation is more effective at low and intermediate energies, where the peak of CCQE lies. It is normally added as a correction to the RFG model and the effect is to reduce the CCQE cross-section at lower energies, as shown in Fig. 3.2.

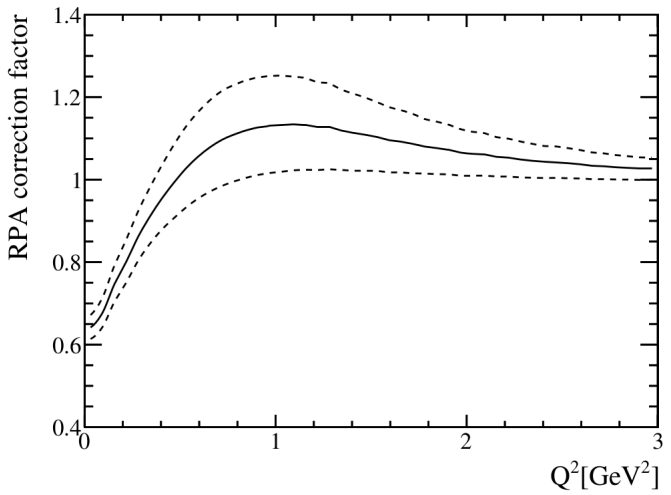


Fig. 3.2 The nominal relativistic RPA correction factor, implemented by Nieves [17], relative to the un-modified CCQE cross-section is shown in solid line, along with the $\pm 1\sigma$ uncertainties (dashed lines). Figure courtesy of [36].

3.3 Spectral Function (SF)

The Spectral Function (SF) model [37] is a more sophisticated description of nuclear matter and provides an alternative for RFG. SF is based on a two-dimensional distribution of the nucleon momentum and energy inside the nucleus. The SF model depends on the target. Therefore we have specific spectral functions for each nucleus.

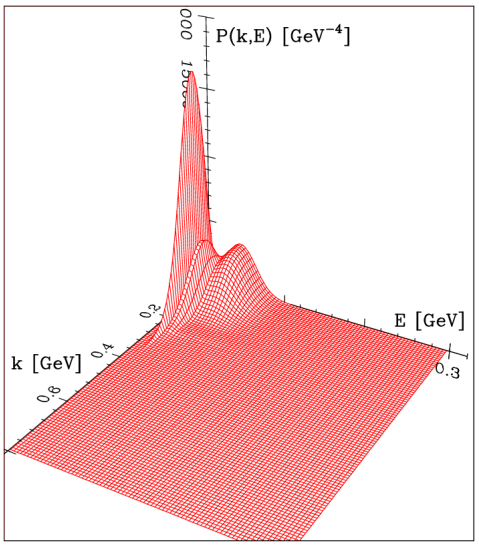


Fig. 3.3 Three-dimensional plot of the spectral function of Oxygen, showing the energy, E, and momentum, K, distribution of the nucleons. Figure courtesy of [37].

Unlike RFG, SF includes the strong nucleon-nucleon (NN) correlation. This correlation results in increasing the tail of the neutrino interaction cross-section at high energy transfer, and reducing the cross-section for small energy transfer. The integrated effect is a suppressed total cross-section by 5-10% compared to that predicted using the RFG model [36]. The increase in the high momentum tail explains the short-range correlation effects, as described in Section 2.8. If the RFG model is replaced by the SF model, the axial mass parameter, M_A, must be increased to account for the decrease of the total cross-section.

3.4 Other Models

There are several other neutrino-nucleus interaction models developed and used by other neutrino experiments, like NOvA [38], Minerva [39], MiniBoone [15], etc. This section will present a very brief overview of the commonly used models.

Green Function Monte Carlo (GFMC)

A non-relativistic Hamiltonian model that works only for Quasi-Elastic scattering, but supports two body currents, meson exchange current (MEC), and nucleon correlation. For more details, refer to [40, 41].

Relativistic Green Function (RGF)

The RGF model is based on the spectral function and supports 2p–2h as well as final state interactions (FSI). For more details, refer to [42].

Valencia Group

The Valencia model is based on the relativistic Fermi gas (RFG) model but supports long-range nucleon correlation using Random Phase Approximation (RPA), as well as 2p–2h. For more details, refer to [43].

Ghent HF-CRPA

The Ghent model is based on a mean-field technique (Skyrme based Hartree Fock) that also supports long-range correlation (RPA). For more details, refer to [44].

Coherent Density Fluctuation Model (CDFM)

CDFM includes the effect of short-range correlation that will result into high momentum tails of the struck hadrons. For more details, refer to [45].

Giessen Boltzmann-Uehling-Uhlenbeck (GiBUU)

GiBUU is a relativistic model and a theoretical framework based on kinetic transport theory [46, 12]. GiBUU was developed to describe general nuclear reactions, e.g. proton-nucleus, electron-nucleus, pion-nucleus, and neutrino-nucleus scattering. It supports a wide range of energies ($\sim$ 20 GeV for hadronic reactions and $\sim$ 100 GeV for leptonic

reactions), making it suitable to be used for all reactions, including resonance and deep inelastic scattering. It includes reasonable bound nuclear ground states and considers the nuclear potential for both incoming and outgoing generated particles (FSI). GiBUU also implements the 2p–2h structure function for all kinematics, not just the Delta region, and is fitted to electron scattering data. It can be used to give both inclusive cross-sections and full event simulation. It has good agreement with T2K and MINERVA data, without any special tuning, but failed to reproduce MiniBooNE results [22].

Lattice QCD (LQCD)

LQCD is a numerical, first-principles calculation approach. In LQCD, the fields are defined on a discrete Euclidean space-time which naturally introduces a cutoff in a non-perturbative way, solving the Ultraviolet UV divergence problem of QCD. In this framework, time is complexified, $t \rightarrow i\tau$, and the position four-vector is discretized, $x_\mu = n_\mu a$, where "a" is the finite lattice spacing, $n \in \mathbb{Z}$, $\mu = 0,..,D-1$ for a "D" dimension space-time, and the QCD path integral is approximated by Monte Carlo integration. In order to obtain fully physical results, the cutoff is removed at the end of the calculation by taking the continuum limit (CL) where $a \rightarrow 0$ in the interacting quantum theory. Lattice QCD provides predictions of the nucleon and nuclear matrix elements. It has a successful history of determining the standard model parameters [47], the neutron-proton mass splitting [48], the CKM matrix elements [47], and single nucleon form factors [49]. It has been also used in neutrino-nucleon interactions to calculate the axial mass "M_A" in QE scattering [49], the $N\Delta$ transition form factor [50], and low momentum PDF in DIS [51]. Nuclear effects have been studied, and LQCD was used to predict the matrix elements in light nuclei ($A < 5$) from first principles. The nucleus on the lattice technique is quite challenging as the statistical uncertainty grows exponentially with the number of nucleons, and the complexity increases substantially, requiring huge computational power. It is expected that LQCD will provide precise calculations with controlled percent-level uncertainties within a few years, which will have a potential

impact on the neutrino energy determinations.

It is worth noting that a hybrid model is used by T2K, also called the “Frankenstein” model. This model normally starts with impulse approximation and then adds nucleon-nucleon interaction effects. It selects a target in the nuclear potential with a certain initial state, using something like RFG, Local Fermi Gas (LFG) or SF. It then adds the long-range correlation and the 2p–2h effect and finally propagates the struck nucleon through the nuclear media to account for the FSI.

Chapter 4

Neutrino Oscillation

Neutrinos are generated in weak decays of nucleons. The standard model (SM) has three flavours of neutrinos, (ν_α, with $\alpha = e, \mu, \tau$) labelled according to their charged lepton partner l_α in the decay of a heavy charged boson $W^\pm$

$$W^- \longrightarrow l_\alpha + \overline{\nu}_\alpha, \text{ and} \tag{4.1}$$

$$W^+ \longrightarrow \bar{l}_\alpha + \nu_\alpha. \tag{4.2}$$

These three flavour eigenstates represent a complete orthonormal basis for the SM neutrinos. Neutrino oscillation, that is changing of flavour during propagation, necessitates that neutrinos have masses in the SM. In the three neutrinos oscillation model, there is a spectrum of three mass eigenstates, ν_i, each with mass m_i that will diagonalize the neutrino free particle Hamiltonian. The flavour and mass eigenbases are not the same, so each neutrino flavour is a superposition of the mass eigenstates, described by the leptonic mixing, Pontecorvo-Maki-Nakagawa-Sakata (PMNS), matrix U:

$$|\nu_\alpha\rangle = \sum_{i=1}^{3} U^*_{\alpha i} |\nu_i\rangle. \tag{4.3}$$

For the three neutrino flavour case, the mixing matrix has four parameters, including three mixing angles θ_{12}, θ_{13}, θ_{23}, and one complex phase δ_{CP}. The mixing matrix is given by

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.4}$$

This matrix can be expressed as

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{4.5}$$

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$.

Figure 4.1 shows a neutrino production of flavour ν_α that propagates through space for a macroscopic distance L, and it interacts at the target side producing a charged lepton of flavour l_β.

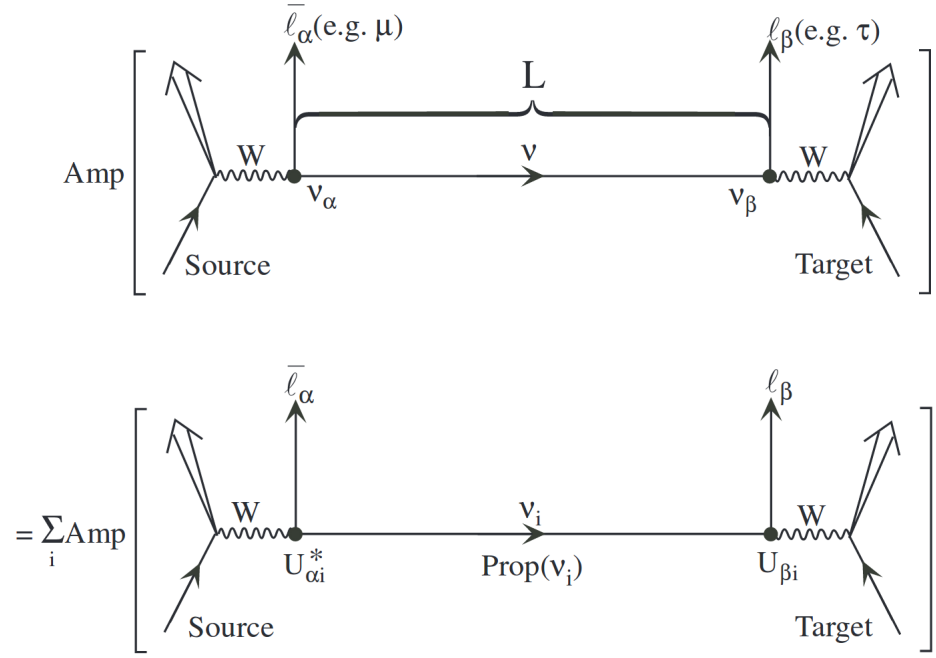


Fig. 4.1 Neutrino oscillations in vacuum. Figure courtesy of [52].

To calculate the probability amplitude of the oscillation, $P(\nu_\alpha \longrightarrow \nu_\beta)$, we start in the rest frame of the ν_i neutrino mass eigenstate and apply Schrödinger's equation[1], which gives

$$|\nu_i(t)\rangle = e^{-i(E_i t - \vec{p}_i \cdot \vec{x})} |\nu_i(0)\rangle, \tag{4.6}$$

where quantities are expressed in natural units, E_i is the Energy, $\vec{p}_i$ is the momentum, $\vec{x}$ the position vector and t is the time. In the ultra-relativistic limit, $|p_i| \gg m_i$, we can approximate the energy as

$$E_i = \sqrt{p_i^2 + m_i^2} \approx p_i + \frac{m_i^2}{2p_i} \approx E_i + \frac{m_i^2}{2E_i}. \tag{4.7}$$

[1] Schrödinger's equation describes the changes over time of the wave function ψ of the system, and is given by

$$i\hbar \frac{\partial}{\partial t} |\psi\rangle = \hat{H} |\psi\rangle,$$

where $\hat{H}$ is the Hamiltonian of the system and t is the time. In our example we assume a free particle.

In this case the wave function becomes

$$|\nu_i(L)\rangle = e^{-i(\frac{m_i^2 L}{2E})}|\nu_i(0)\rangle, \tag{4.8}$$

and the probability amplitude of the oscillation is

$$P(\nu_\alpha \longrightarrow \nu_\beta) = |\langle \nu_\beta | \nu_\alpha(L)\rangle|^2. \tag{4.9}$$

Representing the mass eigenstate as a superposition of the flavour states, one obtains

$$|\nu_i\rangle = \sum_\alpha U_{\alpha i}|\nu_\alpha\rangle. \tag{4.10}$$

Using Eqs. (4.3) and (4.10) in Eq. (4.9)

$$P(\nu_\alpha \longrightarrow \nu_\beta) = |\langle \nu_\beta | \sum_i U^*_{\alpha i} e^{-im_i t} \sum_\gamma U_{\gamma i}|\nu_\gamma\rangle|^2. \tag{4.11}$$

Assuming that the PMNS matrix is unitary, and after some algebra, one can show that

$$\begin{aligned} P(\nu_\alpha \longrightarrow \nu_\beta) =& \delta_{\alpha\beta} \\ & - 4\sum_{i>j} \Re(U^*_{\alpha i}U_{\beta i}U_{\alpha j}U^*_{\beta j}) \sin^2(\Delta m^2_{ij}(L/4E)) \\ & + 2\sum_{i<j} \Im(U^*_{\alpha i}U_{\beta i}U_{\alpha j}U^*_{\beta j}) \sin(\Delta m^2_{ij}(L/2E)), \end{aligned} \tag{4.12}$$

where $\Delta m^2_{ij} = m^2_i - m^2_j$. In commonly used units: $\Delta m^2_{ij}(L/4E) \simeq 1.27\Delta m^2_{ij}(\mathrm{eV}^2)$ (L (km)/E (GeV)).

From Eq. (4.12) it is clear that for oscillation, neutrinos must have masses, i.e. $\Delta m^2_{ij} \neq 0$, and lepton flavour mixes, i.e. PMNS is not a diagonal matrix.

One of the still open questions in neutrino physics is the mass hierarchy. In Fig. 4.2 each

neutrino mass eigenstate is represented as summation of neutrino flavour eigenstates, where the percentage reflects $|U_{\alpha i}|^2$.

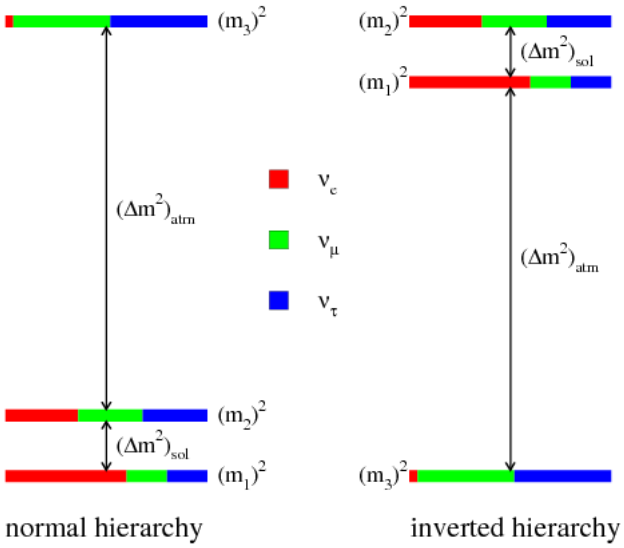


Fig. 4.2 Neutrino mass hierarchy. Figure courtesy of [53].

Neutrino oscillation can be detected either by the appearance of ν_α in a ν_β beam or the disappearance of some ν_β from the beam. Each experiment has a specific (L/E) ratio, making it sensitive to certain Δm^2_{ij}.

Recent experiments agreed on the values of $\Delta m^2_{21} = (7.50 \pm 0.185) \times 10^{-5}$ eV^2 and $\Delta m^2_{32} = (2.47^{+0.069}_{-0.067}) \times 10^{-3}$ eV^2 [54]. We can approximate the three neutrino model in a two neutrino model because of the orders of magnitude difference between the mass squared splitting. In such a case we will have

$$P(\nu_\alpha \longrightarrow \nu_\beta) = \sin^2(2\theta)\sin^2\left(1.27\Delta m^2_{ij}(\text{eV}^2)\frac{\text{L(km)}}{\text{E(GeV)}}\right). \tag{4.13}$$

The oscillation does not change the total neutrino flux, but redistributes it among different flavours. Some experiments suggest there may be additional neutrinos, called sterile neutrinos [54], that do not interact via the weak interactions. The three weakly interacting neutrino flavours may oscillate into a sterile neutrino and vice versa.

4.1 Matter Effects

The propagation of the neutrinos can be significantly affected when they travel through a dense medium, like the Sun or the Earth. This can be explained by coherent neutrino scattering with the particles they encounter in that medium. Mikhaev, Smirnov and Wolfenstein (MSW) were the first to show that the neutrino oscillation probability is different in case of a dense medium than it is in the vacuum.

The MSW effect originates from the fact that electron neutrinos and electron anti-neutrinos have different interactions with matter than the other flavours. For example, ν_e can interact with electrons through charged-current as well as neutral current, while ν_μ and ν_τ can only interact through neutral current with the electrons to conserve the lepton number of the interaction. This will give rise to an extra potential V_e that will only affect ν_e. This potential is proportional to the electron density of matter N_e, as follows

$$V_e = \pm 2 G_F N_e. \tag{4.14}$$

Where G_F is the Fermi constant, and the "+", "-" signs apply for the neutrino and anti-neutrino respectively. To illustrate how MSW affects the oscillation, let's work in the simplest case of two flavours, ν_e and ν_μ. In case of propagation in vacuum and using approximation Eq. (4.7) and knowing that adding any constant to the Hamiltonian, i.e. any multiple of $\mathbb{1}$, does not change the physics as it will just add a global constant phase to the wave equation, the Schrödinger equation can be written as

$$i\frac{\partial}{\partial t}\begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \end{pmatrix} = \frac{1}{2E}\begin{pmatrix} m_1^2 & 0 \\ 0 & m_2^2 \end{pmatrix}\begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \end{pmatrix}. \tag{4.15}$$

Approximating Eq. (4.3) to the case of two flavours only, the unitary matrix U relating mass-eigenstates to the flavour eigenstates can be represented as a simple rotation

$$\begin{pmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} |\nu_1\rangle \\ |\nu_2\rangle \end{pmatrix}. \tag{4.16}$$

Where θ is the mixing angle.

In this case, the Schrödinger equation in the flavour space is res presented by

$$i\frac{\partial}{\partial t}\begin{pmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \end{pmatrix} = U\frac{1}{2E}\begin{pmatrix} m_1^2 & 0 \\ 0 & m_2^2 \end{pmatrix} U^\dagger \begin{pmatrix} |\nu_e\rangle \\ |\nu_\mu\rangle \end{pmatrix}. \tag{4.17}$$

The vacuum Hamiltonian in the flavour space is

$$H_V = \frac{\Delta m^2}{4E}\begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}. \tag{4.18}$$

Now, to move to the Hamiltonian in the medium, we add the potential V_e from Eq. (4.14), which will give

$$H_\mathrm{m} = \frac{\Delta m^2}{4E}\begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix} + \begin{pmatrix} V_e & 0 \\ 0 & 0 \end{pmatrix}. \tag{4.19}$$

Again we can add/subtract any multiples of the unit matrix from the Hamiltonian without changing the physics. So subtracting $\begin{pmatrix} \frac{V_e}{2} & 0 \\ 0 & \frac{V_e}{2} \end{pmatrix}$ will simplify the Hamiltonian form to be

$$H_\mathrm{m} = \frac{\Delta m^2}{4E}\begin{pmatrix} -\cos 2\theta + A & \sin 2\theta \\ \sin 2\theta & \cos 2\theta - A \end{pmatrix}, \tag{4.20}$$

with

$$A = \pm \frac{2\sqrt{2} G_F N_e E}{\Delta m^2}. \tag{4.21}$$

If the matter density is constant, the solution of the corresponding Schrödinger equation is simple and can easily be found by diagonalizing H_{m} to create the new mass-eigenstates in case of propagation through a medium. Then we can repeat the same process similar to equations Eqs. (4.15) to (4.18), to obtain an effective mixing angle in matter θ_{m} and an effective difference of squared masses Δm^2_{m}. The Hamiltonian in matter will have the same form as the vacuum Hamiltonian but with the effective parameters as

$$H_{\mathrm{m}} = \frac{\Delta m^2_{\mathrm{m}}}{4E} \begin{pmatrix} -\cos 2\theta_{\mathrm{m}} & \sin 2\theta_{\mathrm{m}} \\ \sin 2\theta_{\mathrm{m}} & \cos 2\theta_{\mathrm{m}} \end{pmatrix}. \tag{4.22}$$

With

$$\Delta m^2_{\mathrm{m}} = \Delta m^2 \sqrt{\left(\cos 2\theta - A\right)^2 + \sin^2 2\theta}, \tag{4.23}$$

and

$$\sin 2\theta_{\mathrm{m}} = \frac{\sin 2\theta}{\sqrt{\left(\cos 2\theta - A\right)^2 + \sin^2 2\theta}}. \tag{4.24}$$

This formulation will preserve the oscillation pro ability form as in Eq. (4.13) but with the effective parameters

$$P(\nu_e \longrightarrow \nu_\mu) = \sin^2(2\theta_{\mathrm{m}}) \sin^2 \left(1.27 \Delta m^2_{\mathrm{m}21}(\mathrm{eV}^2) \frac{\mathrm{L(km)}}{\mathrm{E(GeV)}}\right). \tag{4.25}$$

From Eqs. (4.14), (4.22) and (4.25), we notice that the oscillation probabilities for neutrinos and anti-neutrinos are different due to the matter effect, since the matter potential will change sign. This is observed even if $\delta_{cp} = 0$ and even if neutrino interactions with matter do not violate CP.

Finally, it is worth mentioning that the matter effect is very small in case of the T2K experiment, as the baseline length is not very long (295 km).

4.2 Neutrino Masses

Neutrinos may have a Dirac and/or a Majorana mass terms in the interaction Lagrangian. The Dirac term is responsible for annihilating a neutrino (anti-neutrino) and creating a neutrino (anti-neutrino), which conserves lepton number. On the other hand, the Majorana term is responsible for annihilating a neutrino and creating an anti-neutrino or vice versa, so it does not conserve lepton number.

Lepton number L is conserved by SM coupling of neutrino to all other particles, so neutrino oscillation, in which L is not conserved, may be explained by the existence of a Majorana mass term.

To construct a Dirac mass term, the Left-Handed (LH) neutrino field is coupled with the Right-Handed (RH) anti-neutrino field in the Lagrangian as

$$\mathscr{L}_D = -m_D \overline{\nu}_R \nu_L + h.c. \tag{4.26}$$

While to construct the Majorana mass term, the LH and RH neutrino fields are not coupled together. The Lagrangian will take the following form

$$\mathscr{L}_{M_L} = -\frac{m_L}{2} \overline{\nu}_L \nu_L + h.c., \text{ and} \tag{4.27}$$

$$\mathscr{L}_{M_R} = -\frac{m_R}{2} \overline{\nu}_R \nu_R + h.c., \tag{4.28}$$

with mass parameters m_L, m_R and m_D.

The Majorana mass term is allowable for neutrinos since they are neutral particles, but it is forbidden for other leptons and quarks due to charge conservation. The Majorana mass term, if it exists, can be used to explain the tiny masses of neutrinos compared to the other particles, through the see-saw mechanism.

In the see-saw mechanism, we assume that neutrinos have both Dirac and Majorana mass terms and that $m_L = 0$, cf. [55]. So the part of the Lagrangian corresponding to the neutrino masses is

$$\mathscr{L}_{m\nu} = -m_D \overline{\nu}_R \nu_L - \frac{m_R}{2} \overline{\nu}_R \nu_R + h.c. \tag{4.29}$$

Which can be arranged in a matrix form as

$$\mathscr{L}_{m\nu} = -\frac{1}{2} \begin{pmatrix} \overline{\nu}_L & \overline{\nu}_R \end{pmatrix} \begin{pmatrix} 0 & m_D \\ m_D & m_R \end{pmatrix} \begin{pmatrix} \nu_L \\ \nu_R \end{pmatrix}. \tag{4.30}$$

Diagonalizing the neutrino mass matrix under the assumption that $m_R \gg m_D$, we can get the neutrino mass eigenvalues corresponding to the mass eigenstates. In a simplified single neutrino flavour model, we have

$$\begin{aligned} m_1 &\simeq \frac{m_D^2}{m_R}, \text{ and} \\ m_2 &\simeq m_R. \end{aligned} \tag{4.31}$$

Where m_1 is the light neutrino mass eigenvalue and m_R is the mass eigenvalue of the hypothetical RH heavy neutrino. If the see-saw mechanism is correct m_R should be in the range of 10^{15} GeV, which will result in m_1 in the range of 10^{-2} eV, as expected by neutrino experiments.

The see-saw mechanism is based on the assumption that neutrinos are Majorana particles, and current experiments looking for neutrino-less double beta decay ($0\nu\beta\beta$) are underway

to verify this hypothesis. In $(0\nu\beta\beta)$, two neutrons transform into two protons, ${}^{A}_{Z}X \longrightarrow {}^{A}_{Z+2}Y + 2e^{-}$ via the exchange of a Majorana neutrino. The exchange particle is created as an anti-neutrino and absorbed as a neutrino, as illustrated by Fig. 4.3.

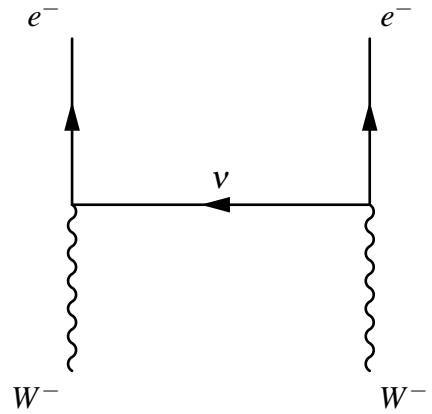


Fig. 4.3 $0\nu\beta\beta$ decay. In this diagram, time flows from bottom to top.

Any non-zero level of occurrence of this process will prove that neutrinos are Majorana particles.

4.3 Sterile Neutrinos

Sterile neutrinos, if they exist, may be heavy right-handed neutrinos. The proposed heavy right-handed neutrino has not been observed yet and does not couple to the SM particles via the weak interaction.

Experimentally, in 1995, the Liquid Scintillator Neutrino Detector (LSND) experiment [54] found some clues, $> 3\sigma$, of rapid, i.e. at small (L/E), oscillation from $\overline{\nu}_{\mu}$ to $\overline{\nu}_{e}$ corresponding to a Δm^2 of around 1 eV^2. The existence of this mass splitting requires at least one new neutrino flavour. This neutrino does not couple to the W and Z bosons in the SM.

This result has also been confirmed by The MiniBooNE experiment [54] in 2012.

The exact number of sterile neutrinos is still unknown. In a $3+1$ model, there are 3 active neutrino flavours and 1 sterile neutrino. The corresponding mixing matrix is given by

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \\ \nu_s \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} & U_{e4} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} & U_{\mu 4} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} & U_{\tau 4} \\ U_{s1} & U_{s2} & U_{s3} & U_{s4} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix}. \tag{4.32}$$

The $3+1$ model has two assumptions: $|U_{\tau 4}|$ is negligible, and that ν_4 is much bigger than the other masses. Applying the short baseline (SBL) approximation $\Delta m^2_{21} \simeq \Delta m^2_{32} \simeq 0$, we can use the appearance probability approximation in Eq. (4.13) with

$$\sin^2(2\theta_{e\mu}) = 4U^2_{e4}U^2_{\mu 4}. \tag{4.33}$$

Similarly, the disappearance probability is the complement of the appearance probability, with

$$\sin^2(2\theta_{\mu\mu}) = 4U^2_{\mu 4}(1-U^2_{\mu 4}), \text{ and} \tag{4.34}$$

$$\sin^2(2\theta_{ee}) = 4U^2_{e4}(1-U^2_{e4}). \tag{4.35}$$

To test the 3+1 model, two data sets are used. The first data set is the appearance and the other is the disappearance. Each of the data sets separately fits for the oscillation parameters, U_{e4}, $U_{\mu 4}$ and Δm^2_{41}. The two separate fits should give the same results. LSND and MiniBooNE data fit this model poorly, motivating the introduction of another sterile neutrino, creating a 3+2 model.

The 3+2 model has a larger number of oscillation parameters, including a fifth mass eigenstate and a complex phase for Charge Parity (CP) violation. The two experiments'

data fit the new 3+2 model better. A 3+3 model with three sterile neutrino states was also proposed but did not improve the quality of the fit of the data.

4.4 Impact of the Neutrino-Nucleus Interaction on the Oscillation Analyses

The oscillation probability, $P(\nu_{\alpha\to\nu_\beta})$, depends on the true neutrino energy which, in turn, depends on the model of interaction. The amplitude of the oscillation determines the mixing angles, θ_{13} and θ_{23}, as can be seen from Eq. (4.12) and depicted in Fig. 4.4.

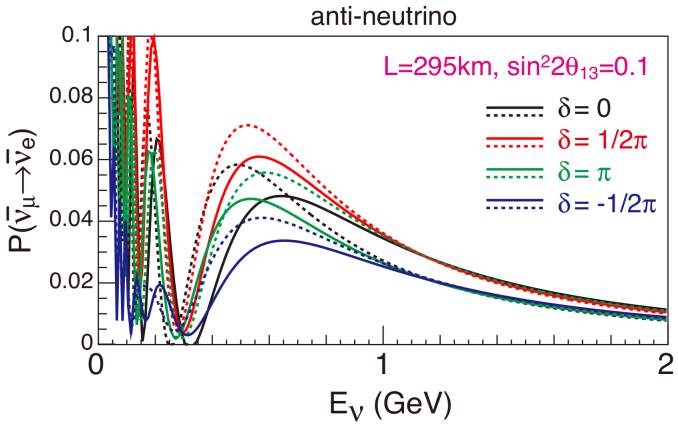


Fig. 4.4 Oscillation probability of anti-neutrino as a function of its energy, for different values of δ_{cp}. Normal and inverted hierarchies are plotted in solid and dotted lines, respectively. Figure courtesy of the T2K collaboration [56].

On the other hand, cross-section systematic uncertainties affect the number of observed events and hence the oscillation amplitude. Precise measurement of the neutrino cross-section at the near detector constrains the systematic uncertainties of the oscillation parameters. The theoretical prediction of the event rate at the far detector due to neutrino oscillation from flavour α to β ($N^{FD}_{\alpha\to\beta}$) can be expressed as follows

$$N^{FD}_{\alpha\to\beta} = \sum_i \phi_\alpha(E_{true}) P_{\alpha\beta}(E_{true}) \sigma^i_\beta(P_{true}) \varepsilon_\beta(P_{true}) R_i(P_{true}; P_{rec}), \tag{4.36}$$

where $\sum_i$ is the sum over all interaction channels, $\phi_\alpha(E_{true})$ is the flux as a function of the true neutrino's energy, $P_{\alpha\beta}(E_{true})$ is the probability of oscillation from flavour α to β, $\sigma^i_\beta(P_{true})$ is the cross-section of the neutrino of flavour β, $\varepsilon_\beta(P_{true})$ is the detector efficiency for charged lepton of flavour β and $R_i(P_{true};P_{rec})$ is the probability of reconstructing the true momentum (P_{true}) from the measured one (P_{rec}).

We use the near detector event rates, (N^{ND}), to measure the unoscillated rate of interaction:

$$N^{ND} = \sum_i \phi_\alpha(E_{true})\sigma^i_\alpha(P_{true})\varepsilon_\alpha(P_{true})R_i(P_{true};P_{rec}). \tag{4.37}$$

The flux, cross-section and detector efficiency are highly correlated between near and far detectors. This correlation is used to constrain the oscillation parameters affecting the far detector event rates (cf. [57]). This goal is achieved by introducing a penalty term to the negative log likelihood ($-\ln(L)$), containing the deviation of the flux (ϕ), cross-section (σ) and efficiency (ε) parameters from their expected values, $\Delta(\phi,\sigma,\varepsilon)$, along with their covariance matrix (V_{ij}). The negative log likelihood to be minimized in the fitting process is

$$-\ln(L) = \sum_i^{\text{Nb Bins}} N_i^p(\phi,\sigma,\varepsilon) - N_i^d + N_i^d ln\frac{N_i^d}{N_i^p(\phi,\sigma,\varepsilon)} + \sum_{i,j}\Delta(\phi,\sigma,\varepsilon)_i V_{ij}^{-1}\Delta(\phi,\sigma,\varepsilon)_j, \tag{4.38}$$

where N_i^p and N_i^d are the predicted and measured number of events in bin i, respectively.

The neutrino's true energy is estimated from the outgoing particle kinematics, leptons and hadrons, using a certain cross-section model that encapsulates the nuclear effects of the target. This procedure suffers from systematic uncertainties coming from: a broad neutrino beam energy, covering more than one interaction channel, unknown four-momentum transfer to the nucleus, final state interactions and detector resolution affecting the reconstructed measured energy.

For the Charged-Current Quasi-Elastic (CCQE) interaction channel, that is one nucleon knockout, the estimated neutrino energy, $E_\nu^{estimated}$, is

$$E_\nu^{estimated} = \frac{2(M-\varepsilon)E_l + M^2 - (M-\varepsilon)^2 - m_l^2}{2(M-\varepsilon - E_l + |K_l| cos\theta)}. \tag{4.39}$$

Here M is the nucleon mass, ε is the separation energy, E_l is the lepton energy, $|K_l|$ is the lepton three momentum, m_l is the lepton mass and θ is the angle between the outgoing lepton and the neutrino beam. In the above equation, we have neglected the recoil of the system and assumed an average separation energy for all nucleons.

More generally, if there are hadronic products at the final state other than just one nucleon, applying conservation of energy, the above equation is modified to

$$E_\nu + AM = E_l + (A-n)M + E_{\text{ex}} + T_{A-n} + \sum_i^n E_{N_i} + \sum_j^m E_{h_j}. \tag{4.40}$$

Here A is the target atomic number, n is the number of knocked out nucleons, E_{ex} is the nucleus excitation, T_{A-n} is kinetic energy of the resultant nucleus, $\sum_i^n E_{N_i}$ is the sum of the nucleon energies and $\sum_j^m E_{h_j}$ is the sum of meson energies. Some practical difficulties arise, as neutrons and neutral mesons may escape detection resulting in an underestimated neutrino energy.

Another source of systematic uncertainty for the oscillation analysis in Eqs. (4.36) and (4.37) is the detector efficiency. This quantity is simulated using event generators, which may have different particle-detector interaction models, and different particle multiplicities.

4.5 Open Questions in Neutrino Physics

Observation of neutrino oscillation infers the discovery of neutrino masses and that lepton flavour is not conserved. The study of neutrinos is a rich area of research in particle physics

that still has a lot of open questions, including but not limited to: the total number of neutrinos, the scale of neutrino masses, the mass hierarchy, Majorana vs. Dirac mass and the status of CP symmetry in the lepton sector. Currently running and future neutrino experiments will try to answer such questions.

Chapter 5

The T2K Experiment

Tokai to Kamiokande (T2K) is a long-baseline neutrino oscillation experiment in Japan. It targets the measurement of the mixing angle between the first and the third neutrino mass eigenstates (θ_{13}), as well as a precision measurement of the mass difference between the second and the third neutrino mass eigenstates (Δm^2_{32}) and their mixing angle (θ_{23}). T2K can also probe anti-neutrino oscillation by looking for the appearance of $\overline{\nu}_e$ in a $\overline{\nu}_\mu$ beam.

The experiment uses two detectors: a near detector at 280 m from the neutrino production target in Tokai, called ND280, and a far detector at 295 km, in Kamiokande, called Super-Kamiokande (SK). ND280 consists of one on-axis detector, INGRID, and one off-axis detector. The off-axis detector is a complex system that includes a Pi-Zero Detector (PØD), two Fine Grained Detectors (FGDs), three Time Projection Chambers (TPCs), a Side Muon Range Detector (SMRD) and Electromagnetic Calorimeters (ECALs). The far detector, SK, consists of 50 ktons water Cherenkov detector [58]. A diagrammatic overview of the T2K experiment setup is shown in Fig. 5.1.

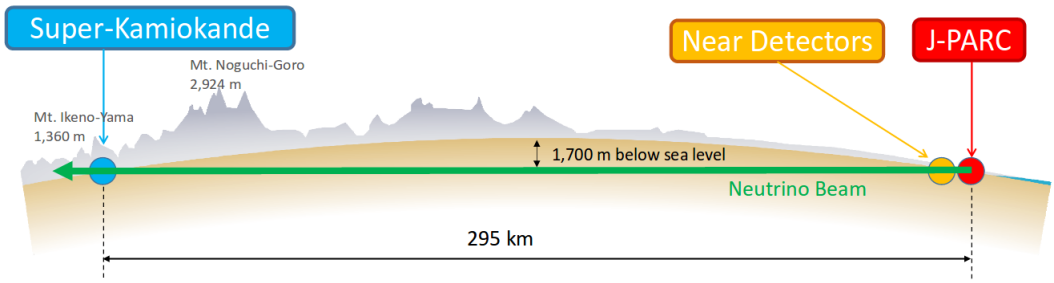


Fig. 5.1 A diagram showing the T2K experiment setup with the near detector in Tokai and the far one in Kamiokande. Figure courtesy of the T2K collaboration [56].

This chapter will provide an overview on the main components of the T2K experiment, with an emphasis on the off axis near detector tracker (TPC, FGD and ECAL).

5.1 Neutrino Beam Production

5.1.1 Proton Accelerator

The neutrinos for T2K are produced by the 30 GeV proton beam generated by the J-PARC accelerator complex [58]. J-PARC consists of three successive accelerators: the linear accelerator (LINAC), the rapid-cycling synchrotron (RCS) and the main ring synchrotron (MR), shown in Fig. 5.2.

The LINAC is a 300 m linear accelerator that accelerates hydrogen anions, H^-, to 181 MeV. Stripping foils remove the electrons to produce H^+ ions (protons). The RCS accelerates the protons to 3 GeV, then injects them to the main ring. The MR accelerates the protons to 30 GeV (up to 50 GeV).

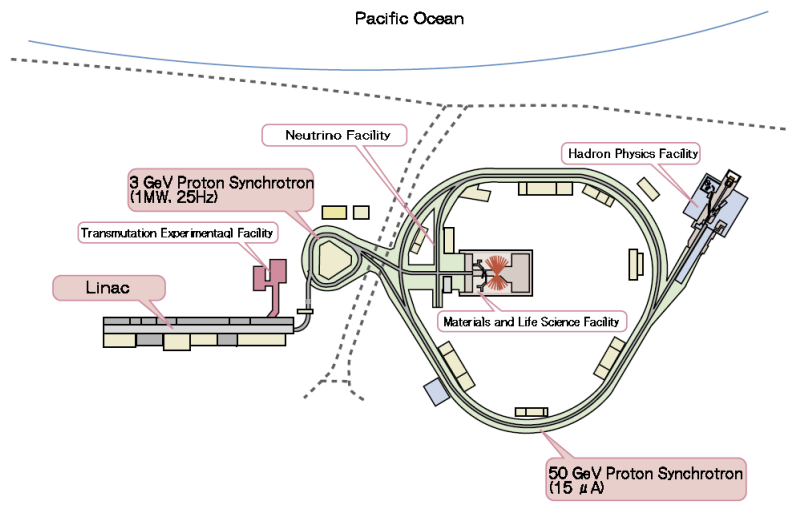


Fig. 5.2 A diagram showing the J-PARC accelerator facility in Tokai, with the tree successive accelerators: LINAC, RCS and MR. The figure is taken from the J-PARC website [59].

The proton beam consists of eight bunches per spill. The spill cycle is $\sim$ 0.5 Hz and each spill lasts less than 5 μs. The time difference between two consecutive bunches is $\sim$ 600 ns and each bunch lasts $\sim$ 60 ns. This time structure of the beam plays an important role in background rejection, such as cosmic ray events at the near and far detectors.

5.1.2 Neutrino Production

The proton beam strikes a 91.4 cm long $\times$ 2.6 cm diameter graphite target, producing hadrons, mainly pions and kaons, as shown in Figs. 5.5a and 5.5c. The selected positively (negatively) charged particles, in case of neutrino (anti-neutrino) mode, are then focused by a series of three magnetic horns, powered by 250-320 kA current pulses. The focused mesons then enter a 96 m long decay volume. The main decay branches producing neutrinos, in the anti-neutrino mode, are shown in Table 5.1.

Decay channel	Branching ratio
$\pi^- \longrightarrow \mu^- + \overline{\nu_\mu}$	$\sim 99.98\%$
$K^- \longrightarrow \mu^- + \overline{\nu_\mu}$	$\sim 63.55\%$
$K^- \longrightarrow \pi^- + \pi^0$	$\sim 20.66\%$
$K^- \longrightarrow \pi^- + \pi^- + \pi^+$	$\sim 5.59\%$
$K^- \longrightarrow e^- + \overline{\nu_e} + \pi^0$	$\sim 5.07\%$
$K^- \longrightarrow \mu^- + \overline{\nu_\mu} + \pi^0$	$\sim 3.35\%$
$K^- \longrightarrow \pi^- + \pi^0 + \pi^0$	$\sim 1.76\%$
$\mu^- \longrightarrow e^- + \overline{\nu_e} + \nu_\mu$	$\sim 100\%$

Table 5.1 Main decay channels in the anti-neutrino mode [8].

A diagram of the beam production process, highlighting the main stages is shown in Fig. 5.3.

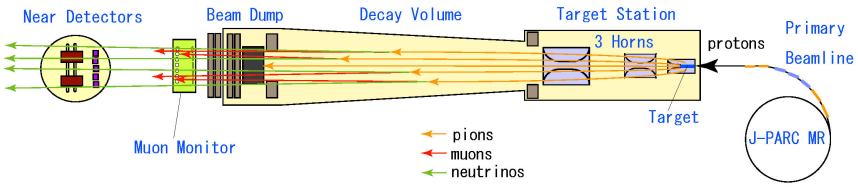
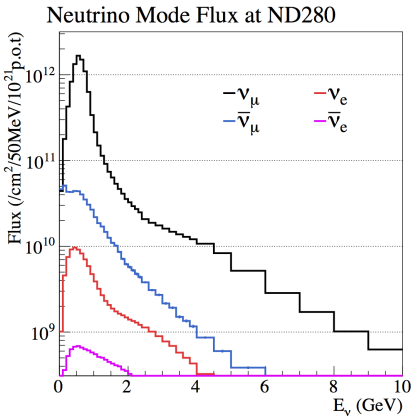


Fig. 5.3 A diagram showing the neutrino beam production with its main composites. Figure courtesy of the T2K collaboration [56].

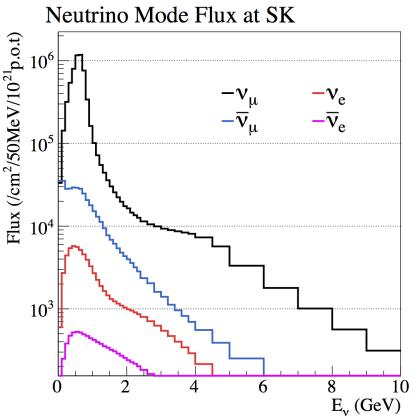
Most of the surviving charged particles are stopped by the beam dump at the end of the decay volume. Only muons (with an energy greater than 5 GeV) and neutrinos can pass through the beam dump. After the beam dump, a muon monitor measures the direction and the flux of the high energy muons. The muon detector helps to characterize the neutrino beam stability, intensity, and direction with high precision (0.25 mrad). Finally, the neutrino beam traverses through the ND280 and reaches SK.

The alignment between the target and the detectors, and the time synchronization between the beam spills and the detectors are performed using a Global Position System (GPS).

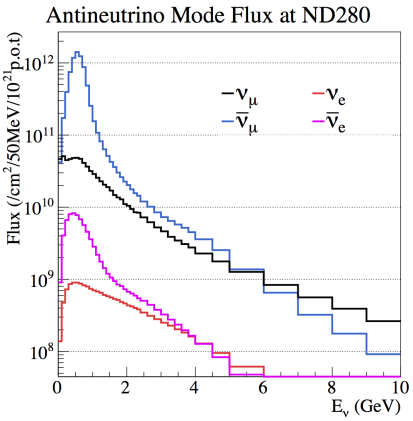
The flux prediction at both the near and far detectors, along with its different components is shown in Fig. 5.4 .



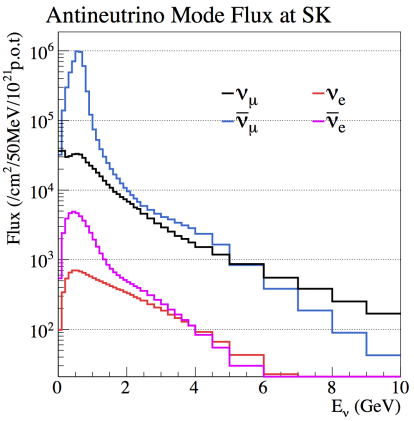
(a) The ND280 (FGD) flux prediction for Runs 1-7c with horns operating in 250 kA mode.



(b) The SK unoscillated flux prediction for Runs 1-7c with horns operating in 250 kA mode.



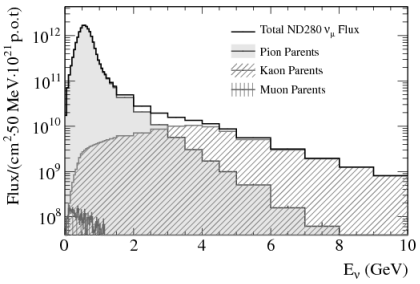
(c) The ND280 (FGD) flux prediction for Run 5c-7b with horns operating in -250 kA mode.



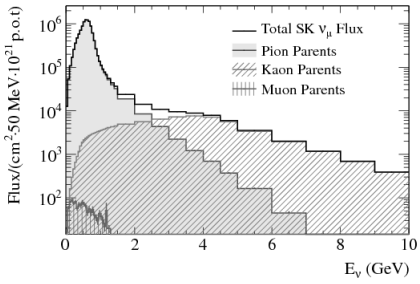
(d) The SK unoscillated flux prediction for Run 5c-7b with horns operating in -250 kA mode.

Fig. 5.4 Plots of neutrino flux prediction at the near detector (top left) and at the far detector (top right), and that of the anti-neutrino at the near detector (bottom left) and at the far detector (bottom right). Flux is normalized to 1e21 POT. Figure courtesy of the T2K flux official plots (2016) [56].

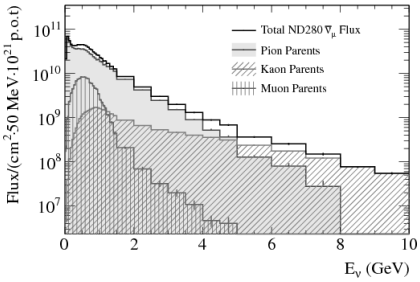
The flux parents decomposition for the muon neutrino and muon anti-neutrino mode at both the near and far detector are shown in Fig. 5.5, while that of electron neutrino and electron anti-neutrino are shown in Fig. 5.6.



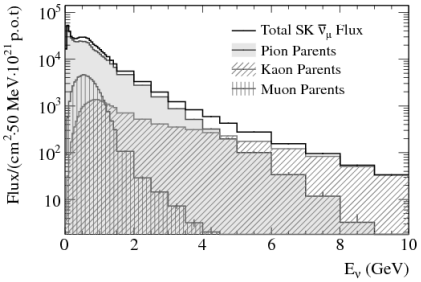
(a) ν_μ flux parents at ND280



(b) ν_μ flux parents at SK

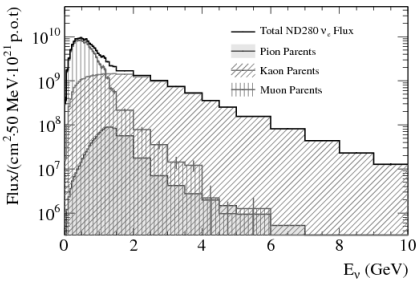


(c) $\overline{\nu}_\mu$ flux parents at ND280

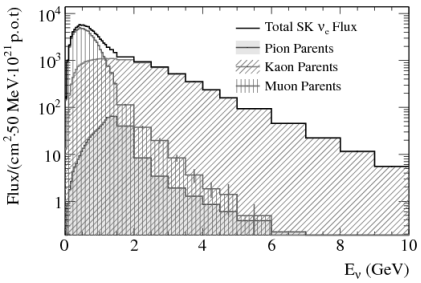


(d) $\overline{\nu}_\mu$ flux parents at SK

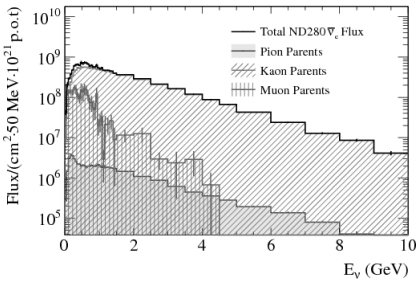
Fig. 5.5 Plots of muon neutrino flux parent decomposition at the near detector (top left) and at the far detector (top right), and that of the muon anti-neutrino at the near detector (bottom left) and at the far detector (bottom right). The figure is taken from [60].



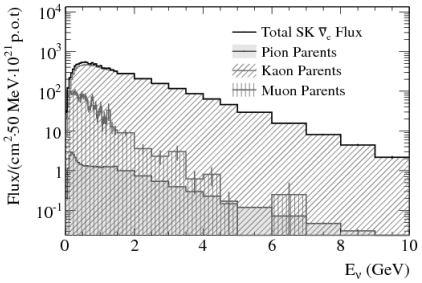
(a) ν_e flux parents at ND280



(b) ν_e flux parents at SK



(c) $\overline{\nu}_e$ flux parents at ND280



(d) $\overline{\nu}_e$ flux parents at SK

Fig. 5.6 Plots of electron neutrino flux parent decomposition at the near detector (top left) and at the far detector (top right), and that of the electron anti-neutrino at the near detector (bottom left) and at the far detector (bottom right). The figure is taken from [60].

5.2 On-axis Near Detector (INGRID)

The Interactive Neutrino Grid (INGRID) is located on the neutrino beam axis, as shown in Fig. 5.7a. It measures beam profile, such as beam intensity and direction. INGRID consists of 14 identical modules arranged as a cross and two modules located outside the main cross, as shown in Fig. 5.7b.

(a) INGRID detector overview. (b) INGRID modules arrangement.

Fig. 5.7 Overview of the INGRID detector located on the neutrino beam axis (left), along with its 16 modules constituents (right). Figure courtesy of the T2K collaboration [56].

The centre of the INGRID cross is made of two overlapping modules and coincides with the centre of the neutrino beam, while the off-axis modules are used to check the neutrino beam cylindrical symmetry.

Each module consists of nine iron plates interleaved with 11 tracking scintillator planes surrounded by veto scintillator planes, to reject interactions outside the module, as shown in Fig. 5.8. The tracking planes are made of 24 scintillator bars in the horizontal direction alternating 24 vertical scintillator bars. This design allows reconstruction of a three-dimensional track of the path of the charged particle.

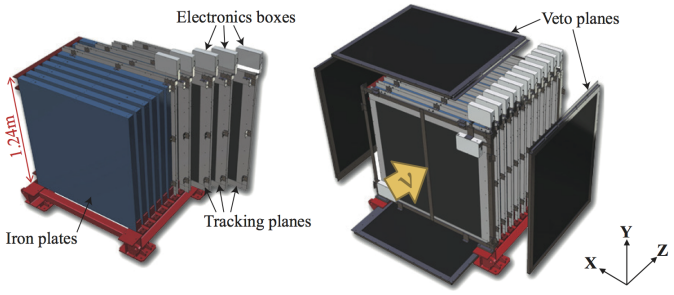


Fig. 5.8 The INGRID module structure, showing the iron plates interleaved with the scintillator planes. Figure courtesy of the T2K collaboration [56].

When a neutrino interacts with the iron through charged-current, it creates a charged lepton of the same neutrino flavour. This charged lepton, e.g μ^- in case of ν_μ beam, will

cause the scintillator to produce light that is guided to the end of each scintillator bar through a wavelength-shifting (WLS) fibre, and detected using a Multi-Pixel Photon Counter (MPPC).

5.3 Off-axis Near Detector (ND280)

The neutrinos in the T2K beam are mainly produced from two-body pion decay, e.g $\pi^+ \rightarrow \mu^+ + \nu_\mu$. The resultant neutrino energy spectrum has a strong angular dependence, as shown by Eq. (5.1)

$$E_\nu = \frac{m_\pi^2 - m_\mu^2}{2(E_\pi - p_\pi \cos\theta)}. \tag{5.1}$$

Where E_ν is the neutrino energy, E_π is the pion energy, m_π is the pion mass, m_μ is the muon mass and θ is the small off-axis angle in rad. The proof of Eq. (5.1) is given in Appendix A.

This feature is used in T2K by placing the detectors (off-axis ND280 and SK) at a 2.5° angle relative to the axis for the beam, to achieve a narrow-band neutrino beam. This setup creates a neutrino energy peak around 0.6 GeV which has two main advantages. First, it is close to the first predicted oscillation maximum at SK, as shown in Fig. 5.9, secondly it reduces the background from non-CCQE processes at both the near and far detectors.

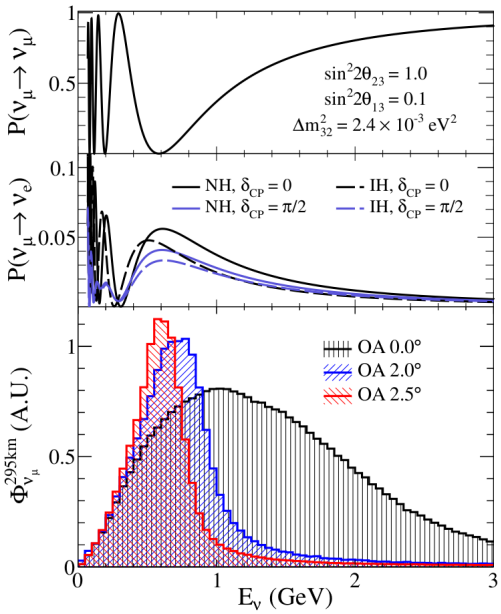


Fig. 5.9 Neutrino energy flux at different off-axis angles (bottom). The 2.5° off-axis angle is chosen in order to tune the maximum neutrino peak around 0.6 GeV where the maximum oscillation probability $P(\nu_\mu \to \nu_e)$ (midlle) coincides with the minimum of $P(\nu_\mu \to \nu_\mu)$ (top). The appearance probability is shown for two values of the CP phase ($\delta_{CP} = 0$ and $\frac{\pi}{2}$), and for normal (NH) and inverted (IH) mass hierarchies. The figure is taken from [57].

The main roles of the off-axis detector are to:

- provide information to determine the ν_μ (neutrino mode) or $\overline{\nu}_\mu$ (anti-neutrino mode) flux;
- measure the ν_e (neutrino mode), $\overline{\nu}_e$ (anti-neutrino mode) component, which represent an intrinsic, non removable, background to the oscillation analysis;
- measure the neutral current (NC1π^0) interaction, which represents the dominant background to the ν_e appearance analysis in SK.

To achieve these goals, the off-axis near detector is built from several sub-detectors, including a Pi-Zero Detector (PØD), two Fine Grained Detectors (FGDs), three Time Projection Chambers (TPCs), a Side Muon Range Detector (SMRD) and Electromagnetic Calorimeters (ECALs) as shown in Fig. 5.10.

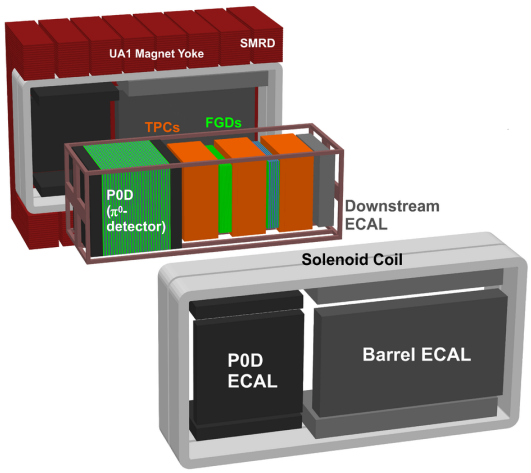


Fig. 5.10 ND280 off-axis detector complex showing the different sub-detectors. The neutrino beam is coming from the left going to the right. Figure courtesy of the T2K collaboration [56].

The design and functionality of these different sub-detectors will be briefly discussed next.

5.3.1 Pi-Zero Detector (PØD)

The main objective of the PØD is to measure the neutral current process on a water target. Understanding the neutrino neutral current interaction (NC1π^0) as well as the charged-current π^0 production ($CC1\pi^0$) cross section is crucial for the analysis. After all, SK cannot distinguish between photons, that may be generated from a π^0, and electrons, as will be shown in Section 5.5.

The PØD consists of 40 scintillator modules (PØDules). Each PØDules has 134 vertical and 126 horizontal scintillator bars. The planes of scintillator bars are interleaved with fillable target water bags and lead and brass sheets. The neutrino cross-section on water is extracted using a subtraction method; where the cross-section in the case of empty bags is subtracted from that of the filled bags. The segmentation of the scintillator bars helps to reconstruct

charged particle tracks and electromagnetic showers that originated from the electrons and photons. Each scintillator bar is read out with a single WLS fibre, that is mirrored at one end and has an MPPC at the other.

The central part of the PØD consists of an "upstream water target" and a "central water target". Each of the targets has 13 P0Dules. The upstream water target has 13 water bag layers and 13 brass sheets, whereas the central water target has 12 water bags and 12 brass sheets. The front and rear sections of the PØD are made up of electromagnetic calorimeters, also called the "upstream ECAL" and "central ECAL" (not to be confused with the tracker ECAL in Section 5.3.4). The PØD's upstream and central ECALs have alternating scintillator planes and lead sheets. The structure of the PØD is shown in Fig. 5.11.

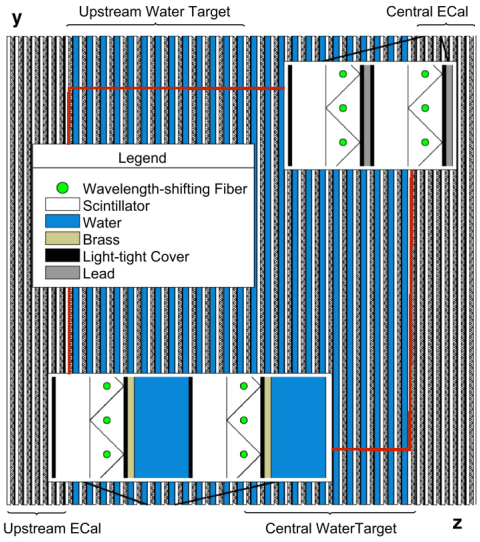


Fig. 5.11 A schematic of the pi-zero detector. The beam is coming from the left and going right. The figure is taken from [61].

This design enhances the containment of electromagnetic showers and provides a veto region before and after the water target to reject events originating from interactions outside the PØD.

5.3.2 Time Projection Chambers (TPC)

There are three TPCs in ND280, arranged as shown in Fig. 5.10. The main uses of the TPCs are:

- to detect the number and orientation of the charged particles generated from neutrino interactions;
- to measure the momenta of the charged particles, with the help of the applied magnetic field; and
- to act as particle identification, by comparing the amount of ionization left by each particle, as seen in Fig. 5.12.

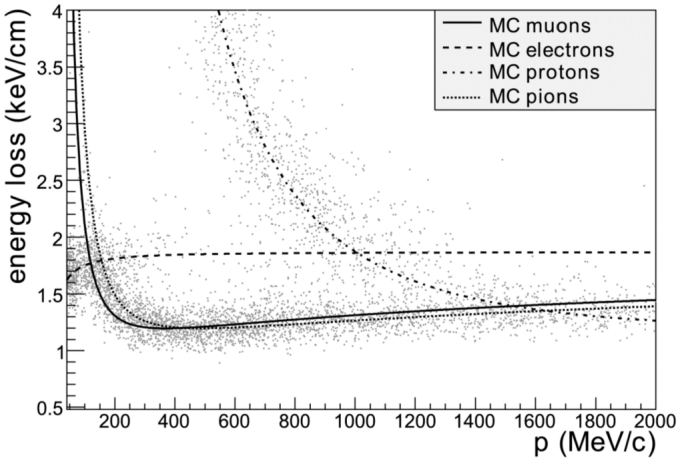


Fig. 5.12 Energy loss as a function of the particle's momentum. Every point represents a measurement of the energy loss and momentum of positively charged particles produced in neutrino interactions. Solid lines represent the expected relationships for muons, positrons, protons. Similar plots exist for negatively charged particles. The figure is taken from [58].

Each TPC consists of an inner box filled with an argon-based drift gas (95% Ar, 3% CF_4 and 2% C_4H_{10}), contained within an outer box filled with an insulating gas (CO_2). A central cathode, connected to $\sim$ -25kV, divides the TPC into two identical drift volumes. The TPC inner walls are covered with copper strips to form the field cage. Each strip is connected to a

gradually higher potential, producing a uniform electric drift field of about 200 V/cm. This electric field is in the same direction as the 0.2 T magnetic field provided by the magnet.

Each side of the TPC is instrumented with 12 Micro Mesh Gaseous Structures, called Micromegas modules (MM), which amplify the signal. Each MM module consists of 1728 pads (7.0 mm $\times$ 9.8 mm) in 48 rows and 36 columns in the (yz) plane; the x-coordinate is determined by measuring the arrival time. Combining the pad pattern with the arrival time gives a complete three-dimensional image of the paths of the traversing charged particles.

The electrical signal from each MM is then sampled and digitized using six front-end electronics (FEC) cards. Each FEC has four custom ASICs, called "AFTER", that shape the signals and buffer 72 pad signals into 511 time-bin switched capacitor arrays (SCA). Each set of six front-end cards is connected to a single front-end mezzanine card (FEM) that collects the data and performs zero suppression. Every four FEMs are connected to one Data Concentration Card (DCC) that collects the data from the FEMs over an optical transceiver and sends the data over Ethernet through a switch to the TPC DAQ computer.

A simple structure of the TPC is shown in Fig. 5.13.

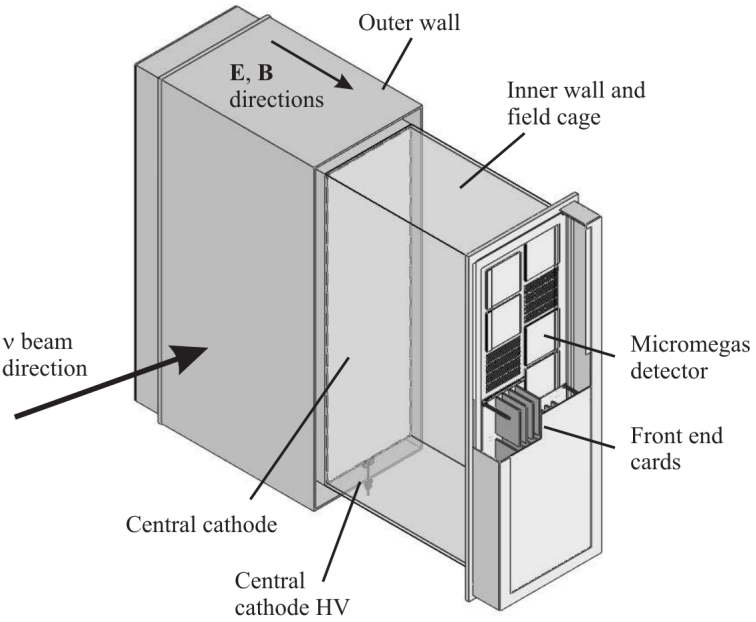


Fig. 5.13 A simple TPC structure. The figure is taken from [58].

The TPC is calibrated using a laser system that shines 266 nm light pulses on thin aluminium dots and strips glued to the copper surface of the cathode. Photoelectrons are then drifted under the influence of the electric field and may be used to determine the electron drift velocity and to measure distortions in the electron drift due to inhomogeneous and misaligned electric and magnetic fields.

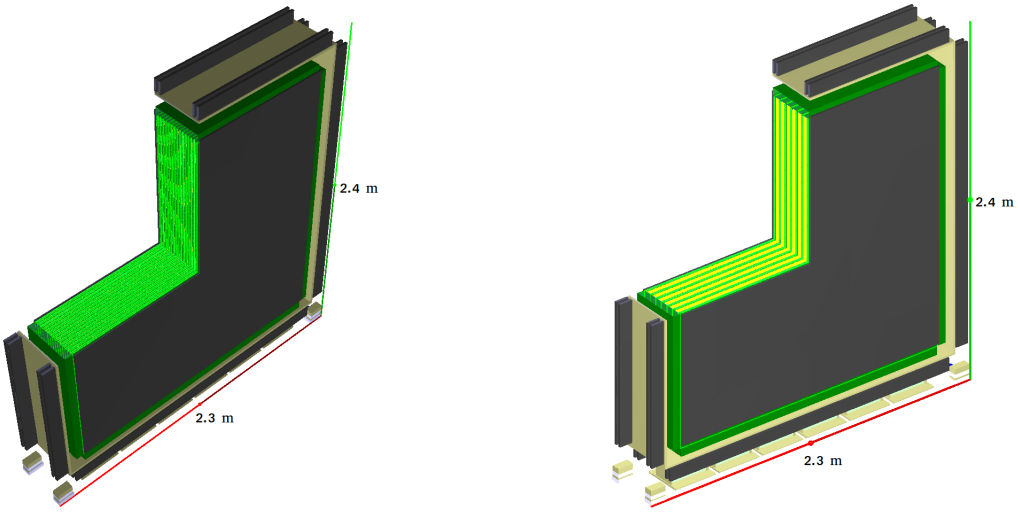
5.3.3 Fine Grained Detectors (FGD)

There are two Fine Grained Detectors (FGD) in ND280, called FGD1 and FGD2, arranged as shown in Fig. 5.10. The main functionalities of an FGD are:

- to provide target mass for neutrino interactions, and
- to track charged particles produced from the interaction vertex.

Each FGD has outer dimensions of 2300 mm (width) $\times$ 2400 mm (height) $\times$ 365 mm (depth). FGD1, consists of 30 layers of plastic scintillator bars, polystyrene $(C_8H_8)_n$, while FGD2 consists of 14 layers of plastic scintillators interleaved with six water layers. The water in FGD2 is maintained under sub-atmospheric pressure so that in case of a leak the system will intake air in rather than release water inside the FGD. The scintillator layer consists of 192 bars of 9.61 mm $\times$ 9.61 mm $\times$ 1864.3 mm. These bars are perpendicular to the beam direction, with subsequent layers alternating between x and y orientations.

Each scintillator bar has a reflective coating and a Wave Length Shifter (WLS) fibre passing through its centre. The fibre is mirrored from one end, and attached to a Multi-Pixel Photon Counter (MPPC) from another. The MPPC is then connected to the Front End Electronics, which digitize the light signal and transfer it to the rest of the FGD DAQ system. The structures of FGD1 and FGD2 are shown in Figs. 5.14a and 5.14b.



(a) FGD1 (b) FGD2

Fig. 5.14 Structure of FGD1, made of 30 layers of plastic scintillator bars (left), and that of FGD2 made of 14 layers of plastic scintillators interleaved with six water layers (right). Scintillator layers are drawn in green and water layers in yellow.

This design helps to extract neutrino cross-sections on water, which is crucial for SK, by comparing interaction rates in FGD1 and FGD2. The analysis in this thesis uses FGD1 as the active interaction target.

5.3.4 Electromagnetic Calorimeters (ECAL)

ND280 has 13 electromagnetic calorimeter modules as shown in Fig. 5.10. The ECAL surrounds the TPCs, FGDs and the PØD. It complements the inner detectors in full event reconstruction, with the following main functionalities:

- measurement of photon energy and direction;
- particle identification for the charged particles;
- reconstruction of π^0 generated inside the tracker, and determining the π^0 escape energy in case of production inside the PØD; and
- acts as a veto for external backgrounds.

The 13 ECAL modules are made of three different types and distributed as follow: six Barrel-ECAL modules surround the tracker volume, one downstream module (Ds-ECAL) at the exit of the tracker volume, and six PØD-ECAL modules that surround the PØD detector. The Ds-ECAL is located inside the basket carrying the inner sub-detectors of the off-axis detector. The other 12 ECAL modules are located inside the magnet.

Each ECAL module consists of layers of plastic scintillator bars, interleaved with lead sheets. This makes the ECAL a sampling calorimeter, i.e. the material that produces the particle shower (lead) is distinct from the material that measures the deposited energy (scintillator). One advantage of this is that the lead, a very dense material, can be used to produce a shower that evolves quickly in a limited space. A disadvantage is that the energy deposited in the lead is not measured; thus the total shower energy must be estimated.

The Barrel ECAL modules consist of 31 layers, with each layer having 50 scintillator bars, interleaved with 1.75 mm thick lead sheets. The Ds-Ecal module has 34 layers of scintillator interleaved with 1.75 mm thick lead sheets. The PØD-ECAL modules each have six scintillator layers that alternate with five 4 mm thick lead sheets. The PØD-ECAL has a coarser granularity as it only serves to provide additional energy information for the π^0 measured by the PØD.

The bars of consecutive layers are at 90° to allow three-dimensional reconstruction of charged particle tracks and electromagnetic clusters. All ECAL scintillator bars have a 4.0 cm $\times$ 1.0 cm cross-section coated with a 0.25 mm thick layer of TiO_2 providing light reflection and isolation. The light signal is transported through WLS fibre and read at one (bars running in the z-direction) or both ends (bars running in the x or y-direction) of each fibre with MPPCs. Those read out at one end only are mirrored at the other end.

5.3.5 The Magnet

The magnet used in the ND280 detector has been refurbished from the CERN UA1 [62] and NOMAD [63] experiments. It produces a 0.2 T dipole magnetic field, with a 0.1% uncertainty for each field component. This magnetic field enables the TPC to measure the momenta and sign of charged particles passing through. [1]

The magnet consists of water-cooled aluminium coils, creating the horizontally oriented dipole field, and a flux return yoke. The yoke helps to ensure a homogeneous magnetic field (as much as possible), and to reduce the field extension outside the detector. The dimensions of the magnet are 7.0 m $\times$ 3.5 m $\times$ 3.6 m and 7.6 m $\times$ 5.6 m $\times$ 6.1 m for the inner and outer structure, respectively. The total weight of the yoke is 850 tons. The magnet is symmetrically halved to facilitate the access to the sub-detectors. Each half of the yoke consists of eight individual C-shaped elements made of steel plates and standing on a movable carriage. This design allows each half of the magnet to be separately moved to an open or closed position.

5.4 Side Muon Range Detector (SMRD)

The side muon range detector (SMRD) consists of 440 scintillator modules that are inserted between the steel plates in the magnet yoke elements. Each scintillator layer is read by a WLS fibre connected to an MPPC as shown in Fig. 5.15.

[1] A charged particle of mass m and charge q moving in a magnetic field B with a constant velocity v will feel a perpendicular force of magnitude F pushing it to move in a circular path with radius r, such that: $F = qvB = \frac{mv^2}{r}$ leading to $p_t \approx 0.3Br$, where p_t is the measured transverse momentum (in GeV/c).

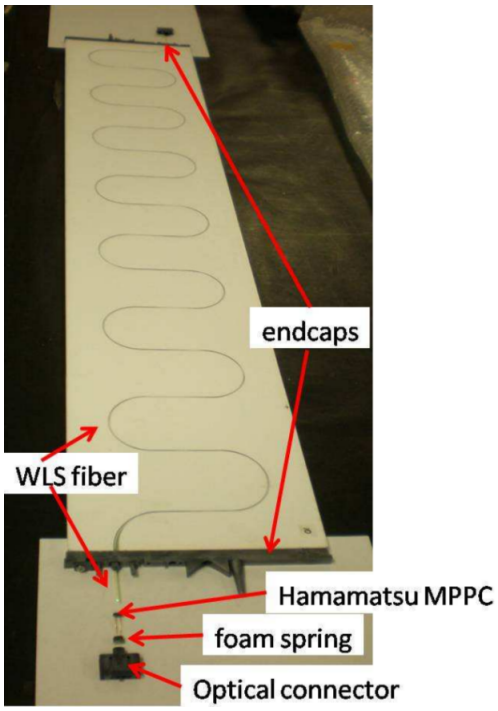


Fig. 5.15 View of SMRD scintillator layer with the WLS fibre connected to an MPPC. The figure is taken from [58].

The main functionalities of the SMRD are:

- to measure the momenta of the high-angle muons with respect to the beam direction;
- to act as cosmic ray trigger; and
- to veto neutrino events that started in the pit walls (sand muons) or in the magnet.

5.5 Super-Kamiokande Far Detector

Super Kamiokande (SK) (see Fig. 5.16) is a large Cherenkov detector, located 295 km from the beam production. It is built within a 1 km deep mine at Kamioka. At that depth, the cosmic ray flux is reduced by five orders of magnitude with respect to the Earth's surface.

SK consists of a cylindrical tank of 41.4 m height and 39.3 m diameter, filled with 50 ktons of pure water. It has an inner and an outer cylindrical detector. The inner detector is

viewed by 11146 inward facing 50 cm diameter photo-multipliers (PMT) and filled with 32 ktons of water. The outer detector surrounds the inner volume and is instrumented with 1885 outward facing PMTs to veto cosmic and out of fiducial volume events.

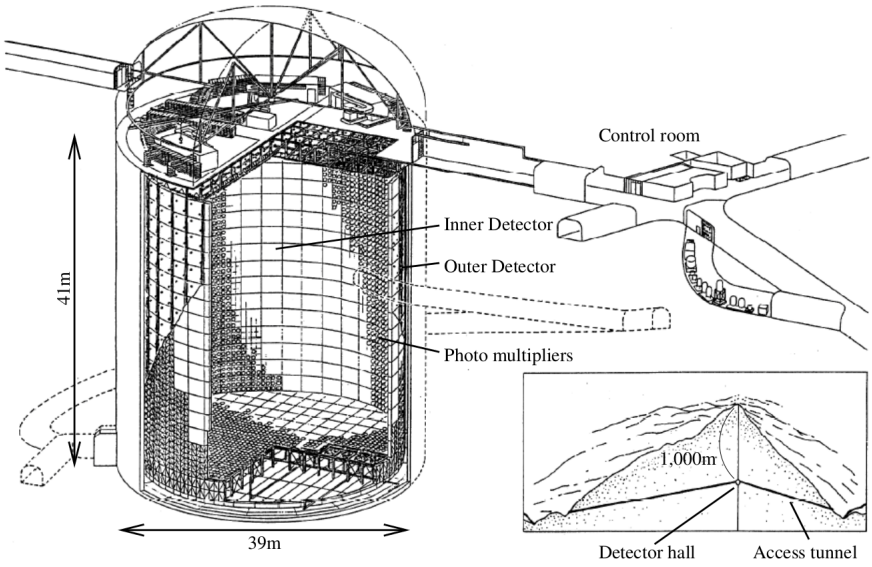


Fig. 5.16 Schematic view of the Super Kamiokande Water Cherenkov detector. The figure is taken from [64].

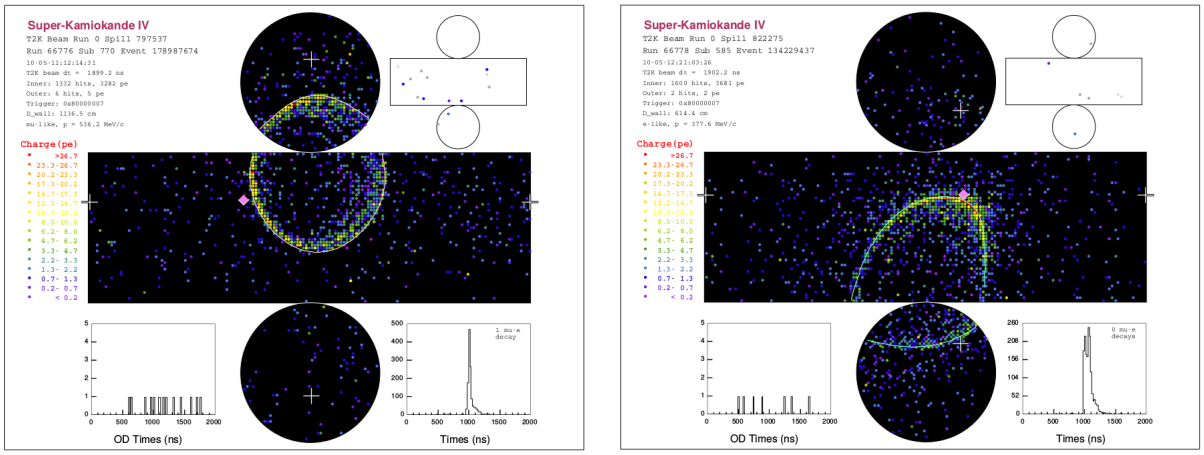
A Cherenkov detector is based on the idea that a particle passing through a material at a velocity greater than that at which light can travel through this material emits coherent light. This light is called Cherenkov radiation and is detected by the photo sensors. The emitted light forms a cone with angle θ_c with respect to the direction of the moving particle, with

$$\cos\theta_c = \frac{c}{nv}, \tag{5.2}$$

where c is the vacuum speed of light, n is the refractive index of the material, and v is the particle speed.

Energetic charged particles generated by neutrino interactions can be distinguished by examining the edges of the light ring they produce. Muons produce a single ring with well-defined edges, as shown in Fig. 5.17a, while electrons produce multiple Cherenkov

rings in slightly different directions due to the electron scattering in the water, as shown in Fig. 5.17b. The misidentification between electrons and muons is ~ 2 % in the sub-GeV energy range.



(a) Muon-like ring.

(b) Electron-like ring.

Fig. 5.17 An event display showing a muon-like ring with clear edges (Left) and an electron-like ring with fuzzy scattered edges (right). The figures are taken from [58].

It is worth noting, that Cherenkov detectors can not distinguish between electrons and photons, as they both produce electromagnetic cascades. For this reason, an electron-neutrino (anti-neutrino) $CC1\pi^0$ event can mimic a $CC0\pi$ event if the produced π^0 decays into two colinear photons, or if one photon is not detected.

Part II

Event Selection

Chapter 6

Multivariate Analysis For Particle Identification

Multivariate data analysis and machine learning are useful tools in high-energy physics (HEP). The need for more sophisticated data analysis algorithms arose with the increased complexity of the classification problems. In ND280, selecting an $\overline{\nu}_e$ interaction event is like picking a needle from a haystack due to the tiny neutrino cross-section and a large number of background events. Increasing the selection purity and efficiency is therefore crucial for a precision measurement of the neutrino cross-section.

In this thesis, a machine learning algorithm called a Boosted Decision Tree (BDT) is used as a particle identification (PID) classifier. Information gained from the ND280 tracker, TPC, FGD and ECAL, is used to discriminate the signal (positron tracks) from the background (any other particle track). Consequently, other selection steps will be applied to only select the positron coming from the $\overline{\nu}_e$ interaction, as will be described in Chapter 8.

The BDT has a successful history in particle physics. It became popular in HEP when used by MiniBooNE as an alternative to the artificial neural networks algorithm for particle identification [65]. It is simpler and easier to visualize and gives results as good as or better than most other classifiers, cf. [66, Section 9.2.4]. The analysis integrates the Toolkit for

MultiVariate Analysis (TMVA) [66] libraries with the standard T2K software libraries in the event selection process. TMVA was chosen because:

1. it is part of ROOT [67], an object-oriented framework for large-scale data analysis, commonly used in T2K and in many HEP experiments,
2. it is free and open source,
3. it includes implementations of several machine learning algorithms, e.g.: Boosted Decision Trees, Neural Network, Support Vector Machine, etc., and
4. it has good documentation and online resources.

6.1 Why Utilize a Multivariate Analysis?

Traditional cuts on individual variables are not the most powerful selection. To illustrate this idea, Fig. 6.1 shows the signal and the background distribution for two measured variables, var0 and var1, of a toy example. Using traditional cuts on var0 or var1 will result in very poor efficiency. However, visualizing the two-dimensional distribution of var0 and var1, one can find a better decision boundary to separate signal from background.

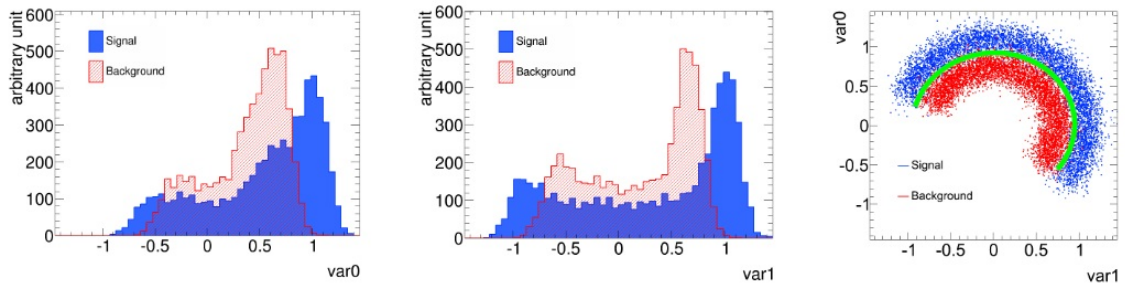


Fig. 6.1 Single and multivariate cut effects on correlated data. Signal (in blue) and background (in red) normalized probability distribution for var0 and var1 of a toy example are shown at the left and the centre plots respectively. A better decision boundary, using variable correlation, is shown (in green) in the right plot. Figure courtesy of [68].

Using the correlation among variables increases the efficiency and purity of the selection. It may be possible to visualize such relationships for two- or three-dimensional problems, yet, a computer algorithm will be needed to optimize the decision boundary for higher-dimensional feature spaces.

6.2 Event Classification

Each event, signal or background, has "D" measured variables that construct a D-dimensional feature space; for instance, the features used in the positron selection are shown in Table 7.1. A machine learning algorithm is a map from the D-dimensional feature space into a one-dimensional output

$$y(x) : \mathbb{R}^D \rightarrow \mathbb{R}, \tag{6.1}$$

where y(x) is the value of the test statistics variable evaluated by the machine.

Distributions of y(x) for both signal, $PDF_{SIG}(y)$, and background, $PDF_{BKG}(y)$, are built using a subset of the training sample. Cutting at a particular value, t_{cut}, represents a surface in the D-dimensional space defining the decision boundary, and is chosen to maximize a certain figure of merit (FoM). For example, purity $\times$ efficiency, such that:

$$y(x) \begin{cases} \geq t_{cut} \text{ (signal)} \\ < t_{cut} \text{ (background)} \end{cases} \tag{6.2}$$

6.3 Boosted Decision Trees

6.3.1 Decision Trees

A decision tree is a supervised machine learning algorithm consisting of a consecutive set of cuts. Each cut has only one of two outcomes based on which a new cut will be applied, as shown in Fig. 6.2.

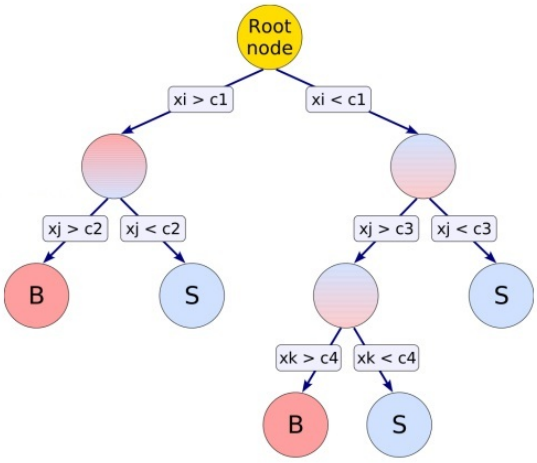


Fig. 6.2 A decision tree with depth of three levels. At each node, a branching occurs as a result of the splitting question. Final leaves are either mostly signal (in blue) or mostly background (in red). Figure courtesy of [68].

The node criterion (split variable) is chosen to maximize the separation gain between nodes, defined as:

$$\begin{aligned} \text{separation gain} = \text{gain(the parent node)} - \text{gain(first daughter node)} \\ - \text{gain(second daughter node)}, \end{aligned} \tag{6.3}$$

where the gain is computed depending on the separation type as shown in Table 6.1 and plotted in Fig. 6.3.

Separation Type	Gain Formula
GiniIndex	$p(1-p)$
CrossEntropy	$-p\ln(p)-(1-p)\ln(1-p)$
MisClassificationError	$1-\max(p,1-p)$
SDivSqrtSPlusB	$\frac{S}{\sqrt{S+B}}$

Table 6.1 Separation criterion for BDT node splitting, supported by TMVA (cf. [66, Section 8.12]), where p is the purity, S and B are the numbers of signal and background in a node respectively.

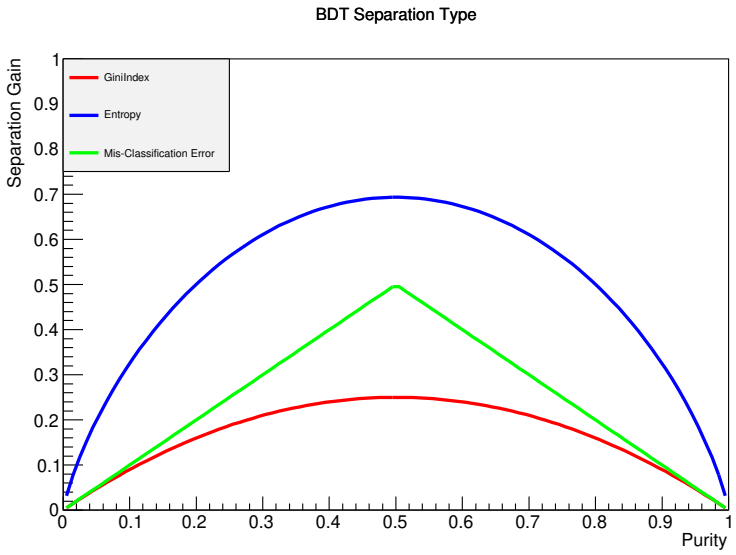


Fig. 6.3 BDT node splitting techniques based on maximizing the gain. GiniIndex, entropy and misclassification error as a function of node purity, are shown in red, blue and green curves, respectively.

Maximizing the gain results into nodes consisting of either mostly signal, or mostly background. Furthermore, background nodes are considered as important as signal nodes. Hence the separation function is symmetric around the 50% purity line. All splitting techniques give similar results. In this work, the GiniIndex, being simple and fast to compute, was chosen.

The node splitting process continues until either the maximum depth is reached, or the minimum number of events in one leaf is achieved.

Decision trees are fast learners, they are easy to understand and to interpret. However, they are vulnerable to overtraining, i.e. learning the data fluctuations. Boosting techniques can be used to overcome this problem.

6.3.2 Boosting

Boosting is a powerful technique for combining multiple classifiers to form a group whose performance is significantly better than that of any of its members alone. Every classifier could be chosen to be "weak", i.e. its performance is slightly better than a random guess, to avoid the overtraining problem. The final output is determined by the majority vote of all classifiers.

For example, if one has three uncorrelated decision trees, each one is correct for 60% of the time; the probability that the majority (at least two trees) is incorrect is given by

$$\begin{aligned} p_{\text{vote wrong}} &= \binom{3}{2} p^2_{\text{single wrong}} \times p_{\text{single right}} + \binom{3}{3} p^3_{\text{single wrong}} \\ &= 0.352 < 0.4. \end{aligned}$$

To form this forest, two common methods exist: Bootstrap Aggregation (Bagging) and boosting. In bagging: each tree is trained on a subset of the training data with replacement[1]. One problem that may arise in such a method, is that the trees choose which variable to split on using an algorithm that minimizes error (or maximizes the gain). Consequently, the decision trees can have a lot of structural similarities and become highly correlated.

[1] In a Bootstrap technique, we produce N copies of the data using random sampling with replacement, such that each copy consists of the same number of data points and points are allowed to be replicated, cf. [69, Section 2.1].

Boosting constructs the forest using a different approach. The Adaptive Boosting (AdaBoost) algorithm is briefly described here and visualized in Fig. 6.4. For more details please refer to [69, Section 14.3]. Boosting consists of the steps enumerated below.

1. Build the first classifier (tree) using the whole training data set, with equal data weighting coefficients, i.e. $w_n = \frac{1}{N}$.

2. Train a classifier y_m by minimizing the weighted error function, J_m, given by

$$J_m = \sum_{n=1}^{N} w_n^{(m)} I(y_m(x_n) \neq t_n), \tag{6.4}$$

   where $I(y_m(x_n) \neq t_n)$ is an indicator function whose value equals 1 in case of misclassification, i.e. when the output of the classifier $y_m(x_n)$ for a certain data point x_n and its target value t_n are different, and 0 otherwise. For a two class classifier, $y_m(x_n)$ and t_n are $\in \{-1, 1\}$.

3. Evaluate the average misclassified error

$$\varepsilon_m = \frac{\sum_{n=1}^{N} w_n^{(m)} I(y_m(x_n) \neq t_n)}{\sum_{n=1}^{N} w_n^{(m)}}, \tag{6.5}$$

   where N is the total number of data entries.

   Also, calculate the error power

$$\alpha_m = \beta \ln \frac{1 - \varepsilon_m}{\varepsilon_m}, \tag{6.6}$$

   where β controls the strength of the misclassification error weight and is normally set to 0.5. A larger β results in a bigger weight for the misclassified event and vice-versa.

4. Update the weights of the next classifier based on the misclassification error of the former using

$$w_n^{(m+1)} = w_n^{(m)} \exp\{\alpha_m I(y_m(x_n) \neq t_n)\}. \tag{6.7}$$

Now the misclassified events have higher weights and the correctly classified events have less weight. The next classifier, trained on the new data set, learns better the misclassified event of the previous pass.

5. Repeat the previous steps (Items 2 to 4) until reaching the required number of trees M is reached. The final output is a weighted average of all classifiers, given by

$$Y_{(M)}(x) = \text{sign}\Big(\sum_{m=1}^{M} \alpha_m y_m(x)\Big). \tag{6.8}$$

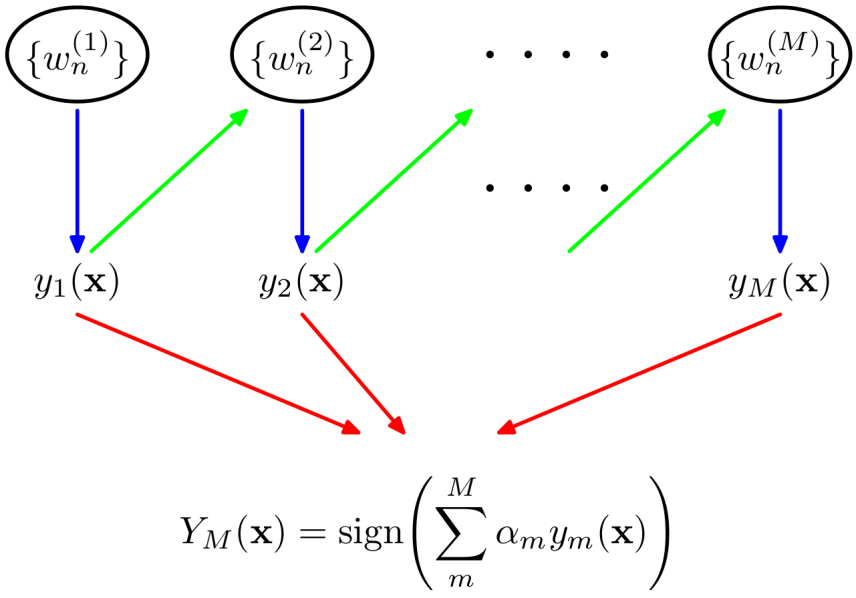


Fig. 6.4 A diagram showing the Adaboost algorithm. Figure courtesy of [69].

In order to protect the Boosted Decision trees from overtraining, either insignificant nodes are pruned or the depth of the tree is limited. In this thesis, the latter option is chosen with the maximum depth of the tree set to seven.

6.3.3 BDT Selection Workflow

The BDT selection has two major phases: a training phase and an evaluation phase, as shown in Fig. 6.5. In the training phase, a Monte Carlo (MC) simulated dataset with labels (signal and background) is used by the machine to determine the decision boundary. Here, the MC particle numbering scheme from the PDG [8] is used as a label, with the signal code of -11 for positron and anything else is background.

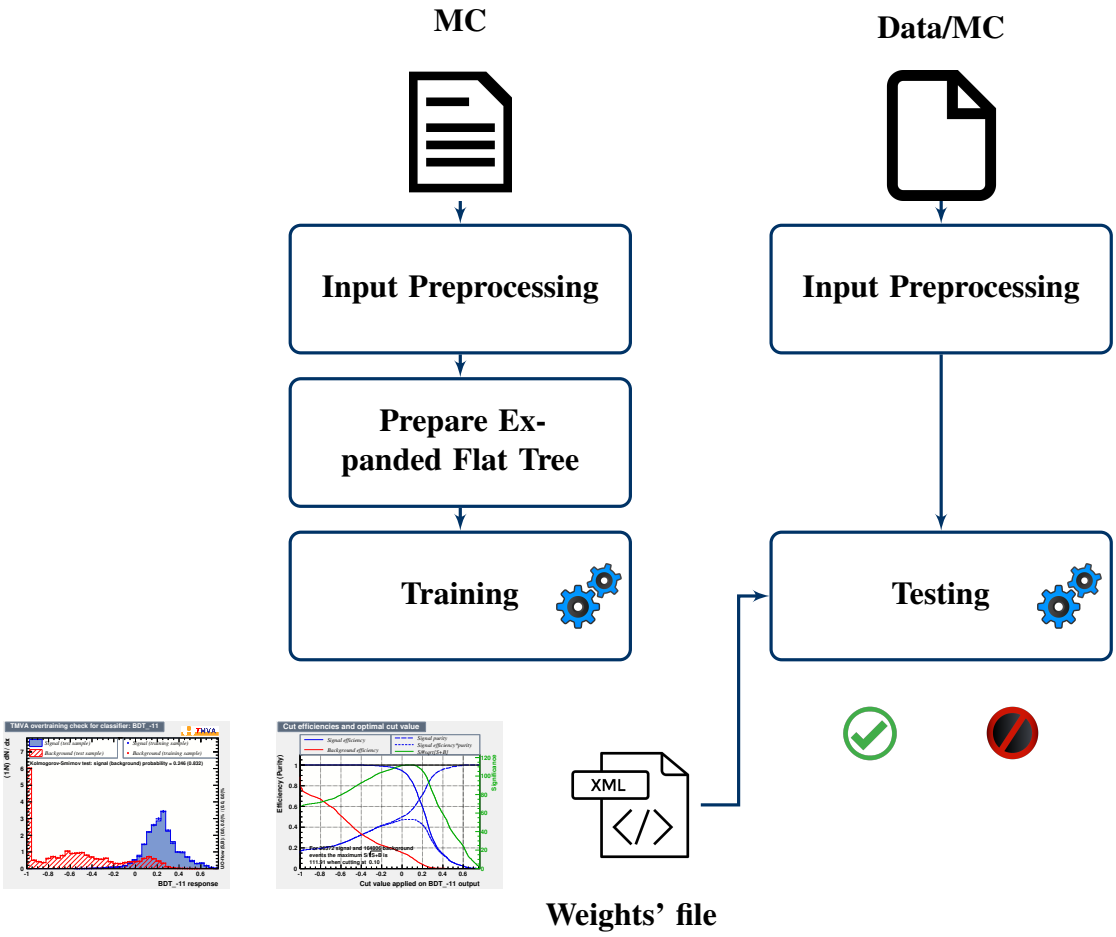


Fig. 6.5 A diagram showing the BDT integration workflow. The left branch represents the training phase, while the evaluation phase is shown on the right.

Input Preprocessing

The MC (data) used for the training (evaluation) must go through a data cleaning (preprocessing) step. In this step, only events with positive tracks, that started in FGD1 (the carbon target), and that have at least one TPC segment will be passed to the BDT. This step is necessary to match the signal definition used in the analysis (evaluation phase) of the electron anti-neutrino cross-section.

Furthermore, the outliers in input data can mislead the training process of the BDT, resulting in longer training times and poorer results. For instance, a high energy track with a short TPC segment, e.g. track starting in FGD1 with a high angle, crossing a short distance in TPC2, then entering the barrel ECAL, can have a tiny curvature, less than the TPC resolution, resulting in an overestimated calculated momentum. Consequently, the range of the momentum variable will be enormous and the BDT will have trouble dividing the range into regions to optimize the cut value of this variable. Another example is the non-filled variables. For instance in the case of a track with no ECAL segment, the default T2K software assigns a non-physical value (e.g -9999) to this variable, which will result in an unrealistic variable range, mean and standard deviation, that affects the BDT performance. Hence, outliers are replaced with a maximum/minimum value.

In the evaluation phase, an extra step called the “data quality cut” is added to make sure that the data to be processed did not suffer from any hardware or software problems when collected.

Expanded Tree Preparation

The main T2K software for data analysis is called HighLAND [70]. It is an event-based selection. The input class considers the event as a whole, with the possibility of extraction of information for each track in an event. Moreover, the MC and data files are ROOT files with a similar format. Each entry in the ROOT tree (called a flat tree) represents an event. Yet,

in order to use the BDT as a PID, features from each particle track should be extracted for every event, then processed, making the BDT selection track-based.

Furthermore, TMVA has a limitation: it can only process float or integer data types, and cannot process more complex C++ data classes such as structures and arrays.

Consequently, this step acts as an interface, transforming the input flat tree into an expanded flat tree format. Every entry of the expanded flat tree is a track for which each measured value is "expanded" into either float or integer data types.

Tuning The BDT's parameters

At this step, we fine-tune the BDT's parameters to obtain the best performance. The parameters of interest are the number of trees in the forest, the tree depth, and the minimum number of events in a leaf calculated as a percentage of the training sample (the minimum node percentage). The tree depth and the minimum node percentage parameters control the splitting of the decision tree.

We found that the BDT performance was stable after 100 trees (see Section 6.3.3). We therefore chose 150 trees in the forest to gain performance stability and to optimize the training and evaluation processing time. Consequently, the tuning parameter space is reduced to only include the tree depth and minimum node percentage. We tried tree depths of 5, 6, 7 and 8, and a minimum node percentage of 0.01, 0.05, 0.10, 0.20, 0.50, 0.75, 1.00, 1.25, and 1.50%, building a matrix for the tuning parameter consisting of 36 elements. For each case, we calculate the signal purity and efficiency after the cut, the product of efficiency and purity, as well as the Kolmogorov-Smirnov (KS) test p-value [71][2]. The numerical values of these tests are summarized in Tables B.1 to B.3.

The BDT purity and efficiency did not change significantly as a function of these parameters as shown in Figs. 6.6 to 6.8. In fact, the calculated FoM (efficiency $\times$ purity) as

[2]Kolmogorov-Smirnov test is a non-parametric test that is used, here, to check the equality of the probability distributions of the training and testing samples. It quantifies the distance between the cumulative distribution functions of the two samples and calculates the p-value

a function of the tree depth and the minimum node percentage parameters for the signal is nearly constant, indicating a stable BDT performance. Whereas, the KS test p-value suffered from large variation, indicating an overtraining problem in some cases, where a large tree depth and/or a small number of events in a leaf were allowed, as shown in Fig. 6.9.

Since we are replacing all PID cuts in our analysis with the BDT PID, we chose the configuration that has a maximum signal purity and a KS p-value > 0.1. Accordingly, the tree depth of 7 and the minimum node size of 0.2% are chosen to build the BDT structure weights file.

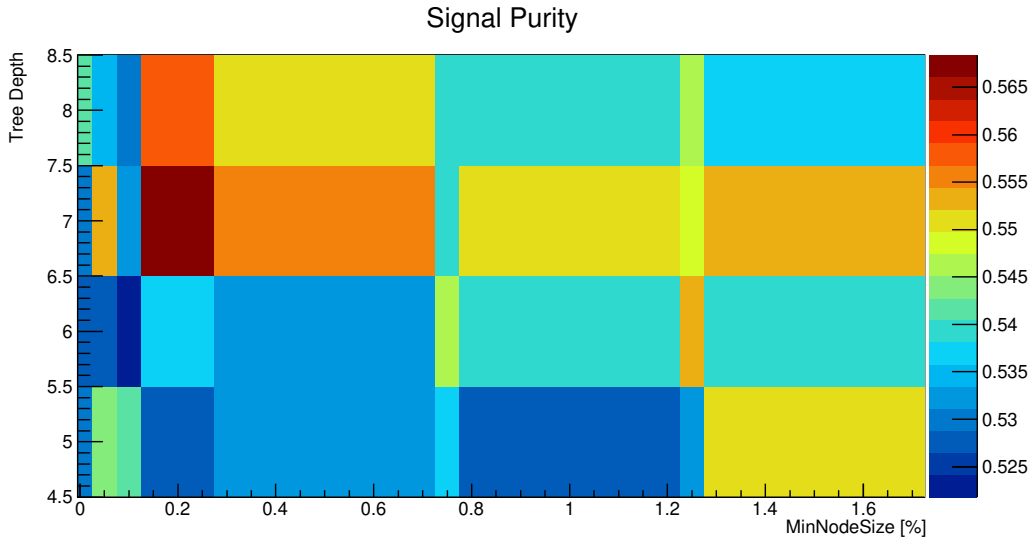


Fig. 6.6 Signal purity as a function of the tree depth and the minimum node percentage parameters.

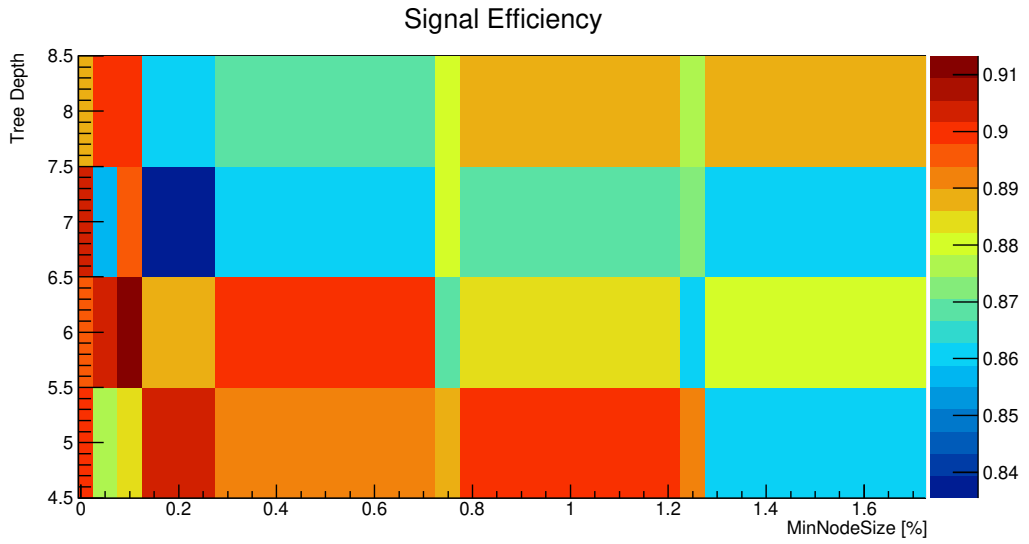


Fig. 6.7 Signal efficiency as a function of the tree depth and the minimum node percentage parameters.

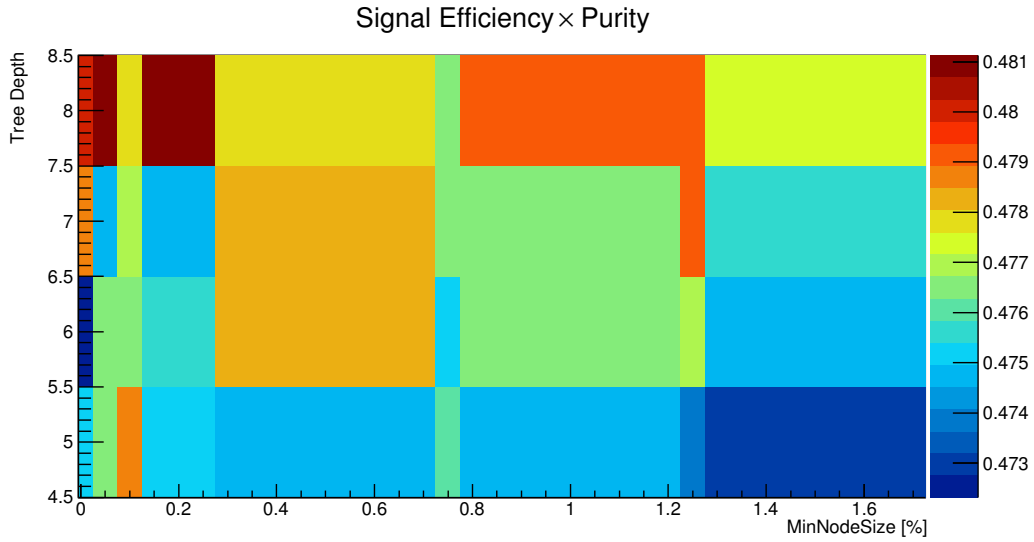


Fig. 6.8 Calculated Figure of Merit (efficiency $\times$ purity) as a function of the tree depth and the minimum node percentage parameters for the signal.

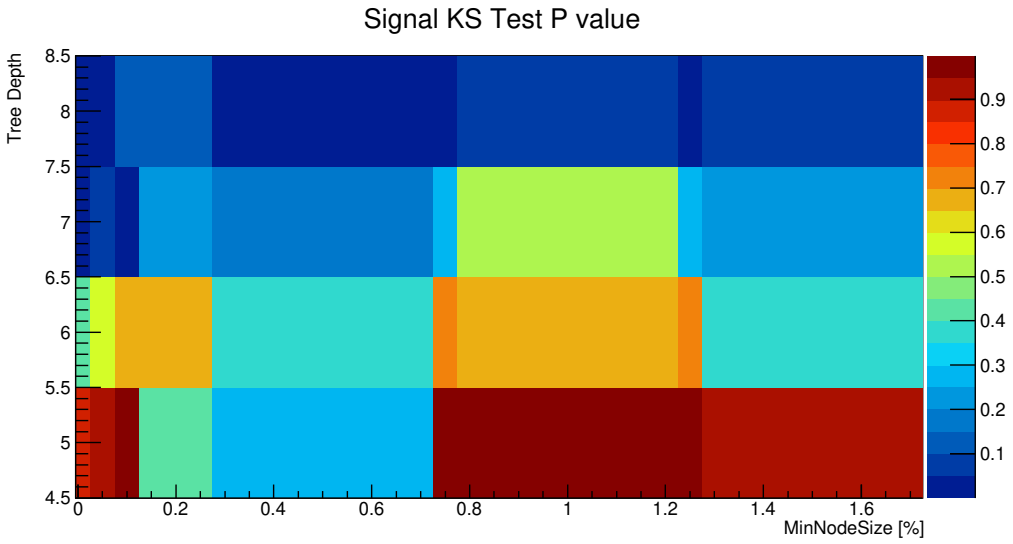


Fig. 6.9 Kolmogorov-Smirnov test p-value as a function of the tree depth and the minimum node percentage parameters for the signal.

Training

The BDT training is done using 150 trees, with a maximum tree depth equal to seven, minimum node percentage of 0.2% and an adaptive boosting (AdaBoost). The most important FoM produced by TMVA are described in this section.

The Receiver Operational Characteristic (ROC) curve represents background rejection (Eq. (6.10)) versus signal efficiency (Eq. (6.9)) obtained when cutting on the classifier outputs, and is shown in Fig. 6.10. The signal efficiency and background rejections are defined as:

$$\text{signal efficiency} = \int_{t_{cut}}^{1} PDF_{Sig}(t)dt, \tag{6.9}$$

and

$$\text{background rejection} = \int_{-1}^{t_{cut}} PDF_{Bkg}(t)dt, \tag{6.10}$$

with PDF_{Sig} and PDF_{Bkg} the probability density function for the signal and background, respectively, and t and t_{cut} are the test statistic and the chosen cut values, respectively. A

good classifier will have high signal efficiency and high background rejection, which gives a maximum area under the ROC curve.

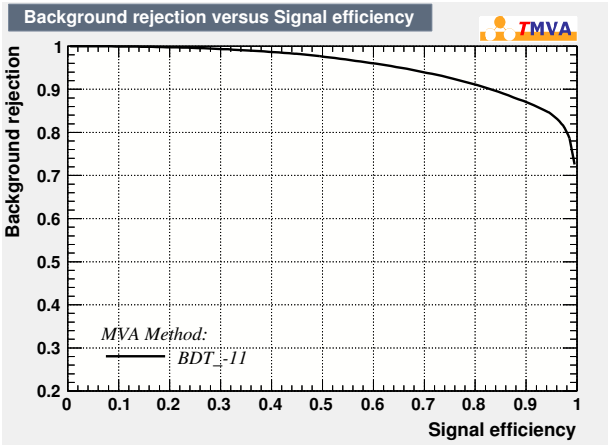


Fig. 6.10 A figure showing the BDT Receiver Operational Characteristic (ROC) curve, representing background rejection versus signal efficiency obtained when cutting on the classifier outputs.

Another useful FoM is the cut efficiency and optimal cut value plot, shown in Fig. 6.11. An optimal BDT test statistic cut value could be inferred from this plot, for instance, by maximizing the signal efficiency $\times$ purity, and/or the $\frac{S}{\sqrt{S+B}}$, where S and B are the signal and background event numbers respectively, cf. [72].

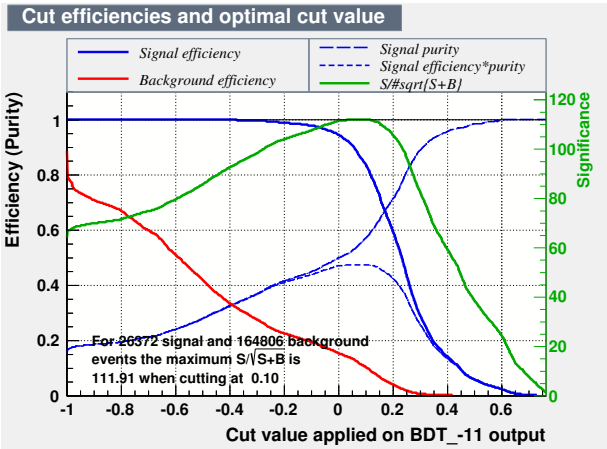


Fig. 6.11 A figure showing the efficiency and purity as a function of the BDT test statistic cut value. Signal and background efficiencies are shown in blue and red solid curves respectively, and signal purity is shown in long dash blue line. Two FoMs are presented: the signal efficiency $\times$ purity in short dash blue line, and $\frac{S}{\sqrt{S+B}}$, where S and B are the numbers of signal and background respectively, in green solid line. A test statistic cut value is chosen such that it maximizes one (or both) of the previous FoMs.

To check the overtraining, the superimposed signal and background distributions for the trained classifiers (test and training samples), are shown in Fig. 6.12. When there is no overtraining, the signal (background) test statistic histogram of the testing dataset will follow the same distribution as the training dataset within the statistical uncertainties.

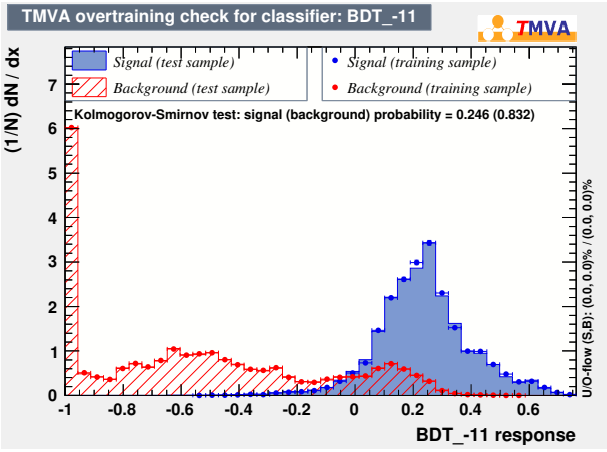


Fig. 6.12 A figure showing the signal and background test statistic value distribution for the training dataset, in blue and red points, respectively, with statistical fluctuations; as well as those of the testing dataset in filled blue and red hatched histogram respectively.

In order to check the BDT performance as a function of the number of trees used, we investigate several BDT control plots, shown in Fig. 6.13.

The integrated area under the ROC curve as a function of the number of trees quickly converges and stabilizes to a number close to one, showing a fast and good background rejection and signal efficiency. The Adaboost weight distribution shows the weights in Eq. (6.7); higher weights mean a misclassified event that failed to be learned by a sequence of trees. A boost weights versus tree plot shows how fast the BDT learns the misclassified events. After the first few tens of trees, the BDT performance is stable. In this thesis, no pruning is used, as recommended for a BDT (cf. [66, Section 8.12]). Instead a smaller tree depth and a minimum number of events in a leaf are forced. Investigating Fig. 6.13, one can conclude that the BDT is a strong, fast and stable classifier.

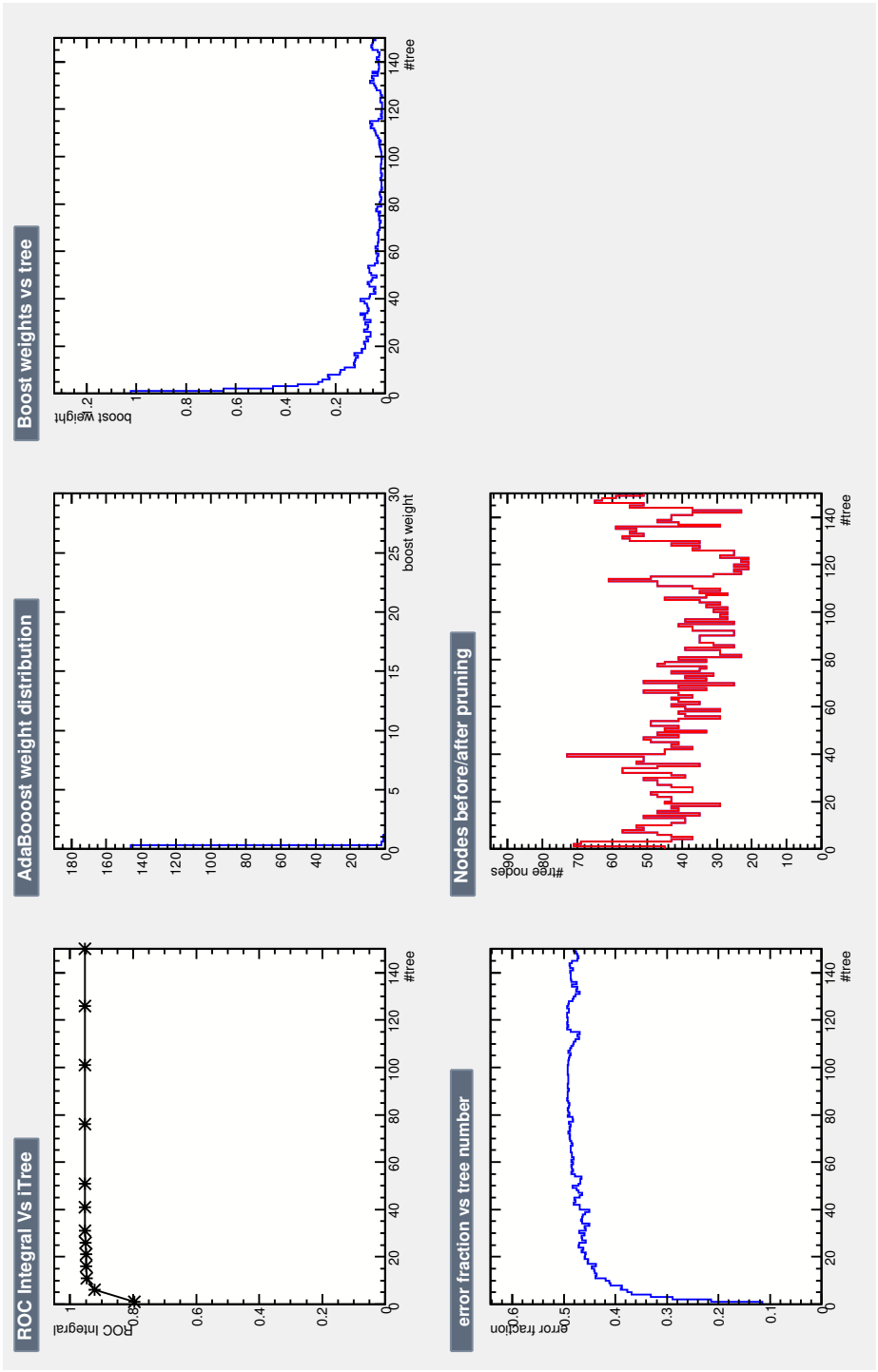


Fig. 6.13 These plots show the BDT performance with respect to the number of trees. The integrated area under the ROC curve, that is close to one, demonstrates a good background rejection and signal efficiency. The Adaboost weight distribution shows the weights in Eq. (6.7), and the boost weights versus tree plot shows how fast the BDT learns the misclassified events; Node before and after pruning is not used in this thesis, instead a smaller tree depth and a minimum number of events in a leaf are forced.

Table 6.2 summarizes the main parameters used to fine tune the BDT along with their chosen values. A full list of the BDT parameters supported by TMVA is in [66, tables 9,22,24].

BDT Parameter Name	Description	Value
NTrees	number of trees in the forest	150
MinNodeSize	minimum node size	0.2%
MaxDepth	maximum tree depth	7
nCuts	number of division for a certain variable value range to optimize the cut	-1 (find optimum cut value)
BoostType	boosting algorithm	AdaBoost
AdaBoostBeta	Learning rate for AdaBoost algorithm (boosting strength)	0.5

Table 6.2 BDT parameters and their tuned values

After fine tuning the BDT parameters, the generated weight file describing the structure of the BDT is used in the evaluation phase.

Figure 6.14 shows the calculated test statistic distribution for different particles, neglecting their charges. It is clear that the trained BDT can easily distinguish the signal (positron) from the background (anything else).

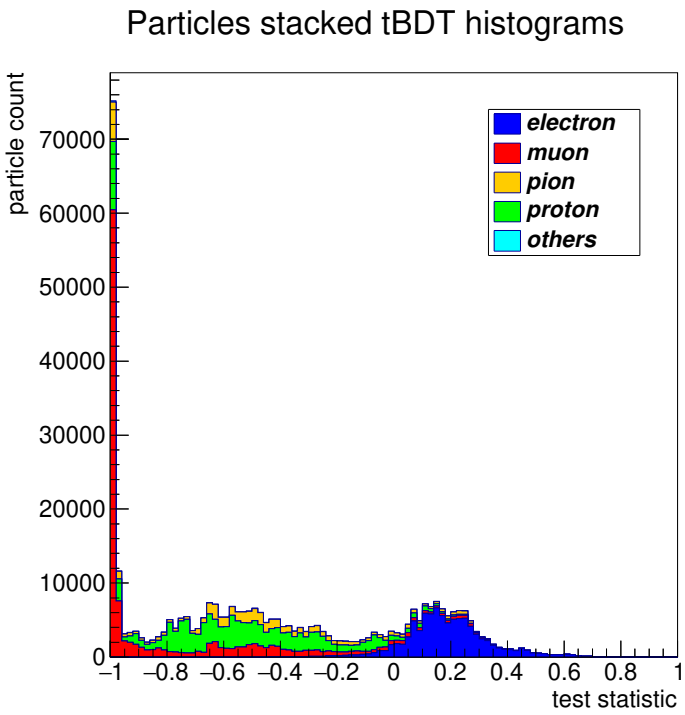


Fig. 6.14 A figure showing the calculated test statistic distribution (tBDT) for different particles.

Testing

The inputs for the testing phase are statistically separate samples of the preprocessed data (or MC), and the weights file generated from the training phase. The BDT can then cut on the calculated test statistic, to separate the signal from the background.

It is crucial for any machine learning method to carefully choose the discriminating variables, fine-tune its parameters, and to evaluate its systematic uncertainties, as will be described in the next chapters.

Chapter 7

BDT Input Variables

In this chapter, the different measured quantities from the TPC, FGD and ECAL, that are used as input for the BDT, will be explained along with their discrimination power. Furthermore, the distribution of the input variables for both signal and background MC simulation, representing the separation power of each variable, will be discussed as well as the data to MC agreement.

7.1 BDT Input Variables Definition

7.1.1 TPC Variables

The TPC provides the momentum measurement of a track, which is directly input to the BDT as a variable. Furthermore, the TPC can distinguish between different particles, by measuring their ionization energy loss per unit length. The average estimated ionization energy loss per unit length, as a function of particle momentum, follows the Bethe-Bloch equation, cf. [8, Section 33.2],

$$\left\langle -\frac{dE}{dx} \right\rangle = kz^2 \frac{Z}{A} \frac{1}{\beta^2} \left(\frac{1}{2} \frac{2m_e c^2 \beta^2 \gamma^2 W_{max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right). \tag{7.1}$$

Here $k = 4\pi N_A r_e^2 m_e c^2$, with N_A Avogadro's number, m_e the electron mass, c the speed of light and r_e the classical relativistic electron radius, z and Z are the atomic number of the incident and absorber particle respectively, A is the atomic mass of the absorber, $\beta = \frac{v}{c}$, with v the particle's velocity, $\gamma = \frac{1}{\sqrt{1-\beta^2}}$, $\delta(\beta\gamma)$ is a parameter that takes into account correction effects based on material density, I is the mean excitation energy, and W_{max} is the Maximum energy transfer in a single collision given by

$$W_{max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + \frac{2\gamma m_e}{M} + (\frac{m_e}{M})^2}. \tag{7.2}$$

In Eq. (7.2), M is the incident particle mass. A pull value, assuming a certain particle, may be constructed as

$$\text{pull}_{\text{particle i}} = \frac{\frac{dE}{dX}^{\text{meas}} - \frac{dE}{dX}^{\text{particle i}}}{\sigma^{\text{particle i}}}, \tag{7.3}$$

where $\frac{dE}{dX}$ is the ionization energy loss per unit length, the superscripts "meas" and "particle i" indicate the measured quantity and the estimated one under a certain particle assumption respectively, σ is the estimated uncertainty, and "particle i" is $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$ or the proton.

The pulls for different particles, electrons, muons, pions and protons, are chosen as inputs for the BDT, rather than just the measured $\frac{dE}{dX}$, to include the information gained from the particle-matter interaction theory, increasing the discrimination power of the input variables.

7.1.2 FGD Variables

A previous work [73] developed an algorithm for identification of particles (protons, electrons, muons and pions) with trajectories contained within the FGD. This was achieved by constructing a distribution of the energy deposited by a particle as a function of its track length using MC simulation. The track length was then divided into 1 cm bins, and the energy distribution was fitted to a Gaussian function for each bin. The mean and the standard deviation of this Gaussian are then used as the mean value of the energy and its uncertainty,

respectively, at a specific track length. The points, with their uncertainty, are then used to fit the parameters of the curve $ax^b + c$. Furthermore, a pull is constructed as

$$\text{pull}_{\text{particle i}} = \frac{E_{meas} - E_{simulated}(\text{particle i, measured length})}{\sigma_{E_{simulated}}(\text{particle i, measured length})}. \tag{7.4}$$

The current limitations of the FGD PID are:

- At the moment of writing this thesis, the FGD PID capability is limited to the tracks that stop inside the FGD.
- Only a small range of track angles were considered since the reconstruction failed at angles $> 70^{\circ}$.

Since we are interested in tracks starting in FGD1 and having a TPC segment, and due to the above-mentioned limitations, we can only use the FGD2 PIDs.

In this thesis, we use proton and pion pulls as inputs for the BDT for two reasons.

- The proton is one of our main backgrounds. Consequently, a dedicated proton pull can enhance the signal purity.
- In the pion pull calculation, the electrons (positrons) and protons have different pull signs (positive for electrons and negative for protons) [73].

7.1.3 ECAL Variables

The ECAL provides useful information to distinguish between electrons, muons, pions and protons. For instance, a muon is a Minimum Ionizing Particle (MIP) while an electron typically creates a shower. Furthermore, the ECAL complements the TPC PID. For instance, a pion behaving as a MIP in the TPC may undergo hadronic interactions in the ECAL. Several quantities extracted from the ECAL reconstruction are used to distinguish between tracks

and showers. The ECAL PID variables are briefly explained here, for more information, please refer to [74].

Circularity

The Circularity variable measures the width of a charge cluster compared to its length. A track-like object will consist of long thin cluster of hits, whereas, a shower-like object will have a short wide cluster of hits. The Principal Component Analysis (PCA)[1] of the two-dimensional hits distribution, where each hit is weighted by its deposited charge, is extracted for the x-y and y-z planes and used to construct the two major axes of the hits cluster. The total circularity is defined as:

$$\text{Circularity} = \text{Circularity}_{x-y} \times \text{Circularity}_{y-z}, \tag{7.5}$$

where, the circularity in any i-j plane is defined as:

$$\text{Circularity}_{i-j} = 2(\text{PCA } 2^{nd} \text{ component}) - 1. \tag{7.6}$$

In this definition, the circularity will be bounded as:

$$0 \leq \mathit{Circularity} \leq 1. \tag{7.7}$$

The simple major axis to minor axis ratio is not used to avoid the problems that may arise from the finite ECAL bar size, since, in certain cases, the reconstructed hits can have the same x or y leading to a minor axis length equals to zero.

[1] The Principal Component Analysis (PCA) is a dimensionality reduction technique. It is the orthogonal projection of the data onto a lower dimensional linear space (formed from the eigenvectors of the data covariance matrix), such that the variance of the projected data is maximized, cf. [69, Section 12.1].

Truncated Maximum Ratio

Truncated Maximum Ratio (TMR) is defined as the ratio of the maximum charge deposited in layer i to the minimum charge deposited in layer j, after truncating the highest and lowest 20% charge deposit layers to reduce the effect of noise.

Charge Root Mean Square (QRMS)

Since the deposited charge quantity is affected by the electronic noise, which is poorly simulated, the Data/MC agreement will be poor for such a variable. Consequently, the charge root mean square (QRMS) is used instead and defined as

$$QRMS = \frac{1}{\overline{q}} \sqrt{\sum_{i}^{N} \frac{(q_i - \overline{q})^2}{N}}. \tag{7.8}$$

Where N is the number of hits in the cluster, q_i is the charge measured at layer i, and $\overline{q}$ is the mean hit charge in the cluster.

Front Back Ratio

As particles slow down, they lose more energy due to ionization. Protons can stop inside the ECAL deposit most of their energy just before coming to rest. In order to include the Bragg peak effect[2], the principal component of the three-dimensional reconstructed hits is extracted using PCA. Consequently, the ratio of the total charge of the back quarter of the track inside the ECAL to the total charge of the front quarter is calculated. The Front Back Ratio (FBR) is defined as

$$FBR = \frac{\text{track's back-quarter charge}}{\text{track's front-quarter charge}}. \tag{7.9}$$

[2]The Bragg curve plots the energy loss of ionizing radiation versus the travelled distance through matter. Protons depose most of their energy immediately before it comes to rest.

These variables are used as inputs for a likelihood function to generate three higher level discriminating variables: the Log Likelihood Ratios (LLR), cf.[74, Section 5]. Each LLR distinguishes between two hypotheses,

$$LLR\_H0\_H1 = \log \frac{P(\vec{x}|H_0)}{P(\vec{x}|H_1)} \approx \sum_{i=1}^{4} \log \frac{P(x_i|H_0)}{P(x_i|H_1)}, \tag{7.10}$$

where, $\vec{x}$ is the vector of the four ECAL variables (Circularity, TMR, QRMS and FBR) and H_0 and H_1 are the two particle hypotheses to evaluate (e.g. proton, electron, muon or pions).

The ECAL PID assumes no correlation between the discriminating variables to calculate the likelihood ratios as in Eq. (7.10).

In this thesis, we use the LLR_MIP_EM and LLR_EM_HIP to distinguish between the muon and electron hypothesis as well as the electron and proton.

Electromagnetic Energy Deposited

The total electromagnetic (EM) energy deposited is also extracted from the ECAL by adding the EM energy deposits from the barrel and downstream ECAL, and then added to the BDT.

7.2 MC Signal/Background Variable Distributions

The input variables for the BDT are seldom equally relevant. Therefore, it is expected that some variables will have a bigger influence on the response than the others. The variables with different distributions for the signal and background, or those that have a different correlation in the signal and background space, are the best discriminators. A quantitative measure for a variable's discrimination power is presented in Section 7.3.

Figure 7.1 shows the distributions of the BDT input variables for both signal (positron track) and background (any other particle).

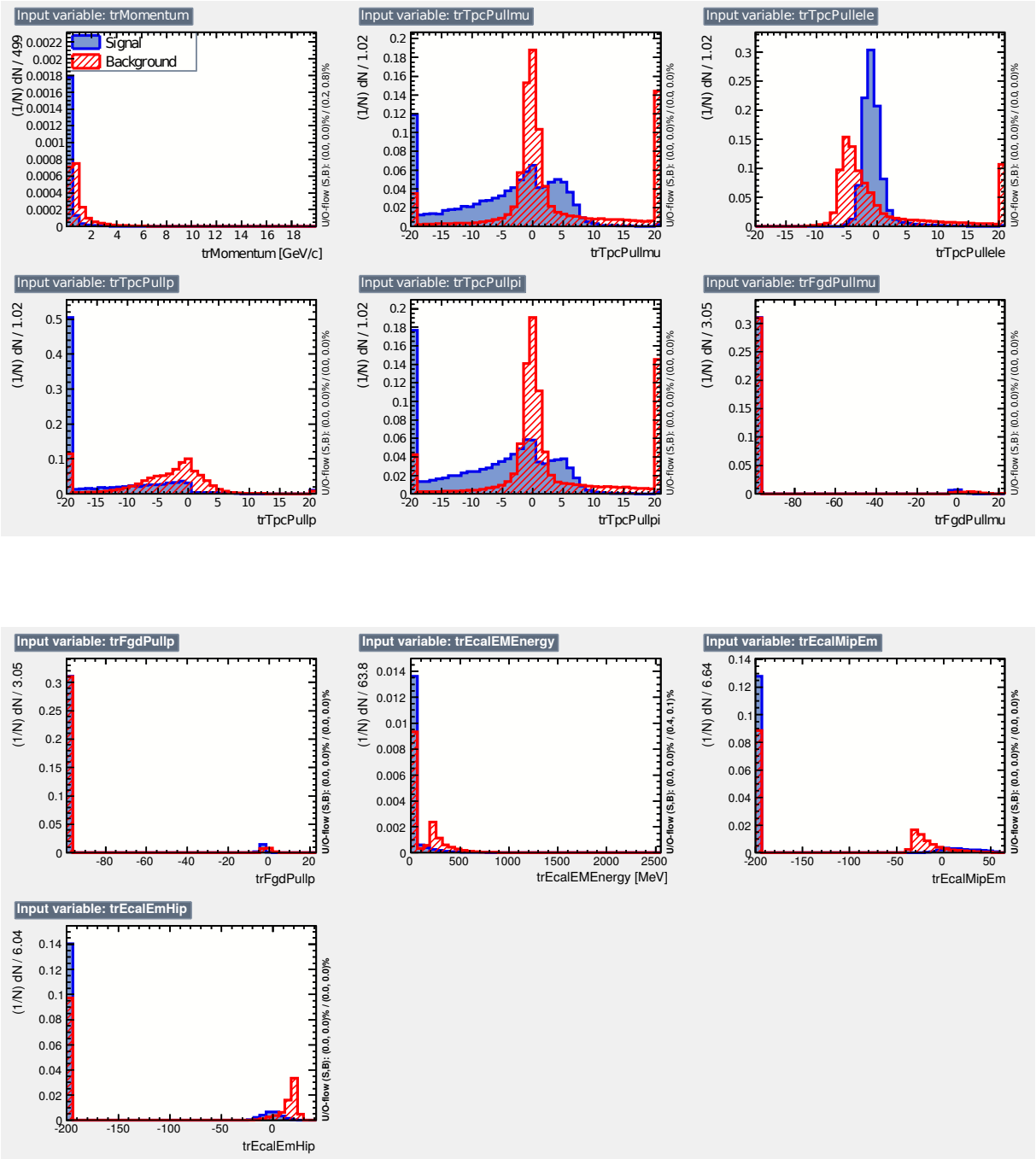


Fig. 7.1 The distribution of the BDT input variables for both signal (in blue) and background (in red), showing the separation power of each individual variable.

The correlations of the input variables for both signal and background are shown in Fig. 7.2 and Fig. 7.3 respectively.

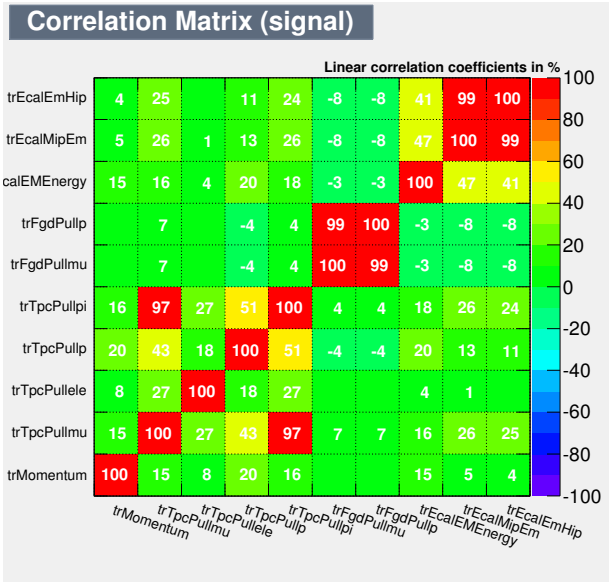


Fig. 7.2 The BDT input variable correlation for positron track signal.

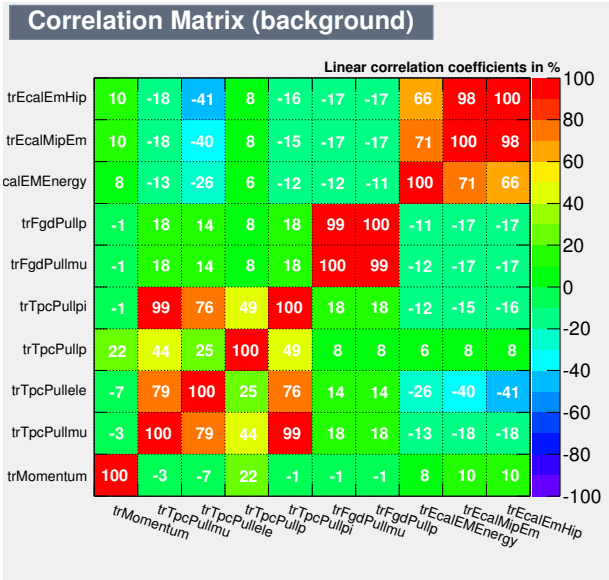


Fig. 7.3 The BDT input variable correlation for background.

7.3 Variables Discrimination Power

The TMVA software provides two quantities, called variable importance and separation, to quantify the input variables discrimination power. The variable importance is based on how

often the variable is used to split the decision tree nodes, weighted by the separation-gain squared, Eq. (6.3), and the square of the number of events in the node, cf. [75, Section 10.13]. The importance of variable x, I_x, is then averaged over the number of trees, M, in the forest as

$$I_x = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{N} g_{i,m}^2 \times n_{i,m}^2, \tag{7.11}$$

where N is the total number of nodes in a tree that have x as a split variable, and $g_{i,m}$ and $n_{i,m}$ are the separation gain and the total number of events (signal + background) in node i for tree m, respectively.

Another measurement of variable discrimination power is the separation, $\langle S \rangle$, defined as

$$\left\langle S^2 \right\rangle = \frac{1}{2} \int \frac{Y(y)_s - Y(y)_B}{Y(y)_s + Y(y)_B} dy. \tag{7.12}$$

Here $Y(y)$ is the PDF of variable y, and the subscripts s and B are for signal and background respectively, cf. [66, Section 3.1.10]. The separation is zero for signal and background with identical distributions, and it is one for non-overlapping distributions.

Table 7.1 summarizes the reconstructed quantities measured from the TPC, FGD and ECAL that are used as input for the BDT, along with their calculated variable separation and importance.

Variable Name	Description	Separation	Variable Importance
trEcalEmHip	ECAL LLR of electrons and protons	1.924×10^{-1}	3.123×10^{-1}
trMomentum	track momentum	3.053×10^{-1}	2.088×10^{-1}
trTpcPullele	TPC electron pull	4.813×10^{-1}	1.103×10^{-1}
trTpcPullpi	TPC pion pull	3.338×10^{-1}	9.140×10^{-2}
trEcalEMEnergy	total energy deposit in the ECAL	1.335×10^{-1}	8.242×10^{-2}
trTpcPullp	TPC proton pull	2.775×10^{-1}	7.747×10^{-2}
trTpcPullmu	TPC muon pull	3.082×10^{-1}	5.127×10^{-2}
trFgdPullp	FGD2 proton pull	1.568×10^{-2}	2.860×10^{-2}
trEcalMipEm	ECAL LLR of muons and electrons	1.822×10^{-1}	2.596×10^{-2}
trFgdPullmu	FGD2 muon pull	1.631×10^{-2}	1.150×10^{-2}

Table 7.1 The reconstructed quantities measured from the TPC, FGD and ECAL that are used as input for the BDT, along with their calculated variable separation and importance (ordered by their importance).

From Table 7.1, we can conclude that the ECAL LLR of electrons and protons is the most important variable to separate our signal (positron) from the background. The track momentum and total ECAL energy play an important role as well since our neutrino beam flux is centred around the 0.6 GeV. Nevertheless, the TPC electron pull is crucial for the selection as we require in our signal definition that all the selected tracks have a TPC segment, and it is very powerful in separating electrons and muons. However, the TPC pull alone cannot separate the electrons (positrons) from the protons as their ionization curves overlap at the 1 GeV range. The FGD2 PIDs were less important, considering that we only use them for the subgroup of tracks that start in FGD1, travel across TPC2, then stop in FGD2.

7.4 Data/MC Agreement

The Data to MC agreement for each of the variables used as discriminators by the BDT is studied to ensure they are (as much as possible) model independent. A scaled MC histogram for each variable, where the scale is the ratio of the integral of data events to the integral of MC events, is overlaid with the data. Furthermore, a ratio of data/MC for each bin is provided, as shown in Figs. 7.4 to 7.13.

The agreement is quantified by a χ^2/Ndof as summarized in Table 7.2. Here, χ^2 is defined as

$$\chi^2 = \sum_i \frac{(\text{data}_i - \text{MC}_i)^2}{\sigma^2_{\text{data}_i} + \sigma^2_{\text{MC}_i}}, \tag{7.13}$$

with data_i and MC_i being the number of data and MC events in bin i, respectively, while σ_{data_i} and σ_{MC_i} are the statistical uncertainty for data and MC assuming a Poissonian model. Moreover, the number of degrees of freedom Ndof is equal to the number of bins - 1 for the data and MC histograms.

In the ratio plots, the error bars are calculated, for each bin, using the error propagation formula[3], and assuming uncorrelated histogram bins.

The poor χ^2/Ndof for the Data/MC is due to the systematic uncertainties, as will be discussed in Chapter 12. Including the systematic uncertainties in the MC, reduces the calculated χ^2. Furthermore, the systematics of some variables are bin-dependent. The large discrepancy in the first bin of the momentum and in the TPC electron pulls can be explained by the large amount of gamma background in the data that is mis-modelled in the MC. Evaluating the systematic on the BDT cut will be discussed in Chapter 9. In addition, these distributions are shown without all of the final cuts applied, and with a particular choice of neutrino-nucleon cross-section model parameters in the simulation (NEUT v5.3.2). The cuts values are chosen to reduce the differences between data and Monte-Carlo by removing parts of the distributions which are less well modelled in the simulation. The cuts for the $\overline{\nu}_e$ event selection is described further in Chapter 8.

[3]If y was a function of x_n with $n = 1,..,N$; the uncertainty in y due to uncertainty in x_n is given by

$$\sigma_y^2 = \frac{\partial y}{\partial x_i}\frac{\partial y}{\partial x_j}V^{ij}. \tag{7.14}$$

Where V^{ij} is the covariance matrix element of the function parameters (x_n).

Variable Name	χ^2/Ndof
trMomentum	371.57/19
trTpcPullele	1675.16/19
trTpcPullmu	224.79/19
trTpcPullp	379.28/19
trTpcPullpi	127.81/19
trFgdPullp	16.18/19
trFgdPullmu	20.13/19
trEcalEMEnergy	1247.01/19
trEcalMipEm	91.74/19
trEcalEmHip	184.48/19

Table 7.2 Data/MC agreement for the variables used as discriminators by the BDT.

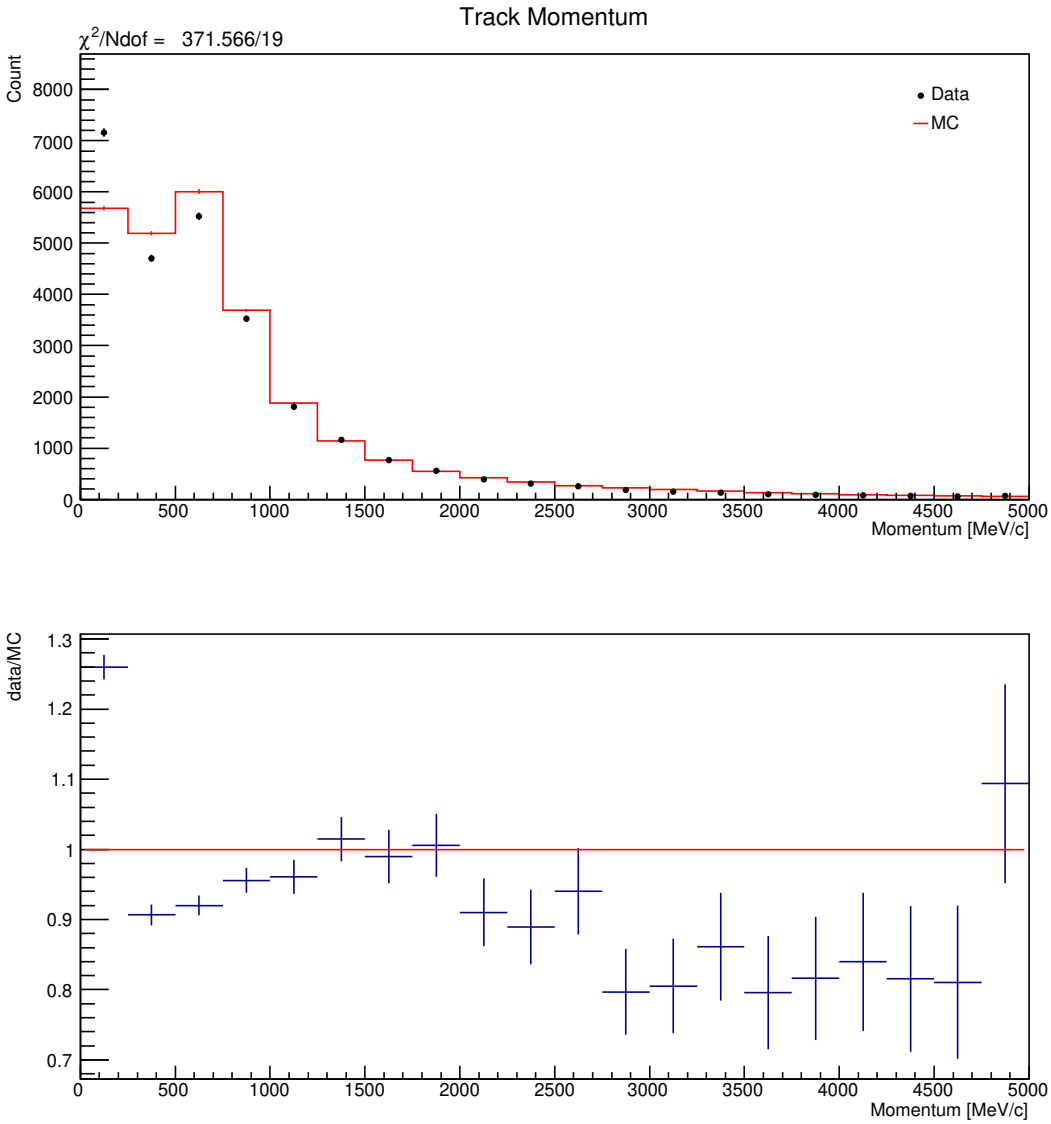


Fig. 7.4 Data/MC agreement for the momentum variable. Top plot represents an overlay of a scaled MC prediction (red histogram) and data points (black circles) with their statistical uncertainty (black lines). Data/MC ratio with the corresponding statistical uncertainty is shown in the bottom plot (in blue) as well as the ideal ratio value of unity (in red).

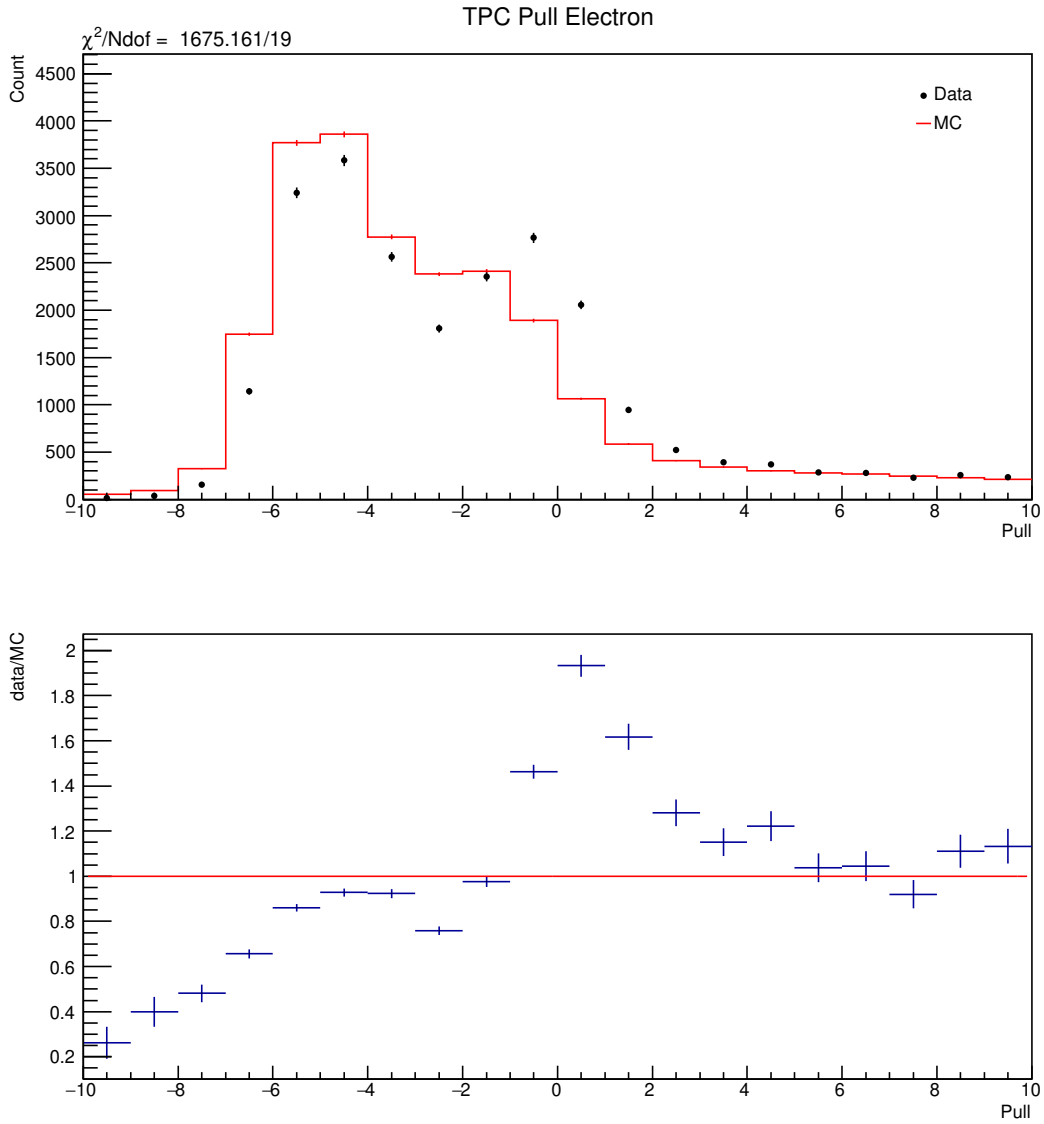


Fig. 7.5 Data/MC agreement for the TPC electron pull variable. Top plot represents an overlay of a scaled MC prediction (red histogram) and data points (black circles) with their statistical uncertainty (black lines). Data/MC ratio with the corresponding statistical uncertainty is shown in the bottom plot (in blue) as well as the ideal ratio value of unity (in red).

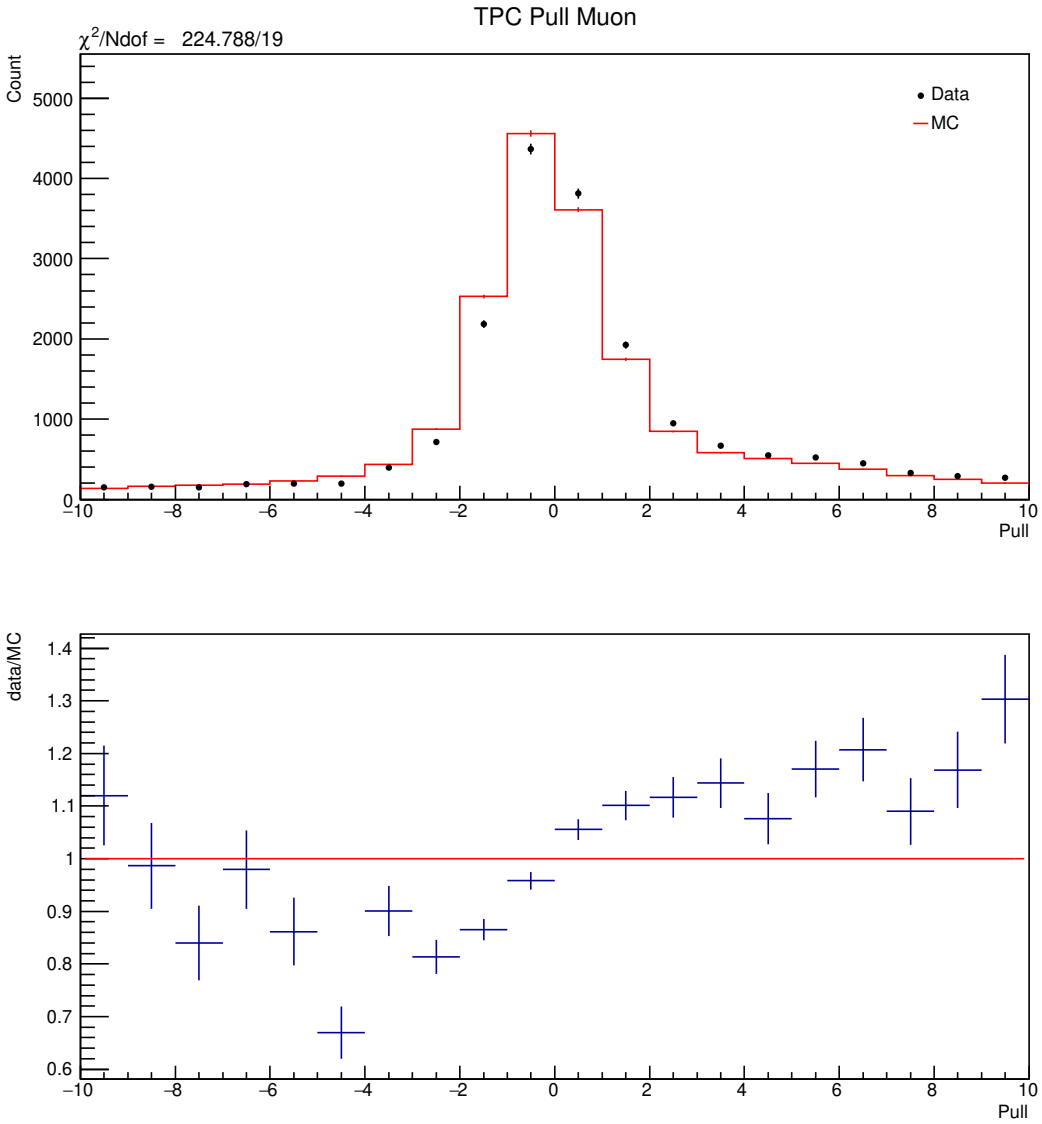


Fig. 7.6 Data/MC agreement for the TPC muon pull variable. Top plot represents an overlay of a scaled MC prediction (red histogram) and data points (black circles) with their statistical uncertainty (black lines). Data/MC ratio with the corresponding statistical uncertainty is shown in the bottom plot (in blue) as well as the ideal ratio value of unity (in red).

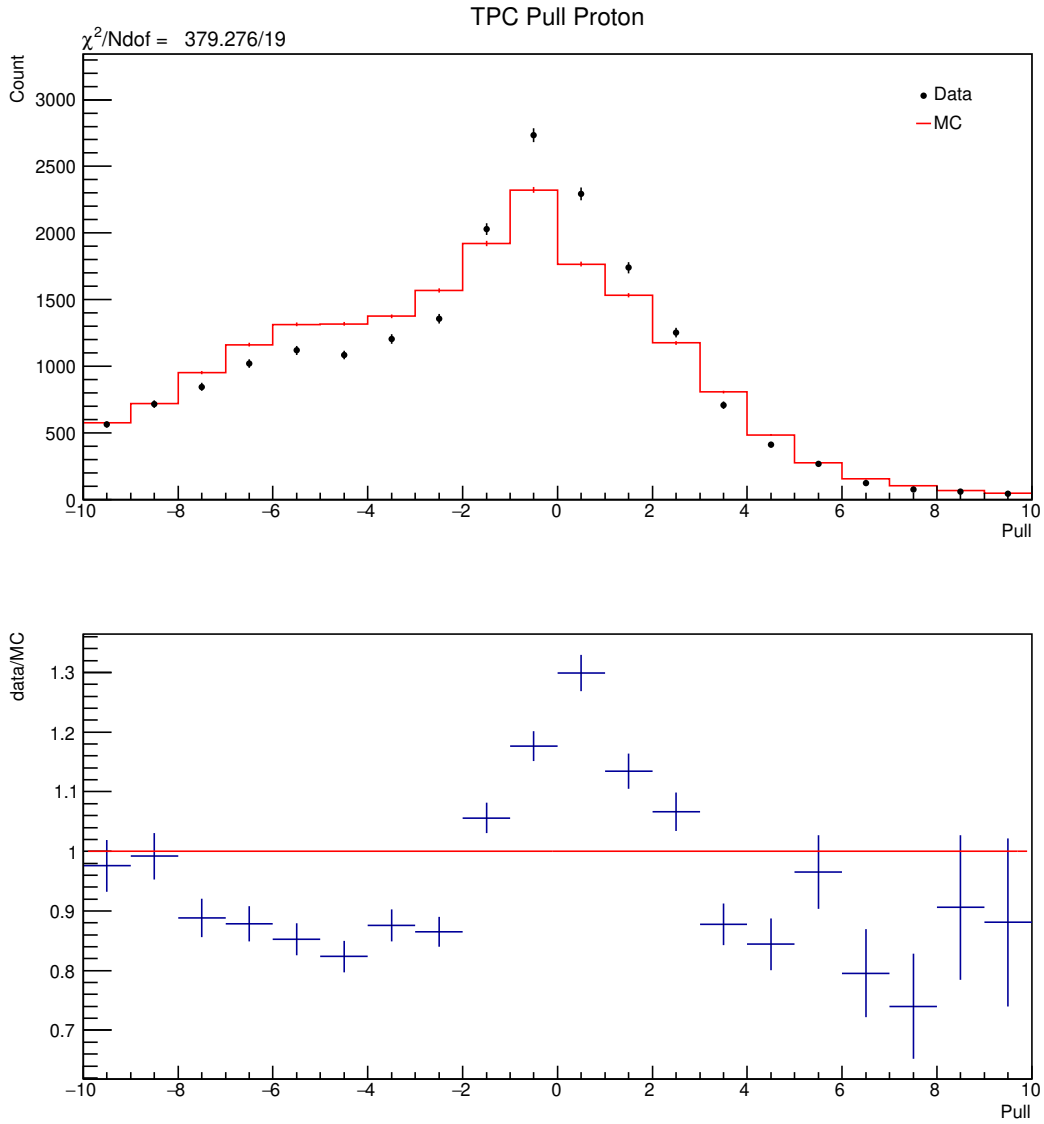


Fig. 7.7 Data/MC agreement for the TPC proton pull variable. Top plot represents an overlay of a scaled MC prediction (red histogram) and data points (black circles) with their statistical uncertainty (black lines). Data/MC ratio with the corresponding statistical uncertainty is shown in the bottom plot (in blue) as well as the ideal ratio value of unity (in red).

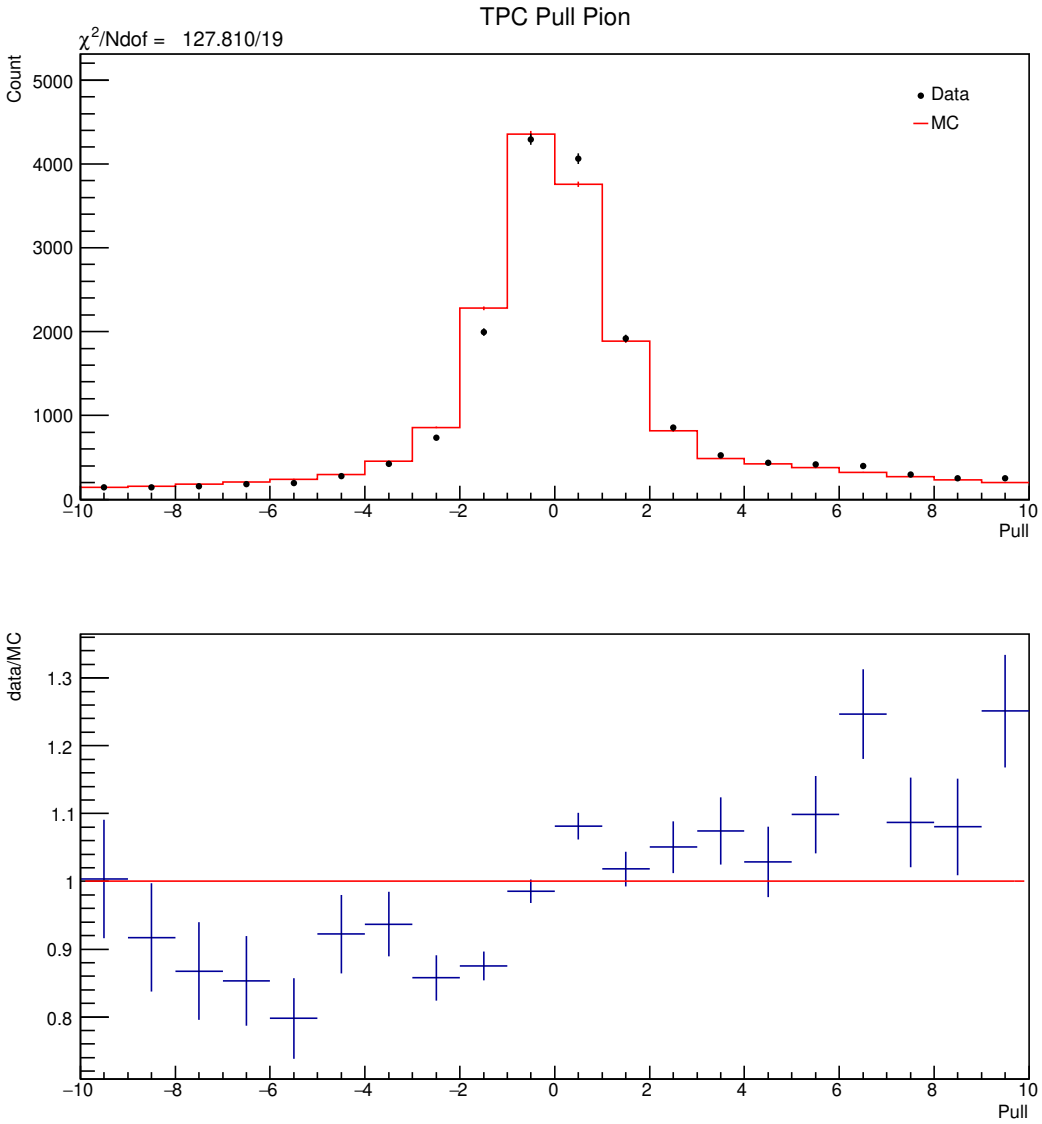


Fig. 7.8 Data/MC agreement for the TPC pion pull variable. Top plot represents an overlay of a scaled MC prediction (red histogram) and data points (black circles) with their statistical uncertainty (black lines). Data/MC ratio with the corresponding statistical uncertainty is shown in the bottom plot (in blue) as well as the ideal ratio value of unity (in red).

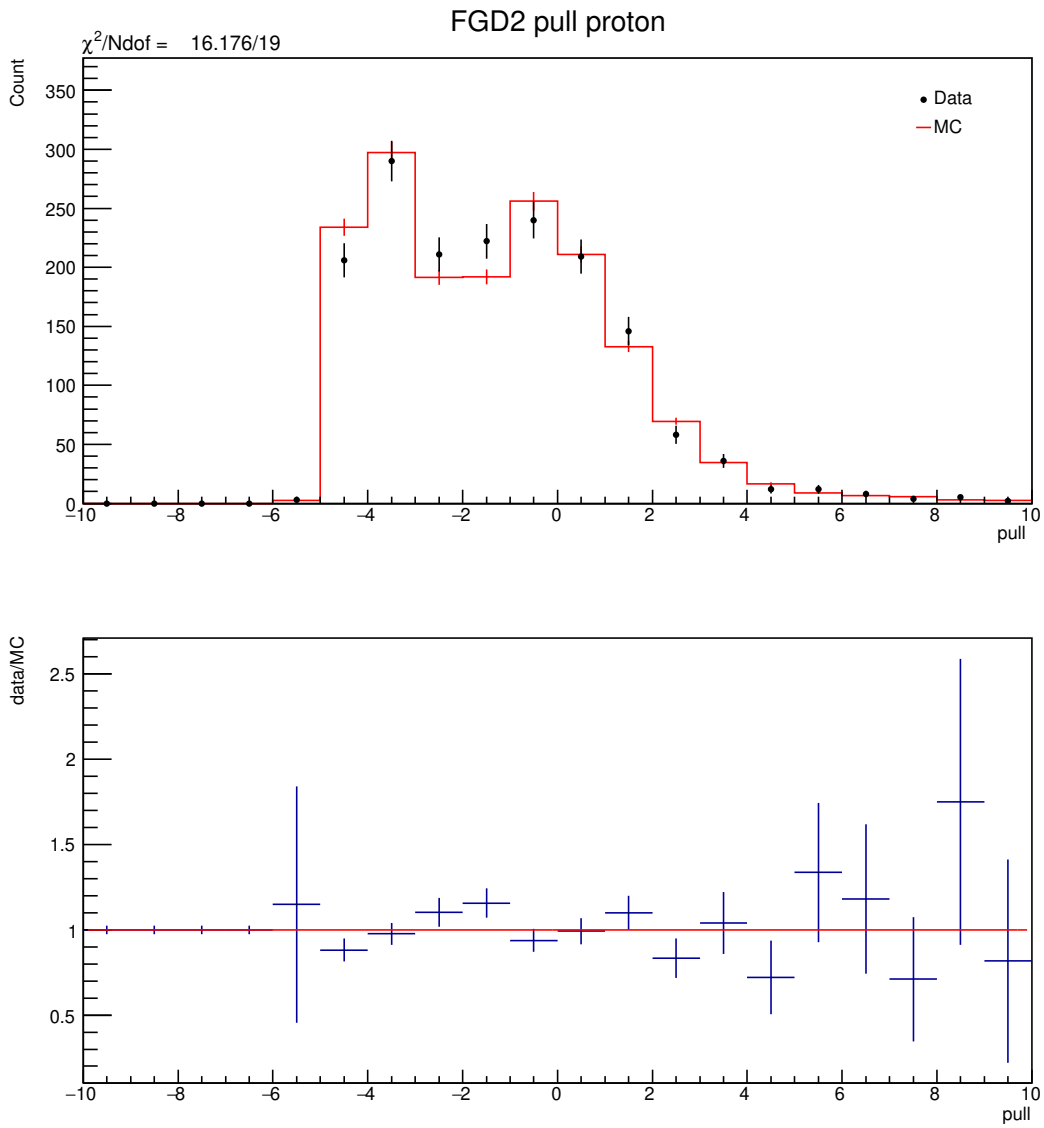


Fig. 7.9 Data/MC agreement for the FGD2 proton pull variable. Top plot represents an overlay of a scaled MC prediction (red histogram) and data points (black circles) with their statistical uncertainty (black lines). Data/MC ratio with the corresponding statistical uncertainty is shown in the bottom plot (in blue) as well as the ideal ratio value of unity (in red).

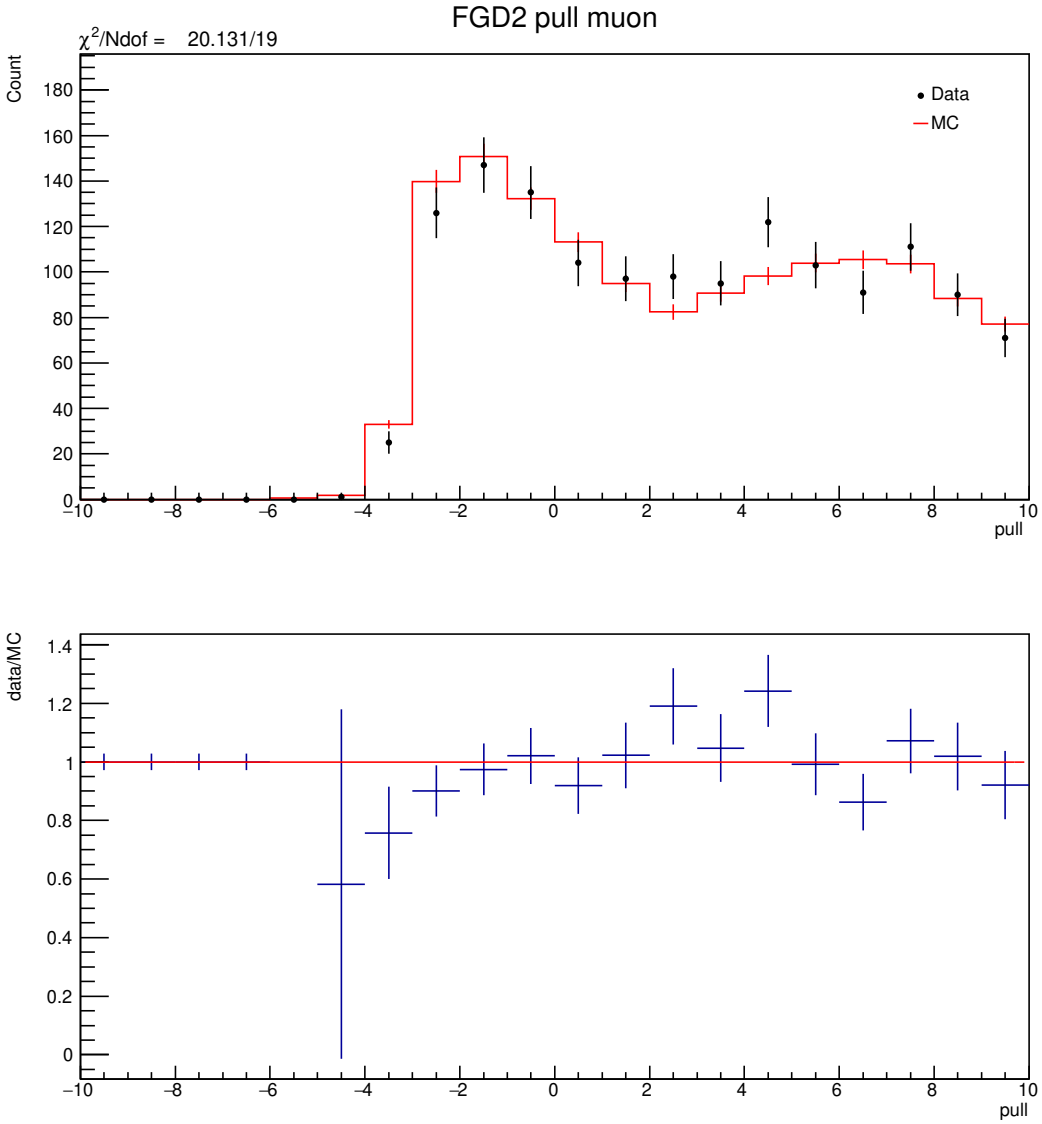


Fig. 7.10 Data/MC agreement for the FGD2 muon pull variable. Top plot represents an overlay of a scaled MC prediction (red histogram) and data points (black circles) with their statistical uncertainty (black lines). Data/MC ratio with the corresponding statistical uncertainty is shown in the bottom plot (in blue) as well as the ideal ratio value of unity (in red).

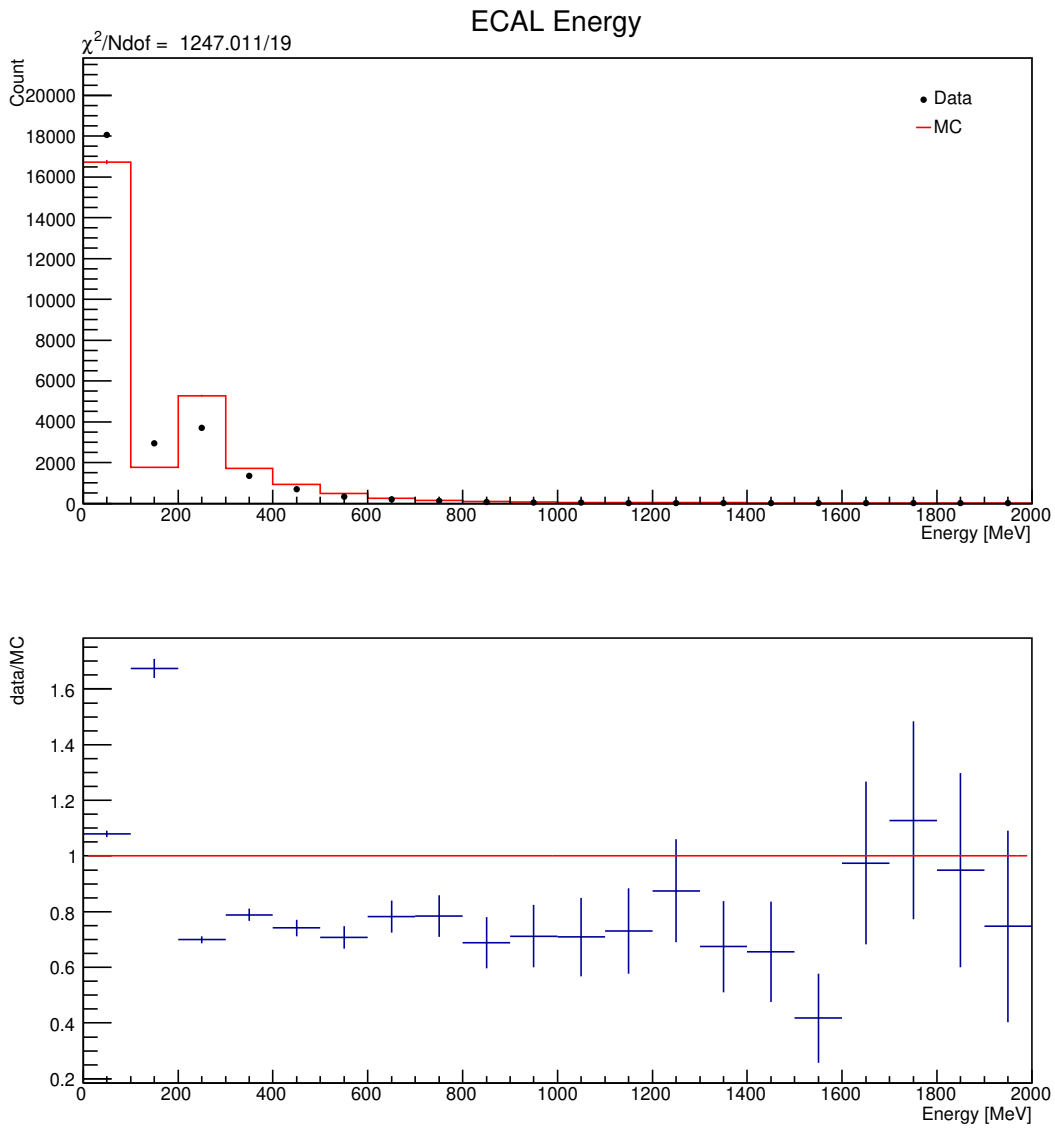


Fig. 7.11 Data/MC agreement for the ECAL energy variable. Top plot represents an overlay of a scaled MC prediction (red histogram) and data points (black circles) with their statistical uncertainty (black lines). Data/MC ratio with the corresponding statistical uncertainty is shown in the bottom plot (in blue) as well as the ideal ratio value of unity (in red).

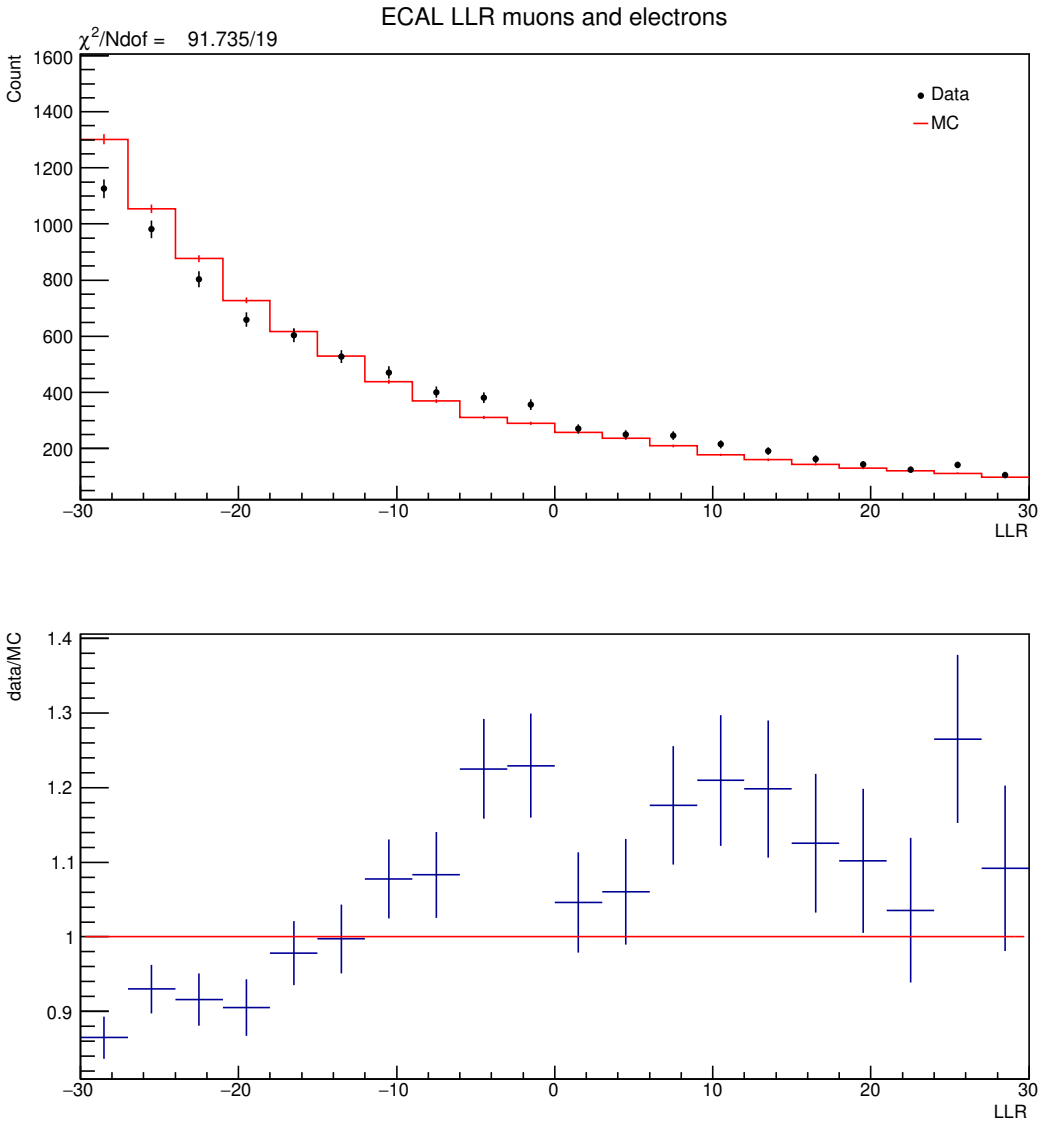


Fig. 7.12 Data/MC agreement for the ECAL LLR_Mip_Em variable. Top plot represents an overlay of a scaled MC prediction (red histogram) and data points (black circles) with their statistical uncertainty (black lines). Data/MC ratio with the corresponding statistical uncertainty is shown in the bottom plot (in blue) as well as the ideal ratio value of unity (in red).

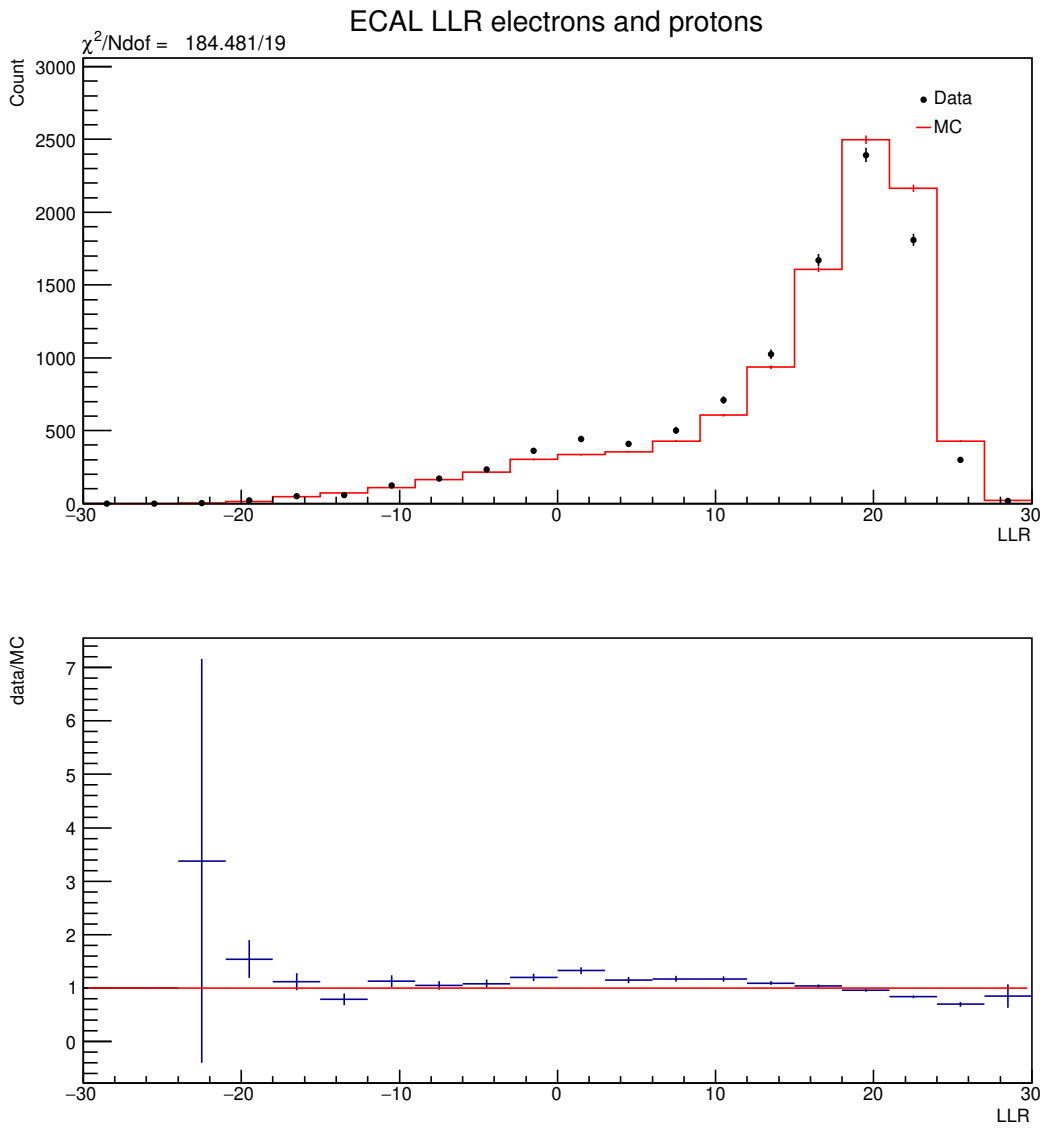


Fig. 7.13 Data/MC agreement plots for the LLR_Em_Hip variable. Top plot represents an overlay of a scaled MC prediction (red histogram) and data points (black circles) with their statistical uncertainty (black lines). Data/MC ratio with the corresponding statistical uncertainty is shown in the bottom plot (in blue) as well as the ideal ratio value of unity (in red).

Chapter 8

Event Selection

The measurement of the intrinsic $\overline{\nu}_e$ component in the $\overline{\nu}_\mu$ T2K beam is crucial for the oscillation analysis, as it represents the largest irreducible background for the ($\overline{\nu}_\mu \rightarrow \overline{\nu}_e$) appearance measurement at Super-Kamiokande. Furthermore, the near detector number of $\overline{\nu}_e$ events is used to constrain the flux and cross-section given detector efficiency, as described in Section 4.4.

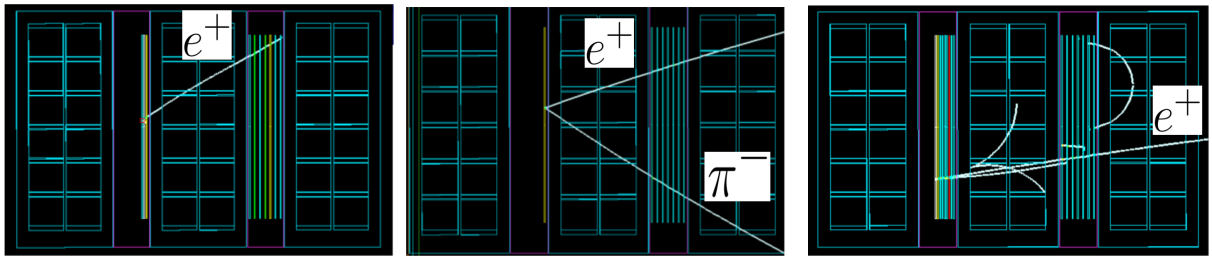
The $\overline{\nu}_e$ event selection will be summarized in this chapter. This work is based on [76], replacing the classical PID cuts with the BDT.

8.1 Signal Definition

In this analysis, the signal is defined as

- a selected highest momentum positive track representing a true positron,
- coming from Charged-Current (CC) $\overline{\nu}_e$ interaction, and
- starting inside FGD1 fiducial volume (FV) (interaction on carbon).

The aim of the selection is to choose events similar to those in Fig. 8.1 which shows CC0π and CC-Other interactions in ND280.



(a) $\overline{\nu}_e$ CC0π. (b) $\overline{\nu}_e$ CC1π. (c) $\overline{\nu}_e$ CCnπ.

Fig. 8.1 ND280 event display showing $\overline{\nu}_e$ CC0π (left), $\overline{\nu}_e$ CC1π (middle), and $\overline{\nu}_e$ CCnπ (right) interactions starting in FGD1 fiducial volume. In this analysis, we group the CC1π and the CCnπ into the CC-Other category.

8.1.1 Corrections to the Number of Events

The nominal number of events undergo two main corrections: the flux tuning and the pile-up correction. The flux tuning is based on the pion and kaon production in proton interactions from measurements by CERN's NA61 experiment [77]. The pile-up correction accounts for the neutrino interactions that happen outside the magnet, producing particles that enter the ND280 (sand interactions). These interactions are not modelled with the magnet MC, but are modelled in a separate simulation. For more details, please refer to [76, Sections 3.2 and 5].

8.2 Electron Anti-neutrino Selection Steps

Every event has to pass each of the following selection steps (in order) to be considered as a signal.

8.2.1 Event Quality Cut

The event quality combines the data quality and the beam checks. The data quality cut ensures that the collected data did not suffer from any detector hardware or calibration problems. The beam check ensures that the event timing is in accordance with the beam bunches described in Section 5.1.1.

8.2.2 TPC Number of Tracks Cut

The track's momentum and charge are measured using the TPC. Consequently, the selected event must have at least one TPC segment.

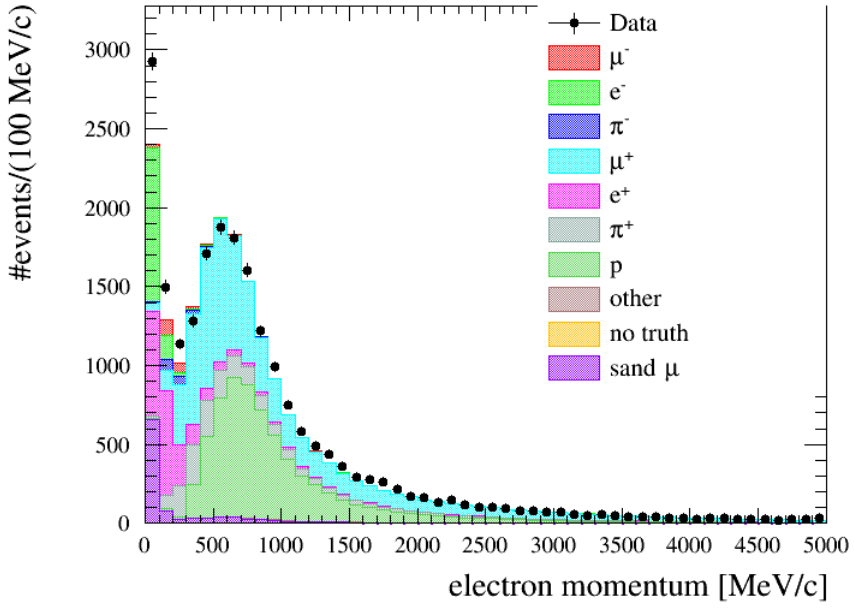
8.2.3 TPC Quality Cut

The TPC segment should consist of at least 18 hits if the track had an ECAL segment, and 36 hits otherwise. If the track crosses the whole TPC horizontally, we expect 72 hits. The number of hits in the TPC segment balances the TPC performance with its systematic uncertainties and was chosen based on previous studies [76].

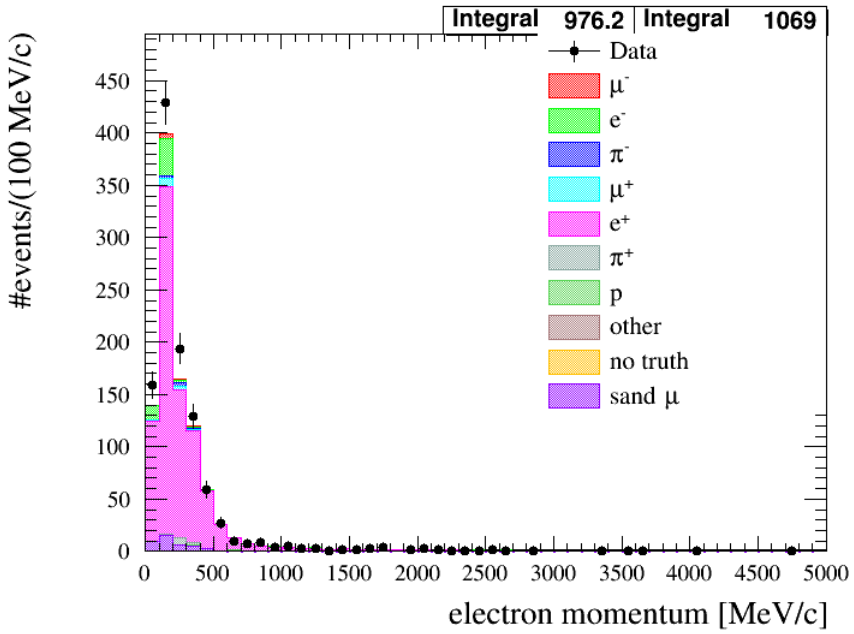
8.2.4 BDT Cut

The BDT works as an optimized cut in a multi-dimensional feature space, 10-D in this analysis, and uses the correlation between input variables to improve the separation power between signal and background. First, it calculates the test statistic (tBDT) for each track. It then checks if it is higher than a certain value (tBDT_cut). If so, it will be considered as a signal, otherwise as a background. The tBDT_cut is obtained from the training phase by optimizing a Figure Of Merit (FoM), e.g. maximizing purity $\times$ efficiency, as described in Section 6.3.3.

Figure 8.2 shows the effect of the BDT cut on the selection, where it successfully removes most of the proton background. This was a subtle problem that was not fixed using the classical cuts due to the overlap of the $\frac{dE}{dx}$ curves of protons and positrons at the 1 GeV/c momentum range [76, cf.Section 3.5].



(a) Before BDT.



(b) After BDT.

Fig. 8.2 Particle momentum distribution before (top) and after (bottom) the BDT cut.

Although the BDT was able to successfully select the positrons, most of these positrons are arising from gammas and not from the electron-antineutrino signal, as shown in Figs. 8.3 and 8.4.

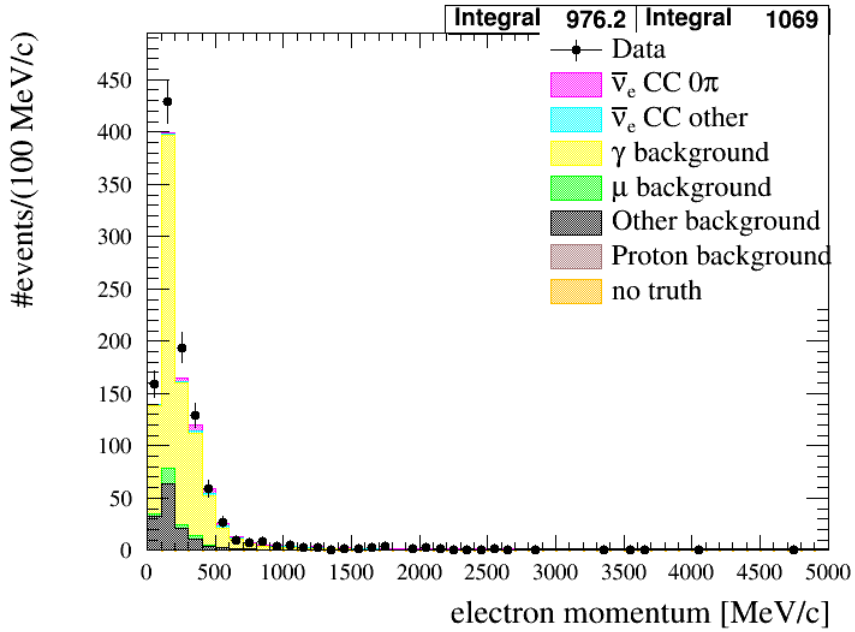
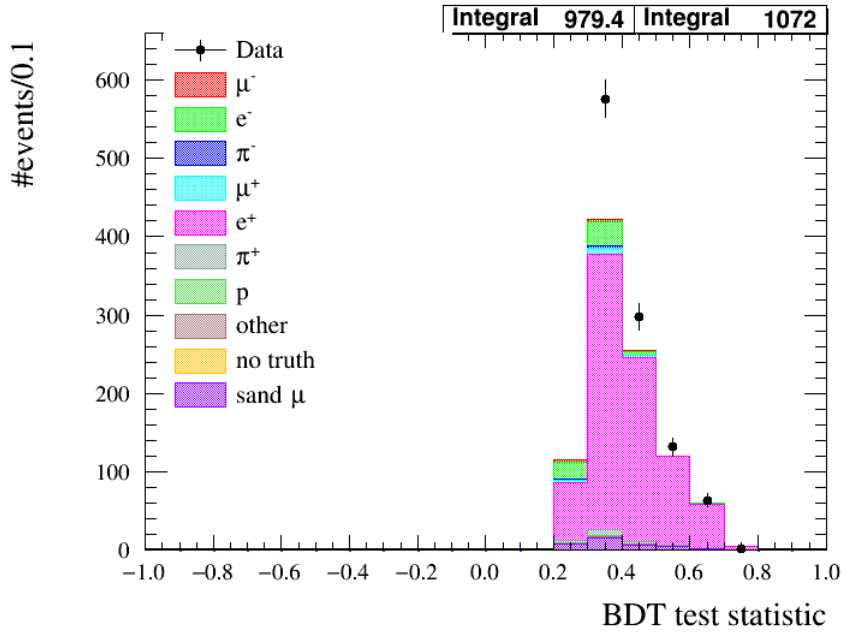
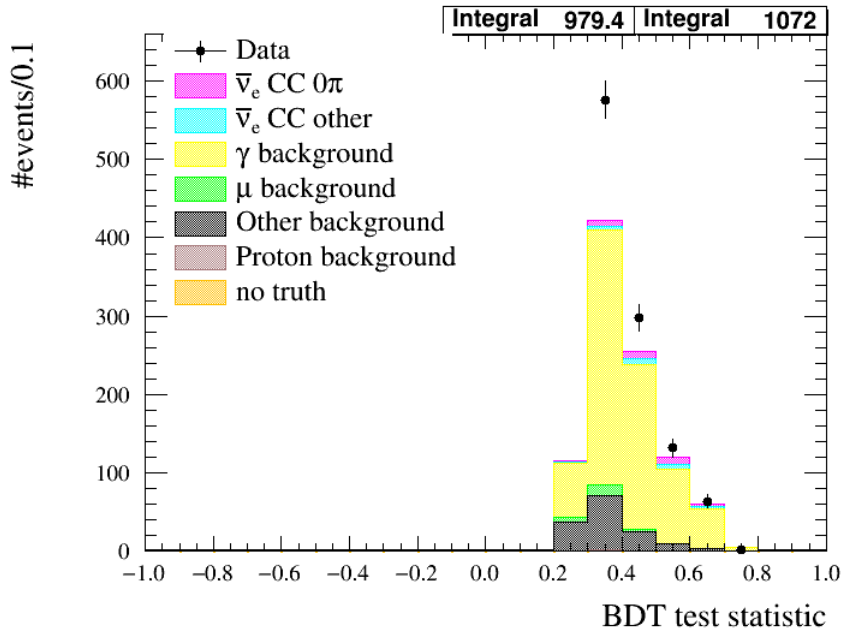


Fig. 8.3 Momentum distribution for different interaction topologies after the BDT cut.



(a)



(b)

Fig. 8.4 The BDT test statistic distribution after the cut, decomposed into its (particle) constituents in (a) and its (topology) constituents in (b).

8.2.5 Momentum Cut

One of the major backgrounds is formed by the positrons or electrons coming from photon conversion. The low momentum region of the phase space is highly contaminated with these particles. Consequently, a momentum cut was introduced, as shown in Fig. 8.5, that rejects any event with momentum less than 200 MeV and does not have Time of Flight (ToF) information.

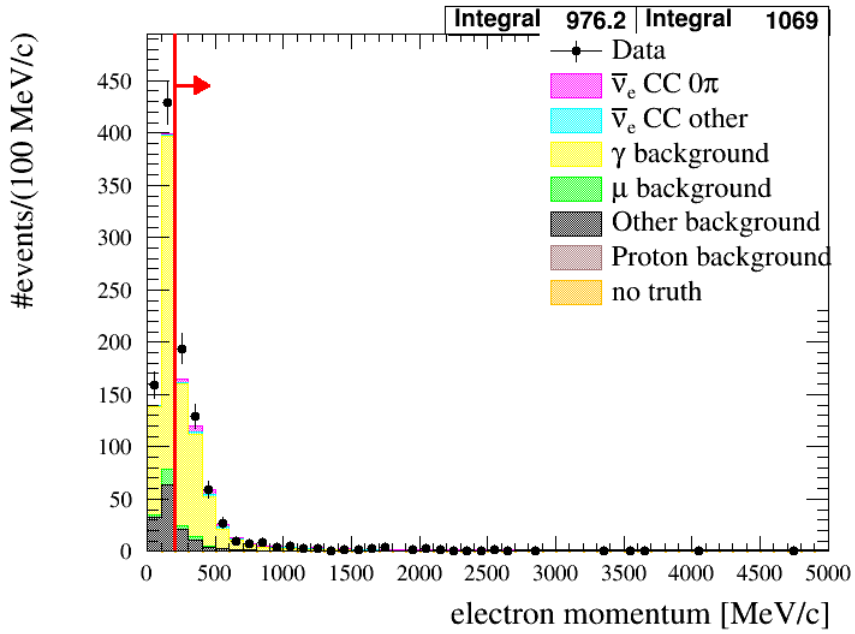


Fig. 8.5 Momentum cut introduced to reduce the gamma background contamination.

The interaction topology distribution after applying the momentum cut is shown in Fig. 8.6.

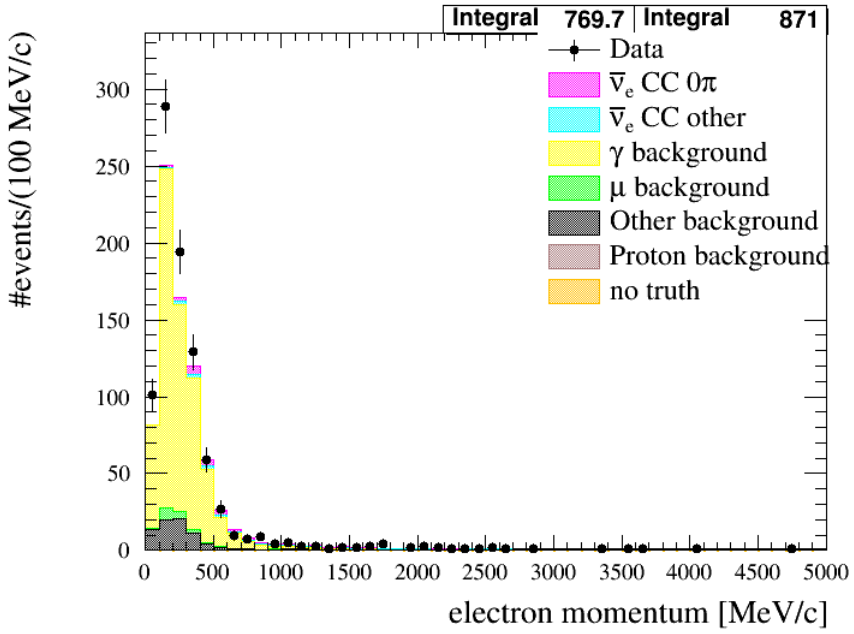


Fig. 8.6 The interaction topology distribution after the momentum cut.

8.2.6 Most Energetic Positive Track

This cut ensures that the selected highest momentum positive track is the most energetic track in the TPC. That is, if there exists another TPC track with a momentum higher than the selected one, the event is discarded.

This cut is based on the study of ν_e CC interaction in [78, Section 2.4], where it was found that 99% of the neutrino events fulfilled this requirement. The interaction topology distribution after applying this cut is shown in Fig. 8.7.

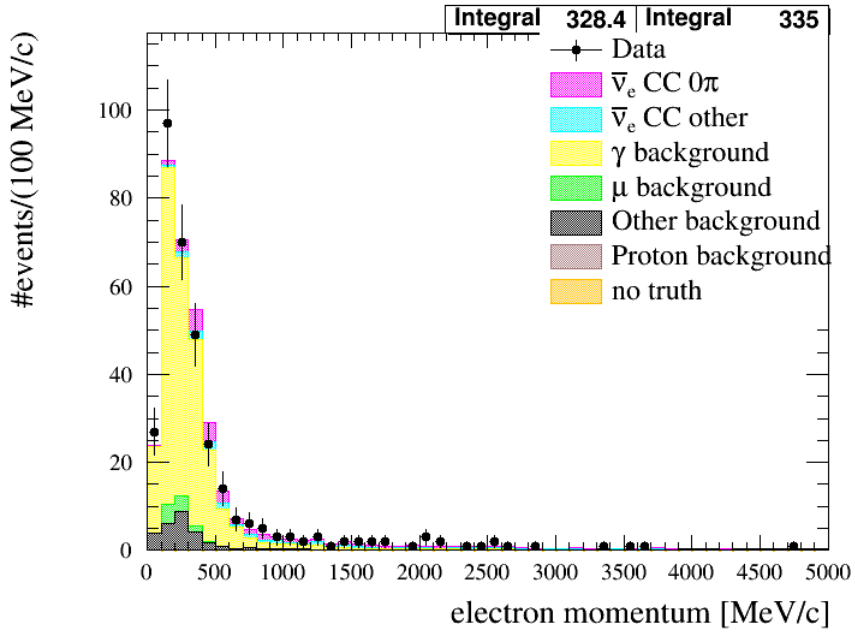


Fig. 8.7 The most energetic positive track cut effect on the data and MC.

8.2.7 TPC Veto

This cut is designed to reduce the gamma background, $\gamma \rightarrow e^- + e^+$, coming from Bremsstrahlung[1], or from pion production due to a neutrino interaction, that happened before the reconstructed vertex.

The Z-position difference between the main track and the second most energetic track in the TPC is checked. If the difference is less than 3 cm, the main track started after the second most energetic one, then the event is rejected. For instance, a neutrino interaction in the barrel ECAL may produce a neutral pion that decays later into two photons, and one of these photons is converted into an electron-positron pair in the FGD. In this case, the second most energetic track could be that of the original neutrino interaction and the selected track could be that of the converted gamma. The TPC veto removes such events.

[1]Bremsstrahlung is electromagnetic radiation produced by the deceleration of a charged particle when scattered off another charged particle. The decelerated particle loses kinetic energy in the form of radiation. cf.[79, Section 15.2]

Figure 8.8 shows the effect of this veto on the rejection of gamma background.

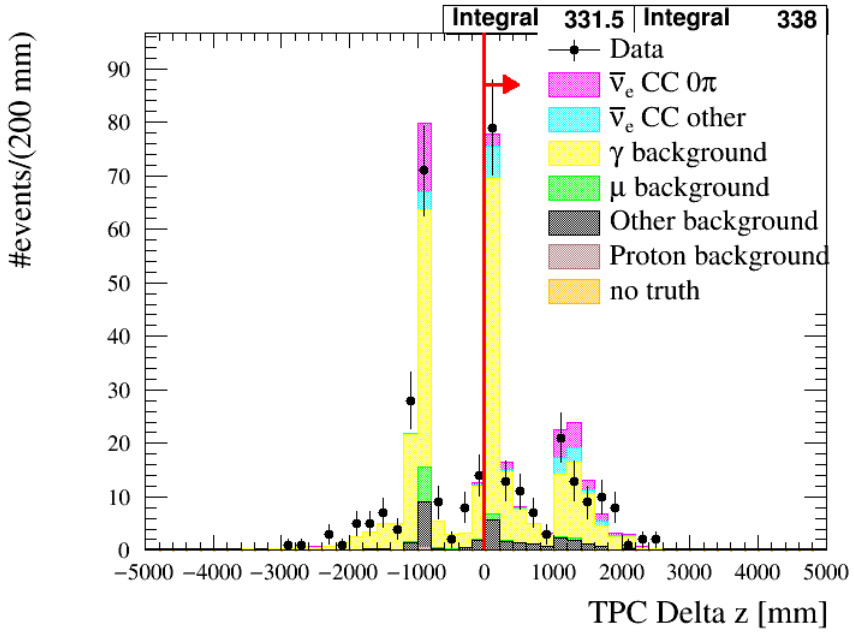


Fig. 8.8 Effect of the TPC veto cut on the gamma background rejection.

The interaction topology distribution after the TPC veto is shown in Fig. 8.9.

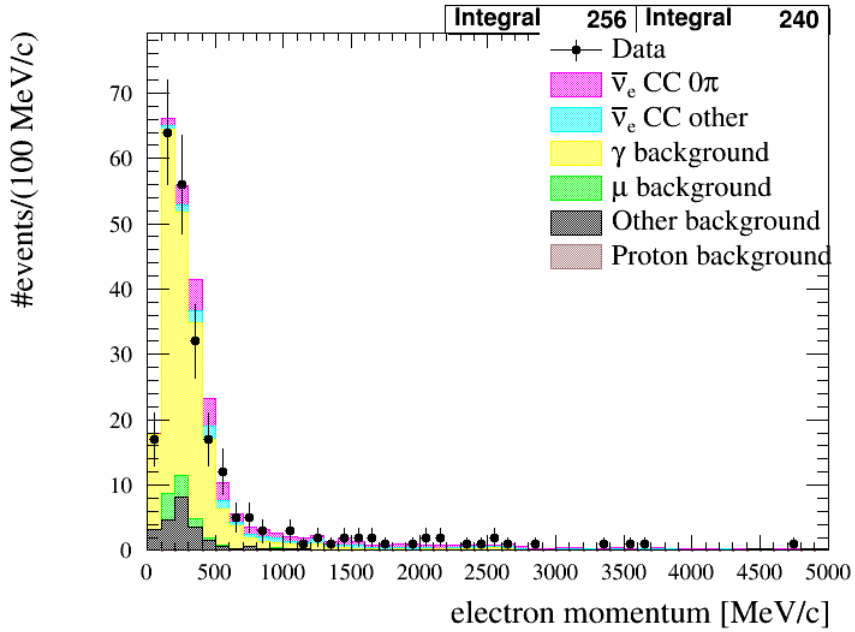


Fig. 8.9 The interaction topology distribution after the TPC veto.

8.2.8 Gamma Isolation

This cut removes any event with a second vertex in FGD1. The second vertex may come, for instance, from a neutral pion decaying into two photons and each photon then decays into an electron and positron pair inside the FGD. The interaction topology distribution after applying the gamma isolation cut is shown in Fig. 8.10.

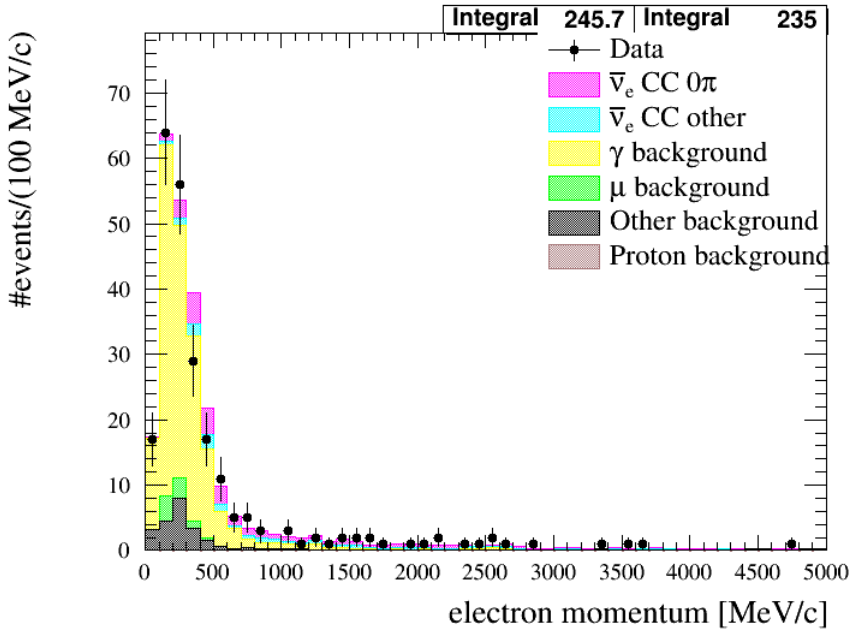


Fig. 8.10 The interaction topology distribution after the gamma isolation cut.

8.2.9 External FGD1

This cut is applied to reject events with mis-reconstructed tracks. Sometimes, the reconstruction software breaks the tracks into two components: one fully contained in the FGD FV and another starting from the last layers of the FGD. We therefore reject events with tracks starting at the last layers of the FGD if there was a fully contained track in the FGD. The interaction topology distribution after applying the external FGD1 layers cut is shown in Fig. 8.11.

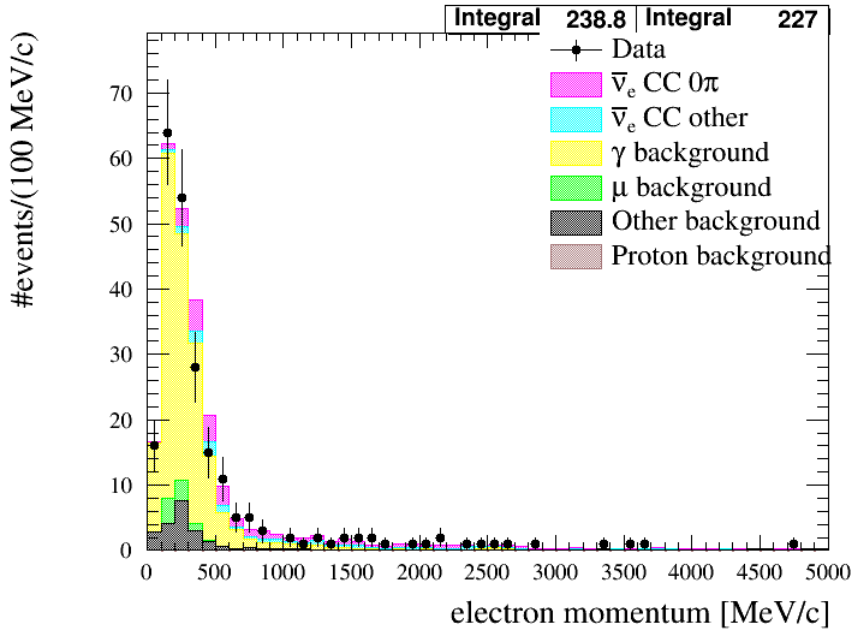


Fig. 8.11 The interaction topology distribution after the external FGD1 veto.

8.2.10 Gamma Invariant Mass Veto

The gamma invariant mass veto is introduced to supplement the gamma background rejection. An event will be rejected if the secondary track fulfills the following four checks:

1. it has a good TPC quality (> 18 TPC hits);
2. it passed a relaxed TPC electron pull (δ_e) test, $-3 < \delta_e < 3$;
3. it started within 10 cm away from the main track starting point; and
4. its measured momentum, under the electron track assumption, with that of the main track, produces an invariant mass less than 110 MeV/c^2. The calculated invariant mass M_{inv} is given by

$$M_{\text{inv}} = \sqrt{2m_e^2 + 2(E_1E_2 - \vec{p}_1 \cdot \vec{p}_2)}, \tag{8.1}$$

where m_e is the mass of the electron, E_i and p_i are the energy and momentum of particle i respectively, with $i = 1$ or 2.

Figure 8.12 shows the effect of this veto on the rejection of gamma background.

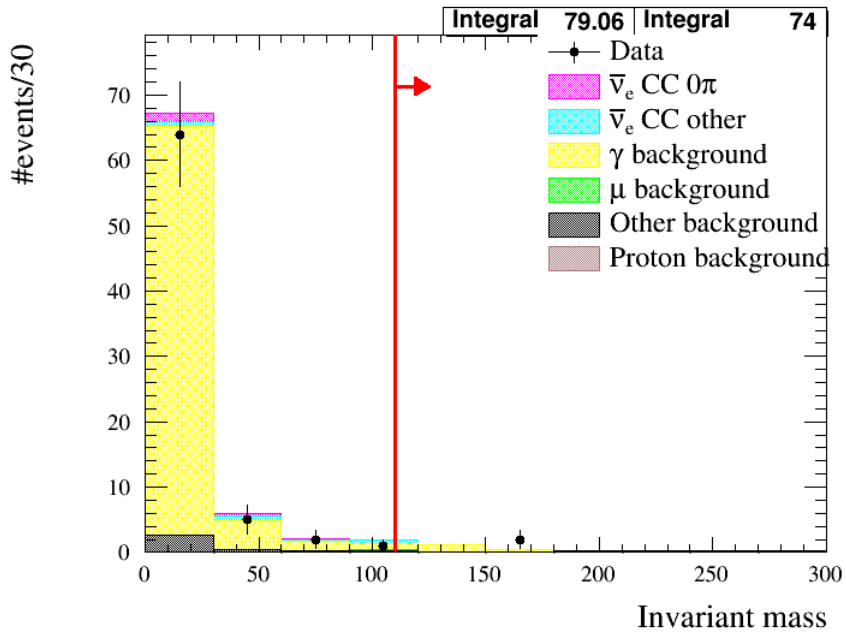


Fig. 8.12 Effect of the gamma invariant mass veto cut on the gamma background rejection.

The interaction topology distribution after the gamma invariant mass veto is shown in Fig. 8.13.

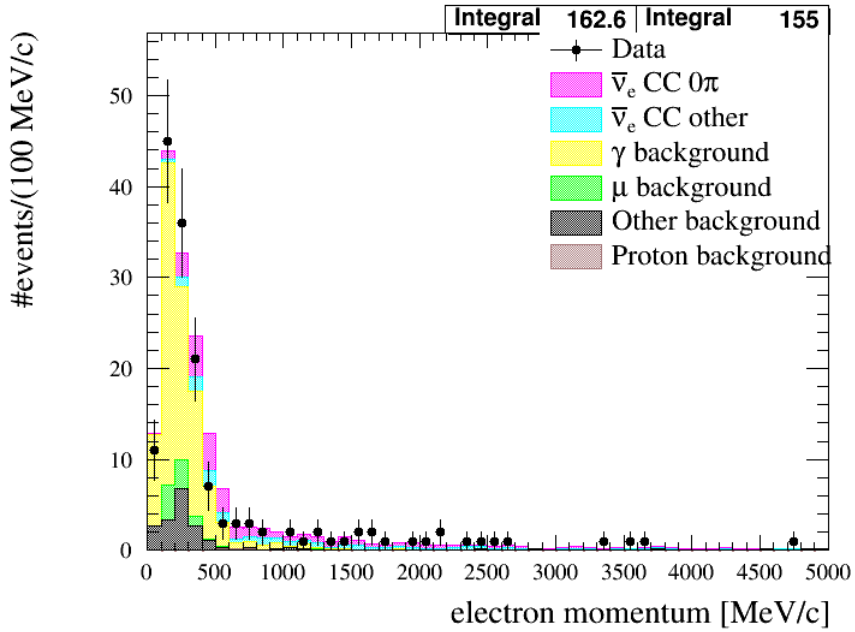


Fig. 8.13 The interaction topology distribution after the gamma invariant mass veto.

8.2.11 PØD Veto

This cut is designed to reject electrons and positrons coming from gammas produced by pions, e.g. $\pi^0 \rightarrow \gamma + \gamma$. Consequently, an event is rejected if the PØD or PØD-ECAL recorded any activity within the same beam bunch of the main track. Figure 8.14 shows the effect of this veto on the rejection of gamma background.

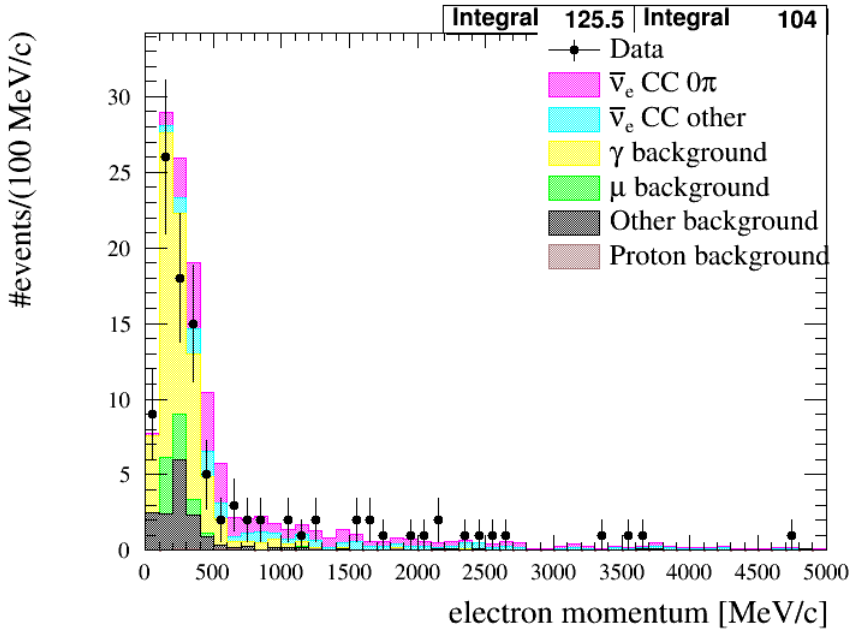


Fig. 8.14 Effect of the PØD veto cut on the gamma background rejection.

8.2.12 ECAL Veto

This cut is similar to Section 8.2.7. Here, we compare the Z-position of the most upstream ECAL cluster with the Z-position of the main track. If the difference is less than 10 cm, in other words if the ECAL activity was before the main track, the event is rejected. The effect of this veto on the rejection of gamma background is shown in Fig. 8.15.

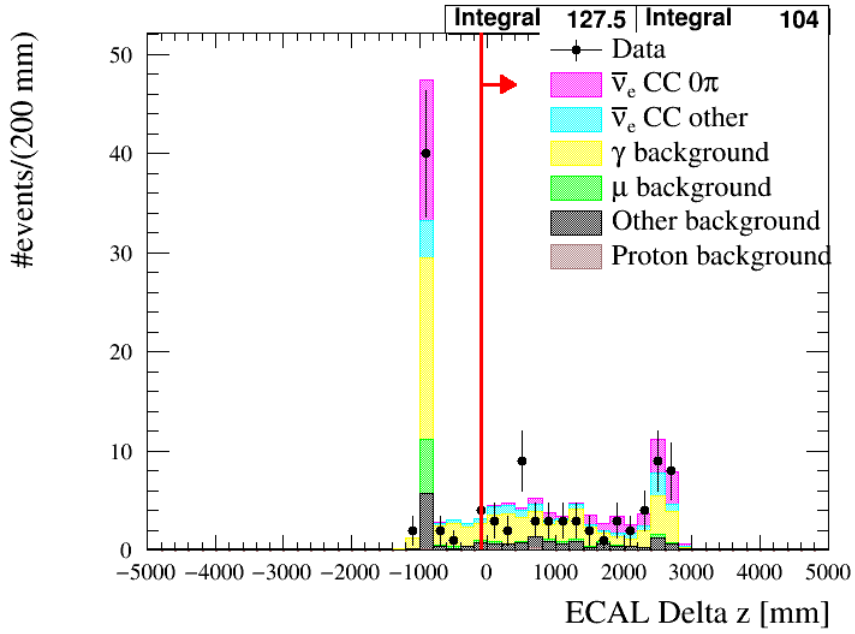


Fig. 8.15 Effect of the ECAL veto cut on the gamma background rejection.

The interaction topology distribution after the ECAL veto is shown in Fig. 8.16.

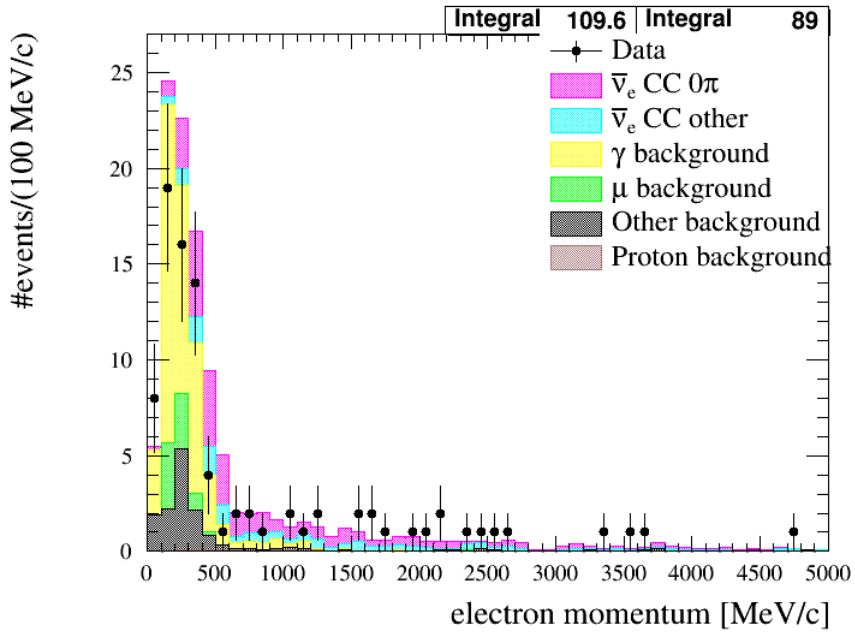


Fig. 8.16 The interaction topology distribution after the ECAL veto.

8.2.13 Time of Flight

For tracks that have an ECAL segment, this cut makes sure that the selected track's Time of Flight (ToF) to the ECAL is positive. That is, the track starts first in FGD1 fiducial volume then travels to the ECAL. The interaction topology distribution after the ToF is shown in Fig. 8.17.

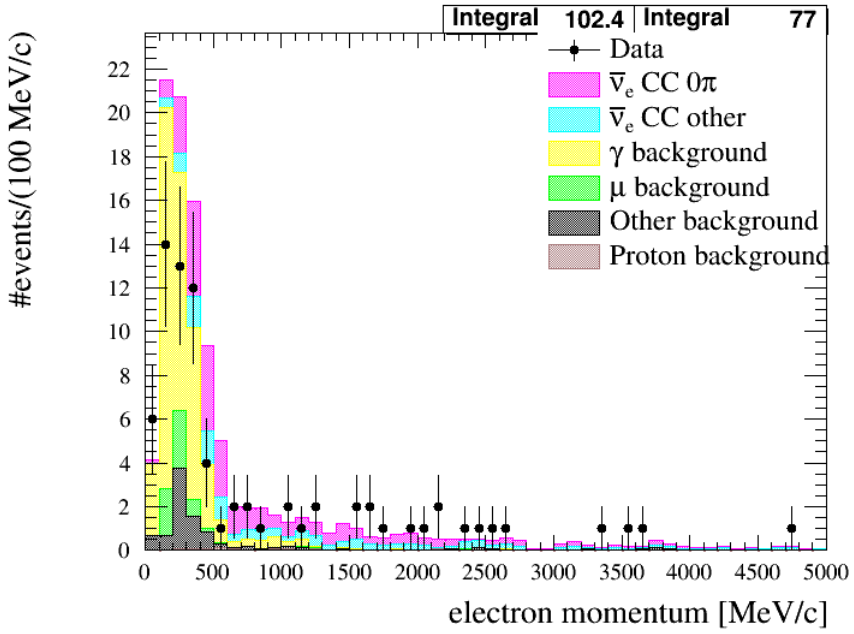


Fig. 8.17 The interaction topology distribution after the ToF.

8.2.14 Momentum Quality

This cut is introduced to reject tracks with a mis-reconstructed higher momentum. If the selected track had a momentum larger than 200 MeV/c and a negative muon or pion TPC pull, it will be far away from the TPC electron hypothesis, and hence, the event will be discarded. The interaction topology distribution after the momentum quality cut is shown in Fig. 8.18.

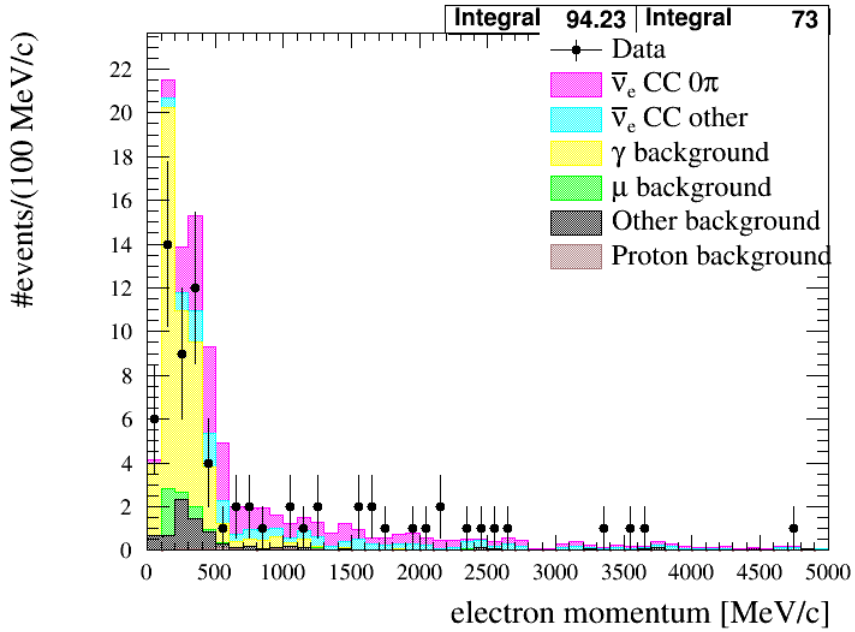
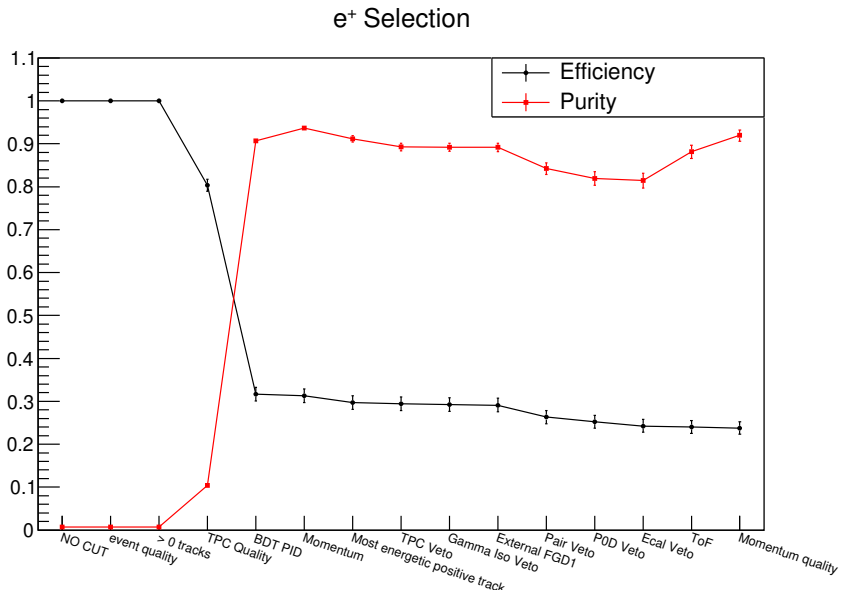


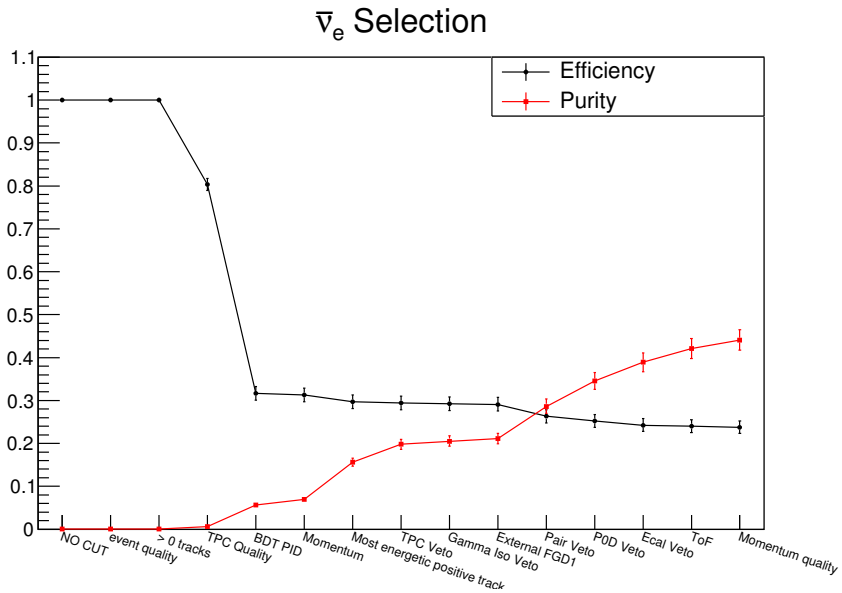
Fig. 8.18 The interaction topology distribution after the external FGD1 veto.

8.3 Selection Performance

All of the selection steps have been designed using MC simulation, and show a good agreement with the data. The selection efficiency and purity as a function of each step are shown schematically in Fig. 8.19, where the BDT cut massively enhanced the purity of the positron selection, as shown in Fig. 8.19a. Although the veto cuts reduced the positron efficiency and purity, they have increased the purity of the $\overline{\nu}_e$ selection by removing positrons coming from gammas interactions, as shown in Fig. 8.19b.



(a) Positrons.



(b) $\overline{\nu}_e$ CCinc interaction.

Fig. 8.19 Efficiency (black curve) and purity (red curve) as functions of selection steps for positron (top) and $\overline{\nu}_e$ inclusive charged-current interaction (bottom) selections.

Furthermore, the purity and efficiency as a function of the true particle momentum in the MC simulation are shown in Fig. 8.20. Ideally, the selection efficiency should not be biased to a certain region of the phase space. However, the selection suffers from a

low efficiency at low particle momentum, due to detector limitations. For instance, a low momentum positron may not have the necessary TPC number of hits or may suffer from charge mis-reconstruction. The reconstruction software used in this analysis was not good at reconstructing low momentum tracks that spiral back on themselves.

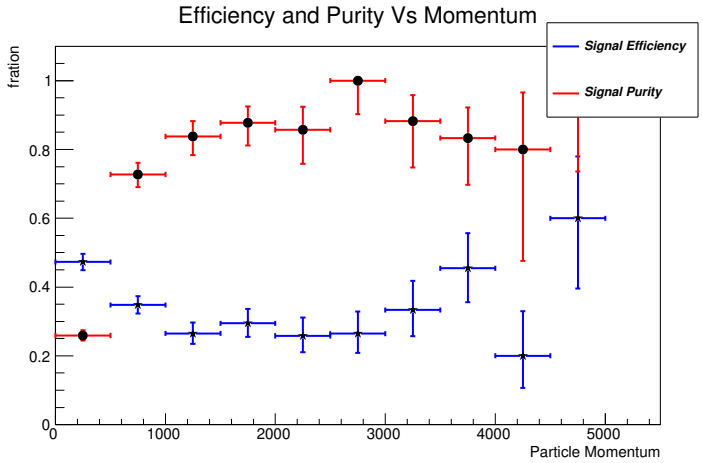
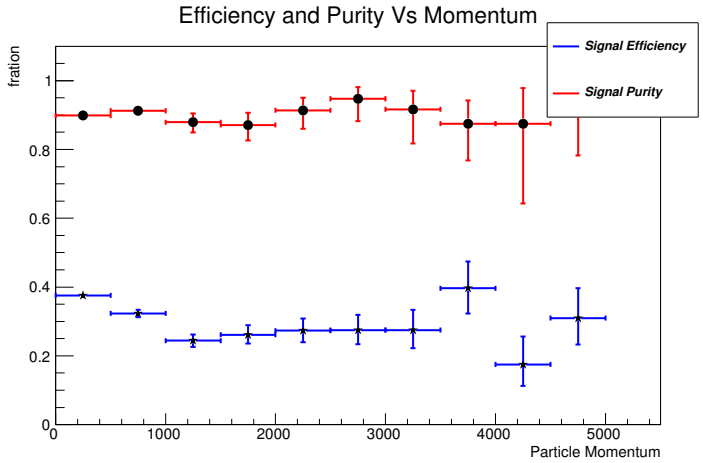
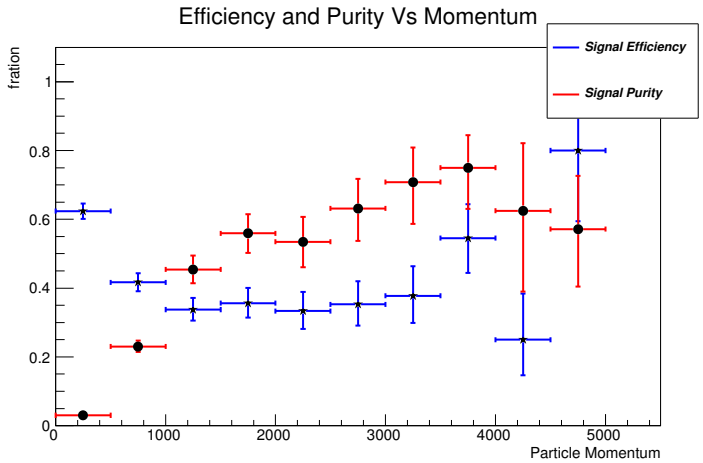


Fig. 8.20 Efficiency (in blue) and purity (in red) of the event selection as a function of the true momentum of the particle .

The selection efficiency and purity, due to the newly introduced BDT cut, as a function of the particle momentum are shown in Fig. 8.21.



(a) Particle Selection.



(b) Interaction Selection.

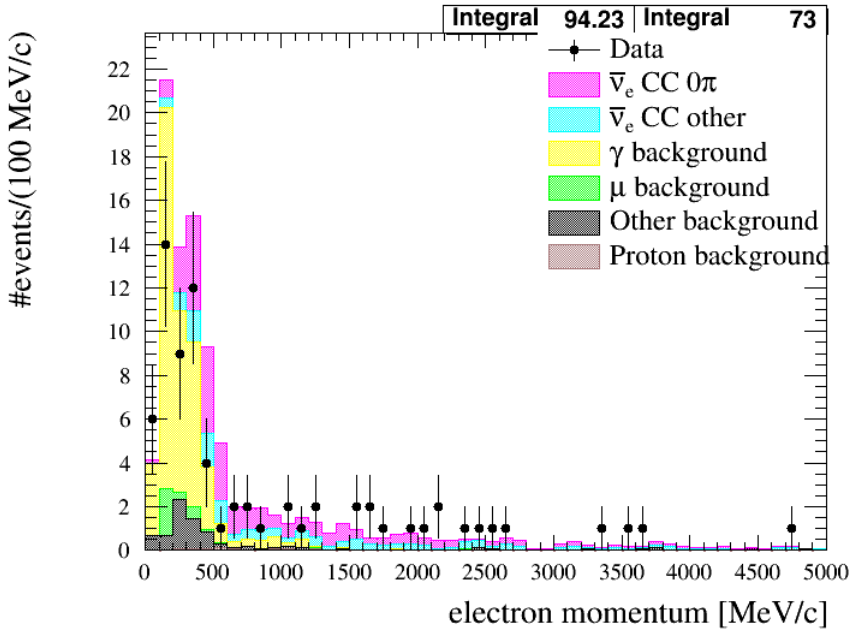
Fig. 8.21 Efficiency (in blue) and purity (in red) of the positron (top) and $\overline{\nu}_e$ charged-current interaction (bottom) selection as a function of the particle's momentum.

Table 8.1 summarizes the performance enhancement due to each step as well as the data to MC agreement.

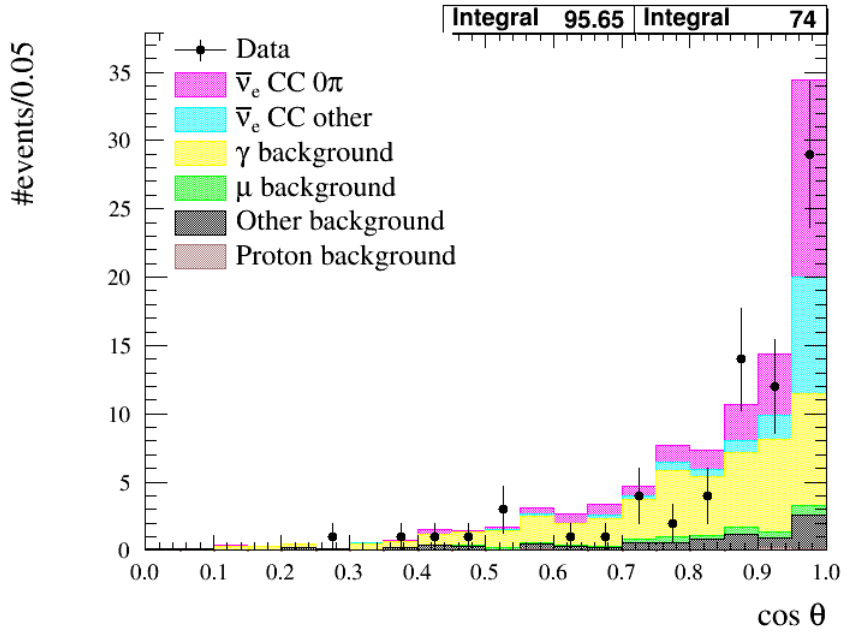
Cut Name	Data Events	MC Events	Data/MC	Purity (MC)	Efficiency (MC)
TPC Quality	21711	20581	1.05	0.01	0.80
BDT	1069	976	1.09	0.06	0.0.32
Momentum	871	769	1.13	0.07	0.31
Most Energetic Positive Track	335	328	1.02	0.16	0.30
TPC Veto	240	256	0.94	0.20	0.29
Gamma Isolation	235	245	0.96	0.21	0.29
External FGD1	227	238	0.95	0.21	0.29
Gamma Invariant Mass Veto	155	162	0.96	0.29	0.26
PØD Veto	104	125	0.83	0.35	0.25
ECAL Veto	89	109	0.82	0.39	0.24
ToF	77	102	0.75	0.42	0.24
Momentum Quality	73	94	0.78	0.44	0.24

Table 8.1 Efficiency, purity and Data/MC agreement for each selection step.

The momentum and angular distribution of the MC, superimposed with the data, for events that passed all the cuts are shown in Fig. 8.22.



(a) Momentum distribution.



(b) $\cos\theta$ distribution.

Fig. 8.22 Momentum (top) and $\cos\theta$ (bottom) distributions for events passing all cuts. Data (in black dots) with its statistical uncertainty overlays the MC simulation (in coloured histograms).

8.4 Surviving Background

Even after all the selection cuts, explained above, some backgrounds can still survive. The sources, decomposition and size expectation of the main backgrounds will be provided here.

8.4.1 Proton

The ionization energy loss dE/dx for the protons at the 1 GeV energy range overlaps with that of the positrons, making it a hard to remove the background, cf. [76]. Here, the BDT, using information from different sub-detectors, was able to successfully remove most of the proton background, showing better performance than the standard cut selection.

Most of these protons come from neutrino interactions, e.g. $\nu_\mu + n \rightarrow \mu + p$, or from resonances and DIS. Figure 8.23 shows the proton background which passed all selection cuts, split into its sources.

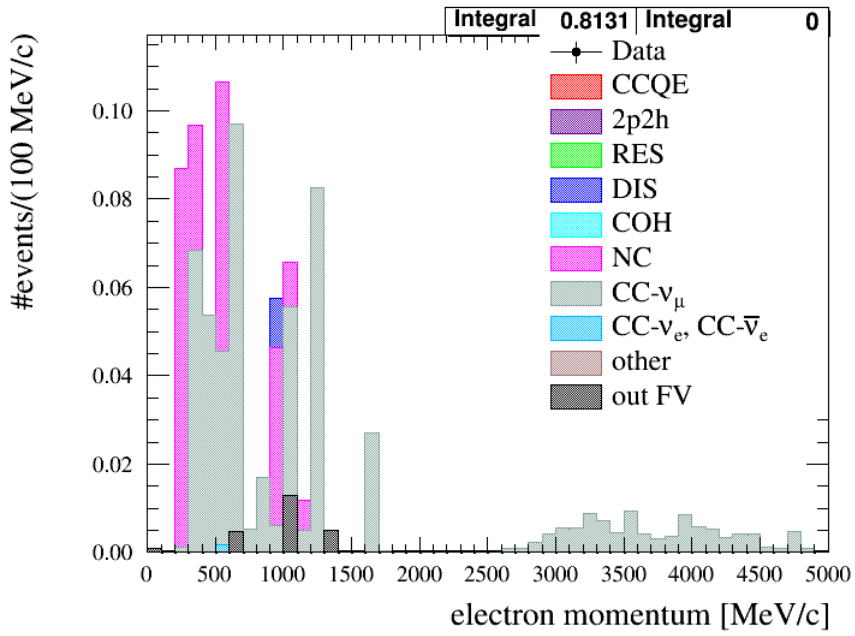


Fig. 8.23 The proton background, after applying all the selection cuts, identified by its sources.

8.4.2 Muons

Most of the muons come from neutrino interactions, e.g. $\overline{\nu}_\mu + p \rightarrow \mu^+ + n$, in which the muon did not have an ECAL segment or had TPC pulls comparable with that of the positrons. The muon background, after applying all selection cuts, is shown in Fig. 8.24.

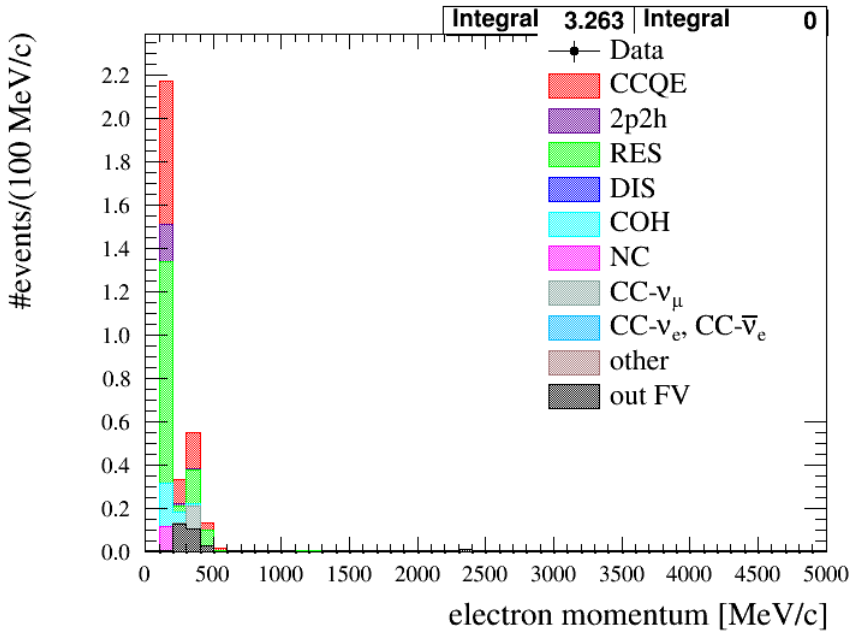


Fig. 8.24 The muon background, after applying all the selection cuts, identified by its sources.

8.4.3 Gamma

Gamma backgrounds are the most challenging to identify and remove. The neutrino interaction vertex can be either inside the FGD Fiducial Volume (FV), out of the FV (OOFV), or completely out of the FGD (OOFGD), but the positron starting point (the assumed vertex position) must be inside the FGD FV.

The main sources of these gammas are the neutral pion production from a neutral current neutrino interaction (e.g. NC1π^0), or a charged-current neutrino interaction (CC1π^0), in which the selected main track originates from the gamma conversion. In this situation, neutral

pions can decay into two photons in the reaction $\pi^0 \rightarrow \gamma + \gamma$. Subsequently, each photon can be converted into an electron-positron pair, mimicking the signal definition.

The background due to gammas is categorized into four groups (see Table 8.2) based on the true neutrino interaction vertex location (InFV, OOFV or OOFGD) and the true gamma conversion point.

Neutrino interaction vertex	Gamma conversion point	Reconstructed (assumed) vertex	Background category
FV	FV	FV	InFV gamma
OOFV	FV	FV	OOFV gamma
OOFGD	FV	FV	OOFGD gamma
OOFV or OOFGD	OOFV or OOFGD	FV	Others

Table 8.2 Categorization of the gamma background based on the neutrino interaction vertex location (source of π^0), the gamma conversion point and the reconstructed vertex.

This categorization helps to better understand the sources of gamma background and is essential for the systematics evaluation. For example, the OOFGD gammas, coming from different sub-detectors, will have different flux and detector systematics. The gamma background from the events in the main selection is shown in Fig. 8.25.

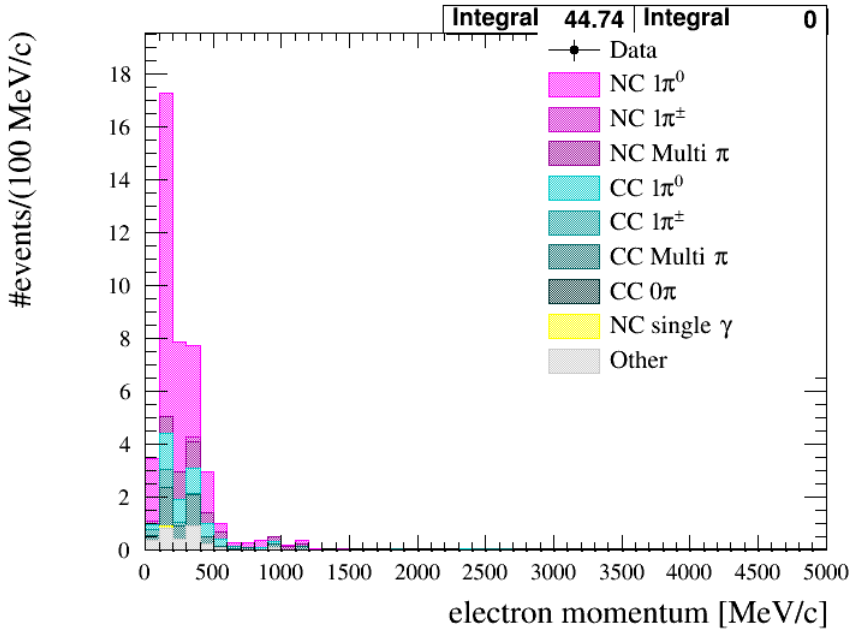


Fig. 8.25 The gamma background, after applying all the selection cuts, decomposed by its sources.

8.4.4 Other

All backgrounds that do not fall into one of the above categories are denoted “other”. The “other” backgrounds consist mainly of misclassified pions and OOFV gammas for which the gamma’s conversion point was outside the FGD FV, but has been mis-reconstructed to be inside the FV (see Fig. 8.26).

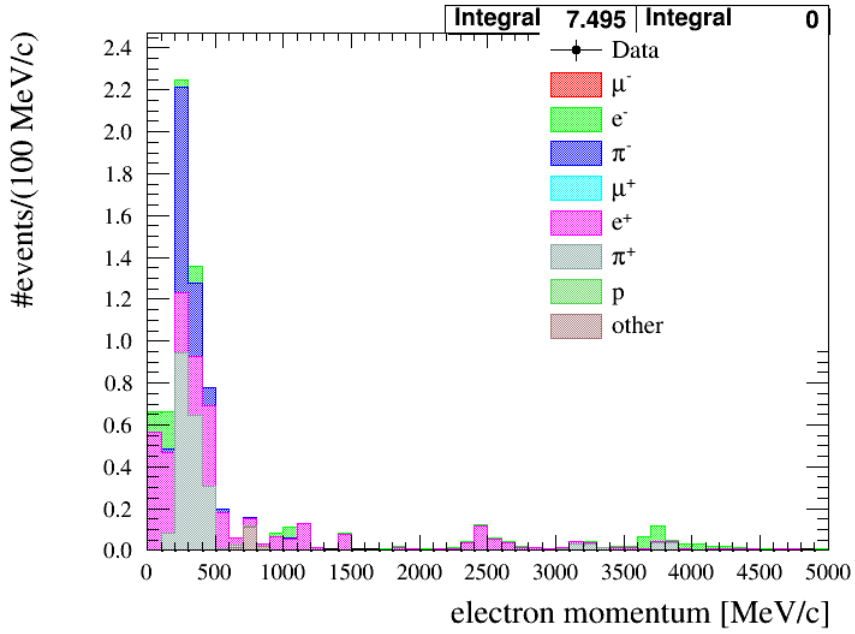


Fig. 8.26 The other background, after applying all the selection cuts.

The selection purity for the MC, scaled to the appropriate data's Protons On Targets (POT) for ND280 RHC runs 5,6 and 7, was $\sim 40\%$. The contributions of each of the backgrounds are summarized in Table 8.3.

Background	size
Gamma	47.6%
Muons	3.4%
Protons	0.9%
Other	8%

Table 8.3 Background decomposition for the $\overline{\nu}_e$ selection.

Chapter 9

BDT Systematics

In this chapter, we examine the BDT test statistics and performance systematics. The rest of the systematic uncertainties for an electron anti-neutrino cross section measurement will be described in Chapter 12.

Important questions in the multivariate classification technique are how to address systematic uncertainties and whether multivariate techniques are more prone to systematics uncertainties than standard cut selections.

Firstly, there is no substantial difference between the estimation of systematic uncertainties in multivariate classification techniques and classical cut analysis. Multivariate classifiers themselves do not introduce any new systematic uncertainties. Even when trained on a wrong MC simulation, they may perform poorly, the same as a poorly tuned classical cut analysis [80].

Secondly, the multivariate techniques can easily identify mis-modelled variable correlations (systematics), while standard cuts need a meticulous study of cut flows to detect these, cf. [81, Section 5.5 and 8.4.3.1].

In order to minimize the systematic uncertainties entering the BDT selection, only variables with good Data/MC agreement and known systematic uncertainties are chosen as inputs, cf. Section 7.4. Therefore, other variables, including weighted track length (a track

length weighted by the average density of the traversed matter) and the number of tracks in an event, were omitted, even though they have a high discrimination power because of their model dependence.

Two systematic measurement approaches are studied for the BDT. One is the effect of the input variables' systematic uncertainties on the BDT output distribution (tBDT). It is important to show the error propagation of the input variables which provides an indirect measurement of the BDT performance. The second approach is to evaluate the effect of the input variables' systematics on the performance of the trained BDT or, more precisely, on the purity and efficiency of the selection.

To achieve these goals, the BDT output distribution and training performance (purity and efficiency of the selection) are calculated using test samples of input variables with all possible systematic variations.

9.1 Uncertainty on the BDT Output Distribution

9.1.1 Data/MC Agreement

So far, only the detector systematics were taken into consideration, since we are considering the same Monte Carlo (MC). The theoretical model and background systematics are studied by overlaying the BDT test statistic (tBDT) distribution for the simulated signal and background with the tBDT distribution calculated for the data (see Fig. 9.1).

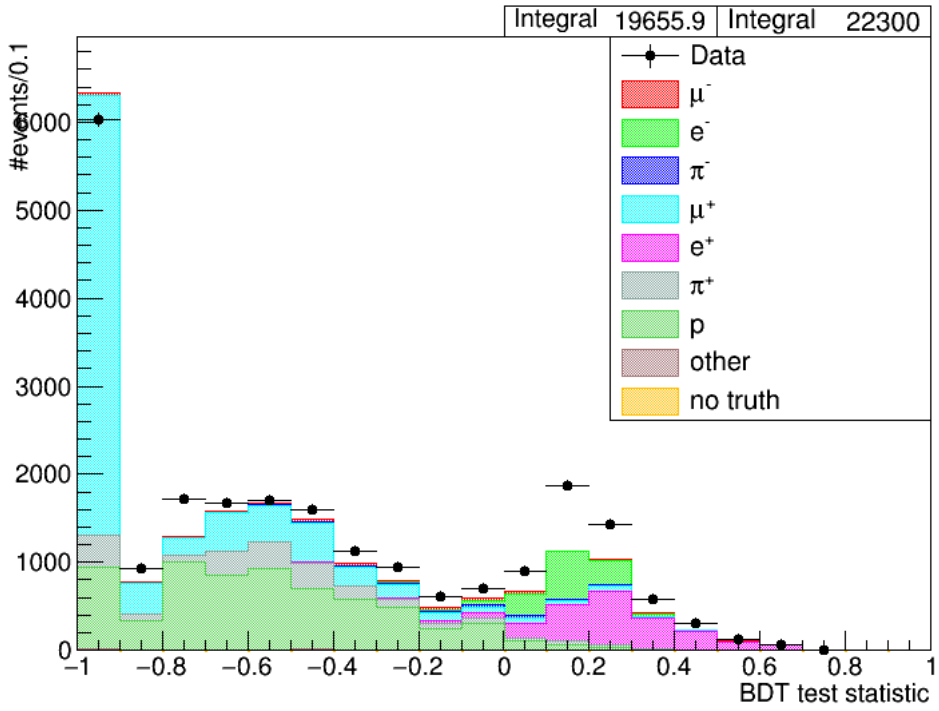


Fig. 9.1 BDT test statistic distribution for data and MC.

In this analysis, we do not use the Data/MC agreement as a measurement for the systematic uncertainty, but we propagate the well known error of the input variables to the BDT output variable (see Section 9.1.2).

9.1.2 Test Statistic Uncertainty

The test statistic (tBDT) uncertainty is quantified using 500 toy-experiments. Each toy-experiment has a modified version of the BDT's input variables. We use the uncertainty on the measured variables to build a Gaussian distribution with a mean equal to the nominal value and a sigma equal to the systematic uncertainty. The tBDT is calculated for every event in each toy experiment as well as the difference between the toy and the nominal tBDT. Finally, we build the distribution of the tBDT difference (Δ tBDT). In this case, the bias

and systematics uncertainty estimates are represented by the mean and width of the Δ tBDT distribution, respectively.

The propagation of the BDT input systematics to the BDT test statistic for the signal (positron), background (any other particle) and their sum (all particles) is shown in Figs. 9.2 to 9.4.

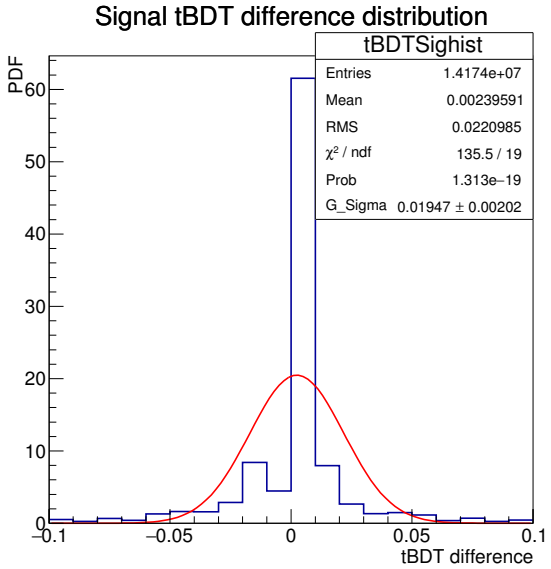


Fig. 9.2 BDT test statistic difference distribution histogram for the signal (positrons) in blue, and a fit to a Gaussian function in red.

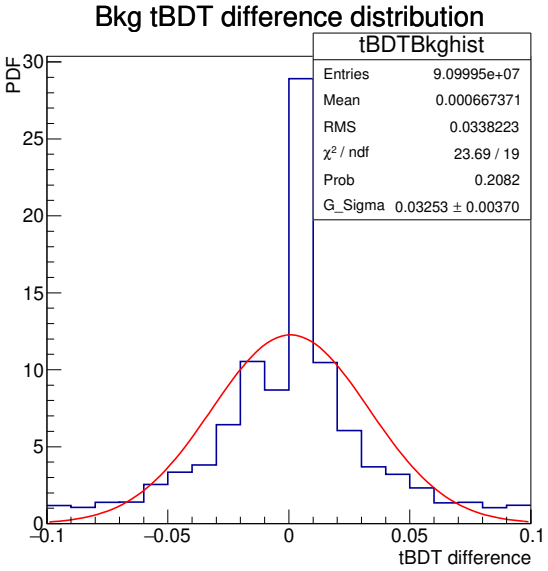


Fig. 9.3 BDT test statistic difference distribution histogram for the background (all particles except positrons) in blue, and a fit to a Gaussian function in red.

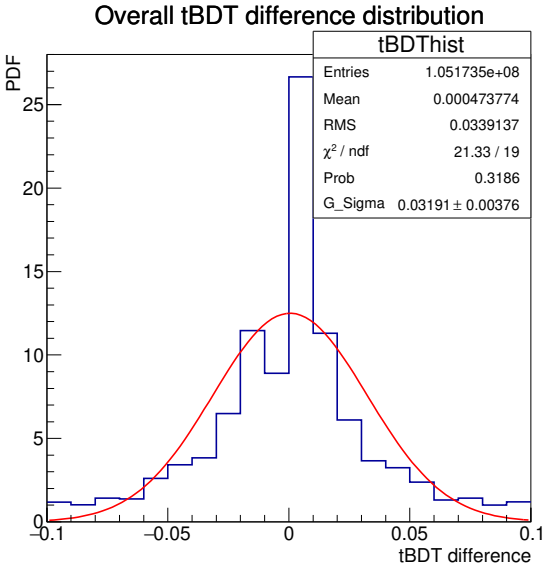


Fig. 9.4 BDT test statistic difference distribution histogram for the signal and the background (all particles) in blue, and a fit to a Gaussian function in red.

The means of each of the tBDT differences is almost zero. Hence, there is no sign of a bias and there is no need for a correction to be applied to the calculated tBDT. The

signal tBDT difference histogram is highly peaked around zero, resulting in a poor fit to a Gaussian function. Moreover, the width of the tBDT for the signal case is less than that of the background, indicating that the tBDT uncertainty could be a function of the calculated test statistic itself.

Accordingly, we divide the test statistic range into five regions, -1 – -0.2, -0.2 – 0.0, 0.0 – 0.2, 0.2 – 0.4 and 0.4 – 1.0. Then, we calculate the tBDT difference for all particles in each region and fit it to a Gaussian function as before. The results are shown in Figs. 9.5 to 9.9 and summarized in Fig. 9.10.

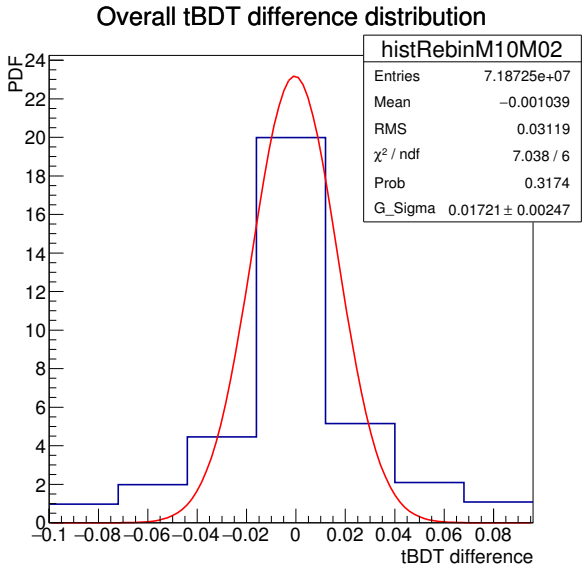


Fig. 9.5 BDT test statistic difference distribution histogram for the signal and the background (all particles) in blue, and a fit to a Gaussian function in red, calculated at tBDT range [-1.0,-0.2].

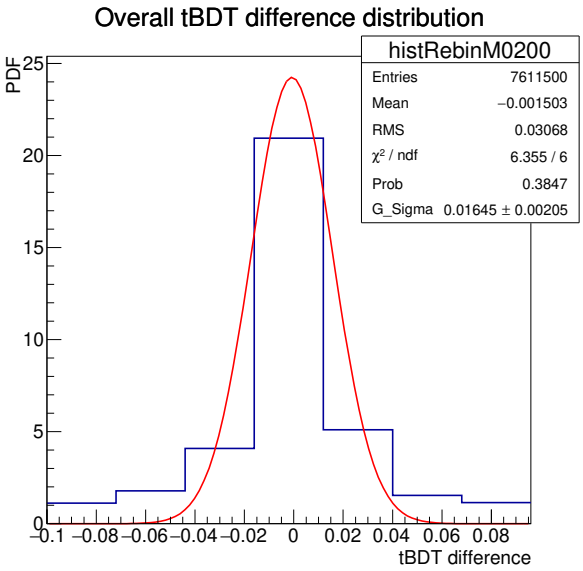


Fig. 9.6 BDT test statistic difference distribution histogram for the signal and the background (all particles) in blue, and a fit to a Gaussian function in red, calculated at tBDT range]-0.2,0.0].

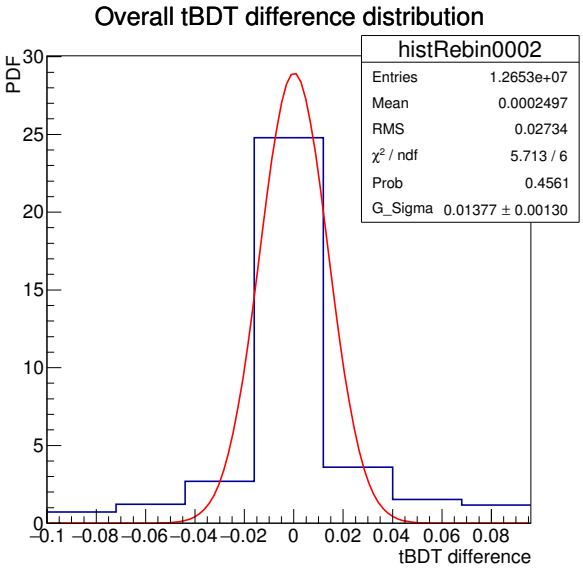


Fig. 9.7 BDT test statistic difference distribution histogram for the signal and the background (all particles) in blue, and a fit to a Gaussian function in red, calculated at tBDT range]0.0,0.2].

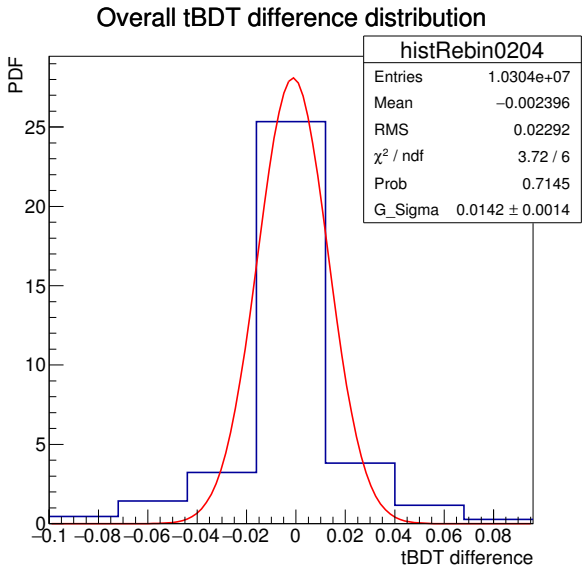


Fig. 9.8 BDT test statistic difference distribution histogram for the signal and the background (all particles) in blue, and a fit to a Gaussian function in red, calculated at tBDT range]0.2,0.4].

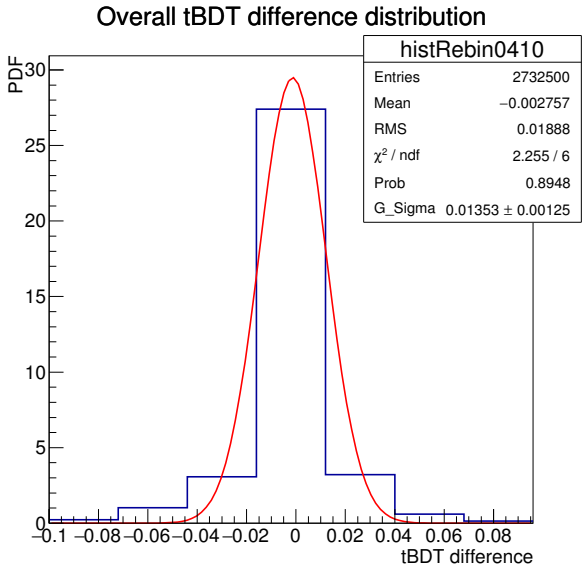


Fig. 9.9 BDT test statistic difference distribution histogram for the signal and the background (all particles) in blue, and a fit to a Gaussian function in red, calculated at tBDT range]0.4,1.0].

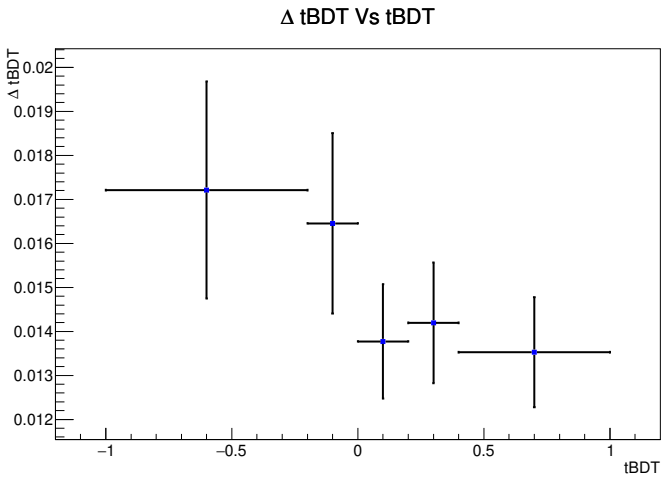


Fig. 9.10 BDT test statistic (tBDT) uncertainty as a function of the calculated tBDT.

The tBDT systematic at the region of interest (0.2–0.4) is 0.014 ± 0.001.

When running the analysis with the tBDT systematic enabled, the tBDT will be assigned an uncertainty, according to its value, as depicted in Fig. 9.10. In that case, all the corresponding BDT input systematics must be disabled (TPC, FGD and ECAL PIDS, and ECAL energy-related systematics) to avoid double counting systematics. However, the momentum-related systematics must still be enabled as it is used by other cuts, for instance to find the second most energetic track in a TPC veto or the gamma invariant mass calculation.

9.2 Uncertainty on the BDT Performance

The systematics on the purity (Δp) and efficiency ($\Delta\varepsilon$) are calculated following a similar recipe to that of Section 9.1.2 but with only 100 toy experiments and on a smaller MC sample due to the substantial computational power required. Here, we calculate the purity (p) and efficiency (ε) once for each toy experiment. We then build the difference distribution as before. Again, the mean of the difference distribution represents the bias estimate while the width represents the systematic uncertainty. The BDT test statistic and performance system-

atics evaluation is visualized using Fig. 9.11, where N is the number of toy experiments, m is the total number of events in one experiment, $tBDT_{ij}$ is the test statistics calculated by the BDT for event i in toy experiment number j ($j = 0$, for the nominal experiment), $\Delta tBDT_{ij} = tBDT_{ij} - tBDT_{i0}$, p and ε are the purity and efficiency for one experiment as before, $\Delta FoM = FoM_{toy} - FoM_{Nominal}$, with FoM =$p$ or ε.

	Toy experiments →		
	Nominal	**Toy 1**	...	**Toy N**
Events ↓	$tBDT_{10}$	$tBDT_{11}$	...	$tBDT_{1N}$
		$\Delta tBDT_{11}$	...	$\Delta tBDT_{1N}$
	$tBDT_{20}$	$tBDT_{21}$	...	$tBDT_{2N}$
		$\Delta tBDT_{21}$	...	$\Delta tBDT_{2N}$
	$tBDT_{m0}$	$tBDT_{m1}$	...	$tBDT_{mN}$
		$\Delta tBDT_{m1}$	...	$\Delta tBDT_{mN}$
	p_0, ε_0	Δp_1, $\Delta\varepsilon_1$	...	Δp_N, $\Delta\varepsilon_N$

Fig. 9.11 Schematic representation for the BDT test statistics and performance systematics evaluation. Here, N is the number of toy experiments, m is the total number of events in one experiment, $tBDT_{ij}$ is the test statistics for event i in toy experiment number j ($j = 0$, for the nominal experiment), $\Delta tBDT_{ij} = tBDT_{ij} - tBDT_{i0}$, and $\Delta FoM = FoM_{toy} - FoM_{Nominal}$, with FoM = purity (p) or efficiency (ε).

The Δp and $\Delta\varepsilon$ distributions are shown in Figs. 9.12 and 9.13 respectively.

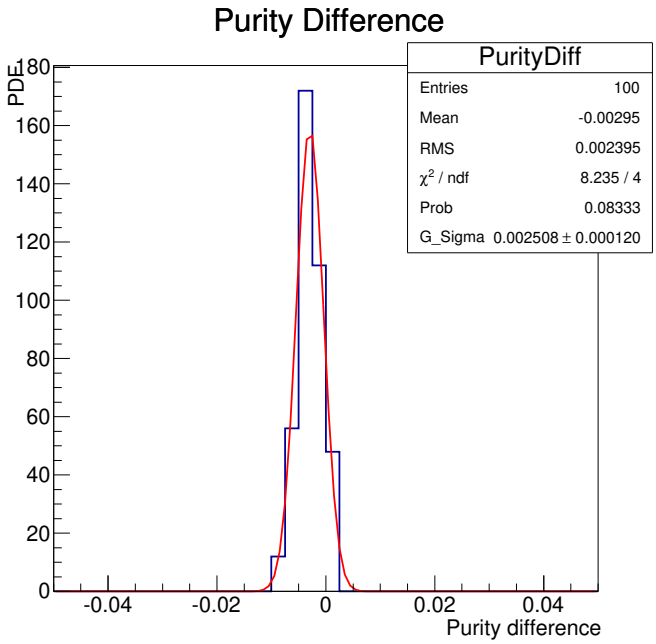


Fig. 9.12 The purity difference distribution histogram in blue, and a fit to a Gaussian function in red.

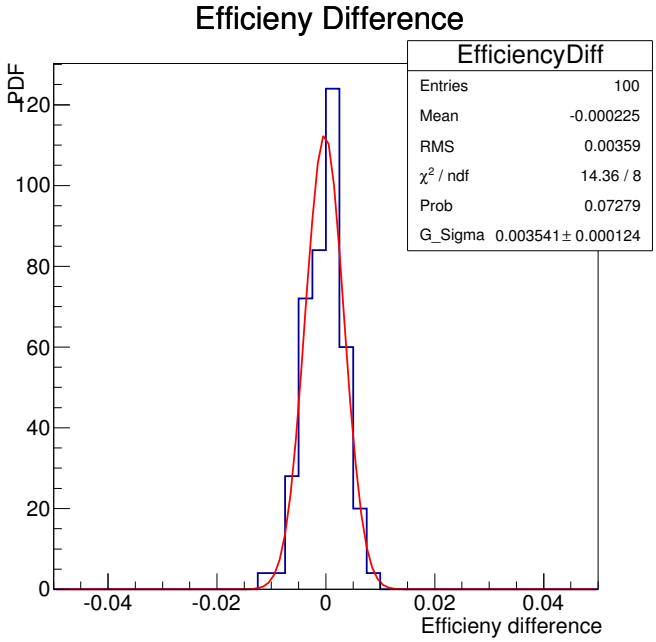


Fig. 9.13 The efficiency difference distribution histogram in blue, and a fit to a Gaussian function in red.

The mean of the FoM distributions is close to zero. Consequently, the BDT cut is not biased. Furthermore, the widths are small, indicating a stable BDT performance.

Part III

Unfolding

Chapter 10

Unfolding

10.1 The Unfolding Problem

An ideal detector will measure a kinematic variable x with complete accuracy. However, a typical detector suffers from limited acceptance, non linearity and finite resolution. These imperfections smear the measured quantity or miss it completely. In addition, any signal measurement is contaminated with background. Consequently, the measured quantity distribution $g(y)$ is related to the true quantity distribution $f(x)$ by

$$g(y) = \int A(y,x) f(x) dx + b(y), \tag{10.1}$$

where $A(y,x)$ is the detector response and $b(y)$ represents the background, cf. [82, Section 1].

There are two main approaches to tackle this problem. The first one is by parametric inference, in which we assume a mathematical function (model) for the signal and background distributions with free parameters, and we use the measurements to fit for the model parameters. An example of such an approache is the maximum likelihood or the least square techniques cf. [83, chapters 6 and 7]. However, we sometimes do not have a good representa-

tion of the functional form of the model, or we want a model-independent solution. In such cases, an alternate approach, unfolding, is often used.

The unfolding process is used to estimate the true variable distribution $f(x)$ from the measured one $g(y)$, having a model for the detector response $A(y,x)$ as well as the background contamination $b(y)$. This is an inverse problem, also known in mathematics by the name of Fredholm integral equation of the first kind, and in signal processing as deconvolution.

In practice, and in order to use numerical techniques, we discretize the measurement, binning to histograms, and transforming the unfolding problem into the linear equation

$$y = Ax + b. \tag{10.2}$$

Here the symbols have the same meaning as above, and the matrix element of A, A_{ij}, represents the probability of measuring the kinematic variable in bin i of the $(y-b)$ vector, i.e. after removing the background, if its true value was in bin j of the x vector.

Usually, determining the direct inverse

$$x = A^{-1}(y-b) \tag{10.3}$$

is not possible as the detector response function is either not invertible or ill-posed. Normally all the measured quantities have statistical and systematic uncertainties, and for an ill-posed inverse problem a small variation in the measured (smeared) space will result in a big variation in the unfolded (truth) space, making the whole inverse process unstable (see Fig. 10.1). Hence, the introduction of a regularization technique is mandatory.

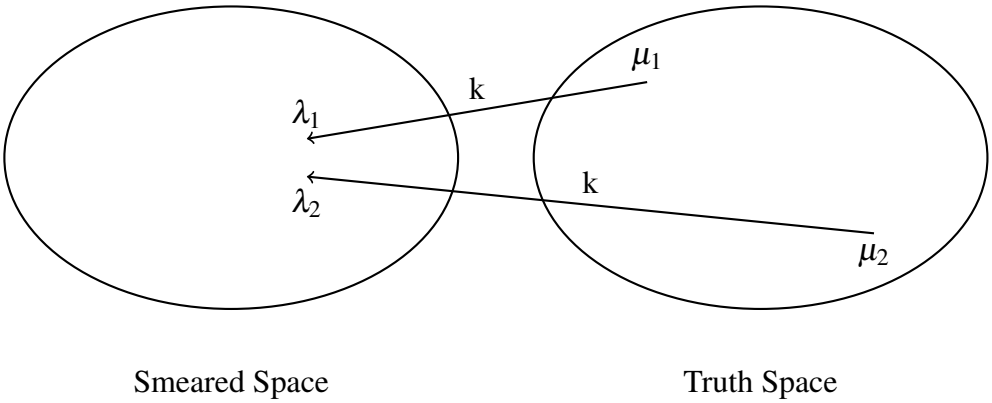


Fig. 10.1 Ill-posed pseudo-inverse, for the $\lambda = k\mu$ or $\mu = k^{-1}\lambda$ problem. Here, λ is the observed quantity vector, k is the detector response (smearing matrix) and μ is the true quantity vector. A small variation in observed quantity (λ) in the smeared space leads to a big change in the unfolded (μ) in the truth space.

The ill-posed pseudo-inverse problem drove the development of different unfolding algorithms (cf.Section 10.2) or to completely abandon the unfolding process by using forward folding. In forward folding, the theoretical model being tested is fed to the detector response matrix. The output is then compared to the measured quantities. Alternatively, one can find the best fit of the model parameters using a likelihood or a chi-squared minimization method [84].

Furthermore, the complexity of the unfolding process increases with limited statistics of the data as an effect of the binning choice and the regularization of a small number of events. Moreover, the statistical uncertainty of the detector response (smearing matrix) obtained from the MC simulation, is often neglected by some unfolding algorithms. Nonetheless, unfolding is essential to compare results from different experiments having different detector responses.

10.2 Unfolding Algorithms

The main software package for cross-section extraction (unfolding) in T2K is called the xsTool [85]. xsTool integrates different unfolding algorithms from RooUnfold [86] and is responsible for the calculation of the error propagation.

Generally, the unfolding algorithms can be categorized into three main groups:

- bin-by-bin,
- matrix inversion, and
- Bayesian methods.

The RooUnfold package supports bin-by-bin, unregularized matrix inversion, Singular Value Decomposition (SVD), TUnfold, and iterative Bayes (based on D'Agostini 1995 paper [87]). This section presents a brief overview of the above-mentioned algorithms along with their limitations. A complete discussion with test results is found in [86] and [88]. First, in Sections 10.2 and 10.2.6, we will investigate the case of no background for simplicity, then, the background treatment will be discussed in Section 10.3.

10.2.1 Bin-by-bin

Bin-by-bin is the simplest unfolding algorithm, where the unfolded value (x) is just a scaled version of the measured value (y). The scale factor is estimated from MC simulations to be the ratio between the generated number of events in bin i ($N_i^{\text{MC gen}}$) to the reconstructed number of events at the same bin ($N_i^{\text{MC reco}}$), as in

$$x_i = \frac{N_i^{\text{MC gen}}}{N_i^{\text{MC reco}}} y_i. \tag{10.4}$$

The results will be significantly biased toward the underlying MC distribution (the assumed model).

10.2.2 Unregularized Matrix Inversion

A direct matrix inversion, as described by Eq. (10.3), is possible by applying techniques like Singular Value Decomposition (SVD)[1], cf. [90, section 4] to obtain a pseudo-inverse[2] for the detector response.

This method is ill-posed, as described in Fig. 10.1, and is therefore not recommended, cf. [86, Section 3.4]. Furthermore, the physical interpretation of the measured spectrum as a vector in a certain vector-space is not logically motivated, especially when considering that some elements of this vector may be due to background that has not been taken into consideration. In this case, the inverse matrix may result in a vector of non-physical negative numbers.

10.2.3 Regularized Singular Value Decomposition

The singular value decomposition (SVD) technique provides a pseudo-inverse for the detector response matrix ($A_{m\times n}$) by factoring

$$A = USV^T, \tag{10.5}$$

[1] A real valued matrix $M_{m\times n}$ has a singular value σ, if and only if, there exist unit-length vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, such that

$$Mv = \sigma u, \text{ and}$$
$$M^T u = \sigma v.$$

Where, T denotes the matrix transpose, and the vectors u and v are the left-singular and right-singular vectors for σ, respectively [89].

[2] A pseudo-inverse matrix A^+ fulfills Moore and Penrose, cf.[91, section 2] conditions:

- $AA^+A = A$
- $A^+AA^+ = A^+$
- $(AA^+)^* = AA^+$
- $(A^+A)^* = A^+A$

i.e. AA^+ and A^+A map the columns of A and A^+ to themselves, besides, AA^+ and A^+A are Hermitian.

where, U and V are orthogonal $m \times m$ and $n \times n$ matrices, respectively, $U^T = U^{-1}$ and $V^T = V^{-1}$, S is an $m \times n$ diagonal matrix of the singular values (non negative), and the columns of U and V represent the left- and right-singular vectors, respectively.

This process could be visualized as a rotation by an orthogonal matrix, then stretching by a diagonal matrix and finally another rotation to bring the truth vector as close as possible to the measured vector.

The problem is formulated as

$$(\widetilde{A}w - y)^T Y^{-1} (\widetilde{A}w - y) + \tau (Cw)^T (Cw) = min, \tag{10.6}$$

where $\widetilde{A}_{ij}$ is the reconstructed number of events in bin i due to the simulated true number of events in bin j, w_i is the weight of true bin i, $w_i = \frac{x_i}{x_i^{\text{MC gen}}}$, with $x_i^{\text{MC gen}}$ the MC simulated number of events in bin i, y the measured vector having a covariance matrix Y. We solve the equation for w_i, from which we can infer the true value of x_i. The regularization term, $\tau (Cw)^T (Cw)$, is added to stabilize the results. Here, τ is the regularization strength, and C defines a prior condition for the solution. If we want to reduce the deviation from the MC expectation, we minimize the length of w by choosing C to be the identity matrix. Alternatively, we can set C to be the second derivative matrix to minimize the curvature of w.

In the original SVD paper [90], the regularization strength is set to the square of the singular value s_k, where k is the index of the element of the vector d after which $d_{j>k}$ is not statistically significant. The vector d is obtained after reformulating the problem as follows.

First, we diagonalize the covariance matrix, $Y = QRQ^T$, with $R_{ii} = r_i^2$. Substituting this in Eq. (10.6), and defining $\hat{A}_{ij} = \frac{1}{r_i}(Q^T \widetilde{A})_{ij}$, and $\hat{y}_i = \frac{1}{r_i}(Q^T y)_i$, we get

$$(\hat{A}w - \hat{y})^T (\hat{A}w - \hat{y}) + \tau (Cw)^T (Cw) = min, \tag{10.7}$$

or, equivalently solving

$$\begin{pmatrix} \hat{A}C^{-1} \\ \sqrt{\tau} \end{pmatrix} Cw = \begin{pmatrix} \hat{y} \\ 0 \end{pmatrix}, \tag{10.8}$$

and decomposing $\hat{A}C^{-1} = U^{'}S^{'}V^{'T}$, where $S^{'}$ has the singular values in a non-increasing order. The d vector will be equal to $U^{'T}\hat{y}$. Finally, the significance of d_i can be estimated by plotting $log|d_i|$ vs i, cf. [90, Section 7].

In practice, the regularization tuning depends on the distribution, binning, as well as the data and MC statistics. Aside from this, the SVD method ignores the Poisson nature of the measurement. Furthermore, the choice of the C matrix may result in a bias toward the underlying model. Moreover, the second derivative matrix is generally degenerate and requires the addition of an artificial small term ζ to its diagonal, e.g. $\zeta = 10^{-4}$, to make it invertible.

10.2.4 TUnfold

TUnfold [92] is an unfolding algorithm based on least square fitting. TUnfold uses Tikhonov regularization [92] for which the strength is normally determined using the L-curve method. The algorithm minimizes the following Lagrangian $\mathscr{L}$ to find the unfolded histogram x,

$$\mathscr{L} = (y-Ax)^T Y^{-1}(y-Ax) + \tau^2(x - f_b x^{\text{MC reco}})^T C^T C(x - f_b x^{\text{MC reco}}) + \lambda(y^{tot} - e^T x). \tag{10.9}$$

Here the first term is the least square fit with A the detector response matrix, x the unfolded histogram, and y the measured histogram having a covariance matrix Y. The second term serves as a regularization to suppress the statistical fluctuations. Here, the regularization strength τ is normally obtained by the L-curve scan method, in which the $\log \frac{\mathscr{L}_2}{\tau^2}$ is plotted versus $\log \mathscr{L}_1$ for different values of τ and the point of maximum curvature (the kink of the curve) is chosen to determine τ. While a small value of τ ($\tau \to 0$) is similar to the

no-regularization case, a larger value of τ will result in a bias toward the model. The kink point represents a balanced choice for τ. f_b represents a normalization factor, normally set to 1, $x^{\text{MC reco}}$ is the expected MC truth value, and C could be either the identity matrix to minimize the norm of the difference vector or a second derivative matrix to minimize the curvature.

Finally, the last term is optional. Here e is the efficiency vector, $e_i = \sum_i A_{ij}$ and $y^{tot} = \sum_i y_i$ is the total number of observed events. The $\mathscr{L}_3$ forces the sum of the regularized unfolded results x to be close to the total number of observed events y^{tot}. The strength of this constraint is set to λ, which is determined by minimizing Eq. (10.9).

This method is similar to the regularized SVD, but with a different way of setting the regularization parameter τ as well as an extra normalization constraint $\mathscr{L}_3$. Hence, it shares the same drawbacks as the regularized SVD.

10.2.5 Iterative Bayes (D'Agostini 1995)

The last algorithm supported by RooUnfold, is the iterative Bayes aproach. It applies Bayes rule[3] to unfold the number of truth events (causes) from the number of reconstructed events (effects). Here, the smearing matrix, $P(E|C)$, is a cause-effect probabilistic relation. The probability of a cause, given an effect is then

$$P(C_i|E_j) = \frac{P(E_j|C_i)P(C_i)}{\Sigma_k P(E_j|C_k)P(C_k)}, \tag{10.10}$$

[3]Bayes rule connects the conditional probability of an event with the likelihood of the condition and the prior knowledge,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}.$$

Where, $P(A_i|B)$ is called the posterior probability and is related to $P(B|A_i)$, the likelihood, and the prior probability of the event $P(A_i)$. The denominator, $\sum_j P(B|A_j)P(A_j)$, is just a normalization condition which is equal to P(B).

where the prior $P(C_i)$ is unknown. The prior, $P(C_i)$ can be initialized using the MC model prediction, or to any other reasonable assumption. The choice of a uniform $P(C_i)$, means that all causes are equally probable. This is obviously wrong and can be corrected using an iterative technique. The iterative Bayes was first introduced in an optics application [93], then in astronomy [94], and adapted for HEP [87]. In this technique, the unfolded histogram is input as an enhanced prior to the next iteration.

In the D'Agostini algorithm, the inefficiency is modelled by an extra effect bin called the trash bin (T) that holds the probability for events that have not been reconstructed for a certain truth bin (see Fig. 10.2).

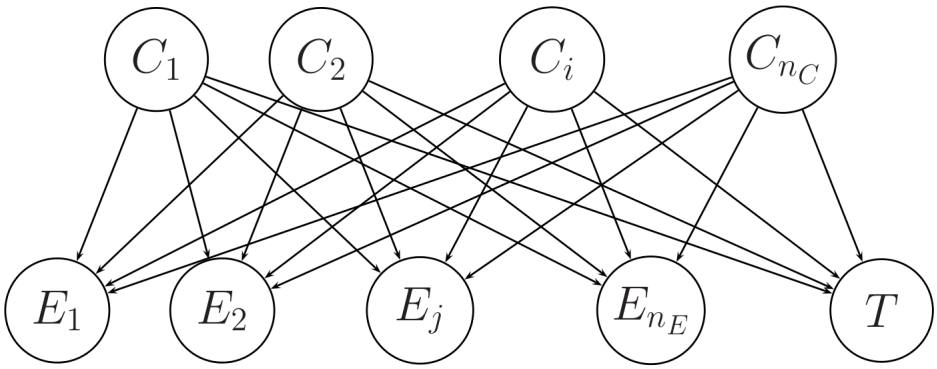


Fig. 10.2 The detector response (smearing) matrix presented as probabilistic links between causes (C_i) and effects (E_i). The inefficiency is modelled by an extra effect bin called the trash bin (T) that holds the probability of all the events that have not been reconstructed for a certain truth bin. Figure courtesy of [1].

The Algorithm is depicted in Fig. 10.3 and explained here. First, we initialize the prior $P(C_i)$ using an educated guess. In addition, the efficiency of the detector for each truth bin ε_i is calculated as

$$\varepsilon_i = \sum_j P(E_j|C_i). \tag{10.11}$$

Where $P(E_j|C_i)$ is the probability of having an effect in the reconstructed bin j due to a cause in bin i.

Subsequently, we use Bayes rule to calculate the probability of a cause, given an effect $P(C_i|E_j)$, from the smearing matrix $P(E_j|C_i)$ and the prior knowledge of the truth histogram $P(C_i)$ as in Eq. (10.10).

Then, we obtain the unfolded histogram, by updating the cause count in bin i, $N(C_i)$, based on the observed number of events in bin j, $N^{obs}(E_j)$, and corrected by the detector efficiency for this bin ε_i, as

$$N(C_i) = (\Sigma_j P(C_i|E_j) \times N^{obs}(E_j)) \times \frac{1}{\varepsilon_i}. \tag{10.12}$$

Afterwards, we update the prior information for bin i to be the ratio of the unfolded number of events in bin i to the total number of unfolded events,

$$P(C_i) = \frac{N(C_i)}{\Sigma_k N(C_k)}. \tag{10.13}$$

Finally, we feed this new prior to the next iteration loop, until the number of iterations is reached.

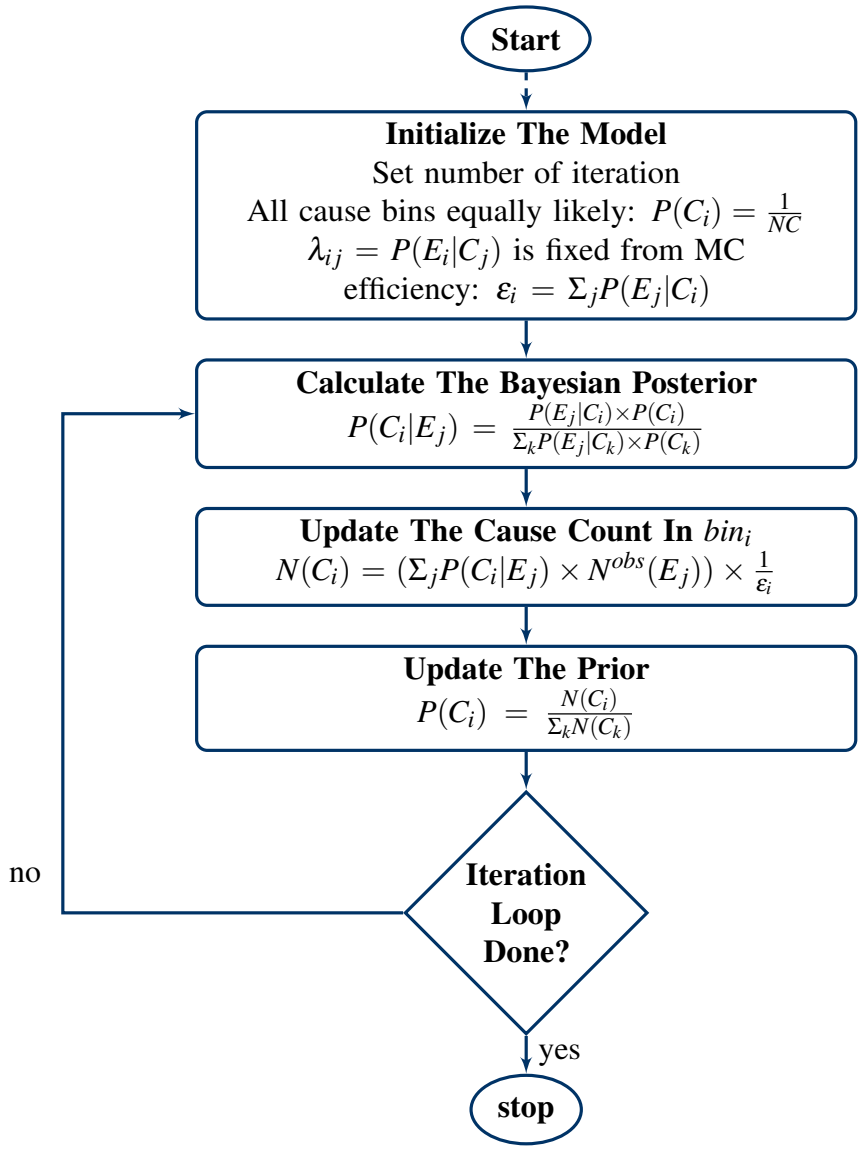


Fig. 10.3 Flow chart representation of the D'Agostini iterative Bayes algorithm, based on [87].

The main criticisms for this algorithm are:

- it results in fluctuations when increasing the number of iterations;
- it suffers from under-coverage [95, Section 5.2.2], as will be discussed in Chapter 11;
- its regularization is achieved by stopping early, with no quantitative measure to choose the number of iterations; and

- it creates a bias toward the underlying model if initiated with the MC as a prior and a single iteration is applied to overcome the previously mentioned problems.

10.2.6 D'Agostini 2010

Despite the criticisms of the iterative Bayes algorithm, it is widely used in T2K. D'Agostini, realizing the inaccuracy of this algorithm, provided an improved version [1], in which he addressed some of the challenges described above.

In this thesis, we have implemented the new D'Agostini algorithm, integrated it with the main T2K cross-section extraction tool (xsTool), supplied it with many regularization techniques, and tested its performance, as will be described in the next sections.

One of the main enhancements in the D'Agostini 2010 algorithm is the natural way of implementing the statistical error propagation using Monte-Carlo integration. In fact, all the input quantities, the observed histogram and smearing matrix, have statistical fluctuations. Accordingly, the algorithm samples the inputs from distributions built using the application of Bayes rule and conjugate priors, where we know the functional form of the distribution, $P(data|\theta)$, but we do not have the exact parameters values θ. Assigning the conjugate prior of the likelihood function, $P(data|\theta)$, to the distribution of the unknown parameter, $P(\theta)$, the probability of the parameter given the data, $P(\theta|data)$, will have the same form as the prior but with parameters constrained from the data, as represented by the following Bayes rule

$$P(\theta|data) \propto P(data|\theta) \times P(\theta). \tag{10.14}$$

For example, we know that the observed number of events follows a Poisson distribution[4] , but we do not know its mean. We can build a distribution of this mean by assigning a flat

[4] A random variable X that is Poisson-distributed with a rate μ, have a probability density function given by [1]

$$P(x;\mu) = \frac{e^{-\mu}\mu^{x}}{x!}.$$

gamma distribution[5] to be the prior probability of the mean and a Poisson distribution for the likelihood of the data. Then, applying Eq. (10.14), we obtain a gamma distribution whose parameters are fixed from the data, as

$$\text{gamma}(\mu;\alpha^{'},\beta^{'}) \propto \text{Poisson}(\text{observed data}|\mu) \times \text{gamma}(\mu;\alpha,\beta), \tag{10.15}$$

with

$$\alpha^{'} = \alpha + N^{obs}(E_j), \text{ and} \tag{10.16}$$

$$\beta^{'} = \beta + 1, \tag{10.17}$$

where $N^{obs}(E_j)$ is the number of observed events (effects) in bin j.

Similarly, for the smearing matrix, with $\lambda_{ij} = P(E_i|C_j)$, where each column, λ_j, is a multinomial distribution[6] with unknown parameters (p_1, p_2,..,p_k), whose conjugate prior is a Dirichlet distribution[7].

$$Dir(\alpha^{'}) \propto Mult(\lambda_j|p) \times Dir(\alpha), \tag{10.18}$$

[5] A random variable X that is gamma-distributed with a shape α and rate β parameters, have a probability density function given by [1]

$$f(x;\alpha,\beta) = \frac{\beta^{\alpha}x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)},$$

where, $\Gamma(\alpha)$ is the gamma function, the generalization of the factorial, and is given by

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}dx.$$

For a flat gamma distribution, $\alpha = 1$ and $\beta \to 0$.

[6] A multivariate random variable X, with $x = \{x_1, x_2, \ldots, x_k\}$, following a multinomial distribution with parameters $p = \{p_1, p_2, \ldots, p_k\}$, is given by [1]

$$Mult(x;p) = \frac{n!}{\prod_{i=1}^{k} x_i!}\prod_{i=1}^{k} p^k,$$

where, $n = \sum_i x_i$ and $\sum_i p_i = 1$.

[7] A multivariate random variable X, with $x = \{x_1, x_2, \ldots, x_k\}$, following a Dirichlet distribution with parameters $\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$, is given by [1]

$$Dir(x;\alpha) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i^k \Gamma(\alpha_i)}\prod_{i=1}^{k} x_i^{\alpha_i-1},$$

and we can sample λ_j from the new Dirichlet distribution (the posterior) as

$$\lambda_j \sim Dir(\alpha^{'}), \tag{10.19}$$

$$\sim Dir(\alpha + x^{MC}(E)|_{x^{MC}(C_j)}), \tag{10.20}$$

where, $x^{MC}(E)|_{x^{MC}(C_i)}$ is the vector of simulated number of effects, $x^{MC}(E)$, due to a certain number of true causes in bin j $x(C_j)$.

The Algorithm is depicted in Fig. 10.4 and explained here. First, we initialize the model the same way as in the older D'Agostini algorithm by either setting the prior based on MC or another educated guess. The code supports supplying any valid arbitrary function as long as it fulfills the requirement of probability theory. Basically, the prior histogram must consist of non-negative numbers whose sum equals unity. We set also the number of samples that will be used in the evaluation of statistical error propagation. A large number of samples is preferred for a better modelling of the underlying distribution. However, this results in a longer execution time. The default value of the number of samples is set to 1000, with the possibility of adjusting it.

We then sample one smearing matrix built from the Dirichlet distribution and the MC simulation as discussed above. Accordingly, the efficiency of each cause bin, ε_j, for this sample is computed as

$$\varepsilon_j = \Sigma_i \lambda_{ij}, \tag{10.21}$$

where λ_{ij} is the i^{th} element of the multivariate random number λ_j sampled from the Dirichlet distribution. We after calculate the posterior, with elements $\Theta_{ij} = P(C_i|E_j)$, from the smearing matrix sample and the prior using Bayes rule as before (see Eq. (10.10)).

where $\Gamma(.)$ is the gamma function, $\sum_i x_i = 1$, $x_i \geq 0$, and $\alpha_i > 0$.
An uninformative Dirichlet prior (uniform) has $\alpha_i = 1 \; \forall i \in [1,k]$.

Consequently, we sample the number of observed events, μ, from a gamma distribution. The number of events is potentially the effect of several different cause bins due to event migration (the detector response). The corresponding number of causes follows a multinomial distribution with parameters equal to Θ_{ij}.

Since the gamma function is continuous, μ could be any non-negative real number, and not necessarily an integer. However, the multinomial distribution is only defined for non-negative integer values. Consequently, we round μ to the nearest integer m, and calculate the scale factor s, as $s = \frac{\mu}{m}$.

We then sample a multinomial distribution of the number of observed effects in bin j, m_j, over the possible causes with probabilities Θ_{ij}. We correct the number of causes in each true bins by the inverse of the scale factor. We sum all the spectra generated by the multinomial throws, for all the effect bins, and finally we correct the number of true causes in each bin by the corresponding detector efficiency using

$$N(C_i) = \Big(\sum_{N_E} Mult(m_j, \Theta_{ij}) \times \frac{1}{s_j}\Big)_i \times \frac{1}{\varepsilon_i}. \tag{10.22}$$

Here $N(C_i)$ is the number of cause events in bin i, s_j is the scale factor of the sampled effect in bin j, and the sum is over all the effect bins.

The generation of possible truth spectra from the input samples continues until reaching the maximum number of samples set in the initialization. From the generated truth spectra, we compute the covariance matrix as well as the average spectrum as the result of this iteration.

If the maximum number of iterations has not been reached, the average spectrum is smoothed and fed as an enhanced prior to the next iteration. The algorithm introduces this regularization step (smoothing the prior), motivated by the fact that the physical distribution should not have an abrupt change from one bin to the other. This means that the unfolded

distribution has a shape constraint: it must be smooth. Different smoothing algorithms have been investigated and tested, as will be described in Section 10.2.6.

Furthermore, the new D'Agostini algorithm promises a better treatment of the small number of observed and simulated events, a common problem in an $\overline{\nu}_e$ measurement. This is achieved by adjusting the priors such that they do not treat all zeros equally (using non-flat priors). For example, a zero number of events can appear in either the measured number of events or the simulated smearing matrix. However, one may believe that an obtained zero may turn up into one or more events if we let the experiment run long enough. For other zeros we believe they should remain as they are. Nevertheless, this method can result in a biased solution when performing an MC study since we are forcing zeros to be some positive values based on underlying assumptions. A full discussion of this procedure is found in [1].

In this thesis, we rely on a more pragmatic solution, where we simply re-bin the observed histogram to avoid the small number or zero counts, and we build the smearing matrix from a MC simulation that is at least one order of magnitude larger than the data.

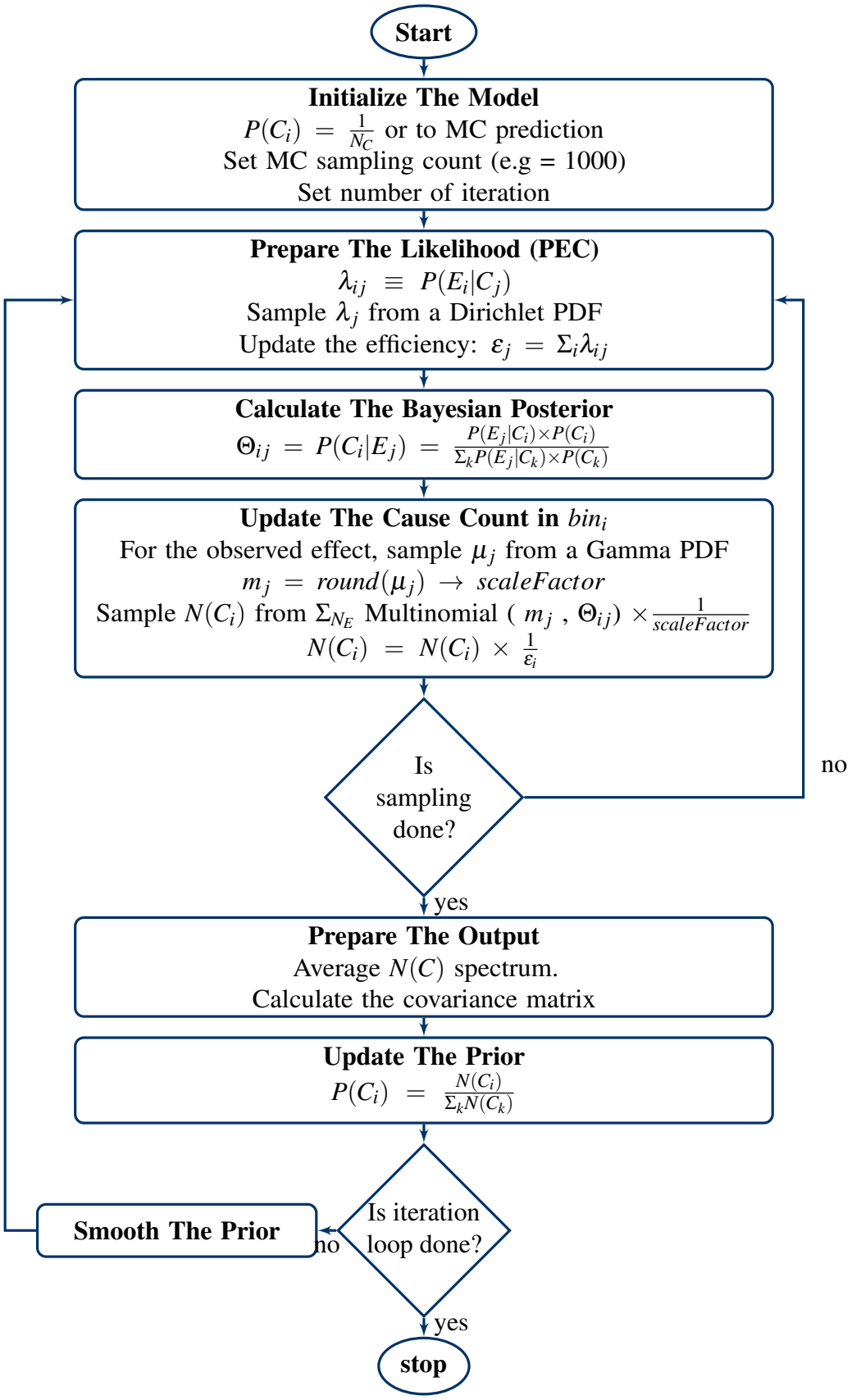


Fig. 10.4 Flow chart representation of the Improved D'Agostini algorithm, based on [1] .

In essence, the three major improvements results from the D'Agostini 2010 algorithm:

- accounting for the uncertainty of the observed effects and the smearing matrix (including efficiencies) early in the unfolding;
- adding a regularization step (smoothing the prior); and
- providing a better treatment for the limited statistics of the observed effects.

Smoothing Regularization

Smoothing is a major element of the new D'Agostini algorithm. A perfect smoother would:

- be MC shape independent, to avoid a bias toward certain model;
- be binning independent, as we expect to have fewer bins in an electron antineutrino analysis; and
- have a stable output.

At this step, the unfolded histogram probability $P(C_i)$ is smoothed before being input as a prior to the next iteration. It is worth mentioning that the unfolded result is never smoothed, only the prior that represents our best guess. Adding an extra requirement on the prior therefore does not contradict the Bayesian spirit of the algorithm. Different algorithms have been developed and tested as described in this section.

n^{th} Order Polynomial Fit The prior is fit to a polynomial with a predefined user order. The order of the polynomial should be carefully chosen to avoid overfitting, and must be less than the number of histogram bins minus one. A specific choice depends on the expected shape of the unfolded distribution as well as the available statistics. It is clear that this method is MC shape dependent.

Gaussian Convolution This smoothing method is widely used in image processing. The bin content is replaced by its weighted average along with that of its neighbours. For instance, in case of a one-dimensional three-bit mask, the smoothed value of bin i, b'_i, will be related to the neighbouring bins and mask values M_j as

$$b'_i = b_{i-1}M_{-1} + b_iM_0 + b_{i+1}M_1. \tag{10.23}$$

The mask size determines the number of neighbours to affect the smoothed value. On the other hand, the sigma parameter determines the weights of the neighbouring bin contributions; a smaller sigma results in a smaller contribution and vice versa. An example of such three-bit Gaussian masks for different values of sigma is shown in Table 10.1.

Sigma	Mask
1.00	(0.27901, 0.44198, 0.27901)
0.50	(0.15773, 0.68453, 0.15773)
0.25	(0.02275, 0.9545, 0.02275)

Table 10.1 Example of Gaussian three-bits mask with sigmas equal to 1, 0.5 and 0.25 respectively. The mask triplet represents (M_{-1}, M_0, M_{+1}) of Eq. (10.23).

This method is not binning independent and consequently a non-equal binning scheme would have the wrong smoothing.

Fast Fourier Transform (FFT) Here, we treat the $P(C_i)$ as a signal in the time domain, with an amplitude equal to the content of the bin happening at a time equal to the centre of the bin. The FFT of such a signal is then obtained and we can filter either the high-frequency components or the components with less power, where the filtering is done by setting the corresponding bins of the FFT signal to zero before applying an inverse FFT to return back to

the time domain with a smoothed signal. The code is implemented using the GNU Scientific Library (GSL)[96, Chapter 16].

Although this method is MC shape and binning independent, the limited statistics of the $\overline{\nu}_e$ forces the usage of a small number of bins, making the application of FFT impractical. For instance, removing one harmonic from a four-bin analysis will result in an unstable output due to the oscillatory nature of the FFT components.

353QH A wrapper for the 353QH-Twice algorithm presented by J.Friedman [97] was implemented. The algorithm is supported by ROOT[67] to smooth histograms and arrays.

The 353QH treats the data as a sum of smooth and rough information. The smooth portion corresponds to the signal and the rough portion to the noise to be eliminated. The algorithm performs three consecutive running medians[8] [98], smoothing the data with a window size equal to 3, then 5, and 3 again. It then applies a Quadratic interpolation (Q) for any set of three bins with identical contents (flat). The interpolation uses the two adjacent points to the flat area and its centre, to set the values of the other two flat points. Finally, it executes a Hanning (H), i.e. a convolution with a symmetric kernel, with weights (0.25, 0.50, 0.25), hence the name 353QH. The "twice" version of this algorithm applies the smoothing to the rough part of the data and adds it to the final smoothed histogram.

This method is also not binning independent, and thus not general enough for any binning scheme.

Spline We implemented a wrapper for the spline interpolator of ROOT[67]. The spline interpolator builds a cubic function at each data point,

$$s_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, \tag{10.24}$$

[8] A running median is proposed by Tukey, cf. [98], to overcome the vulnerability to outliers of the moving average. Here, each value of the histogram is replaced by the median of the neighbour bins covered by the window. Endpoints require a special treatment as described by Tukey's extrapolation method.

where a,b,c, and d are the coefficients of the cubic spline $s(x)$, and $x_i \leq x \leq x_{i+1}$, with i the histogram index. The algorithm ensures the continuity of the curve as well as the first and second derivatives of the spline,

$$s_{i-1}(x_i) = s_i(x_i),\ s'_{i-1}(x_i) = s'_i(x_i),\ \text{and}\ s''_{i-1}(x_i) = s''_i(x_i). \tag{10.25}$$

Two further conditions need to be applied in order to fix all the spline coefficients. A common choice is to set the second derivative of the first and last point to be zero (minimizing the curvature of the function) (normal cubic spline), or to apply a more sophisticated technique, e.g Akima, where another condition on each point slope ($s'_i = t_i$) is imposed:

$$t_i = \frac{|m_{i+1} - m_i|m_{i-1} + |m_{i-1} - m_{i-2}|m_i}{|m_{i+1} - m_i| + |m_{i-1} - m_{i-2}|}. \tag{10.26}$$

Here m is the slope of the straight line connecting the points (the secant), cf. [99].

Although the Akima is more stable than the cubic spline, it is still binning dependent and performs poorly in smoothing a noisy histogram.

Adaptive Kernel Density Estimator A Kernel Density Estimator (KDE) is a non-parametric (model independent) estimator of the probability density function from which we have obtained the measured data points. It works on unbinned data by superimposing a symmetric function (the kernel) on each data point, then summing them to build a probability density function.

The adaptive kernel automatically calculates the width of the kernel from the local density of events, ensuring a small width in the high statistics regions, to preserve the precise function details, and a large width in the low statistics regions to remove statistical fluctuations. Figure 10.5 illustrates the density estimation process for a toy example.

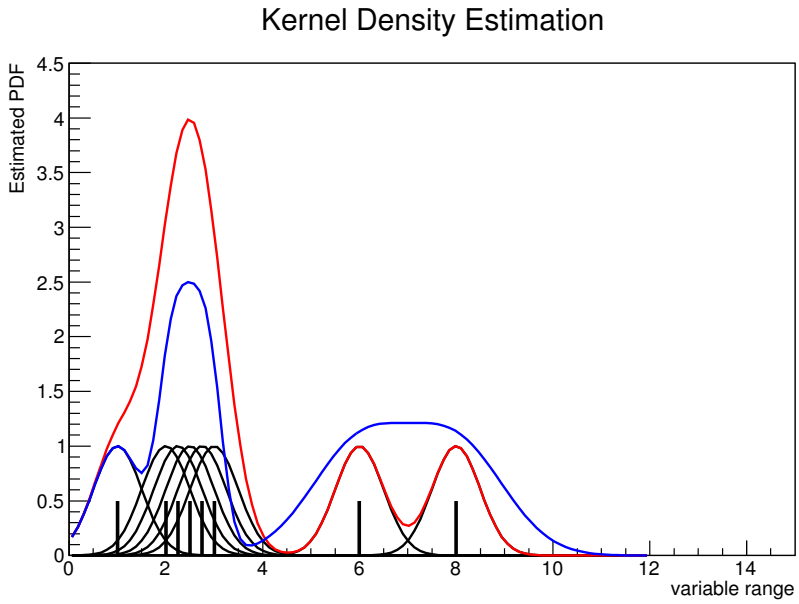


Fig. 10.5 KDE smoothing for a toy example consisting of eight data points, where data is represented by black vertical bars. In the fixed-width mode, a Gaussian function (in black curves), with a constant width, is superimposed on each data point. The sum of the fixed-width Gaussian functions is represented by the red curve. In the Adaptive mode, the kernel width varies, such that it is higher at low statistics and lower at high statistics. The sum of the Gaussian functions with adaptive widths is represented by the blue curve.

The KDE technique is summarized here. For more details please refer to [100, 101]. The KDE builds the density estimator for variable x, $\hat{p}(x)$, as

$$\hat{p}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_i}K\left(\frac{x-x_i}{w_i}\right), \tag{10.27}$$

where n is the total number of data points, K is the kernel function with width w, for a Gaussian kernel $K(x,w) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2w^2}}$. Here, w_i is chosen to minimize the Mean Integrated

Square Error (MISE) defined as

$$\begin{aligned} MISE &= E\Big[\int (\hat{p}(x) - p(x))^2 dx\Big] \\ &= \int E\big[(\hat{p}(x) - p(x))^2 dx\big] \\ &= \int MSE(\hat{p}(x)) dx. \end{aligned} \tag{10.28}$$

Here $E[.]$ is the expectation value, $p(x)$ is the true probability distribution and the Mean Square Error (MSE) of our estimator $\hat{p}(x)$ is related to its variance *Var* and bias, cf.[101, Section 2.1], by

$$\begin{aligned} MSE &= E\big[(\hat{p}(x) - p(x))^2\big] \\ &= Var(\hat{p}(x)) + bias^2(\hat{p}(x)). \end{aligned} \tag{10.29}$$

The true distribution $p(x)$ is unknown and hence the calculation of the bias and variance of the estimator is performed using re-sampling techniques such as cross-validation[9], Jackknife[10] and bootstrapping[11] [102]. It is worth noting that we can not minimize both the bias and the variance of the estimator at the same time. This is due to the bias–variance tradeoff that makes the estimator with a lower bias have a higher variance and vice versa, cf. [83, Sections

[9] In cross-validation re-sampling, we divide the dataset into k disjoint groups. Then, we perform K experiments, using $K-1$ groups to evaluate the density estimate and the remaining one for testing. Finally, the error (bias or variance) is estimated as the average error of the k experiments.

[10] In Jackknife re-sampling, we divide the dataset into k disjoint groups as in cross-validation. Then, we perform $k+1$ experiments, one on each K group and one on the original dataset, to evaluate our estimator. Finally, we compare the average of estimates on the k groups with estimator based on the full dataset.

[11] In bootstrapping, we re-sample uniformly n values from the data x_i ; $i = 1,...,n$ with replacement. We repeat the re-sampling process k times, generating k dataset similar to our original one. Then, we evaluate the estimator $\hat{p}^*(x)$ on each of these data sets. Finally, The distribution of the estimator $\hat{p}^*(x)$ about the original data estimator $\hat{p}(x)$ mimics the distribution of $\hat{p}(x)$ about the true distribution $p(x)$.

11.6 and 11.7]. For instance, the expectation value of the kernel estimator in Eq. (10.27) is

$$\begin{aligned} E\left[\hat{p}(x)\right] &= \frac{1}{n}\sum_{i=1}^{n}\int K(x-x_i;w_i)p(x_i)dx_i \\ &= \int K(x-y)p(y)dy. \end{aligned} \tag{10.30}$$

Unless $K(x-y)=\delta(x-y)$, $\hat{p}(x)$ will be biased for any $p(x)$, where $\delta(x-y)$ is the Dirac delta function, cf. [103, Chapter 5].

For Gaussian kernel and normally distributed data, the optimum kernel width (w_i^*) for data point x_i, cf.[103, Section 5.9], is given by

$$w_i^* = \left(\frac{4}{3}\right)^{\frac{1}{5}}\sqrt{\frac{\rho\sigma}{\hat{p_0}(x_i)}}n^{\frac{-1}{5}}, \tag{10.31}$$

where σ is the data standard deviation, n is the total number of data points, ρ is an extra smoothing factor, usually set to one, and $\hat{p_0}(x_i)$ is our initial estimate for the estimator, obtained from a fixed width kernel.

The code implements this smoothing functionality using the available libraries of RooFit[104]. Since we want to smooth a probability histogram, we have developed an unbinning strategy to create separate data events from the available histogram. The unbinning process provides our best estimate of the local data points destroyed by histogramming them. First, we build a probability histogram, by diving the number of events in each bin by the total number of events. Then, we build the PDF histogram by diving each of the probability histogram bins by its variable range, as

$$PDF_{bin_i} = \frac{n_i}{n_t \times r_i}. \tag{10.32}$$

Here n_i and n_t are the number of events in bin i and the total number of events respectively, and r_i is the variable range covering bin i, and is equal to the difference between the higher edge and the lower edge of the bin.

We rely on the fact that the underlying physics producing the histogram is continuous, making the probability density of events at the boundaries of adjacent bins continuous. The densities outside the binning range are chosen under the assumption that they will produce a correct normalization. For instance, the left density of the first bin and the right density of the last bin were assumed to be zero in this analysis, as the binning variable range covers almost all the available phase-space. We create a piecewise linear PDF from the PDF histogram by connecting the central bin values as well as the endpoints using linear curves[12]. The piecewise linear technique creates a PDF described by two linear curves $L_{i,1}$ and $L_{i,2}$ connected at the central bin value of bin i. Consequently, we assume that the bin contents are distributed inside each half-bin following the ratio of the area of the trapezoids created by the PDF's piecewise curves, with the histogram bin edges

$$\begin{aligned} n_{i,1} &= \frac{a_1}{a_1 + a_2}, \text{ and} \\ n_{i,2} &= n_i - n_{i,1}. \end{aligned} \tag{10.33}$$

Here $n_{i,1}$ and $n_{i,2}$ are the number of events in bin i having variable range from the low bin edge to the bin centre, and from the bin centre to the high bin edge respectively. Further, a_1 is the trapezoid area formed by the linear PDF curve $L_{i,1}$ from the top and the bin half range from the low bin edge to the centre, and a_2 is the trapezoid area formed by the linear PDF curve $L_{i,2}$ from the top and the bin half range from the centre to the high bin edge. After, we throw $n_{i,1}$ number of events following a PDF described by the linear curve $L_{i,1}$ in the first half-range of bin i, and similarly for $n_{i,2}$ in the second half-range. It is worth noting that the PDF normalization is automatically done by RooFit. Finally, we apply the Adaptive KDE on the thrown number of events to create a smooth PDF.

[12] If we want to fit a second-order continuous curve to pass through the central bin value of the current bin and the one before and after, a lot of subtleties will be faced including the discrete derivative of non-equal binning, discussed in Appendix C and the possibility that such curve has one or more zeros giving rise to negative probability density region. The piecewise linear is hence a simpler and better approximation.

We compute the prior probability of causes required by the D'Agostini algorithm as

$$P(C_i) = \int_{\text{low edge}_i}^{\text{high edge}_i} \hat{p}(x) \tag{10.34}$$

using numerical integration, where $P(C_i)$ is the probability of cause bin i, and $\hat{p}(x)$ is the smoothed PDF created by the KDE. The unbinning process is illustrated in Fig. 10.6.

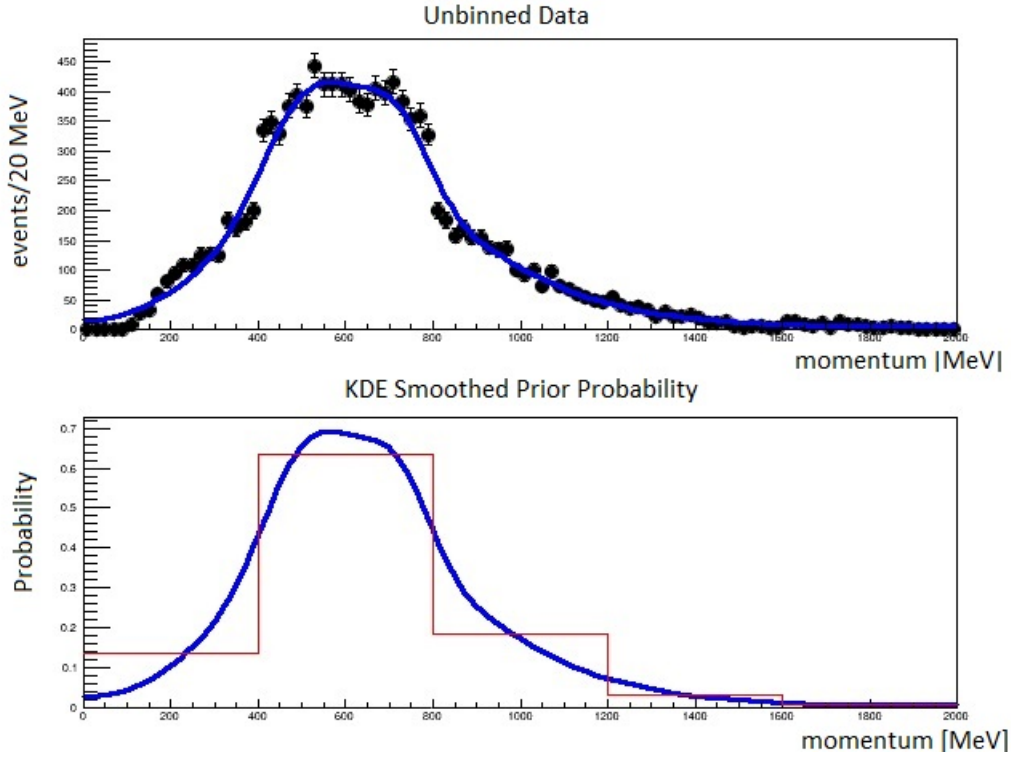


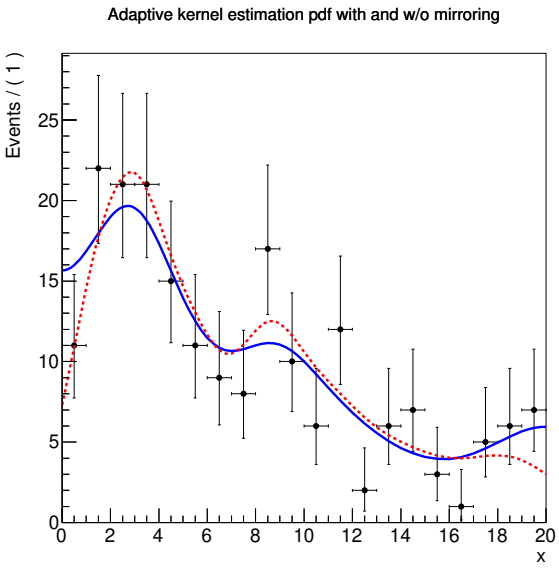
Fig. 10.6 Smoothing a probability histogram of the momentum distribution of a toy example, using the unbinning process and the adaptive kernel density estimation. The x-axis of the plot represents the momentum of the muon (in MeV) obtained from a ν_μ interaction sample, while the y-axis shows the PDF. The top figure shows the unbinned variable data points (in black) along with their statistical uncertainties overlaid with the estimated smooth probability density curve in blue. The bottom figure shows the probability histogram in red, overlaid with the smooth probability curve in blue.

Although numerical integration is computationally expensive, KDE smoothing fulfills the requirement of a perfect smoother mentioned in Section 10.2.6, and provided the best

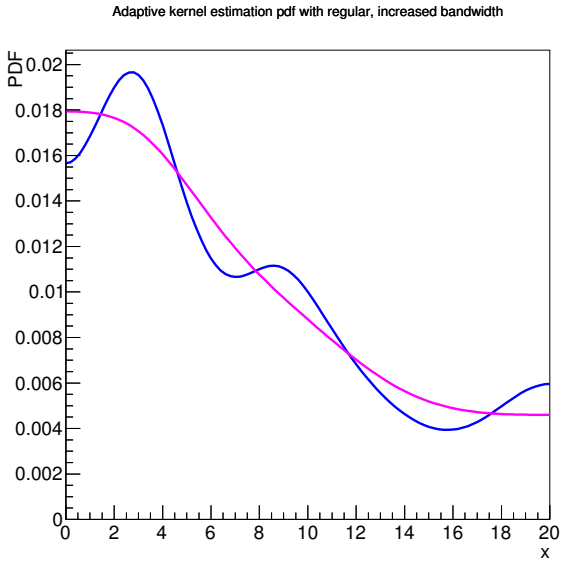
results when compared to other algorithms, hence, it is used as the default prior smoothing in this analysis.

Smoothing with Non-equal Bin Widths

RooFit[104] implements the one-dimensional adaptive KDE via a class called RooKeysPdf, with two main parameters, the mirror and the bandwidth scale factor. The mirror parameter models the possibility of non-negligible densities at the boundaries. The bandwidth scale factor, ρ, controls the strength of the smoothing, as in Eq. (10.31). Figure 10.7 shows the effect of these parameters on a toy example.



(a) Adaptive KDE Mirror effect.



(b) Adaptive KDE bandwidth scale factor effect.

Fig. 10.7 Adaptive-KDE applied to 200 points thrown from a toy example following a second-order polynomial $f(x) = 0.01 - 0.01x + 0.0004x^2$. Figure 10.7a shows the estimated PDF using adaptive KDE with no mirror (red dotted curve) and with mirror effect on both sides (blue solid curve). Figure 10.7b shows the Adaptive KDE with regular bandwidth scale parameter, set to one, (blue curve) and an increased bandwidth scale, set to two, (magenta curve).

Using the KDE to smooth a histogram with variable bin width and low statistics, that is not necessarily normally distributed, is challenging. The bandwidth scale parameter

will start to play an important role in the accuracy of the result, even when using adaptive techniques. For example, in our $\overline{\nu}_e$ analysis, we may have three or four momentum bins with the following bin edges {0, 0.6, 1.65, 30} GeV. The last bin is very large so as to include any $\overline{\nu}_e$ event with a mis-reconstructed momentum. Having low statistics, around 10-15 events/bin, prevents us from picking a better binning scheme.

We notice that the last bin is sparse and ~ 50 times wider than the first one. Keeping the KDE width scale equal to unity with the density in the last bin being small, will drive the algorithm to set the kernel width to be large, resulting in a biased result. On the other hand, setting the scale factor ρ to be the ratio of the smallest bin width to the biggest one, to be able to resolve the PDF structure at the lower bin width (in our example, set $\rho' = \rho \frac{0.6}{30} = 0.02$) will result in an extremely slow computation.

We either therefore have to neglect the last bin in our smoothing, or smooth a transformed variable, as explained below.

Smoothing a Transformed Variable

We developed a technique to smooth a transformed variable on a log instead of a linear scale. First, we build g(y) as the PDF of $y(x) = ln(x)$. Since the PDF $f(x)$ was inferred from the densities of the neighbour bins, and takes the form of two adjacent trapezoids for each bin, $f(x) = c_0 + c_1 x$, we have

$$\begin{aligned} y(x) &= ln(x), \\ x &= e^y, \end{aligned} \tag{10.35}$$

and

$$\frac{dx}{dy} = e^y. \tag{10.36}$$

The PDF of the transformed variable $g(y)$ is related to the original PDF $f(x)$ as

$$\begin{aligned} g(y) &= f(x(y))\frac{dx}{dy} \\ &= (c_0 + c_1 e^y)e^y \\ &= c_0 e^y + c_1 e^{2y}. \end{aligned} \tag{10.37}$$

The bin widths are comparable and the effect of the scale width is minimized while the computation speed is boosted. Finally, we compute the prior probability of causes required by the D'Agostini algorithm from the transformed variable y as

$$P(C_i) = \int_{\text{y low edge}_i}^{\text{y high edge}_i} \hat{p}_y(x), \tag{10.38}$$

where $\hat{p}_y(x)$ is the density estimate of the transformed variable, and $g(y)dy = f(x)dx$.

10.3 Background Treatment

The T2K software package for cross-section extraction (xsTool) has three options to treat the backgrounds: background subtraction, purity correction, and simultaneous unfolding. All these methods support the use of sidebands[13], and will be briefly described here.

10.3.1 Background Subtraction

In the background subtraction method, the background events expected from the MC simulation (b^{MC}) are subtracted from the measured data sample (y) before unfolding, as

$$d_i = y_i - b_i^{MC}. \tag{10.39}$$

[13] A sideband, sometimes called control sample, is a background selection, that means that it contains the events that represent one (or more) backgrounds. A sideband selection must be mutually exclusive with the signal selection. However, it can contain a small number of signal events.

Here d_i is the number of measured data events in bin i after subtracting the corresponding expected number of background events.

This method assumes that the MC model of the background, which is normally less well known, is correct. Furthermore, the systematic uncertainty in this method is larger than others because it sums from the right-hand side of the equation, e.g. the flux uncertainty.

In the case when a sideband (control sample) is used, the MC estimate for the background is scaled by a factor calculated from the number of data events that passed the sideband selection (SB^{data}) and the number of events expected from MC simulation of this sideband (SB^{MC}), giving

$$d_i = y_i - b_i^{MC} \times \frac{SB_i^{data}}{SB_i^{MC}}. \tag{10.40}$$

The usage of a sideband constrains the background MC prediction, yet background subtraction methods may result in a (non-physical) negative number of events, especially for a signal with low statistics.

10.3.2 Purity Correction

In the purity correction method we multiply the measured data in bin i (y_i) by an estimate of the signal purity (p) obtained from the MC simulation before unfolding. Consequently, the number of signal events in bin i (d_i) is

$$d_i = y_i \times p_i, \tag{10.41}$$

with

$$p_i = \frac{s_i^{MC}}{N_{\text{i tot}}^{MC}}. \tag{10.42}$$

Here s_i^{MC} and N_{tot}^{MC} are the estimated number of signal and the total number of events from the MC simulation in bin i, respectively.

When using a sideband, an updated purity value ($p^{'}$) will be calculated, where

$$p^{'} = \frac{s_i^{MC}}{\widetilde{N}_{\text{i tot}}^{MC}}, \tag{10.43}$$

with

$$\widetilde{N}_{\text{i tot}}^{MC} = s_i^{MC} + b_i^{MC} \times \frac{SB_i^{data}}{SB_i^{MC}}. \tag{10.44}$$

The purity value is bounded, $0 \leq p \leq 1$, and the corrected number of events is therefore never negative. Furthermore, the purity correction results in a smaller systematic uncertainty than background subtraction, due to the cancellation of the systematic uncertainties in the purity factor.

10.3.3 Simultaneous Unfolding

The D'Agostini unfolding algorithm provides a natural way of treating the background by including it in the smearing matrix as a possible cause of the observed effects. In fact, the smearing matrix is augmented by adding columns of different backgrounds or sidebands, as sketched in Fig. 10.8. The algorithm then simultaneously unfolds both the signal and background.

In case of sideband usage, the sideband selection must be mutually exclusive with that of the signal, otherwise, the background correction, as well as the detector efficiency calculation, will be wrong. Although the implemented algorithm supports having different binning for the sideband and the signal, the xsTool does not allow that. Another limitation in the current version of the xsTool is the requirement that a sideband describes a single type of background. In this thesis, we apply the simultaneous unfolding technique.

Trash ⟶ T		$1-\varepsilon_{C_{S_1}}$	…	$1-\varepsilon_{C_{S_m}}$	$1-\varepsilon_{C_{SB_1}}$	…	$1-\varepsilon_{C_{SB_k}}$	$1-\varepsilon_{C_{BKG}}$
Sideband Effects	E_{SB_l}	$P(E_{SB_l}\|C_{S_1})$	…	$P(E_{SB_l}\|C_{S_m})$	$P(E_{SB_l}\|C_{SB_1})$	…	$P(E_{SB_l}\|C_{SB_k})$	$P(E_{SB_l}\|C_{BKG})$
	…							
	E_{SB_1}	$P(E_{SB_1}\|C_{S_1})$	…	$P(E_{SB_1}\|C_{S_m})$	$P(E_{SB_1}\|C_{SB_1})$	…	$P(E_{SB_1}\|C_{SB_k})$	$P(E_{SB_1}\|C_{BKG})$
Signal Effects	E_{S_n}	$P(E_{S_n}\|C_{S_1})$	…	$P(E_{S_n}\|C_{S_m})$	$P(E_{S_n}\|C_{SB_1})$	…	$P(E_{S_n}\|C_{SB_k})$	$P(E_{S_n}\|C_{BKG})$
	…							
	E_{S_1}	$P(E_{S_1}\|C_{S_1})$	…	$P(E_{S_1}\|C_{S_m})$	$P(E_{S_1}\|C_{SB_1})$	…	$P(E_{S_1}\|C_{SB_k})$	$P(E_{S_1}\|C_{BKG})$
		C_{S_1}	…	C_{S_m}	C_{SB_1}	…	C_{SB_k}	C_{BKG}
		Signal Causes			Sideband Causes			Other BKG

Fig. 10.8 The augmented smearing matrix with one sideband. Sideband and signal selection must be mutually exclusive. Here, $P(E_i|C_j)$ is the probability of observing an effect in bin i due to a cause in bin j. S, SB, BKG and T denote the signal, sideband, other backgrounds and trash bins respectively. The trash bin holds all the events that have not been reconstructed, and is related to the detector efficiency ε, as $T = 1-\varepsilon$. m and n are the number of cause and effect bins for the signal, and k and l are the number of cause and effect bins for the sideband respectively.

Chapter 11

Unfolding Tests

In order to measure the performance of the D'Agostini 2010 unfolding algorithm, and to compare it with other algorithms, we have developed an unfolding test suite. The main inputs of these tests are the MC microtree file and the algorithm to be tested. The microtree is a ROOT [67] file generated by our main analysis software, HighLAND [70], containing all the necessary information about the events that entered the selection process.

Knowing the signal definition and the binning for the truth and reconstructed variable distributions, we process the microtree file to build the truth and reconstructed signal histograms, as well as the smearing matrix. We then input these histograms and the smearing matrix into the unfolding test engine. The test engine subsequently creates a distribution of all the possible reconstructed spectra. It treats each bin of the reference reconstructed histogram as the mean of a Poisson distribution and, sampling from this distribution, builds a fluctuated version of the reconstructed histogram. The engine passes these samples to the unfolding algorithm under evaluation to test the unfolded results. Finally, the engine computes the different figures of merit on the unfolded spectra, as will be discussed in the next sections.

All the test results shown in this chapter are for the unfolded momentum distribution of the $\overline{\nu}_e$ sample unless otherwise stated. The workflow of the performance testing of the

unfolding algorithm is depicted in Fig. 11.1, while the output of the unfolding process due to different throws is shown in Fig. 11.2.

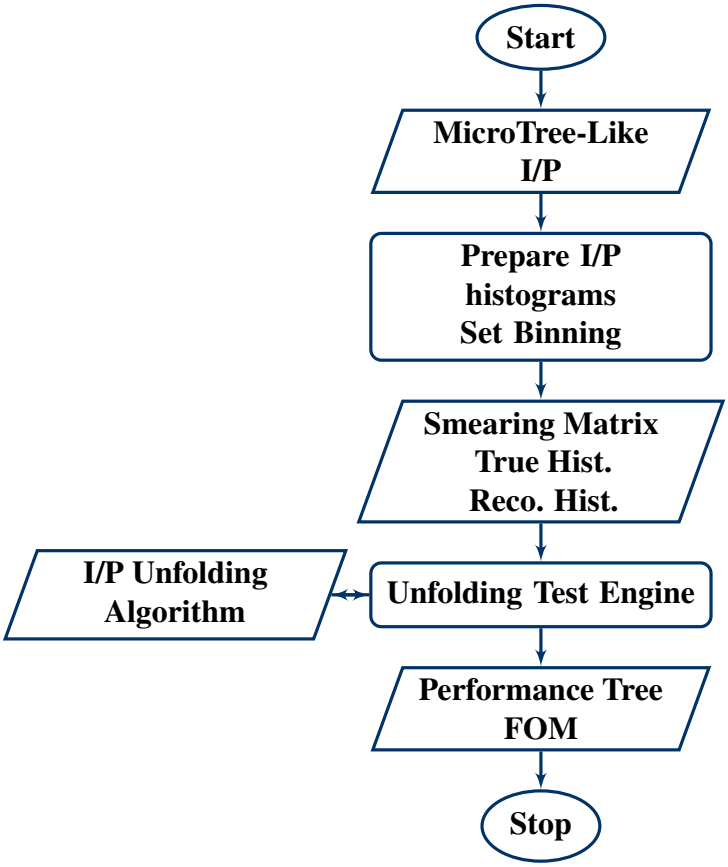


Fig. 11.1 Workflow of the unfolding performance testing.

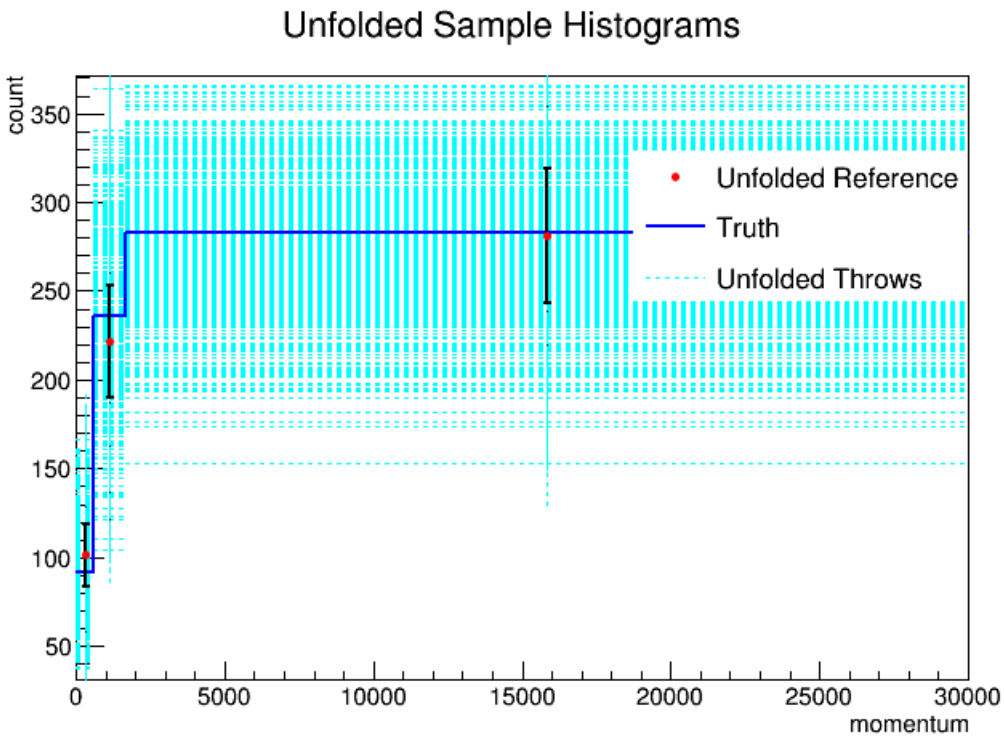


Fig. 11.2 Unfolding performance summary plot showing the truth distribution of the momentum (in MeV) of an electron antineutrino sample in dark blue, the unfolded results of 500 throws in cyan, and the unfolded results of the reference histogram in red superimposed with the statistical error in black bars.

The next sections show the results of various tests applied to the D'Agostini unfolding engine, configured with 10 iterations and a KDE smoothing for a transformed logarithmic variable as described in Section 10.2.6.

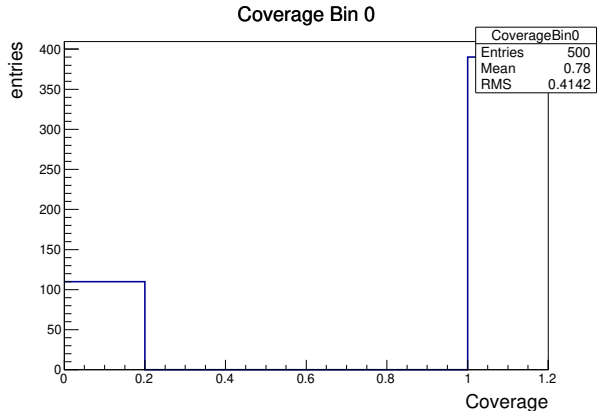
11.1 Coverage

11.1.1 Bin-by-bin Coverage

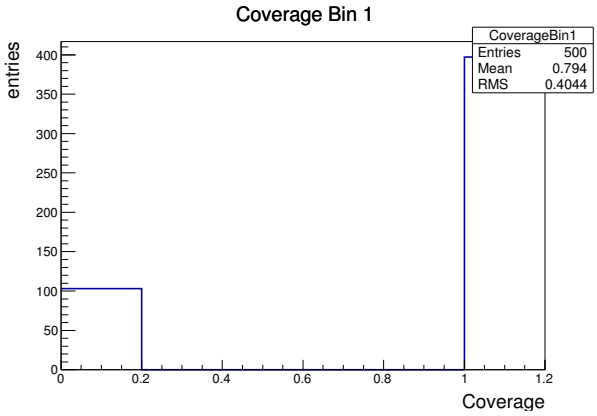
Bin-by-bin coverage represents the percentage of the truth values that lie within a certain error from the unfolded values at the corresponding Confidence Level (CL) for one bin. For instance, at 68% CL, we expect $\sim$ 68% of the truth values to be contained within one

sigma from the unfolded values. The developed test engine has the flexibility to calculate the bin-by-bin coverage at any CL set by the user.

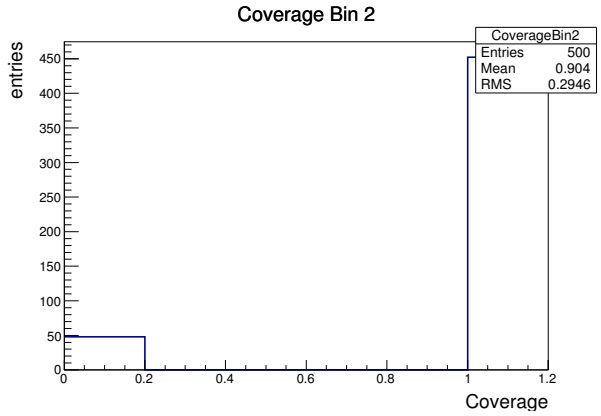
Figure 11.3 shows the bin-by-bin coverage at 95% CL of the unfolded three-bins momentum distribution of the electron antineutrino for 500 throws. Taking the statistical fluctuation of the 500 throws into consideration, we expect to have a coverage of 95 $\pm$ 4.5 % at the 95% CL.



(a) Coverage for bin 0.



(b) Coverage for bin 1.



(c) Coverage for bin 2.

Fig. 11.3 Bin-by-bin coverage of the unfolded momentum distribution of the electron antineutrino sample at 95% CL for 500 throws.

The results of this test indicate that the D'Agostini 2010 unfolding algorithm still suffers from an under-coverage problem due to the under-estimated uncertainties. However, it has much better results than the old D'Agostini 1995, as will be shown in Section 11.5.

11.1.2 Simultaneous Coverage

We rely on the empirical simultaneous coverage for the whole histogram with the Bonferroni correction, cf. [95, Section 5.2.2]. The Bonferroni correction accounts for the increase of type I error[1] of multiple coverage tests, one per bin, by decreasing the significance level of each test to be $\frac{\alpha}{n}$. Here α is the type I error level per bin and n is the total number of bins in the histogram.

Consequently, the confidence interval for each bin is then given by

$$\text{confidence interval} = \left[\hat{\lambda}_k - z_{1-\frac{\alpha}{2n}}\sqrt{\text{Var}(\hat{\lambda}_k)},\ \hat{\lambda}_k + z_{1-\frac{\alpha}{2n}}\sqrt{\text{Var}(\hat{\lambda}_k)}\right], \tag{11.1}$$

where $\hat{\lambda}_k$ is the expected value of the unfolded histogram for bin k, $Var(\hat{\lambda}_k)$ is its variance and $z_{1-\frac{\alpha}{2n}}$ is the standard normal quantile.

The simultaneous coverage test passes if the coverage test for all the bins passes at the modified significance level. The result of the simultaneous coverage test is shown in Fig. 11.4.

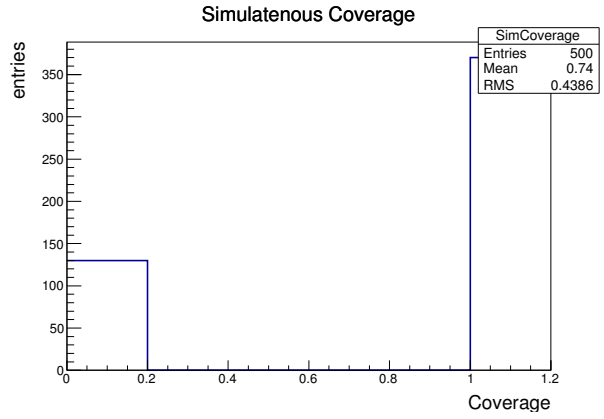


Fig. 11.4 Simultaneous coverage at 95% CL after 10 iterations.

The bin-by-bin under-coverage discussed in this section resulted in an overall simultaneous under-coverage.

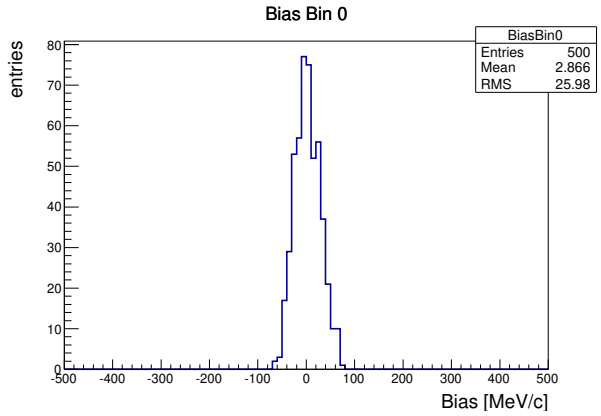
[1] A type I error rejects the hypothesis while it is true.

11.2 Bias

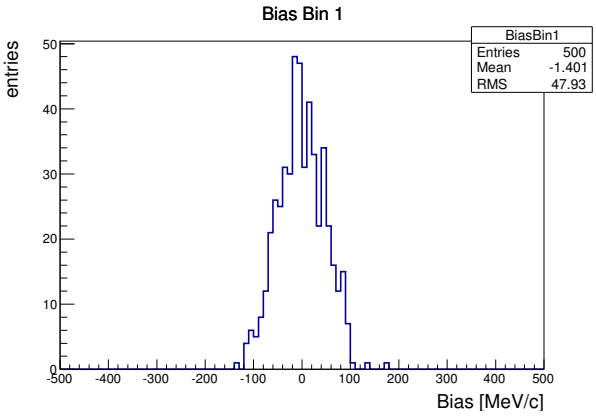
A bias-test measures the difference between the expected value of the unfolded histogram bins and its true values. The bias is defined as

$$\text{bias}_i = x_i^{unf} - x_i^{true}, \tag{11.2}$$

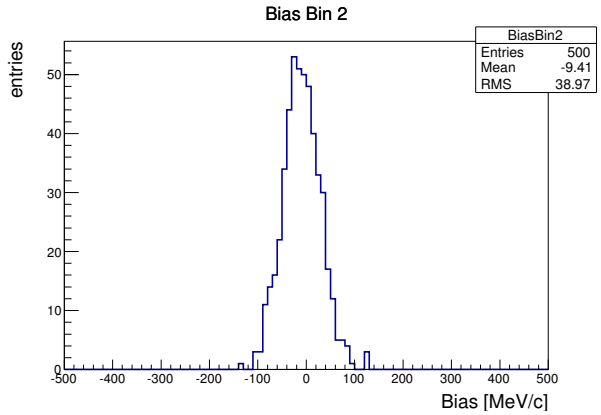
where x_i^{unf} and x_i^{true} are the unfolded and truth values of bin i, respectively. Figure 11.5 shows the bias test results for unfolded momentum distribution of the $\overline{\nu}_e$ sample.



(a) Bias for bin 0.



(b) Bias for bin 1.



(c) Bias for bin 2.

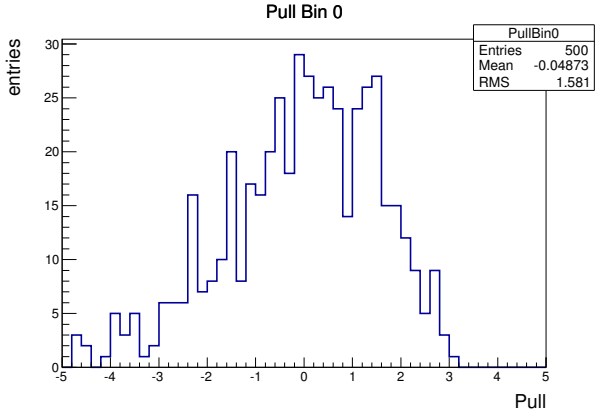
Fig. 11.5 Bin-by-bin Bias of the unfolded momentum distribution of the $\overline{\nu}_e$ sample.

The bias measurement depends on the unit of the measured quantity and does not give an intuition about how far the unfolded histogram is from the truth. Therefore, we construct

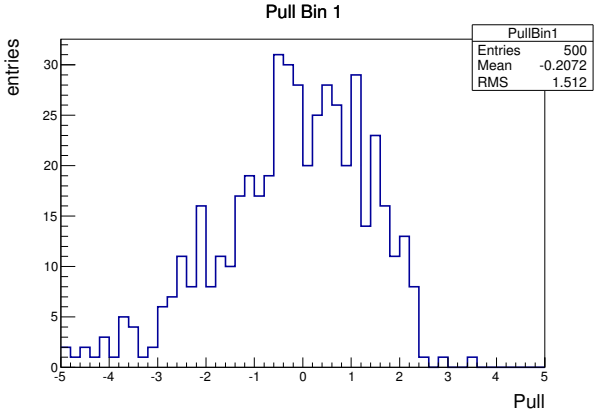
another dimensionless figure of merit, the pull, defined as

$$\text{pull}_i = \frac{\text{bias}_i}{\sigma_i}, \tag{11.3}$$

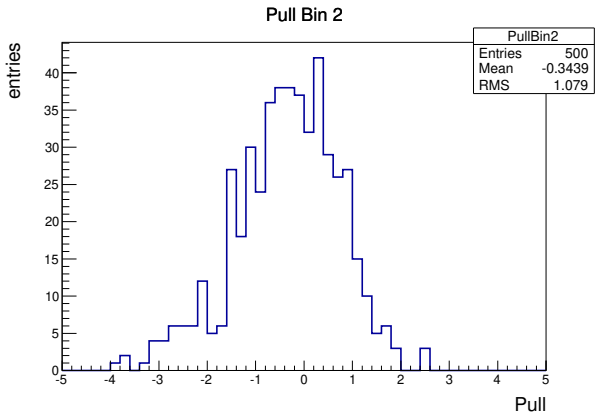
where σ_i is the uncertainty of the measured variable at bin i. Figure 11.6 shows the pull test results for the same unfolded histogram.



(a) Pull for bin 0.



(b) Pull for bin 1.



(c) Pull for bin 2.

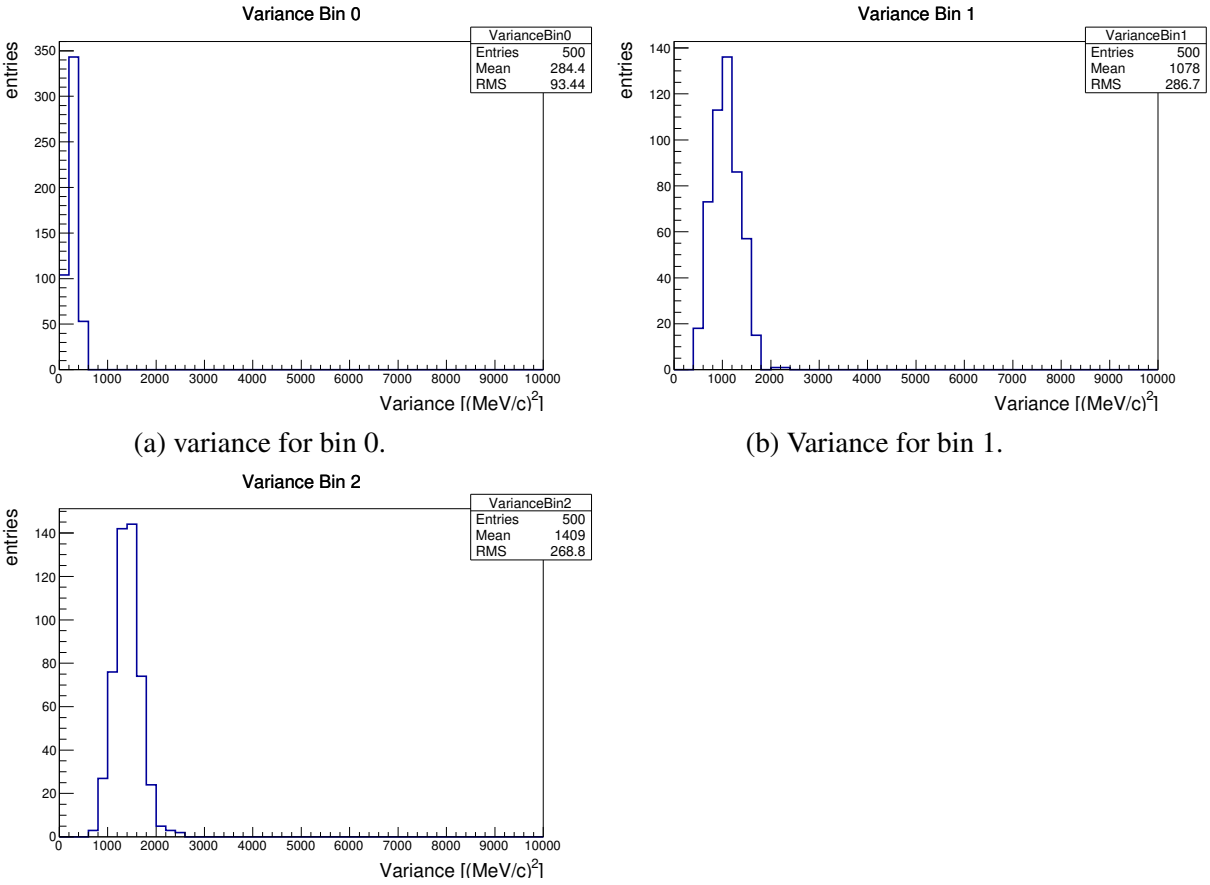
Fig. 11.6 Bin-by-bin pull of the unfolded momentum distribution of the $\overline{\nu}_e$ sample.

11.3 Variance

The variance test measures the expected value of the squared difference between the unfolded bin value and its mean value as

$$\mathrm{Var}_i = \frac{1}{M-1}\sum_{m=1}^{M}(x_{i,m}-\mu_i)^2. \tag{11.4}$$

Here M is the total number of unfolded histograms, μ_i is the mean unfolded value at bin i and the "-1" correction in the denominator gives an unbiased estimate of the sample variance. Figure 11.7 shows the variance test results for the same unfolded histogram.



(a) variance for bin 0.

(b) Variance for bin 1.

(c) Variance for bin 2.

Fig. 11.7 Bin-by-bin variance of the unfolded momentum distribution of the $\overline{\nu}_e$ sample.

Similar to the bias, the variance value depends on the measured quantity units and, hence, is less informative.

11.4 Chi-square

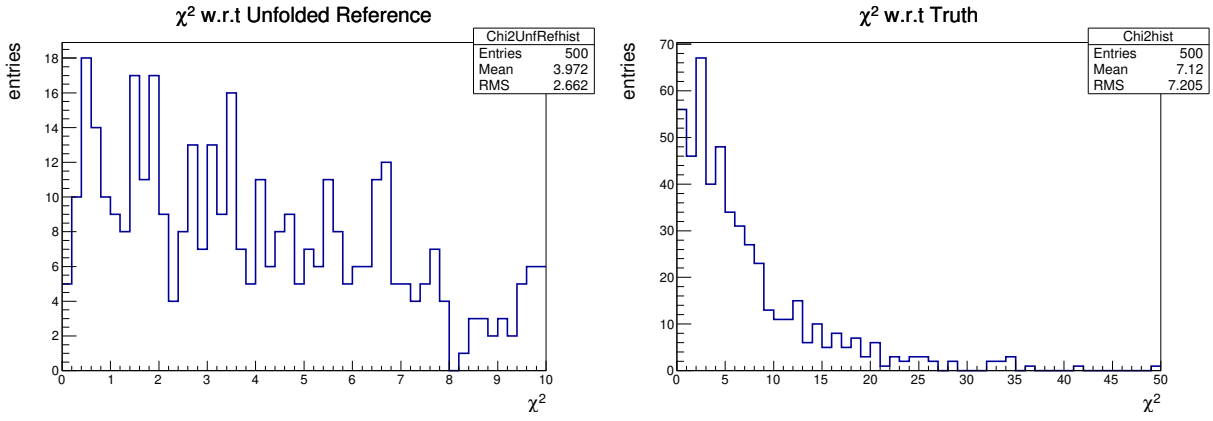
Two χ^2 tests are performed: one with respect to the reference unfolded histogram $\chi^{2^{ref}}$, and another with respect to the truth histogram $\chi^{2^{true}}$. The reference unfolded histogram is obtained by unfolding the reference input histogram before applying the Poissonian fluctuation. We implement a full χ^2 test, including the covariance matrix as

$$\chi^{2^{true}} = (x_i^{unf} - x_i^{true}) V_{ij}^{-1} (x_j^{unf} - x_j^{true}), \tag{11.5}$$

and

$$\chi^{2^{ref}} = (x_i^{unf} - x_i^{unf\ ref}) V_{ij}^{-1} (x_j^{unf} - x_j^{unf\ ref}). \tag{11.6}$$

Here x_i^{unf} , x_i^{true} and $x_i^{unf\ ref}$ are the unfolded, truth and unfolded reference value of bin i, respectively, and V_{ij}^{-1} is the ij element of the inverse of the covariance matrix. The results of the χ^2 tests are shown in Fig. 11.8.



(a) χ^2 w.r.t the reference histogram.

(b) χ^2 w.r.t the truth histogram.

Fig. 11.8 χ^2 of the unfolded momentum distribution of the $\overline{\nu}_e$ sample with respect to the reference unfolded histogram (left) and the truth histogram (right).

The average χ^2 with respect to the truth histogram is 7.12 for a three-bins unfolded histogram, leading to a χ^2/NDF of 2.373 (see Fig. 11.8b).

Furthermore, we expect that the calculated χ^2 with respect to the reference histogram, due to statistical fluctuation, follows a χ^2 distribution where the NDF (the mean of the distribution) is similar to the total number of unfolded bins. The correlation between the unfolded bins as well as the constraint that the total number of unfolded events is fixed, causes the NDF of the χ^2 distribution to deviate from the total number of unfolded bins (see Fig. 11.8a).

11.5 Effect of Number of Iterations

The number of iterations is a crucial parameter in the old D'Agostini algorithm. In particular, a small number of iterations results in a bias toward the prior and in under-coverage, cf.[95, Section 5.2.2] and Fig. 11.11. Whereas a larger number of iterations amplifies the statistical fluctuation, as discussed in Section 10.2.5.

The aim of these tests is to check the effect of the number of iterations on both the χ^2 value (convergence test), and on the coverage. If we set our initial prior of the D'Agostini 2010 algorithm to the MC truth, we will still expect to have the best χ^2 and coverage after applying only one iteration. Nevertheless, to avoid a model bias we perform the test starting from a wrong prior where all cause bins are equally probable. Subsequently, we chose the number of iterations at which the unfolded histogram is as close as possible to the truth. At this point, the algorithm forgets the prior, to a good extent, and is driven by the observed data. This happens when the measured χ^2 reaches its minimum and the coverage is at its maximum.

Figure 11.9 shows that the number of iterations is not vital in the new D'Agostini algorithm as the χ^2 of the unfolded histogram almost reaches the minimum after the first

5-10 iterations. Furthermore, the dominance of the statistical fluctuations at a large number of iterations is removed by the added smoothing step.

The effect of the number of iterations on the χ^2 and on the simultaneous coverage at 95% CL is shown in Fig. 11.9 and Fig. 11.10 respectively.

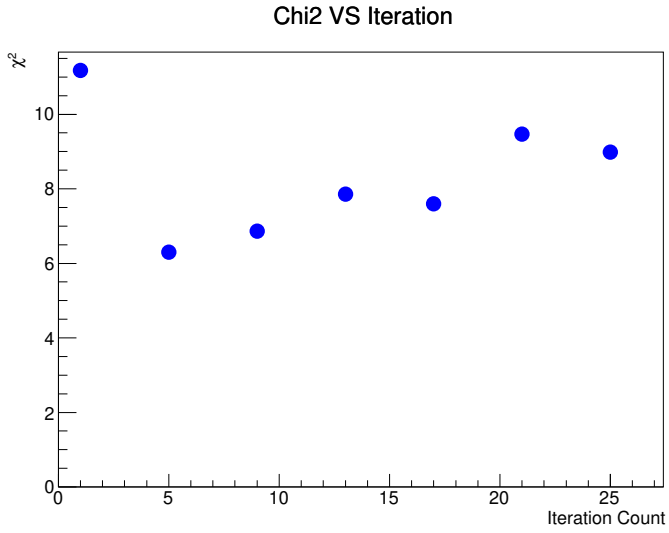


Fig. 11.9 χ^2 w.r.t the truth histogram versus number of iterations.

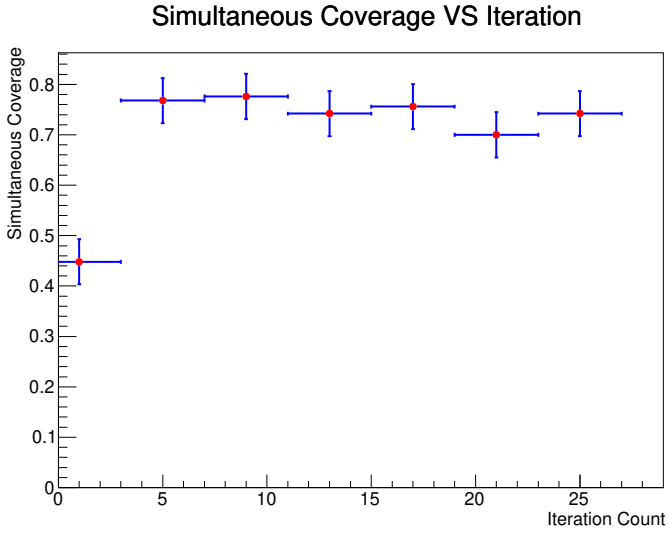


Fig. 11.10 Simultaneous coverage at 95% CL versus number of iterations.

In order to avoid a bias toward the MC prior, to maximize the coverage and to minimize the chi-square, we chose to apply 10 iterations in our cross-section analysis.

The old D'Agostini 1995 unfolding algorithm simultaneous coverage is shown in Fig. 11.11. Comparing this with Fig. 11.10 we conclude that the new algorithm has a better coverage.

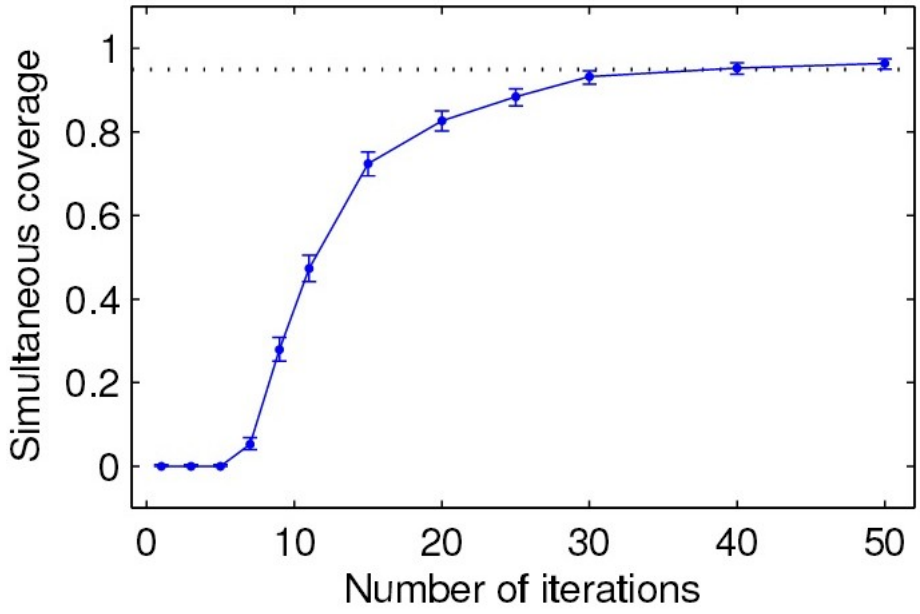


Fig. 11.11 Simultaneous coverage at 95% CL versus the number of iterations for the old D'Agostini 1995 algorithm. Figure courtesy of [95].

11.6 Conclusion

The previous tests show that the performance of the new D'Agostini 2010 algorithm is superior to the 1995 version. The KDE played an important role in stabilizing the unfolding process. Table 11.1 summarizes the results of the performance tests applied to the new D'Agostini algorithm with 10 iterations and a KDE smoother.

χ^2 w.r.t Truth /Nb Bins	χ^2 w.r.t unfolded reference /Nb Bins	Simultaneous Coverage (at CL=95%)
7.12/3	3.97/3	0.74

Table 11.1 Summary of the performance tests results applied on the new D'Agostini algorithm configured with 10 iterations and a KDE smoother with no mirror and a $\rho = 0.6$.

Chapter 12

Systematic Uncertainties and Error Propagation

The main sources of systematics uncertainties are:

- flux uncertainties,
- uncertainties in the underlying interaction theoretical models,
- uncertainties in detector acceptances, calibration or resolution, and
- mis-modelled background.

The total effect of each of these systematic sources on the extracted cross-section will be shown in Section 13.2. This chapter provides a brief overview of the main systematic uncertainties in an $\overline{\nu}_e$ cross-section measurement.

12.1 Flux Uncertainties

In T2K, the largest source of systematic uncertainty in a cross-section measurement comes from the flux uncertainty. The (anti)neutrino flux production is described in Section 5.1.2.

The flux uncertainty is dominated by the hadron interaction; mainly, the pion and kaon kinematics and multiplicities. Other sources of flux uncertainties include the number of protons that hit the target, the proton beam profile, the target material modelling, the horn current and magnetic field, the beam alignment with the horn and the target, and the neutrino off-axis angle.

Figure 12.1 shows the different sources of flux uncertainties, as a function of the true neutrino energy.

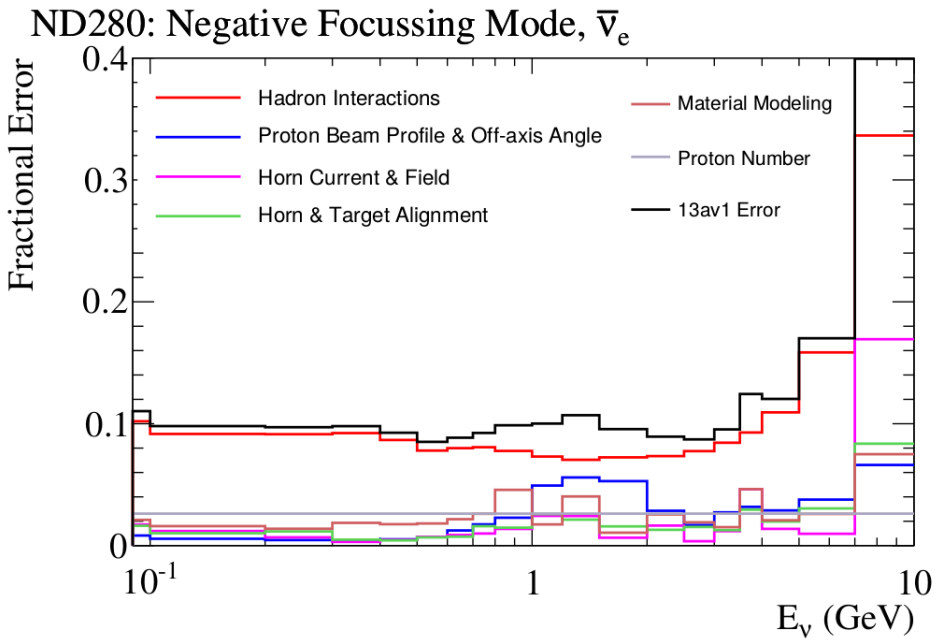


Fig. 12.1 The total uncertainties on the ND280 $\overline{\nu}_e$ flux prediction, in case of antineutrino operation mode, as a function of the true neutrino energy E_ν, is shown in black line. The main contributors to this uncertainty, including the hadron interactions, beam profile, horn current and field, the beam alignment, the material modelling and the number of protons are represented by the red, blue, pink, green, break and grey lines respectively. Figure courtesy of [105].

T2K uses external data from the NA61/SHINE [106] and the HARP [107] experiments to fine tune the nominal flux and constrain the associated uncertainties of the hadron interactions.

12.2 Model Uncertainties

The cross-section and Final State Interaction (FSI) uncertainties are evaluated based on the default NEUT [9] parameterization complemented by the recommendation of the Neutrino Interaction Working Group (NIWG) in T2K, as described in [36].

12.2.1 Cross-section Model Uncertainties

NIWG used external data from the Minerva [39] and MiniBooNE [15] experiments to fit three cross-section models: SF + 2p–2h, RFG + relativistic RPA + 2p–2h and RFG + non relativistic RPA + 2p–2h. For more details about these models, please refer to Chapter 3. The fit results show poor agreement between any of these models and the data [108]. Nevertheless, the RFG + relativistic RPA + 2p–2h model performed slightly better than the others, and hence was used since then in this T2K cross-section analysis.

Table 12.1 summarizes the cross-section model systematics used in this analysis.

Systematic Name	Description
NXSec_MaCCQE	Axial mass in CCQE interaction. It parameterizes the nucleon form factor as in Eq. (2.15).
NIWG2014a_pF_C12	Carbon Fermi momentum, as described in Section 3.1.
NIWGMEC_Norm_C12	Carbon 2p–2h normalization, as described in Section 2.3.
NIWG2014a_Eb_C12	Carbon binding energy, cf. Section 3.1.
NIWG2014a_pF_O16	Oxygen Fermi momentum, cf. Section 3.1.
NIWGMEC_Norm_O16	Oxygen 2p–2h normalization, cf. Section 2.3.
NIWG2014a_Eb_O16	Oxygen binding energy, cf. Section 3.1.
NXSec_CA5RES	C_5^A form factor for single pion resonance, as described in Section 2.4.
NXSec_MaNFFRES	Resonance axial mass, as in Eq. (2.25).
NXSec_BgSclRES	Isospin = $\frac{1}{2}$ background scale factor for single pions resonance.
NIWG2012a_ccnueE0	Normalization factor for CC-nue.
NIWG2012a_dismpishp	Normalization factor for DIS multi-pi, cf. Section 2.5.
NIWG2012a_cccohE0	Normalization factor for CC-Coherent, as described in Section 2.7.
NIWG2012a_nccohE0	Normalization factor for NC-Coherent, cf. Section 2.7.
NIWG2012a_ncotherE0	Normalization factor for NC-Other normalization, cf. Section 2.7.

Table 12.1 Cross-section model systematics affecting the $\overline{\nu}_e$ analysis.

12.2.2 Final State Interaction Uncertainties

The final state interaction (FSI) is discussed in Section 2.6. Table 12.2 summarizes the main FSI uncertainties affecting the $\overline{\nu}_e$ cross-section measurement.

Systematic Name	Description
NCasc_FrInelLow_pi	Inelastic re-scattering probability for low energy pions.
NCasc_FrInelHigh_pi	Inelastic re-scattering probability for high energy pions.
NCasc_FrPiProd_pi	Pion production probability.
NCasc_FrAbs_pi	Pion absorption probability.
NCasc_FrCExLow_pi	Charge exchange probability for low energy pions.
NCasc_FrCExHigh_pi	Charge exchange probability for high energy pions.

Table 12.2 FSI systematics affecting the $\overline{\nu}_e$ analysis.

12.3 Detector Systematics

There are two types of detector systematics: variation and weight. The variation systematics are applied by modifying the reconstructed quantity according to its uncertainty for each event and rerunning the selection process on the new values. For instance, a modified reconstructed quantity x' is related to its original value x by

$$x' = x(s + \alpha\delta_x). \tag{12.1}$$

Here s is the scale factor, normally set to one, α is a random number following the systematic PDF, normally a Gaussian with mean equal to zero and variance equal to one (except in case of magnetic field systematic where a uniform PDF is used instead) and δ_x is the uncertainty of the corresponding quantity x.

The weight systematics are implemented via an event-by-event re-weighting, without the need to rerun the selection. For example, a normalization weight systematic w for a quantity x, is given by

$$w = 1 + \alpha\delta_x, \tag{12.2}$$

where α and δ_x have the same interpretation as in Eq. (12.1).

The detector uncertainties include the detector mis-reconstruction and inefficiencies affecting both the rate and shape of the cross-section.

ND280 detector systematics for the $\overline{\nu}_e$ analysis is studied in [76, Section 5] and shares many systematics with that of the ν_μ [109]. Moreover, the ECAL systematics are described in detail, in [110].

Magnetic Field Distortions

The track momentum is calculated from its curvature due to the applied magnetic field, as discussed in Section 5.3.5. Any non-uniformity in the magnetic field results in an uncertainty in the measured momentum.

TPC Momentum Resolution

Finite detector resolution smears the measured momentum, creating a difference between the data and MC.

TPC Momentum Scale

The uncertainty on the absolute strength of magnetic field results in an uncertainty in the momentum scale.

TPC PID

The uncertainty on the measured dE/dx in the TPCs affects the selection efficiency and purity. This systematic is included in the BDT test statistic systematic evaluation and the TPC PID variables data to MC agreement is discussed in Section 7.4.

FGD PID

The uncertainty on this quantity creates a difference between the selected number of events in data and MC. This systematic is included in the BDT test statistic systematic evaluation and the FGD PID variables data to MC agreement is discussed in Section 7.4.

ECAL PID

For the ECAL, any uncertainty on the measured electromagnetic energy or the log likelihood ratio distinguishing between muons and electrons or between electrons and protons (LLR_MIP_EM and LLR_EM_HIP), results in a difference between the number of selected events in data and MC. The ECAL variables are described in detail in Section 7.1.3 and their data to MC agreement is discussed in Section 7.4.

Time of Flight

The Time of Flight (ToF) measures the average time between hits in two different sub-detectors. Knowing the travelled distance and the particle momentum, ToF can be used to estimate the particle mass. The uncertainty on ToF affects the purity and efficiency of the selection.

TPC Cluster Efficiency

The probability of finding a cluster of TPC hits in the data could be different compared to the MC, affecting the TPC track quality cut. Consequently, this uncertainty results in a different number of events that passes the selection process.

TPC Track Efficiency

The probability of a successful TPC track reconstruction could be different in data and MC for each of the three TPCs. This uncertainty affects the number of selected events.

TPC Charge ID Efficiency

The TPC reconstruction fails to successfully compute the track charge from its curvature, if the track was very straight, backwards-going or spiral. Consequently, the selected number of events and the cross-section will be affected by this uncertainty.

FGD-TPC Matching Efficiency

The uncertainty on the efficiency of matching a track across an FGD and a TPC affects the number of selected events and hence the cross-section measurement.

TPC-ECAL Matching Efficiency

The uncertainty on the efficiency of matching a track across a TPC and an ECAL affects the number of selected events and hence the cross-section measurement.

Pion Secondary Interactions

The pion secondary interaction refers to the pion produced from the neutrino interaction in the ND280 that interacts elsewhere in the detector. These pions could suffer from absorption and charge exchange when interacting in the detector resulting in detection inefficiency. The

probability of a pion secondary interaction depends on its energy and on the target nuclei of the detector, as described in [111].

Proton Secondary Interactions

Protons represent one of the main backgrounds in the $\overline{\nu}_e$ selection as discussed in Section 8.4.1. The uncertainty on the interaction of this background in the detector results in an uncertainty of the purity and the efficiency of the selected events, as discussed in [78].

FGD Mass

The FGDs provide target mass for neutrino interactions, as described in Section 5.3.3. Consequently, the uncertainty of on the number of nucleons in the FGD directly affects the measured cross-section (see Eq. (1.24)).

Pile-up

The ND280 MC does not simulate the cosmic interactions nor the neutrino interactions that happen outside the ND280 detector (sand interactions). Consequently, if a cosmic ray or a sand interaction was coincident with a neutrino interaction from the beam bunch, the neutrino event will be wrongly discarded and considered as a background by one of the veto cuts described in Section 8.2. Although the pile-up correction is applied to the MC, there remains uncertainty on this correction, resulting in a systematic error. Furthermore, the pile-up systematic is correlated with the beam intensity, horn current and sand interaction. The pile-up systematic is broken down into TPC Pile-up, PØD-PØD and ECAL-FGD1 Pile-up, and ECAL Pile-up according to the associated veto cut.

Out of Fiducial Volume Background (OOFV)

The main OOFV background is the OOFV gamma, described in detail in Section 8.4.3. The uncertainty on this background directly affects the extracted number of signal events.

Sand Interactions

The interaction rate of the neutrino outside the ND280 detector and inside the pit walls is modelled by a separate MC (Sand-MC). The uncertainty on the sand interaction is taken to be the difference between the sand-data and sand-MC. This uncertainty affects the signal event rate.

BDT PID

The uncertainties of the input variables are propagated to the calculated test statistic (tBDT). Consequently, the uncertainty of the test statistic affects the number of events passed. Ideally, after evaluating the uncertainty of the tBDT, as described in Chapter 9, we should turn off the systematic uncertainties of the BDT's input variables to avoid double counting, and only enable the tBDT systematic. However, in this analysis, we left the momentum-related systematics on, as the momentum variable is used by other cuts and vetos that we need to evaluate their systematics.

Table 12.3 summarizes the ND280 detector systematics used in this analysis along with their relative errors on the selected $\overline{\nu}_e$ signal number of events. For each of these systematics, the relative error as a function of the selected positron momentum is shown in Fig. 12.2 to Fig. 12.16.

The BDT systematic error is the dominant error source since it includes all the TPC, FGD and ECAL systematics along with their correlations, which is normally neglected in a variable-by-variable systematics error evaluation. Nevertheless, the bin-to-bin error may be bigger than the integrated uncertainty.

Systematic	Type	Integrated relative error (%)
B Field Distortions	Variation	0.81
TPC Momentum Resolution	Variation	2.00
TPC Momentum Scale	Variation	0.57
TPC PID	Variation	(included in the tBDT)
ECAL EM Energy	Weight	(included in the tBDT)
ECAL PID	Weight	(included in the tBDT)
FGD PID	Variation	(included in the tBDT)
BDT PID	Variation	3.38
Time Of Flight	Variation	0.69
TPC Cluster Efficiency	Weight	~ 0
TPC Track Efficiency	Weight	0.83
TPC Charge ID Efficiency	Weight	0.55
TPC-FGD Matching Efficiency	Weight	0.11
TPC-ECAL Matching Efficiency	Weight	2.77
Pion Secondary Interactions	Weight	0.05
Proton Secondary Interactions	Weight	0.05
FGD Mass	Weight	0.61
TPC Pile-up	Weight	0.12
P0D-P0D ECAL-FGD1 Pile-up	Weight	0.23
ECAL Pile-up	Weight	0.08
OOFV	Weight	~ 0
Sand Interactions	Weight	~ 0

Table 12.3 ND280 detector systematics affecting the $\overline{\nu}_e$ analysis along with their types and effects on the selected number of the $\overline{\nu}_e$ charged-current events.

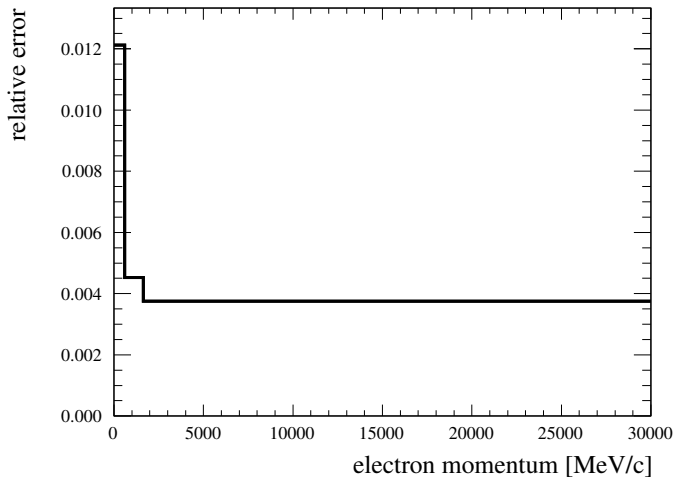


Fig. 12.2 The B field systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

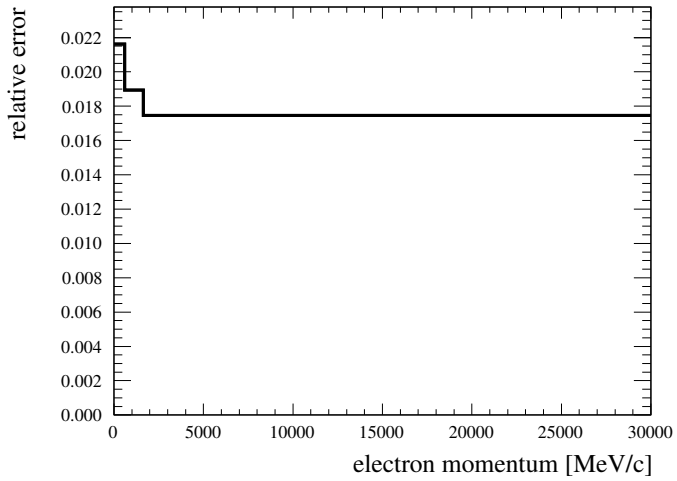


Fig. 12.3 The TPC momentum resolution systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

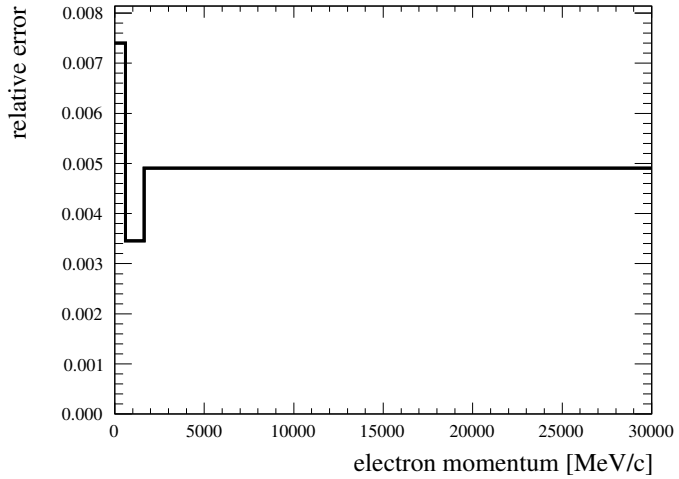


Fig. 12.4 The TPC momentum scale systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

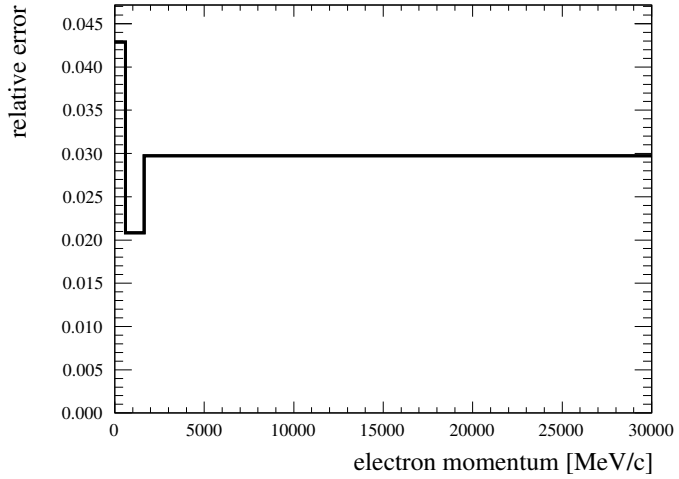


Fig. 12.5 The BDT test statistic systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

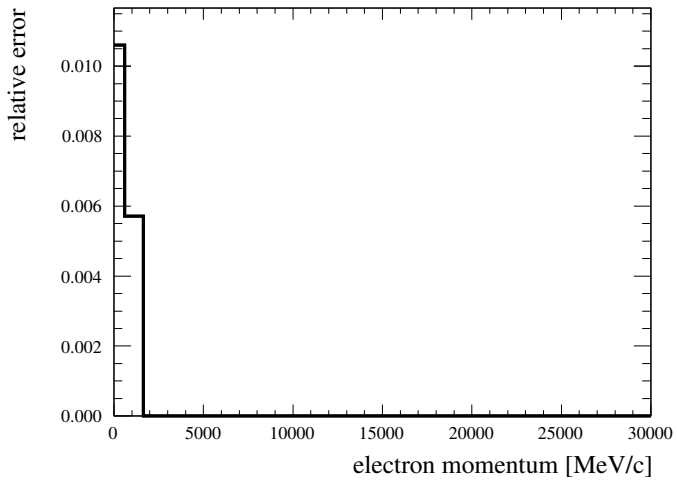


Fig. 12.6 The time of flight (ToF) systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

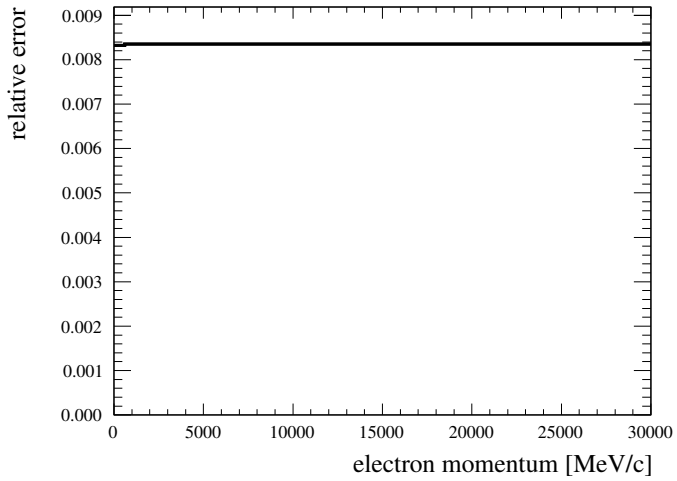


Fig. 12.7 The TPC track efficiency systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

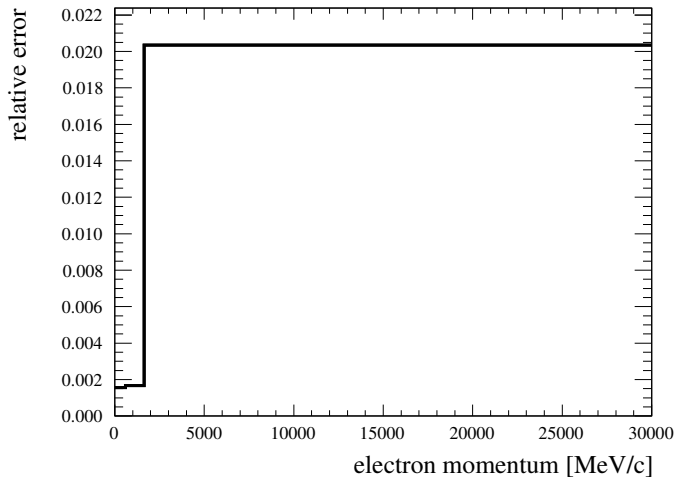


Fig. 12.8 The charge ID efficiency systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

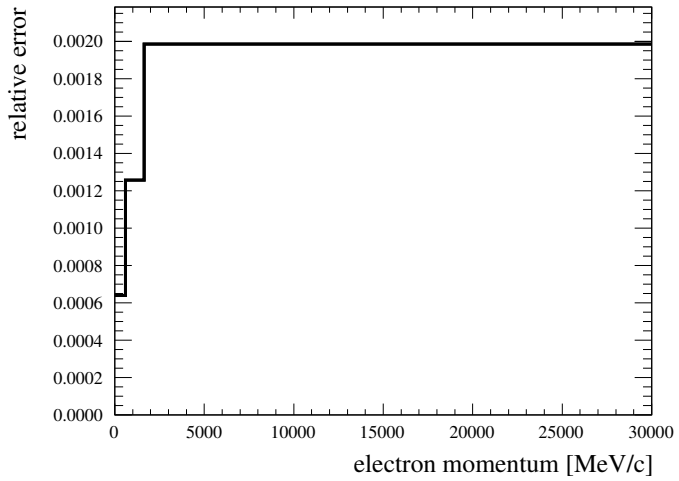


Fig. 12.9 The TPC-FGD matching efficiency systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

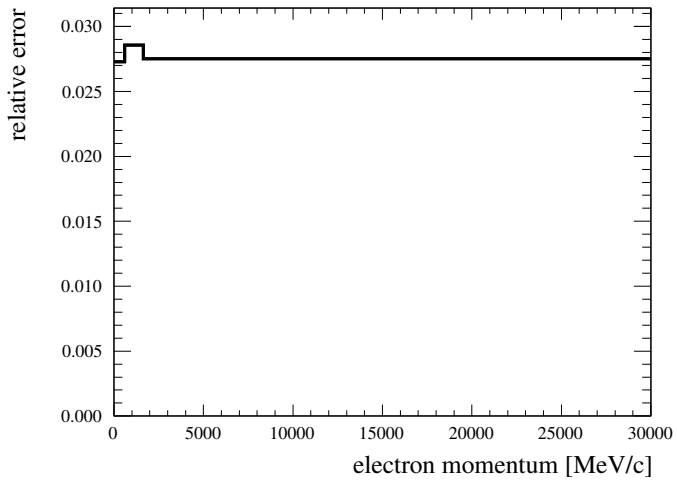


Fig. 12.10 The TPC-ECAL matching efficiency systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

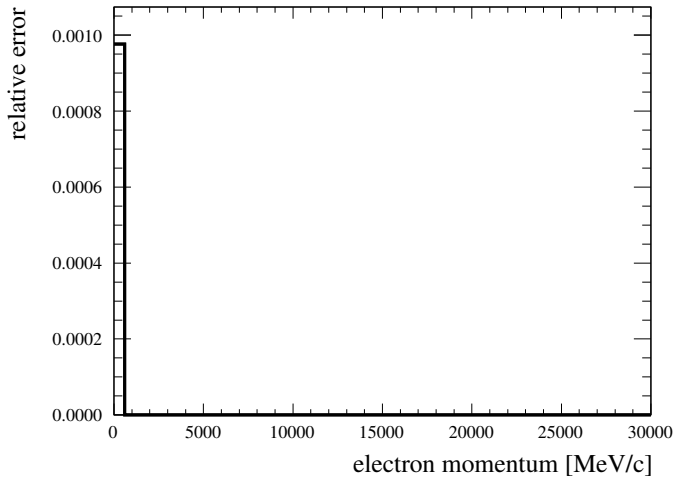


Fig. 12.11 The pion secondary interaction systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

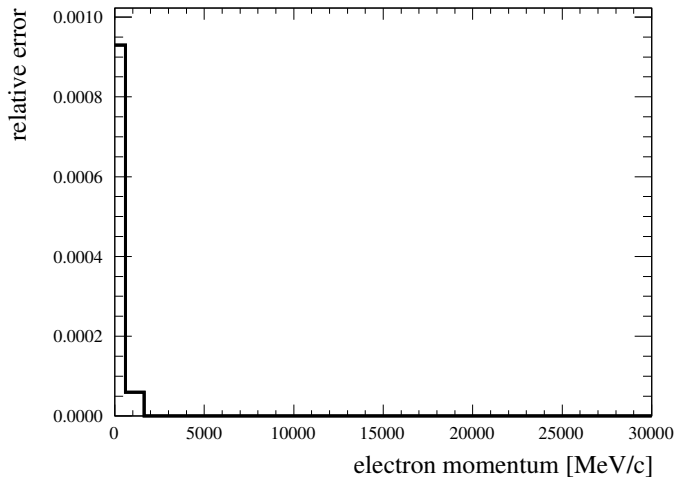


Fig. 12.12 The proton secondary interaction systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

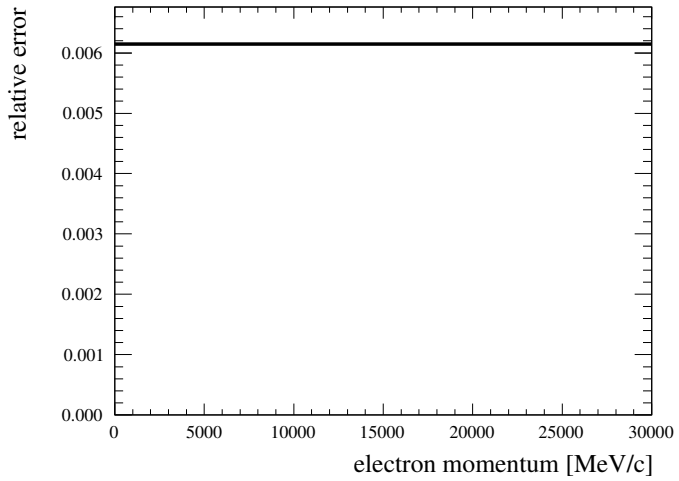


Fig. 12.13 The FGD mass systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

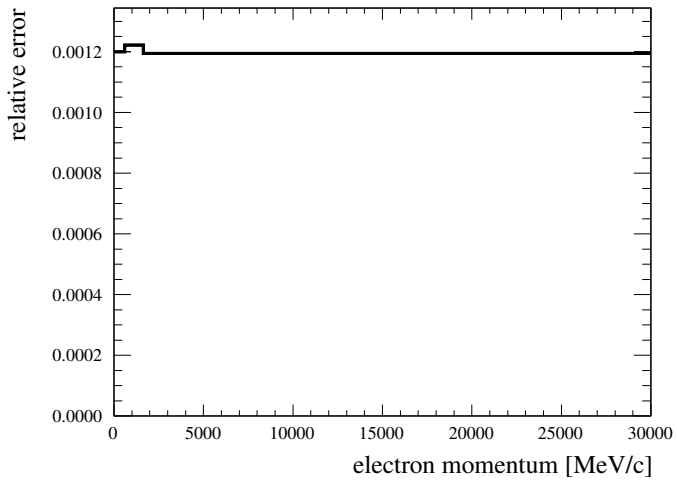


Fig. 12.14 The TPC pile-up systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

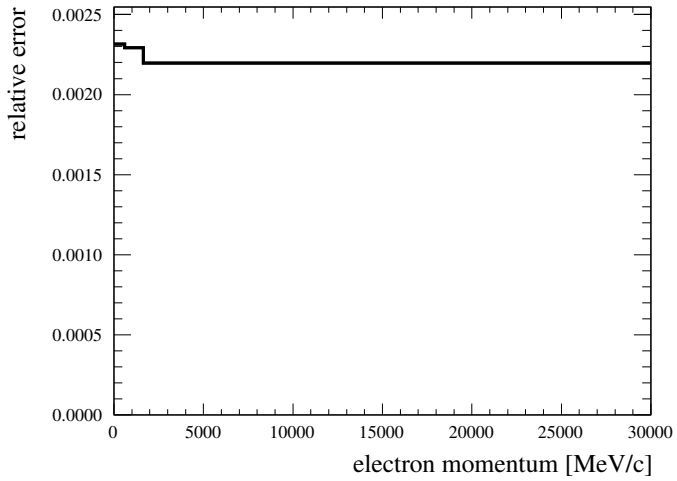


Fig. 12.15 The P0D pile-up systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

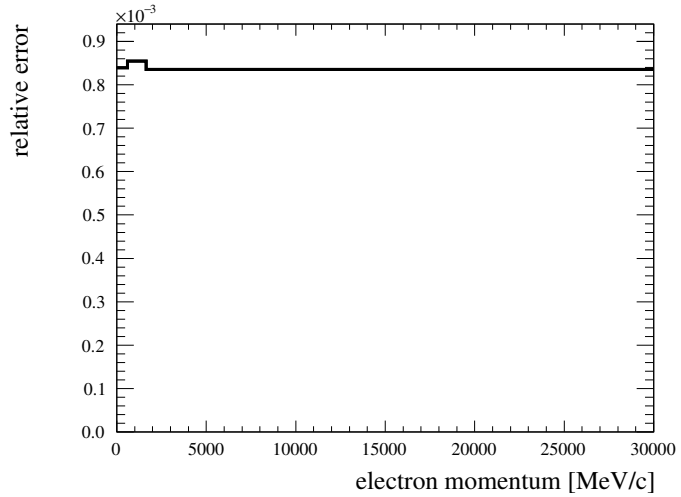


Fig. 12.16 The ECAL pile-up systematic, as a function of the selected positron's momentum, on the selected CCinc $\overline{\nu}_e$ events.

12.4 Error Propagation

The xsTool evaluates the estimated errors on the unfolded quantity by creating a modified version of the experiment (a throw) and changing the affected input of the unfolding process, smearing matrix, background or observed histogram, following its uncertainty. Subsequently, it repeats the unfolding process for all throws and calculates the covariance matrix due to the input variation. The xsTool categorizes the possible sources of uncertainties into five independent groups: statistical errors, the systematic uncertainties on the flux, cross-section model, FSI, and detector response. The covariance matrix is computed individually for each group of errors. Subsequently, the covariance matrices are summed to form the overall covariance matrix, where the total uncertainty on the unfolded bin will be given by the square root of the corresponding element of the diagonal of the overall covariance matrix.

The data statistical errors affect the number of events in the observed histogram and are treated by the xsTool using Poisson throws, where all the bins are assumed to be independent of each other. The xsTool evaluates the MC statistical error by throwing different truth

and reconstructed histograms and building from them different smearing matrices having different efficiencies and background histograms. Furthermore, the xsTool evaluates the flux, cross-section, and FSI systematics using a software module called T2KReWeight. The T2KReWeight supports the change of systematic parameters by implementing a dial to tune the values of these parameters to those other than nominal. The correlation between these parameters is taken into consideration by using the corresponding parameter covariance matrix. The re-weight engine assigns different weights to each selected event according to the parameters' covariance matrix. The flux uncertainties and its covariance matrix are provided by the beam group while the cross-section model and FSI parameters' and covariance matrix are provided by the NIWG.

The detector uncertainties, variation or weight systematics, are calculated and propagated via the ND280 software package called Propagation of SYstematics and CHaracterization of Events (PSYCHE). PSYCHE is a part of the high-level analysis framework in ND280 (HighLAND2)[70]. In PSYCHE, all the detector systematics are thrown simultaneously, allowing the manifestation of the systematics correlation. The detector systematic affects the cross-section measurement by modifying the smearing matrix, described in Eq. (10.2).

Chapter 13

Results and Conclusion

13.1 Data Runs

This analysis was performed on the data collected during the anti-neutrino operation mode characterized by the Reverse Horn Current (RHC). Normally, the simulated MC is ~ 10 times the corresponding data and the total amount of processed good Protons On Target (POT) for data and MC are 6.11×10^{20} and 7.20×10^{21}, respectively. The detailed decomposition of the collected data is shown in Fig. 13.1 and summarized in Table 13.1.

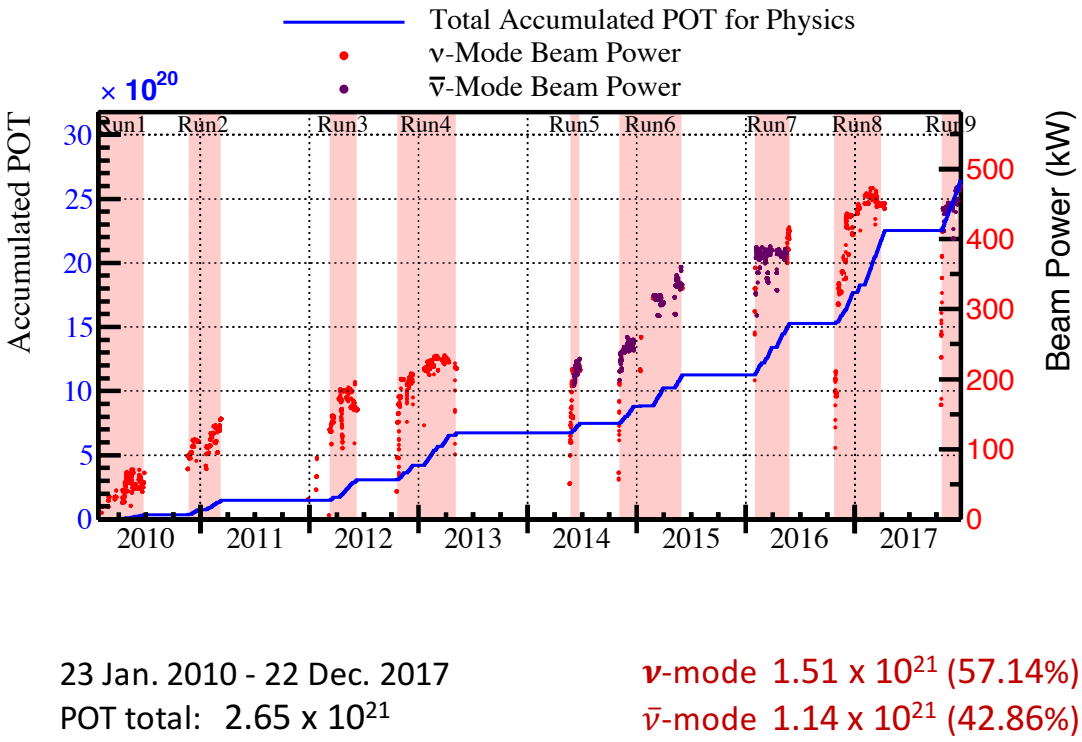


Fig. 13.1 ND280 data taking periods from 2010 until 2017. The neutrino and the anti-neutrino mode beam power are shown in purple and red dots respectively. The blue curve represents the accumulated Proton On Target (POT) that passed the data-quality checks.

ND280 run	Data POT	MC POT	Data/MC
Run 5c	0.43×10^{20}	4.7×10^{20}	0.091
Run 6b	1.27×10^{20}	1.42×10^{21}	0.090
Run 6c	0.50×10^{20}	5.27×10^{20}	0.095
Run 6d	0.64×10^{20}	6.88×10^{20}	0.093
Run 6e	0.83×10^{20}	8.59×10^{20}	0.096
Run 7b	2.44×10^{20}	3.24×10^{21}	0.075

Table 13.1 ND280 data (with good data quality) and MC runs used in this analysis.

13.2 Unfolding Results

We use the developed improved D'Agostini unfolding engine to extract the $\overline{\nu}_e$ cross-section. In this process, we apply a simultaneous unfolding for the signal and the background, where all different background sources in the smearing matrix are grouped in one cause bin.

The xsTool propagates different sources of systematic and statistical errors in the final cross-section results. The systematic sources include 500 toy experiments of statistical fluctuation for the observed data, 500 for the statistical fluctuation of the input MC, 100 for the Final State Interaction (FSI) systematics, 100 for the cross-section systematics, 100 for the flux, and another 100 for the detector systematics. Each of these toy experiments represents an unfolding problem and the unfolded histogram and systematic covariance are calculated.

The xsTool generates all these types of error sources except for that of the detector response, which is taken from the main ND280 analysis software (HighLand).

The binning is selected as follows: bin 0 represents the out-of-range bin that collects any underflow or overflow reconstructed events. This bin should be empty or have very few events. The other three bins represent positron momenta of 0-600 MeV/c, 600-1650 MeV/c and 1650-30000 MeV/c. This specific binning was chosen such that there are sufficient statistics in each bin, not less than 15 data events, and to study specific background effects on the selection. The first bin was in fact highly contaminated with the gamma background, while the second bin contains most of the surviving protons.

The unfolded differential cross section as a function of the selected positron's momentum $\frac{\partial\sigma}{\partial p_{e^+}}$ is shown in Fig. 13.2.

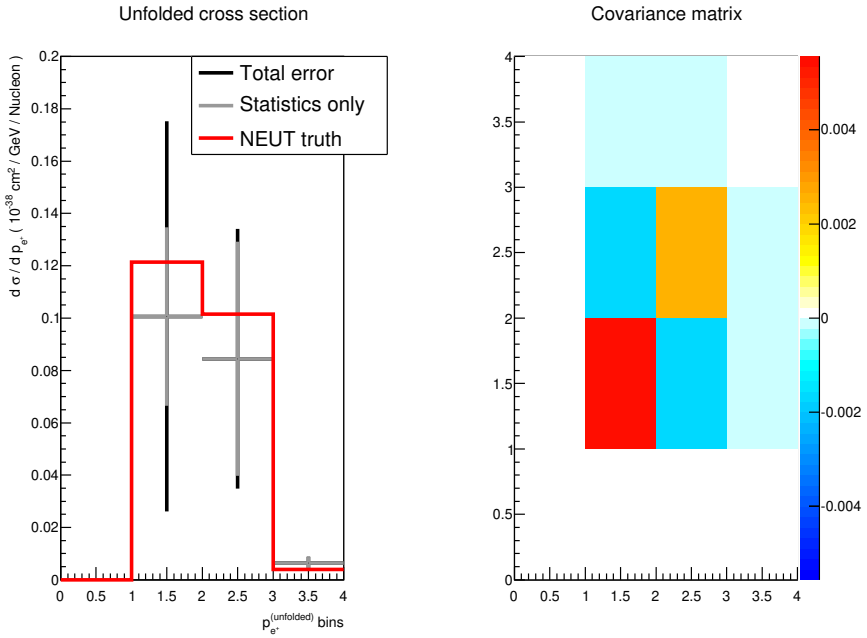


Fig. 13.2 Unfolded differential cross-section of the electron anti-neutrino with respect to the positron momentum (left). The plot shows the truth (in red), unfolded cross-section value (in black point), and the gray error bars represent the statistical error, while the total error bars are shown in black. On the right, the total covariance matrix due to statistical, detector, FSI, cross-section and flux errors.

Shown in Fig. 13.3 is the total covariance matrix decomposed into its different systematic covariance matrices sources.

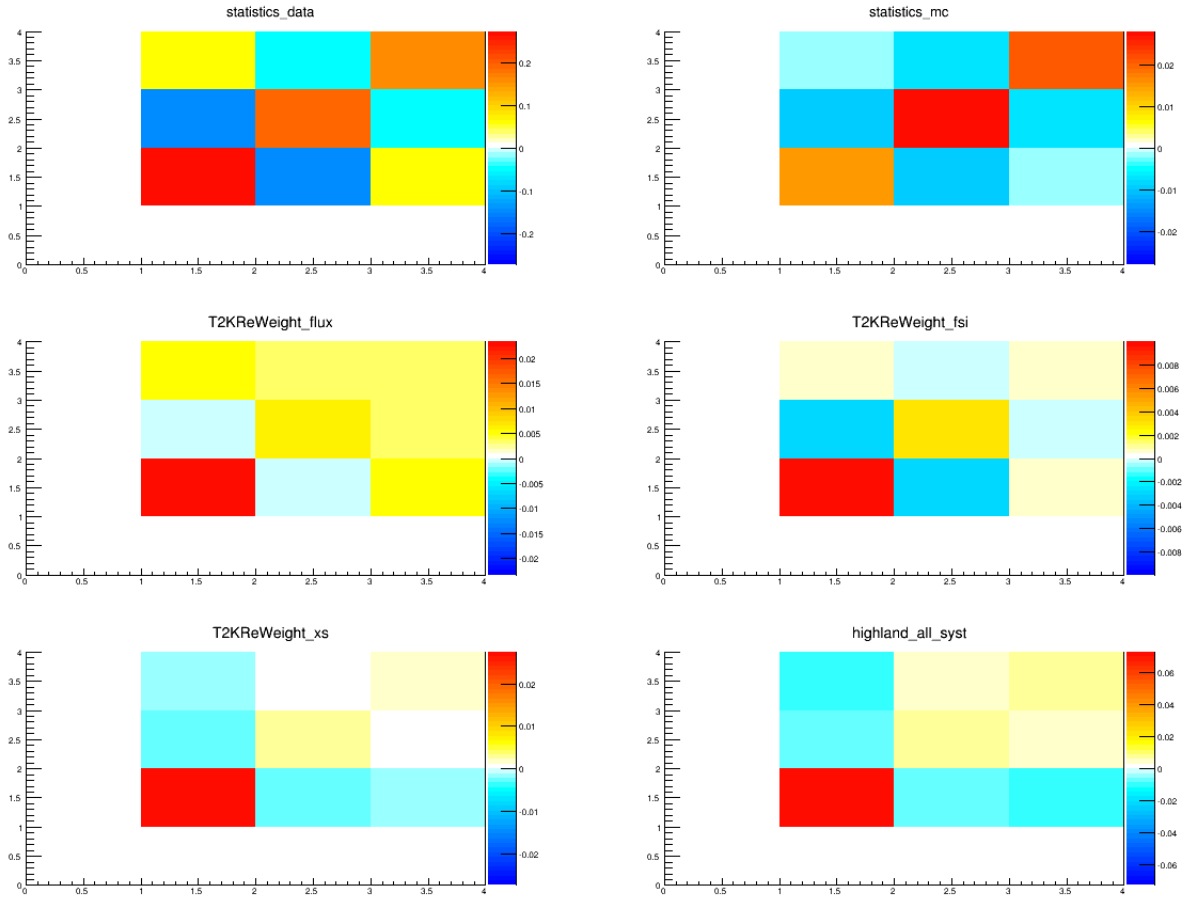


Fig. 13.3 The sources of errors in an unfolding cross-section extraction process. The statistical error in data (top left) and MC (top right), as well as the flux (middle left) , FSI (middle right), cross-section model parameters (bottom left) and detector systematics (bottom right) are shown.

The total measured electron anti-neutrino cross-section is 0.332 ± 0.052 (15.8% stat) $\pm$ 0.025 (7.5% syst) $\times 10^{-38}$ cm^2/nucleon. In comparison, the simulated cross-section is 0.290×10^{-38} cm^2/nucleon. The measured cross-section agrees with the MC simulation within statistical and systematic errors.

In this measurement, the statistical uncertainty is dominant and $\sim$ 14.7% for data and 5.6% for MC. Furthermore, the detector, flux, cross-section and Final State Interaction (FSI) systematics are $\sim$ 5.0%, 4.1%, 3.0% and 1.8%, respectively.

It is worth noting that the flux uncertainty is overestimated. As per the time of writing this thesis, the main analysis software, HighLand, required enabling the flux systematics in order to apply the appropriate flux correction to the data. Highland does not take the correlation of flux bins into account and we thus need to calculate this systematic properly using the xsTool. This Highland "bug" will be fixed in an upcoming analysis software release.

13.3 Sideband Effect

13.3.1 Gamma Control Sample

The gamma events are the dominant source of background for our signal, as described in Section 8.4.3. In our analysis, we therefore constructed a gamma control sample (sideband) to select these events and to use it in our simultaneous unfolding framework to constrain the poorly modelled gamma background. In the gamma selection, we pick the events that:

- passed the BDT cut, as described in Section 8.2.4,
- have a TPC secondary track of opposite charge within 3 cm of the main track,
- and the invariant mass of the main track and its TPC secondary track is less than 55 $\mathrm{MeV/c^2}$.

These conditions guarantee that the selected gamma events have shape and phase space close to the gamma background of the $\overline{\nu}_e$ selection. Furthermore, both selections are mutually exclusive since, in the $\overline{\nu}_e$ selection, we reject events that have an invariant mass less than 110 $\mathrm{MeV/c^2}$, as described in Section 8.2.10.

The gamma selection is implemented as a branch of our main Charged-Current Inclusive (CCinc) $\overline{\nu}_e$ selection. In fact, splitting the analysis into two selections at the last cut will result in a sideband that describes (more closely) the gamma background in our main selection. In

doing so, however, the gamma selection will suffer from a very low statistics. Consequently, we chose to perform the split just after our main PID, the BDT cut.

The gamma selection efficiency and purity as a function of the applied cut are shown in Fig. 13.4.

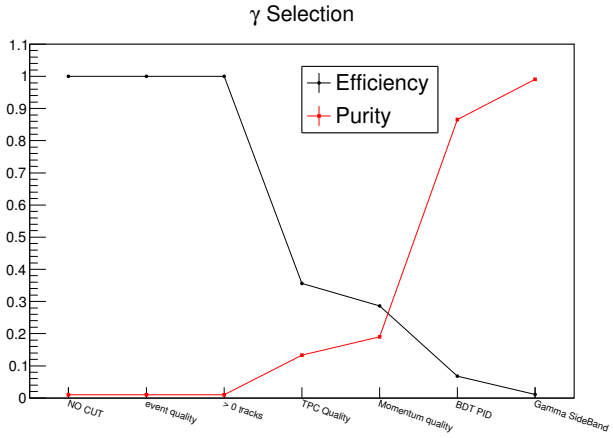


Fig. 13.4 Gamma control sample selection efficiency and purity and a function of the applied cut.

The gamma control sample has an efficiency of 1.1% and a purity of 99%. The very low efficiency of the gamma selection is not a problem for our analysis since the number of initial gamma events is much larger than our main $\overline{\nu}_e$ CCinc signal events. Consequently, the total number of gamma events passing the control sample selection is statistically significant to constrain the background of our $\overline{\nu}_e$ CCinc selection.

The interaction topology distribution of all events, data and MC, passing the gamma control sample selection is shown in Fig. 13.5. The data/MC agreement of this selection is measured to be 1.24, indicating that the MC prediction underestimates the gamma background. Hence, the use of the gamma control sample in the simultaneous unfolding framework is expected to reduce the total cross-section.

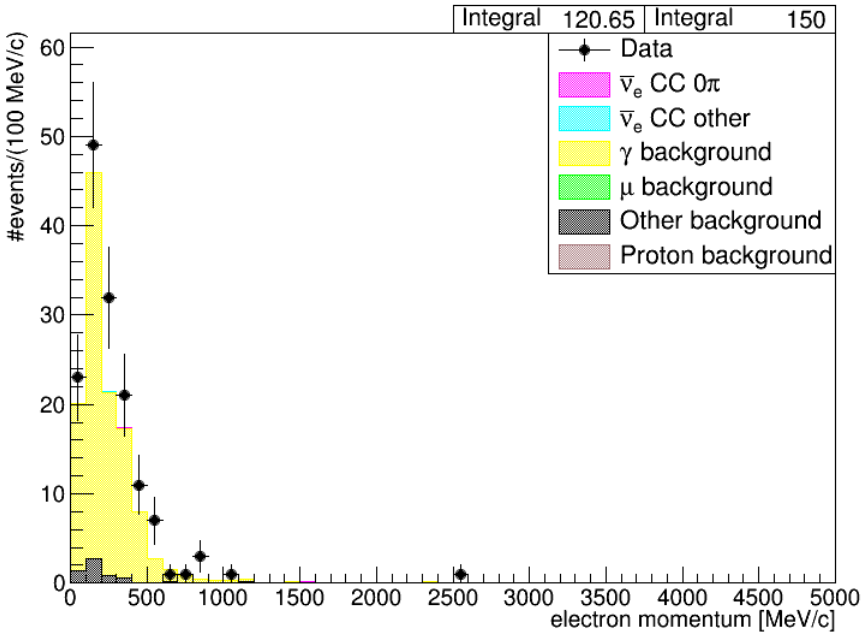


Fig. 13.5 The interaction topology distribution of all events passing the gamma control sample selection. The data points are in black circle, with statistical uncertainty, overlaying the MC prediction, and binned in the selected positron momentum.

13.3.2 Simultaneous Unfolding Using Gamma Control Sample

We unfolded both the $\overline{\nu}_e$ Charged-Current Inclusive (CCinc) differential cross-section and the gamma control sample simultaneously.

The $\overline{\nu}_e$ CCinc total cross-section is measured to be 0.279 $\pm$ 0.054 (19.4% stat) $\pm$ 0.027 (9.7% syst) $\times 10^{-38}$ cm^2/nucleon. The differential cross-section with respect to the positron momentum, constrained by the gamma control sample, is shown in Fig. 13.6. The first bin is the out-of-range bin. That it is empty is evidence that the reconstructed events fulfilled the reconstruction phase space and signal definition. The next three bins represent the $\overline{\nu}_e$ CCinc cross-section. Although, in general, the binning of the control sample could be different than that of the main signal, we chose the same binning for both.

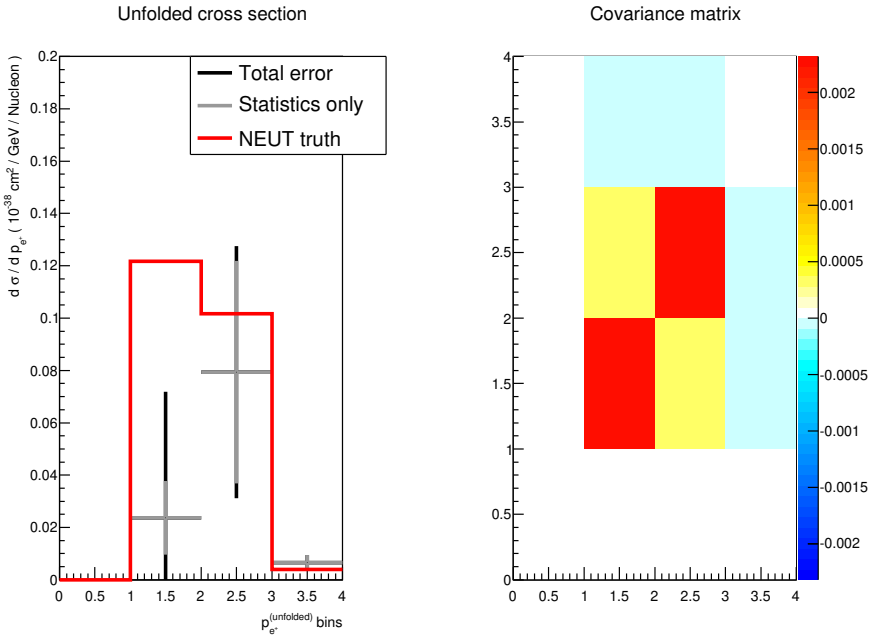


Fig. 13.6 Unfolded differential cross-section of the $\overline{\nu}_e$ with respect to the positron momentum constrained by a gamma sideband (left). The plot shows the truth (in red), unfolded cross-section value (in black point), and the gray error bars represent the statistical error, while the total error bars are shown in black. On the right, the total covariance matrix due to statistical, detector, FSI, cross-section and flux errors.

As expected, the measured total cross-section when using the gamma sideband is lower than without it. This is another indication that the modelled gamma background is highly underestimated, as previously shown in the data/MC agreement (Section 7.4) and the data/MC plots for each selection step (Section 8.2). When using a gamma sideband, the unfolded number of gamma events is more than that predicted by the MC simulation, which will lower the algorithm's expectation for the $\overline{\nu}_e$ cross-section. The unfolded cross-section results are in agreement with that of no sideband use as well as the MC predication within statistical and systematic errors. Nevertheless, until the gamma background prediction is corrected the results in the case of no sideband should be more trusted as it has better data/MC agreement. Furthermore, we could not apply all the designed vetos of the $\overline{\nu}_e$ selection in the gamma selection due to the statistical limitations described in Section 13.3.1.

13.4 Conclusion

In this analysis we process 6.11×10^{20} of data POT and 7.20×10^{21} of MC POT to extract the $\overline{\nu}_e$ charged-current inclusive cross-section. First, we select the $\overline{\nu}_e$ component of the $\overline{\nu}_\mu$ beam (expected to be $\sim 1\%$). Combined with the very small neutrino cross-section, the statistical uncertainty is the main source of error.

We developed a cut, based on a machine learning algorithm, Boosted Decision Tree (BDT), and integrated it with the main ND280 analysis software. The BDT combines information from the different sub-detectors, TPC, FGD and ECAL, and optimizes the cut in a higher dimensional feature space, benefiting from the difference between the signal and background variables correlations. The BDT tuning parameters were optimized to maximize the selection purity and efficiency as well as the KS test. We evaluated the systematics of the output test statistic by propagating the uncertainties of the BDT's input parameters. We then cut sharply on the test statistic, it being our only PID. The chosen cut value was 0.300 $\pm$ 0.014. Even with such a hard cut, the selection purity was $\sim 40\%$ at an efficiency of $\sim$ 24%. The gamma background represented 47.6% of the passed events.

In order to account for the detector inefficiencies and mis-reconstruction, we have developed an unfolding algorithm based on [1] and have integrated it with our main cross-section extraction tool. A framework to test the unfolding algorithm performance was developed. The new D'Agostini 2010 algorithm shows better coverage than the old D'Agostini 1995 algorithm. Furthermore, the added smoothing step (regularization) in the new algorithm stabilized the unfolding process and made the results less dependent on the chosen number of iterations. We chose 10 iterations for the new algorithm to avoid MC model bias, minimizing the chi-squared and maximizing the coverage.

The measured $\overline{\nu}_e$ cross-section is 0.332 $\pm$ 0.052 (15.8% stat) $\pm$ 0.025 (7.5% syst) $\times 10^{-38}$ cm^2/nucleon, which agrees within statistical and systematic errors with the MC simulation (0.290×10^{-38} cm^2/nucleon), as well as the Gargamelle measurement (0.25 $\pm$

0.07 $\times 10^{-38}$ cm^2/nucleon) [2]. Since the largest background is due to the gamma events, we constructed a gamma control sample with an efficiency of $\sim 1.1\%$ and a purity of $\sim 99\%$, and simultaneously unfolded the signal ($\overline{\nu}_e$) and background (γ) events. In this case, the $\overline{\nu}_e$ CCinc total cross-section is measured to be 0.279 $\pm$ 0.054 (19.4% stat) $\pm$ 0.027 (9.7% syst) $\times 10^{-38}$ cm^2/nucleon, which agrees within statistical and systematic errors with the MC simulation, the unfolded result without a sideband, and with the Gargamelle measurement mentioned above.

Figure 13.7 shows the measured $\overline{\nu}_e$ charged-current inclusive cross-section, unfolded with and without sideband, along with the T2K flux, NEUT MC prediction, and the Gargamelle experiment measurement. The Gargamelle data points are taken from [2].

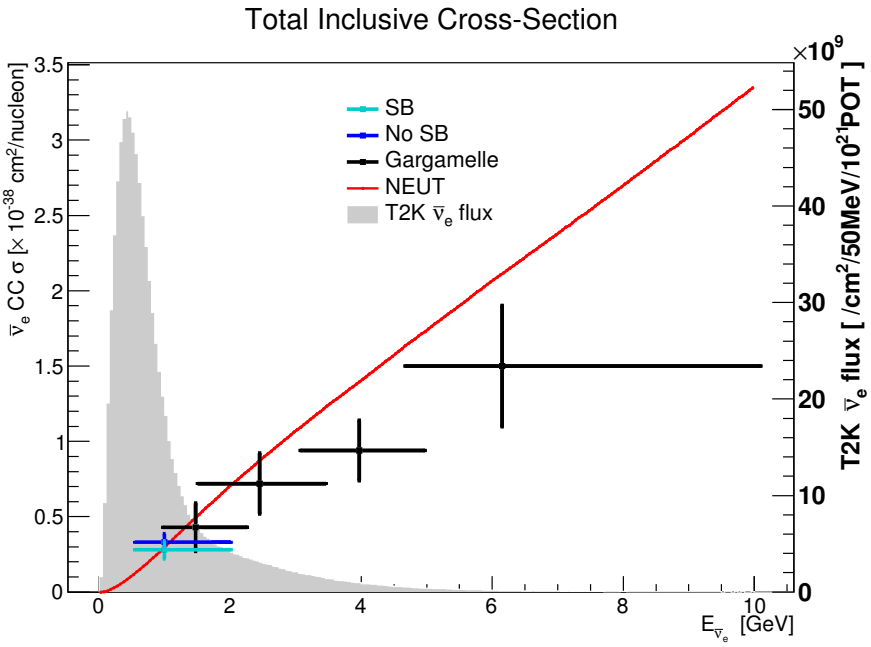


Fig. 13.7 The charged-current inclusive cross-sections for $\overline{\nu}_e$ measured by the ND280. The figure shows the T2K electron anti-neutrino flux in grey, NEUT MC prediction in red, unfolded cross-section result with sideband (SB) in cyan and without sideband (No SB) in dark blue, as well as the Gargamelle measurement in black [2].

13.5 Future Work

Although the BDT cut shows a superior performance as a particle identification method, it was not optimized to select positrons coming from a neutrino interaction. Furthermore, the currently developed vetos did not remove all of the gamma background. A BDT trained on the neutrino interaction, and taking into account different tracks, should perform better. The problem with this approach is the resultant larger bias toward the MC interaction and cross-section model. A careful systematic study using different MC generators could solve this problem.

An alternative approach, that is less model dependent, is to train other machine learning algorithms as a gamma classifier. Solving the large gamma background challenge guarantees a higher selection purity, efficiency and better statistics, that may allow finer bin choice and more precise cross-section measurement.

References

[1] G. D'Agostini. Improved iterative Bayesian unfolding. *arXiv:1010.0632*, 2010.

[2] J. Blietschau et al. Total cross sections for ν_e and $\overline{\nu}_e$ interactions and search for neutrino oscillations and decay. *Nuclear Physics B*, 133(2):205 – 219, 1978.

[3] David Griffiths. *Introduction to Elementary Particles, Second Revised Edition*. WILEY-VCH, New York, USA, 2008.

[4] Wikipedia. Standard Model — Wikipedia, The Free Encyclopedia, 2017. [Online; accessed 15-September-2017].

[5] Robert D. Klauber. *Student Friendly Quantum Field Theory: Basic Principles & Quantum Electrodynamics*. Sandtrove Press, Iowa, USA, 2013.

[6] Anthony Hey Ian Aitchison. *Gauge Theories in Particle Physics: Volume I: From Relativistic Quantum Mechanics to QED, Third Edition*. Institue of Physics, London, UK, 2003.

[7] Anthony Hey Ian Aitchison. *Gauge Theories in Particle Physics: Volume II: Non-Abelian Gauge Theories:QCD and the Electroweak Theory, Third Edition*. Institue of Physics, London, UK, 2004.

[8] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.

[9] Y. Hayato. NEUT. *Nucl. Phys. Proc. Suppl.*, 112:171–176, 2002. [,171(2002)].

[10] Tatsuya Kikawa. *Measurement of Neutrino Interactions and Three Flavor Neutrino Oscillations in the T2K Experiment*. PhD thesis, Kyoto U. (main), 2014-09.

[11] C. H. Llewellyn Smith. Neutrino Reactions at Accelerator Energies. *Phys. Rept.*, 3:261–379, 1972.

[12] T. Leitner, O. Buss, L. Alvarez-Ruso, and U. Mosel. Electron- and neutrino-nucleus scattering from the quasielastic to the resonance region. *Phys. Rev. C*, 79:034601, Mar 2009.

[13] Tina Leitner. *Neutrino-Nucleus interactions in a coupled-channel hadronic transport model*. PhD thesis, Giessen University, 2009.

[14] Veronique Bernard, Latifa Elouadrhiri, and Ulf-G. Meissner. Axial structure of the nucleon: Topical Review. *J. Phys.*, G28:R1–R35, 2002.

[15] A. A. Aguilar-Arevalo et al. Measurement of Neutrino-Induced Charged-Current Charged Pion Production Cross Sections on Mineral Oil at $E_\nu \sim 1$ GeV. *Phys. Rev.*, D83:052007, 2011.

[16] I. Ruiz Simo, J. E. Amaro, M. B. Barbaro, A. De Pace, J. A. Caballero, and T. W. Donnelly. Relativistic model of 2p-2h meson exchange currents in (anti)neutrino scattering. *J. Phys.*, G44(6):065105, 2017.

[17] J. Nieves, I. Ruiz Simo, and M. J. Vicente Vacas. Inclusive Charged–Current Neutrino–Nucleus Reactions. *Phys. Rev.*, C83:045501, 2011.

[18] D. Rein and L. M. Sehgal. Neutrino-excitation of baryon resonances and single pion production. *Annals of Physics 133*, pages 79–153, 1981.

[19] M. Rafi Alam, M. Sajjad Athar, S. Chauhan, and S. K. Singh. Weak charged and neutral current induced one pion production off the nucleon. *Int. J. Mod. Phys.*, E25(02):1650010, 2016.

[20] T. Kitagaki, H. Yuta, S. Tanaka, A. Yamaguchi, K. Abe, K. Hasegawa, K. Tamai, S. Kunori, Y. Otani, H. Hayano, H. Sagawa, K. Akatsuka, K. Furuno, N. J. Baker, A. M. Cnops, P. L. Connolly, S. A. Kahn, H. G. Kirk, M. J. Murtagh, R. B. Palmer, N. P. Samios, M. Tanaka, M. Higuchi, and M. Sato. Charged-current exclusive pion production in neutrino-deuterium interactions. *Phys. Rev. D*, 34:2554–2565, Nov 1986.

[21] G. M. Radecky et al. Study of Single Pion Production by Weak Charged Currents in Low-energy Neutrino *d* Interactions. *Phys. Rev.*, D25:1161–1173, 1982. [Erratum: Phys. Rev.D26,3297(1982)].

[22] L. Alvarez-Ruso et al. NuSTEC White Paper: Status and Challenges of Neutrino-Nucleus Scattering. 2017.

[23] C. Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator. *Nucl. Instrum. Meth.*, A614:87–104, 2010.

[24] Cezary Juszczak, Jaroslaw A. Nowak, and Jan T. Sobczyk. Simulations from a new neutrino event generator. *Nucl. Phys. Proc. Suppl.*, 159:211–216, 2006. [,211(2005)].

[25] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[26] A. Airapetian et al. Hadronization in semi-inclusive deep-inelastic scattering on nuclei. *Nucl. Phys.*, B780:1–27, 2007.

[27] J. A. Formaggio and G. P. Zeller. From eV to EeV: Neutrino Cross Sections Across Energy Scales. *Rev. Mod. Phys.*, 84:1307–1341, 2012.

[28] Tomasz Golan. What is inside mc generators...and why it is wrong. NuSTEC conference, Okayama 2015.

[29] L. L. Salcedo, E. Oset, M. J. Vicente-Vacas, and C. Garcia-Recio. Computer Simulation of Inclusive Pion Nuclear Reactions. *Nucl. Phys.*, A484:557–592, 1988.

[30] Stephen L. Adler. Tests of the conserved vector current and partially conserved axial-vector current hypotheses in high-energy neutrino reactions. *Phys. Rev.*, 135:B963–B966, Aug 1964.

[31] Dieter Rein and Lalit M. Sehgal. Coherent pi0 Production in Neutrino Reactions. *Nucl. Phys.*, B223:29–44, 1983.

[32] R. A. Smith and E. J. Moniz. Neutrino Reactions On Nuclear Targets. *Nucl. Phys.*, B43:605, 1972. [Erratum: Nucl. Phys.B101,547(1975)].

[33] Bogdan Povh, Klaus Rith, Christoph Scholz, and Frank Zetsche. *Particles and Nuclei: An Introduction to the Physical Concepts*. Springer-Verlag, Berlin, 2008.

[34] David Bohm and David Pines. A collective description of electron interactions: Iii. coulomb interactions in a degenerate electron gas. *Phys. Rev.*, 92:609–625, Nov 1953.

[35] Giampaolo Co' and Stefano De Leo. Hartree-Fock and Random Phase Approximation theories in a many-fermion solvable model. *Mod. Phys. Lett.*, A30(36):1550196, 2015.

[36] S. Bolognesi, C. Bronner, A. Cudd, M. Dunkman, P. Dunne, Y. Hayato, A. Kaboth, M. Kabirnezhad, K. Mahn, P. Martins, K. McFarland, J. Morrison, E.S. Pinzon Guerra, M. Scott, P. Stowell, C. Wilkinson, and C. Wret. NIWG model and uncertainties for 2017 oscillation analysis. *T2K Technical Notes*, 2017.

[37] Omar Benhar, Nicola Farina, Hiroki Nakamura, Makoto Sakuda, and Ryoichi Seki. Electron- and neutrino-nucleus scattering in the impulse approximation regime. *Phys. Rev.*, D72:053005, 2005.

[38] P. Adamson, C. Ader, and et. al. Andrews. First measurement of muon-neutrino disappearance in nova. *Phys. Rev. D*, 93:051104, Mar 2016.

[39] L. Fields et al. Measurement of Muon Antineutrino Quasielastic Scattering on a Hydrocarbon Target at $E_\nu \sim 3.5$ GeV. *Phys. Rev. Lett.*, 111(2):022501, 2013.

[40] A. Lovato, S. Gandolfi, J. Carlson, Steven C. Pieper, and R. Schiavilla. Electromagnetic response of ^{12}C: A first-principles calculation. *Phys. Rev. Lett.*, 117:082501, Aug 2016.

[41] A. Lovato, S. Gandolfi, J. Carlson, Steven C. Pieper, and R. Schiavilla. Electromagnetic and neutral-weak response functions of ^{4}He and ^{12}C. *Phys. Rev. C*, 91:062501, Jun 2015.

[42] Andrea Meucci, Franco Capuzzi, Carlotta Giusti, and Franco Davide Pacati. Inclusive electron scattering in a relativistic green's function approach. *Phys. Rev. C*, 67:054601, May 2003.

[43] A. Gil, J. Nieves, and E. Oset. Many body approach to the inclusive (e, e-prime) reaction from the quasielastic to the Delta excitation region. *Nucl. Phys.*, A627:543–598, 1997.

[44] V. Pandey, N. Jachowicz, T. Van Cuyck, J. Ryckebusch, and M. Martini. Low-energy excitations and quasielastic contribution to electron-nucleus and neutrino-nucleus scattering in the continuum random-phase approximation. *Phys. Rev.*, C92(2):024606, 2015.

[45] A. N. Antonov, M. V. Ivanov, M. B. Barbaro, J. A. Caballero, and E. Moya de Guerra. Longitudinal and Transverse Scaling Functions within the Coherent Density Fluctuation Model. *Phys. Rev.*, C79:044602, 2009.

[46] O. Buss, T. Gaitanos, K. Gallmeister, H. van Hees, M. Kaskulov, O. Lalakulich, A. B. Larionov, T. Leitner, J. Weil, and U. Mosel. Transport-theoretical Description of Nuclear Reactions. *Phys. Rept.*, 512:1–124, 2012.

[47] Ruth S. Van de Water. The CKM matrix and flavor physics from lattice QCD. *PoS*, LAT2009:014, 2009.

[48] R. Horsley et al. Isospin splittings of meson and baryon masses from three-flavor lattice QCD + QED. *J. Phys.*, G43(10):10LT02, 2016.

[49] Gupta Rajan, Jang Yong-Chull, Lin Huey-Wen, Yoon Boram, and Bhattacharya Tanmoy. Axial Vector Form Factors of the Nucleon from Lattice QCD. 2017.

[50] C. Alexandrou, G. Koutsou, J. W. Negele, Y. Proestos, and A. Tsapalis. Nucleon to Delta transition form factors with $N_F = 2 + 1$ domain wall fermions. *Phys. Rev.*, D83:014501, 2011.

[51] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, C. Kallidonis, G. Koutsou, A. Vaquero Avilés-Casco, and C. Wiese. The nucleon spin explained using lattice QCD simulations. 2017.

[52] Boris Kayser. Neutrino physics. *eConf*, C040802:L004, 2004.

[53] J. L. Hewett et al. Fundamental Physics at the Intensity Frontier. 2012.

[54] Janet M. Conrad, William C. Louis, and Michael H. Shaevitz. The LSND and MiniBooNE Oscillation Searches at High Δm^2. *Ann. Rev. Nucl. Part. Sci.*, 63:45–67, 2013.

[55] Boris Kayser. Neutrino mass, mixing, and flavor change. pages 1–24, 2002.

[56] t2k.org. T2k intranet, the website for members of the t2k collaboration, 2017. [Online; accessed 8-September-2017].

[57] K. Abe et al. Evidence of Electron Neutrino Appearance in a Muon Neutrino Beam. *Phys. Rev.*, D88(3):032002, 2013.

[58] T2K Collaboration, K .Abe, et al. The t2k experiment. *Nuclear Instruments and Methods in Physics Research*, 659:106–135, 2011.

[59] j-parc.jp. Japan proton accelerator research complex website., 2017. [Online; accessed 8-September-2017].

[60] T2K Collaboration, K .Abe, et al. The t2k neutrino flux prediction. *Phys. Rev. D 87, 012001*, 2013.

[61] T2K Collaboration, K .Abe, et al. First measurement of the ν_μ charged-current cross section without pions in the final state on a water target. 2017.

[62] Claude Amsler. Applications of the standard model. In *Nuclear and Particle Physics*, 2053-2563, pages 17–1 to 17–38. IOP Publishing, 2015.

[63] J Altegoer, M Anfreville, C Angelini, P Astier, M Authier, D Autiero, A Baldisseri, M Baldo-Ceolin, Giuseppe Ballocchi, M Banner, S Basa, G Bassompierre, K Benslama, I Bird, B Blumenfeld, F Bobisut, J Bouchez, S Boyd, A Bueno, and H Zaccone. The nomad experiment at the cern sps. *Nuclear Instruments and Methods in Physics Research*, pages 96–128, 01 1998.

[64] Y. Itow, T. Kajita, K. Kaneyuki, et al. The jhf-kamioka neutrino project. 2001.

[65] Byron P. Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor. Boosted decision trees, an alternative to artificial neural networks. *Nucl. Instrum. Meth.*, A543(2-3):577–584, 2005.

[66] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[67] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth.*, A389:81–86, 1997.

[68] Christian Böser, Simon Fink, Steffen Röcker. Introduction to boosted decision trees – a multivariate approach to classification problems, 2014. KSETA Doctoral Workshop, Freudenstadt, Germany.

[69] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, USA, 2006.

[70] The T2K Collaboration, Anselmo Cervera et. al. The highland analysis framework, 2017.

[71] Wikipedia contributors. Kolmogorov–smirnov test — wikipedia, the free encyclopedia, 2018. [Online; accessed 5-April-2018].

[72] Benno List. Why and when to optimize efficiency times purity, 2002.

[73] Anezka Popovski Kolaceke. Particle identification with the nd280 fine grained detectors for the t2k experiment. Master's thesis, University of Regina, Canada, 2014.

[74] Gary Barker, Steven Boyd, Stephen Dennis, Andrew Furmanski, David Hadley, Callum Lister, Phillip Litchfield, Daniel Scully, and Leigh Whitehead. Implementation of the Second Generation PID for the ND280 Tracker ECals. *T2K Technical Notes*, 2012.

[75] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning, 2^{nd} edition*. Springer Series in Statistics. Springer New York Inc., New York,USA, 2017.

[76] G. Christodoulou, B. Jamieson, S. King, P. Lasorak, and N. McCauley. Selection of ν_e and $\overline{\nu}_e$ in the ND280 tracker using anti-neutrino beam data from Run5 and Run6. *T2K Technical Notes*, 2016.

[77] Mark Hartz, Asher Kaboth, Kendall Mahn, Jordan Myslik, Christine Nielsen, Mark Scott, and Callum Wilkinson. Constraining the Flux and Cross Section Models with Data from the ND280 Detector for the 2014/15 Oscillation. *T2K Technical Notes*, 2015.

[78] G. Christodoulou, B. Jamieson, S. King, P. Lasorak, N. McCauley, and F. Shaker. Measurement of electron (anti-)neutrino cross-sections in the ND280 tracker using (anti-)neutrino beam data up to Run8. *T2K Technical Notes*, 2018.

[79] John David Jackson. *Classical electrodynamics*. Wiley, New York, NY, 3rd ed. edition, 1999.

[80] Helge Voss. What to expect from tmva the toolkit for multivariate analysis in root, 2013.

[81] Olaf Behnke, Kevin Kröninger, Thomas Schörner-Sadenius, and Gregory Schott, editors. *Data analysis in high energy physics*. Wiley-VCH, Weinheim, Germany, 2013.

[82] Volker Blobel. An Unfolding method for high-energy physics experiments. In *Advanced Statistical Techniques in Particle Physics. Proceedings, Conference, Durham, UK, March 18-22, 2002*, pages 258–267, 2002.

[83] G. Cowan. *Statistical data analysis*. Oxford, UK, 1998.

[84] Jakub Cuth, Kyrylo Merkotan, Matthias Schott, and Samuel Webb. Determination of the Transverse Momentum of W Bosons in Hadronic Collisions via Forward Folding Techniques. *Mod. Phys. Lett.*, A31(14):1650069, 2016.

[85] Martin Hierholzer. xstool – a multi-purpose cross-section tool, 2014. ND280 plenary meeting, Toyama, Japan.

[86] Tim Adye. Unfolding algorithms and tests using RooUnfold. In *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, pages 313–318, Geneva, 2011. CERN, CERN.

[87] G. D'Agostini. A Multidimensional unfolding method based on Bayes' theorem. *Nucl. Instrum. Meth.*, A362:487–498, 1995.

[88] Stefan Schmitt. Data Unfolding Methods in High Energy Physics. *EPJ Web Conf.*, 137:11008, 2017.

[89] James E. Gentle. *Solution of Linear Systems*, pages 87–121. Springer New York, New York, NY, 1998.

[90] Andreas Hocker and Vakhtang Kartvelishvili. SVD approach to data unfolding. *Nucl. Instrum. Meth.*, A372:469–481, 1996.

[91] Vladimir Rakocevic. On continuity of the moore-penrose and drazin inverses. *Matematichki Vesnik*, 49(3-4):163–172, 1997.

[92] Stefan Schmitt. TUnfold: an algorithm for correcting migration effects in high energy physics. *JINST*, 7:T10003, 2012.

[93] William Hadley Richardson. Bayesian-based iterative method of image restoration∗. *J. Opt. Soc. Am.*, 62(1):55–59, Jan 1972.

[94] L. B. Lucy. An iterative technique for the rectification of observed distributions. *Astron. J.*, 79:745–754, 1974.

[95] Mikael Kuusela and Philip B. Stark. Shape-constrained uncertainty quantification in unfolding steeply falling elementary particle spectra. 2015.

[96] Brian Gough. *GNU Scientific Library Reference Manual - Third Edition*. Network Theory Ltd., 3rd edition, 2009.

[97] Jerome H. Friedman. Data Analysis Techniques for High-Energy Particle Physics. In *1974 CERN School of Computing, Godoysund, Norway, 11-24 Aug 1974: Proceedings*, page 271, 1974.

[98] Volker Blobel. Smoothing or Fitting without a parametrization, 2005.

[99] Hiroshi Akima. A new method of interpolation and smooth curve fitting based on local procedures. *J. ACM*, 17(4):589–602, October 1970.

[100] Kyle S. Cranmer. Kernel estimation in high-energy physics. *Comput. Phys. Commun.*, 136:198–207, 2001.

[101] David W. Scott and Stephan R. Sain. Multi-dimensional density estimation, 2004.

[102] Bradley Efron and Gail Gong. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36–48, 1983.

[103] Ilya Narsky and Frank C. Porter. *Statistical analysis techniques in particle physics*. Wiley-VCH, Weinheim, Germany, 2014.

[104] Wouter Verkerke and David P. Kirkby. The RooFit toolkit for data modeling. *eConf*, C0303241:MOLT007, 2003. [,186(2003)].

[105] Megan Friend, Alexis Haesler, Mark Hartz, Atsuko K. Ichikawa, Scott Johnson, Alexander Korzenev, Katarzyna Kowalik, Andrew Missert, Takeshi Nakadaira, Boris Popov, Ken Sakashita, Kento Suzuki, Takahiro Hiraki, Magdalena Posiadala-Zezula, Davide Sgalaberna, Martin Tzanov, Mitchell Yu, and Laura Zambelli. Flux Prediction and Uncertainty Updates with NA61 2009 Thin Target Data and Negative Focussing Mode Predictions. *T2K Technical Notes*, 2016.

[106] Marek Gazdzicki and for the NA61/SHINE Collaboration. Ion program of na61/shine at the cern sps. *Journal of Physics G: Nuclear and Particle Physics*, 36(6):064039, 2009.

[107] M. G. Catanesi et al. The HARP detector at the CERN PS. *Nucl. Instrum. Meth.*, A571:527–561, 2007.

[108] A. Bercellie, Y. Hayato, K. Ieki, A. Kaboth, K. Mahn, K. McFarland, P. Rodrigues, R. Terri, M. Wascko, and C. Wilkinson. Cross section parameters for 2014 oscillation analysis. *T2K Technical Notes*, 2015.

[109] P. Bartet, S. Bordoni, R. Castillo, G. Catanesi, A. Cervera, L. Cremonesi, A. Fiorentini, L. Haegel, A. Hillairet, A. Izmaylov, J. Kim, J. Lagoda, T. Lindner, L. Magaletti, J. Myslik, C. Nielsen, S. Oser, W. Oryszczak, M. Posiadala-Zezula, Y. Petrov, M. Ravonel, F. Sanchez, E. Scantamburlo, M. Scott, W. Warzycha, J. Wilson, and J. Zalipskan. ν_μ CC event selections in the ND280 tracker using Run 2+3+4 data. *T2K Technical Notes*, 2015.

[110] D. Brailsford, A. Chappell, P. Denner, D. R. Hadley, P. Martins, G. Christodoulou, S. King, and I. Lamont. Study of the tracker ECal systematic uncertainties. *T2K Technical Notes*, 2017.

[111] Jordan Myslik. Determination of pion secondary interaction systematics for the ND280 tracker analysis. *T2K Technical Notes*, 2013.

Appendix A

Off-Axis Angle Dependence for Neutrino Energy

In this chapter, we prove Eq. (5.1) connecting the neutrino energy to the off-axis angle. Consider the following main decay producing neutrinos

$$\pi^+ \to \mu^+ + \nu_\mu. \tag{A.1}$$

Applying simple conservation of four momentum to this two-body decay gives

$$P_\pi = P_\mu + P_\nu, \tag{A.2}$$

where: P_π, P_μ and P_ν are the four-momentum for the pion, muon and neutrino respectively. Rearranging Eq. (A.2), $P_\mu = P_\pi - P_\nu$, then squaring both sides, and replacing the four-momentum squared with the invariant mass squared gives

$$m_\mu^2 = m_\pi^2 + m_\nu^2 - 2P_\pi \cdot P_\nu. \tag{A.3}$$

Neglecting neutrino mass term gives

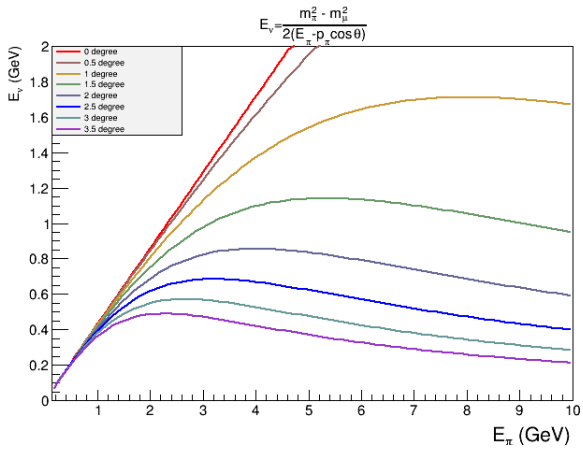
$$\begin{aligned} m_\mu^2 &= m_\pi^2 - 2(E_\pi E_\nu - \vec{p_\pi}.\vec{p_\nu}) \\ &= m_\pi^2 - 2(E_\pi E_\nu - p_\pi p_\nu \cos\theta) \\ &= m_\pi^2 - 2E_\nu(E_\pi - p_\pi \cos\theta), \end{aligned} \tag{A.4}$$

where $\vec{p_\pi}$, $\vec{p_\nu}$, p_π and p_ν are the three-momentum for the pion and neutrino along with their magnitude; θ is the angle between the outgoing neutrino and the incoming pion in the lab frame; again the neutrino mass has been neglected, and $E_\nu \approx p_\nu$.

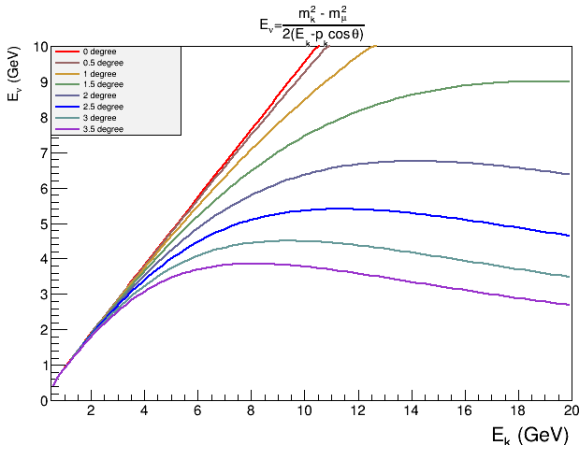
Rearranging Eq. (A.4) gives the required equation

$$E_\nu = \frac{m_\pi^2 - m_\mu^2}{2(E_\pi - p_\pi \cos\theta)}. \tag{A.5}$$

The neutrino energy as a function of the parents' (pions and kaons) energies is shown in Figs. A.1a and A.1b respectively. E_ν is almost linear with E_π (or E_k) at 0°, and does not change much with the parent's energy after certain energy threshold. Neutrino beam at 2.5° off-axis angle, the blue curve in Fig. A.1a, has an energy around 0.6 GeV for a large range of the pion energies.



(a) Neutrino Energy Vs. Pion Energy at different off-axis angles.



(b) Neutrino Energy Vs. Kaon Energy at different off-axis angles.

Fig. A.1 Neutrino energy as a function of the parent pion energy (left), and kaon energy (right), at different off-axis angles.

Appendix B

Tuning the BDT Parameters

Node Size \Depth	5	6	7	8
0.01%	0.8978	0.8942	0.9047	0.8861
0.05%	0.8770	0.9021	0.8570	0.8988
0.10%	0.8853	0.9133	0.8945	0.9008
0.20%	0.9022	0.8873	0.8355	0.8614
0.50%	0.8917	0.8982	0.8598	0.8690
0.75%	0.8866	0.8704	0.8818	0.8820
1.00%	0.9013	0.8823	0.8666	0.8866
1.25%	0.8917	0.8618	0.8720	0.8759
1.50.1%	0.8596	0.8787	0.8613	0.8879

Table B.1 Signal efficiency for different BDT configuration parameters.

Node Size \Depth	5	6	7	8
0.01%	0.5292	0.5282	0.5290	0.5417
0.05%	0.5432	0.5282	0.554	0.5353
0.10%	0.5406	0.5218	0.5331	0.5303
0.20%	0.5265	0.5358	0.5684	0.5581
0.50%	0.5326	0.5326	0.5562	0.5498
0.75%	0.5370	0.5457	0.5402	0.5400
1.00%	0.5266	0.5402	0.5498	0.5402
1.25%	0.5314	0.5534	0.5497	0.5468
1.50%	0.5500	0.5402	0.5523	0.5377

Table B.2 Signal purity for different BDT configuration parameters.

Node Size \Depth	5	6	7	8
0.01%	0.89067	0.422849	0.0269633	0.00025784
0.05%	0.905797	0.549312	0.0970727	0.000903589
0.10%	0.968815	0.690125	0.0250923	0.123026
0.20%	0.440611	0.697263	0.245907	0.127032
0.50%	0.297363	0.361634	0.19038	0.0453518
0.75%	0.997554	0.745421	0.254738	0.0197337
1.00%	0.992393	0.696003	0.535161	0.0619011
1.25%	0.99825	0.747178	0.256597	0.0150874
1.50%	0.947698	0.38493	0.23309	0.067415

Table B.3 KS p-value (signal) for different BDT configuration parameters.

Appendix C

Approximate 2D Discrete Differentiation For Non-equal Binning

C.1 Motivation

A good smoother for a one-dimensional distribution with low statistics and non-equal binning is difficult to build, as discussed in Section 10.2.6. A smoother for a two-dimensional distribution is even more challenging.

In order to unbin a two-dimensional histogram, we assume that the underlying PDF is continuous and differentiable at each data point and we rely on the Taylor expansion to approximate the PDF as a polynomial series. Then, we throw the number of events in each bin following this PDF. Subsequently, we apply a multidimensional KDE on the unbinned data to produce the smooth PDF as in the one-dimensional case.

Dealing with histograms requires a discrete version of Taylor expansion that depends on discrete differentiation. There are different methods (like forward, backward and central difference) to compute the discrete differentiation. All of these methods assume equal steps in each dimension. In the next sections, we present a method to deal with the discrete differentiation in the case of non-equal binning.

C.2 Taylor Expansion

Expanding a continuous and differentiable bi-variate function, f(x,y), around its value at the point (x_0, y_0), F(x_0, y_0), and keeping only up to the 2^{nd} order terms, gives

$$\begin{aligned} f(x,y) \approx & F(x_0,y_0) + \frac{\partial f}{\partial x}(x-x_0) + \frac{\partial f}{\partial y}(y-y_0) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x-x_0)^2 + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(y-y_0)^2 \\ & + \frac{\partial^2 f}{\partial x \partial y}(x-x_0)(y-y_0). \end{aligned} \tag{C.1}$$

The error in this case will be of order $\mathscr{O}(\Delta x^3, \Delta y^3, \Delta x^2 \Delta y, \Delta y^2 \Delta x)$, where $\Delta x = x - x_0$ and $\Delta y = y - y_0$.

C.2.1 First Order Derivatives

The first-order partial derivatives, in the discrete case, could be approximated using forward, backward or central difference as

forward difference:

$$\frac{\partial f}{\partial x} \approx \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}, \tag{C.2}$$

backward difference:

$$\frac{\partial f}{\partial x} \approx \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}, \tag{C.3}$$

and

central difference:

$$\frac{\partial f}{\partial x} \approx \frac{f(x_{i+1}) - f(x_{i-1})}{x_{i+1} - x_{i-1}}, \tag{C.4}$$

where x_i, x_{i+1} and x_{i-1} are the centres of bins i, $i+1$ and $i-1$, respectively, and $f(x_i)$, $f(x_{i+1})$ and $f(x_{i-1})$ are their corresponding histogram values. Similarly, the first order derivative is evaluated for the y variable.

In fact, it is preferred to use the central difference method, since its error is in $\mathscr{O}(\Delta x^3)$.

For non-equal binning, we refer to the more general expression

$$\begin{aligned}\frac{\partial f}{\partial x} &\approx \frac{f(x_{i+1}) - f(x_{i-1})}{x_{i+1} - x_{i-1}} \\ &= \frac{f(x_i + \frac{h_i}{2}) - f(x_i - \frac{h_i}{2})}{h_i},\end{aligned} \tag{C.5}$$

where h_i is a tiny step in one direction in bin i. For a histogram, we can take h_i to be our bin width. Consequently, we use linear interpolation with the next and earlier bins, to calculate the function values at the bin boundaries. Here, we assume that the bin values are evaluated at the bin centres. One may also reduce the number of interpolations by setting $\frac{h_i}{2}$ to be the minimum difference between the bin centres, that is

$$\frac{h_i}{2} = min(\ (x_{i+1} - x_i),\ (x_i - x_{i-1})\ ). \tag{C.6}$$

This assumption will fix one of the function values, $f(x_i + \frac{h_i}{2})$ or $f(x_i - \frac{h_i}{2})$, to be equal to the next or the previous histogram bin value.

C.2.2 Second Order Derivatives

Using the central difference method and applying the same concept twice, with a step equals to half of the bin size, we have

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2} &= \frac{\partial}{\partial x}\Big(\frac{\partial f}{\partial x}\Big) \\
&\approx \frac{\partial}{\partial x}\Big(\frac{f(x_i+\frac{h_i}{2})-f(x_i-\frac{h_i}{2})}{h_i}\Big) \\
&\approx \frac{1}{h_i}\left(\frac{\partial}{\partial x}\Big(f(x_i+\frac{h_i}{2})\Big)-\frac{\partial}{\partial x}\Big(f(x_i-\frac{h_i}{2})\Big)\right) \\
&\approx \frac{1}{h_i}\left(\frac{f(x_i+h_i)-f(x_i)}{h_i}-\frac{f(x_i)-f(x_i-h_i)}{h_i}\right) \\
&\approx \frac{f(x_i+h_i)-2f(x_i)+f(x_i-h_i)}{{h_i}^2}.
\end{aligned} \tag{C.7}$$

For general binning $f(x_i+h_i) \neq f(x_{i+1})$ and $f(x_i-h_i) \neq f(x_{i-1})$. Accordingly, we have to interpolate for $f(x_i+h_i)$ and $f(x_i-h_i)$.

In order to avoid negative values that may result from the interpolation, we first calculate the bin locations of the x_i+h_i and x_i-h_i points, then we interpolate with the values of the corresponding bins, not necessarily the adjacent ones.

The same differentiation rule is applicable for $\frac{\partial^2 f}{\partial y^2}$ just with replacing $x \to y$.

Now for $\frac{\partial^2 f}{\partial x \partial y}$, we use the same concept with h_x the bin size in the x-direction and h_y the bin size in the y-direction.

$$\begin{aligned}
\frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial}{\partial x}\Big(\frac{\partial f}{\partial y}\Big) \\
&\approx \frac{\partial}{\partial x}\Big(\frac{f(x,y+\frac{h_y}{2}) - f(x,y-\frac{h_y}{2})}{h_y}\Big) \\
&\approx \frac{1}{h_y}\Big(\frac{f(x+\frac{h_x}{2},y+\frac{h_y}{2}) - f(x-\frac{h_x}{2},y+\frac{h_y}{2})}{h_x} - \frac{f(x+\frac{h_x}{2},y-\frac{h_y}{2}) - f(x-\frac{h_x}{2},y-\frac{h_y}{2})}{h_x}\Big) \\
&\approx \frac{1}{h_x h_y}\Big(f(x+\frac{h_x}{2},y+\frac{h_y}{2}) - f(x-\frac{h_x}{2},y+\frac{h_y}{2}) - f(x+\frac{h_x}{2},y-\frac{h_y}{2}) + f(x-\frac{h_x}{2},y-\frac{h_y}{2})\Big).
\end{aligned} \tag{C.8}$$

Again, we need to interpolate the function values at the x- and y-boundaries, taking into account the bin location of the function value.