Doctoral Dissertation

博士論文

Search for CP violation through a joint oscillation analysis of atmospheric and accelerator neutrinos

(大気ニュートリノと加速器ニュートリノの
合同振動解析を通じた CP 対称性の破れの探索)

A Dissertation Submitted for the Degree of Doctor of Philosophy
December 2023

令和５年１２月博士（理学）申請

Department of Physics, Graduate School of Science,
The University of Tokyo

東京大学大学院理学系研究科物理学専攻

Aoi Eguchi

江口碧

Abstract

The neutrino oscillation is known as an important probe of physics beyond the Standard Model, which was first experimentally confirmed by the Super-Kamiokande (SK) experiment in 1998. It suggests that neutrinos have masses and provides an opportunity to investigate the properties of neutrinos and to search for the violation of CP symmetry in the lepton sector.

In this thesis, we performed a joint neutrino oscillation analysis using the SK atmospheric neutrinos and the T2K accelerator neutrinos for the first time in the history of these experiments. We established a method to analyze the atmospheric neutrinos and accelerator neutrinos in a single framework by taking into account the correlations of systematic uncertainties. Using this framework, we analyzed the SK-IV atmospheric neutrino data corresponding to 3244.4 days of data taking and the T2K Run 1-10 data corresponding to 1.97×10^{21} protons-on-target in the neutrino mode and 1.63×10^{21} protons-on-target in the antineutrino mode.

The Bayesian analysis results showed improvements in the constraints of the CP phase and neutrino mass ordering determination compared to the individual experiments. The CP-conserving values of δ_{CP} ($\delta_{\mathrm{CP}} = 0, \pi$) are excluded at 2σ under the flat prior in δ_{CP} and at 1.5σ under the flat prior in $\sin \delta_{\mathrm{CP}}$. The CP-conserving value of the Jarlskog invariant ($J_{\mathrm{CP}} = 0$) is excluded at 2σ under both priors. The Jarlskog invariant is a parameterization-independent measure of CP violation, and it is the first time in the world that we consistently exclude the CP-conserving value of J_{CP} from the 2σ credible intervals. The Bayes factor for the normal ordering over inverted ordering is computed to be 8.98 ± 0.06 under the flat δ_{CP} prior, which corresponds to the significance level of 1.64σ. It suggests a moderate preference for normal ordering, but it is not enough to claim a discovery.

This thesis also includes a variety of detailed discussions to understand the systematic uncertainties in the analysis, the contribution of each sample, and future sensitivity, which will be valuable input for future analysis.

Preface

The analyses presented in this thesis are the results of efforts by many people in the Super-Kamiokande (SK) and T2K Collaborations although the thesis itself is written by a single author. The contributions of the author are clarified in the following.

Chapters 1 to 3 contain reviews of the historical context of this analysis, theories, experimental setups, and the detectors, which is necessary for understanding the analysis carried out in this thesis but not part of the work done by the author.

Chapter 4 presents the Monte-Carlo (MC) simulations and the systematic uncertainty models that are used in the joint oscillation analysis between the SK atmospheric neutrinos and T2K accelerator neutrinos. Most of the MC and systematic uncertainty models were taken from the analyses of individual experiments with some adaptions that are necessary to take into account the correlations between the two experiments. The author contributed to updating the systematic uncertainty estimations of the detector responses. The method to take into account the correlations between the atmospheric and beam detector response systematic uncertainties was originally developed by Dr. Adrien Blanchet and was adopted by the author to reevaluate the correlations. The additional uncertainties due to the worse PID performance in the low-momentum region were introduced by the author to cover the data/MC excess in the atmospheric down-going events.

Chapter 5 presents the analysis method and sensitivity, which were originally developed and studied by the SK+T2K joint fit working group. The author introduced the multicanonical method for improving the performance of Markov-Chain Monte Carlo with the help of Dr. Lukas Berns. The sensitivity studies were updated by the author using the improved systematic uncertainty models.

Chapter 6 presents the robustness test of the model. The framework of the robustness study was developed by the T2K Collaboration, and the list of alternative models to be studied was determined by the SK+T2K joint fit working group through discussions with the neutrino interaction experts. The T2K near-detector side of the study was performed by Dr. Ciro Riccio and the SK side of the study was performed by the author.

Chapters 7 and 8 present the Bayesian analysis results and various additional discussions. A framework for calculating the posterior predictive p-values was developed by Dr. Lukas Berns and the values were reevaluated by the author. All the other studies presented in these chapters were performed by the author.

Chapter 9 presents the future sensitivities and prospects. The future sensitivity studies assuming the expected statistic increase were performed by the author, but the Hyper-Kamiokande (HK) sensitivities were made by the HK Collaboration.

Acknowledgments

I am extremely grateful to my supervisor Professor Masashi Yokoyama for guiding me in my research for five years and sharing with me his deep insight into physics, which always helped me to understand the big picture of the analysis. I also would like to express my appreciation to Kota Nakagiri and Daniel Ferlewicz for their patient review of my thesis and to the members of the Yokoyama-Nakajima Group for making a comfortable environment for pursuing research: Professor Yasuhiro Nakajima, Yoshimi Yoshimoto, Li Cheng Feng, Shoma Kodama, Wataru Okinaga, Hokuto Kobayashi, Eiichiro Watanabe, Tomochika Arai, Sanshiro Goto, Yusuke Mizuno, Tomoki Muro, and Kuniko Kono.

I would like to send my special gratitude to Christophe Bronner and Lukas Berns for guiding me in the SK and T2K joint fit group. I could not have achieved anything in this thesis without their countless and patient support. It was one of the most interesting times in my research life to discuss with both of you about physics and statistics. I appreciate Clarence Wret for sharing with me a great wealth of knowledge and insight as an expert in MCMC and cross-sections, and Ciro Riccio for supporting me in the fake data studies despite his busy time. I would like to thank Junjie Jiang, Zhenxiong Xie, and Jianrun Hu for working together with me on the almost never-ending cross-validations. The results presented in this document would have never been achieved without their hard work. I also would like to thank all the members of the joint fit and p-theta group: Professors Megan Friend, Michael Wilking, Roger Wendell, and Daniel Barrow, Adrien Blanchet, Mathieu Guigue, Kenji Yasutome, Yohei Noguchi, Lucile Mellet, and Junjie Xie.

Although the studies did not go into this thesis, I spent half of my Ph.D. period on the studies of the T2K near detector upgrade analysis, which helped me to gain a lot of knowledge about the experiments and to train myself as a researcher. I would like to express my special thanks to my colleagues at ETH Zurich: Professors André Rubbia and Davide Sgalaberna, and Saúl Alonso Monsalve, Xingyu Zhao, Botao Li, and Till Dieminger. It was my pleasure to spend time with you in such a wonderful environment and to have many insightful discussions. I also would like to thank all the members of the SuperFGD simulation and reconstruction group: Professor Clark McGrew, Lorenzo Magaletti, Mahesh Jakkapu, Katharina Lachner, and Weijun Li.

I would like to express my thanks to the members of the SK and T2K collaborations: Professors Atsuko Ichikawa, Masaki Ishitsuka, Kendall Mahn, Akihiro Minamino, Ken Sakashita, Yoshinari Hayato, and Patrick Duune, Claudio Giganti, Tatsuya Kikawa, Tsunayuki Matsubara, Yohei Noguchi, Hikaru Tanigawa, Takuji Arihara, Takashi Honjo, Seidai Tairafune, Masaki Kawaue, and Sakiko Nishimori.

Finally, no amount of text can describe how grateful I am to my family, Shunji Eguchi, Toshiko Eguchi, and Shizuka Eguchi for supporting me throughout entire my life.

Contents

List of abbreviations

Category	Abbreviation	Description
Experiments	T2K	Tokai to Kamioka
	SK	Super Kamiokande
Neutrino physics	PMNS	Pontecorvo-Maki-Nakagawa-Sakata
	NO	Normal Ordering
	IO	Inverted Ordering
	LO	Lower Octant
	UO	Upper Octant
Super-Kamiokande	ID	Inner Detector
	OD	Outer Detector
	FC	Fully Contained
	PC	Partially Contained
	UpMu	Upward-going Muon
T2K neutrino beam	J-PARC	Japan Proton Accelerator Research Complex
	LINAC	Linear accelerator
	RCS	Rapid-Cycling Synchrotron
	MR	Main Ring synchrotron
	FHC	Forward Horn Current (neutrino mode)
	RHC	Reversed Horn Current (antineutrino mode)
	POT	Protons On Target
T2K near detector	ND280	T2K Near Detector complex
	ECal	Electromagnetic Calorimeter
	FGD	Fine-Grained Detector
	TPC	Time Projection Chamber
	P0D	π^0 Detector
	SMRD	Side Muon Range Detector
Analysis method	MCMC	Markov Chain Monte Carlo
	fiTQun	SK reconstruction fitter name
	NEUT	Neutrino intereaction event generator name
	Xsec	Cross-section
	PID	Particle Identification
Neutrino interaction	CC	Charged Current
	NC	Neutral Current
	CCQE	Charged-Current Quasi-Elastic
	2p2h	Two particle two hole
	DIS	Deep Inelastic Scattering
	FSI	Final State Interactions
	RFG	Relativistic Fermi gas
	LFG	Local Fermi gas
	SF	Spectral Function
	CRPA	Continuum Random Phase Approximation

Chapter 1

Introduction

The elementary particles of nature have been the target of interest in a long history of humans studying the world around us. In the 20th century, the Standard Model (SM) of particle physics was rapidly developed and was very successful in describing most of the phenomena observed in particle physics experiments within a single framework. However, there are some phenomena that cannot be explained by the SM. For example, the SM does not include a theory to describe gravity, which is one of the fundamental forces. In addition, the SM cannot explain dark matter and dark energy, and a mechanism of the baryon asymmetry of the Universe either.

Neutrinos are the second most abundant particles in the Universe next to photons and are assumed to be massless and chargeless in the SM. However, in 1998, a phenomenon called *neutrino oscillation*, in which neutrinos change their flavors while propagating, was first experimentally confirmed by the Super-Kamiokande experiment [1]. The existence of neutrino oscillation suggests that neutrinos have masses, and is inconsistent with the assumptions in the SM. It therefore started a new era of searching for new physics beyond the Standard Model in the neutrino sector.

Although the existence of neutrino oscillation itself has already been proved by many experiments, it still provides us with various opportunities to search for new physics. One of the open questions we try to investigate in this thesis is whether CP symmetry is violated in the lepton sector, where C and P denote the charge conjugate and parity conjugate, respectively. Although CP symmetry is known to be violated in the quark sector [2, 3], it has never been experimentally proved in the lepton sector. Since CP symmetry is one of the most fundamental symmetries, it is of great interest whether it is violated in the lepton sector as well. If CP violation in the lepton sector is proved, it also has the possibility of providing a key to understanding the origin of baryon asymmetry in the Universe through scenarios such as leptogenesis [4].

In addition to the search for CP violation, there are several open questions in neutrino oscillation physics. Although it was proved that neutrinos have non-zero masses, the ordering of the three neutrino mass eigenstates ($m_1 < m_2 < m_3$ or $m_3 < m_1 < m_2$) is still unknown, and this is called the neutrino mass ordering problem. The magnitude of the neutrino flavor mixing angles, which define the conversion from the neutrino mass eigenstates to the neutrino flavor eigenstates, is also of interest.

The Super-Kamiokande (SK) and the Tokai-to-Kamioka (T2K) experiments are neutrino oscillation experiments conducted in Japan and have been playing leading roles in

the field of neutrino oscillation physics. These two experiments share the same water Cherenkov detector, Super-Kamiokande, located in the Gifu prefecture. However, they are independent collaborations and have been studying neutrinos from different sources independently. SK started its operation in 1996 and has been observing neutrinos produced from natural sources such as the atmosphere around the Earth and the Sun. As mentioned earlier, it is known as the experiment that confirmed neutrino oscillations for the first time in the world. On the other hand, T2K started data taking in 2010 and has been measuring the oscillations of muon neutrinos and antineutrinos produced at the Japan Proton Accelerator Research Complex (J-PARC) in the Ibaraki prefecture. T2K discovered the oscillation from muon neutrinos to electron neutrinos in 2013 [5], and has the world's best sensitivity to CP violation in the lepton sector [6].

In this thesis, we present a joint analysis between SK and T2K for the first time in the history of these experiments. We establish a method to analyze the atmospheric neutrinos and accelerator neutrinos in a single framework by correctly taking into account the correlations of systematic uncertainties. Since these two neutrino oscillation measurements are complementary to each other, the joint analysis of the two experiments is expected to further improve the sensitivities to the open questions in neutrino oscillation physics including CP violation, mass ordering, and the magnitude of neutrino mixing. Through the joint analysis, we try to address these questions in this thesis.

This thesis is organized as follows. Chapter 2 describes the motivation of the joint analysis, as well as the history and theoretical backgrounds of neutrino oscillation physics. The electroweak model in the Standard Model is first reviewed, and then a minimal extension is introduced to include the massive neutrinos and neutrino oscillations in the model.

The experimental setup including the neutrino sources, detectors, and samples used in the analysis are described in Chapter 3. The Monte-Carlo simulation and systematic uncertainties used in the analysis are described in Chapter 4.

In Chapter 5, the analysis framework of joint analysis is introduced including the statistical analysis methods and the fitting methods. The expected sensitivity of the joint analysis is also studied using the simulated Monte-Carlo data set.

Chapter 6 describes the robustness test of our systematic uncertainty model. Here we test whether there could be a bias in the oscillation parameter measurements when our systematic uncertainty model is incorrect.

In Chapter 7, the results of the actual data analysis are presented. We perform a Bayesian analysis based on the Markov-Chain Monte Carlo (MCMC) method and construct credible intervals for the parameters of interest. The Bayes factors for the mass ordering hypothesis and goodness of the fit are also tested. In Chapter 8, we discuss the data fit results in more detail through studies of the contribution of each sample, the systematic uncertainty constraints, and the CP-conserving hypothesis test. The future sensitivities and prospects will also be discussed in Chapter 9. Finally, the conclusions of the analyses performed in this thesis are described in Chapter 10.

Chapter 2

Physics of neutrinos

2.1 History of neutrino oscillations

2.1.1 Discovery of neutrinos

In 1914, Chadwick measured the energy spectrum of electrons emitted from the β decay [7]. If one assumes that there are only electrons emitted from the beta decay and that the recoil energy of the nucleus can be neglected, the electron should have energy with a single peak near the mass difference of neutron and proton. However, Chadwick discovered that the emitted electron has a continuous energy spectrum. A solution to this problem was proposed by Pauli in 1930 in his famous letter; a new spin-$\frac{1}{2}$ particle is produced together with the electron, which is nowadays known as a neutrino [8]. The actual β decay can be described as the three-body decay process

$$n \rightarrow p + e^- + \overline{\nu}_e, \tag{2.1}$$

where neutrinos carry away some energy and escape from the detectors.

After it was first predicted by Pauli, it took more than 20 years to directly observe neutrinos. Reines and Cowan conducted experiments for detecting the neutrinos produced at nuclear reactors using a water tank with dissolved $CdCl_2$ [9, 10]. The experiments took place in 1953 at the Hanford reactor and were later repeated in 1956 at the Savannah River reactor. In these experiments, they confirmed the existence of neutrinos through the detection of positrons and neutrons produced in the scattering process of electron antineutrinos and protons

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

which is known as the inverse beta decay. The inverse beta decay process is being used as an important process to detect neutrinos in many of the current neutrino experiments.

In 1962, groups from Columbia University and Brookhaven National Laboratory observed muon neutrinos produced in the pion decay [11]

$$\pi^\pm \rightarrow \mu^\pm + (\nu/\overline{\nu}). \tag{2.3}$$

They confirmed that only muons are produced from the interaction of neutrinos produced in the pion decay, which implies that the neutrinos observed here (ν_μ) are different from

the ones observed in the previous experiments via the inverse beta decay (ν_e). From these results, they confirmed that there are at least two types of neutrinos.

After the discovery of the τ lepton in 1975 [12], the existence of tau neutrino ν_τ was postulated. In 2001, the DONUT experiment detected four ν_τ events with an estimated background of 0.34±0.05, using an emulsion target [13]. The tau neutrinos were produced in the decay of D_S mesons into τ and ν_τ and the subsequent decay of τ into ν_τ using 800 GeV protons from the Fermilab Tevatron. With these results, all three flavors of neutrinos (ν_e, ν_μ, ν_τ) have been experimentally observed[1].

2.1.2 Discovery of neutrino oscillation

Around 1968, Davis started searching for electron neutrinos produced in the thermonuclear reactions in the Sun (known as *solar neutrinos*). They used a detector based on the reaction

$$^{37}\text{Cl} + \nu_e \rightarrow e^- + {}^{37}\text{Ar}, \tag{2.4}$$

which was suggested by Pontecorvo and known as the chlorine-argon reaction [15, 16]. The experiment led by Brookhaven National Laboratory used a 100,000-gallon chlorine-argon neutrino detector in the Homestake Gold Mine, in Lead, South Dakota. The product of the neutrino flux and cross-sections was measured to be

$$\sigma(^{37}\text{Cl} + \nu_e \rightarrow e^- + {}^{37}\text{Ar}) \leq 0.3 \times 10^{-35} \text{ sec}^{-1} \text{ per atom}, \tag{2.5}$$

which is smaller by approximately a factor of 7 than the predicted value of $(2.0 \pm 1.2) \times 10^{-35}$ sec^{-1} per atom obtained from the Standard Solar Model[2] [18]. This discrepancy between the measurement and prediction was called the solar neutrino problem and triggered discussions among physicists.

To explain the deficit in the observed number of neutrino events in the solar neutrino experiment, Gribov and Pontecorvo (1968) suggested neutrino oscillation in which neutrinos change their types while traveling long distance [19]. Later, Wolfenstein (1978) suggested that we should take into account the effect of the coherent scattering process when considering the neutrino oscillation, which is now known as the *matter effect* [20]. This effect was further developed by Mikheyev and Smirnov (1985) [21] and called the Mikheyev-Smirnov-Wolfenstein (MSW) effect.

The existence of neutrino oscillation was first experimentally proved by the Super-Kamiokande experiment in 1998, through the measurement of neutrinos produced in the atmosphere (known as *atmospheric neutrinos*) [1]. Although atmospheric neutrinos are produced uniformly around the Earth, it was found that the number of ν_μ that come from the opposite side of the Earth is fewer than the ones that come from the atmosphere above the detector. This is due to the oscillation of ν_μ into ν_τ and the observed event rates were consistent with the prediction assuming the two flavor $\nu_\mu \leftrightarrow \nu_\tau$ oscillation.

---

[1]The number of neutrinos was determined to be 2.9840 ± 0.0082 by the LEP experiment through the measurement of the total decay width of the Z resonance [14], which agrees well with the observed generations of neutrinos.

[2]The latest measurement and prediction are 2.56±0.16±0.16 SNU and $8.46^{+0.87}_{-0.88}$ SNU, respectively [17]. Here SNU denotes the solar neutrino unit and is defined as 1 SNU = 10^{-37} captures/cm^2/s.

In 2002, Sudbury Neutrino Observatory measured solar neutrinos via the charged current (CC) and neutral current (NC) reaction on deuterium (d) and elastic scattering (ES) on electrons [22, 23]

$$\nu_e + d \rightarrow p + p + e^- \quad \text{(CC)}, \tag{2.6}$$

$$\nu_x + d \rightarrow p + n + \nu_x \quad \text{(NC)}, \tag{2.7}$$

$$\nu_x + e^- \rightarrow e^- + \nu_x \quad \text{(ES)}. \tag{2.8}$$

The CC reaction is only caused by ν_e while the NC and ES reactions can be caused by ν_μ and ν_τ as well, which allowed them to study the deficit in ν_e event rates while ensuring the total event rates of all the flavors of neutrinos are consistent with the prediction. The observed neutrino flux showed clear proof of the appearance of ν_μ and ν_τ in the solar neutrinos and an agreement of total neutrino flux with the prediction based on the Standard Solar Model. In 2002, the KamLAND experiment also confirmed the deficit of $\bar{\nu}_e$ events due to the neutrino oscillation using the neutrinos produced at the reactors (known as *reactor neutrinos*) at the 99.95% confidence level [24].

Furthermore, the Tokai-to-Kamioka (T2K) experiment reported the evidence of $\nu_\mu \rightarrow \nu_e$ oscillation with a significance of 7.3σ by detecting 28 ν_e appearance signal events in the ν_μ-dominant beam in 2014 [5]. In 2013, SK detected atmospheric ν_τ with a confidence level of 3.8σ in the atmospheric neutrinos [25]. The OPERA experiment also detected 5 ν_τ appearance events in 2015 and confirmed $\nu_\mu \rightarrow \nu_\tau$ oscillation at a 5.1σ confidence level [26]. With these results, all the oscillation (and non-oscillation) channels of $\nu_\mu \rightarrow \nu_\alpha$ for $\alpha = e, \mu, \tau$ were confirmed to exist.

2.2 Physics of neutrinos

2.2.1 Neutrinos in the Standard Model

The Standard Model (SM) of particle physics has been developed to describe the properties of elementary particles and the strong, weak, and electromagnetic interactions. These three interactions can be described based on the gauge symmetry under $SU(3)_C \times SU(2)_L \times U(1)_Y$ where C stands for color, L for left-handedness, and Y for hypercharge.

The strong interaction is described by the $SU(3)_C$ gauge theory, which is called the Quantum chromodynamics (QCD). The theory for dealing with the electromagnetic and weak interaction in a single framework has been developed by Glashow, Salam, and Weinberg in the 1960s [27, 28, 29]. This is called the electroweak model (GSW model or WS model) and it is associated with the $SU(2)_L \times U(1)_Y$ gauge symmetry.

The elementary particles described by the SM are classified into three groups: fermions, gauge bosons, and the Higgs boson. The fermions have a half spin in the unit of Planck's constant $\hbar$ and compose the matter around us. Among the fermions, particles that participate in QCD are called quarks, and the others that are involved in the electroweak interaction but not in the strong interaction are called leptons. The gauge bosons are introduced as spin-1 force-carrying particles in each gauge symmetry. Finally, the Higgs boson, a spin-0 scalar particle, provides a mechanism for the gauge bosons and fermions to obtain their mass via spontaneous symmetry breaking. In the following, the electroweak model in the lepton sector is briefly reviewed.

In the electroweak model, left-handed neutrinos and their counterpart left-handed charged leptons form doublets, while right-handed charged leptons form singlets under the $SU(2)_L \times U(1)_Y$ gauge symmetry:

$$\begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \ \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \ \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix}, \ e_R, \ \mu_R, \ \tau_R. \tag{2.9}$$

The right-handed neutrinos have never been experimentally observed, and only the left-handed neutrinos are included in the SM[3].

Let us denote the left-handed lepton doubet as $L_\ell = \begin{pmatrix} \nu_{\ell L} & \ell_L \end{pmatrix}^T$ for $\ell = e, \mu, \tau$ and the electroweak Lagrangian density can be written as

$$\mathcal{L}_\ell = \overline{L}_\ell D_\mu \gamma^\mu L_\ell + \overline{\ell}_R i \gamma^\mu \partial_\mu \ell_R. \tag{2.10}$$

The covariant derivative is defined as

$$D_\mu := \partial_\mu - i \left(\sum_{a=1}^{3} g W_\mu^a \frac{\sigma^a}{2} - \frac{1}{2} g' B_\mu \right), \tag{2.11}$$

where W_μ^a and B_μ are the gauge bosons of $SU(2)_L$ and $U(1)_Y$, respectively. $\sigma^a (a = 1, 2, 3)$ are Pauli matrices which are the three generators of the $SU(2)$ group, and g, g' are the coupling constants. To better describe the observed interactions, these gauge bosons can be rewritten as

$$W_\mu^\pm = \frac{W_\mu^1 \mp i W_\mu^2}{2\sqrt{2}}, \tag{2.12}$$

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}, \tag{2.13}$$

where θ_W is the Weinberg angle (weak mixing angle) which defines the mixture between $SU(2)_L$ and $U(1)_Y$ and is measured to be [17]:

$$\sin^2 \theta_W \simeq 0.23. \tag{2.14}$$

Using these notations, the Lagrangian density can also be rewritten as

$$\mathcal{L}_\ell = \overline{\nu}_{\ell L} i \slashed{\partial} \nu_{\ell L} + \overline{\ell} i \slashed{\partial} \ell + W_\mu^+ J_+^\mu + W_\mu^- J_-^\mu + Z_\mu J_Z^\mu + A_\mu J_{\text{EM}}^\mu, \tag{2.15}$$

where the currents are given as

$$J_+^\mu = \frac{g}{\sqrt{2}} \overline{\nu}_{\ell L} \gamma^\mu \ell_L, \tag{2.16}$$

$$J_-^\mu = (J_+^\mu)^\dagger = \frac{g}{\sqrt{2}} \overline{\ell}_L \gamma^\mu \nu_{\ell L}, \tag{2.17}$$

$$J_Z^\mu = \frac{g}{\cos\theta_W} \left[\frac{1}{2} \overline{\nu}_{\ell L} \gamma^\mu \nu_{\ell L} - \frac{1}{2} \overline{\ell}_L \gamma^\mu \ell_L + \sin^2\theta_W \overline{\ell} \gamma^\mu \ell \right], \tag{2.18}$$

$$J_{\text{EM}}^\mu = -\overline{\ell} \gamma^\mu \ell. \tag{2.19}$$

---

[3]The existence of right-handed neutrinos is suggested from the observation of neutrino oscillations as discussed in Section 2.2.2.

In conclusion, the weak interaction is mediated by either charged $W^\pm$ bosons or charge-less Z bosons, and the electromagnetic interaction is mediated by the photon field A. The weak interaction mediated by $W^\pm$ is called the "charged current" (CC) interaction as it does not conserve the charge of incoming and outgoing leptons in the process. When the particles exchange Z, the interaction is called the "neutral current" (NC) interaction.

2.2.2 Neutrino mass

The masses of the charged leptons are generated via a Yukawa coupling of the right-handed charged lepton ℓ_R and left-handed lepton doublet L_L with the Higgs doublet ϕ

$$\mathcal{L}_{\text{Yukawa},\ell} = -\sum_{\alpha,\beta}^{e,\mu,\tau} Y_{\alpha\beta}^\ell \overline{L}_{\alpha L} \phi \ell_{\beta R} + \text{h.c.}, \tag{2.20}$$

where $Y_{\alpha\beta}^\ell$ is the Yukawa-coupling constant and h.c. is an abbreviation for the Hermitian conjugate. Since the SM does not contain right-handed neutrinos, the Yukawa coupling cannot be built for neutrinos, and they are assumed to be massless. Nevertheless, it was found that neutrinos actually have non-zero masses through the observation of neutrino oscillation [1].

One of the simplest extensions of the SM to include non-zero neutrino masses is to introduce right-handed neutrinos and add a neutrino mass term. The right-handed neutrinos are introduced as the gauge singlets (ν_{eR}, $\nu_{\mu R}$, and $\nu_{\tau R}$) and do not participate in any of the SM interactions[4].

Let us consider the Lagrangian density of Yukawa couplings between Higgs and the neutrinos:

$$\mathcal{L}_{\text{Yukawa},\nu} = -\sum_{\alpha,\beta}^{e,\mu,\tau} Y_{\alpha\beta}^\nu \overline{L}_{\alpha L} \tilde{\phi} \nu_{\beta R} + \text{h.c.}. \tag{2.21}$$

After spontaneous symmetry breaking, the Higgs doublet in the unitary gauge is

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+H \end{pmatrix}, \quad \tilde{\phi} \equiv i\sigma^2 \phi^* = \frac{1}{\sqrt{2}} \begin{pmatrix} v+H \\ 0 \end{pmatrix} \tag{2.22}$$

and the Yukawa Lagrangian can be written in the matrix format:

$$\mathcal{L}_{\text{Yukawa},\nu} = -\frac{v+H}{\sqrt{2}} \overline{\nu}_L Y^\nu \nu_R + \text{h.c.}, \tag{2.23}$$

where

$$\nu_L = \begin{pmatrix} \nu_{eL} \\ \nu_{\mu L} \\ \nu_{\tau L} \end{pmatrix} \quad \text{and} \quad \nu_R = \begin{pmatrix} \nu_{eR} \\ \nu_{\mu R} \\ \nu_{\tau R} \end{pmatrix} \tag{2.24}$$

---

[4]They are called *sterile* neutrinos as they do not even participate in the weak interaction. Theoretically, any number of right-handed neutrinos can be introduced in the model [17], but here the number is fixed to be three for simplicity.

are the chiral spinors of neutrino's weak eigenstates. The matrix Y^ν of neutrino Yukawa couplings can be diagonalized with two unitary matrices V_L^ν and V_R^ν as

$$V_L^{\nu\dagger} Y^\nu V_R^\nu = Y'^\nu \quad \text{where} \quad Y'^\nu_{ij} = y_i^\nu \delta_{ij} \quad (i, j = 1, 2, 3), \tag{2.25}$$

and the chiral spinors of neutrino mass eigenstates can be denoted as

$$\nu_L^{(m)} = V_L^{\nu\dagger} \nu_L = \begin{pmatrix} \nu_{1L}^{(m)} \\ \nu_{2L}^{(m)} \\ \nu_{3L}^{(m)} \end{pmatrix}, \quad \nu_R^{(m)} = V_R^{\nu\dagger} \nu_R = \begin{pmatrix} \nu_{1R}^{(m)} \\ \nu_{2R}^{(m)} \\ \nu_{3R}^{(m)} \end{pmatrix}. \tag{2.26}$$

Finally, denoting $\nu_k = \nu_{kL} + \nu_{kR}$, we obtain

$$\mathcal{L}_{\text{Yukawa},\nu} = -\frac{v + H}{\sqrt{2}} \sum_{k=1}^{3} y_k^\nu \overline{\nu}_{kL}^{(m)} \nu_{kR}^{(m)} + \text{h.c.} \tag{2.27}$$

$$= -\sum_{k=1}^{3} \frac{y_k^\nu v}{\sqrt{2}} \overline{\nu}_k^{(m)} \nu_k^{(m)} - \sum_{k=1}^{3} \frac{y_k^\nu}{\sqrt{2}} \overline{\nu}_k^{(m)} \nu_k^{(m)} H, \tag{2.28}$$

which gives the neutrino masses of

$$m_k = \frac{y_k^\nu v}{\sqrt{2}} \quad (k = 1, 2, 3). \tag{2.29}$$

Therefore, the neutrino masses are proportional to the Higgs vacuum expectation value v and the eigenvalues of the Yukawa coupling matrix y_k^ν.

Although the neutrino masses have not been directly measured, the current best upper limit of the neutrino mass is set to be $m_\nu < 0.8 \text{ eV}/c^2$ (90% CL) by the Karlsruhe Tritium Neutrino (KATRIN) experiment [30]. The fact that neutrinos have significantly small mass implies that y_k^ν is very small compared to the other fermions, but there is no clear explanation of why.

It is also possible to introduce Majorana neutrinos which satisfy the Majorana condition

$$\psi = \psi^c := C\overline{\psi}^T, \tag{2.30}$$

where C is the charge conjugate. With the Majorana neutrinos, we can form the Majorana mass term

$$\mathcal{L}_{\text{mass}}^M = -\frac{1}{2} m \left(\overline{\nu}_L^c \nu_L + \overline{\nu}_L \nu_L^c \right), \tag{2.31}$$

and the smallness of the neutrino masses can be explained naturally by the see-saw mechanism [31, 32, 33]. However, the neutrino oscillation experiments do not have sensitivities to the neutrino types (Dirac or Majorana) as the Majorana phases do not participate in the oscillation probabilities [34]. Therefore, we will treat neutrinos as Dirac particles and ignore the Majorana phases in the following sections.

2.2.3 Neutrino mixing

Using the formalization in Eq. (2.26) (and its equivalent for the charged leptons), the charged current of leptons can be written as

$$J_+^\mu = \frac{g}{\sqrt{2}} \bar{\nu}_L \gamma^\mu \ell_L = \frac{g}{\sqrt{2}} \bar{\nu}_L^{(m)} V_L^{\nu\dagger} \gamma^\mu V_L^\ell \ell_L^{(m)}. \tag{2.32}$$

Denoting $U = V_L^{\ell\dagger} V_L^\nu$ and redefining the neutrino flavor eigenstates as $\nu_L^{(f)} := U\nu_L^{(m)} = V_L^{\ell\dagger} \nu_L$, the current becomes

$$J_+^\mu = \frac{g}{\sqrt{2}} \bar{\nu}_L^{(m)} U^\dagger \gamma^\mu \ell_L^{(m)} = \frac{g}{\sqrt{2}} \bar{\nu}_L^{(f)} \gamma^\mu \ell_L^{(m)}, \tag{2.33}$$

which is the same format as the standard notation of the leptonic charged current in the SM. The unitary matrix $U = V_L^{\ell\dagger} V_L^\nu$ is known as the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [35]:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = U \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}, \tag{2.34}$$

which defines the conversion between the flavor eigenstates and mass eigenstates[5].

Let us think about the number of free parameters in the PMNS matrix. In general, an $n \times n$ unitary matrix has $2n^2 - n^2 = n^2$ independent real parameters where n^2 parameters are constrained by the unitary condition $UU^\dagger = I$. These parameters can be divided into $n(n-1)/2$ mixing angles and $n(n+1)/2$ phase parameters[6]. In addition, $2n - 1$ phases can be absorbed by rephasing the lepton fields and they do not appear as physical observables. It therefore leaves $n(n+1)/2 - (2n-1) = (n-1)(n-2)/2$ physical phases that cannot be absorbed. Since the PMNS matrix has a dimension of 3×3, it has three mixing angles and one physical phase.

Denoting these angle and phase parameters as $\theta_{12}, \theta_{13}, \theta_{23}$, and δ_{CP}, it is convenient to rewrite the PMNS matrix as

$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta_{\mathrm{CP}}} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta_{\mathrm{CP}}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.35}$$

$$= \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{\mathrm{CP}}} \\ -s_{12}c_{23} - c_{12}s_{13}s_{23}e^{i\delta_{\mathrm{CP}}} & c_{12}c_{23} - s_{12}s_{13}s_{23}e^{i\delta_{\mathrm{CP}}} & c_{13}s_{23} \\ s_{12}s_{23} - c_{12}s_{13}c_{23}e^{i\delta_{\mathrm{CP}}} & -c_{12}s_{23} - s_{12}s_{13}c_{23}e^{i\delta_{\mathrm{CP}}} & c_{13}c_{23} \end{pmatrix}, \tag{2.36}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$. The three mixing angles can be defined in the first quadrant $\theta_{ij} \in [0, \pi/2]$ without losing generality. δ_{CP} is known as the CP-violating phase and it takes a value in the range $\delta_{\mathrm{CP}} \in [-\pi, \pi]$.

---

[5]The subscripts for the left-handedness L, flavor state (f), and mass state (m) are removed for simplicity.

[6]The number of angles in the $n \times n$ unitary matrix is equal to the number of parameters in a real $n \times n$ orthogonal matrix.

In the SM, only the charged current interaction $\mathcal{L}_{\text{CC}}$ has the possibility of violating CP conservation. It can be confirmed by applying the CP conversion to $\mathcal{L}_{\text{CC}}$ as

$$\mathcal{L}_{\text{CC}} = -\frac{g}{\sqrt{2}} \left[\overline{\nu}_L U^\dagger \gamma^\mu \ell_L W_\mu^+ + \overline{\ell}_L U \gamma^\mu \nu_L W_\mu^- \right] \tag{2.37}$$

$$\xrightarrow{\text{CP}} \mathcal{L}'_{\text{CC}} = -\frac{g}{\sqrt{2}} \left[\overline{\nu}_L U^T \gamma^\mu \ell_L W_\mu^+ + \overline{\ell}_L U^* \gamma^\mu \nu_L W_\mu^- \right], \tag{2.38}$$

where we use the following formula for the conversion

$$\overline{\psi}_1 \gamma^\mu (1 - \gamma^5) \psi_2 \xrightarrow{\text{CP}} -\overline{\psi}_2 \gamma_\mu (1 - \gamma^5) \psi_1, \tag{2.39}$$

$$W_\mu^\pm \xrightarrow{\text{CP}} -W^{\mp\mu}. \tag{2.40}$$

Comparison of Eq. (2.37) and Eq. (2.38) yields a fact that it can break CP conservation when U has a physical complex phase and satisfies $U^* \neq U$. In other words, only the δ_{CP} phase can bring CP violation in the lepton sector.

2.2.4   Neutrino oscillation in vacuum

Neutrino oscillation is a quantum phenomenon in which neutrinos change their flavors. In this subsection, we will derive the neutrino oscillation probabilities using the PMNS matrix.

Let us assume that a neutrino with flavor α is created in a charged-current weak interaction process. The flavor eigenstate can be expressed as the superposition of mass eigenstates using the PMNS matrix:

$$|\nu_\alpha\rangle = \sum_{i=1}^{3} U^*_{\alpha i} |\nu_i\rangle \quad (\alpha = e, \mu, \tau). \tag{2.41}$$

The mass eigenstates of neutrinos evolve in time as plane waves

$$|\nu_i(t)\rangle = e^{-iE_i t} |\nu_i\rangle. \tag{2.42}$$

Therefore, after traveling a distance L ($L \simeq ct$ for relativistic neutrinos), Eq. (2.41) evolves as

$$|\nu_\alpha(t)\rangle = \sum_i U^*_{\alpha i} |\nu_i(t)\rangle = \sum_i U^*_{\alpha i} e^{-iE_i t} |\nu_i\rangle. \tag{2.43}$$

The transition probability of $\nu_\alpha \to \nu_\beta$ is given by

$$P(\nu_\alpha \to \nu_\beta) = |\langle \nu_\beta | \nu_\alpha(t) \rangle|^2 = \left| \sum_i U^*_{\alpha i} U_{\beta j} e^{-iE_i t} \right|^2. \tag{2.44}$$

For ultrarelativistic neutrinos, its energy can be approximated by $E_i = \sqrt{\boldsymbol{p}^2 + m_i^2} \simeq |\boldsymbol{p}| + m_i^2/2|\boldsymbol{p}|$. Denoting $E \equiv |\boldsymbol{p}|$ and $\Delta m_{ij}^2 := m_i^2 - m_j^2$, the oscillation probability is

$$P(\overset{(-)}{\nu}_\alpha \to \overset{(-)}{\nu}_\beta) = \delta_{\alpha\beta} - 4 \sum_{i<j} \text{Re}[U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}] \sin^2 \left(\frac{\Delta m_{ij}^2 L}{4E} \right)$$

$$\overset{(-)}{+} 2 \sum_{i<j} \text{Im}[U_{\alpha i} U^*_{\beta i} U^*_{\alpha j} U_{\beta j}] \sin \left(\frac{\Delta m_{ij}^2 L}{2E} \right), \tag{2.45}$$

where the sign of the second term flips for the antineutrino case. It is sometimes called the *appearance* oscillation probability as a neutrino with a different flavor appears during the propagation. When $\alpha = \beta$, it describes the non-oscillation (called the *survival* or *disappearance*) probability

$$P(\nu_\alpha \to \nu_\alpha) = P(\overline{\nu}_\alpha \to \overline{\nu}_\alpha) = 1 - 4 \sum_{i<j} \left| U_{\alpha i} U_{\alpha j}^* \right|^2 \sin^2 \left(\frac{\Delta m_{ij}^2 L}{4E} \right), \qquad (2.46)$$

which is identical for neutrinos and antineutrinos. This is because $U_{\alpha i} U_{\beta i}^* U_{\alpha j}^* U_{\beta j}$ is real for $\alpha = \beta$ and the second term in Eq. (2.45) becomes 0.

The leading terms of the dominant oscillation channels in the T2K accelerator neutrino measurements and the SK atmospheric neutrino measurements can be written as

$$P(\nu_\mu \to \nu_e) \simeq \sin^2 \theta_{23} \sin^2 2\theta_{13} \sin^2 \left(\frac{1.27 \Delta m_{32}^2 L}{E} \right), \qquad (2.47)$$

$$P(\nu_\mu \to \nu_\tau) \simeq \sin^2 2\theta_{23} \cos^4 \theta_{13} \sin^2 \left(\frac{1.27 \Delta m_{32}^2 L}{E} \right), \qquad (2.48)$$

$$P(\nu_\mu \to \nu_\mu) \simeq 1 - 4 \cos^2 \theta_{13} \sin^2 \theta_{23} (1 - \cos^2 \theta_{13} \sin^2 \theta_{23}) \sin^2 \left(\frac{1.27 \Delta m_{32}^2 L}{E} \right), \qquad (2.49)$$

$$P(\nu_e \to \nu_e) \simeq 1 - \sin^2 2\theta_{13} \sin^2 \left(\frac{1.27 \Delta m_{32}^2 L}{E} \right), \qquad (2.50)$$

where the units of the squared mass difference Δm^2, distance L, and energy E are [eV^2], [km], and [GeV], respectively [36]. The constant 1.27 is obtained as

$$\frac{\Delta m^2 L}{4E} = \frac{\text{eV}^2 \cdot \text{km}}{4 \cdot \text{GeV}} \frac{\Delta m^2 / \text{eV}^2 \cdot L/\text{km}}{E/\text{GeV}} = 1.27 \times \frac{\Delta m^2 / \text{eV}^2 \cdot L/\text{km}}{E/\text{GeV}}, \qquad (2.51)$$

where $c\hbar = 197$ fm $\cdot$ MeV.

The oscillation probabilities for different flavors of neutrinos as a function of the neutrino energy E_ν are shown in Fig. 2.1. It illustrates that the transition probability oscillates as a function of the neutrino energy.

2.2.5 CP violation in neutrino oscillation

Since the disappearance probability is identical for neutrinos and antineutrinos as shown in Eq. (2.46), the effect of CP violation does not appear in the disappearance channel. As a consequence, CP violation is only observable through the appearance probabilities $P(\nu_\alpha \to \nu_\beta)$ and $P(\overline{\nu}_\alpha \to \overline{\nu}_\beta)$.

In order to study CP violation in neutrino oscillation in a vacuum, it is convenient to introduce the quantity $\Delta P_{\alpha\beta}$ as

$$\Delta P_{\alpha\beta} := P(\nu_\alpha \to \nu_\beta) - P(\overline{\nu}_\alpha \to \overline{\nu}_\beta)$$
$$= -16 J_{\alpha\beta} \sin \left(\frac{\Delta m_{32}^2 L}{4E_\nu} \right) \sin \left(\frac{\Delta m_{31}^2 L}{4E_\nu} \right) \sin \left(\frac{\Delta m_{21}^2 L}{4E_\nu} \right), \qquad (2.52)$$
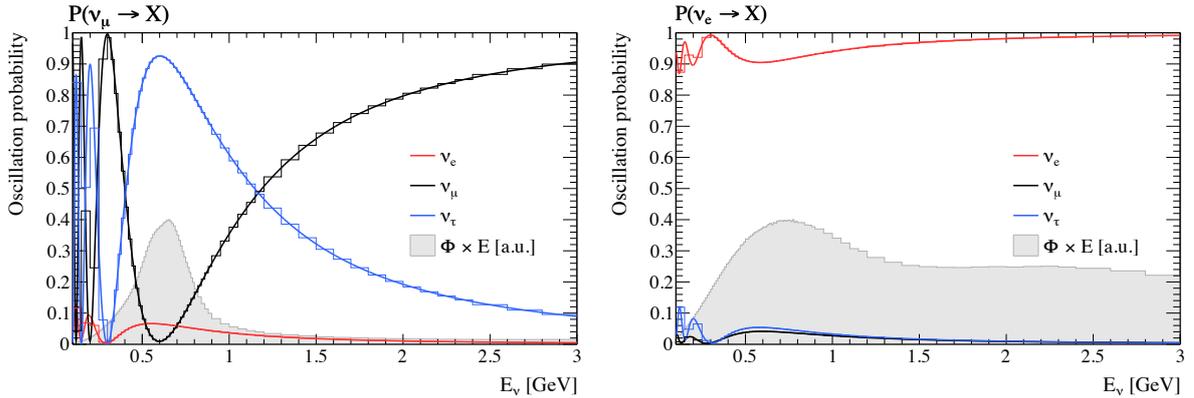
Figure 2.1. ν_μ and ν_e oscillation probabilities assuming the values of neutrino mixing angles, squared mass differences, and δ_{CP} obtained in Ref. [37]. The oscillation baseline length of 295 km is assumed. The neutrino flux weighted by the neutrino energy is overlaid to illustrate the relevant energy ranges for the T2K beam neutrinos. Since the neutrino-nucleus interaction cross-section is roughly proportional to the neutrino energy, the product of the neutrino flux and neutrino energy gives a rough approximation of the shape of the observed event spectra at the detector.

where $J_{\alpha\beta} := \mathrm{Im}\left[U_{\alpha1}U_{\alpha2}^*U_{\beta1}^*U_{\beta2}\right]$ is known as the Jarlskog invariant [38]. The condition for CP violation in neutrino oscillation in a vacuum is written as $\Delta P_{\alpha\beta} \neq 0$.

The Jarlskog invariant itself is a parameterization-independent measure of CP violation, but under the PMNS parameterization, it can be written as

$$J_{\alpha\beta} = \pm\frac{1}{8}\cos\theta_{13}\sin(2\theta_{13})\sin(2\theta_{12})\sin(2\theta_{23})\sin\delta_{\mathrm{CP}}$$
$$= \pm\sin\theta_{13}\cos^2\theta_{13}\sin\theta_{12}\cos\theta_{12}\sin\theta_{23}\cos\theta_{23}\sin\delta_{\mathrm{CP}}, \tag{2.53}$$

where $\pm$ denotes the cyclic (anti-cyclic) permutation of $(\alpha, \beta) = (e, \mu), (\mu, \tau), (\tau, e)$. From Eq. (2.52) and Eq. (2.53), the necessary conditions to obtain CP violation in neutrino oscillations are

$$\theta_{ij} \neq 0 \quad \cap \quad m_i \neq m_j \quad \cap \quad \delta_{\mathrm{CP}} \neq 0, \pi, \tag{2.54}$$

or equivalently,

$$m_i \neq m_j \quad \cap \quad J_{\alpha\beta} \neq 0. \tag{2.55}$$

2.2.6   Neutrino oscillation in matter

When neutrinos propagate through matter, the oscillation probabilities are affected by the interaction between neutrinos and particles in matter such as electrons, protons, and neutrons.

Neutrino propagation in matter can be described by the Schrödinger equation

$$i\frac{d}{dt}\left|\nu(t)\right\rangle = H\left|\nu(t)\right\rangle, \quad H = H_0 + V, \tag{2.56}$$

where the Hamiltonian H is split into the kinetic energy part H_0 and matter potential term V. H_0 describes the free propagation of neutrinos in a vacuum and gives the vacuum oscillation probability described in Eq. (2.56) when $V = 0$. The matter potential V is induced by coherent forward scatterings of neutrinos and the electrons and nucleons in the propagation medium [20]. The potential can be further broken down into the charged current contribution V_{CC} and the neutral current contribution V_{NC}. V_{CC} only affects the electron neutrinos because usually matter contains only electrons and no muons nor taus, and V_{NC} affects all three flavors equivalently. The matter potential is diagonal in the flavor basis and it can be written as

$$V = V_\alpha \delta_{\alpha\beta} = (V_{CC,\alpha} + V_{NC,\alpha})\delta_{\alpha,\beta} \tag{2.57}$$

$$V_{CC,\alpha} = \begin{cases} \sqrt{2}G_F n_e(x) & (\alpha = e) \\ 0 & (\alpha = \mu, \tau) \end{cases} \tag{2.58}$$

$$V_{NC,\alpha} = -\frac{G_F}{\sqrt{2}} n_n(x) \quad (\alpha = e, \mu, \tau), \tag{2.59}$$

where $G_F = 1.166 \times 10^{-5}$ GeV^{-2} is the Fermi constant, and $n_e(x)$ and $n_n(x)$ are the electron and neutron densities in the propagation medium.

Due to the corrections from the matter effect, the $\nu_\mu \to \nu_e$ oscillation probability formula in Eq. (2.47) is modified to be

$$P_M(\nu_\mu \to \nu_e) \simeq \sin^2\theta_{23} \sin^2 2\theta_M \sin^2\left(\frac{1.27\Delta m_M^2 L}{E}\right), \tag{2.60}$$

where Δm_M^2 and $\sin^2 2\theta_M$ are the effective oscillation parameters defined as

$$\Delta m_M^2 = \Delta m_{32}^2 \sqrt{(\cos 2\theta_{13} - \alpha/\Delta m_{32}^2)^2 + \sin^2 2\theta_{13}}, \tag{2.61}$$

$$\sin^2 2\theta_M = \frac{\sin^2 2\theta_{13}}{(\cos 2\theta_{13} - \alpha/\Delta m_{32}^2)^2 + \sin^2 2\theta_{13}}, \tag{2.62}$$

$$\alpha = \pm 2\sqrt{2}G_F n_e E_\nu, \tag{2.63}$$

where the $\pm$ sign of α is defined to be positive for neutrinos and negative for antineutrinos[7].

Equation (2.61) implies that a resonance occurs when $\cos 2\theta_{13} \approx \alpha/\Delta m_{32}^2$ depending on the sign of α (neutrino or antineutrino) and the sign of $\Delta m_{32}^2 = m_3^2 - m_2^2$ ($m_2 < m_3$ or $m_3 < m_2$). In other words, the resonance effect only appears for neutrinos when $m_2 < m_3$ and only appears for antineutrinos when $m_3 < m_2$.

The $\nu_\mu \to \nu_e$ ($\bar\nu_\mu \to \bar\nu_e$) oscillation probability approximated to first order in the

---

[7]Let us calculate α in more detail. The electron density n_e can be written as $n_e = N_A \times \frac{Z}{A} \times \rho$ cm^{-3} where $N_A = 6.02 \times 10^{23}$ mol^{-1} is the Avogadro constant, Z and A are the atomic number and the mass number, respectively. Using the relation cm $\cdot$ eV $= \frac{1}{1.97} \times 10^5$ and assuming $A \simeq 2Z$, α can be calculated as $\alpha = 7.56 \times 10^{-5}\rho(\text{g/vm}^3)E_\nu(\text{GeV})$. For the T2K oscillation analysis, we assume the constant density of $\rho = 2.6$ g/cm^3 [39].

matter effect can be written as [40]

$$
\begin{aligned}
P(\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_e) = \ & 4c_{13}^2 s_{13}^2 s_{23}^2 \sin^2 \Phi_{31} \\
& + 8c_{13}^2 s_{12} s_{13} s_{23}(c_{12} c_{23} \cos \delta_{\mathrm{CP}} - s_{12} s_{13} s_{23}) \cos \Phi_{32} \sin \Phi_{31} \sin \Phi_{21} \\
& \overset{(+)}{-} 8c_{13}^2 c_{12} c_{23} s_{12} s_{13} s_{23} \sin \delta_{\mathrm{CP}} \sin \Phi_{32} \sin \Phi_{31} \sin \Phi_{21} \\
& + 4s_{12}^2 c_{13}^2 (c_{12}^2 c_{23}^2 + s_{12}^2 s_{23}^2 s_{13}^2 - 2c_{12} c_{23} s_{12} s_{13} \cos \delta_{\mathrm{CP}}) \sin^2 \Phi_{21} \\
& - 8c_{13}^2 s_{13}^2 s_{23}^2 \frac{\alpha L}{4E_\nu}(1 - 2s_{13}^2) \cos \Phi_{32} \sin \Phi_{31} \\
& + 8c_{13}^2 s_{13}^2 s_{23}^2 \frac{\alpha}{\Delta m_{31}^2}(1 - 2s_{13}^2) \sin^2 \Phi_{31},
\end{aligned}
\tag{2.64}
$$

where the abbreviations are defined as follows

$$
c_{ij} \equiv \cos \theta_{ij}, \quad s_{ij} \equiv \sin \theta_{ij}, \quad \Phi_{ij} \equiv \frac{\Delta m_{ij}^2 L}{4E}.
\tag{2.65}
$$

We should note that the constant α (defined in Eq. (2.63)) in the last two terms in Eq. (2.64) changes the sign for neutrinos and antineutrinos, as well as the term proportional to $\sin \delta_{\mathrm{CP}}$. In the neutrino oscillation in a vacuum, CP violation was simply defined as the difference of the neutrino and antineutrino oscillation probabilities as described in Eq. (2.52). However, when we take into account the matter effect, the oscillation probabilities for neutrinos and antineutrinos are different even if $\delta_{\mathrm{CP}} = 0$, which makes the search for CP violation more complex.

Therefore, instead of simply comparing the neutrino and antineutrino oscillation probabilities as shown in Section 2.2.5, we rely on the PMNS parameterization and constrain all the relevant oscillation parameters simultaneously. This is done by measuring both the appearance and disappearance oscillation probabilities taking into account the contribution of the matter effect.

2.3 Review of the neutrino oscillation analysis

2.3.1 Latest oscillation analysis results

It has been shown that the neutrino oscillation probabilities can be parametrized using the mixing angles θ_{ij}, squared mass differences Δm_{ij}^2, and the leptonic CP phase δ_{CP}. It means that we can constrain these oscillation parameters through the neutrino oscillation experiments.

Figure 2.2 shows the evolution of the constraints on the six oscillation parameters ($\sin^2 \theta_{12}$, $\sin^2 \theta_{23}$, $\sin^2 \theta_{13}$, Δm_{21}^2, Δm_{32}^2, and δ_{CP}) that have been experimentally measured. They are sorted in the order of the first measurements. These six parameters based on the PMNS parametrization are in general regarded as the target of measurements in the neutrino oscillation analysis.

Nowadays oscillation experiments can be divided into four categories depending on the neutrino sources they use: solar neutrinos, atmospheric neutrinos, accelerator neutrinos, and reactor antineutrinos. The current major neutrino oscillation experiments are listed
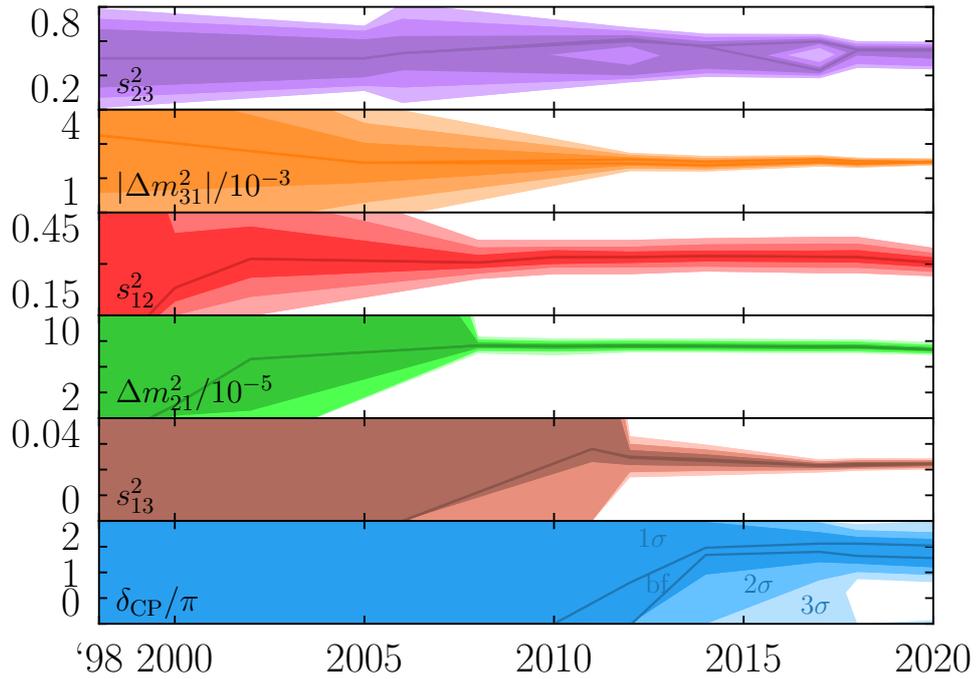
Figure 2.2. The evolution of constraints on the six neutrino oscillation parameters. The mixing parameters are denoted as $s_{ij}^2 = \sin^2\theta_{ij}$. The unit for the squared mass difference parameters is eV^2. The figure is taken from Ref. [41].

in Table 2.1. These neutrino sources provide different energies of neutrino fluxes and different oscillation baseline distances, which results in different sensitivities to the oscillation parameters in each experiment. In the following, the latest status of the neutrino oscillation parameter measurements is briefly reviewed.

Table 2.1. Neutrino sources of selected neutrino oscillation experiments. The baseline length and mean neutrino energy are also listed for the accelerator-based long-baseline oscillation experiments.

Experiment	Neutrino source	Reference	Baseline and mean ν energy
T2K	Accelerator	[37]	$L = 295$ km, $\langle E_\nu \rangle = 0.6$ GeV
NOνA	Accelerator	[42]	$L = 810$ km, $\langle E_\nu \rangle = 2.0$ GeV
IceCube	Atmospheric	[43]	
SK	Atmospheric, Solar	[36, 44]	
Borexino	Solar	[45]	
KamLAND	Reactor, Solar	[46]	
RENO	Reactor	[47, 48]	
Daya Bay	Reactor	[49]	
Double Chooz	Reactor	[50]	

θ_{12} **and** Δm_{21}^2 The mixing angle θ_{12} and mass squared difference Δm_{21}^2 are known as the *solar* parameters as they are relevant to the solar neutrino oscillation, but the reactor antineutrino measurements can also constrain them. The sign of Δm_{21}^2 was determined to be positive (i.e. $m_1 < m_2$) from the solar neutrino oscillation measurements by making use of the matter effect[8].

The best constraint comes from the KamLAND experiment [46] for Δm_{21}^2, while the solar experiments have the best constraints on $\sin^2 \theta_{12}$. The latest constraints on these parameters from KamLAND and the solar neutrino experiments are [17]

$$\sin^2 \theta_{12} = 0.307^{+0.013}_{-0.012}, \tag{2.66}$$
$$\Delta m_{21}^2 = 7.53 \pm 0.18 \times 10^{-5} \text{ eV}^2. \tag{2.67}$$

θ_{23} **and** Δm_{32}^2 The mixing angle θ_{23} and squared mass difference Δm_{32}^2 are known as the *atmospheric* parameters as they are relevant to the atmospheric neutrino oscillation. The accelerator-based neutrino oscillation experiments also have good sensitivities to these parameters.

The atmospheric mixing angle θ_{23} is known to be close to maximal mixing ($\theta_{23} \sim 45°$). However, whether this parameter is above $45°$ or below $45°$ is still unknown, which is called the *octant* problem and is one of the open questions in neutrino oscillation physics. The octant of θ_{23} defines the amplitude of the neutrino flavor mixing in a mass eigenstate. When θ_{23} is in the upper octant ($\theta_{23} > 45°$), the mass eigenstate ν_3 (which has the least ν_e contribution) has more ν_μ components, and it has more ν_τ components otherwise.

The atmospheric squared mass difference is measured to be [17]

$$\Delta m_{32}^2 = m_3^2 - m_2^2 = \begin{cases} (2.453 \pm 0.033) \times 10^{-3} & \text{eV}^2 \\ (-2.536 \pm 0.034) \times 10^{-3} & \text{eV}^2 \end{cases} \tag{2.68}$$

from the accelerator and atmospheric neutrino experiments. The sign of Δm_{32}^2 is not known because the leading terms in the oscillation probabilities (Eqs. (2.45) and (2.46)) are proportional to $\sin^2 \frac{\Delta m_{ij}^2 L}{4E}$, which does not give the sign of squared mass difference. Therefore, there remains uncertainty about whether the neutrino masses are ordered as $m_1 < m_2 < m_3$ or $m_3 < m_1 < m_2$. The former hypothesis is called the "normal ordering" (NO) and the latter is called the "inverted ordering" (IO). Fig. 2.3 shows the illustration of two possible mass orderings.

The comparison of the 90% confidence level regions in $\sin^2 \theta_{23}$-Δm_{32}^2 assuming normal ordering from some experiments is shown in Fig. 2.4. The results of the long baseline neutrino oscillation experiments such as T2K and NOνA have the best sensitivities to these atmospheric parameters. These results are overall very consistent but have slight differences in the preferred region of Δm_{32}^2. For θ_{23}, none of the experiments have a strong preference for either of the octants.

θ_{13} The mixing angle θ_{13} is the *reactor* mixing angle. The constraints mostly come from the reactor experiments such as Double Chooz [50], RENO [48, 47], and Daya Bay [49, 54].

---

[8]See for example Ref. [51].
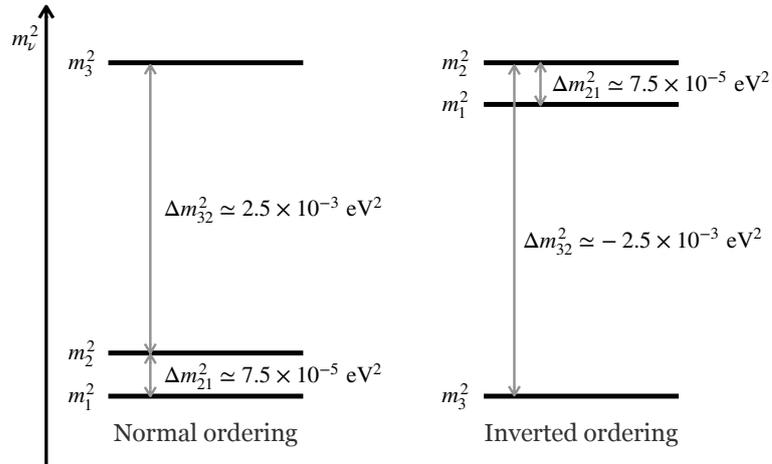
Figure 2.3. Scheme of the neutrino mass ordering hypotheses. The squared mass difference values are taken from Ref. [17].
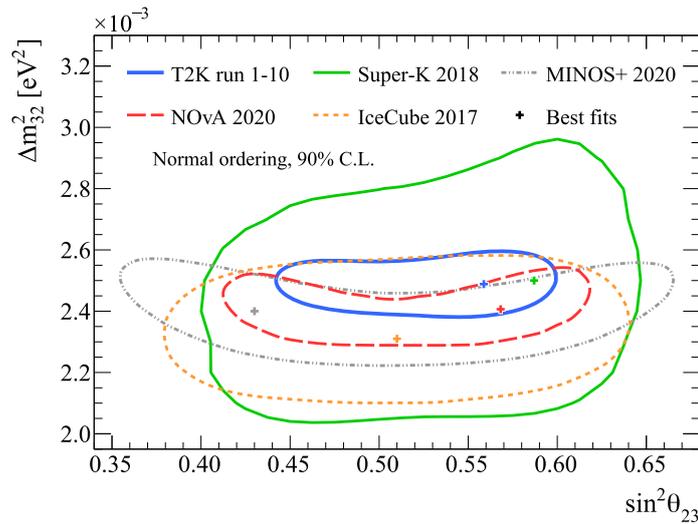
Figure 2.4. Comaprison of the 90% confidence level regions in $\sin^2 \theta_{23}$-Δm_{32}^2 among different neutrino oscillation experiments. The normal ordering is assumed. The data is taken from T2K [37], NOνA [42], SK [52], IceCube [43], and MINOS+ [53]. The figure is taken from Ref. [37].

The current world average (the weighted average of these reactor experiments) is [17]

$$\sin^2 \theta_{13} = 2.20 \pm 0.07 \times 10^{-2}. \tag{2.69}$$

$\delta_{\mathbf{CP}}$ The CP violation phase δ_{CP} is mostly constrained by the accelerator-based long-baseline oscillation experiments. Figure 2.5 shows the comparison of 68% and 90% confidence level regions in δ_{CP}-$\sin^2 \theta_{23}$ for the two leading long-baseline experiments T2K and NOνA. T2K has a better constraint on δ_{CP} and has a best-fit point around $\delta_{\mathrm{CP}} \approx 3\pi/2$.

However, the current measurements show a large discrepancy in the preferred region of δ_{CP} for normal ordering between T2K and NOνA, which is therefore of particular interest in future analyses.
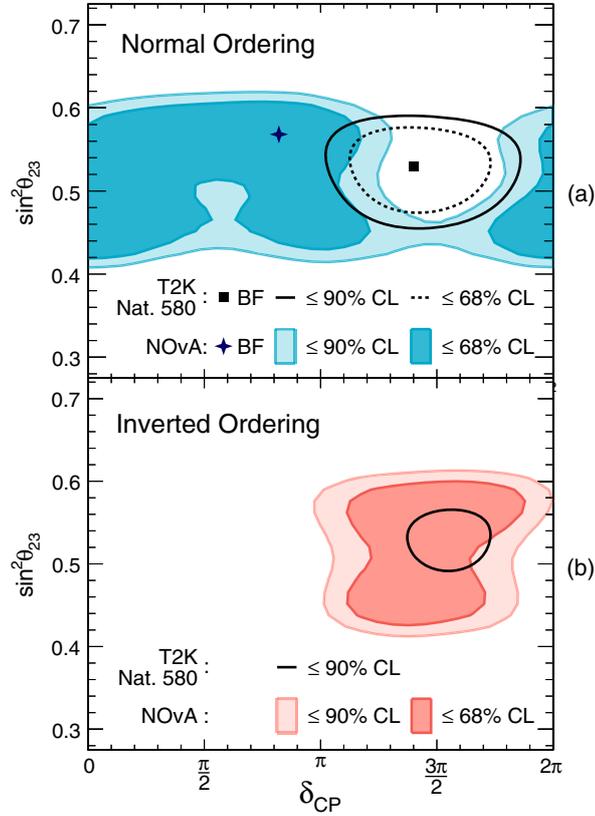
Figure 2.5. Comaprison of the 68% and 90% confidence level regions in δ_{CP}-$\sin^2\theta_{23}$ between T2K [6] and NOνA [42] for each mass ordering. The figure is taken from Ref. [42].

2.3.2 Open questions in neutrino oscillation physics

The neutrino oscillation provides a key to search for CP violation in the lepton sector. One of the unanswered questions in the present physics field is the origin of the baryon number asymmetry of the Universe[9], where it is mostly composed of matter, and only a small fraction of antimatter exists. Although a specific mechanism to explain this asymmetry is still unknown, three general conditions to be satisfied are known as the Sakharov conditions [55]:

1. C and CP violation

2. Baryon number violation

3. Existence of non-equilibrium process.

_____

[9]It is also known as the matter-antimatter asymmetry of the Universe.

One of the possible scenarios to generate such an asymmetry is called the leptogenesis [4]. In this scenario, the lepton number asymmetry is created by the Majorana neutrinos and it is converted into the baryon number asymmetry through sphaleron processes [56]. A study has shown that the baryon number asymmetry can be explained even exclusively by CP violation in the Dirac phase in a certain scenario [57]. Therefore, if CP violation in the lepton sector is proved through the precise measurements of neutrino oscillations, it has the possibility of contributing to the explanation of the origin of the baryon number asymmetry of the Universe.

In addition to the CP phase measurements, the fundamental properties of neutrinos, such as the squared mass differences and mixing angles are the target of measurements in the neutrino oscillation analysis. The mass ordering of neutrinos and the octant of θ_{23} are particularly of interest among them. In summary, there are three open questions in neutrino oscillation physics:

- CP symmetry in the lepton sector

- Neutrino mass ordering (normal ordering or inverted ordering)

- Octant of θ_{23}.

2.4 Motivation of the joint analysis of atmospheric and accelerator neutrinos

As shown in Section 2.3, different experiments have different sensitivities to the oscillation parameters. Therefore, it is valuable to combine the analyses of two or more experiments to get better constraints on the parameters of interest. In this thesis, a joint analysis of the SK atmospheric neutrinos and the T2K accelerator neutrinos is performed to try to answer the open questions in neutrino oscillation physics.

2.4.1 Overview of the experiments

T2K beam neutrino analysis The T2K (Tokai-to-Kamioka) experiment is a long baseline neutrino oscillation experiment ongoing in Japan. It aims to search for CP violation and measure the neutrino mixing angles and square mass differences such as $\sin^2 \theta_{23}$, $\sin^2 \theta_{13}$, and Δm_{32}^2. T2K uses the neutrinos produced at the Japan Proton Accelerator Research Complex (J-PARC) in Ibaraki and detects them at the Super-Kamiokande detector which is located 295 km away from the neutrino production point. The schematic view of the T2K experiment baseline is shown in Fig. 2.6. J-PARC provides very pure ν_μ and $\bar{\nu}_\mu$ neutrino sources, which allows us to precisely measure the $\nu_\mu \to \nu_e$ ($\bar{\nu}_\mu \to \bar{\nu}_e$) appearance probabilities as well as the $\nu_\mu \to \nu_\mu$ ($\bar{\nu}_\mu \to \bar{\nu}_\mu$) disappearance probabilities. The near detectors located near the neutrino production target are also used to measure the neutrino spectra before the oscillation, which helps to constrain the uncertainties in the neutrino flux and neutrino-nucleus interaction cross-sections.

As for the oscillation parameter sensitivities, T2K has good sensitivity to δ_{CP} through the measurement of $\nu_\mu \to \nu_e$ and $\bar{\nu}_\mu \to \bar{\nu}_e$ appearance probabilities. In addition, the

shape and normalization of the $\nu_\mu \to \nu_\mu$ ($\bar{\nu}_\mu \to \bar{\nu}_\mu$) disappearance probabilities give the sensitivities to $\sin^2 \theta_{23}$ and Δm_{32}^2. In the latest analysis [37], T2K reported that CP conserving values of $\delta_{\text{CP}} = 0, \pi$ are excluded at more than 90% confidence level.
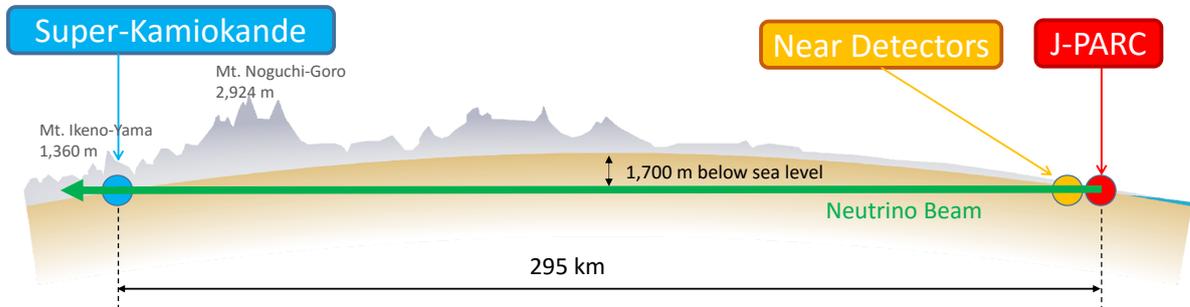
Figure 2.6. Schematic overview of the T2K experiment.

SK atmospheric neutrino analysis The Super-Kamiokande (SK) detector is a large water Cherenkov detector located in Gifu, Japan. The detector is used to perform various physics programs including not only the accelerator-based long-baseline neutrino oscillation measurements, but also the atmospheric and solar neutrino measurements [52, 36, 44], diffuse supernova neutrino background (DSNB) searches [58], nucleon decay searches [59], and dark matter searches [60].

The SK atmospheric neutrino analysis also has sensitivities to δ_{CP} and the atmospheric oscillation parameters $(\Delta m_{32}^2, \sin^2 \theta_{23})$. Figure 2.7 shows the oscillation probabilities of atmospheric muon neutrinos for normal ordering. The zenith angle (Θ_z) is defined as the direction of neutrino origin to the normal vector on Earth's surface at the location of SK. Therefore, $\cos \Theta_z > 0$ corresponds to the down-going events and $\cos \Theta_z < 0$ corresponds to the upward-going events. The propagation distance of atmospheric neutrinos depends on the zenith angle and spans from ~ 15 km for the down-going events to $\sim 13,000$ km for the upward-going events. The oscillation probabilities (which are dependent on L/E_ν) are therefore dependent on both the neutrino energies and zenith angles as shown in Fig. 2.7. The baseline distance for T2K is optimized to observe the first oscillation maximum, whereas SK has a more complex oscillation probability distribution as it includes the following oscillation maxima. In the actual observation, however, only the overall normalizations can be measured due to the detector resolutions. The sensitivity to δ_{CP} comes from the normalization of event rates around the sub-GeV energy regions.

2.4.2 Motivation of the joint analysis

Although T2K has a good sensitivity to δ_{CP}, the contribution of δ_{CP} to the oscillation probability is partially degenerated with the mass ordering effect, which makes the search for CP violation more complicated. Fig. 2.8 shows the event rates of the observed ν_e and $\bar{\nu}_e$ candidates in T2K overlaid with the expectation at different combinations of true oscillation parameter values. The different markers in each ellipse correspond to different values of δ_{CP}, the solid and dashed ellipses correspond to the normal and inverted ordering,
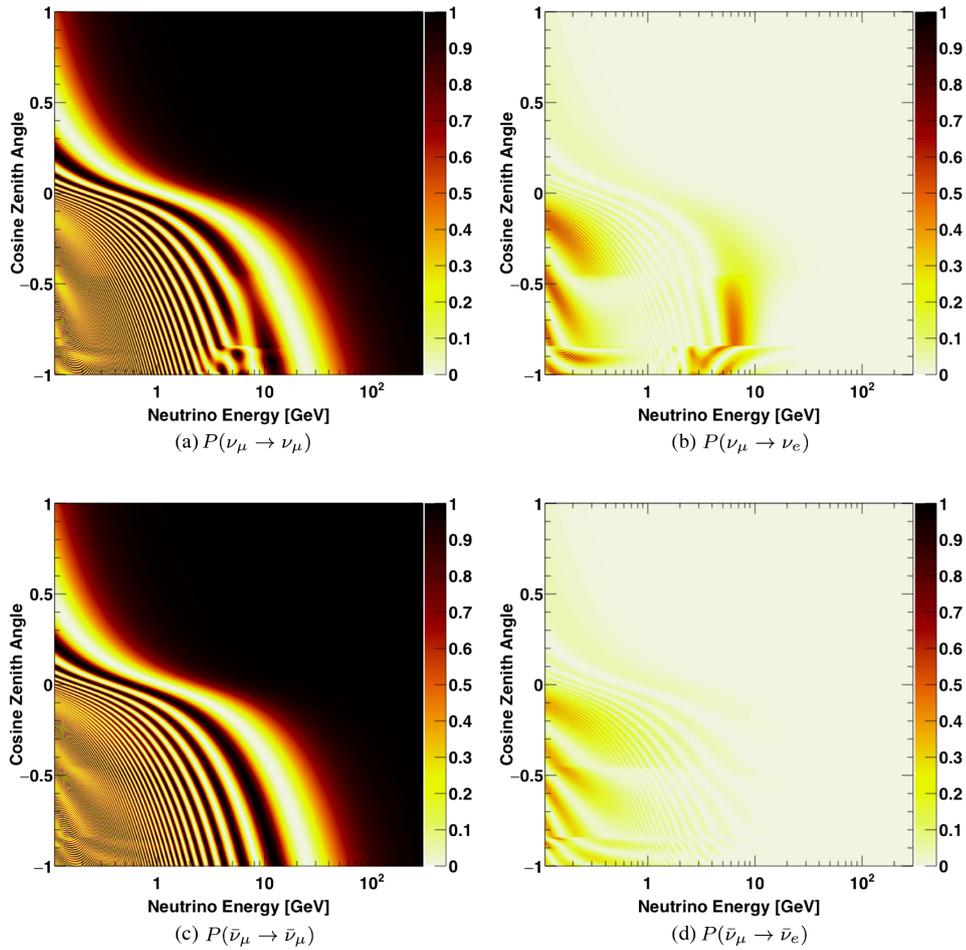
Figure 2.7. $\nu_\mu \to \nu_\mu$ (left) and $\nu_\mu \to \nu_e$ (right) oscillation probabilities for neutrinos (top) and antineutrinos (bottom) as a function of true neutrino energy and direction (in zenith angle) for normal mass ordering. The oscillation parameters are assumed to be $\Delta m_{32}^2 = 2.5 \times 10^{-3}$ eV^2, $\sin^2\theta_{23} = 0.5$, $\sin^2\theta_{13} = 0.0219$, and $\delta_{\rm CP} = 0$. The figures are taken from Ref. [52].

and different colors correspond to different values of $\sin^2\theta_{23}$. The different values of $\delta_{\rm CP}$ appear as an anticorrelated change in the ν_e and $\bar{\nu}_e$ event rates (i.e. we will have more ν_e events when $\delta_{\rm CP} = -\pi/2$ and more $\bar{\nu}_e$ events when $\delta_{\rm CP} = \pi/2$). The plot clearly illustrates that the latest T2K data lies near the maximal violation of CP ($\delta_{\rm CP} = -\pi/2$). However, some values of $\delta_{\rm CP}$ are overlapped with different values of $\delta_{\rm CP}$ in the other mass ordering. If the data falls into this area, we cannot distinguish these overlapped $\delta_{\rm CP}$ values. This is called the *degeneracy* of $\delta_{\rm CP}$ and mass ordering, and it limits the sensitivity to $\delta_{\rm CP}$ in the T2K oscillation analysis.

On the other hand, in the SK atmospheric neutrino analysis, the matter effect could provide a key to distinguishing the neutrino mass ordering. Figure 2.7 demonstrates that we have a distortion in the $\nu_\mu \to \nu_e$ oscillation probability around a few GeV due to the resonance effect discussed in Section 2.2.6. The distortion only happens in the
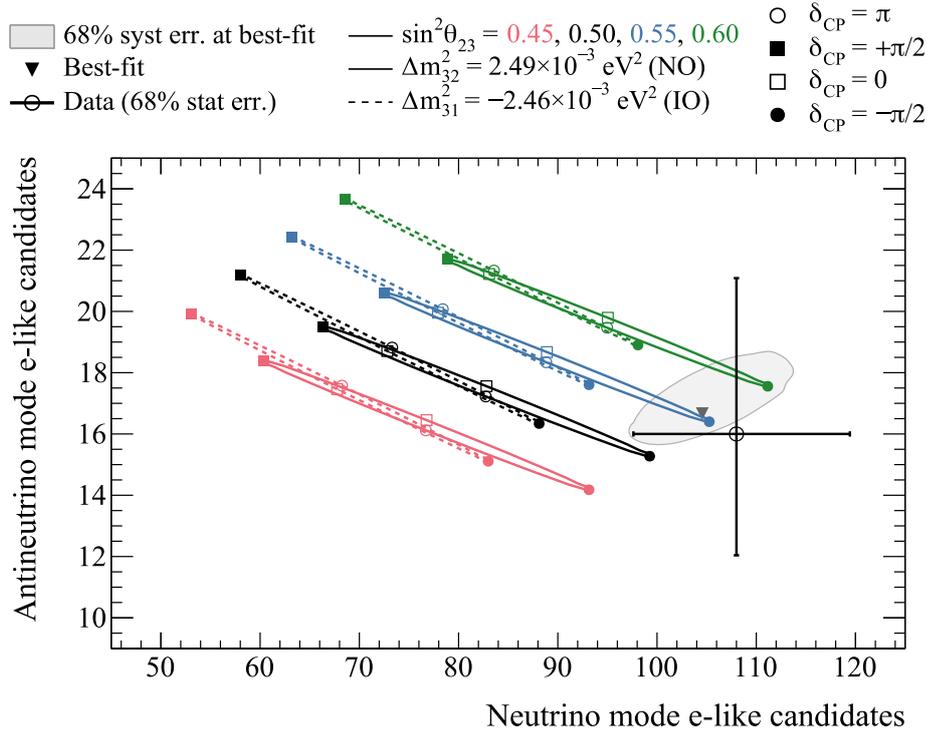
Figure 2.8. Number of observed ν_e candidate events and $\bar{\nu}_e$ candidate events with expectation at the several combinations of the true oscillation parameters. The error bars on the data point show the statistical errors. The figure is taken from Ref. [37].

neutrino oscillations if the true mass ordering is normal, while it only happens in the antineutrino oscillations when the inverted ordering is true. Therefore, it may be possible that combining these two experiments could solve the degeneracy between $\delta_{\rm CP}$ and mass ordering, and improve the sensitivities.

In addition, sensitivities of the neutrino oscillation experiments are in general limited by the statistical uncertainties. It is therefore beneficial to combine the data from two experiments simply in terms of the increase in the statistics. It is also possible to apply the T2K near detector constraints to the SK atmospheric neutrino analysis to better constrain the neutrino cross-section systematic uncertainties.

In summary, the first joint analysis is performed using the T2K accelerator neutrinos and the SK atmospheric neutrinos in this thesis. There are three main motivations for performing the joint analysis between T2K accelerator neutrinos and SK atmospheric neutrinos:

1. Solve the degeneracy between $\delta_{\rm CP}$ and mass ordering

2. Increase the data statistics

3. Constrain the systematic uncertainties better

Chapter 3

Experimental setup

3.1 Neutrino sources

In this thesis, we perform a joint analysis of the atmospheric neutrinos and the accelerator neutrinos. This section presents a description of each neutrino source.

3.1.1 Atmospheric neutrinos

The primary cosmic rays The atmospheric neutrinos are produced by the primary cosmic rays hitting the nucleons in the atmosphere of the Earth. The primary cosmic rays are dominated by the protons ($\approx 85\%$), but have some contributions from α-particles ($\approx 11\%$), electrons ($\approx 2\%$), and heavier nuclei ($\approx 2\%$) as well. The fluxes of these cosmic rays can be directly measured by experiments in space (e.g. on the international space station [61]) or balloons [62]. Since the primary cosmic rays are composed of charged particles, their trajectories can be affected by magnetic fields. Two types of effects are known to modulate their fluxes. One of them is the solar activity, which suppresses the flux of lower energy ($\lesssim 10$ GeV) cosmic rays reaching the Earth [63]. Since solar activity has a periodicity of 11 years, it gives an anticorrelated 11-year periodic effect to the cosmic ray flux. The second effect is the geomagnetic field of the Earth, which causes both the latitude (the cosmic ray flux is larger near the geomagnetic poles) and longitude effects (the cosmic ray flux is larger for east-going particles). The latter is especially called the "east-west effect" and was experimentally observed [64]. These effects are taken into account in the calculation of the atmospheric neutrino flux.

Production of atmospheric neutrinos The primary cosmic rays reaching the atmosphere interact with nucleons and produce secondary particles such as pions and kaons, which decay into neutrinos. The dominant decay chains are

$$\pi^{\pm} \to \mu^{\pm} + \nu_\mu(\bar{\nu}_\mu), \tag{3.1}$$

$$K^{\pm} \to \mu^{\pm} + \nu_\mu(\bar{\nu}_\mu), \tag{3.2}$$

$$K_L \to \pi^{\mp} + \mu^{\pm} + \nu_\mu(\bar{\nu}_\mu), \tag{3.3}$$

$$K_L \to \pi^{\mp} + e^{\pm} + \nu_e(\bar{\nu}_e), \tag{3.4}$$

which are then followed by the muon decays

$$\mu^{\pm} \to e^{\pm} + \nu_e(\bar{\nu}_e) + \bar{\nu}_\mu(\nu_\mu). \tag{3.5}$$

The subsequent decay chains of charged pions (kaons) and muons produce a roughly twice larger flux of muon neutrinos compared to that of electron neutrinos

$$\frac{N(\nu_\mu) + N(\bar{\nu}_\mu)}{N(\nu_e) + N(\bar{\nu}_e)} \approx 2. \tag{3.6}$$

However, the contribution of electron neutrinos from the muon decays is suppressed above a few GeV because the muon with higher energies can reach the surface of the Earth before decaying into electrons and neutrinos. Therefore, at higher energies, the contribution from the kaon decays becomes more important for the electron neutrino flux.

3.1.2 J-PARC accelerator neutrinos

In general, the neutrino beam generation has two main steps: acceleration of the protons and production of the neutrinos. At J-PARC, the former step is done by a subsequent chain of three accelerators, and the latter is done at the neutrino beamline. The J-PARC proton accelerator consists of three accelerators: a linear accelerator (LINAC), a rapid-cycling synchrotron (RCS), and the main ring (MR) synchrotron [65]. A H$^-$ ion (1 proton and 2 electrons) is first injected into LINAC and is accelerated up to 400 MeV. Then it is converted to H$^+$ (proton) by removing the two electrons using the charge-stripping foils. The proton beam is accelerated up to 3 GeV by the RCS and to 30 GeV by the MR, and provided to each experimental facility including the neutrino beamline. The J-PARC accelerator can produce proton beam spills with a time spread of ~ 5 μsec at the repetition rate of 2.48 s (0.4 Hz). Each spill has an 8-bunch structure with a narrow bunch width (~ 15 ns), which helps us to identify the events that originated from the beam neutrinos. By requiring the timing cut, most of the background events (e.g. cosmic rays and atmospheric neutrinos) can be rejected at the first stage of the event selection at the far detector.

Fig. 3.1 shows the schematic view of the neutrino beamline at J-PARC. The protons injected from the MR strike a graphite target and create secondary pions and other hadrons. These pions are focused by the three magnetic horns and decay inside the 96 m-long decay volume, producing muons and muon neutrinos in flight via the following processes:

$$\pi^{\pm} \to \mu^{\pm} + \nu_\mu(\bar{\nu}_\mu). \tag{3.7}$$

For 3 GeV or higher energy neutrinos, dominant contributions are from kaon decays such as

$$K^{\pm} \to \mu^{\pm} + \nu_\mu(\bar{\nu}_\mu), \tag{3.8}$$
$$K^{\pm} \to \pi^0 + \mu^{\pm} + \nu_\mu(\bar{\nu}_\mu). \tag{3.9}$$

Since the type of produced neutrinos (neutrino or antineutrino) depends on the charge of the parent pion, we can produce a neutrino-dominant beam or antineutrino-dominant

beam by focusing only pions and kaons with the selected charge. The polarity of the magnetic field made by the horns for the neutrino-dominant beam mode is called the forward horn current (FHC; +250 kA), and for the antineutrino-dominant beam mode, it is called the reversed horn current (RHC; −250 kA). Although the neutrino beams are dominated by ν_μ ($\bar{\nu}_\mu$) for FHC (RHC), there are some contaminations of the *wrong-sign* neutrinos ($\bar{\nu}$ for FHC and ν for RHC) and electron neutrinos. The wrong-sign neutrino background mainly comes from the imperfection of the horn focusing, including the contaminations from pions produced outside the target and from forward-going wrong-sign pions that are not de-focused by the magnetic field.
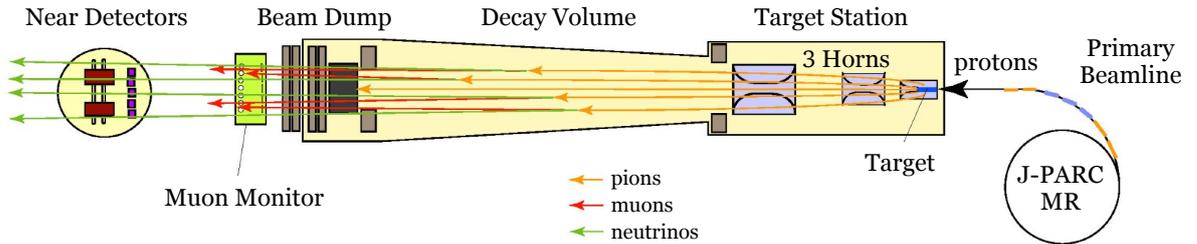
Figure 3.1. Schematic view of the J-PARC neutrino beamline.

Hadrons are stopped by a beam dump located 109 m downstream from the graphite target. High-energy muons can penetrate the beam dump and are measured by the muon monitor to monitor the two-dimensional profile of the beam direction and the beam intensity [66]. The neutrino beam is directed to 2.5° off from the SK direction. This off-axis angle provides us with a narrower neutrino energy spectrum than that at the on-axis, which has a peak around the oscillation maximal point as shown in Fig. 3.2.
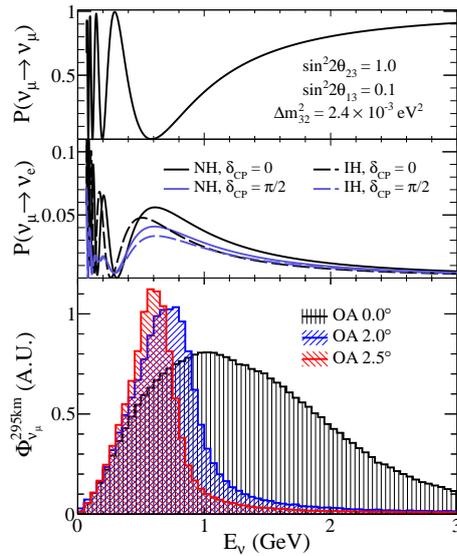
Figure 3.2. T2K flux predictions at the far detector at different off-axis angles in an arbitrary unit. The $\nu_\mu \to \nu_e$ appearance and $\nu_\mu \to \nu_\mu$ disappearance probabilities are shown in the top two panels.

3.2 Detectors

In this analysis, we use two sets of detectors. One is the Super-Kamiokadne detector which is used for both the T2K beam and SK atmospheric neutrino measurements. The other is the T2K near detector ND280 which is used to measure the T2K beam neutrinos before they oscillate and to constrain the flux and cross-section systematic uncertainties.

3.2.1 Super-Kamiokande detector

3.2.1.1 Overview

The SK detector is located 1 km underground of Mt. Ikenoyama [67]. The detector is a cylindrical tank with a height and a diameter of 41.4 m and 39.3 m, respectively. The detector is filled with 50 kton of ultrapure water (and gadolinium since 2020) and is equipped with Photo-Multiplier Tubes (PMT) on the wall. The detector consists of two parts: the inner detector (ID) and the outer detector (OD). The ID has a height of 36.3 m and a diameter of 33.8 m and is equipped with 11,129 inward-facing 20-inch PMTs. This is used as the main target volume in the analysis. The OD surrounds the ID and has a thickness of 2.2 m at the barrel and 2.06 m on the top and bottom. The OD is equipped with 1,885 outward-facing 8-inch PMTs. It is used to detect the charged particles coming from outside the detector or charged particles escaping from the ID. The schematic view of the SK detector is shown in Fig. 3.3.
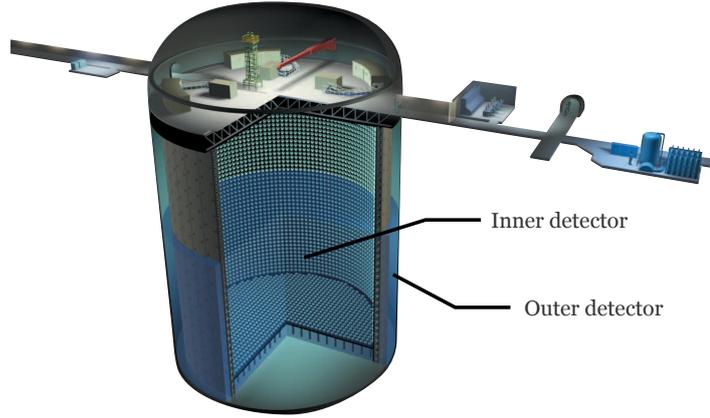
Figure 3.3. Schematic view of the Super-Kamiokande detector.

3.2.1.2 Cherenkov radiation

When a charged particle goes through a medium, it excites the atoms inside the medium along its trajectory. The excited atoms emit light isotropically such that the waves of the light cancel each other. However, when the charged particle moves faster than the speed of light in the medium, the excited atoms emit deexcitation light continuously before being canceled out, which makes a conical shape of light toward the direction of the charged particle. This phenomenon is called the Cherenkov radiation, and its light is called the Cherenkov light. It is used to detect the charged particles traversing the SK detector. Since the conical light forms ring-shaped hit patterns when projected onto the detector walls, these signals are often referred to as a *ring* in SK. Denoting the velocity of the particle as $\beta = v/c$ and the refractive index of the medium as n, the opening angle of the Cherenkov light (θ_C) can be written as

$$\cos\theta_C = \frac{c/n \times t}{\beta c \times t} = \frac{1}{n\beta}. \tag{3.10}$$

The refractive index of pure water is $n \simeq 1.34$, so the opening angle is $\theta_C \sim 42°$ when the charged particle's velocity is close to the speed of light $\beta \sim 1$.

The condition to emit Cherenkov light is $v > c/n$ (i.e. $\beta > 1/n$). Therefore, the momentum threshold of the Cherenkov emission depends on the mass of the charged particles (m) as

$$p_C = \frac{mc}{\sqrt{n^2 - 1}}, \tag{3.11}$$

which is roughly $p_C = 1.12 \times m$ for pure water. The momentum thresholds for charged particles of interest are summarized in Table 3.1. When charged particles have momentum below the threshold, they cannot be detected as a Cherenkov light signal at SK. Pions and muons can also be identified by detecting delayed signals from the decay electrons even though their momentum is below the threshold.

Table 3.1. Cherenkov thresholds for charged particles of interest. The particle masses are taken from Ref. [17] and the threshold values are computed from Eq. (3.11).

Particle	Mass [MeV/c^2]	Momentum threshold [MeV/c]
$e^{\pm}$	0.511	0.573
$\mu^{\pm}$	105.7	118.4
$\pi^{\pm}$	139.6	156.5
p	938.3	1052

3.2.1.3 Photo-Multiplier Tubes

The Cherenkov light is detected by the PMTs attached to the surface of the detector walls. The effective photocathode coverage of the ID is around 40%. The schematic view of the ID PMT and its quantum efficiency as a function of wavelength is shown in Fig. 3.4. The quantum efficiency is the number of photoelectrons emitted from the photocathode of the PMT divided by the number of photons reaching the surface of the cathode and is about 21% at the peak. These PMTs are operated with a gain of 10^7, and the average dark noise rate is about 3 kHz at a threshold of 0.25 photoelectron. The signals detected at the PMTs are transported to the front-end electronics via 70 m coaxial cables.

(a) Schematic view

(b) Quantum efficiency
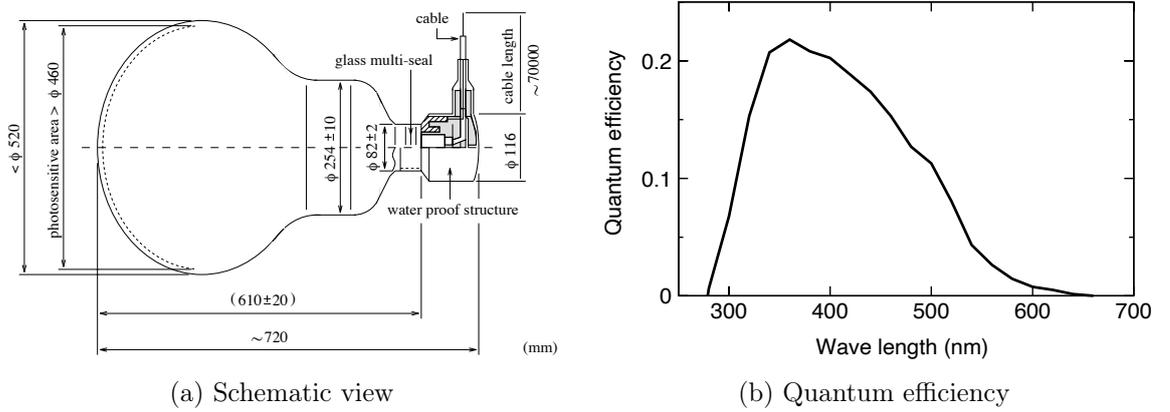
Figure 3.4. Schematic view of the inner detector PMT and its quantum efficiency as a function of the wavelength of the incoming light. The figures are taken from Ref. [67].

3.2.1.4 Electronics

The main function of the electronics is to convert the analog pulse signals obtained by the PMTs into the digitized information of integrated charge and the arrival timing of the signal. As the front-end electronics, SK uses the QTC-Based Electronics with Ethernet (QBEE) [68].

QBEE has an application-specific integrated circuit (ASIC), the high-speed charge-to-time converter (QTC) IWATSU CLC101, which converts the input signal to the special pulse whose rising edge represents the timing and width represents the integrated charge, respectively. The ASIC provides three output signals with different amplification ratios

(i.e. different dynamic ranges), and a field-programable gate array (FPGA) selects one with the most appropriate dynamic range. Therefore, it allows us to choose the best dynamic range event-by-event. Another benefit of using QBEE is that we can obtain a long data acquisition window for each event. It allows us to implement an additional trigger that saves data in the time window of $[35, 535]$ μs following the signal trigger.

3.2.1.5   Reconstruction

The SK reconstruction aims to convert the observed charge and timing information from the PMT hits into the variables used in the analysis. The variables include the position of the neutrino interaction vertex, number of rings, particle identification (PID), and kinematics (direction and energy) of each reconstructed particle. For the PID, only three particle hypotheses, electron, muon, and pion are considered.

The reconstruction tool used in the SK analysis is called `fiTQun` [36]. The `fiTQun` reconstruction process consists of four steps: vertex pre-fit, PMT hit clustering, single-ring reconstruction, and multi-ring reconstruction.

Vertex pre-fit and hist clustering   The vertex pre-fit is a fast algorithm that can search for the approximate position of the event vertex only from the timing information of the PMTs. The pre-fit vertex found in this algorithm will be used as an input in the following steps and will be fit again with higher precision. Using the pre-fit vertex, the PMT hits are clustered based on the timing and separated into subevents which represent the prompt signal and delayed signals. Since the timing measurements also depend on the vertex position due to the time of flight, we need to find a vertex for each subevent (e.g. a Michel electron can have a different vertex from the parent particle). Therefore, the vertex pre-fitting step is applied to each hit cluster.

Single-ring fit   The single-ring fit is then applied to each subevent where the particle's kinematics is reconstructed under different lepton PID hypotheses (i.e. e-like or μ-like). The PID is determined by comparing the likelihood of the e-like hypothesis and the μ-like hypothesis and taking the larger one. The misidentification rate of e/μ PID is less than $\sim 1\%$ when the visible energy is $E_{\mathrm{vis}} < 1330$ MeV [36]. Here the visible energy is the reconstructed momentum of the observed ring under the electron hypothesis ($E_{\mathrm{vis}} := p_e$), which represents approximately the total energy observed by the PMTs. When there is more than one reconstructed ring, the visible energy is the sum of them ($E_{\mathrm{vis}} := \sum p_e$).

Multi-ring fit   The multi-ring fit aims to find events with more than one charged particle (e.g. one lepton and 1 or more pions). The basic algorithm of the multi-ring fit is the same as the single-ring fit, but here we repeat the fit by adding a new ring to the previous fit hypothesis. In the multi-ring fit, a π^+-like hypothesis is also included. The fit is repeated until the likelihood will not be improved by the newly added ring, or the number of rings reaches the limit of six. The results of these reconstructions such as the number of rings, PID, and particle kinematics will be used in the sample classification and oscillation analysis.

Decay electron tagging   Decay electron signals produced by the pion or muon decay chains are tagged using the delayed signals. When neutrons produced in the neutrino interaction are captured by protons in nuclei, they emit a gamma ray with 2.2 MeV, which can be a background for the decay electron tagging. Therefore, the decay electron signals are selected by requiring both the timing and the number of hits in the delayed hit cluster to satisfy the condition as shown in Fig. 3.5.
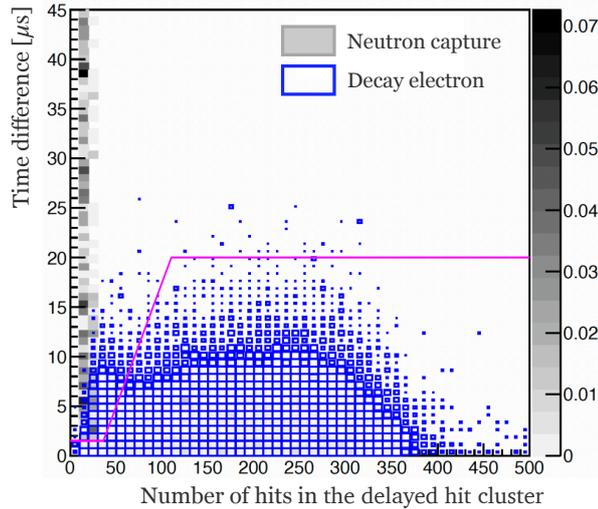
Figure 3.5. Two-dimensional distributions of the time difference between the prompt signal and the delayed signal and the number of hits within the 50 ns time window for the delayed signals produced from the decay electrons and neutron captures. The events under the purple line are selected as the decay electron candidates.

3.2.2   T2K near detector (ND280)

The T2K near detector complex, ND280, is located roughly 280 m downstream of the graphite target. The schematic view of ND280 is shown in Fig. 3.6. ND280 consists of several subdetectors placed in the magnet[1] which provides a 0.2 T magnetic field so that we can measure the charge and momentum of the particles produced in the neutrino interactions. The description of each subdetector is given in the following.

FGD   The Fine Grained Detectors (FGD) are plastic scintillator (and partially water) target detectors [71]. There are two FGDs, upstream and downstream which are called FGD1 and FGD2, respectively. FGDs consist of several sub-modules that are made with fine-grained scintillator bars of size 184 cm × 0.96 cm × 0.96 cm that are oriented perpendicular to the beam direction. One sub-module is composed of two orthogonal layers that consist of 192 scintillator bars in the horizontal and vertical directions. While FGD1 has fifteen sub-modules, FGD2 has seven modules and six water sub-modules with thin-walled

---

[1]The ND280 magnet was originally used in the UA1 experiment [69] and the NOMAD experiment [70].
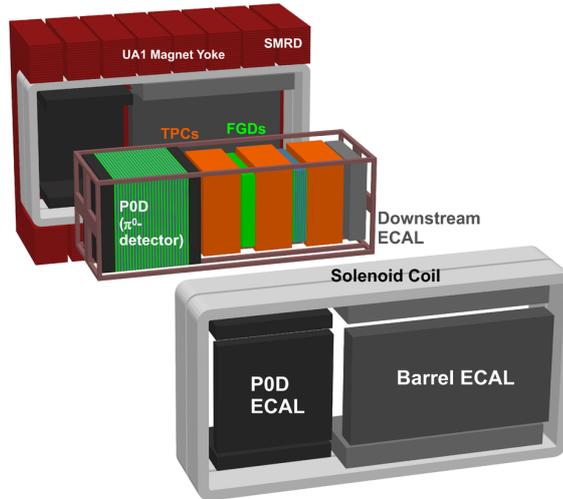
Figure 3.6. Schematic view of the T2K near detector ND280. The neutrino beam is injected from the left-hand side of the figure.

hollow polycarbonate sheets filled with water. Therefore, by measuring the neutrino interactions at both FGD1 and FGD2, we can measure the cross-sections on both the carbon and oxygen targets. This is important for constraining the cross-section systematic uncertainties at the far detector as our far detector SK is a water target.

TPC   Three Time Projection Chambers (TPC) play critical roles in the tracking of the charged particles produced from the neutrino interactions in FGDs. They are filled with a mixture of $Ar : CF_4 : iC_4H_{10}$ gasses (95%:3%:2%) [72], and a uniform electric field is applied in a horizontal direction aligned with the magnetic field direction. When charged particles pass through the TPC, they ionize the gas molecules and create electron-ion pairs along its trajectory. Ionized electrons drift toward the anode and they are detected by the detectors attached to the wall of the drift region (so-called Micromegas modules [73]). We can measure the charged particles' momentum from the trajectories' curvature. By combining the reconstructed momentum and energy loss along its trajectory, we can also perform the PID. Figure 3.7 shows the distribution of energy loss (dE/dx) as a function of the momentum for external muons crossing the detector or charged particles produced from the neutrino interactions in ND280. It illustrates that the TPC has a very good separation of muons and protons, while it is difficult to distinguish pions from muons as they have almost the same dE/dx curves. A few low-energy electrons are also observed but the separation efficiency becomes worse for them.

ECal   The Electromagnetic Calorimeters (ECal) are composed of lead and plastic scintillators and surround the tracking detectors (FGDs and TPCs) [74]. They provide additional PID information for the particles penetrating the TPCs in addition to the calorimetric energies for the electromagnetic showers caused by electrons and gammas.

(a) Negatively charged particles   (b) Positively charged particles
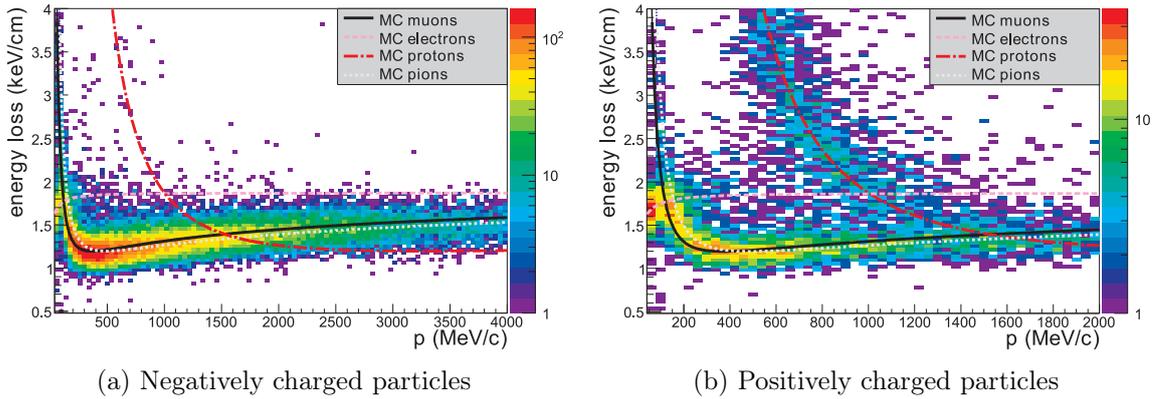
Figure 3.7. Distribution of the energy loss as a function of the momentum for the through-going muons and charged particles produced from the neutrino interactions in ND280. For comparison, expected curves for muons, electrons, protons, and pions are shown together. The figures are taken from Ref. [72].

P0D   The π^0 detector (P0D) is located at the upstream of the tracker detectors. It aims at observing neutral current interactions that contain π^0 in the final state. The central region, which is referred to as the water target region, is made from alternating scintillator planes, water bags, and brass sheets [75]. Brass sheets are inserted to increase the detector mass and to stop more gammas produced from the π^0 decay, which allows us to reconstruct the calorimetric energy of π^0 with a higher resolution.

SMRD   The Side Muon Range Detectors (SMRD) are placed around the magnet and used to detect the high-energy muon tracks penetrating through the tracking detectors and magnet [76].

3.3   Data set

3.3.1   Data set used in the analysis

In this analysis, we perform a simultaneous fit to both the SK atmospheric neutrinos and the T2K accelerator neutrinos. For the SK atmospheric neutrinos, 18 samples from the full SK-IV data are used, which corresponds to 3244.4 days of data taking [52]. The SK operating period can be broken down into seven phases (I-VII). The SK-IV phase is the longest one and contains roughly $\sim 46\%$ of the data. More details of the SK data set will be discussed in Section 9.1 including the differences between SK-IV and other phases.

The accumulated T2K data in terms of protons on target (POT) is shown in Fig. 3.8. In this analysis, we use five samples from the T2K Run 1-10 data corresponding to 1.97×10^{21} POT in FHC and 1.63×10^{21} POT in RHC [37].

The description of the samples used in the analysis is given in the following. The neutrino flavors can only be identified through the charged current (CC) interactions, which in general produces one charged lepton ($e^{\pm}, \mu^{\pm}$) and 0 or more pions. Therefore, the sample classification aims to classify events with similar event topologies (i.e. the
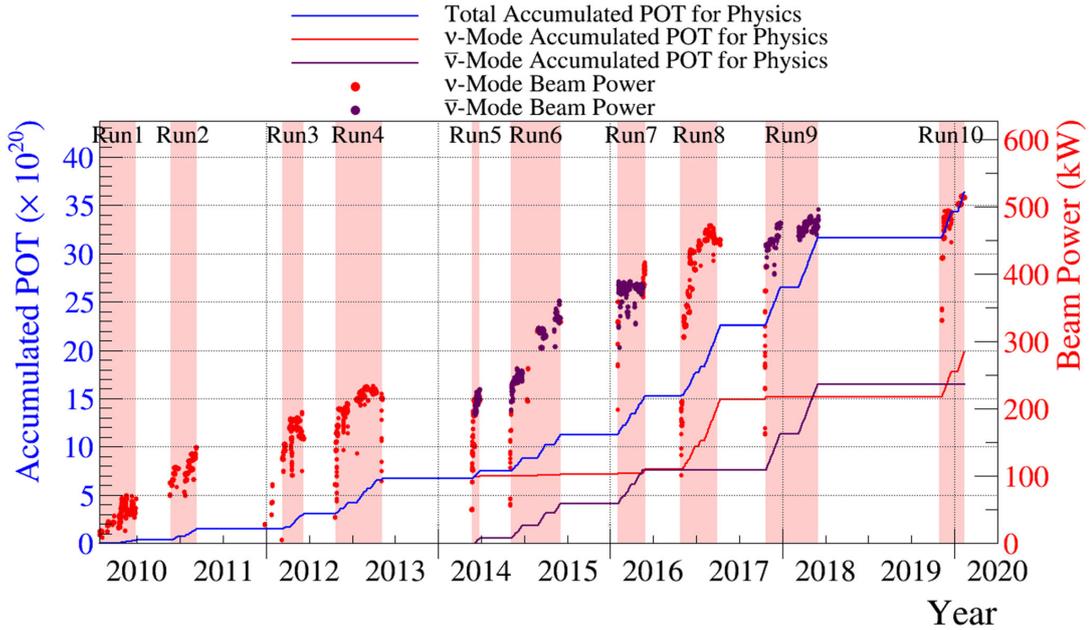
Figure 3.8. The accumulated T2K data in terms of the protons on target (POT). The history of the beam power is shown together with the red line. The figure is taken from Ref. [37].

lepton types, number of pions, etc) by using the PID and the number of reconstructed particles. In addition, the number of delayed signals is used to identify pions or muons that produce a decay electron (denoted as de or dcy) in their decay chains. The samples targeting the CC interactions with n pions are denoted as the CC$n\pi$ samples (e.g. CC0π, CC1π). The neutral current (NC) interactions can also be used to detect neutrinos, but these events are thought to be a background as they do not provide the flavor information of the neutrino. More detailed descriptions of the neutrino-nucleus interactions will be given in Section 4.3.

3.3.2 Atmospheric neutrino samples

In the SK atmospheric neutrino analysis, events are first classified into three categories: Fully Contained (FC), Partially Contained (PC), and Upward-going Muon (UpMu). When an event takes place inside the inner detector (ID) and leaves no Cherenkov light in the outer detector (OD), it is classified as FC. An event is classified as PC when charged particles escape the ID and are detected in the OD. To increase the statistics for the events coming from the opposite side of the Earth, we have an UpMu category which targets the neutrino interactions that happened in the rock below the detector with the produced muon track reaching the ID. These out-of-detector events are only selected for upward-going events to avoid the contamination of cosmic ray backgrounds. The schematic views of these three categories are shown in Fig. 3.9.

The event is then classified into 18 samples based on the number of rings, PID, and other event topologies, as summarized in Table 3.2. Further details of the sample classi-
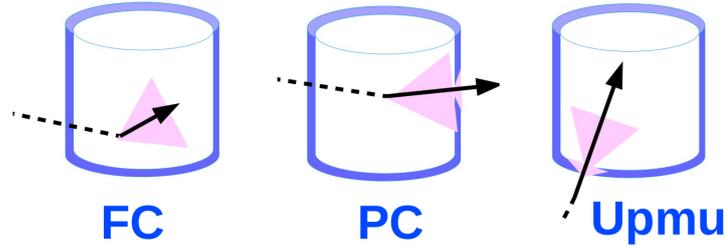
Figure 3.9. Three categories used in the SK atmospheric sample classification.

fication are given in the following.

Table 3.2. Summary of the atmospheric sample classification. See the text for more details on the classification.

Sample name	Category	Selection			
SubGeV elike 0de		Sub-GeV	Single ring	e-like	0 decay-e
SubGeV elike 1de					1 decay-e
SubGeV mulike 0de				μ-like	0 decay-e
SubGeV mulike 1de					1 decay-e
SubGeV mulike 2de					≥ 2 decay-e
SubGeV pi0like			Two rings	Two e-like rings and pass M_{inv} cut	
MultiGeV elike nue	Fully Contained (FC)	Multi-GeV	Single ring	e-like	≥ 1 decay-e
MultiGeV elike nuebar					0 decay-e
MultiGeV mulike				μ-like	
MultiRing elike nue			Multi rings	e-like	Pass MME likelihood cut and ν_e-like
MultiRing elike nuebar					Pass MME likelihood cut and $\bar{\nu}_e$-like
MultiRingOther					Fail MME likelihood cut
MultiRing mulike				μ-like	
PCStop	Partially Contained (PC)	Smaller charge deposition in outer detector			
PCThru		Larger charge deposition in outer detector			
UpStop mu	Up-going Muon (UpMu)	Stopping			
UpThruNonShower mu		Through-going non-showering			
UpThruShower mu		Through-going showering			

3.3.2.1 Fully Contained

The FC events are broken down into two groups depending on the visible energy of the events, sub-GeV ($E_{\mathrm{vis}} < 1330$ MeV) and multi-GeV ($E_{\mathrm{vis}} > 1330$ MeV), and further divided into two groups based on the number of rings (single-ring or multi-ring).

Sub-GeV single-ring   When a sub-GeV event has a single ring, it is classified into five samples depending on the PID of the primary ring and the number of decay electrons:

e-like 0 de and 1 de, and μ-like 0 de, 1 de, and ≥ 2 de samples. Since electrons do not have a delayed signal, e-like 0 de and e-like 1 de samples are targeting the CC0π and CC1π samples, respectively, where the delayed signal comes from the decay chain of pions. On the other hand, muons can also contribute to the delayed signals, and therefore, the μ-like 0 de and 1 de samples are dominated by the CC0π events while the μ-like 2 de sample is dominated by the CC1π events.

Sub-GeV two-rings   When the FC sub-GeV event has two primary rings, we apply a π^0 selection. When π^0 decays into two gamma rays, these gamma rays cause electromagnetic showers and are detected as e-like rings. Therefore, we require that both of the two rings should be e-like and the reconstructed invariant mass (M_{inv}) should be in the range $[85, 215]$ MeV (the π^0 mass of 135 MeV with some residuals). The invariant mass cut is necessary to reduce the coincidence background where we have two gamma (e-like) rings by chance and make sure the two gammas originated from a single π^0.

Multi-GeV single-ring   The FC multi-GeV events are classified into single-ring and multi-ring groups too. The single-ring events are classified into three samples depending on the PID and number of the decay electrons: e-like ν_e (≥ 1 de) and $\bar{\nu}_e$ (0 de), and μ-like samples. The event-basis separation of ν_e and $\bar{\nu}_e$ is difficult at SK as the detector is not magnetized. However, it is important to separate them because the ν_e and $\bar{\nu}_e$ event rates in the multi-GeV regions provide sensitivity for mass ordering. In general, the ν_e ($\bar{\nu}_e$) interaction produces π^+ (π^-), and π^- is more likely to be captured by oxygen nuclei than π^+, which produces no decay electron. Therefore, we use the number of decay electrons to separate the ν_e and $\bar{\nu}_e$ samples.

Multi-GeV multi-ring   For the FC multi-GeV multi-ring events, we use the PID of the most energetic ring to classify the e-like and μ-like events. Then, a special cut called the Multi-GeV Multi-ring E-like cut (MME) is applied to the e-like events to reduce further the CCν_μ and NC backgrounds[2]. Reducing these non-CCν_e backgrounds is also important for the mass ordering sensitivity.

The MME cut consists of the e-like/other separation and $\nu_e/\bar{\nu}_e$ separation. At the first stage, a dedicated likelihood ($\mathcal{L}_{e/\mathrm{other}}$) is constructed by calculating the probability density functions (p.d.f.) of four input parameters and taking the ratio of p.d.f. under the signal (CCν_e and CC$\bar{\nu}_e$) and background (CCν_μ and NC) hypotheses. The four input parameters used in the MME likelihood constructions are

1. Fraction of the most energetic e-like ring to the total visible energy

2. Fraction of the $\pi^\pm$-like ring to the total visible energy

3. Number of the decay electrons

4. Distance between the event vertex and the furthest decay electron vertex

---

[2]More details on the MME cut can be found in Refs. [52, 77].

The obtained e-like/other separation likelihood distributions for data and MC are shown in Fig. 3.10a. The events are classified into the multi-ring "other" sample when $\mathcal{L}_{e/\text{other}} < -0.25$, and are regarded as passed the cut otherwise.

For the events that have passed the first stage of the MME cut, an additional cut is applied to separate the ν_e-like and $\bar{\nu}_e$-like events. Here we construct another likelihood ($\mathcal{L}_{\nu_e/\bar{\nu}_e}$) using the number of decay electrons, the number of rings, and the transverse momentum of the most energetic ring. The constructed likelihood is shown in Fig. 3.10b. The events are classified as ν_e-like when the likelihood is positive and as $\bar{\nu}_e$-like otherwise.

(a) e-like and other separation   (b) ν_e and $\bar{\nu}_e$ separation
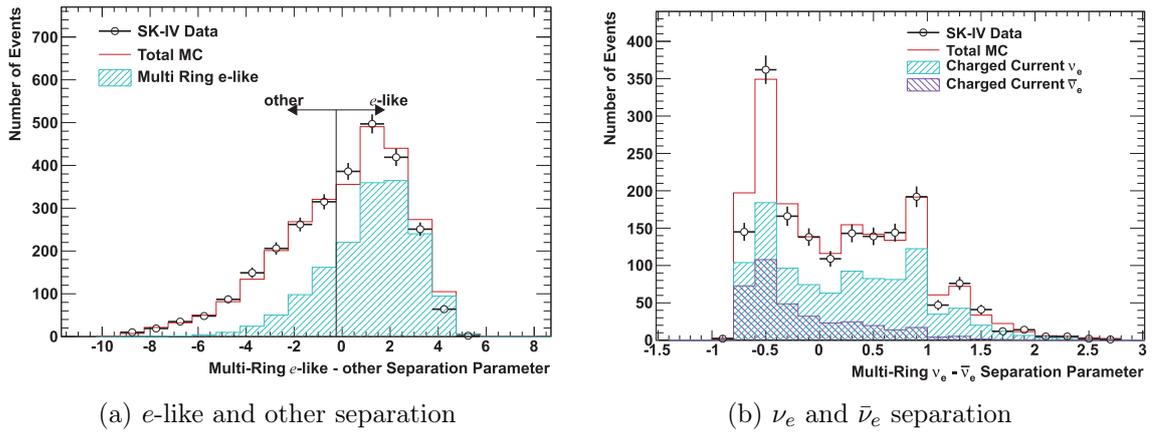
Figure 3.10. Likelihood distributions used in the atmospheric multi-GeV multi-ring e-like sample classification. The figures are taken from Ref. [36].

3.3.2.2   Partially Contained and Upward-going Muon

The PC events are classified into PC stop and PC through-going depending on the amount of energy deposition in the OD. The UpMu events are first classified into stopping and through-going, and through-going events are further divided into showering and non-showering samples. Some quality cuts are imposed on these samples to reduce the backgrounds, including the energy cut and track length cut. More detailed descriptions can be found in Refs. [78, 79].

The contributions of the true neutrino types in each sample for the simulated Monte Carlo data set are summarized in Table 3.3. The purities in the multi-GeV single-ring or multi-ring e-like $\bar{\nu}_e$ samples are still dominated by the ν_e events, but they contain a larger fraction of the $\bar{\nu}_e$ events than their counterpart ν_e samples.

3.3.2.3   Sample binning

We use the reconstructed lepton momentum and the zenith angle $\cos\Theta_z$ for the binning of each sample as shown in Fig. 3.11. The different momentum binnings are used depending on the samples. The number of the $\cos\Theta_z$ bins is fixed to 10 except for the FC sub-GeV e-like 1 de, μ-like 2 de, and π^0-like samples where a single bin covers the entire $\cos\Theta_z$

Table 3.3. Summary of the true neutrino type contributions in each atmospheric sample for the simulated MC data set. The opacity of the background color in each cell corresponds to the fraction size of the component.

Sample	ν_e	$\bar{\nu}_e$	ν_μ	$\bar{\nu}_\mu$	$\nu_\tau + \bar{\nu}_\tau$	NC
SubGeV elike 0de	0.724	0.230	0.009	0.003	0.002	0.032
SubGeV elike 1de	0.831	0.019	0.080	0.025	0.003	0.042
SubGeV mulike 0de	0.007	0.002	0.794	0.141	0.001	0.055
SubGeV mulike 1de	0.000	0.000	0.693	0.297	0.001	0.010
SubGeV mulike 2de	0.000	0.000	0.968	0.026	0.001	0.005
SubGeV pi0like	0.073	0.025	0.015	0.002	0.001	0.885
MultiGeV elike nue	0.713	0.073	0.053	0.005	0.076	0.081
MultiGeV elike nuebar	0.557	0.367	0.002	0.000	0.027	0.046
MultiGeV mulike	0.000	0.000	0.600	0.391	0.008	0.000
MultiRing elike nue	0.499	0.105	0.057	0.003	0.156	0.182
MultiRing elike nuebar	0.490	0.290	0.022	0.002	0.075	0.122
MultiRingOther	0.004	0.001	0.723	0.252	0.011	0.010
MultiRing mulike	0.195	0.016	0.319	0.027	0.222	0.224
PCStop	0.059	0.024	0.600	0.287	0.000	0.031
PCThru	0.005	0.002	0.644	0.348	0.000	0.004
UpStop mu	0.006	0.003	0.663	0.330	0.000	0.002
UpThruNonShower mu	0.006	0.002	0.676	0.323	0.000	0.001
UpThruShower mu	0.006	0.002	0.682	0.322	0.000	0.001

regions. A single momentum bin is used for the UpMu showering and non-showering samples.

For reference, the SK-IV data distributions overlaid with the best-fit MC predictions in the published SK atmospheric neutrino analysis are shown in Fig. 3.12. We should note that the data and MC are not exactly the same as what is used in this analysis. The difference between the published SK analysis and this analysis is detailed in Appendix F.

3.3.3 T2K beam neutrino far detector samples

The beam neutrino events at SK are classified into five samples depending on the neutrino mode (FHC and RHC), PID, and the number of decay electrons. The selection steps for these samples are summarized in Table 3.4 and described in more detail in the following.

Timing To select the neutrino events originating from the T2K neutrino beam, we first apply the timing cut requiring the event should take place within the neutrino beam bunch timing. Figure 3.13 shows the timing distributions of the fully contained events overlaid with the expected beam bunch timing. It shows that the events from the T2K neutrino beam data are distributed around the beam bunch timings. As a first step of
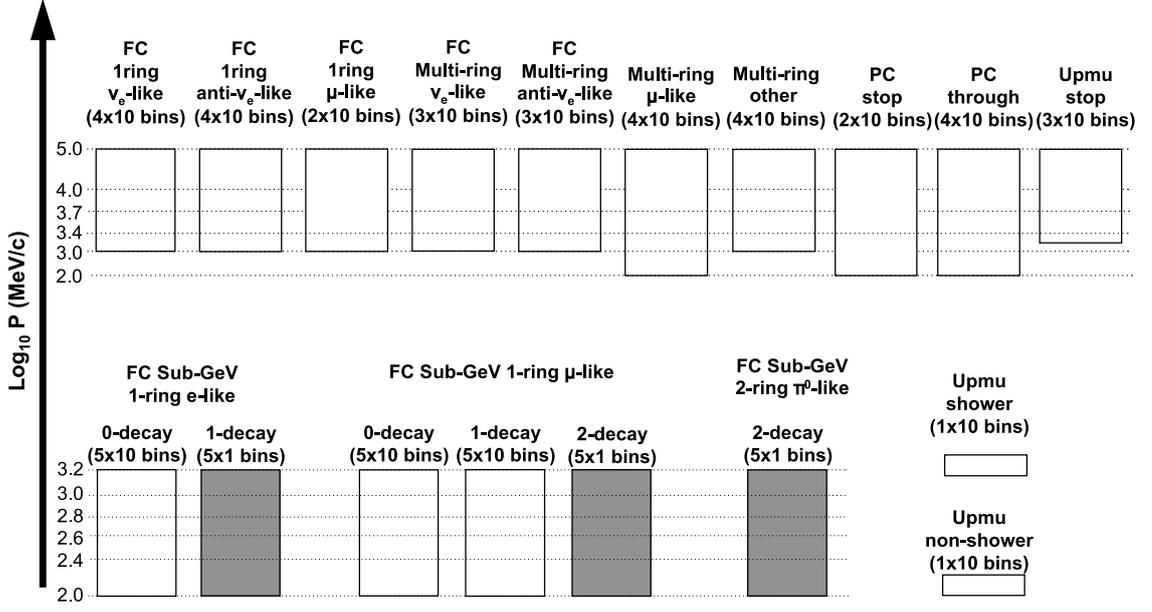
Figure 3.11. Summary of the sample binning for all the atmospheric samples. The color of the box corresponds to the $\cos\Theta_z$ binning: the black means a single $\cos\Theta_z$ bin covering $[-1, 1]$ and the white means 10 bins with an equal bin width for $\cos\Theta_z \in [-1, 1]$ ($\cos\Theta_z \in [-1, 0]$ for the UpMu samples). The vertical axis shows the momentum binning of each sample where the momentum is expressed in the logarithmic scale. The figure is taken from Ref. [78].

Table 3.4. Summary of the T2K beam sample classification.

Cut	FHC/RHC 1Rμ	FHC/RHC 1Re	FHC 1Re1de
Min. E_{vis}		$E_{\mathrm{vis}} > 30$ MeV	
Min. momentum	$p_\mu > 200$ MeV	$p_e > 100$ MeV	
Fiducial volume	$d_{\mathrm{wall}} > 50$ cm $d_{\mathrm{to\text{-}wall}} > 250$ cm	$d_{\mathrm{wall}} > 80$ cm $d_{\mathrm{to\text{-}wall}} > 170$ cm	$d_{\mathrm{wall}} > 50$ cm $d_{\mathrm{to\text{-}wall}} > 270$ cm
Decay electron	≤ 1	$= 0$	$= 1$
Max. E_{rec}	-	$E_{\mathrm{rec}} < 1.25$ GeV	
Additional cut	π^+ rejection	π^0 rejection	

the selection, fully contained events within these timing windows are selected.

Kinematics and fiducial volume We first require the event to have a single ring and visible energy (E_{vis}) larger than 30 MeV. The fiducial volume cut is defined using two parameters, d_{wall} (distance from the vertex to the nearest detector wall) and $d_{\mathrm{to\text{-}wall}}$ (distance
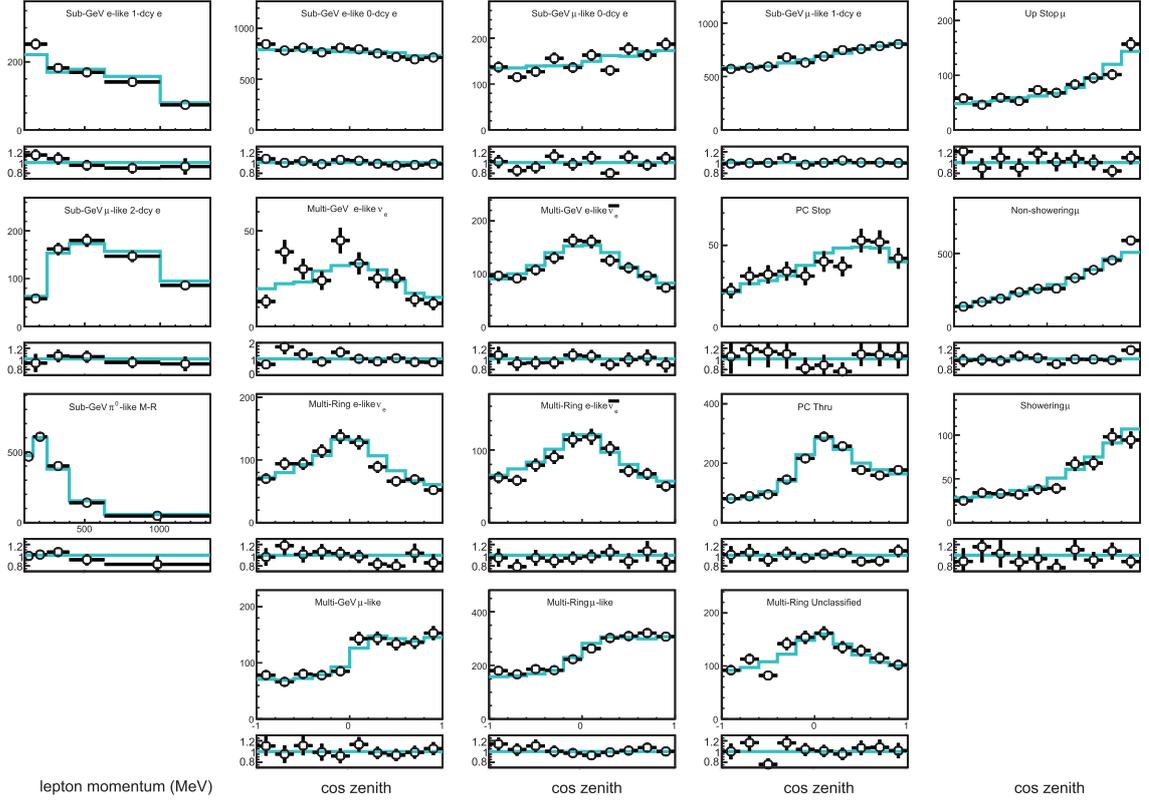
Figure 3.12. SK-IV atmospheric neutrino data distributions from the published SK IV analysis [36] overlaid with the data best-fit MC predictions. Samples with more than one cosine zenith bin are shown in the $\cos \Theta_z$ binning and the others are shown in the reconstructed momentum binning. These plots are only for reference, and the data and MC are not exactly the same as what is used in this analysis. The figures are taken from Ref. [36].

along the reconstructed direction to the nearest detector wall), with the threshold values optimized differently for each sample as summarized in Table 3.4. $d_{\text{to-wall}}$ is defined to be larger to make sure the charged particles travel enough distance to produce Cherenkov light before reaching the detector wall.

PID and event topology   The particle identification is then applied to separate e-like and μ-like samples, using the likelihood ratio ($\log(\mathcal{L}_e/\mathcal{L}_\mu)$) and reconstructed momentum under the e-like hypothesis (p_e) computed by the `fiTQun` single-ring fit. The distributions of the single-ring e/μ PID discriminators are shown in Fig. 3.14.

For the FHC e-like samples, an additional cut on the number of decay electrons is applied to separate CC0π and CC1π samples. For RHC e-like and FHC/RHC μ-like samples, we have no CC1π-dedicated samples. The μ-like events with 1 de are also included in the μ-like samples because muons can also produce decay electrons. Then, a cut on the reconstructed neutrino energy is applied to the e-like samples to reduce the π^0 background and intrinsic beam ν_e ($\bar{\nu}_e$) background which dominates the higher-
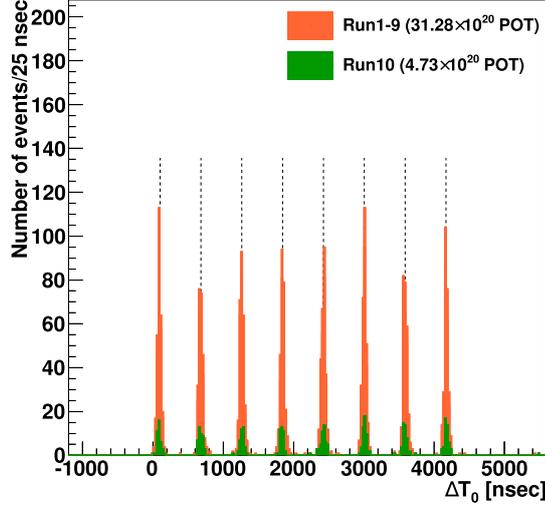
Figure 3.13. SK event timing distributions of the fully contained events collected during T2K run 1-9 and run 10. The horizontal axis shows the relative timing to the start of the spill at J-PARC. The beam bunch timing is overlaid with the dotted lines. The figure is taken from Ref. [37].
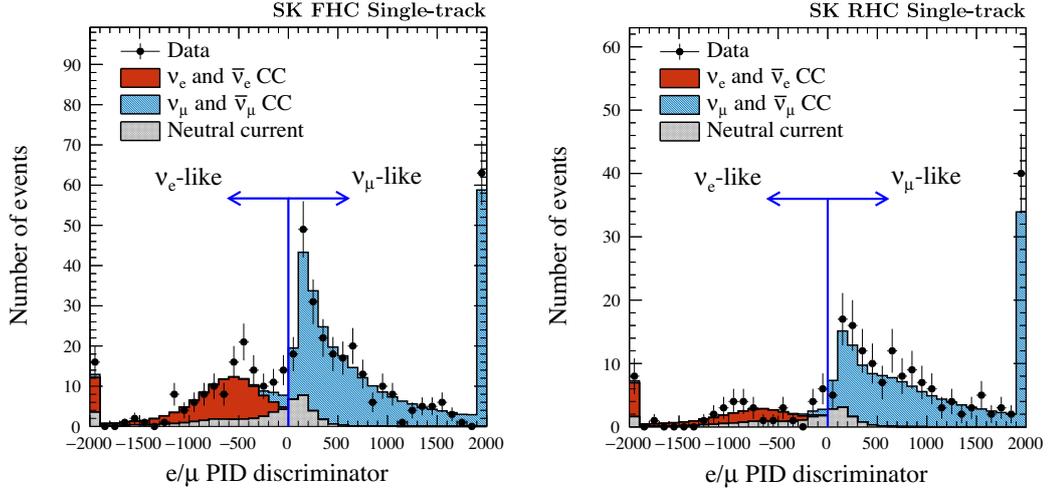
Figure 3.14. Distributions of the single-ring e/μ PID discriminators in FHC (left) and in RHC (right). The figures are taken from Ref. [80].

energy ranges (see Section 4.2.2). An additional π^0 cut is applied to the e-like samples to further reduce the π^0 background. This uses the likelihood ratio between the e-like and π^0-like hypotheses and the reconstructed invariant mass under π^0 hypothesis as shown in Fig. 3.15. For the μ-like samples, we apply the π^+ rejection cut instead of the π^0 cut, using the likelihood ratio between the μ-like and π^+-like hypothesis.

To summarize, we have two μ-like CC0π samples in FHC and RHC mode (FHC 1Rμ and RHC 1Rμ), two e-like CC0π samples (FHC 1Re and RHC 1Re), and one e-like CC1π sample only in FHC (FHC 1Re 1de). The selected events are binned into

histograms using the reconstructed kinematics such as the reconstructed neutrino energy ($E_{\rm rec}$), lepton momentum (p), and lepton direction with respect to the beam direction (θ). The $E_{\rm rec}$-θ binning is used for the μ-like samples, whereas the p-θ binning is used for the e-like samples, respectively.
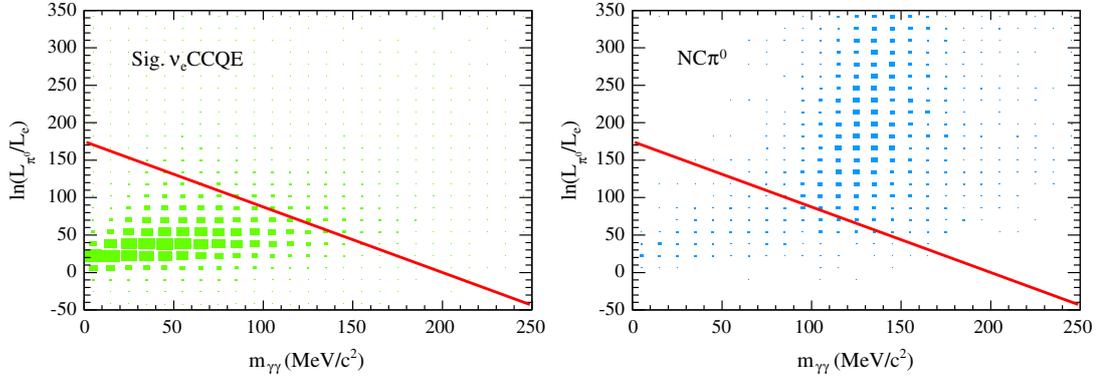
Figure 3.15. Distribution of the likelihood ratio between the e-like and π^0-like hypotheses and the reconstructed invariant mass under π^0 hypothesis for ν_e signal (left) and π^0 background (right). The cut threshold is shown as the red line. The figure is taken from Ref. [81].

For reference, the T2K Run 1-10 data distributions overlaid with the best-fit MC predictions in the published T2K analysis are shown in Fig. 3.16. We should note that the data and MC are not exactly the same as what is used in this analysis. The difference between the published T2K analysis and this analysis is detailed in Appendix F.

3.3.4   T2K beam neutrino near detector samples

In the T2K near detector analysis, we select events with the presence of a muon-like track produced inside the target detector FGD1 or FGD2. The muon track must enter one of the TPCs and must be identified as muon-like by the TPC PID. When there are several tracks entering TPCs, the muon PID is applied to the highest momentum track. The selected events are divided into 18 samples based on the beam horn current (FHC and RHC) and the number of observed pions (0π, 1π, and others). Events are classified into CC0π when they have no reconstructed pion, whereas the events with a single charged pion with opposite charge to the muon are classified into CC1π (i.e. π^+ in ν_μ selection and π^- in $\bar\nu_\mu$ selection). Events with any other number of charged pions or at least one neutral pion are classified as CC other. Figure 3.17 shows the schematic view of the event topologies in the ND280 samples.

Although the samples are statistically dominated with neutrinos in FHC mode and antineutrinos in RHC mode, there are non-negligible fractions of neutrinos in RHC mode (wrong-sign neutrinos). Therefore, in RHC mode, we have both the ν_μ and $\bar\nu_\mu$ samples to constrain the wrong-sign neutrino background. The events are categorized into 18 samples in total, as summarized in Table 3.5.

These events are binned into histograms using the momentum and angle ($\cos\theta$) with respect to the beam axis of the selected muon-like track.

(a) FHC 1Rμ

(b) RHC 1Rμ

(c) FHC 1Re

(d) RHC 1Re

(e) FHC 1Re 1de
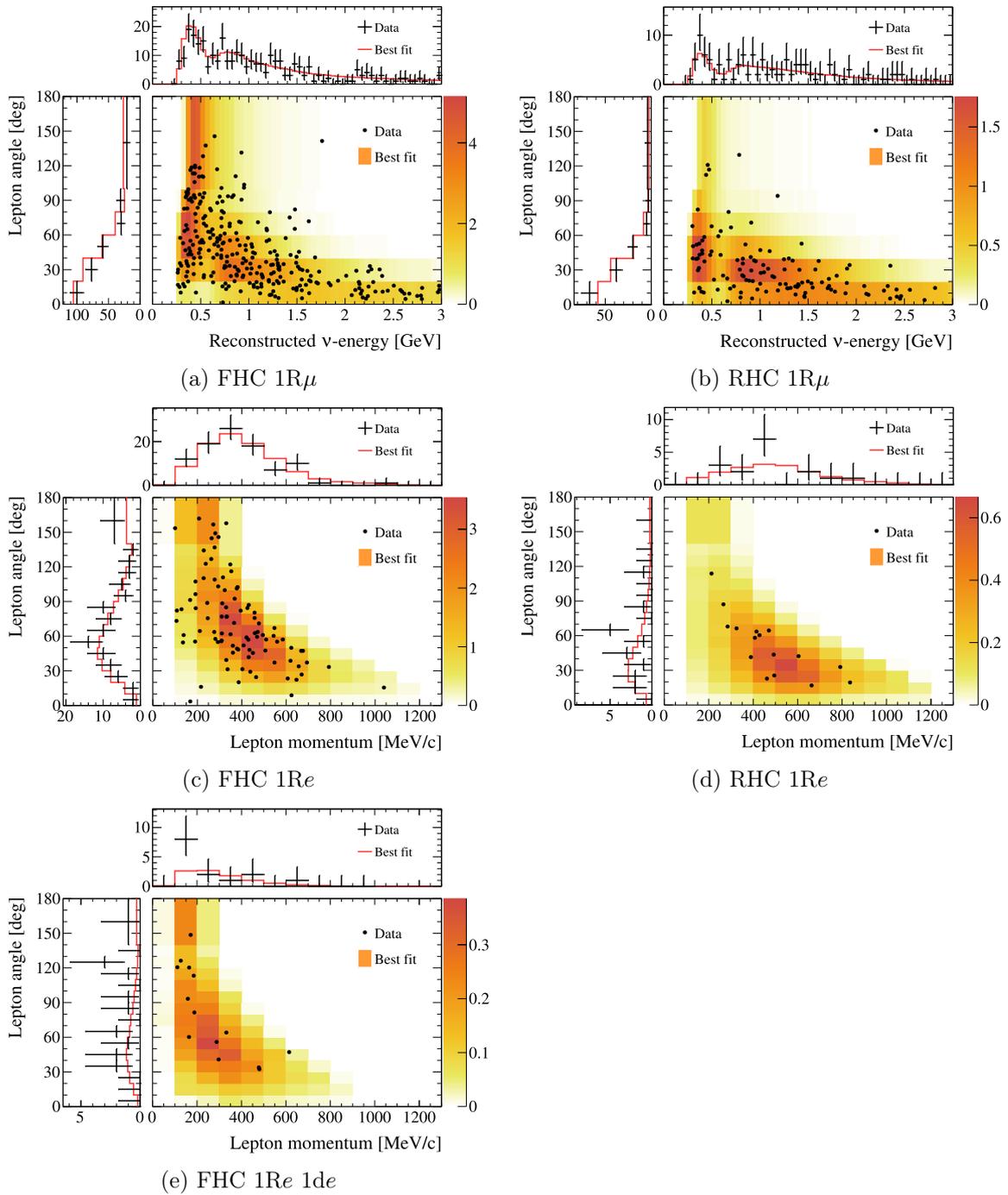
Figure 3.16. T2K Run 1-10 accelerator neutrino data distributions from the published T2K analysis [37] overlaid with the data best-fit MC predictions. The μ-like samples are shown in the E_{rec}-θ binning, and the e-like samples are shown in the p-θ binning. These plots are only for reference, and the data and MC are not exactly the same as what is used in this analysis. The figures are taken from Ref. [37].
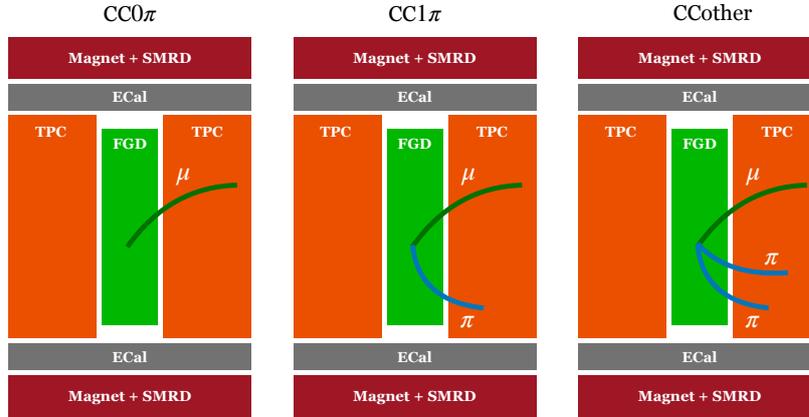
Figure 3.17. Schematic view of the event topologies in the ND280 samples. The samples are classified into 0π, 1π, and others depending on the number of observed pions.

Table 3.5. Summary of the near detector samples used in the near detector fit. The efficiency and purity of each sample selection are taken from Ref. [37].

Beam Mode	Neutrino Type	Topology	Target	Efficiency (%)	Purity (%)
FHC	ν_μ	0π	FGD1	48.0	71.3
			FGD2	48.0	68.2
		$1\pi^+$	FGD1	29.0	52.5
			FGD2	24.0	51.3
		Other	FGD1	30.0	71.4
			FGD2	30.0	71.2
RHC	$\bar{\nu}_\mu$	0π	FGD1	70.0	74.5
			FGD2	69.0	72.7
		$1\pi^-$	FGD1	19.3	45.4
			FGD2	17.2	41.0
		Other	FGD1	26.5	26.3
			FGD2	25.2	26.0
	ν_μ	0π	FGD1	60.3	55.9
			FGD2	60.3	52.8
		$1\pi^+$	FGD1	30.3	44.4
			FGD2	26.0	44.8
		Other	FGD1	27.4	68.3
			FGD2	27.1	69.5

Chapter 4

Monte Carlo simulation and systematic uncertainties

4.1 Overview of the Monte-Carlo method

The number of observed events in the neutrino oscillation experiments can be written as the product of neutrino flux Φ, neutrino-nucleus cross-section σ, number of target nuclei N_{target}, detection efficiency ε, and oscillation probability P_{osc}:

$$N_{\text{obs}}(x) = \Phi(x) \times \sigma(x) \times N_{\text{target}} \times \varepsilon(x) \times P_{\text{osc}}(x), \tag{4.1}$$

where x represents the properties of the events, such as true neutrino types and energies, interaction processes and event topologies, and the kinematics of the particles produced in the neutrino interaction. In the analysis, we measure the number of neutrino events at the detector and constrain the oscillation parameters using the above relation.

Due to the complexities of the models of neutrino flux, cross-sections, detector responses, and oscillation probabilities, it is difficult to analytically solve Eq. (4.1). Therefore, we use the Monte Carlo (MC) simulation to obtain inferences for the parameters of interest.

We first generate a large number of MC events that can cover all the possible combinations of the true and reconstructed properties of the neutrino interaction events at the detector. The true properties of the events include the neutrino types (neutrino or antineutrino) and flavors at both the production and interaction (which can be different due to the oscillation), the energy and direction of the neutrinos, and the kinematics of the outgoing particles from the neutrino interaction. The reconstructed characteristics include the observed event topologies (i.e. the number and type of particles produced in the neutrino interaction) and the reconstructed kinematics of the produced particles. The important point at the MC generation stage is to cover all the possible event properties with enough statistics, rather than producing a realistic sample of the "true" distribution where the actual data is produced from.

Then, the generated MC events are given an overall scaling and event-by-event weights. The overall scaling is determined simply from the amount of generated MC and data we obtained. The event-by-event weights change the composition of the events in the MC data set so that the resulting distribution represents the effect of the oscillation and systematic

uncertainties. To obtain the event-by-event weights, we need to construct a model that is parametrized with the oscillation parameters and the systematics uncertainty parameters that correspond to each source of systematic uncertainty. The model is defined as the response of the event-by-event weights as a function of the oscillation and the systematic uncertainty parameters (referred to as a *response function*). The weight of event i can therefore be written as

$$w_i(\boldsymbol{o}, \boldsymbol{s}) = f_i^{\mathrm{osc}}(\boldsymbol{o}) \times f_i^{\mathrm{syst}}(\boldsymbol{s}) \tag{4.2}$$

$$= f_i^{\mathrm{osc}}(\boldsymbol{o}) \times \prod_{j=1}^{n_s} f_{i,j}^{\mathrm{syst}}(s_j), \tag{4.3}$$

where f_i^{osc} and f_i^{syst} represent the response functions for event i as a function of the oscillation ($\boldsymbol{o}$) and systematic uncertainty ($\boldsymbol{s}$) parameters, respectively. We generally assume that the systematic uncertainty response functions can be expressed as the product of the response function of each parameter s_j for $j = 1, \cdots, n_s$, where n_s denotes the number of systematic uncertainty parameters. Then, the event count in a certain reconstructed kinematic bin b can be written as

$$n_b(\boldsymbol{o}, \boldsymbol{s}) = \alpha \times \sum_i^N I_{i,b} \times w_i(\boldsymbol{o}, \boldsymbol{s}) \tag{4.4}$$

$$= \alpha \times \sum_i^N \left[I_{i,b} \times f_i^{\mathrm{osc}}(\boldsymbol{o}) \times \prod_j^{n_s} f_{i,j}^{\mathrm{syst}}(s_j) \right], \tag{4.5}$$

where α denotes the overall scaling constant of MC, N denotes the number of simulated MC events, and $I_{i,b}$ is an indicator function that returns 1 when the event i falls into the bin b and takes 0 otherwise.

If one assumes an arbitrary set of oscillation and systematic uncertainty parameter values, Eq. (4.4) will give us binned event spectra. The set of parameter values is sometimes called the *tuning*. A parameter set tuned to our best prior understanding of the physics and other systematic sources is referred to as the *nominal* tuning, and the MC data set weighted to the nominal tuning is called the nominal MC data set.

The systematic uncertainty parameters that we use in the neutrino oscillation analysis can be grouped into three categories: neutrino flux, neutrino-nucleus interaction cross-sections, and detector responses. The systematic models used to produce the systematic response functions and the nominal MC are described in the following.

4.2   Neutrino flux simulation

4.2.1   Atmospheric neutrino flux

Nominal MC simulation   In the SK atmospheric neutrino analysis, we use the neutrino flux calculated by Honda et al. [82]. The Honda model uses the primary cosmic ray model based on the balloon [62] and the international space station experiments [61] with the corrections from the solar activity and geomagnetic field described in Section 3.1.1.

Then, the neutrino flux is calculated by simulating the interactions between the primary cosmic rays and nucleons in the atmosphere. The hadronic interaction models JAM [83] and DPMJET-III [84] are used for the cosmic ray energies below and above 32 GeV, respectively. The properties of the atmosphere, such as the molecular compositions and densities, are taken from the U.S.-standard '76 model [85].

The predicted atmospheric neutrino flux from the Honda model is shown in Fig. 4.1 with comparisons to the predictions from other alternative models, Bartol [86] and FLUKA [87]. The flux has a peak around a few hundred MeV and slowly decreases in the higher energy regions. The ratio of the muon and electron neutrinos is roughly ≈ 2 up to a few GeV, but it becomes larger at higher energies due to the suppressed electron neutrino flux.
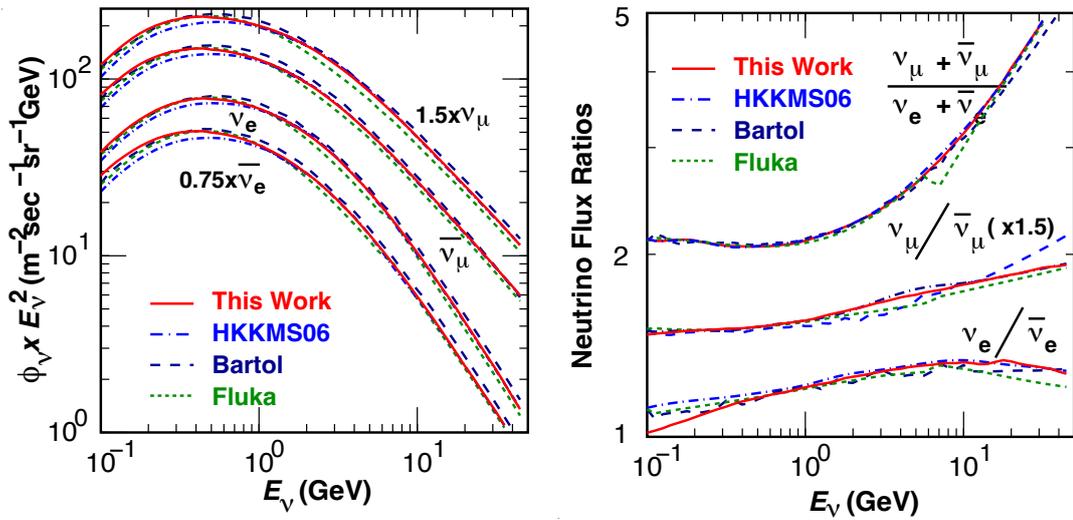
Figure 4.1. Predicted atmospheric neutrino flux by Honda et al. (denoted as "This Work") [82] in comparison to other predictions by the Bartol group [86], the FLUKA group [87], and the previous work by Honda et al. (denoted as "HKKMS06") [88]. The left plot shows the flux of each neutrino type calculated for Kamioka by averaging over all directions, and the right plot shows the flux ratios. The figures are taken from Ref. [82].

Systematic uncertainties The flux systematic uncertainties are introduced to cover the possible flux differences in the data and MC predictions. The size of the uncertainties is estimated by comparing the Honda model with the predictions from Bartol and FLUKA.

The systematic model consists of 18 parameters in total as summarized in Table 4.1. First of all, two overall energy-dependent normalization uncertainties are introduced separately for the neutrino energies below and above 1 GeV which account for the uncertainties on hadron productions, molecule densities in the atmosphere, and hadron interactions. On top of that, relative normalization uncertainties from comparisons of the three models are applied to the higher energy samples (FC multi-GeV, PC, and stopping UpMu) as the differences among the models become larger above 10 GeV. The uncertainties on the neutrino flavor ratios $\nu/\bar{\nu}$, $\nu_\mu/\bar{\nu}_\mu$, and $\nu_e/\bar{\nu}_e$ are also included. These flavor ratio uncertainties are divided into three energy regions ($E_\nu < 1$ GeV, $1 < E_\nu < 10$ GeV, and $E_\nu > 10$ GeV)

to take into account the energy dependence. The 1σ uncertainties of the overall normalization and the flavor ratio normalizations are shown in Fig. 4.2. We should note that for the flavor ratio systematics, a normalization of $1 + \sigma/2$ is applied to the neutrino types in the numerator, and a normalization of $1 - \sigma/2$ is applied to the neutrino types in the denominator.
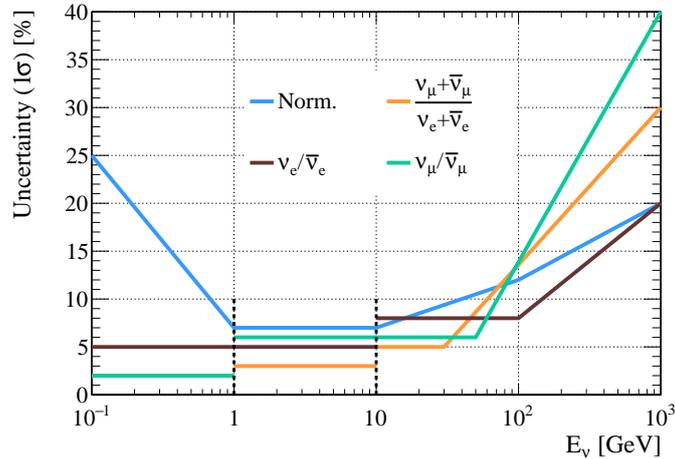
Figure 4.2. The size of atmospheric flux normalization uncertainties as a function of the neutrino energy.

The uncertainty on the geomagnetic field effect that can change the low-energy neutrino flux is included as the up/down ratio uncertainty. When neutrinos fly horizontally, secondary hadrons and muons travel longer through the atmosphere than when neutrinos fly into the atmosphere along a direction normal to the earth. Therefore, an additional horizontal/vertical ratio uncertainty is applied to cover the uncertainty on the horizontal flux where this effect can change the predicted flux. Since the contribution of kaon decays becomes important above a few GeV, the uncertainty on the parent hadron types (K/π ratio) is included. Additional uncertainties are calculated from the neutrino path length taking into account the uncertainty in the atmospheric density profile, and the periodic flux fluctuations due to the solar activity.

4.2.2 T2K accelerator neutrino flux

Nominal MC simulation The simulation of the T2K neutrino beam properties consists of two steps. The first step is the hadronic interaction between the accelerated protons and the target, which produces the secondary hadrons such as pions and kaons. It is simulated using FLUKA 2011.2x [89, 90]. In the FLUKA simulation, the proton beam with measured position and divergence properties is injected into the simulated target material, and the secondary hadrons produced in the interaction are recorded.

The second step is the transportation of the secondary hadrons and the production of the neutrinos in their decays, which is taken care of by the `GEANT3`-based JNUBEAM simulation [91]. In JNUBEAM, the particles are propagated from the production point determined by FLUKA to the beam dump until they decay or interact.

Table 4.1. Summary of the SK atmospheric flux uncertainties. The values are taken from Ref. [36].

Systematics	Sample		σ (%)
Flux normalization	$E_\nu < 1$ GeV		7-25
	$E_\nu > 1$ GeV		7-20
$(\nu_\mu + \bar{\nu}_\mu)/(\nu_e + \bar{\nu}_e)$	$E_\nu < 1$ GeV		2
	$1 < E_\nu < 10$ GeV		3
	$E_\nu > 10$ GeV		5-30
$\bar{\nu}_e/\nu_e$	$E_\nu < 1$ GeV		5
	$1 < E_\nu < 10$ GeV		5
	$E_\nu > 10$ GeV		8-20
$\bar{\nu}_\mu/\nu_\mu$	$E_\nu < 1$ GeV		2
	$1 < E_\nu < 10$ GeV		6
	$E_\nu > 10$ GeV		6-40
Up/down ratio	< 400 MeV	e-like	0.1
		μ-like	0.3
		0-decay μ-like	1.1
	> 400 MeV	e-like	0.8
		μ-like	0.5
		0-decay μ-like	1.7
	Multi-GeV	e-like	0.7
		μ-like	0.2
	Multi-ring Sub-GeV	e-like	0.4
		μ-like	0.2
	Multi-ring Multi-GeV	e-like	0.3
		μ-like	0.2
	PC		0.2
Horizontal/vertical ratio	< 400 MeV	e-like	0.1
		μ-like	0.1
		0-decay μ-like	0.3
	> 400 MeV	e-like	1.4
		μ-like	1.9
		0-decay μ-like	1.4
	Multi-GeV	e-like	3.2
		μ-like	2.3
	Multi-ring Sub-GeV	e-like	1.4
		μ-like	1.3
	Multi-ring Multi-GeV	e-like	2.8
		μ-like	1.5
	PC		1.7
K/π ratio in flux calculation	All		10
Neutrino path length	All		10
Relative normalization	FC Multi-GeV		5
	PC + Stopping Up-μ		5
Solar activity	All		7

The simulated flux is then reweighted to match the data taken by NA61/SHINE[1] experiment at CERN. The NA61/SHINE experiment uses the T2K replica target and measures the $\pi^\pm$, $K^\pm$, and ptoron yields [92, 93].

Figure 4.3 shows the simulated neutrino flux at SK for FHC and RHC mode, tuned to the NA61/SHINE data. The flux has a narrow peak around 0.6 GeV, where the $\nu_\mu \to \nu_e$ ($\bar{\nu}_\mu \to \bar{\nu}_e$) appearance probability becomes roughly the maximum at the location of SK (i.e. baseline of 295 km). The wrong-sign neutrinos ($\bar{\nu}$ for FHC and ν for RHC) and electron neutrinos are also taken into account in the simulation. The dominant contributions to the electron neutrino background are from the kaon and muon decays:

$$K^\pm \to \pi^0 + e^\pm + \nu_e(\bar{\nu}_e), \tag{4.6}$$

$$\mu^\pm \to e^\pm + \nu_e(\bar{\nu}_e) + \bar{\nu}_\mu(\nu_\mu). \tag{4.7}$$

The $\nu_e/\bar{\nu}_e$ flux is roughly $\sim 1\%$ of the total flux, but the fraction becomes larger at the higher energy regions. This is because $\nu_e/\bar{\nu}_e$ from the kaon decays have slightly higher energy peaks. These intrinsic electron neutrinos will be the dominant background in the $\nu_\mu \to \nu_e$ ($\bar{\nu}_\mu \to \bar{\nu}_e$) appearance measurement at the high energy.

(a) FHC

(b) RHC
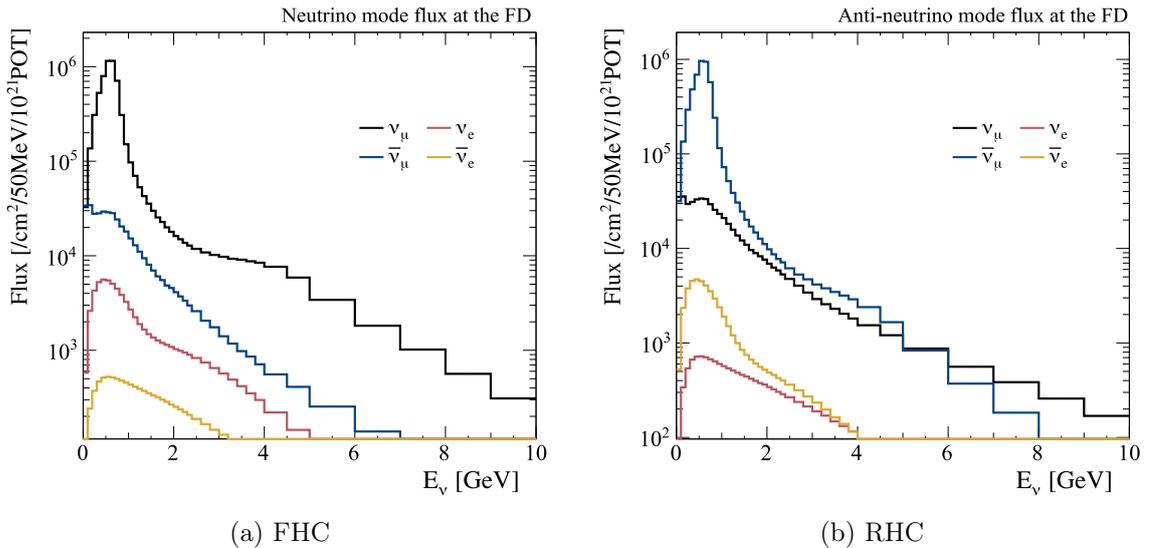
Figure 4.3. The predicted unoscillated neutrino flux at the far detector for FHC and RHC modes. The predictions are tuned to the NA61/SHINE measurements using the T2K replica target. The figures are taken from Ref. [37].

Systematic uncertainties For the flux systematic uncertainties, the uncertainties on the NA61/SHINE measurements are used along with the other uncertainty sources outside the target, such as the proton beam profiles and reinteraction of the hadrons. Figure 4.4 shows the flux uncertainties on the FHC/RHC ν_μ and RHC $\bar{\nu}_\mu$ as a function of the neutrino energy. Overall, the uncertainties from the hadron interactions have the largest contributions, but they are relatively well-constrained around the flux peak. The uncertainties

---

[1] SPS Heavy Ion and Neutrino Experiment

on the proton beam profile and the off-axis angle have the largest contributions near the upper edge of the flux peak. These systematic uncertainties on the flux are combined and implemented as the normalization of the event rates. A normalization parameter is assigned to each sub-divided energy region separately for the FHC and RHC modes, yielding 50 parameters in total. The beam flux systematic uncertainties are summarized in Table 4.2 with the constraints before and after the near detector fit.

(a) FHC ν_μ

(b) RHC $\bar{\nu}_\mu$

(c) RHC ν_μ
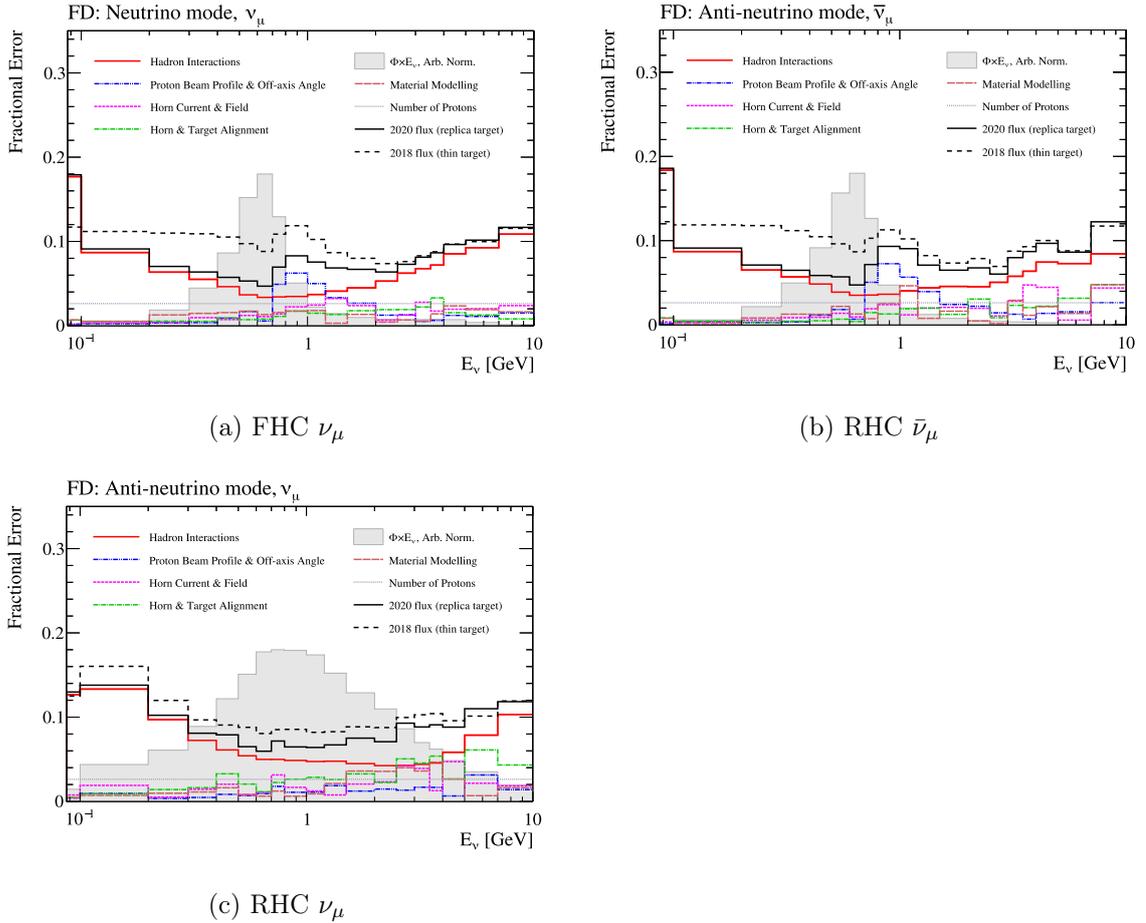
Figure 4.4. The uncertainties on the FHC/RHC ν_μ and RHC $\bar{\nu}_\mu$ fluxes as a function of the neutrino energy. The black solid line shows the total flux uncertainty for the latest T2K analysis tuned to the NA61/SHINE T2K replica target data [92]. The black dashed line shows the total uncertainty for the previous T2K analysis tuned to the thin target data [94] for comparison. The solid lines in other colors show the contribution of each systematic uncertainty source. The figures are taken from Ref. [37].

Table 4.2. Summary of the T2K beam flux systematic uncertainties. The constraints on each parameter before and after the near detector (ND) fit are also shown.

Horn current	Neutrino type	E_ν (GeV)	Pre ND fit	Post ND fit	Horn current	Neutrino type	E_ν (GeV)	Pre ND fit	Post ND fit
		$0.0 - 0.4$	1.00 ± 0.07	1.11 ± 0.05			$0.0 - 0.7$	1.00 ± 0.09	1.11 ± 0.06
		$0.4 - 0.5$	1.00 ± 0.06	1.10 ± 0.04			$0.7 - 1.0$	1.00 ± 0.06	1.07 ± 0.05
		$0.5 - 0.6$	1.00 ± 0.05	1.08 ± 0.03		ν_μ	$1.0 - 1.5$	1.00 ± 0.06	1.07 ± 0.04
		$0.6 - 0.7$	1.00 ± 0.05	1.07 ± 0.03			$1.5 - 2.5$	1.00 ± 0.07	1.07 ± 0.04
		$0.7 - 1.0$	1.00 ± 0.07	1.06 ± 0.04			$2.5 - 30.0$	1.00 ± 0.07	1.02 ± 0.04
	ν_μ	$1.0 - 1.5$	1.00 ± 0.07	1.03 ± 0.04			$0.0 - 0.4$	1.00 ± 0.07	1.09 ± 0.05
		$1.5 - 2.5$	1.00 ± 0.06	1.03 ± 0.04			$0.4 - 0.5$	1.00 ± 0.06	1.09 ± 0.04
		$2.5 - 3.5$	1.00 ± 0.07	1.01 ± 0.04			$0.5 - 0.6$	1.00 ± 0.06	1.07 ± 0.04
		$3.5 - 5.0$	1.00 ± 0.09	0.98 ± 0.04			$0.6 - 0.7$	1.00 ± 0.05	1.06 ± 0.03
		$5.0 - 7.0$	1.00 ± 0.10	0.92 ± 0.04			$0.7 - 1.0$	1.00 ± 0.08	1.09 ± 0.04
		$7.0 - 30.0$	1.00 ± 0.12	0.91 ± 0.04		$\bar\nu_\mu$	$1.0 - 1.5$	1.00 ± 0.08	1.06 ± 0.04
		$0.0 - 0.7$	1.00 ± 0.09	1.06 ± 0.08			$1.5 - 2.5$	1.00 ± 0.06	1.01 ± 0.04
FHC		$0.7 - 1.0$	1.00 ± 0.06	1.04 ± 0.05	RHC		$2.5 - 3.5$	1.00 ± 0.07	1.01 ± 0.05
	$\bar\nu_\mu$	$1.0 - 1.5$	1.00 ± 0.07	1.04 ± 0.06			$3.5 - 5.0$	1.00 ± 0.09	0.95 ± 0.06
		$1.5 - 2.5$	1.00 ± 0.08	1.05 ± 0.07			$5.0 - 7.0$	1.00 ± 0.09	0.96 ± 0.06
		$2.5 - 30.0$	1.00 ± 0.08	1.04 ± 0.06			$7.0 - 30.0$	1.00 ± 0.12	0.93 ± 0.09
		$0.0 - 0.5$	1.00 ± 0.06	1.09 ± 0.04		ν_e	$0.0 - 2.5$	1.00 ± 0.09	1.03 ± 0.07
		$0.5 - 0.7$	1.00 ± 0.05	1.08 ± 0.04			$2.5 - 30.0$	1.00 ± 0.08	1.03 ± 0.07
		$0.7 - 0.8$	1.00 ± 0.05	1.06 ± 0.04			$0.0 - 0.5$	1.00 ± 0.06	1.08 ± 0.04
	ν_e	$0.8 - 1.5$	1.00 ± 0.06	1.04 ± 0.04			$0.5 - 0.7$	1.00 ± 0.05	1.07 ± 0.04
		$1.5 - 2.5$	1.00 ± 0.08	1.00 ± 0.04			$0.7 - 0.8$	1.00 ± 0.06	1.06 ± 0.04
		$2.5 - 4.0$	1.00 ± 0.09	0.98 ± 0.04		$\bar\nu_e$	$0.8 - 1.5$	1.00 ± 0.06	1.04 ± 0.04
		$4.0 - 30.0$	1.00 ± 0.09	0.98 ± 0.05			$1.5 - 2.5$	1.00 ± 0.08	1.01 ± 0.06
	$\bar\nu_e$	$0.0 - 2.5$	1.00 ± 0.10	1.02 ± 0.09			$2.5 - 4.0$	1.00 ± 0.09	1.01 ± 0.07
		$2.5 - 30.0$	1.00 ± 0.13	1.09 ± 0.11			$4.0 - 30.0$	1.00 ± 0.15	1.09 ± 0.13

4.3 Neutrino-nucleus interaction

There are various types of interaction processes depending on the neutrino energy. Figure 4.5 shows the contributions of several interaction processes to the inclusive cross-section in the typical energy ranges of accelerator neutrinos and atmospheric neutrinos. These cross-sections are calculated by NEUT, which is a neutrino-nucleus interaction simulation program library used in the T2K and SK analyses [95, 96].

In the energy range below a few GeV, which is relevant to the T2K accelerator neutrinos and low-energy part of the SK atmospheric neutrinos, the dominant contribution comes from the charged current quasi-elastic (CCQE) interaction that has one charged lepton and one nucleon in the final state. The cross-section of the charged current resonant 1π production process increases above 0.5 GeV and becomes dominant above 1.2 GeV. There are also small but non-negligible contributions of charged current two-particle two-hole (CC 2p2h) interaction where two nucleons in the nucleus are involved in the process. The atmospheric neutrinos have wider energy ranges compared to that of T2K, and the multi-pion process and the charged current deep inelastic scattering process dominate the interaction in the high-energy region. The details of these processes are described in the
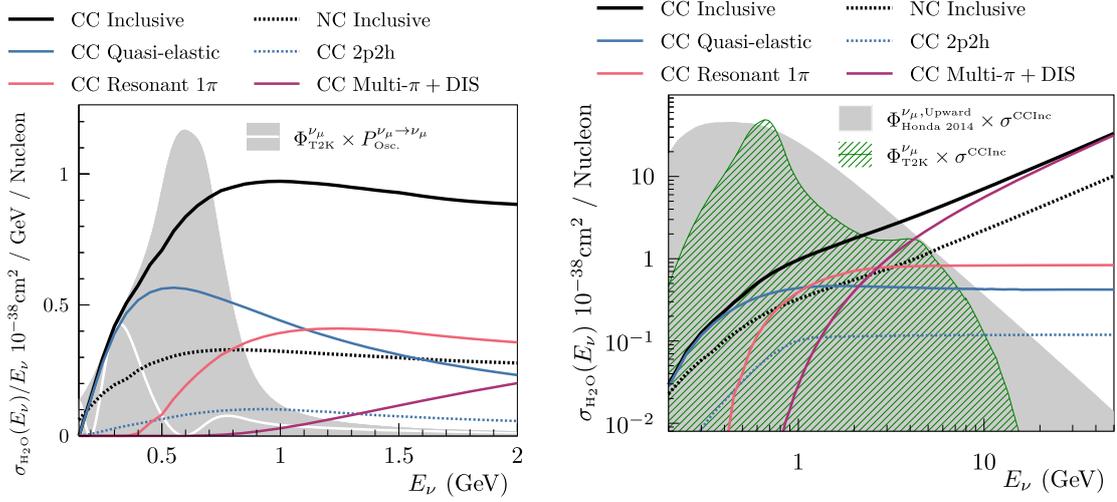
Figure 4.5. The NEUT-predicted ν_μ–water cross sections overlaid on the T2K muon neutrino flux [91] (left) and upward-going atmospheric ν_μ flux [97] (right). CC inclusive (black solid line) and NC inclusive (black dotted line) are the total cross-sections for the charged-current and neutral-current processes, respectively. The CC cross-sections are further broken down into the quasi-elastic, two-particle two-hole (2p2h), resonant 1π production, and multi-pion + deep inelastic (DIS) interactions. It should be noted that the cross-sections on the left panel are divided by the neutrino energy but not on the right panel. The predicted distributions of the T2K ν_μ flux and atmospheric ν_μ flux are overlaid to get an idea of the most relevant energy ranges for each experiment. The figures are taken from Ref. [98].

following.

4.3.1 Nominal MC simulation

4.3.1.1 Neutrino-nucleon interaction

In this subsection, various interaction processes between a neutrino and a single nucleon are detailed. The effects coming from having multiple nucleons inside a nucleus will be discussed in Section 4.3.1.2.

Quasi-elastic scattering The charged current quasi-elastic (CCQE) scattering is a two-body process between a neutrino and a nucleon. This is a dominant process in the regions where neutrino energy is below a few GeV, which is relevant to the T2K beam and atmospheric sub-GeV samples. It produces a charged lepton and a nucleon in the final state

$$\nu_l + n \rightarrow l^- + p \tag{4.8}$$
$$\overline{\nu}_l + p \rightarrow l^+ + n. \tag{4.9}$$

This process is mediated by the charged $W^\pm$ bosons as shown in the Feynman diagram in Fig. 4.6.
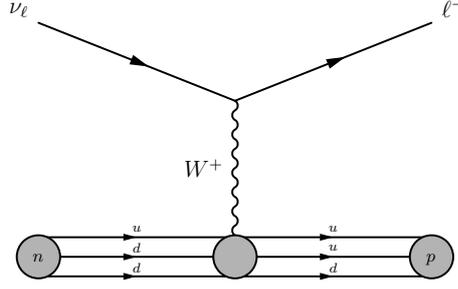
Figure 4.6. Feynman diagram of the charged current quasi-elastic (CCQE) process.

Since this is a pure two-body process, the initial neutrino energy can be reconstructed from the final-state lepton kinematics. Assuming that the nucleon is at rest in the initial state and applying the conservation of energy and momentum, the reconstructed neutrino energy is computed as

$$E_\nu = \frac{2m_n E_l - m_l^2 + (m_p^2 - m_n^2)}{2(m_n - E_l + p_l \cos\theta)},$$ (4.10)

where E_l, p_l, and m_l are the energy, momentum, and mass of the outgoing lepton. m_p and m_n are the mass of the proton and neutron, respectively. θ is the scattering angle of the lepton with respect to the incoming neutrino direction.

The differential scattering cross-section can be expressed, in terms of Mandelstam variables $s = (p_\nu + p_n)^2$, $t = (p_\nu - p_\ell)^2 = q^2 = -Q^2$, $u = (p_\nu - p_p)^2$ and nucleon mass M, as [99]

$$\frac{d\sigma}{dQ^2}\left(\begin{matrix}\nu n \to \ell^- p \\ \bar\nu n \to \ell^+ p\end{matrix}\right) = \frac{M^2 G_F^2 \cos^2\theta_C}{8\pi E_\nu^2}\left(A(Q^2) \mp B(Q^2)\frac{s-u}{M^2} + C(Q^2)\frac{(s-u)^2}{M^4}\right).$$ (4.11)

The factors $A(Q^2)$, $B(Q^2)$, and $C(Q^2)$ are given as

$$A(Q^2) = \frac{m_\ell^2 + Q^2}{M^2}\left[\left(1 + \frac{Q^2}{4M^2}\right)F_A^2(Q^2) - \left(1 - \frac{Q^2}{4M^2}\right)F_1^2(Q^2)\right.$$
$$+ \frac{Q^2}{4M^2}\left(1 - \frac{Q^2}{4M^2}\right)F_2^2(Q^2) + \frac{Q^2}{M^2}F_1(Q^2)F_2(Q^2)$$
$$\left. - \frac{m_\ell^2}{4M^2}\left((F_1(Q^2) + F_2(Q^2))^2 + (F_A + 2F_p)^2 - \left(4 + \frac{Q^2}{M^2}\right)(F_P)^2\right)\right],$$ (4.12)

$$B(Q^2) = \frac{Q^2}{M^2}F_A\left[F_1(Q^2) + F_2(Q^2)\right],$$ (4.13)

$$C(Q^2) = \frac{1}{4}\left[(F_A)^2 + F_1^2(Q^2) + \frac{Q^2}{4M^2}F_2^2(Q^2)\right],$$ (4.14)

$$F_i(Q^2) = F_i^p(Q^2) - F_i^n(Q^2) \qquad (i = 1, 2)$$ (4.15)

where the electromagnetic form factors $F_1^N(Q^2)$ and $F_2^N(Q^2)$ are known as the Dirac and Pauli form factors of the nucleon N ($N = p, n$), respectively. These form factors are well

constrained by electron-nucleon scattering data [100]. Since $m_\ell \ll M$ holds for $\ell = e, \mu$, the contribution of the third line in Eq. (4.12) can be neglected, which implies that the cross-section does not depend on the pseudoscalar form factor F_p. The remaining axial form factor $F_A(Q^2)$ only exists in the neutrino-nucleon cross-section and therefore gives the largest contribution to the uncertainty of the neutrino-nucleon cross-section.

The axial form factor is usually parametrized with a dipole form

$$F_A(Q^2) = \frac{g_A}{\left(1 + Q^2 / \left(M_A^{\mathrm{QE}}\right)^2\right)^2}, \tag{4.16}$$

where the normalization constant $g_A = F_A(0) = 1.2673 \pm 0.0035$ is well-known from the neutron β decay [101]. M_A^{QE} is called the axial mass and is constrained from neutrino-deuterium scattering in the bubble chamber experiments to be 1.026 ± 0.021 GeV [101].

Single pion production   As shown in Fig. 4.5, at the neutrino energy region around a few GeV, the most dominant interaction process is resonance scattering (RES). In the resonance scattering process, a nucleon struck by a neutrino can be a baryon resonant state and it decays into a nucleon and a pion in the final state [102].

An example of the resonant pion production process can be expressed as

$$\nu_l + p \rightarrow \ell^- + \Delta^{++} \rightarrow \ell^- + p + \pi^+. \tag{4.17}$$

The Feynman diagram of this process is shown in Fig. 4.7. The resonant pion production can occur via neutral-current (NC) as well.
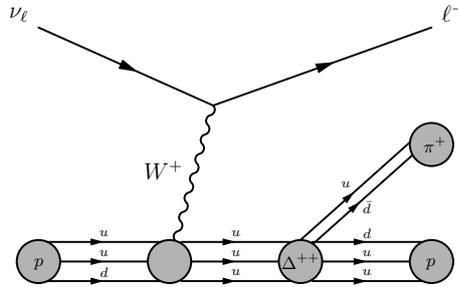
Figure 4.7. Feynman diagram of the charged current resonant 1π process.

Similarly to the CCQE case, one can obtain the reconstructed neutrino energy by replacing the outgoing nucleon with the excited baryon as

$$E_\nu = \frac{2m_p E_l + -m_l^2 + (m_\Delta^2 - m_p^2)}{2(m_p - E_l + p_l \cos\theta)} \tag{4.18}$$

where m_Δ is the mass of Δ^{++}.

The form factor for the resonant interaction can be written as

$$F_A^{\mathrm{RES}}(Q^2) = \frac{C_5^A(0)}{\left(1 + Q^2 / (M_A^{\mathrm{RES}})^2\right)}, \tag{4.19}$$

where $C_5^A(0)$ gives the normalization factor and M_A^{RES} is the axial mass [103].

There are also non-resonant isospin-1/2 and isospin-2/3 processes ($I_{1/2}$ and $I_{3/2}$) that can produce a single pion in the final state [104, 105]. These processes are called the non-resonant backgrounds. In NEUT, only the $I_{1/2}$ process is simulated as the contribution of $I_{2/3}$ is understood to be less dominant from the data obtained in the bubble chamber experiments such as BNL [106, 105]. In the low-energy model, we use the Rein-Sehgal model [102] as the baseline model for the resonant single pion production.

In the coherent pion production process, a neutrino interacts with the entire nucleus and produces a pion without exciting the nucleus [107, 108]

$$\nu_l + N \to l + N' + \pi. \tag{4.20}$$

This interaction occurs with low energy transfer. The cross-section of the coherent pion process is a few percent of CCQE and it is not a dominant process. The Feynman diagram of the coherent process is shown in Fig. 4.8. We use the Berger-Sehgal model [108] for the coherent single pion interactions with a small contribution of diffractive pion production modeled by Rein [109].
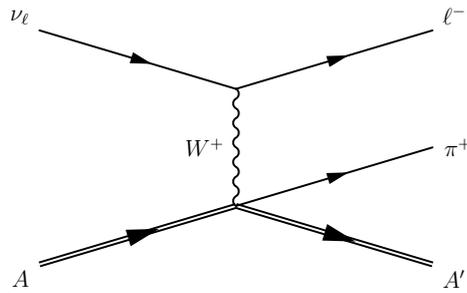
Figure 4.8. Feynman diagram of the charged current coherent 1π process.

Multi-pion and deep inelastic scattering At the higher neutrino energy region above ~ 5 GeV, the multi-pion and deep inelastic scattering (DIS) processes dominantly contribute. In these processes, neutrinos directly interact with quarks inside a nucleon and produce a jet of hadrons [110]

$$\nu_l + N \to l + N' + \mathrm{hadrons}. \tag{4.21}$$

The Feynman diagram of the DIS process is shown in Fig. 4.9.

These processes have only a subdominant effect in the T2K energy ranges, but become crucial for the higher-energy atmospheric samples. In the nominal interaction model, the GRV98 parton distribution functions [111] is used with the corrections by Bodek-Yang [112, 113]. The multi-pion mode is used to describe the events with a hadronic invariant mass[2] of $W \leq 2$ GeV and the DIS mode is used to describe the events with $W \geq 2$ GeV. To avoid double-counting the single-pion production process described above, the multi-pion mode is modified to simulate only the events with more than one pion in the final state.

---

[2]The hadronic invariant mass W is defined as $W = \sqrt{|p_{\mathrm{hadron}}|^2} = \sqrt{|p_\nu + p_{\mathrm{nucleon}} - p_{\mathrm{lepton}}|^2}$
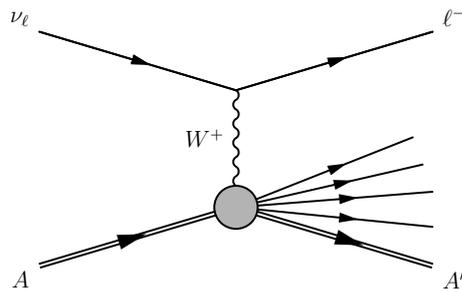
Figure 4.9. Feynman diagram of the charged current deep inelastic scattering (CC DIS) process.

4.3.1.2 Nuclear effects

The processes discussed in Section 4.3.1.1 assume a pure interaction between a neutrino and a nucleon and that the other nucleons do not participate in the interaction. However, since the nuclei used in our experiments (i.e. oxygen and carbon) have several nucleons, we should consider the effect of having multiple nucleons in the nucleus. These effects are often referred to as *nuclear effects.*

Initial state of nucleus   When nucleons are in a nucleus, they behave differently from the "free" state because they are bound and have finite momentum. Therefore, we need to take into account the initial state of nucleons such as the energy, momentum, and spatial distributions in the nucleus to deal with the neutrino-nucleus interaction. The Global Fermi Gas model (also called the Relativistic Fermi Gas model; RFG) is the simplest model for describing these effects, in which the momentum of initial state nucleons is assumed to be uniform in three-dimensional space for energy, momentum, and radial position up to the Fermi momentum p_F [114]. It should be noted that a neutrino interaction only happens when the energy transfer is larger than the binding energy E_b. The reference values of E_b used in the past T2K analysis are 25 MeV and 27 MeV for carbon and oxygen, respectively. The final state momentum of the struck nucleon should also be larger than the Fermi momentum p_F because the momentum states below p_F are filled by other nucleons, which is known as Pauli blocking. The Fermi momentum for the oxygen nucleus is assumed to be 225 MeV in `NEUT`.

In the Local Fermi Gas (LFG) model, effects of the finite size of the nucleus are included by introducing a radial nuclear matter density function, which makes the momentum distribution not flat in radial position anymore [115, 116]. These Fermi-gas models are known to overpredict the cross-sections for the events with forward-going leptons [117]. Therefore, these models usually take into account the so-called random phase approximation (RPA) and suppress the cross-sections for low momentum transfer [118].

The Spectral Function (SF) model proposed by Benhar et al. is a more sophisticated model based on a two-dimensional distribution of the nucleon momentum and binding energy [119, 120]. The spectral function $f(E_b, \boldsymbol{k})$ represents the probability density function of a nucleon in a nucleus as a function of its binding energy E_b and momentum $\boldsymbol{k}$. In SF,

it is assumed that the incoming neutrino interacts with a single nucleon inside the nucleus while the other nucleons behave as spectators, which is called the impulse approximation. Under this assumption, the neutrino-nucleus cross-section can be factorized as

$$\frac{d^6\sigma}{dE_\ell d\Omega_\ell d^3p} \propto \sigma_{\ell n} \times f(E_b, \boldsymbol{k}), \qquad (4.22)$$

where $\sigma_{\ell n}$ is the cross-section of a neutrino and a single nucleon. The spectral function is mostly tuned to electron-nucleus scattering data and partially relies on theoretical calculations.

The comparison of the predicted kinematic distribution of the nucleons in a nucleus is shown in Fig. 4.10. In RFG, the initial nucleon momentum distribution has a quadratic shape as it is integrated over the spherical nucleus, and has a cut-off around 225 MeV due to the Pauli blocking. LFG and SF have more diffused distributions in both the initial momentum and removal energy. The two sharp shell structures (corresponding to $E_b \approx 12$ MeV and $E_b \approx 18$ MeV) are seen in the SF distributions.

SF and LFG show significantly better agreement with the electron scattering data [121] and neutrino-nucleus cross-section measurements [122, 123], compared to RFG. The CCQE cross-section measured by the MINERνA and T2K experiments is shown in Fig. 4.11 together with the predictions from different nuclear initial state models. It demonstrates that SF and LFG outperform RFG in either measurement, but neither SF nor LFG is superior to the other at the current level of precision. In this analysis, we use SF as the nominal model to describe the initial state of the nucleus.

Multi-nucleon process When the neutrino interacts with a nucleon bounded in a nucleus, it can go through a process in which two nucleons are involved and two holes are produced. This process is called two-particle two-hole (2p2h) [127, 116, 128]. The contribution of 2p2h is not dominant but non-negligible in the SK and T2K analyses as shown in Fig. 4.5. There are two main contributions to the 2p2h processes: meson exchange currents (MEC) and nucleon-nucleon correlation (NN). MEC consists of various processes such as seagull, pion-in-flight, pion-pole, and Δ-pole as shown in Fig. 4.12. In NN, a photon is coupled to a pair of correlated nucleons.

In the analysis, we use the Nieves et al. model [116] as the basis of 2p2h modeling. However, the total 2p2h cross-section can differ from an alternative model such as Martini et al. [127] up to a factor of two as shown in Fig. 4.13, which implies that our analysis results could be biased if we use a "incorrect" model (i.e. different from the "true" physics model where the data is generated). Therefore, we study the possible bias coming from the mis-modeling of the neutrino-nucleus interaction in Chapter 6.
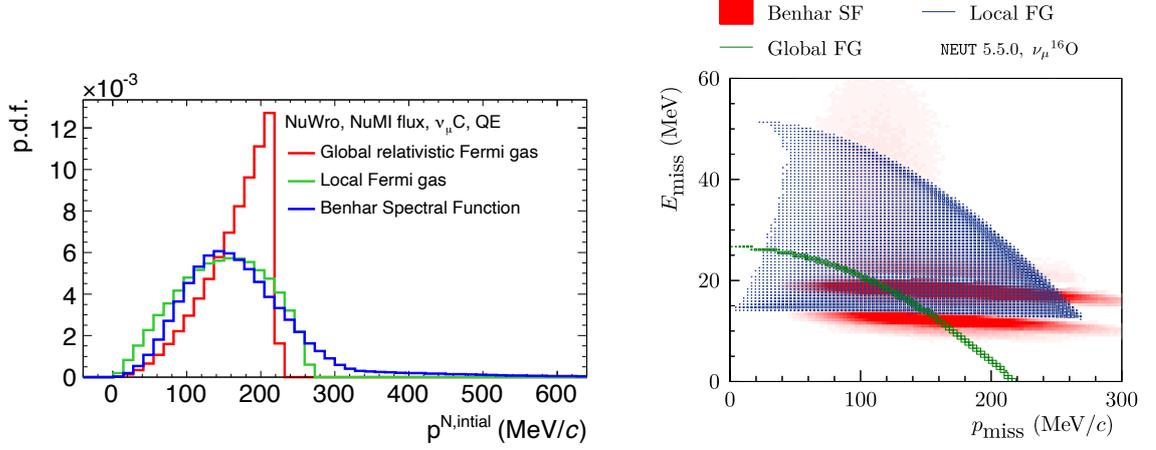
Figure 4.10. Comparison of the nucleon kinematic distributions for the three nuclear models: RFG, LFG, and SF. The left plot shows the initial momentum of the nucleon in the carbon nucleus simulated by the neutrino event generator NuWro [124]. The right plot shows the reconstructed missing momentum (p_{miss}) and missing energy (E_{miss}) distributions for ν_μ interactions on the oxygen nucleus simulated by NEUT. The missing momentum is defined as $p_{\mathrm{miss}} = |\boldsymbol{p}_\nu - \boldsymbol{p}_\ell - \boldsymbol{p}_p|$, and the missing energy is defined as $E_{\mathrm{miss}} = E_\nu + M_n - E_\ell - E_p - T$ using the reconstructed kinetic energy $T = \sqrt{p_{\mathrm{miss}}^2 + M^2} - M$ where M is the ground-state mass of nuclear remnant. The missing momentum and the missing energy are approximately equal to the initial momentum of the struck nucleon and the binding energy, respectively. The figures are taken from Refs. [125, 98].

(a) MINERνA (δp_T)   (b) MINERνA (p_n)   (c) T2K (δp_T)
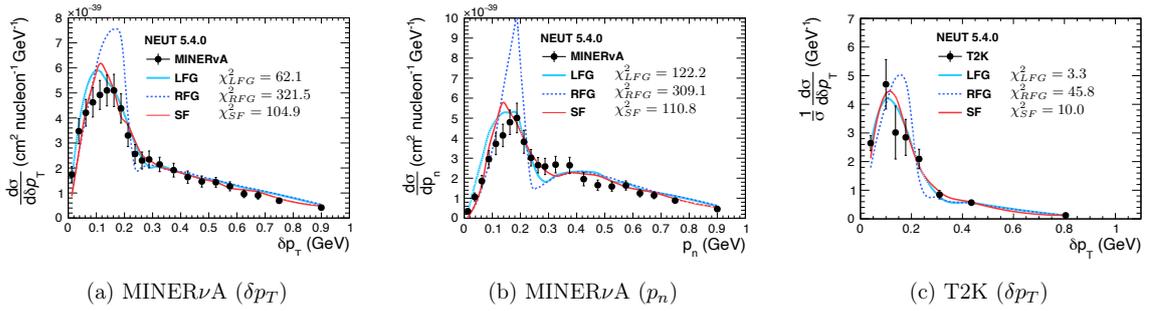
Figure 4.11. The results of the MINERνA and T2K CCQE cross-section measurements compared to the NEUT 5.4.0 [96] predictions for different initial state models: RFG [99], LFG [126], and SF [119]. The cross-sections are measured as a function of the transverse kinematics $\delta p_T = |\boldsymbol{p}_T^\mu + \boldsymbol{p}_T^p|$ and the reconstructed neutron momentum p_n, where $\boldsymbol{p}_T^\mu$ and $\boldsymbol{p}_T^p$ are the momentum of the outgoing muon and highest momentum proton in the plane transverse to the incoming neutrino direction. The agreement of the model and data is shown in terms of χ^2. The figures are taken from Ref. [123].
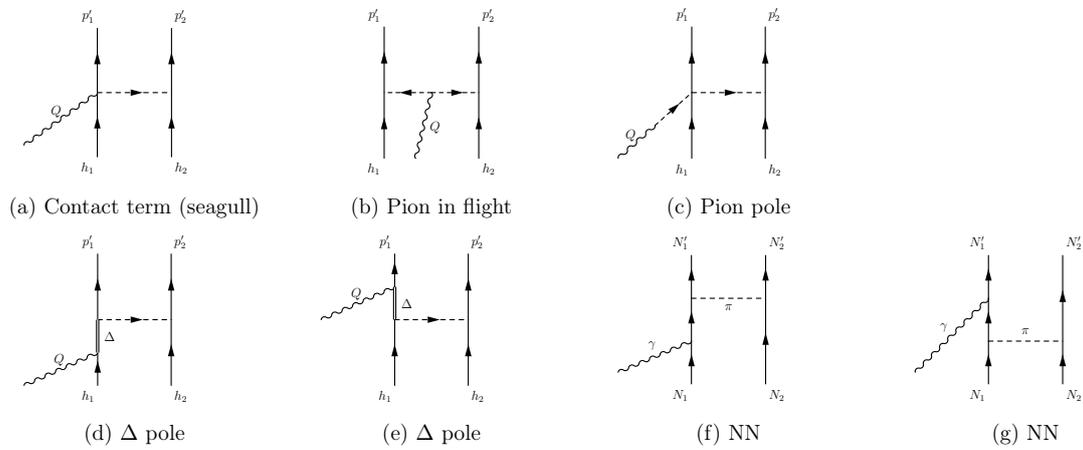
Figure 4.12. Feynman diagrams of various 2p2h processes. The single lines are nucleons, the double lines are the Δ resonances, the dashed lines are pions, and the curly lines are the W boson, respectively. (a)-(e) represent the MEC processes and (f)-(g) represent the NN proccesses. The figures are based on Refs. [129, 130].
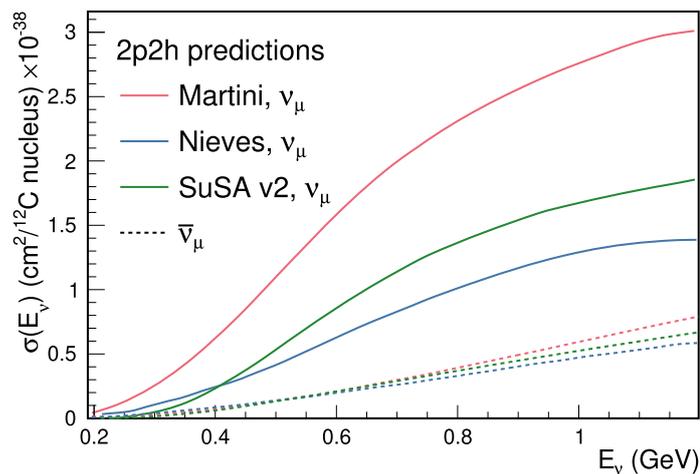
Figure 4.13. Comparison of the cross-sections of 2p2h interactions on ^{12}C from Martini et al. [127], Nieves et al. [126], and SuSAv2 [131, 129]. The figure is taken from Ref. [37].

4.3.1.3 Final state interaction

The final state particles such as pions and nucleons produced at a neutrino interaction process can re-interact with the nuclear medium while propagating [96, 98]. This is called a final state interaction (FSI).

In the neutrino interaction generator NEUT, the pion FSIs are simulated using the cascade model in which pions are propagated through a nucleus with finite steps. At each step, the probability of having various processes is computed and the process to be simulated is determined stochastically. The processes taken into account in our analysis are

1. Quasi-elastic scattering (low pion momentum $p_\pi < 500$ MeV)

2. Quasi-elastic scattering (high pion momentum $p_\pi > 500$ MeV)

3. Pion absorption

4. Charge exchange

5. Hadron production,

as illustrated in Fig. 4.14. For the nominal MC production, we use the Salcedo-Oset model [132, 133] tuned to the π-A scattering data [134] for simulating the cascade model.
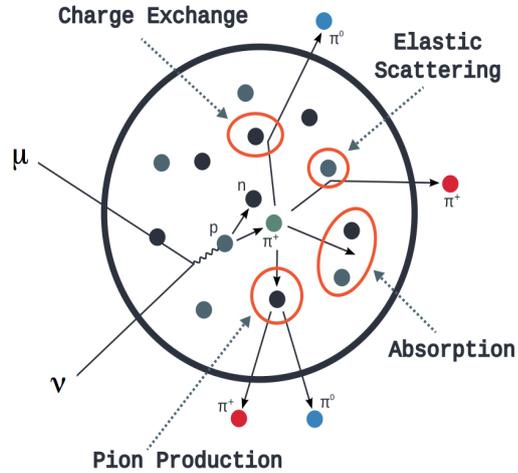
Figure 4.14. Illustration of various final state interaction (FSI) processes. The figure is taken from Ref. [135].

In addition, the momentum of the leptons leaving the nucleus is affected by Coulomb attraction and repulsion [136]. Therefore, the lepton momentum is shifted by $-3.6(-4.6)$ MeV for the negative-charge leptons and by $+2.6(+3.3)$ for the positive-charge leptons.

4.3.2 Systematic uncertainties

Although we use the common neutrino interaction model for the nominal MC production for the SK atmospheric and T2K accelerator neutrinos, we apply a different treatment in the systematic uncertainty model. As shown in Fig. 4.15, the atmospheric sub-GeV samples have neutrino energies below a few GeV, which is common to the T2K beam samples. It is therefore natural to use the same cross-section systematic uncertainty model for these samples as the physics should be common regardless of the neutrino sources or samples. Another benefit of using the same cross-section model between the beam and atmospheric neutrinos is that the T2K near detector constraint can also be applied to the SK atmospheric samples. On the other hand, the rest of the atmospheric samples have higher energies than the T2K beam samples and expand to significantly wider energy ranges. Considering the strong dependence of the cross-section on the neutrino energies, we use a different model for the higher-energy atmospheric samples. Accordingly, the T2K near detector constraint is not applied to these samples except for the CCQE cross-sections which are well-constrained by the T2K model and have less contribution in the high-energy samples.
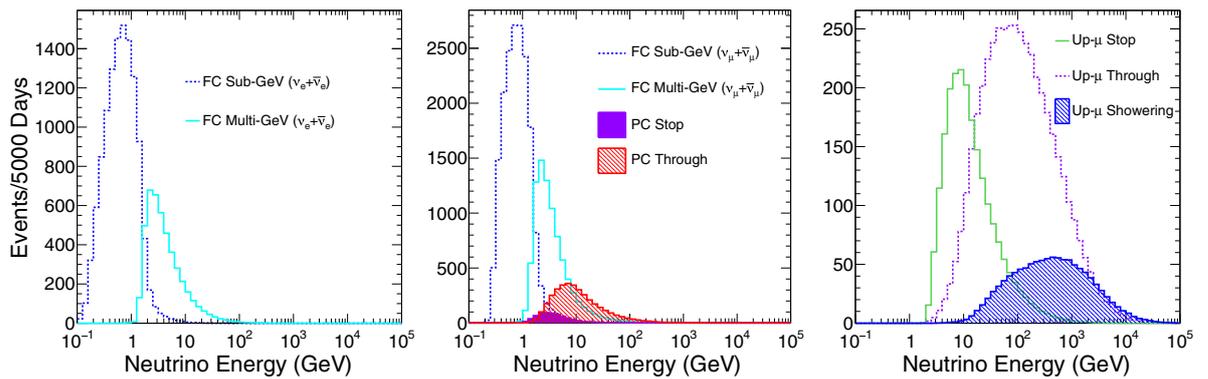
Figure 4.15. Simulated atmospheric neutrino energy spectra for each sample without oscillation. The figure is taken from Ref. [52].

We refer to the systematic uncertainty model applied to the (beam + atmospheric sub-GeV) samples as the "low-energy model" and the one applied to the rest of the atmospheric samples as the "high-energy model", respectively. The summary of the cross-section systematic uncertainty model is shown in Fig. 4.16. In the following subsections, the details of the cross-section systematic uncertainty model will be described.

4.3.2.1 Cross-section systematic uncertainties

CCQE As described in Section 4.3.1.1, the axial form factor plays the most important role in the formulation of the neutrino-nucleus CCQE cross-section. For the parameterization of the systematic uncertainty model, we use the dipole form of the axial form factor as shown in Eq. (4.16) and assign an error of 1.03 ± 0.06 to the axial mass M_A^{QE}. In addition, to give enough freedom to cover the suppression of the events in the recent cross-section measurements by NIMERνA [117, 137] and T2K [138, 139], we have five

	Low-energy sub-GeV atm + beam	**High-energy** multi-GeV atm
CCQE	T2K model with ND280 constraint, correlated in low-E/highE (except for high-Q^2)	
	high-Q^2 params w/ND280	high-Q^2 params w/o ND
	add ν_e/ν_μ ratio unc. (CRPA)	
2p2h	T2K model w/ND280	SK model (100% error) + T2K-style shape
Resonant	T2K model w/ND280 + new pion momentum dial + NC1π0 uncertainties	SK model for 3 dials common with T2K, use more recent larger T2K priors
DIS	T2K model w/ND280	SK model
ν_τ	SK model (25% norm on top of other syst) for other systematics checked that we have no numerically unstable values	
FSI	T2K model w/ND280	T2K model w/o ND280 should be mostly same as SK model
SI	T2K model, correlated in low-E/high-E only applied to FC and PC for atm, PN not applied to atm	

Figure 4.16. Summary of the neutrino interaction model used in this analysis.

energy-dependent normalization parameters for each momentum transfer (Q^2) region (referred to as "low-Q^2" parameters). Each of these parameters gives the normalization to the events with $Q^2 \in [0, 0.25]$ GeV^2 in 0.05 GeV^2 increments, respectively. To account for the higher momentum transfer, we have three additional "high-Q^2" parameters that cover $Q^2 \in [0.25, 0.50]$, $[0.50, 1.0]$, and $[1.0, \infty]$, respectively.

The same set of parameters is applied to the CCQE interactions in the high-energy model as well, except for the high-Q^2 parameters. Since the SK atmospheric samples have wider ranges of neutrino energies than T2K and are expected to have different constraints on these parameters, we use uncorrelated high-Q^2 parameters for the high-energy model without applying the T2K near detector constraints.

The systematic uncertainties for the binding energy are implemented as four correlated parameters applied to each combination of the neutrino type ($\nu, \bar{\nu}$) and target nucleus (carbon, oxygen). The binding energy parameters on carbon are not used in the SK fit but used in the T2K near detector fit, and they constrain the binding energies on oxygen through the correlation.

2p2h The 2p2h interaction has two nucleons in the final state but they mostly have momentum below the Cherenkov threshold. Therefore, it can mimic the signals from CCQE and can bias the measurements if the wrong reconstruction is applied assuming the CCQE process. It is important to model the uncertainties of this process properly.

For the low-energy model, four overall normalization parameters and four energy-dependent parameters are used to cover the possible systematic uncertainties. The normalization parameters are applied to neutrinos and antineutrinos, and carbon and oxygen, respectively. The other four energy-dependent parameters are applied to neutrinos/antineutrinos at $E_\nu < 0.6$ GeV and $E_\nu > 0.6$ GeV, respectively. For the high-energy model, 2p2h interactions are found to have less impact than the low-energy events. Therefore, only a simple 100% normalization uncertainty and an additional shape uncertainty are used to parametrize the uncertainties on 2p2h in this energy region.

Single pion production As defined in Eq. (4.19), we use the normalization factor C_5^A and axial mass M_A^{RES} in the parameterization of the systematic model along with the non-resonant isospin background $I_{1/2}$. In addition, we also have an isospin uncertainty for antineutrino events with low pion momentum ($p_\pi < 200$ MeV), which is uncorrelated with the normal isospin background. For the high-energy model, we also use the axial mass (M_A^{RES}), normalization factor (C_5^A), and isospin background ($I_{1/2}$). The low pion momentum isospin background is not included as it is not relevant to the high-energy neutrino interactions.

In the low-energy model, we assign three simple normalization uncertainties for CC coherent interactions on carbon and oxygen, and NC coherent interaction with a size of 30%, respectively. CC coherent parameters are 100% correlated between carbon and oxygen, and NC coherent parameter is totally uncorrelated with the CC coherent parameters. For the high-energy model, we have a single parameter that gives 100% uncertainty for CCν_μ and CCν_τ, and 50% uncertainty for CCν_e and NC interactions.

Multi pions and deep inelastic scattering (DIS) The systematic uncertainties for the multi-pion and DIS processes are assigned to both the total cross-section and the corrections by Bodek and Yang. For the total cross-section, we have overall normalization uncertainties of 3.5% and 6.5% for neutrinos and antineutrinos, respectively, with an additional uncertainty for the multi-pion mode to account for the uncertainty on the number of produced pions. For the Bodek-Yang corrections, we assign two independent parameters for the multi-pion mode and DIS mode, respectively.

In the high-energy model, we use a similar but more developed systematic model for these multi-pion and DIS modes. We have three separate uncertainties for the Bodek-Yang correction to the multi-pion mode, which account for the axial and vector part of the structure functions and overall normalization, respectively. An additional parameter is included to cover the uncertainty that comes from the comparison of the nominal model to an alternative model called CKMT [140].

ν_τ interaction For tau neutrinos, we assign a normalization uncertainty of 25% to all the ν_τ events on top of the other interaction uncertainties to account for the theoretical uncertainty on the ν_τ cross-sections.

Final state interaction and secondary interaction The FSI uncertainties are directly taken from the tuning to the π-A scattering data as a set of five correlated parameters that correspond to the five simulated processes of FSI (QE scattering for low and high energy, pion absorption, charge exchange, and hadron production). The uncertainties on the secondary interactions (SI), which are the reinteractions of the outgoing hadrons outside the nucleus, are also included using the same parameterization with the FSI. For the T2K samples, these SI uncertainties are implemented as the normalization of the event rates in each momentum bin together with the other detector systematic uncertainties described in Section 4.4.2. For the atmospheric samples, the SI systematic uncertainties are implemented as a set of five parameters that are correlated with the SI contributions in the T2K beam sample uncertainties.

Other interaction uncertainties In the low-energy model, an uncertainty of 100% is applied to the NC1γ interaction. In addition, a single uncertainty of 30% is applied to the cross-sections of "other" NC interactions including NC elastic, NC resonant kaon/eta production, and NC DIS interactions. These NC-other systematics are uncorrelated between the T2K near detector and SK. The uncertainties of cross-sections of ν_e and $\bar{\nu}_e$ are defined by combining the 2% uncorrelated uncertainty and 2% anticorrelated uncertainty. Furthermore, additional uncertainties on the ν_e and $\bar{\nu}_e$ cross-sections are included based on the comparison between SF and an alternative model called CRPA [141, 142]. The total cross-sections of CC interactions that are not affected by the other interaction uncertainties are given a single parameter with 100% uncertainty, which includes CC1γ, CC resonant kaon and eta production, and diffractive pion production. The Coulomb correction uncertainties of 2% (1%) are applied to the (anti-)neutrino cross-sections for $E_\nu \in [0.4, 0.6]$ GeV.

4.3.2.2 Effect of applying the T2K near detector constraint

To test the validity of applying the T2K near detector constraint to the sub-GeV atmospheric samples, we checked the agreement between the data and MC-predicted event spectra for these samples. To avoid seeing data before actually performing the fit, we only selected the down-going events defined as

$$\cos\Theta_z > 0.4 \tag{4.23}$$

where Θ_z is the zenith angle of the incoming neutrino direction ($\Theta_z = -1$ for neutrinos coming from the opposite side of the Earth and $\Theta_z = 1$ for neutrinos coming from the atmosphere above SK). As shown in Fig. 2.7, these down-going events are almost unaffected by the neutrino oscillations because of the relatively shorter traveling distances compared to the upward-going events. For the three μ-like samples, we also ignore the events with lepton momentum $p_\ell < 631$ MeV as they can be slightly affected by the oscillation. We treat these samples as pseudo-control samples and use them to test the agreement between data and MC (they are also used in the actual data fit).

Figure 4.17 shows the data and the variation of the predicted event spectra when varying the flux and cross-section systematic uncertainty parameters. The systematic uncertainty parameters are varied according to their prior uncertainties and correlations

with and without the T2K near detector constraints (denoted as "Pre-ND" and "Post-ND" in the plots, respectively). The oscillation probability is applied to the "Pre-ND" and "Post-ND" predictions (not to the raw MC) assuming the predefined oscillation parameter set A in Table 5.3 which is close to the result of the past T2K analysis. The overall predictions show a good agreement in the CC0π-enriched samples (e-like 0de, μ-like 0de and 1de), while data shows a clear excess in the CC1π-enriched samples (e-like 1de, μ-like 2de) and the π^0-like sample. The excess is larger in the low-momentum regions for the e-like and π^0-like samples while it looks relatively constant in the μ-like sample. The discrepancy becomes worse when we apply the T2K near detector constraints which reduces the predicted number of events in the CC1π samples. These results imply that our originally proposed model did not have a good enough prediction for these CC1π samples, and therefore, we added some additional systematics to account for the possible mis-modeling.
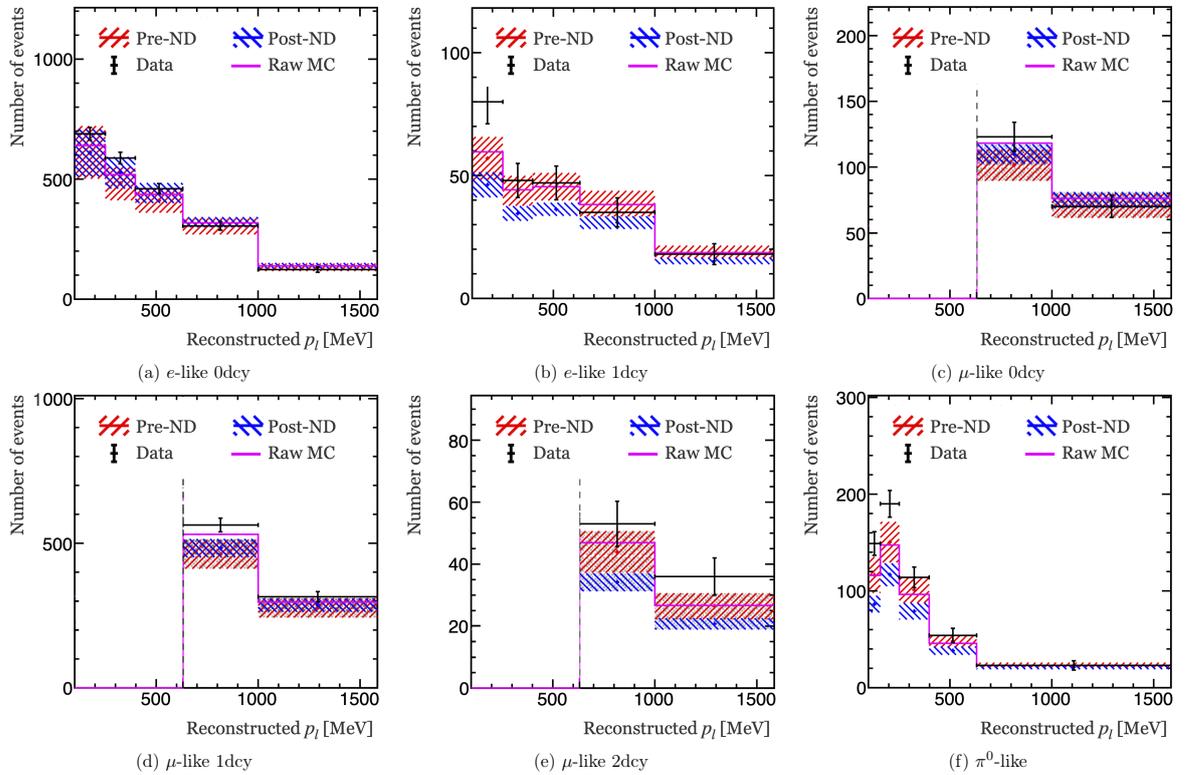
Figure 4.17. Variation of the predicted spectra for the down-going events ($\cos\Theta_z > 0.4$) of each sub-GeV atmospheric sample when varying the flux and cross-section parameters. "Pre-ND" and "Post-ND" show the spectra before and after applying the constraints from the T2K near detector fit. The oscillation probability is applied assuming the predefined oscillation parameter set A in Table 5.3. For the μ-like samples, the energy regions below 631 MeV are not shown as they could be sensitive to the oscillation.

4.3.2.3 Additional cross-section systematic uncertainties

In the SK selection, CC1π samples are identified by requiring one primary ring produced by the outgoing lepton and additional decay electron signals produced by the pion (and the primary muon). If pions have momentum above the Cherenkov threshold, they are identified as an additional primary ring, and the event is classified into the multi-ring samples. Therefore, by modifying the pion momentum and making it cross the Cherenkov threshold, we can increase or decrease the number of CC1π events. To add extra freedom to the pion momentum without changing the lepton kinematic, we added uncertainties that can account for the Adler angle distortions [143, 144, 145]. Here we assume that pions are produced from the resonance state of a nucleon. We first move to the rest frame of the resonance state nucleon. Then, introduce a parameter that can modify the angle of the pion emission in this frame. Thanks to the boost of the resonance state nucleon, we achieve the modified pion momentum which simply depends on the pion emission angle. We introduced two of these Adler angle uncertainties in the low and high momentum regions, respectively, to account for the fact that we observed a momentum-dependent excess in the e-like CC1π sample.

To mitigate the excess seen in the π^0-like sample, we added a parameter of 30% normalization uncertainty for the resonant NC1π^0 interaction which was constrained by the CC parameter in the previous T2K analysis and therefore did not exist as an independent parameter. We also inflated the uncertainty on the coherent NC1π^0 parameter from 30% (in the T2K analysis) to 100%. The values of these NC1π^0 resonant/coherent interaction uncertainties are tuned to the MiniBooNE data [146].

4.4 Detector response simulation

Once the neutrino flux and neutrino-nucleus interactions are simulated, the detector response simulation is responsible for converting the simulated output particles into the visible output of the detectors.

In SK, the detector response is simulated using a `GEANT3`-based [147] simulation software. It simulates the propagation of charged particles through the water and the emission of the Cherenkov light. The secondary interactions of the hadrons with water are simulated using the `NEUT` FSI model below the hadron momentum of $p < 500$ MeV and using `GEANT3 CALOR` [148] at $p > 500$ MeV. The Cherenkov light emitted from the charged particles is also propagated through the detector geometry until it reaches the PMTs or the detector wall. For photons reaching PMTs, the response of the PMTs is simulated by taking into account the quantum efficiencies. Finally, the response of the PMT is converted into the digitized hit information with timing and charge information which can be used in the following analysis.

A similar detector response simulation is implemented for the T2K near detectors using `GEANT4` instead of `GEANT3`. The particles generated from the neutrino interactions are propagated through the detector geometries which include not only the active detectors but also the non-active materials surrounding the detectors such as boxes and frames. Complex detector response simulations are implemented for each subdetector component and they are applied to the charged particles passing these active detectors. Finally, a

set of digitized hits from all the subdetectors is returned as the output of the detector response simulation.

4.4.1 SK atmospheric detector systematic uncertainties

The SK atmospheric detector systematic uncertainties are parametrized with 27 parameters that correspond to each step of the event reconstruction, calibration, and selection. In addition to the cosmic ray muon data, the atmospheric neutrino data is used to evaluate some systematic uncertainties because no control sample can cover the same energies and event topologies as the atmospheric samples. The evaluated size of systematic uncertainties is translated into the normalization of the event rates in the corresponding samples as summarized in Table 4.3. The size of some uncertainties is defined separately for the events within $50 < d_{\mathrm{wall}} < 200$ cm and events in $d_{\mathrm{wall}} > 200$ cm to take into account the relatively larger uncertainties for the events with the vertex near the wall. The details of the evaluation of the atmospheric detector systematic uncertainties can be found in Appendix A.1.

In addition to the systematic uncertainties that affect the event rate normalizations, there are uncertainties in the energy calibration. We implement this systematic uncertainty as a parameter that can directly modify the reconstructed lepton kinematics. The size of the uncertainty is the same for all the samples independent of the event topology and energy. There is one additional energy-scale uncertainty due to the up/down asymmetry of the detector light transmission, which is only applied to the FC and PC samples.

Table 4.3. Summary of the SK atmospheric detector uncertainties. The systematic uncertainty parameters are implemented as the normalization of the events in the corresponding sample, except for the energy-scale systematic uncertainty parameters which are implemented as the scaling of the reconstructed lepton momentum. The values in the parenthesis represent the uncertainties for events close to the wall ($50 < d_{\mathrm{wall}} < 200$ cm). The values are taken from Ref. [36].

Systematics	Sample	σ (%)
FC reduction	FC	1.3
FC/PC separation	FC Multi-GeV μ-like/PC	0.02
PC reduction	PC	1.0
PC stopping/through-going separation top	PC	6.8
PC stopping/through-going separation bar	PC	8.5
PC stopping/through-going separation bot	PC	40
Up-μ reduction	UpMu stopping	0.5
	UpMu through-going	0.3
Up-μ stopping/through-going separation	UpMu stopping/through-going	0.6
Up-μ energy cut	UpMu stopping	1.7
Up-μ path length cut	UpMu through-going	1.5
Up-μ shower/non-shower separation	UpMu showering/non-showering	3.0
Up-μ stopping BG subtraction	UpMu stopping	11
Up-μ through-going non-showering BG subtraction	UpMu non-showering	17
Up-μ through-going showering BG subtraction	UpMu showering	24
Fiducial volume	FC/PC	2.0
Decay-e tagging	Sub-GeV	0.7 (0.7)
	Multi-GeV	0.7 (2.1)
Two-ring π^0 selection	Sub-GeV π^0-like	1.03
Cosmic ray background	Sub-GeV/Multi-GeV μ-like	0.02
	Multi-ring μ-like	0.07
	PC	0.49
FC cosmic muon subtraction	Multi-GeV/Multi-Ring μ-like	67
Flasher background	Sub-GeV e-like	0.03
	Multi-GeV e-like	0.07
Energy scale		2.17
Up/down asymmetry energy scale		0.67

Table 4.3. Summary of the SK atmospheric detector uncertainties. The systematic uncertainty parameters are implemented as the normalization of the events in the corresponding sample. The anticorrelated normalization effects are denoted as negative values. "MME" refers to the multi-GeV multi-ring e-like sample separation described in Section 3.3.2.

Systematics	Sample		σ (%)	
			$50 < d_{\text{wall}} < 200$	$d_{\text{wall}} > 200$
Ring separation	Sub-GeV, Single-ring	e-like, $p < 400$ MeV	-1.31	0.74
		e-like, $p > 400$ MeV	-4.88	1.99
		μ-like, $p < 400$ MeV	-3.85	-0.73
		μ-like, $p > 400$ MeV	3.81	0.70
	Multi-GeV, Single-ring	e-like	-13.1	-11.6
		μ-like	10.3	9.29
	Multi-GeV, Multi-ring	e-like	4.53	5.06
		μ-like	-6.23	-6.10
Single-ring PID	Sub-GeV	e-like	3.88	1.82
		μ-like	-6.60	-1.88
	Multi-GeV	e-like	0.078	-0.036
		μ-like	-0.086	0.038
Multi-ring PID	Sub-GeV	e-like	6.76	-0.27
		μ-like	-8.45	0.30
	Multi-GeV	e-like	-0.85	-0.90
		μ-like	1.51	1.33
MME stage 1	e-like		-0.88	-0.67
	other		0.50	0.53
MME stage 2	ν_e-like		-3.64	-2.33
	$\bar{\nu}_e$-like		4.51	2.10

4.4.2   T2K detector systematic uncertainties

4.4.2.1   Detector systematic uncertainty binning

The detector systematic uncertainties in the T2K far detector analysis are defined by combining the uncertainties corresponding to each step of the reconstructions and selections. The estimated uncertainties are summarized as a covariance matrix which represents the event rate normalization for binned energy (momentum) ranges. The binning is defined differently for each combination of the reconstructed samples and true event topologies as summarized in Table 4.4. In addition to the event topology, the true ν_e events in the e-like samples are separated into the oscillated ν_e components (from $\nu_\mu \to \nu_e$ oscillation) and the intrinsic ν_e components in the neutrino beam.

4.4.2.2   Adding correlations to the detector systematics

Since both T2K and SK use the same detector and reconstruction, it is natural to correlate the detector systematic uncertainties between these two experiments. We include the correlations with the following two steps: propagation of the atmospheric detector systematics to the beam and addition of the beam-specific detector systematics.

Propagation of the atmospheric detector systematic uncertainties   To propagate the atmospheric systematic uncertainties to the beam samples, the atmospheric MC is first reweighted as a function of the flux and true neutrino types to obtain the T2K-like MC data set. Then, we vary the atmospheric detector systematic uncertainty parameters and see how the number of events in the T2K-like MC changes. Since the T2K detector systematic uncertainties are defined as the normalization of the event rates in each kinematic bin, this procedure allows us to get a new error matrix for the T2K part with correlations to the atmospheric detector uncertainties naturally included.

T2K beam-specific detector systematic uncertainties   Since several T2K event selection steps are not included in the atmospheric analysis, we should add the uncertainties related to these selections. These "beam-specific" uncertainties, related to the π^0 rejection, ν_μ CC and NC backgrounds in the e-like samples, and ν_e CC backgrounds in the μ-like samples, are added on top of the event rate variations estimated from the atmospheric detector uncertainties. More details on the T2K detector systematic uncertainty estimation can be found in Appendix A.2.

   The correlation matrices of the detector systematics with and without including the correlation are shown Fig. 4.18. The bottom-left part is the beam detector systematic uncertainties and the top-right part is the atmospheric detector systematic uncertainties. The top-left and bottom-right regions show the evaluated correlations between the beam and atmospheric systematics. Among the atmospheric systematics, fiducial volume, ring separation, and single-ring PID systematics have strong correlations with the beam detector systematics.

   Figure 4.19 shows the breakdown of the T2K detector systematic uncertainties. The uncertainties in bins corresponding to signal-dominated categories (ν_e channels for 1Re

Table 4.4. Summary of the T2K detctor systematic uncertainty binning. The T2K detector systematic uncertainties are extracted for each combination of the reconstructed sample and true event topology. The true event topologies are listed in the "Component" rows, where "Osc. ν_e" means ν_e components from the $\nu_\mu \rightarrow \nu_e$ oscillation, and "Beam ν_e" means the intrinsic ν_e components in the neutrino beam.

Sample	Component	Kinematic range	Sample	Component	Kinematic range
FHC 1Re (p_ℓ [MeV])	Osc. ν_e	$0 - 300$ $300 - 700$ $700 -$	RHC 1Re (p_ℓ [MeV])	Osc. ν_e	$0 - 300$ $300 - 700$ $700 -$
	ν_μ	$0 - 300$ $300 - 700$ $700 -$		ν_μ	$0 - 300$ $300 - 700$ $700 -$
	Beam ν_e	$0 - 300$ $300 - 700$ $700 -$		Beam ν_e	$0 - 300$ $300 - 700$ $700 -$
	NC	$0 - 300$ $300 - 700$ $700 -$		NC	$0 - 300$ $300 - 700$ $700 -$
FHC 1Rμ ($E_{\rm rec}$ [MeV])	ν_μ CCQE	$0 - 400$ $400 - 1100$ $1100 - 30000$	RHC 1Rμ ($E_{\rm rec}$ [MeV])	ν_μ CCQE	$0 - 400$ $400 - 1100$ $1100 - 30000$
	ν_μ CC non-QE	$0 - 30000$		ν_μ CC non-QE	$0 - 30000$
	ν_e CC	$0 - 30000$		ν_e CC	$0 - 30000$
	NC	$0 - 30000$		NC	$0 - 30000$
FHC 1Re 1de (p_ℓ [MeV])	Osc. ν_e	$0 - 300$ $300 - 700$ $700 -$			
	ν_μ	$0 - 300$ $300 - 700$ $700 -$			
	Beam ν_e	$0 - 300$ $300 - 700$ $700 -$			
	NC	$0 - 300$ $300 - 700$ $700 -$			

samples and ν_μ channels for 1Rμ samples) are mostly determined by the effect of atmospheric detector systematic variations, while for the background categories, the main

(a) Uncorrelated   (b) Correlated
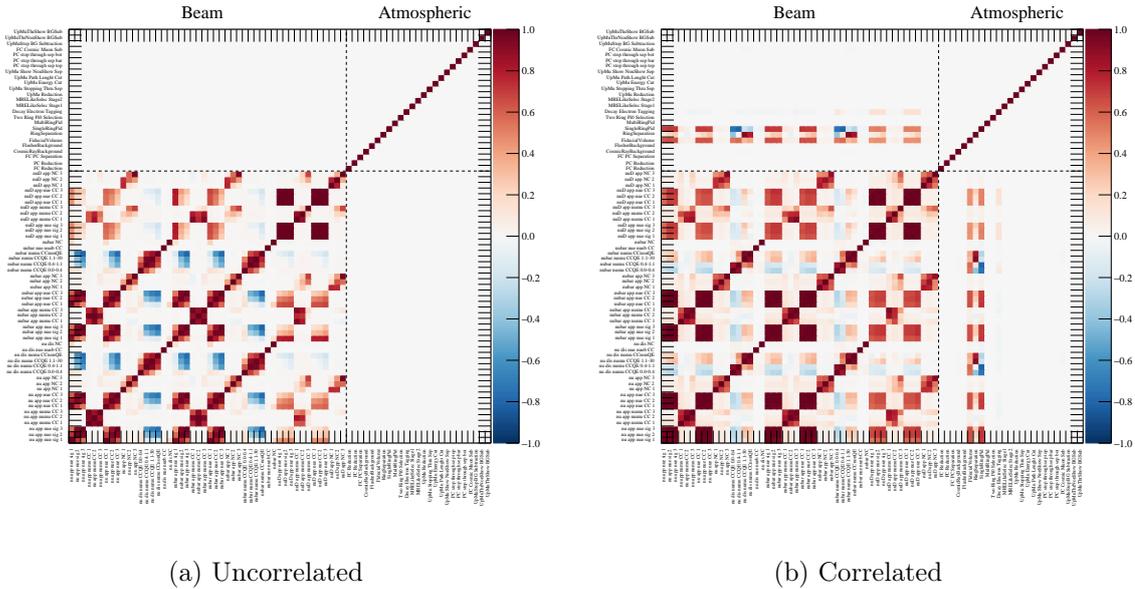
Figure 4.18. Correlation matrix of the detector systematic uncertainties with (left) and without (right) the correlations between SK atmospheric and T2K beam systematics. The bottom-left part is the beam detector systematic uncertainties and the top-right part is the atmospheric detector systematic uncertainties. The top-left and bottom-right parts show the correlations between beam and atmospheric detector systematics.

contribution comes from the beam-specific uncertainties.

Figure 4.20 shows the comparison of the T2K detector systematic uncertainties with and without correlations. In the correlated case, the uncertainties on the ν_e components (the oscillated ν_e and beam intrinsic ν_e) in the FHC 1Re 1de sample are significantly reduced compared to the uncorrelated case[3]. The correlated detector systematics have larger uncertainties for the NC components in the 1Re samples and the signal components (ν_μ CCQE) in the 1Rμ samples. The former comes from the beam-specific uncertainties and the latter comes from the correlated systematics with the atmospheric parameters such as the ring separation systematics.

Energy-scale systematics   We also correlate the two energy-scale parameters for the T2K beam and SK atmospheric samples with 100% prior correlations. For both parameters, the size of the uncertainty is defined as 2.17%.

---

[3]In the uncorrelated case, the errors on the e-like samples are estimated from the atmospheric ν_e CC0π and ν_e other control samples. The former is used to evaluate the uncertainties on the ν_eCC0π signals in the 1Re samples while the latter is used for the ν_eCC1π signals in the 1Re 1de sample. Since the latter atmospheric control sample also contains the background components rather than signal CC1π events, the uncertainties on the signal components in the 1Re 1de sample are estimated larger than that in the FHC 1Re samples. In the correlated detector systematics, these ν_e components in FHC 1Re 1de are also evaluated separately from the background components, and therefore, the size of uncertainties becomes more consistent with the other samples. More details on the T2K detector systematic uncertainty estimation can be found in Appendix A.2.
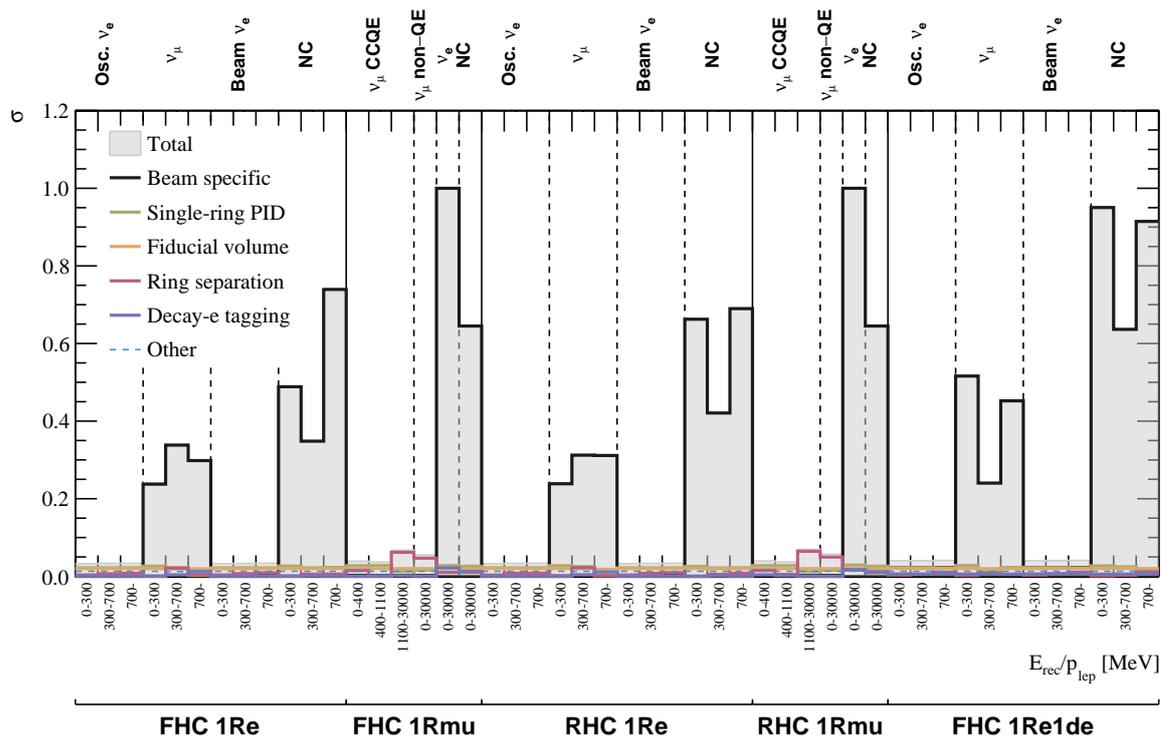
Figure 4.19. Breakdown of the detector systematic uncertainties in T2K error matrix binning (the square root of diagonal elements in the covariance matrix). Only the systematic uncertainties with larger bin contents than 0.02 are shown. The other systematic uncertainties are integrated and shown as the "Other" category with a dotted line. The gray-shaded histogram shows the total detector systematic uncertainties.
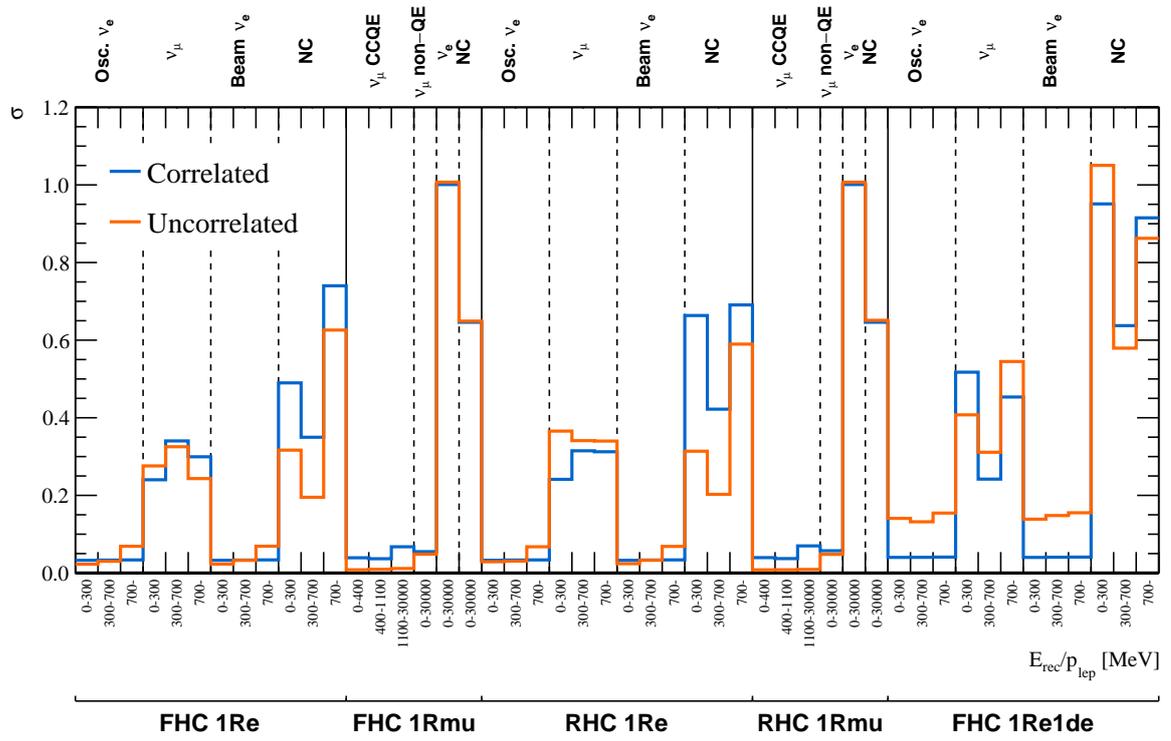
Figure 4.20. Comparison of the size of T2K detector systematic uncertainties (square root of diagonal elements in the covariance matrix) with and without correlations.

4.4.2.3 Additional PID systematics

To mitigate the effect of the data/MC excess seen in the atmospheric down-going CC1π samples shown in Section 4.3.2.2, we introduce empirical systematic uncertainties. The excess in the atmospheric sub-GeV e-like 1de sample has a momentum dependency and is maximum at the lowest momentum bin. The possible explanations for this excess are (1) event migrations between e-like and μ-like samples due to worse PID performance in the low momentum region, (2) different π production rates in ν_e/ν_μ interactions due to the difference in the masses of produced charged leptons ($m_e \ll m_\mu$) or polarization effects, and (3) contamination from $1\mu1\pi^0$ due to invisible muon and misidentified gamma rays. Although the actual reason for the excess is not clear, we introduced extra PID systematic uncertainties by assuming the excess is coming from (1).

If event migrations due to PID issues are considered between samples with the same decay electron count, the changes in event rates can be largely affected by the asymmetry of event rates between e-like and μ-like samples. The simulated event rate in each sub-GeV sample is shown in Table 4.5. As explained in Section 3.3.2, events are classified into samples based on the PID and the number of reconstructed decay electrons. For e-like samples, a decay-electron signal comes from an invisible charged pion, so the 0 de sample is dominated by CCQE and the 1 de sample is dominated by CC1π^+. For μ-like samples, however, the muon produces a decay electron as well, so both 0 and 1 de samples are dominated by CCQE, and the 2 de sample is dominated by CC1π^+. Since the CCQE cross-section is larger than the CC1π^+ cross-sections, there is a large difference in the number of events in the e/μ samples with 1 de. The e-like 1 de sample only has 0.12 times as many events as the μ-like 1 de sample, such that migration of even a small fraction of the events from the μ-like 1 de sample can result in a significant change of the number of events in the e-like 1 de sample. This is particularly relevant at low momenta where the PID performance is worse.

Table 4.5. Summary of the simulated event rates in each atmospheric FC sub-GeV sample with the neutrino oscillations are taken into account (500 years MC). The CCQE-dominated and CC1π^+-dominated samples are shown with red and blue background colors, respectively.

	0 de	1 de	2 de	Total
e-like	376276	37144	-	413420
μ-like	71444	314474	23586	409504
e-like/μ-like ratio	5.27	0.12	-	1.03

To define the properties of the extra PID systematic uncertainty parameters, we inspect the event rates and likelihood distributions of atmospheric sub-GeV samples separately for each analysis momentum bin: $[100.2, 251.2, 398.1, 631.0, 1000, 1585]$ MeV. The fractions of the true ν_μ CC, ν_e CC, and NC events in each sample and each analysis momentum bin are summarized in Table 4.6. The lowest momentum bin of the e-like 1de

sample has the largest contamination from the background ($\sim 19\%$) and the fraction of background events decreases with increasing the lepton momentum bin. This makes it plausible that the momentum dependence in the excess seen in the atmospheric sub-GeV sample comes from a worse PID at low momentum.

Table 4.6. The fraction of the event rates in each sub-GeV atmospheric sample and each analysis momentum bin broken down into true neutrino flavors. The signal and background components in each sample are shown with red and blue background colors. The deeper colors show worse signal efficiencies and greater background contaminations.

PID	e-like						μ-like					
Decay-e	0 de			1 de			0 de			1 de		
True neutrino flavor	ν_e	ν_μ	NC	ν_e	ν_μ	NC	ν_e	ν_μ	NC	ν_e	ν_μ	NC
$[100.2, 251.2]$ MeV	0.966	0.008	0.026	0.814	0.139	0.047	0.025	0.880	0.095	0.000	0.985	0.015
$[251.2, 398.1]$ MeV	0.948	0.012	0.040	0.877	0.086	0.037	0.007	0.953	0.040	0.000	0.993	0.007
$[398.1, 631.0]$ MeV	0.949	0.009	0.042	0.904	0.052	0.044	0.002	0.985	0.013	0.000	0.996	0.003
$[631.0, 1000]$ MeV	0.966	0.002	0.032	0.920	0.034	0.046	0.002	0.995	0.003	0.000	0.999	0.001
$[1000, 1585]$ MeV	0.971	0.002	0.027	0.930	0.027	0.043	0.002	0.997	0.001	0.000	0.999	0.001
Total	0.958	0.008	0.034	0.881	0.076	0.044	0.011	0.916	0.073	0.000	0.986	0.014

Figure 4.21 shows the distribution of the `fiTQun` single-ring PID likelihood for the sub-GeV 1 de events in the lowest and highest lepton momentum bin broken down by the true CC ν_e, CC ν_μ and NC components. Events are selected as e-like if $\log(L_e/L_\mu) \gtrsim 0$, and we can see that there is relatively large contamination from true ν_μ and NC events in this lowest momentum bin of the e-like 1de sample[4]. A lager background contamination is also seen in the lowest momentum bin of the μ-like 0 de sample, but in this case, the contamination seems to come more from the NC events than from the e/μ PID issue. As there is no problematic excess seen for this sample in the down-going data/MC comparison, we do not add extra systematic uncertainties for this sample.

The size of the extra PID systematic uncertainty for e-like samples is defined to be 20% of the number of events in the lowest momentum bin. This comes from the observed down-going data/MC discrepancy in the lowest momentum bin of the atmospheric e-like CC1π sample. The counter-effect of this parameter in the μ-like samples is defined based on the relative populations of 1de e-like and μ-like events in the lowest momentum bin: $20.0\% \times (9691/84109) = 2.3\%$. We introduce this extra PID systematic as two separate parameters for the T2K beam and SK atmospheric samples, with a 50% correlation between them.

---

[4]Here, the `fiTQun` "single-ring" fit results are shown for the atmospheric events. In the actual analysis, T2K uses the `fiTQun` single-ring fit results for event selections, whereas SK atmospheric uses the multi-ring one. Therefore, there will be small differences between the actual SK atmospheric event selection and a selection based on the single-ring $\log(L_e/L_\mu)$ shown here. This can be seen in Fig. 4.21b, which shows the distribution for this single-ring $\log(L_e/L_\mu)$ for events selected as e-like and μ-like in the real atmospheric event selection. This difference appears as a small overlap around 0, but it is negligible for seeing the impact of PID in this study and we can simply assume that events are separated into e/μ-like samples at $\log(L_e/L_\mu) \simeq 0$.

(a) $p_e \in [100, 251.2]$ MeV (true flavor)

(b) $p_e \in [100, 251.2]$ MeV (sample)

(c) $p_e \in [1000, 1585]$ MeV (true flavor)

(d) $p_e \in [1000, 1585]$ MeV (sample)
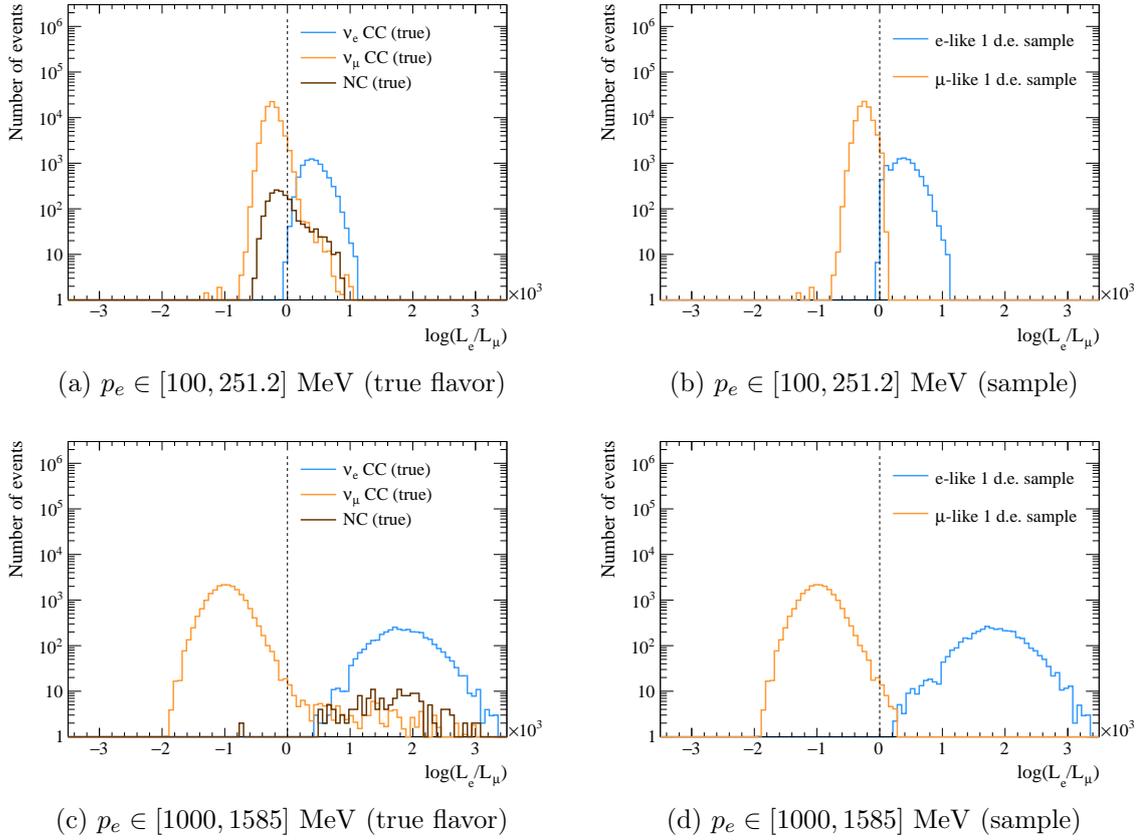
Figure 4.21. The logarithm of likelihood ratio between the e-like and μ-like hypotheses from the `fiTQun` single-ring fit results. The results are broken down into the true neutrino flavors (left) and the reconstructed sample (right). Only the atmospheric sub-GeV 1 de events in the lowest momentum bin $p_e \in [100, 251.2]$ MeV and the highest momentum bin $p_e \in [1000, 1585]$ MeV are shown.

The correlation is not set to 100% to take into account differences in event selections and populations between the samples of the two experiments, as well as differences in the existing PID uncertainty assignment.

Chapter 5

Analysis method and sensitivity

5.1 Likelihood and prior constraints

5.1.1 Likelihood definition

In the analysis, the observed data is first binned into one- or two-dimensional histograms according to their reconstructed kinematic properties for each data sample. Then, the number of observed data events in each kinematic bin is compared to the expected number of events. The expected number of events in bin i is given as a function of the oscillation parameters ($\boldsymbol{o}$) and systematic uncertainty parameters ($\boldsymbol{s}$) as

$$\lambda_i(\boldsymbol{o}, \boldsymbol{s}) = c \times \Phi_i(\boldsymbol{s}) \times \sigma_i(\boldsymbol{s}) \times d_i(\boldsymbol{s}) \times P_i^{\mathrm{osc}}(\boldsymbol{o}), \tag{5.1}$$

where c is an overall scaling constant and Φ_i, σ_i, d_i and P_i^{osc} are the neutrino flux, cross-section, detector response, and oscillation probability in bin i, respectively. The prediction step is done not analytically but using the MC data set as described in Section 4.1.

We use the binned-likelihood method to obtain the constraints on the parameters of interest from the data. In the frequentist analysis, the parameter best-fit values or confidence intervals are constructed by maximizing the (log-)likelihood. On the other hand, in the Bayesian analysis, the likelihood $\mathcal{L}$ is related to the posterior probability distribution $p(\theta|n)$ and prior $\pi(\theta)$ as

$$p(\theta|n) = \frac{\mathcal{L}(\theta|n)\pi(\theta)}{p(n)}, \tag{5.2}$$

where the parameters and data are denoted as θ and n, respectively. This relation, called Bayes' theorem, implies that our result (posterior probability) depends on our prior knowledge (prior probability) about the parameter of interest.

In our analysis, the likelihood consists of the statistical term and the constraint term

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{stat.}} \times \mathcal{L}_{\mathrm{const.}}, \tag{5.3}$$

or equivalently,

$$\log \mathcal{L}_{\mathrm{total}} = \log \mathcal{L}_{\mathrm{stat.}} + \log \mathcal{L}_{\mathrm{const.}}. \tag{5.4}$$

Statistical term   The statistical part of the likelihood can be defined assuming the Poisson distribution in each bin

$$\mathcal{L}_{\text{stat.}}(\lambda|n) \equiv \frac{P_{\text{Poisson}}(n|\lambda)}{P_{\text{Poisson}}(n|n)} = \frac{e^{-\lambda}\lambda^n}{e^{-n}n^n}, \tag{5.5}$$

where n is the number of observed events and λ is the expected number of events. Note that we normalize the Poisson probability by its value at $\lambda = n$ for simplicity.

Constraint term   To take into account the correlation between parameters and prior uncertainties, we have to add a constraint term to the likelihood. The constraint term for the systematic uncertainty parameters is defined as the multivariate Gaussian

$$\mathcal{L}_{\text{const.}}^{\text{syst.}} \equiv \frac{f(\boldsymbol{s}|\boldsymbol{\mu},\Sigma)}{f(\boldsymbol{\mu}|\boldsymbol{\mu},\Sigma)} = \exp\left[-\frac{1}{2}(\boldsymbol{s}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{s}-\boldsymbol{\mu})\right], \tag{5.6}$$

where $\boldsymbol{\mu}$ and Σ are the mean and covariance matrix, respectively. Similarly to the statistical term, the constraint term is also defined by dividing the Gaussian probability with its value at $\boldsymbol{s} = \boldsymbol{\mu}$ so that we can ignore the normalization term $(1/\sqrt{2\pi|\Sigma|^2})$. We should note that there are a few systematic uncertainty parameters that have flat priors and these parameters do not have the constraint terms. Some oscillation parameters are also given a prior uncertainty in the fit and their constraint terms are also added to the likelihood

$$\mathcal{L}_{\text{const.}}^{\text{osc.}} \equiv \prod_{i}^{\text{prior osc.}} \exp\left[-\frac{1}{2}\frac{(o_i - \mu_i)^2}{\sigma_i^2}\right], \tag{5.7}$$

where i only loops over the oscillation parameters that are given the prior uncertainties.

Full likelihood   To summarize, the full log-likelihood is defined as follows:

$$-2\log\mathcal{L}_{\text{total}}(\boldsymbol{o},\boldsymbol{s}|n) = -2\log\mathcal{L}_{\text{stat.}}(\lambda(\boldsymbol{o},\boldsymbol{s})|n) - 2\log\mathcal{L}_{\text{const.}}(\boldsymbol{o},\boldsymbol{s}) \tag{5.8}$$

$$= \sum_{s}^{\text{sample}}\sum_{b}^{\text{bin}} 2\left[\lambda_{s,b}(\boldsymbol{o},\boldsymbol{s}) - n_{s,b} + n_{s,b}\log\frac{n_{s,b}}{\lambda_{s,b}(\boldsymbol{o},\boldsymbol{s})}\right] \tag{5.9}$$

$$+ (\boldsymbol{s}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{s}-\boldsymbol{\mu}) \tag{5.10}$$

$$+ \sum_{i}^{\text{prior osc.}} \frac{(o_i - \mu_i)^2}{\sigma_i^2}. \tag{5.11}$$

When plotting the results, we select one or two oscillation parameter(s) and profile or marginalize the other parameters. The parameters that are profiled or marginalized and are not plotted in the results are called the *nuisance* parameters. Details of the methods to compute the marginal and profiled likelihood are described in Section 5.3.

Finally, let us define $\Delta\chi^2$ computed from the marginal or profiled likelihood. Here we denote the parameter of interest as θ. Then, from Wilks' theorem [149], the test statistic D,

$$D = -2\log\frac{\mathcal{L}(\theta|x)}{\mathcal{L}(\hat{\theta}|x)}, \tag{5.12}$$

is known to follow the χ^2 distribution with the degrees of freedom equal to the number of parameters under some assumptions[1]. Therefore, in the following analysis, we define χ^2 and $\Delta\chi^2$ as

$$\chi^2(\theta) \equiv -2\log\mathcal{L}(\theta|x), \tag{5.13}$$

$$\Delta\chi^2 \equiv \chi^2(\theta) - \chi^2(\hat{\theta}), \tag{5.14}$$

where $\hat{\theta}$ denotes the maximum likelihood estimator, i.e.

$$\hat{\theta} \equiv \arg\max_{\theta}\mathcal{L}(\theta|x) = \arg\min_{\theta}\left[-2\log\mathcal{L}(\theta|x)\right]. \tag{5.15}$$

5.1.2 Oscillation parameter prior constraints

We apply the Gaussian constraints from the external experiments on some oscillation parameters where we do not have strong sensitivities. In particular, the constraint on $\sin^2 2\theta_{13}$ is taken from the results of the reactor neutrino measurements (the weighted average of Daya Bay, RENO, and Double Chooz) and is referred to as "reactor constraint". The prior values used in the analysis are summarized in Table 5.1.

Table 5.1. Summary of the oscillation parameter prior distributions used in the data fit.

Parameter	Prior	Reference
δ_{CP}	$\delta_{\mathrm{CP}} \sim \mathrm{Uniform}(-\pi, \pi)$ $\sin\delta_{\mathrm{CP}} \sim \mathrm{Uniform}(-1, 1)$	
$\sin^2(2\theta_{13})$	$\mathrm{Gaussian}(\mu = 0.0853, \sigma = 0.0027)$	[151]
$\sin^2\theta_{12}$	$\mathrm{Gaussian}(\mu = 0.307, \sigma = 0.013)$	[151]
$\sin^2\theta_{23}$	$\mathrm{Uniform}(0, 1)$	
Δm_{21}^2 [10^{-5} eV^2]	$\mathrm{Gaussian}(\mu = 7.53, \sigma = 0.18)$	[151]
Δm_{32}^2 (NO)/$|\Delta m_{31}^2|$ (IO) [eV^2]	$\mathrm{Uniform}(0, \infty)$	
Mass ordering	$P(\mathrm{NO}) = P(\mathrm{IO}) = 0.5$ (equal)	

Since the results of the Bayesian analysis depend on the choice of the prior, and violation (or conservation) of CP is one of the most important questions we seek to answer in this analysis, we use two different prior probability distributions for δ_{CP} in our Bayesian analysis. The effects of the priors uniform in δ_{CP} and uniform in $\sin\delta_{\mathrm{CP}}$ are shown in Fig. 5.1. More detailed discussions on the prior choice can be found in Appendix B.

---

[1]In general, the conditions to apply the Wilks' theorem are not met in the neutrino oscillation analysis. Thus, we need special treatments to obtain proper coverage for the frequentist analysis [150].
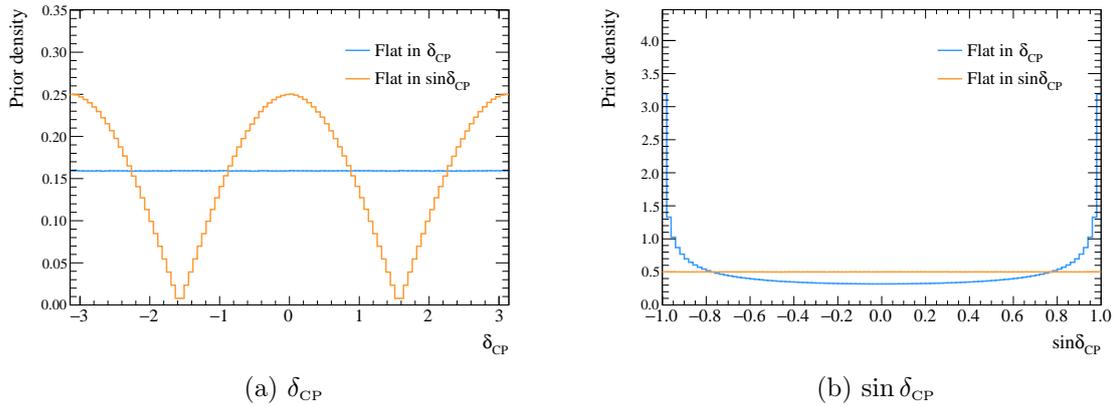
Figure 5.1. Comparison of the effect of the different priors on CP parameters.

5.2 Oscillation probability calculation

The T2K oscillation probability is calculated using the three-flavor oscillation probabilities given in Section 2.2 by taking into account the matter effect. The propagation medium is assumed to be uniform, and the fixed matter density of $\rho = 2.6$ g/cm^3 and the fixed propagation length of $L = 295$ km are used.

The oscillation probability calculation for the atmospheric neutrinos is more complex as the propagation distance depends on the direction of the neutrinos. Since the density of the matter is not uniform inside the Earth, the contribution of the matter effect also varies depending on the neutrino trajectories. In addition, atmospheric neutrinos can be produced up to about 50 km in the atmosphere, so the production height can also modify the oscillation probabilities.

To take into account these effects, the conventional SK analysis uses an approximated version of the preliminary reference Earth model (PREM) [152], where the Earth is modeled with four layers of medium with different densities[2]. The property of each layer is summarized in Table 5.2. In this analysis, we adopted a sophisticated treatment of the Earth's density to obtain a more precise estimation of the density than the fixed layer approximation [145]. In this method, the Earth's density is defined to be a quadratic function of radius (R) as

$$\rho_i(R) = \alpha_{i,2}R^2 + \alpha_{i,1}R + \alpha_{i,0}. \tag{5.16}$$

The value of density at each layer is derived for each neutrino event by averaging the density over its trajectory:

$$\langle \rho \rangle_i := \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} dt \rho(t), \tag{5.17}$$

where t_i is the intersection of trajectory t with the layer boundaries shown in Fig. 5.2b. The coefficients $\alpha_{i,j}$ are extracted by fitting the PREM with the quadratic function, which

---

[2]The full PREM model has 82 layers, but the SK analysis reduces them to four to reduce the computational time.

is summarized in Table 5.2. In the sensitivity study, the effect of the change in the Earth density treatment is found to be small, and it makes a shift in the $\Delta\chi^2$ curve of δ_{CP} by up to 0.2 rad in the fit to only the atmospheric samples.

Table 5.2. Summary of the properties of the Earth density model used in the conventional SK atmospheric neutrino analysis and this analysis. The Earth is modeled as a sphere of radius 6371 km with four layers of medium. The constant densities are used for the conventional SK analysis, whereas the quadratic densities are used for this analysis. The corresponding figure is shown in Fig. 5.2a. The constant density values are taken from Ref. [52].

Region	R_{min} (km)	R_{max} (km)	Constant density (g/cm^3)	Quadratic density (g/cm^3)
Inner core	0	1220	13.0	$13.09 - 8.84x^2$
Outer core	1220	3480	11.3	$12.31 + 1.09x - 1.25x^2$
Mantle	3480	5701	5.0	$6.78 - 1.56x - 1.25x^2$
Crust	5701	6371	3.3	$-50.42 + 123.33x - 69.95x^2$
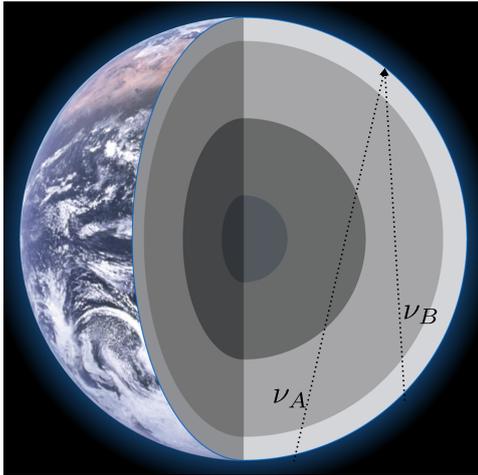
(a) Simplified Earth model
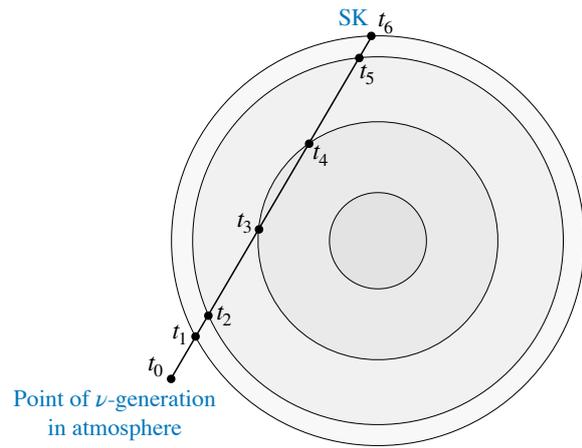
(b) Definition of the intersections t_i

Figure 5.2. Illustration of the propagation of atmospheric neutrinos. (a) Simplified Earth model that is used in the conventional SK analysis, where the four layers shown in different colors are the inner core, outer core, mantle, and crust from the inside. The figure is taken from Ref. [52]. (b) Definitions of the variables used in the average density calculations.

5.3 Fitting method

In our main analysis described in Chapter 7, we use Markov-Chain Monte Carlo (MCMC) to obtain the marginal likelihood and the posterior probability distributions. However, several other methods are also prepared to make inputs for MCMC and to perform validation and additional studies. In particular, the profiling method is used to find the

best-fit parameter set that maximizes the likelihood and to make a Gaussian approximation of the prior constraints on the systematic uncertainty parameters as described in Section 5.3.1. A numerical marginalization based on the importance sampling method, which is described in Section 5.3.2, is used to validate the result of MCMC and to perform many fits at the same time, as it can produce similar results but faster than MCMC.

5.3.1 Profiling method

The best-fit values are obtained by profiling over all the oscillation ($\boldsymbol{o}$) and systematic ($\boldsymbol{s}$) parameters using the gradient-descent method in `MINUIT` [153]:

$$\hat{\boldsymbol{o}}, \hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{o}, \boldsymbol{s}} \left[-\log \mathcal{L}_{\text{total}}(\boldsymbol{o}, \boldsymbol{s}|n) \right]. \tag{5.18}$$

The set of parameters that gives the maximum likelihood is known as the maximum likelihood estimator (MLE) and is denoted with a hat (i.e. $\hat{\boldsymbol{o}}, \hat{\boldsymbol{s}}$).

In general, the Hessian matrix $H(\boldsymbol{s})$, a square matrix of second-order partial derivatives, can be computed at the best-fit point (MLE) as

$$H_{i,j}(\boldsymbol{s}) = \frac{\partial^2 (\log \mathcal{L})}{\partial s_i \, \partial s_j}. \tag{5.19}$$

This is equivalent to the negative Fisher information matrix $I(\boldsymbol{s})$ at MLE:

$$I_{i,j}(\boldsymbol{s}) = \frac{\partial^2 (-\log \mathcal{L})}{\partial s_i \, \partial s_j} = -H_{i,j}(\boldsymbol{s}). \tag{5.20}$$

The inverse of the Fisher information (and therefore the inverse of the negative Hessian) at MLE is known to give the asymptotic covariance matrix [154]:

$$\Sigma \equiv \text{Cov}(\boldsymbol{s}) = I(\boldsymbol{s})^{-1} = -H(\boldsymbol{s})^{-1}. \tag{5.21}$$

This method is used to propagate the T2K near detector fit constraints to the analysis at SK, where the systematic uncertainty parameters are approximated to be the multivariate Gaussian function

$$f(\boldsymbol{s}|\hat{\boldsymbol{s}}, \Sigma) = \frac{1}{\sqrt{2\pi|\Sigma|^2}} \exp\left[-\frac{1}{2}(\boldsymbol{s} - \hat{\boldsymbol{s}})^T \Sigma^{-1} (\boldsymbol{s} - \hat{\boldsymbol{s}}) \right]. \tag{5.22}$$

One can also compute the standard error of the parameters by taking the square root of diagonal elements of the obtained covariance matrix.

A profile likelihood function is computed by fixing the parameter of interest ($\boldsymbol{\theta}$) at grid points and maximizing the likelihood at each grid by profiling the nuisance parameters ($\boldsymbol{\eta}$) using `MINUIT`:

$$\mathcal{L}_{\text{prof.}}(\boldsymbol{\theta}|n) = \mathcal{L}_{\text{total}}(\boldsymbol{\theta}, \hat{\boldsymbol{\eta}}|n), \quad \text{where} \quad \hat{\boldsymbol{\eta}} = \min_{\boldsymbol{\eta}} \left[-\log \mathcal{L}_{\text{total}}(\boldsymbol{\theta}, \boldsymbol{\eta}|n) \right]. \tag{5.23}$$

This method is not used for the main analysis but is used for some validation studies. Unless explicitly mentioned, we always use the marginal likelihood for our analysis.

5.3.2 Importance sampling method

In the numerical marginalization method, one can marginalize the nuisance parameters $\boldsymbol{\eta}$ by randomly throwing them many times and computing the prior-weighted average of the throws. This method has been used in the T2K analyses by using the Gaussian approximation of the near detector fit for the systematic constraints and sampling the throws from it (i.e. $\boldsymbol{\eta} \sim \pi(\boldsymbol{\eta}) \equiv \mathcal{L}_{\text{const.}}$). In this case, the marginal likelihood can be written as

$$\mathcal{L}_{\text{marg.}}(\boldsymbol{\theta}|n) = \int d\boldsymbol{\eta} \ \mathcal{L}_{\text{stat.}}(\boldsymbol{\theta}, \boldsymbol{\eta}|n)\mathcal{L}_{\text{const.}}(\boldsymbol{\eta}) \tag{5.24}$$

$$= \int d\boldsymbol{\eta} \ \pi(\boldsymbol{\eta})\mathcal{L}_{\text{stat.}}(\boldsymbol{\theta}, \boldsymbol{\eta}|n) \tag{5.25}$$

$$\xrightarrow[\eta \sim \pi(\boldsymbol{\eta})]{N \ \text{throws}} = \frac{1}{N_{\text{throw}}} \sum_{i}^{N_{\text{throw}}} \mathcal{L}_{\text{stat.}}(\boldsymbol{\theta}, \boldsymbol{\eta}_i|n). \tag{5.26}$$

However, this does not work well for the fit including the atmospheric samples as the atmospheric systematic uncertainty parameters have larger prior uncertainties than the post-fit uncertainties (i.e. they are not well-constrained in the prior). Since the volume of the parameter space increases in proportion to the power of the number of parameters, throwing unconstrained systematic uncertainty parameters from their priors will give us very sparse throws around the parameter region of interest. This significantly worsens the efficiency of the numerical marginalization (i.e. many throws fall in the small-likelihood regions). Therefore, for the joint analysis, we use an updated importance sampling method using the best-fit results.

We first obtain the best fit to the T2K e-like samples and the SK atmospheric samples using `MINUIT`. To avoid finding the best fit at local minima, we run the fits at several combinations of the octant, mass ordering, and δ_{CP}. Then, we construct the sampling distribution based on the Hessian of these best-fit results and throw the nuisance parameters from this sampling distribution. The final likelihood is obtained by reweighting each throw to obtain the weights from the original prior constraints. This method gives us a volume reduction of roughly $\sim 10^{18}$ and therefore allows us to perform the numerical marginalization with better efficiency (i.e. we can reduce the number of throws that fall in the small likelihood region). We do not include the T2K μ-like samples in the profiling fit when obtaining the best-fit results because they give too strong constraints on $\sin^2 \theta_{23}$ and Δm^2_{32} and give a poor estimation of the posterior parameter spaces.

To illustrate the impact of importance sampling, examples of the numerical marginalization results based on the importance sampling from the prior and the best-fit results are shown in Fig. 5.3. The errors are computed from the jackknife resampling method [155], where we repeat removing 1/10 of the throws to compute the variance with the rest of the throws and rescale the averaged variance to get an estimate of the variance of full throws. We use $N = 100,000$ throws to compute the marginal likelihood for both fits. When we sample the atmospheric systematic uncertainty parameters from their priors, we have large throw statistical errors due to the fewer number of throws that resulted in the large likelihood region. This issue is improved when we sample these systematic uncertainty parameters from the best-fit results.
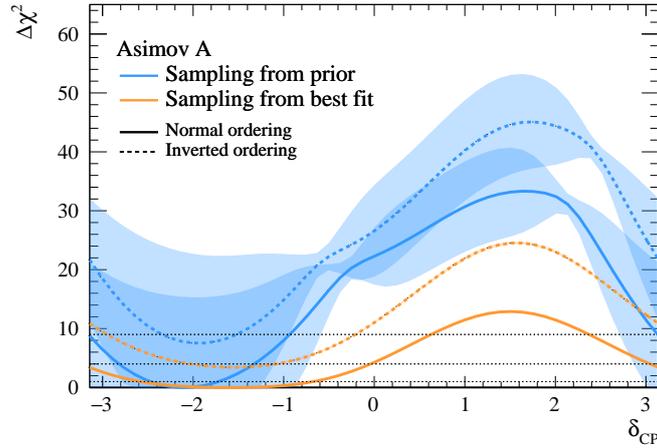
Figure 5.3. Comparison of the δ_{CP} $\Delta\chi^2$ obtained from the fit to MC data set using the numerical marginalization based on different sampling distributions. The errors estimated using the jackknife method are shown as the shaded bands. See Section 5.5.1 for the definition of the data set (Asimov A).

Despite the improvement in the sampling method, there still remains a possibility of getting large throw statistical errors in this method. The issue will be more crucial for the data fit where we have statistical fluctuations, and it is difficult to further improve the stability of the fit. Therefore, the use of the importance sampling method is limited to the validations and additional studies where we need to run many fits at the same time. For the main analysis, we use the MCMC method.

5.3.3 Markov-Chain Monte Carlo

Overview of MCMC Markov-Chain Monte Carlo (MCMC) is a frequently used approach to obtain Bayesian inferences with a complex model [156]. The Markov chain refers to the process where the probability of obtaining a certain state at one step only depends on the state at the previous step. This probability is called the *transition probability*. One can obtain the posterior probability distributions by running a chain of random walks that stochastically explore the parameter space according to the transition probabilities. For the implementation, we adopt a method called the random walk Metropolis-Hastings algorithm [157].

Let us denote the parameter set with $\boldsymbol{x}$ (corresponds to the state at each step) and the target posterior distribution with $f(\boldsymbol{x})$. From Bayes' theorem in Eq. (5.2), the posterior distribution is defined using the likelihood $\mathcal{L}(\boldsymbol{x})$ and prior distributions $\pi(\boldsymbol{x})$ as $f(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x})\pi(\boldsymbol{x})$ (here the normalization constant is ignored). To obtain the target posterior distribution with MCMC, the transition probability $p(\boldsymbol{x}'|\boldsymbol{x})$ from the state $\boldsymbol{x}$ to $\boldsymbol{x}'$ must satisfy the *detailed balance equation*:

$$p(\boldsymbol{x}'|\boldsymbol{x})f(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{x}')f(\boldsymbol{x}'), \tag{5.27}$$

where the transition probability can be further broken down into the *proposal function* $q(\boldsymbol{x}'|\boldsymbol{x})$ and the *acceptance rate* $\alpha(\boldsymbol{x}'|\boldsymbol{x})$ as

$$q(\boldsymbol{x}'|\boldsymbol{x})\alpha(\boldsymbol{x}'|\boldsymbol{x})f(\boldsymbol{x}) = q(\boldsymbol{x}|\boldsymbol{x}')\alpha(\boldsymbol{x}|\boldsymbol{x}')f(\boldsymbol{x}'). \tag{5.28}$$

Here, the proposal function gives a proposal of the next step in a stochastic way, and the acceptance rate decides whether we should accept or reject that proposal. When the proposal is rejected, the next step will stay the state place. Equation (5.28) can be rewritten as

$$\frac{\alpha(\boldsymbol{x}'|\boldsymbol{x})}{\alpha(\boldsymbol{x}|\boldsymbol{x}')} = \frac{q(\boldsymbol{x}|\boldsymbol{x}')f(\boldsymbol{x}')}{q(\boldsymbol{x}'|\boldsymbol{x})f(\boldsymbol{x})}. \tag{5.29}$$

In the random walk Metropolitan-Hastings algorithm, we define the acceptance rate as

$$\alpha(\boldsymbol{x}'|\boldsymbol{x}) = \min\left(1, \frac{q(\boldsymbol{x}|\boldsymbol{x}')f(\boldsymbol{x}')}{q(\boldsymbol{x}'|\boldsymbol{x})f(\boldsymbol{x})}\right). \tag{5.30}$$

Under this definition, since one of the acceptance rates $\alpha(\boldsymbol{x}'|\boldsymbol{x})$ or $\alpha(\boldsymbol{x}|\boldsymbol{x}')$ always take the value of 1 and the other takes $q(\boldsymbol{x}|\boldsymbol{x}')f(\boldsymbol{x}')/q(\boldsymbol{x}'|\boldsymbol{x})f(\boldsymbol{x})$ (or its inverse), it automatically satisfies the modified detailed balance equation in Eq. (5.29).

In our analysis, we use the Gaussian distribution obtained from the best fit to the atmospheric samples as the proposal function. Since the proposal function is symmetric between $\boldsymbol{x}$ and $\boldsymbol{x}'$ under the Gaussian definition ($q(\boldsymbol{x}|\boldsymbol{x}') = q(\boldsymbol{x}'|\boldsymbol{x})$), the acceptance rate will be equal to the ratio of the posterior probabilities at $\boldsymbol{x}$ and $\boldsymbol{x}'$ as

$$\alpha(\boldsymbol{x}'|\boldsymbol{x}) = \min\left(1, \frac{f(\boldsymbol{x}')}{f(\boldsymbol{x})}\right). \tag{5.31}$$

In the actual fitting, the algorithm works in the following way:

1. Choose an initial state $\boldsymbol{x}_0$ with a random seed.

2. Propose the next step $\boldsymbol{x}'$ according the proposal function $q(\boldsymbol{x}'|\boldsymbol{x})$.

3. Decide whether to accept the proposed step $\boldsymbol{x}'$ or not depending on the acceptance rate $\alpha(\boldsymbol{x}'|\boldsymbol{x})$. One can generate a random number a uniformly distributed in $[0,1]$ and define the next step as follows:

$$\boldsymbol{x}_{n+1} = \begin{cases} \boldsymbol{x}' & (a \leq \alpha(\boldsymbol{x}'|\boldsymbol{x}_n)) \\ \boldsymbol{x}_n & (\text{otherwise}) \end{cases}. \tag{5.32}$$

    Note that the acceptance rate is always 1 when the posterior probability at the proposed step is larger than that at the current step, and therefore, the proposed step will always be accepted in that case.

4. Repeat steps 2 and 3 many times until the chain reaches the stationary state.

We should note that the choice of the first step is arbitrary and therefore it could bias our result if we choose the initial point far from the stationary point. Therefore, we need to remove some steps at the beginning of the chain (these steps are called the "burn-in"). More details about the tuning of MCMC can be found in Appendix C.

Multicanonical method   Using the algorithm described above, we can sample the steps directly from the target posterior distribution. However, in this oscillation analysis, certain regions of oscillation parameters (e.g. $\delta_{\mathrm{CP}} \sim \pi/2$ at inverted ordering) get significantly smaller probabilities to be explored. For example, when $\Delta\chi^2$ is above $\Delta\chi^2 > 25$ in some region, the corresponding posterior probability is $< 4 \times 10^{-6}$, which ends up with only a few steps in this region even if we have a few million steps. This causes a large statistical error and makes the posterior distributions around this region unreliable. For $\sin^2\theta_{23}$, it is also important to have enough transitions between the lower and upper octant to get reliable posterior probabilities for both octants. Due to the octant problem of θ_{23}, the posterior probability distribution of $\sin^2\theta_{23}$ is known to have a bimodal shape (see for example Fig. 5.8). In that case, the transition between the two octants can be suppressed due to the lower probability around the valley between the two octants.

To get enough number of steps in all the regions for the parameters of interest (especially for δ_{CP}) and to get enough transitions between octant, we use a so-called "multicanonical method"[3] [158]. To adapt this method in our analysis, we construct a sampling probability distribution $f_{\mathrm{sample}}(x)$ by reweighting the target posterior probability $f(x)$ with the predicted posterior probability $f_{\mathrm{pred.}}(x)$:

$$f_{\mathrm{sample}}(x) = f(x)/(f_{\mathrm{pred.}}(x))^{1-\beta}, \tag{5.33}$$

where $\beta = 1.0$ conserves the sampling distribution as the same as the original posterior probability. Although any predicted probability distributions are acceptable in principle, the reweighting works efficiently when it is closer to the true target posterior probability. In this study, we use the two-dimensional marginal likelihood for δ_{CP}-$\sin^2\theta_{23}$ in both mass ordering (MO) obtained from the importance sampling method as the predicted posterior probability $f_{\mathrm{pred.}}(\delta_{\mathrm{CP}}, \sin^2\theta_{23}, \mathrm{MO})$, which we denote as the "2D multicanonical method" hereafter.

Figure 5.4 shows the posterior distributions of δ_{CP} and $\sin^2\theta_{23}$ from the raw output of MCMC with different values of β. Given a good approximation of the target posterior as the prediction, $\beta = 0.0$ gives an almost uniform (flat) distribution over all the regions of δ_{CP} and $\sin^2\theta_{23}$. In the actual fit, however, we use an intermediate value around $\beta = 0.5$ to avoid losing the steps in the high posterior regions and having an unexpected concentration of the sampling distributions due to the mismatch in $f_{\mathrm{pred.}}(x)$ and $f(x)$.

To get the target posterior distribution from the raw output of the MCMC, we should reweight each step s with the weight

$$w_s = \left[f_{\mathrm{pred.}}(\delta_{\mathrm{CP}s}, \sin^2\theta_{23s}, \mathrm{MO}_s) \right]^{1-\beta}. \tag{5.34}$$

The reweighted posterior distributions from the raw posterior (in Fig. 5.4) are shown in Fig. 5.5, which demonstrates that we can get roughly consistent distributions from the different values of β within statistical fluctuations. In the actual data fit, however, it could give incorrect results when the multicanonical β is not well tuned, so we should tune β properly to obtain a reliable result. More detailed discussions on the tuning of β, including the MCMC convergence diagnostics and autocorrelations, are given in Appendix C.

---

[3]The multicanonical method was originally developed in the context of statistical physics to study the behavior of the system (e.g. Ising model) at the first-order transition point where the transition probability is suppressed.

(a) δ_{CP} (b) $\sin^2 \theta_{23}$
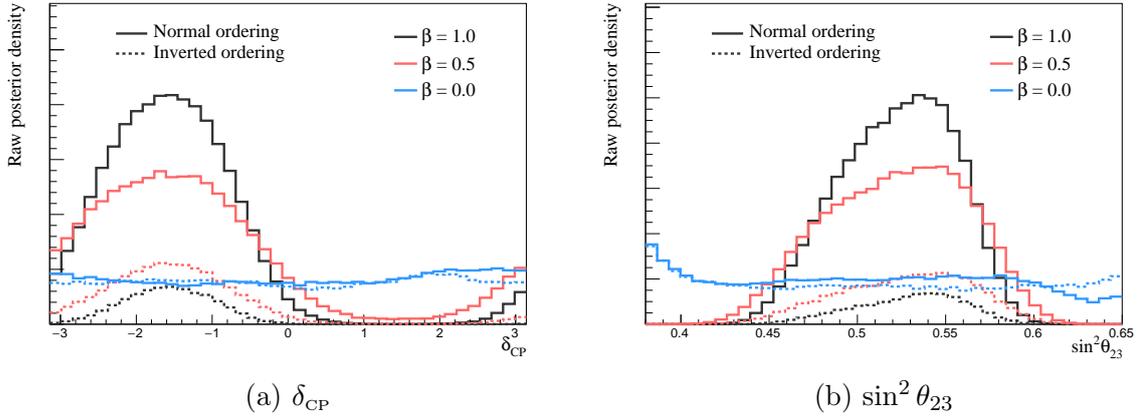
Figure 5.4. Comparison of the raw posterior distributions for normal and inverted ordering with different values of β for the 2D multicanonical method. See Section 5.5.1 for the definition of the MC data set (Asimov A).

(a) δ_{CP} (b) $\sin^2 \theta_{23}$
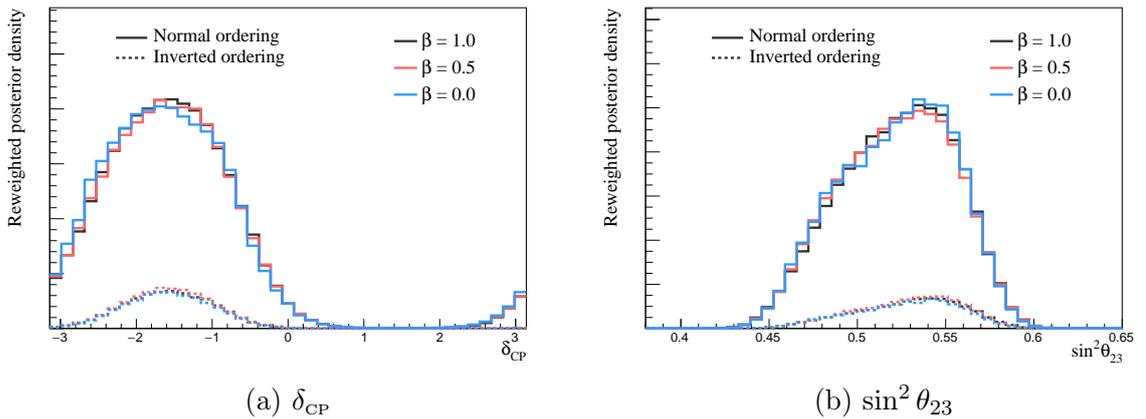
Figure 5.5. Comparison of the reweighted posterior distributions for normal and inverted ordering with different values of β for the 2D multicanonical method. See Section 5.5.1 for the definition of the MC data set (Asimov A).

5.4   T2K near detector analysis

The T2K near detector (ND) data is used to constrain the flux and cross-section systematic uncertainties presented in Chapter 4. The joint analysis between the SK atmospheric and T2K accelerator neutrinos does not modify the ND part of the analysis. Therefore, we do not repeat the ND analysis and simply use the results from the past T2K analysis reported in Ref. [37]. An overview of the ND analysis results is briefly described in the following.

The likelihood used in the ND analysis is defined as

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{stat.}} \times \mathcal{L}_{\mathrm{MC\ stat.}} \times \mathcal{L}_{\mathrm{syst.}}, \tag{5.35}$$

where $\mathcal{L}_{\mathrm{stat.}}$ is the statistical likelihood, $\mathcal{L}_{\mathrm{MC\ stat.}}$ is the MC statistical uncertainty likelihood, and $\mathcal{L}_{\mathrm{syst.}}$ is the prior constraints on the systematic uncertainties. The MC statis-

tical uncertainty likelihood is included to take into account the uncertainty due to using a finite size of MC simulation data set. This is defined as

$$- 2 \log \mathcal{L}_{\text{stat.}} - 2 \log \mathcal{L}_{\text{MC stat.}}$$
$$= 2 \sum_i^{\text{samples}} \sum_j^{\text{bins}} \left[\left(N_{\text{MC}} - N_{\text{Data}} + N_{\text{Data}} \log \frac{N_{\text{Data}}}{N_{\text{MC}}} \right) + \frac{(\beta_j - 1)^2}{2\sigma_{\beta_j}^2} \right], \qquad (5.36)$$

where β_j gives the scaling from the raw MC ($N_{\text{MC}}^{\text{raw}}$) to the weighted MC ($N_{\text{MC}} = \beta_j N_{\text{MC}}^{\text{raw}}$), and σ_j gives the MC statistical uncertainty in bin j. The value of β_j is obtained by analytically solving the following equation at run time:

$$\beta_j^2 + (N_{\text{MC}}^{\text{raw}} \sigma_j^2 - 1)\beta_j - N_{\text{Data}} \sigma_j^2 = 0. \qquad (5.37)$$

This MC statistical uncertainty estimation is based on the method proposed by Barlow and Beeston [159] with an approximation proposed by Conway [160].

The ND fit is performed by maximizing the likelihood defined in Eq. (5.35) over the systematic uncertainties using MINUIT. The flux and cross-section systematic uncertainties are mostly common to the far detector (SK) analysis described in Chapter 4, and the other systematic uncertainty parameters unique to the ND analysis (e.g. detector systematics) are detailed in Ref. [37].

Figure 5.6 shows an example of the pre-fit and post-fit event spectra (projected onto p_μ) of the selected near detector samples. The pre-fit spectrum for the FHC FGD1 ν_μ CC0π sample shows that data is underpredicted by the MC. After the fit, the CCQE component increased while the CC resonant 1π component decreased, which resulted in a better agreement between data and MC. In the RHC FGD2 $\bar{\nu}_\mu$ CC0π sample, the agreement between data and pre-fit MC is not as bad as the FHC FGD1 ν_μ CC0π sample, but it becomes even better after the fit. It is important to note that the momentum (energy) dependence or target dependence (CH or H_2O) of the MC/data discrepancies is also taken into account in the fit, which allows us to achieve different sizes of adjustment in different samples.

The results of these near-detector fits allow us to give precise predictions of flux and cross-sections at SK. To propagate the constraints on the systematic uncertainties to the SK fit, we assume that the systematic uncertainty parameters can be approximated with a multivariate Gaussian function defined in Eq. (5.22).

(a) FHC FGD1 ν_μ CC0π (b) RHC FGD2 $\bar{\nu}_\mu$ CC0π
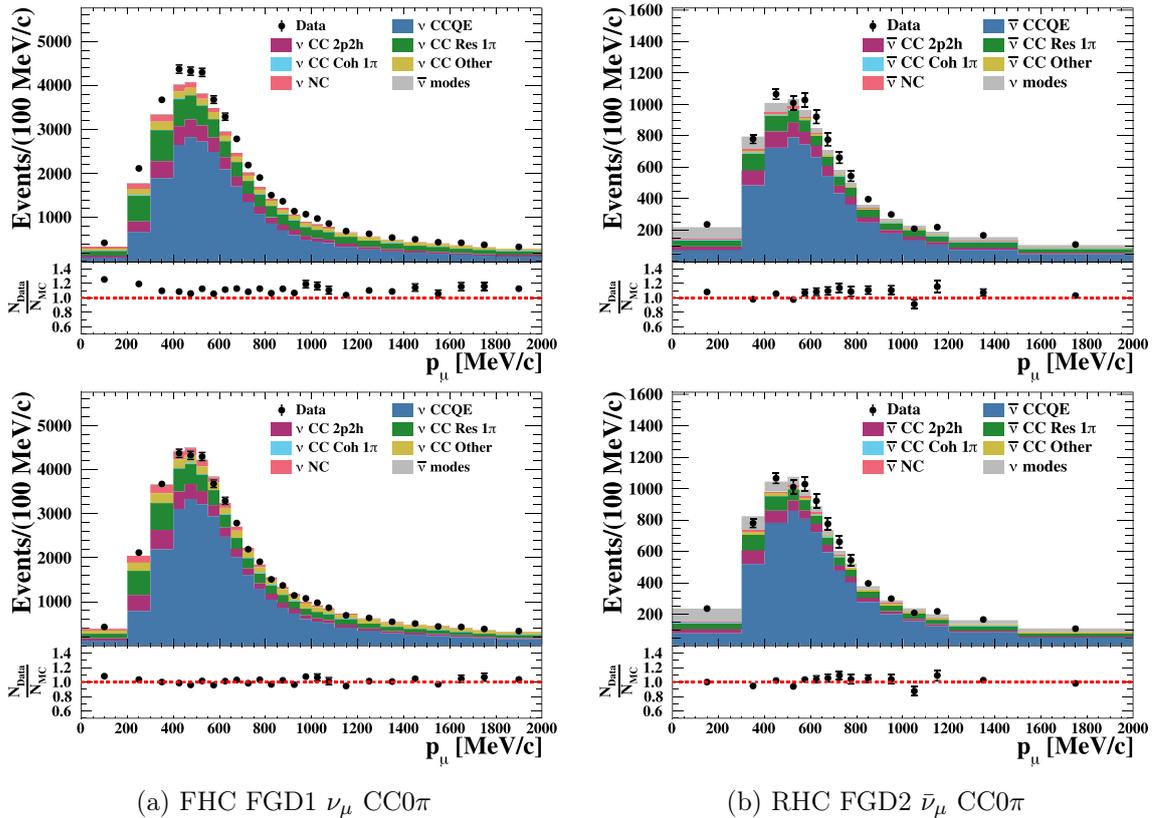
Figure 5.6. The pre-fit (top) and post-fit (bottom) event spectra of selected near detector samples. The data/MC ratio is 1.10 (pre-fit) and 1.00 (post-fit) for the FHC FGD1 ν_μ CC0π sample, and 1.06 (pre-fit) and 1.02 (post-fit) for the RHC FGD2 $\bar{\nu}_\mu$ CC0π sample. The figures and MC/data ratio values are taken from Ref. [37].

5.5   Analysis sensitivity

We evaluate the sensitivity of our analysis using the predefined MC data set before performing the data analysis. The MC data set we use for the sensitivity study is defined in Section 5.5.1, and the results of the sensitivity studies are described in Section 5.5.2.

5.5.1   Definition of the Asimov data set

For the sensitivity study, we use an MC data set constructed by setting the oscillation parameters at a predefined set of values, the beam flux and cross-section systematic uncertainty parameters at the T2K ND post-fit values, and the other systematic uncertainty parameters at the nominal values of the model. The MC data set is produced without having any additional statistical fluctuations and is scaled to be the same statistics as the real data we analyze (see Section 3.3). This data set is called the "Asimov" data set and we assume that the result of the fit to this data set represents the expected sensitivities of our analysis[4]. The predefined sets of oscillation parameters are summarized in Table 5.3.

---

[4] The name is taken after the American writer Isaac Asimov who wrote a short science fiction story in which a computer determines the result of the presidential election. In this story, public opinion is

The Asimov A oscillation parameter set is closer to the best-fit values in the past T2K analysis which takes the maximal violation of CP and upper octant of $\sin^2\theta_{23}$. On the other hand, Asimov B is a hypothetical set of oscillation parameter values complementary to Asimov A where CP is assumed to be conserved and $\sin^2\theta_{23}$ is set to be the lower octant.

Table 5.3. Summary of the Asimov oscillation parameter set used in the MC sensitivity study.

Parameter	Asimov A	Asimov B
$\sin^2\theta_{12}$	0.0307	
$\sin^2\theta_{13}$	0.0218	
$\sin^2\theta_{23}$	0.528	0.45
Δm^2_{21}	7.53×10^{-5} eV^2	
Δm^2_{32} (NO)/ $|\Delta m^2_{31}|$ (IO)	2.509×10^{-3} eV^2	
δ_{CP}	-1.601	0
Mass ordering	Normal ordering	

5.5.2 Result of the sensitivity study

5.5.2.1 Comparison of the different sample fits

Here we perform three types of fits to compare the joint fit sensitivity with the contributions from individual experiments. The performed fits are (1) a fit to only the T2K beam samples (denoted as "T2K"), (2) a fit to only the SK atmospheric samples using the T2K near detector constraints (denoted as "SK (+ND)"), and (3) a joint fit to both the SK atmospheric and T2K beam samples (denoted as "SK + T2K").

Figure 5.7 shows the sensitivity results from the importance sampling method for the oscillation parameter set A. For normal ordering, the constraints on δ_{CP}, $\sin^2\theta_{23}$, and Δm^2_{32} mostly come from the T2K samples. The addition of the SK samples improves the constraints on δ_{CP} and $\sin^2\theta_{23}$ slightly. A more visible effect can be seen in the rejection of the inverted ordering, which appears as the offset in the $\Delta\chi^2$ curves. The constraints on $\sin^2\theta_{13}$ are dominated by the prior constraints from the reactor experiments, and we do not have much improvement in this parameter except for the inverted ordering rejection. Overall, we do not have a significant improvement beyond a simple sum of $\Delta\chi^2$ from the beam and atmospheric samples at this oscillation parameter set A.

Figure 5.8 shows the sensitivity results for the oscillation parameter set B. For δ_{CP}, the T2K-only fit has a weak rejection of $\delta_{\mathrm{CP}} = \pi$ due to the symmetric feature of $\sin\delta_{\mathrm{CP}}$ at $\delta_{\mathrm{CP}} = 0$ (true value) and $\delta_{\mathrm{CP}} = \pi$. However, in the joint fit, we gain stronger rejection power for $\delta_{\mathrm{CP}} = \pi$ thanks to the addition of the SK samples. For $\sin^2\theta_{23}$, we gain a stronger rejection of the wrong octant (upper octant in this case, i.e. $\sin^2\theta_{23} > 0.5$) than the T2K-only fit.

---

represented by only one selected citizen's opinion [161].
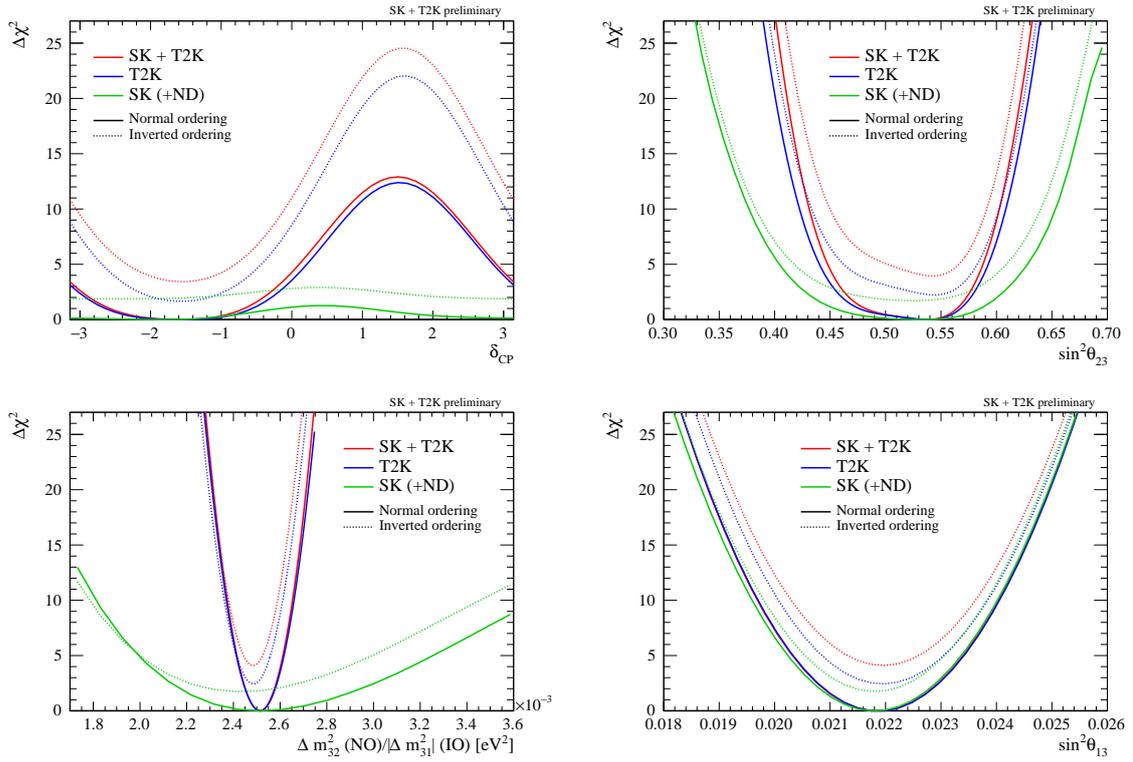
Figure 5.7. Comparison of the Asimov sensitivity at true oscillation parameter set A among the SK + T2K, T2K, and SK (+ND) fits within the joint analysis framework.

5.5.2.2   Power of rejecting CP conservation

Since one of the most important questions we seek to answer in this analysis is whether the CP symmetry is violated, we examined the power of rejecting CP conservation (CPC). Here we construct many Asimov data sets at different values of true δ_{CP} for true normal ordering (NO) and inverted ordering (IO) while fixing the other oscillation parameters at Asimov A. We define the power of CPC rejection using the difference between χ^2 at CP conserving values ($\delta_{\mathrm{CP}} = 0, \pi$) and the minimum χ^2 in all the δ_{CP} regions and both mass ordering (MO):

$$f(\delta_{\mathrm{CP}}^{\mathrm{true}}, \mathrm{MO}^{\mathrm{true}}) = \min_{\alpha} \chi^2(\alpha | \delta_{\mathrm{CP}}^{\mathrm{true}}, \mathrm{MO}^{\mathrm{true}}) - \min_{\delta_{\mathrm{CP}}, \mathrm{MO}} \chi^2(\delta_{\mathrm{CP}}, \mathrm{MO} | \delta_{\mathrm{CP}}^{\mathrm{true}}, \mathrm{MO}^{\mathrm{true}}), \qquad (5.38)$$

where α runs over the four points of the CPC hypothesis: $\delta_{\mathrm{CP}} = 0, \pi$ for NO and IO, respectively.

The results of the power of rejecting CP conservation are shown in Fig. 5.9. Some examples of the fit $\Delta\chi^2$ distributions at different values of true δ_{CP} and different true mass ordering are also shown in Fig. 5.10. In Fig. 5.9a, the SK (+ND) fit shows small CPC rejecting power in all the regions, which implies that the atmospheric samples do not have strong CP sensitivity. The T2K-only fit shows good CPC rejection around $\delta_{\mathrm{CP}} = -\pi/2$ but significantly loses the rejection power around $\delta_{\mathrm{CP}} = \pi/2$. As discussed in Section 2.4, this is due to the degeneracy between δ_{CP} and mass ordering in the T2K measurement. The effect of the degeneracy can be seen in the ($\delta_{\mathrm{CP}}^{\mathrm{true}} = 1.57$, true normal ordering) plot in
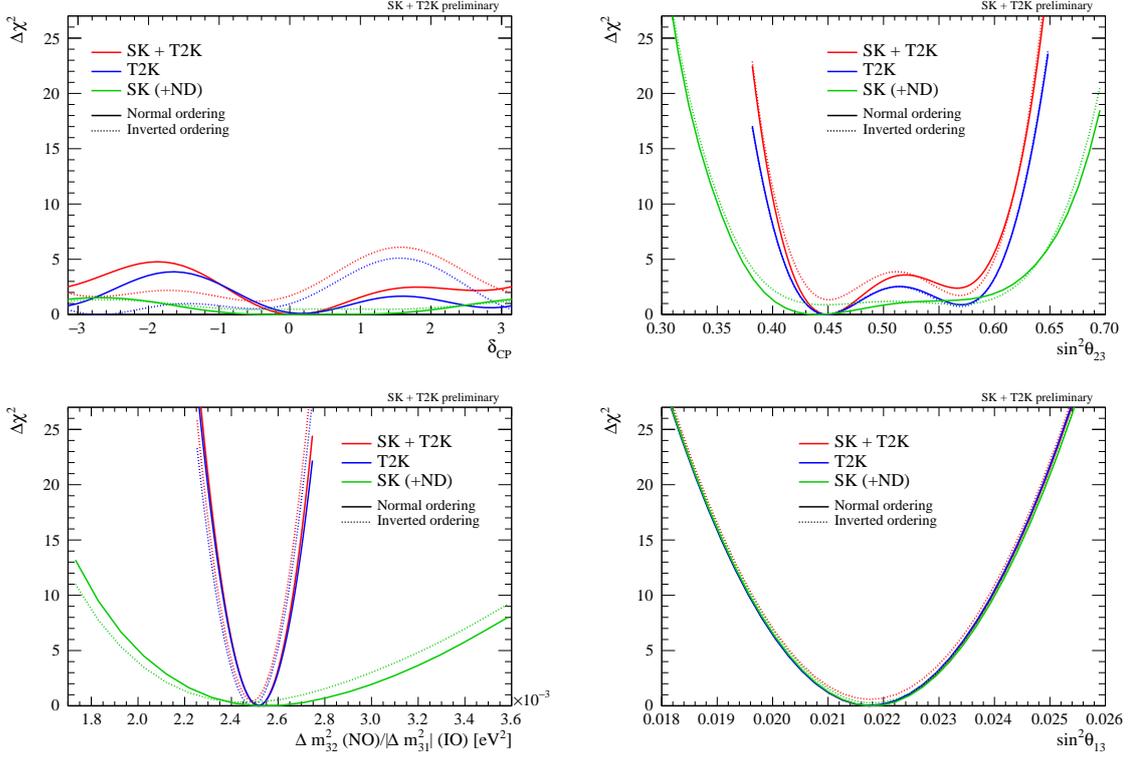
Figure 5.8. Comparison of Asimov sensitivity at true oscillation parameter set B among the SK + T2K, T2K, and SK (+ND) fits within the joint analysis framework.

Fig. 5.10, where the T2K-only fit has similar χ^2 at both the normal and inverted ordering.

This degeneracy, however, can be broken when we fit the T2K and SK samples simultaneously since the atmospheric neutrinos give almost constant wrong-ordering rejection which is independent of true δ_{CP}. This effect appears as an offset in the $\Delta\chi^2$ curves for the wrong ordering (e.g. IO plots in the true NO) in Fig. 5.10. As a result, SK+T2K shows a significant improvement in CPC rejection power around $\delta_{\mathrm{CP}}^{\mathrm{true}} = \pi/2$. The opposite behavior is also seen for the true inverted ordering case as shown in Fig. 5.9b.

The benefit of degeneracy breaking is pronounced at the combinations of a certain region of true δ_{CP} and true mass ordering. For the other regions, the result of the joint analysis becomes closer to a simple sum of $\Delta\chi^2$ from the beam and atmospheric samples.

5.5.2.3 Sample likelihood scanned over oscillation parameters

Using the sensitivity framework, we study how each parameter gets constrained from the samples. For each parameter of interest θ, we scan the sample likelihood defined as

$$-2\log\mathcal{L}_{\mathrm{sample}}(\theta|n, \boldsymbol{\eta} = \boldsymbol{\eta}_0) = -2\log\mathcal{L}_{\mathrm{stat.}}(\lambda(\theta, \boldsymbol{\eta}_0)|n) \tag{5.39}$$

$$= \sum_b^{\mathrm{bin}} 2\left[\lambda_b(\theta, \boldsymbol{\eta}_0) - n_b + n_b\log\frac{n_b}{\lambda_b(\theta, \boldsymbol{\eta}_0)}\right], \tag{5.40}$$

where the nuisance parameters $\boldsymbol{\eta}$ are fixed at the values used to define the Asimov data set and n_b is the number of MC events in each bin b. We call this process the *likelihood*

(a) True normal ordering   (b) True inverted ordering
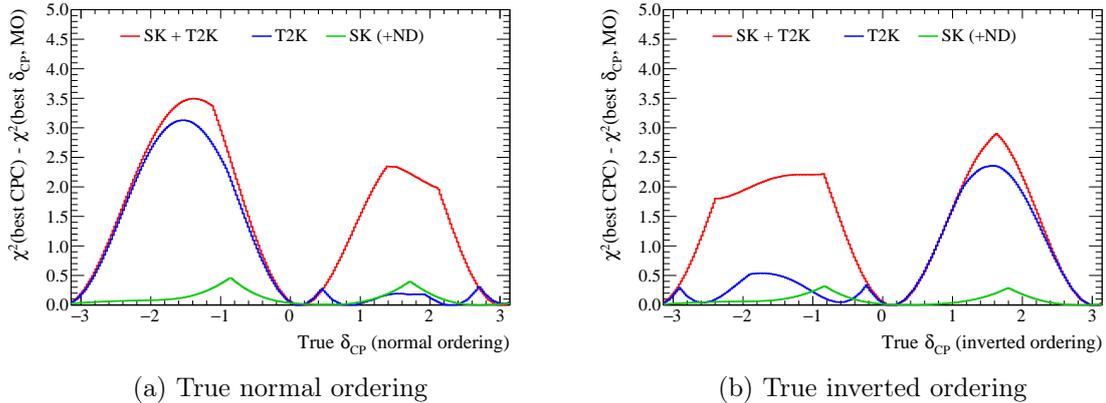
Figure 5.9. Power for rejecting CP conservation (CPC) as a function of the true δ_{CP}.

scan.

The results of the likelihood scan for δ_{CP}, $\sin^2\theta_{23}$, and Δm^2_{32} are shown in Fig. 5.11. Only the total likelihood and the likelihood of the five samples that have the largest contributions are shown. Since the MC data set is constructed using the same parameter tuning, the sample likelihood always takes 0 at the Asimov A value (i.e. $\delta_{\mathrm{CP}} = -1.601$, $\sin^2\theta_{23} = 0.528$, and $\Delta m^2_{32} = 2.509 \times 10^{-3}$ eV^2).

For δ_{CP}, the largest contribution comes from the T2K FHC 1Re and the atmospheric sub-GeV e-like 0de samples. These are mostly dominated by the $\nu_\mu \to \nu_e$ appearance probability which gives the sensitivity to δ_{CP}. The contributions from the RHC 1Re and FHC 1Re 1de samples are relatively small compared to these samples due to the relatively small statistics. For $\sin^2\theta_{23}$ and Δm^2_{32}, the largest contributions come from the T2K FHC/RHC 1Rμ samples. The shapes of the event spectra in these samples are subject to the $\nu_\mu \to \nu_\mu$ disappearance probability, which gives strong constraints to these parameters. Some contributions also come from the atmospheric samples such as the sub-GeV μ-like 1 de sample.

However, we have correlations among these parameters in the actual fit, which is not taken into account in these likelihood scan studies. In addition, the likelihood distribution also depends on the data set n used when computing Eq. (5.39). Therefore, although these studies are helpful to get an idea of how each parameter is constrained by each sample, it is possible that we would get different constraints in the actual data fit.

5.5.2.4 Comparison of the different fitting methods

As described in Section 5.3, we have two methods to compute the marginal likelihood: MCMC and the importance sampling method. These two methods are used to evaluate the same marginal likelihood but are based on different methodologies. Therefore, we make a comparison of the likelihood obtained from these two methods as validation before the data fit.

The comparison of the 1D $\Delta\chi^2$ contours obtained from both the importance sampling method and MCMC are shown in Figs. 5.12 and 5.13 for Asimov A and B, respectively. For MCMC, we use the 2D multicanonical method with $\beta = 0.5$ by taking the 2D δ_{CP}-
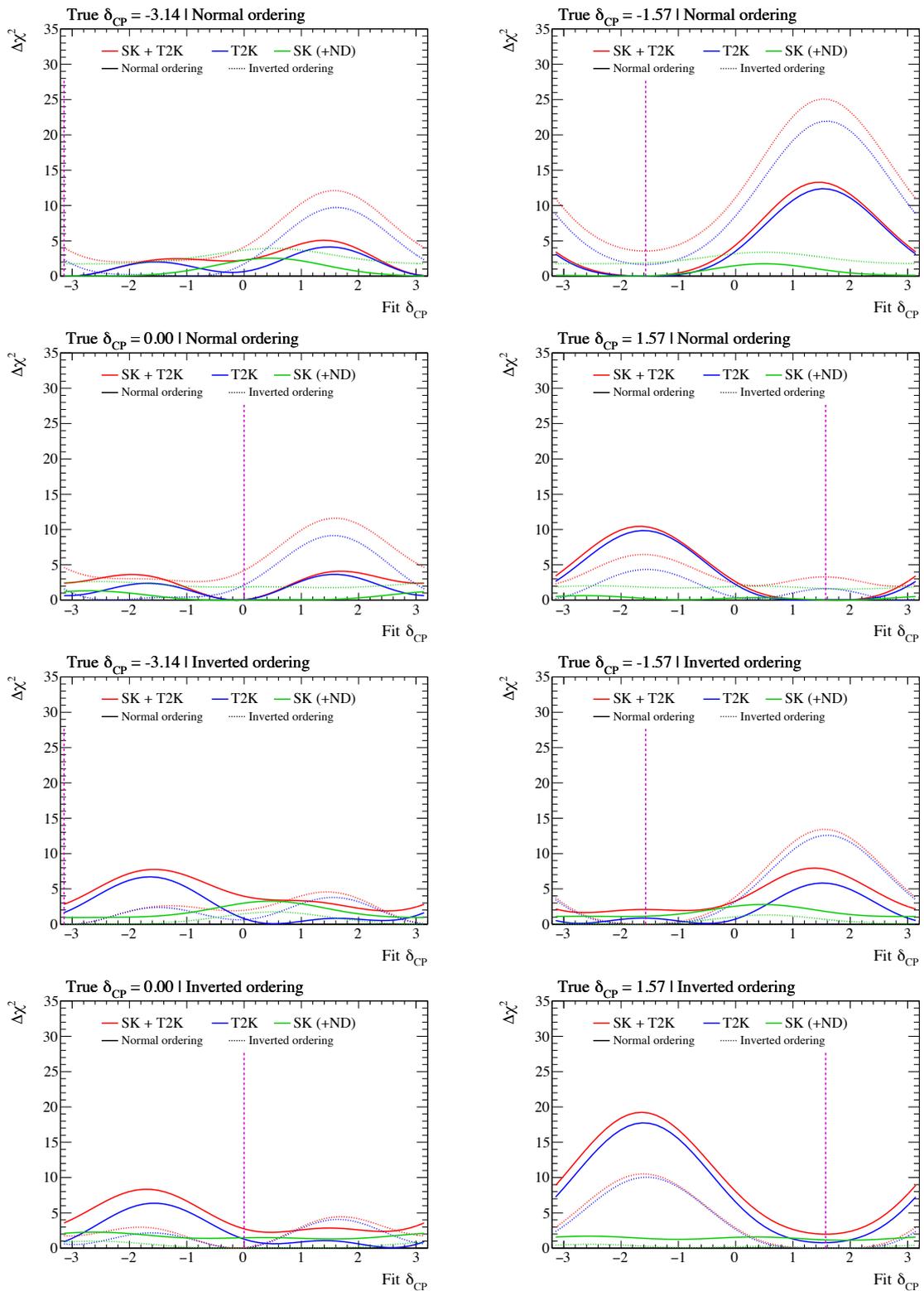
Figure 5.10. Comparison of Asimov sensitivity at different values of true δ_{CP} and different true mass ordering. The other oscillation parameters are fixed at Asimov A. The dotted vertical line shows the true δ_{CP} value.
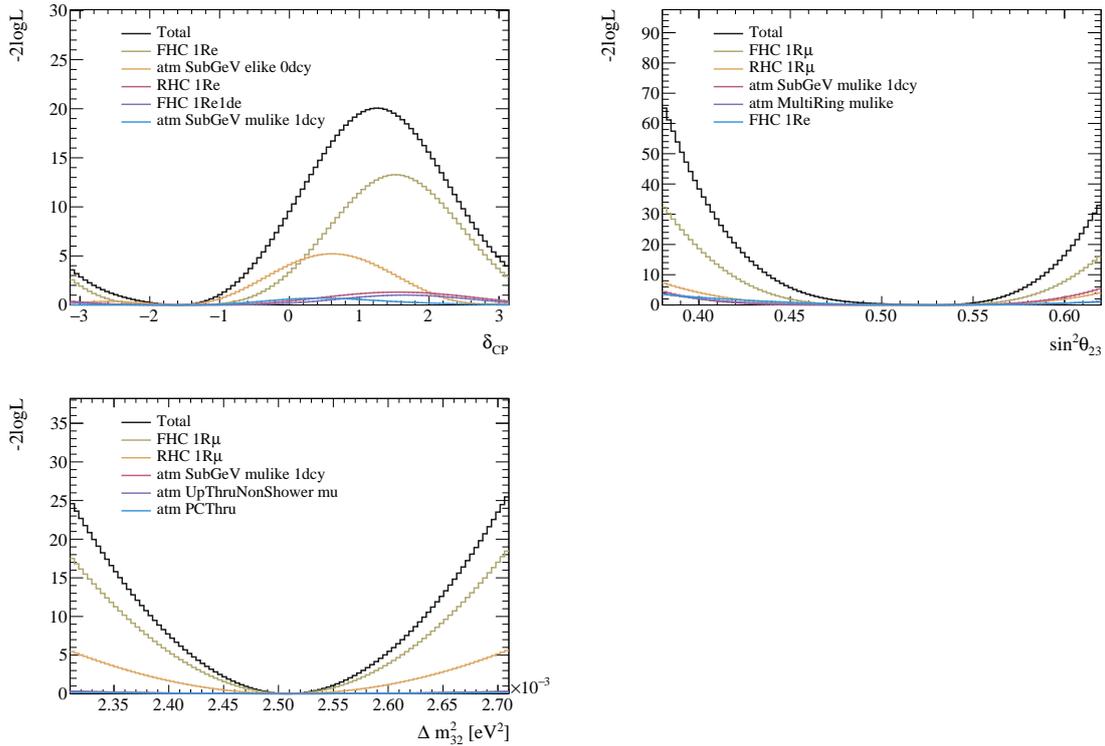
Figure 5.11. Sample likelihood distributions when varying the selected oscillation parameters. While one parameter is varied, the other oscillation parameters and the systematic parameters are fixed at the Asimov A parameter set and the Post-ND tuning. Only the total likelihood and likelihood of the five samples with the largest contributions are shown for each parameter.

$\sin^2 \theta_{23}$ likelihood distributions from the importance sampling method as the predicted distribution. The number of steps used in MCMC is two million steps after removing the burn-in period. The errors are estimated using the jackknife resampling for the importance sampling method and MCMC statistical error with a correction of autocorrelation for MCMC[5].

The overall agreement between the importance sampling and MCMC is good, except for the upper octant region in $\sin^2 \theta_{23}$ for the Asimov B sensitivity (Fig. 5.13c). We think this is because of the insufficient transitions between the upper and lower octant in MCMC. Since we have a deeper valley between the two octants in the posterior distribution of $\sin^2 \theta_{23}$ at Asimo B, it causes a less frequent transition between the upper and lower octants (the steps might have stayed longer in the upper octant in this case). We need to either increase the step size or optimize the value of β to fix this issue.

Figure 5.14 shows the $\sin^2 \theta_{23}$ 1D $\Delta \chi^2$ curve for MCMC with $\beta = 0.0$ (the step size is the same). The agreement between the importance sampling and MCMC becomes better, which implies we can fix the octant transition problem by tuning β properly. We will use $\beta = 0.5$ as the baseline for data fit as it did not show a problem in the Asimov A fit, but a similar validation between the importance sampling and MCMC will be repeated

---

[5]More details on the MCMC error estimation can be found in Appendix C.

in Chapter 7, as there could be more difference in a data fit than the Asimov data set which has more smooth distributions.

We should note that the importance sampling method shows a relatively stable performance in the Asimov fit (MC data set without any statistical fluctuations), but it may have larger errors in the data fit due to the less efficient sampling, which is difficult to fix. On the other hand, it is easier to reduce the MCMC errors as we can simply increase the number of steps. We will use 200 million steps in the actual data fit to get more stable MCMC results ($\times 100$ more than that for the sensitivity study).

(a) δ_{CP}

(b) $\Delta m_{32}^2 / |\Delta m_{31}^2|$

(c) $\sin^2 \theta_{23}$

(d) $\sin^2 \theta_{13}$
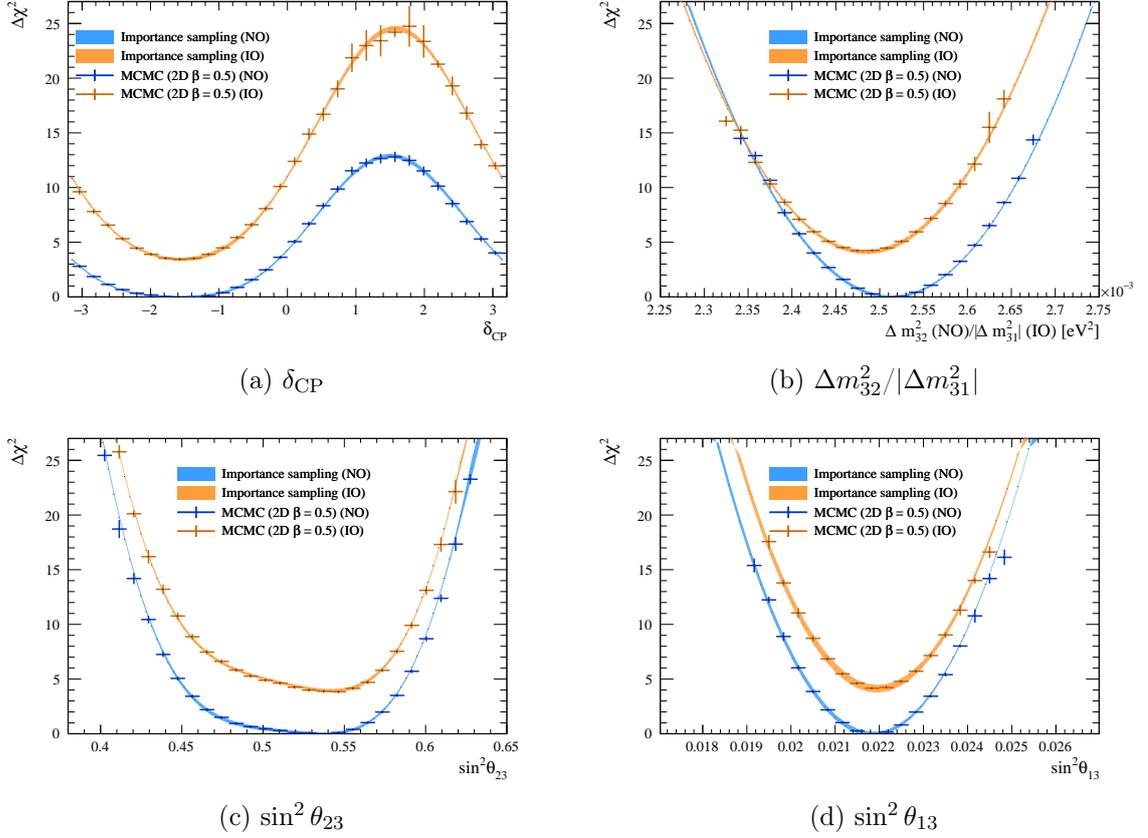
Figure 5.12. Comparison of Asimov A sensitivity 1D $\Delta \chi^2$ contours for the importance sampling method and MCMC. The 2D multicanonical is used for MCMC with $\beta = 0.5$. For ease of interpretation, MCMC $\Delta \chi^2$ values are not shown for the bins that have less than five raw MCMC steps where it has a large MCMC fluctuation.
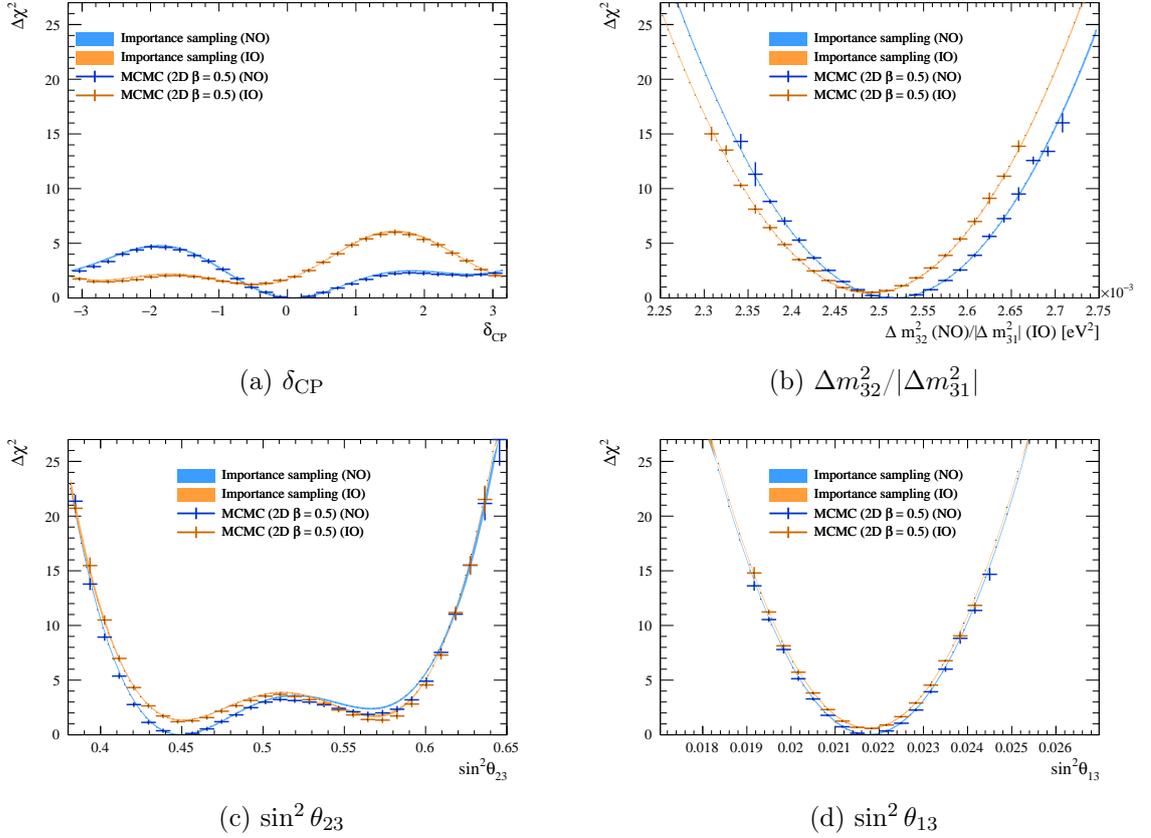
Figure 5.13. Comparison of Asimov B sensitivity 1D $\Delta\chi^2$ contours for the importance sampling method and MCMC. The 2D multicanonical is used for MCMC with $\beta = 0.5$. For ease of interpretation, MCMC $\Delta\chi^2$ values are not shown for the bins that have less than five raw MCMC steps where it has a large MCMC fluctuation.
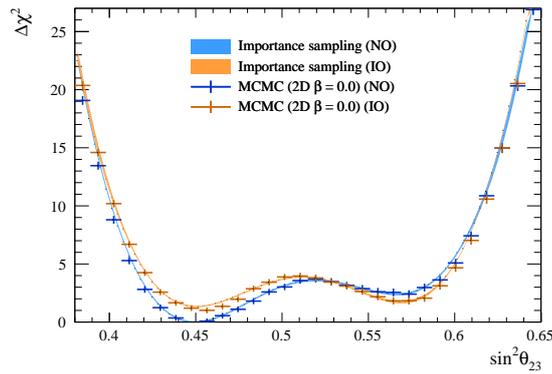
Figure 5.14. Comparison of Asimov B sensitivity $\sin^2\theta_{23}$ 1D $\Delta\chi^2$ contour for the importance sampling method and MCMC. The 2D multicanonical is used for MCMC with $\beta = 0.0$. For ease of interpretation, MCMC $\Delta\chi^2$ values are not shown for the bins that have less than five raw MCMC steps where it has a large MCMC fluctuation.

Chapter 6

Robustness test of the model

6.1 Overview of the robustness test of model

This study is the first joint analysis between the beam and atmospheric neutrinos. Therefore, we developed a combined systematic model for the neutrino fluxes, cross-sections, and detector responses as described in Chapter 4. These systematic models are mostly based on the models used in the individual experiments but include some modifications to take into account the correlations between beam and atmospheric samples. Possible problems that would come with the joint analysis have been addressed during the construction of these models. However, our understanding of neutrino-nucleus interaction cross-sections and sources of other systematic uncertainties are limited in accuracy, and their impact on the oscillation analysis needs to be carefully studied.

In this section, the motivations and procedures of the model robustness test are described including the definitions of the metrics and criteria that need to be fulfilled to ensure our model is robust against the possible out-of-model effects. The alternative models to be tested will be described in Section 6.2 and the results of the robustness test will be discussed in Section 6.3.

6.1.1 Motivations and procedures

In the model robustness test, we test whether we would have a potential bias in the measured oscillation parameters when our systematic model is incorrect (i.e. different from the true nature that the actual data follows). The study is done by generating a "fake" data set by modifying the input MC beyond the uncertainties assigned in our nominal model and fitting them using the nominal systematic model that we use for the actual analysis. This process allows us to test whether a possible out-of-model effect can be absorbed in the current systematic model or whether it would leave a bias in the oscillation parameter measurements. In addition, since the fake data fit is performed at both the T2K near detector (ND) and SK, it also allows us to check whether our extrapolation scheme of the near detector constraints to the SK analysis is robust.

The study is done by first generating fake data sets at the ND and SK. In practice, this is done by reweighting the MC events according to an alternative theoretical model or a data-driven assumption. Then, we fit the T2K ND fake data with our nominal model.

The results of the ND fake data fit are extrapolated to make the predictions at SK and are used in the SK fake data fit to extract the constraints on the oscillation parameters. The SK side of the fit is performed using the importance sampling method instead of MCMC. The importance sampling method produces similar results with MCMC but in 1/10 amount of time and therefore is chosen in this test. Once the fit is performed, we compare the results of the SK fake data fit to the results of the nominal Asimov fit (i.e. we fit the nominal MC Asimov data set with the nominal model) and check whether there would be a bias or not. The metrics and criteria that are used to evaluate the size of the bias will be described in Section 6.1.3. Since we use the "fake" data sets to study the impact of out-of-model effects, this study is called the "Fake Data Study" (FDS). In addition, for some studies, a so-called "scaled Asimov" method is used to mitigate the possible biases due to the change in the statistics in the fake data set. More details on this method and a criterion to apply this treatment are detailed in Section D.1.

The ideal situation is that any out-of-model effects can be absorbed into our nominal systematic model and the oscillation parameters get no bias, but it is difficult to model everything perfectly. When the fake data study fails the predefined criteria, we take the following approaches. If a bias is observed in an oscillation parameter that can be approximated with a Gaussian, we manually inflate the error on the obtained data fit result by applying Gaussian smearing. The size of the smearing factor is computed from the size of the biases we obtained from all the robustness tests. This approach is straightforward but only applicable to the oscillation parameter that has a (near-)parabolic likelihood, which is Δm_{32}^2 in our analysis.

If we observe a bias in the other parameters ($\delta_{\rm CP}$, $\sin^2 \theta_{23}$, and mass ordering), we take another approach, where we add or modify the systematic parameters so that our model can cover such a bias. For example, the extra PID systematics described in Section 4.4.2.3 were added after the first iteration of the robustness test to mitigate the effect of data/MC excess seen in the atmospheric down-going CC1π events.

Finally, after the data fit, we will perform additional tests to see whether our conclusion of CP violation could be affected by the out-of-model effects.

6.1.2   Parameter set

For the fake data studies, we use both the Asimov A and B oscillation parameter sets defined in Table 5.3. Asimov A is closer to the best-fit values of the past T2K analysis and Asimov B is a hypothetical set of values that are complementary to Asimov A. Although we require the fake data studies to pass the criteria at both parameter sets, the result with Asimov A is regarded as the main result of the fake data study, as it is expected to be closer to our data fit results.

For the systematic parameters, we mostly use the set of values used as the prior in the T2K near detector fit, which is called the "Pre-ND" tuning. Since the systematic parameters are mostly implemented near their prior values, this treatment gives better control of the systematic uncertainties that are not relevant to the out-of-model effect that we want to test. However, depending on the alternative model we choose, we need to use different tunings for the construction of the fake data set. One alternative tuning is the set of best-fit values obtained in the near detector real data fit, which is called the "Post-ND"

tuning. In the "Pre-ND" tuning, we also have two possible approaches to deal with the low-Q^2 parameters. The first approach is to define these variables simply to 1 which is the nominal value of the SF prediction and does not change the MC normalization. The other one is to set these parameters to values smaller than 1 to account for the observed deficit in the external cross-section measurements[1]. Therefore, we denote the former as the "Pre-ND ($Q^2 = 1$)" tuning to distinguish it from the latter.

Since we have three tunings of the systematic parameters, the three different nominal Asimov fits are defined by setting the systematic parameters at these tunings. The choice of tuning is mentioned in each fake data study's description in the following section. The intervals obtained from the fit results of each nominal Asimov are described in Section 6.3.

6.1.3 Test metrics and criteria

In this section, the metrics and criteria that we use to evaluate the effects of fake data studies on the oscillation parameters are introduced. For Δm_{32}^2, $\sin^2 \theta_{23}$, and δ_{CP}, we define two metrics to assess the bias in the central values of the oscillation parameters and the change in the size of uncertainties. For mass ordering, we define one metric based on the Bayes factor to assess the change in the preference for normal ordering over inverted ordering. Finally, additional checks will be performed for δ_{CP} and Jarlskog invariant J_{CP} after the data fit.

Bias   The bias and interval size metrics are defined using the confidence intervals obtained from the one-dimensional $\Delta \chi^2$ curves for each oscillation parameter. The fixed-$\Delta \chi^2$ method is used for calculating the confidence interval sizes, where $\Delta \chi^2 = 1$ and $\Delta \chi^2 = 4$ correspond to 1σ and 2σ intervals, respectively[2].

Let us denote the size of the $1(2)\sigma$ interval of the fake data fit as $1(2)\sigma_{\mathrm{tot.}}^{\mathrm{FDS}}$ and that of the reference Asimov fit as $1(2)\sigma_{\mathrm{tot.}}^{\mathrm{ref}}$, where tot. refers to the fit using the total likelihood (Eq. (5.8)). We denote the center of the 2σ confidence interval as $\bar{x}_{2\sigma}$. We perform a statistics-only reference Asimov fit by fixing the systematic parameters at their central values (where the Asimov data set is defined) and compute the statistics-only interval size $1(2)\sigma_{\mathrm{stat.}}^{\mathrm{ref}}$. By combining these two values, we can define the systematic-only uncertainty as

$$1\sigma_{\mathrm{syst.}}^2 \equiv 1\sigma_{\mathrm{tot.}}^2 - 1\sigma_{\mathrm{stat.}}^2. \tag{6.1}$$

The bias metric is defined as the fractional shift in the center of the 2σ confidence intervals between the fake data fit and the reference Asimov fit, divided by the 1σ systematic uncertainty in the reference Asimov fit:

$$B_x^{\mathrm{syst.}} = \frac{\bar{x}_{2\sigma}^{\mathrm{FDS}} - \bar{x}_{2\sigma}^{\mathrm{ref}}}{1\sigma_{\mathrm{syst.}}^{\mathrm{ref}}}. \tag{6.2}$$

---

[1]Here we use the values of [0.495, 0.695, 0.780, 0.890, 0.930] for the low-Q^2 parameters that are tuned to the MINERνA data [117, 137].

[2]This fixed-$\Delta \chi^2$ method does not guarantee the actual frequentist coverage for δ_{CP} and $\sin^2 \theta_{23}$ in the oscillation analysis as Wilks' theorem does not hold in this case. Therefore, the fixed-$\Delta \chi^2$ is only used for the construction of the fake data study metrics and not used for the main analysis.

We require that the observed bias is less significant compared to the estimated size of systematic uncertainty $1\sigma_{\text{syst.}}^{\text{ref}}$. Therefore, using the bias metric defined above, we require the following criterion should be satisfied for $\sin^2\theta_{23}$ and Δm_{32}^2:

$$\left|B_x^{\text{syst.}}\right| < 0.5. \tag{6.3}$$

If this criterion is broken, it means that the observed bias is at a similar level to the systematic uncertainty, and it should be mitigated.

Size of the error   A secondary metric used to assess the change in the size of uncertainty is defined as the relative size of the $2\sigma_{\text{tot.}}^{\text{FDS}}$ interval to the $2\sigma_{\text{tot.}}^{\text{ref}}$ interval:

$$R_x^{2\sigma} = \frac{2\sigma_{\text{tot.}}^{\text{FDS}}}{2\sigma_{\text{tot.}}^{\text{ref}}}. \tag{6.4}$$

We require the change in the interval size should be small compared to the full interval size. Therefore, using the interval size metric defined above, we require the following criterion should be satisfied for δ_{CP}, $\sin^2\theta_{23}$, and Δm_{32}^2:

$$\left|R_x^{2\sigma} - 1\right| < 0.1. \tag{6.5}$$

Mass ordering Bayes factor   The sensitivity to mass ordering is where we expect the main gain in sensitivity compared to the individual beam and atmospheric analyses. Therefore, we define a metric to assess the change in the preference between the normal and inverted orderings. From the results of the marginal likelihood for δ_{CP} at fixed mass ordering, we compute the Bayes factor as follows:

$$\text{BF}_{\text{NO/IO}} = \frac{\mathcal{L}(\text{NO}|x)}{\mathcal{L}(\text{IO}|x)} = \frac{\int_{-\pi}^{\pi} \mathcal{L}(\delta_{\text{CP}}, \text{NO}|x)\pi(\delta_{\text{CP}}|\text{NO})d\delta_{\text{CP}}}{\int_{-\pi}^{\pi} \mathcal{L}(\delta_{\text{CP}}, \text{IO}|x)\pi(\delta_{\text{CP}}|\text{IO})d\delta_{\text{CP}}},$$

where $\mathcal{L}(\text{MO}|x)$ is the marginal likelihood for each mass ordering hypothesis. In this study, the prior distribution for δ_{CP} under each mass ordering hypothesis, $\pi(\delta_{\text{CP}}|\text{MO})$, is defined to be uniform in δ_{CP}. Thus, one can simply integrate the likelihood for all values of δ_{CP}. To assess the change in this Bayes factor, we compute the ratio of them in the fake data fit and Asimov fit as

$$R_{\text{MO}} = \frac{\text{BF}_{\text{NO/IO}}^{\text{FDS}}}{\text{BF}_{\text{NO/IO}}^{\text{ref}}}. \tag{6.6}$$

For changes to the mass ordering preference, we require the following criterion

$$|R_{\text{MO}} - 1| < 0.3. \tag{6.7}$$

This criterion was made by looking at one of the frequently used scales of the Bayes factor: Lee and Wagenmakers' scale [162]. In this scale, change from one category to another roughly corresponds to a factor of 3 increase in Bayes factor as summarized in Table 6.1. Thus, we set the criterion at a 30% change in the Bayes factor ratio taking 10% of this change to the next category.

Bayes factor	Interpretation
1	No evidence
1 to 3	Anecdotal evidence
3 to 10	Moderate evidence
10 to 30	Strong evidence
30 to 100	Very strong evidence
> 100	Extreme evidence

Table 6.1. Lee and Wagenmakers' scale on Bayes factor [162].

$\delta_{\rm CP}$ and $J_{\rm CP}$ interval edges   Due to the cyclic nature and current statistical limitation in the sensitivity to $\delta_{\rm CP}$, the bias metric defined in Eq. (6.2) is not well suited for assessing changes to the $\delta_{\rm CP}$ intervals. Therefore, instead of the bias metric defined in Eq. (6.2), we directly test the effects of the fake data studies on the data fit results after performing the data fit.

The test is done by imposing a shift coming from the fake data effects to the $\delta_{\rm CP}$ and $J_{\rm CP}$ credible interval edges and checking whether it changes our conclusion on CP violation at an interesting level of significance (i.e. 1σ, 2σ, 1.5σ, and 3σ). Let us denote each interval edge as $b_x^{s,\downarrow}$, $b_x^{s,\uparrow}$ for $x = \delta_{\rm CP}, J_{\rm CP}$ and $s = 1\sigma, 1.5\sigma, 2\sigma, 3\sigma$. A shift in each edge, $\Delta b_x^{s,\downarrow}$ and $\Delta b_x^{s,\uparrow}$, is computed by taking the ratio between the posterior distributions obtained from the fake data fit and reference fit at Asimov A and scaling the data fit posterior distributions with that ratio. We should note that we perform this test for both the flat $\delta_{\rm CP}$ prior and flat $\sin\delta_{\rm CP}$ prior as we have different credible intervals corresponding to these priors.

The calculation of the Jarlskog invariant depends on four variables $(\delta_{\rm CP}, \theta_{23}, \theta_{13}, \theta_{12})$, and it cannot be directly computed in the current importance sampling framework that we use for the fake data study. Therefore, we employ some special treatments to obtain the $\delta_{\rm CP}$ and $J_{\rm CP}$ posterior distributions for each fake data fit. Here we assume that the posterior distribution of θ_{13} and θ_{12} is independent of the other two parameters and almost identical to their priors (denoted as $\pi(\cdot)$). On the other hand, $\delta_{\rm CP}$ and $\sin^2\theta_{23}$ have strong correlations, so we cannot separate these two variables. Therefore, the posterior distribution can be written as

$$p(\delta_{\rm CP}, \theta_{23}, \theta_{13}, \theta_{12}) \simeq p(\delta_{\rm CP}, \theta_{23}) \times p(\theta_{13}) \times p(\theta_{12})$$
$$\simeq p(\delta_{\rm CP}, \theta_{23}) \times \pi(\theta_{13}) \times \pi(\theta_{12}). \tag{6.8}$$

Based on this assumption, we generate 10 million sets of these oscillation parameter values and construct the posterior distributions of $\delta_{\rm CP}$ and the Jarlskog invariant. Each parameter is randomly sampled from

$$\pi(\sin^2 2\theta_{13}) \sim \text{Gaussian}(\mu = 0.085, \sigma = 0.0027),$$
$$\pi(\sin^2 \theta_{12}) \sim \text{Gaussian}(\mu = 0.307, \sigma = 0.013),$$
$$p(\delta_{\rm CP}, \sin^2 \theta_{23}) \sim \mathcal{L}(\delta_{\rm CP}, \sin^2 \theta_{23}) \times \pi(\delta_{\rm CP}),$$

where we take both the flat $\delta_{\rm CP}$ prior and flat $\sin\delta_{\rm CP}$ prior for $\pi(\delta_{\rm CP})$. After constructing the posterior distributions, we calculate the shifts induced by each fake data study in each

edge of four credible intervals: $b^{s,\downarrow}_{\delta_{\mathrm{CP}}}$, $b^{s,\uparrow}_{\delta_{\mathrm{CP}}}$, $b^{s,\downarrow}_{J_{\mathrm{CP}}}$, $b^{s,\uparrow}_{J_{\mathrm{CP}}}$ for $s = [1\sigma, 1.5\sigma, 2\sigma, 3\sigma]$. We should note that we use the MCMC result to obtain the data fit posterior distributions of δ_{CP} and J_{CP}. The method described above is only used for the test of fake data effects based on the importance sampling method and it will not be used in our main analysis.

6.2   Alternative models

In this analysis, we tested the robustness of our nominal model using 12 alternative models in total, which are summarised in Table 6.2.

The alternative models can be classified into three categories. The first category is the theory-driven interaction model which is used to cover the possible weakness of our nominal model. The second category is the data-driven model for testing the possible biases coming from the observed data/MC discrepancies. The last category is the analysis validation to test the validity of our analysis method itself.

Theory-driven models   Two alternative nuclear models, Continuum Phase Space Approximation (CRPA) [141, 163] and Local Fermi Gas (LFG) [115], are tested as alternates of the nominal nuclear model, Spectral Function (SF). To test whether our systematic model for the removal (binding) energy can absorb the variation of possible E_b values, a fake data study to test an extreme alternate value of $E_b = 15$ MeV is used instead of the nominal value of 4 (0) MeV for neutrons (protons) in oxygen. For the CCQE cross-section, we assume a dipole-shape axial form factor as discussed in Section 4.3.1.1. There are several possible choices of the form factor parameterizations and here we test the so-called 3-component form factor as an alternative model (expansion of [164]). We use the Nieves model [126] for the 2p2h process but there are also several possible choices of the model that give a different cross-section prediction up to a factor of 2. Here we test the Martini model [127] as an alternative 2p2h model. For the simulation of the multi-pion mode, the probability of having a certain number of outcoming pions is given by the pion multiplicity model. An alternative model can give different predictions of the number of pions and can change the cross-section of the multi-pion mode accordingly. Therefore, we tested an alternative pion multiplicity model to test whether our nominal model can absorb these differences. As discussed in Section 4.3, in the joint analysis, the same electron neutrino cross-section σ_{ν_e} uncertainty parameters (which is defined as scaling from the muon neutrino cross-section σ_{ν_μ}) are applied to both the low-energy and high-energy samples under the assumption that $\sigma_{\nu_e}/\sigma_{\nu_\mu}$ does not have energy dependence. One robustness test is used to test whether it could affect the oscillation parameter measurements in case there is energy dependence in the $\sigma_{\nu_e}/\sigma_{\nu_\mu}$ difference.

Data-driven models   As discussed in Section 4.3.2.2, we observed a data/MC excess in the atmospheric down-going CC1π samples. A possible bias in the oscillation parameters from this excess is tested in the atmospheric CC1π fake data study. Two additional fake data studies are performed to test the effect of MC/data discrepancies observed in the T2K near detector measurements.

Analysis validity test The validity of using the cross-section constraints from the near detector (ND) fit in the low-energy atmospheric samples is tested through the ND extrapolation fake data study. In addition, an effect of a bug found in the simulation of the pion secondary interactions (SI) in the T2K near detector analysis is also tested.

Among these fake data studies, three selected studies from each category are described in detail in the following sections: CRPA, atmospheric CC1π excess, and ND extrapolation. The rest of the studies are summarized in Appendix D.

Table 6.2. Summary of the robustness tests studied in this analysis.

Alternative model name	Model component	Description	Results
CRPA	Nuclear model	Section 6.2.1	Section 6.3.1
LFG	Nuclear model	Section D.2.1	Section D.3
Removal energy	Nuclear model	Section D.2.2	Section D.3
Axial form factors	CCQE	Section D.2.3	Section D.3
Martini 2p2h	2p2h	Section D.2.4	Section D.3
Pion multiplicity	CC$n\pi$	Section D.2.5	Section D.3
Energy-dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$	$\sigma_{\nu_e}/\sigma_{\nu_\mu}$	Section D.2.6	Section D.3
Atmospheric CC1π	CC1π	Section 6.2.2	Section 6.3.2
ND Non-QE CC0π	CC0π	Section D.2.7	Section D.3
ND CC1π	CC1π	Section D.2.8	Section D.3
ND extrapolation	Fit	Section 6.2.3	Section 6.3.3
Pion SI bug fix	CC1π, CC$n\pi$	Section D.2.9	Section D.3

6.2.1 CRPA

The main neutrino interaction target nuclei at SK and T2K ND are oxygen and carbon. Since these nuclei are composed of multiple nucleons, the modeling of the nucleus plays a key role in the prediction of the neutrino interaction cross-sections. In this analysis, we use SF as the baseline model, but there are several other theoretical models proposed, such as Hartree-Fock with Continuum Random Phase Approximation (HF + CRPA) [141], Local Fermi Gas with Random Phase Approximation corrections (LFG + RPA) [115, 116], and Super-Scaling Approach (SuSAv2 [165, 131]). This CRPA model robustness study (and LFG described in Section D.2.1) is developed to test a possible bias from the wrong choice of nuclear model.

In the HF method, a nucleus is described using an effective potential that summarizes all the nucleons in the nucleus. This potential allows us to calculate the neutrino-nucleus cross-section as well as the final state interaction of the outgoing leptons and hadrons inside the nucleus. The model CRPA refers to HF with the Random Phase Approximation which deals with the correction of the W propagator due to the long-range nuclear screening. This HF + CRPA model was found to give a good agreement with the cross-sections of electron scattering measurements [163].

On the other hand, SuSAv2 can calculate the QE cross-section of both the electron scattering and neutrino-nucleus scattering using the scaling functions which gives the scaling from single-nucleon cross-section to multi-nucleon cross-section. SuSAv2 is an updated version of SuSA including more realistic and accurate treatments of relativistic mean field theory [165].

The comparison of the T2K flux integrated ν_μ and $\bar{\nu}_\mu$ double-differential cross-sections for HF (+CRPA), SuSAv2, and LFG (+RPA) is shown in Fig. 6.1. The HF + CRPA model (simply referred to as CRPA hereafter) was found to give different predictions compared to the other nuclear models especially in the low momentum and energy transfer regions (i.e. smaller scattering angles) [166]. Therefore, if CRPA is the true model that describes the nature of the neutrino interaction, using SF as the baseline of the fit could cause a bias in the oscillation parameter measurements.

(a) Neurino (b) Antineutrino
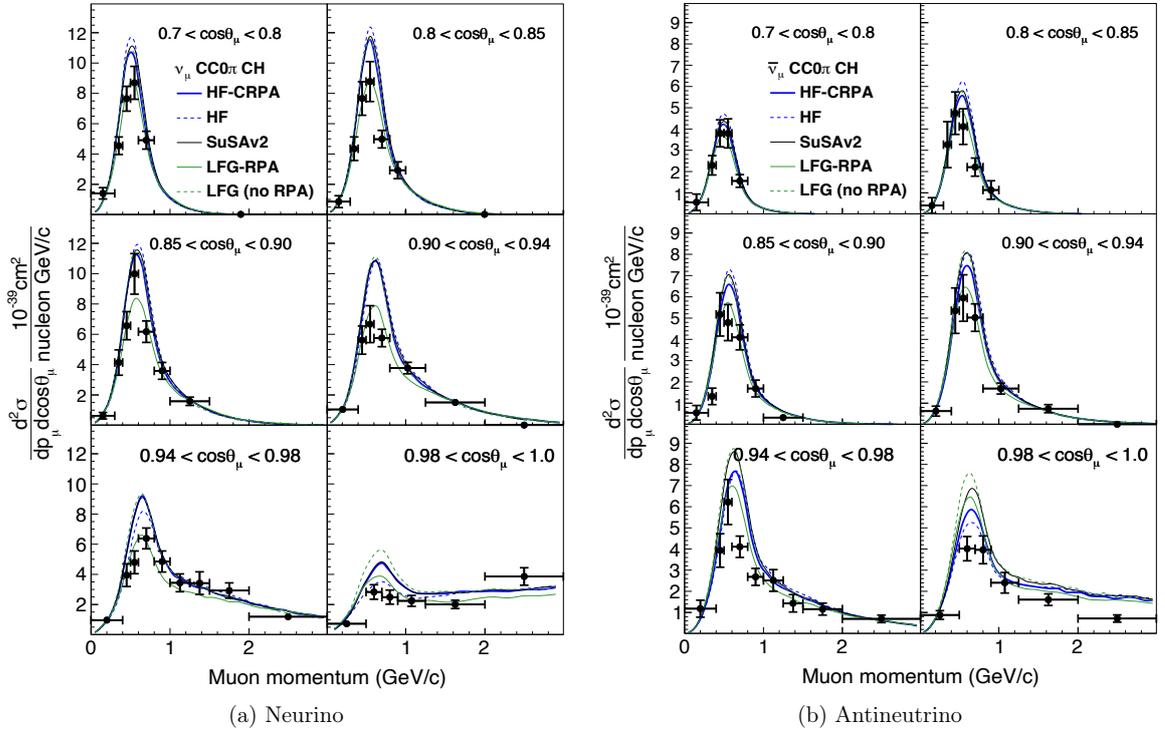
Figure 6.1. Comparison of the T2K flux integrated CC0π ν_μ and $\bar{\nu}_\mu$ double-differential cross-sections on carbon predicted by several 1p1h models. The figures are taken from Ref. [166].

The input MC at the Pre-ND ($Q^2 = 1$) tuning is reweighted to obtain the event spectra predicted by CRPA. Since CRPA is not valid at higher energy transfer q_0, we use CRPA only at $q_0 < 0.5$ GeV and use SuSAv2 at $q_0 \geq 1$ GeV. In the intermediate region of $q_0 \in [0.5, 1.0]$ GeV, we use a linear interpolation between these two models defined as

$$\sigma_{\text{Hybrid}} = \sigma_{\text{CRPA}} + (\sigma_{\text{SuSAv2}} - \sigma_{\text{CRPA}}) \times \frac{q_0 - 0.5}{1.0 - 0.5}. \tag{6.9}$$

The reweighting factors are obtained separately for different neutrino types (neutrino and antineutrino), different neutrino flavors (muon and electron), and different target nucleus

species (carbon and oxygen) by taking the ratio of SF and CRPA cross-section predictions. Since we use the Pre-ND ($Q^2 = 1$) tuning for the generation of this fake data set, we use the nominal Asimov defined at the same Pre-ND ($Q^2 = 1$) tuning for the reference fit.

6.2.2 Atmospheric CC1π excess

This fake data study is developed to test a possible bias from the large data/MC excess seen in the atmospheric down-going ($\cos \Theta_z > 0.4$) sub-GeV CC1π samples. The data/MC comparisons of the down-going events in the e-like and μ-like samples are shown in Fig. 6.2.

(a) Sub-GeV e-like 1de (b) Sub-GeV μ-like 2de
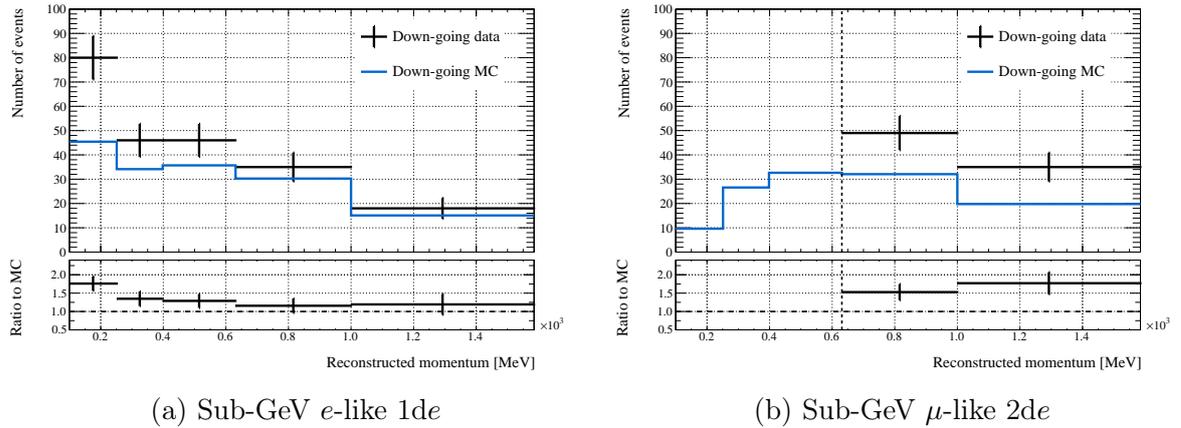
Figure 6.2. The data/MC comparison in the atmospheric sub-GeV CC1π samples down-going events ($\cos \Theta_z > 0.4$). The MC prediction is tuned at the best-fit values of the ND fit and the oscillation probabilities are applied by assuming the oscillation parameter set Asimov A. For the μ-like sample 2de, the data in the momentum region below $p_\ell < 631$ MeV is not shown as it could be sensitive to the oscillation.

The fake data set is constructed by reweighting both the T2K beam CC1π and atmospheric sub-GeV CC1π samples assuming the excess observed in the down-going events will also appear in the entire samples. All the other samples are kept the same. The weights are defined by fitting the data/MC ratio as an exponential function of the reconstructed lepton momentum for e-like samples and as a constant weight for μ-like samples to take into account the shape of the excess:

$$w_{\text{down-going}} = \begin{cases} 2.07 \times \exp(-7.4 \times p_{\text{lep}}^{\text{rec}}/\text{GeV}) + 1.2 & (e\text{-like}) \\ 1.61 & (\mu\text{-like}) \end{cases} \tag{6.10}$$

The data/MC excess is not seen in the T2K near detector data, so here we assume that these effects only appear at SK and do not apply the reweighting to the near detector MC data set. Since we use the Pre-ND tuning for the generation of this fake data set, we use the nominal Asimov fit defined at the same Pre-ND tuning.

6.2.3 Near detector fit extrapolation

As explained in Section 4.3, we use the T2K ND fit results to constrain not only the flux and cross-section uncertainties of the T2K samples but also the cross-section uncertainties

of the atmospheric sub-GeV samples through the common low-energy cross-section model. It is, however, not trivial whether we can apply these constraints to the atmospheric samples. In the typical T2K analysis, the flux and cross-section systematic uncertainties achieve significant anticorrelations after the near detector fit as demonstrated in Fig. 6.3. This anticorrelation helps to reduce the overall uncertainties on the predicted number of events at SK. However, since we only extrapolate the near detector constraints to the cross-section part of the uncertainties in the atmospheric samples, this treatment could cause an unexpected pull in the event rate predictions in these samples.

	$0.0 < E_\nu < 0.4$	$0.4 < E_\nu < 0.5$	$0.5 < E_\nu < 0.6$	$0.6 < E_\nu < 0.7$	$0.7 < E_\nu < 1.0$	M_A^{QE}	$0.00 < Q^2 < 0.05$	$0.05 < Q^2 < 0.10$	$0.10 < Q^2 < 0.15$	$0.15 < Q^2 < 0.20$	$0.20 < Q^2 < 0.25$	$0.25 < Q^2 < 0.50$	$0.50 < Q^2 < 1.00$
$0.50 < Q^2 < 1.00$	-0.13	-0.15	-0.17	-0.19	-0.19	-0.62	0.19	0.31	0.38	0.20	0.41	0.46	1.00
$0.25 < Q^2 < 0.50$	-0.23	-0.27	-0.32	-0.38	-0.37	-0.54	0.39	0.52	0.44	0.52	0.05	1.00	0.46
$0.20 < Q^2 < 0.25$	-0.17	-0.21	-0.24	-0.28	-0.26	-0.17	0.29	0.25	0.55	-0.44	1.00	0.05	0.41
$0.15 < Q^2 < 0.20$	-0.17	-0.21	-0.24	-0.28	-0.26	-0.16	0.27	0.45	-0.13	1.00	-0.44	0.52	0.20
$0.10 < Q^2 < 0.15$	-0.30	-0.37	-0.43	-0.49	-0.46	-0.14	0.55	0.37	1.00	-0.13	0.55	0.44	0.38
$0.05 < Q^2 < 0.10$	-0.36	-0.43	-0.50	-0.59	-0.55	-0.02	0.61	1.00	0.37	0.45	0.25	0.52	0.31
$0.00 < Q^2 < 0.05$	-0.31	-0.38	-0.45	-0.54	-0.52	0.10	1.00	0.61	0.55	0.27	0.29	0.39	0.19
M_A^{QE}	-0.16	-0.20	-0.21	-0.21	-0.17	1.00	0.10	-0.02	-0.14	-0.16	-0.17	-0.54	-0.62
$0.7 < E_\nu < 1.0$	0.49	0.49	0.49	0.71	1.00	-0.17	-0.52	-0.55	-0.46	-0.26	-0.26	-0.37	-0.19
$0.6 < E_\nu < 0.7$	0.54	0.72	0.89	1.00	0.71	-0.21	-0.54	-0.59	-0.49	-0.28	-0.28	-0.38	-0.19
$0.5 < E_\nu < 0.6$	0.67	0.87	1.00	0.89	0.49	-0.21	-0.45	-0.50	-0.43	-0.24	-0.24	-0.32	-0.17
$0.4 < E_\nu < 0.5$	0.86	1.00	0.87	0.72	0.49	-0.20	-0.38	-0.43	-0.37	-0.21	-0.21	-0.27	-0.15
$0.0 < E_\nu < 0.4$	1.00	0.86	0.67	0.54	0.49	-0.16	-0.31	-0.36	-0.30	-0.17	-0.17	-0.23	-0.13

Figure 6.3. Correlations between FHC ν_μ flux and CCQE cross-section uncertainties constrained by the T2K near detector fit. The neutrino energy E_ν and four-momentum transfer Q^2 ranges are shown in the unit of GeV and GeV^2, respectively. The flux and cross-section parameters have clear anticorrelations. The figure is taken from Ref. [37].

This fake data study is used to test the possible bias coming from using the near detector constraints to the atmospheric samples. The fake data set is constructed using the two systematic parameter tunings. The beam part of the fake data set is constructed by using the Post-ND tuning (best-fit values of the T2K ND fit to the real data, not to the simulated Asimov data set). The atmospheric part is then constructed by using the values of the special ND data fit which is done by fixing the flux parameters at their prior values. In this special ND fit, the systematic effects are absorbed only by the cross-section parameters. The atmospheric fake data, therefore, are tuned to these modified cross-section parameters that do not have correlations with the beam flux parameters. Finally, we fit the fake data set using the covariance matrix obtained from the normal ND data fit, which has the anticorrelations between flux and cross-section systematics. If this fake data fit shows a bias in the measured oscillation parameter values, it would imply that we cannot simply apply the near detector constraints to the atmospheric samples due to the correlations between beam flux and cross-section uncertainties. Since we use

the Post-ND tuning for the generation of this fake data set, we use the nominal Asimov fit defined at the same Post-ND tuning.

6.3 Result of the robustness test

In this section, the result of each fake data study (FDS) discussed in Section 6.2 are detailed. For the oscillation parameters, the one-dimensional (1D) $\Delta\chi^2$ curves are reported along with the changes and biases observed in every FDS.

As described in Section 6.1.1, the results of the fake data fits are compared to the results of the corresponding nominal Asimov fit. The three different nominal Asimov fits are defined by setting the systematic parameters at Pre-ND tuning, Pre-ND ($Q^2 = 1$) tuning, and Post-ND tuning. The intervals obtained from the fit results of each nominal Asimov are shown in Fig. 6.4, and the summary tables are shown in Section D.3. For the nominal Asimov at Pre-ND (with and without $Q^2 = 1$), the $\Delta\chi^2$ surface of δ_{CP} is below the 2σ threshold in all the regions in the Asimov B fit and the bias metric cannot be defined properly. Therefore, we do not require δ_{CP} to pass the criterion of the bias metric at Asimov B. In fact, for the CP conclusions, we will later investigate whether the fake data effects would affect the data fit results, which is supposed to be supplemental to the bias check.
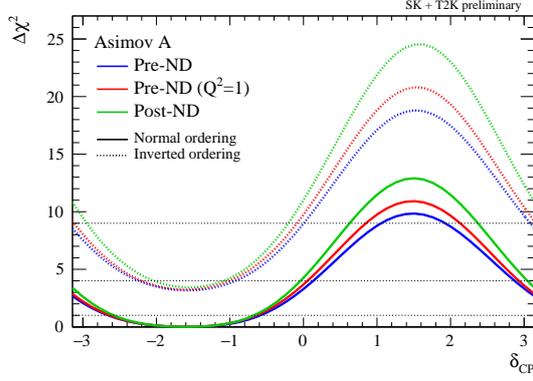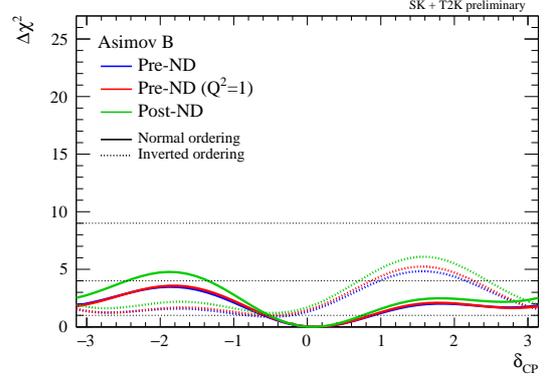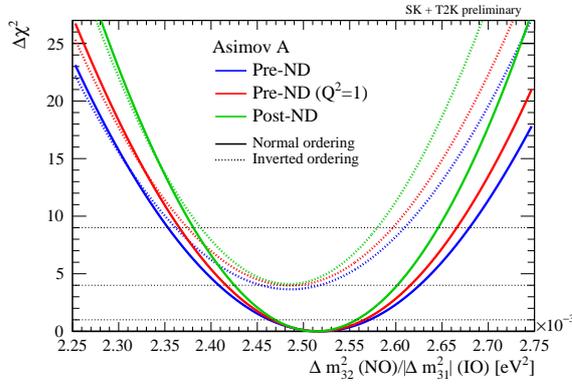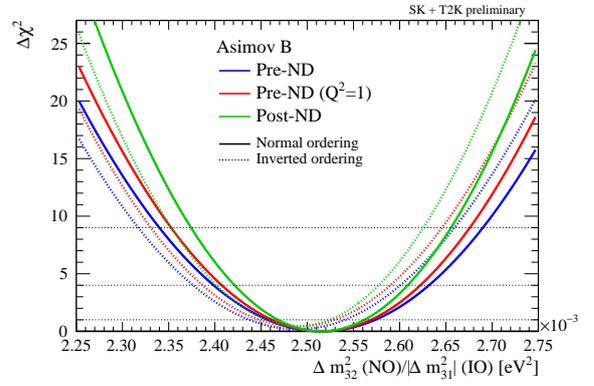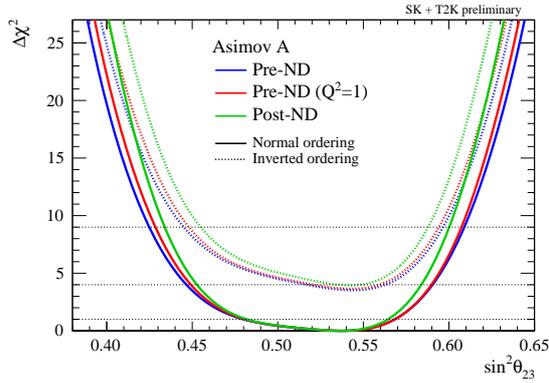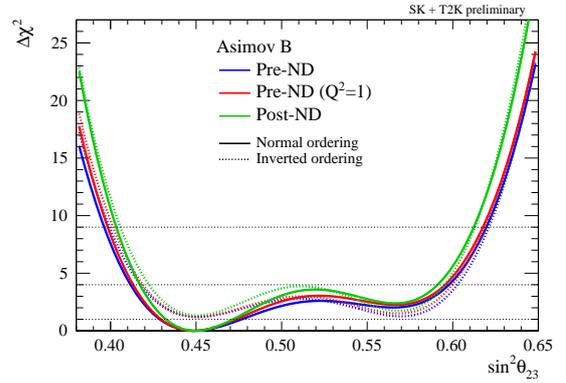
Figure 6.4. 1D $\Delta\chi^2$ curves for the nominal Asimov fit with different systematic parameter tunings. The true oscillation parameter set of Asimov A is used on the left column and Asimov B is used on the right column. The results of the nominal Asimov defined at Pre-ND tuning, Pre-ND ($Q^2 = 1$) tuning, and Post-ND tuning are shown in red, blue, and green, respectively.

6.3.1   CRPA

The generated FDS spectra and ND prediction error bands are shown in Fig. 6.5. The event spectra of the nominal Asimov (green solid line), the fake data (blue solid line), and the prediction from the ND fit to the CRPA fake data (red band) are compared. As CRPA gives different predictions of CCQE cross-sections at low momentum transfer, the fake data event spectra have different shapes compared to the nominal spectra mostly below 1 GeV in the T2K FHC/RHC 1Re/μ samples. The predictions from the ND fake data fit capture these features only partially, and the fake data spectra are not well covered by the predictions.

(a) FHC 1Rμ (b) RHC 1Rμ

(c) FHC 1Re (d) RHC 1Re (e) FHC 1Re 1de
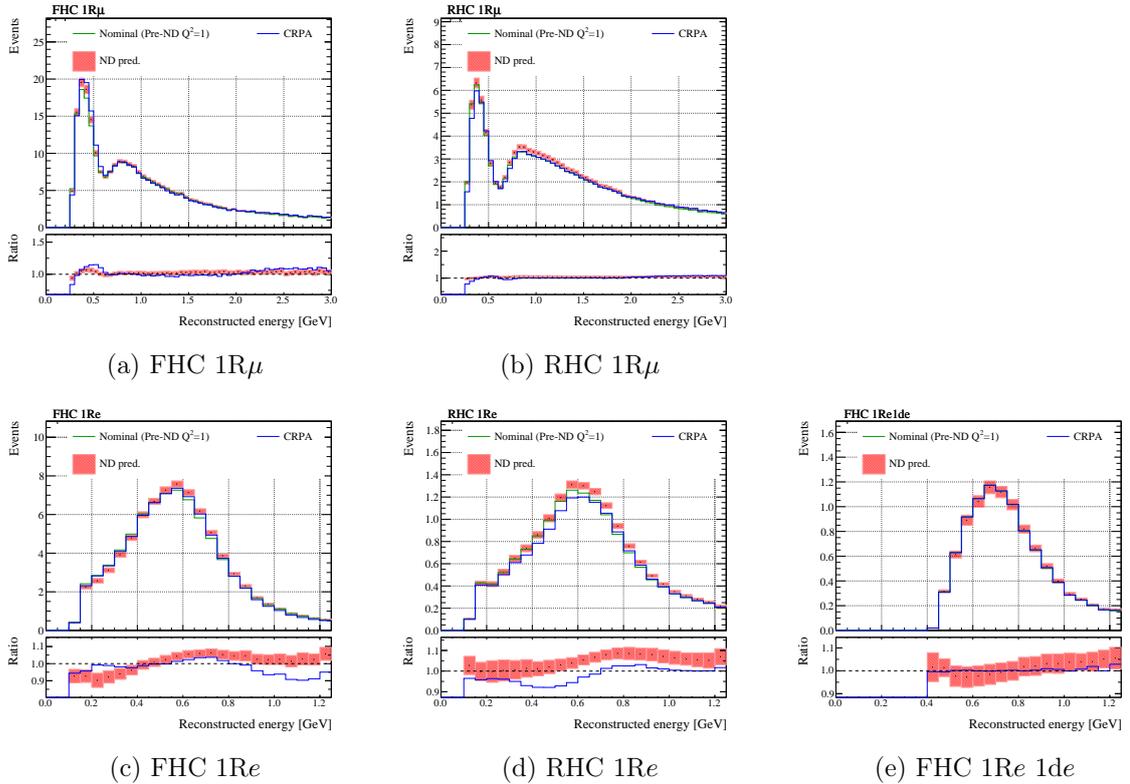
Figure 6.5. Comparison between the nominal event spectra at Pre-ND ($Q^2 = 1$) (solid green) and the CRPA fake data event spectra (solid blue) at the Asimov A oscillation parameter set. The prediction from the T2K near detector (ND) fake data fit is shown with the red band. The bottom panel shows the ratio to the nominal spectra.

The 1D $\Delta\chi^2$ curves for $\sin^2\theta_{23}$, Δm^2_{32}, and δ_{CP} are shown in Fig. 6.6, and the size of the intervals and biases are quantified in Table 6.3. This fake data study gives a significant bias on Δm^2_{32} and fails the FDS criterion. We considered adding an extra systematic for which the 1σ variation would be this fake data study but did not adopt this approach because this FDS is just one possible variation at low transferred energy. Therefore, we simply apply smearing to Δm^2_{32}. For $\sin^2\theta_{23}$, it has a deeper valley shape around the maximal mixing point in Asimov B fit, and the 2σ intervals are divided into two parts. The interval size is therefore defined as the sum of two 2σ intervals. This causes a slightly

smaller interval ratio when compared to the nominal, but it still satisfies the criterion of the interval ratio. Therefore, we do not require an action for $\sin^2\theta_{23}$.

The mass ordering Bayes factors for the fake data fit and reference Asimov fits and their ratios are summarized in Table 6.4 and they satisfy the criteria.

(a) δ_{CP} (Asimov A)

(b) δ_{CP} (Asimov B)

(c) Δm_{32}^2 (Asimov A)

(d) Δm_{32}^2 (Asimov B)

(e) $\sin^2\theta_{23}$ (Asimov A)

(f) $\sin^2\theta_{23}$ (Asimov B)

Figure 6.6. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND ($Q^2 = 1$) and the CRPA fake data fit. The left column shows the results at the oscillation parameter set Asimov A and the right column shows the results at Asimov B.

Table 6.3. Summary of the bias for the CRPA fake data study at both Asimov A and B. The values of interest are highlighted in blue and the values that break the criterion are highlighted in red.

		δ_{CP}	Δm^2_{32}	$\sin^2\theta_{23}$
AsimovA	Middle of the 1σ interval	-1.66	0.002541	0.5206
	1σ interval size: 1σ	0.9982	4.77e-05	0.04787
	1σ interval ratio to Nominal	1.013	0.949	1.156
	Fractional change in the 1σ interval size wrt the syst interval $\left(1\sigma^{\mathrm{FDS}}_{\mathrm{tot.}} - 1\sigma^{\mathrm{ref}}_{\mathrm{tot.}}\right)/1\sigma^{\mathrm{ref}}_{\mathrm{syst.}}$	2.1%	-12%	31.6%
	Bias in the middle of 1σ interval wrt the size of the $1\sigma^{\mathrm{ref}}_{\mathrm{tot.}}$ interval	0.465%	54%	-13.6%
	Bias in the middle of 1σ interval wrt the size of the $1\sigma^{\mathrm{ref}}_{\mathrm{syst.}}$ interval	0.774%	128%	-27.6%
	Middle of the 2σ interval	-1.678	0.002541	0.5164
	2σ interval size: 2σ	1.701	9.55e-05	0.07159
	2σ interval ratio to Nominal	1	0.9486	1.031
	Fractional change in the 2σ interval size wrt the syst interval $\left(2\sigma^{\mathrm{FDS}}_{\mathrm{tot.}} - 2\sigma^{\mathrm{ref}}_{\mathrm{tot.}}\right)/2\sigma^{\mathrm{ref}}_{\mathrm{syst.}}$	-7.18e-14%	-12%	6.21%
	Bias in the middle of 2σ interval wrt the size of the $2\sigma^{\mathrm{ref}}_{\mathrm{tot.}}$ interval	-1.05%	27%	-4.41%
	Bias in the middle of 2σ interval wrt the size of the $2\sigma^{\mathrm{ref}}_{\mathrm{syst.}}$ interval	-1.93%	63.3%	-8.88%
	Bias in the middle of 2σ interval wrt the size of the $1\sigma^{\mathrm{ref}}_{\mathrm{tot.}}$ interval	-1.82%	54.1%	-7.4%
	Bias in the middle of 2σ interval wrt the size of the $1\sigma^{\mathrm{ref}}_{\mathrm{syst.}}$ interval	-3.03%	128%	-15%
AsimovB	Middle of the 1σ interval	0.06925	0.002548	0.4481
	1σ interval size: 1σ	0.6841	5.052e-05	0.0198
	1σ interval ratio to Nominal	0.9663	0.9457	0.884
	Fractional change in the 1σ interval size wrt the syst interval $\left(1\sigma^{\mathrm{FDS}}_{\mathrm{tot.}} - 1\sigma^{\mathrm{ref}}_{\mathrm{tot.}}\right)/1\sigma^{\mathrm{ref}}_{\mathrm{syst.}}$	-5.42%	-12.9%	-20.6%
	Bias in the middle of 1σ interval wrt the size of the $1\sigma^{\mathrm{ref}}_{\mathrm{tot.}}$ interval	-11.9%	60.9%	-16.9%
	Bias in the middle of 1σ interval wrt the size of the $1\sigma^{\mathrm{ref}}_{\mathrm{syst.}}$ interval	-19.1%	145%	-30.1%
	Middle of the 2σ interval	1.237	0.002547	0.4635 and 0.5605
	2σ interval size: 2σ	3.053	0.0001011	0.05076 and 0.03101
	2σ interval ratio to Nominal	0.9718	0.9458	-
	Fractional change in the 2σ interval size wrt the syst interval $\left(2\sigma^{\mathrm{FDS}}_{\mathrm{tot.}} - 2\sigma^{\mathrm{ref}}_{\mathrm{tot.}}\right)/2\sigma^{\mathrm{ref}}_{\mathrm{syst.}}$	-3.41%	-13%	-
	Bias in the middle of 2σ interval wrt the size of the $2\sigma^{\mathrm{ref}}_{\mathrm{tot.}}$ interval	-	30.4%	-
	Bias in the middle of 2σ interval wrt the size of the $2\sigma^{\mathrm{ref}}_{\mathrm{syst.}}$ interval	-	72.7%	-
	Bias in the middle of 2σ interval wrt the size of the $1\sigma^{\mathrm{ref}}_{\mathrm{tot.}}$ interval	-	60.7%	-
	Bias in the middle of 2σ interval wrt the size of the $1\sigma^{\mathrm{ref}}_{\mathrm{syst.}}$ interval	-	144%	-

Table 6.4. Summary of the Bayes factors for the CRPA fake data fit and reference Asimov fit, and their ratio.

	Fake data	Reference	Ratio (fake data/reference)
Asimov A	7.02	7.08	0.99
Asimov B	1.33	1.29	1.04

6.3.2   Atmospheric CC1π excess

The generated FDS spectra and ND prediction error bands are shown in Fig. 6.7. In this fake data study, the reweighting is applied only to the CC1π-enhanced samples (i.e. T2K FHC 1Re1de, atmospheric sub-GeV e-like 1de, and μ-like 2de), and clear differences are seen in the event spectra compared to the nominal.

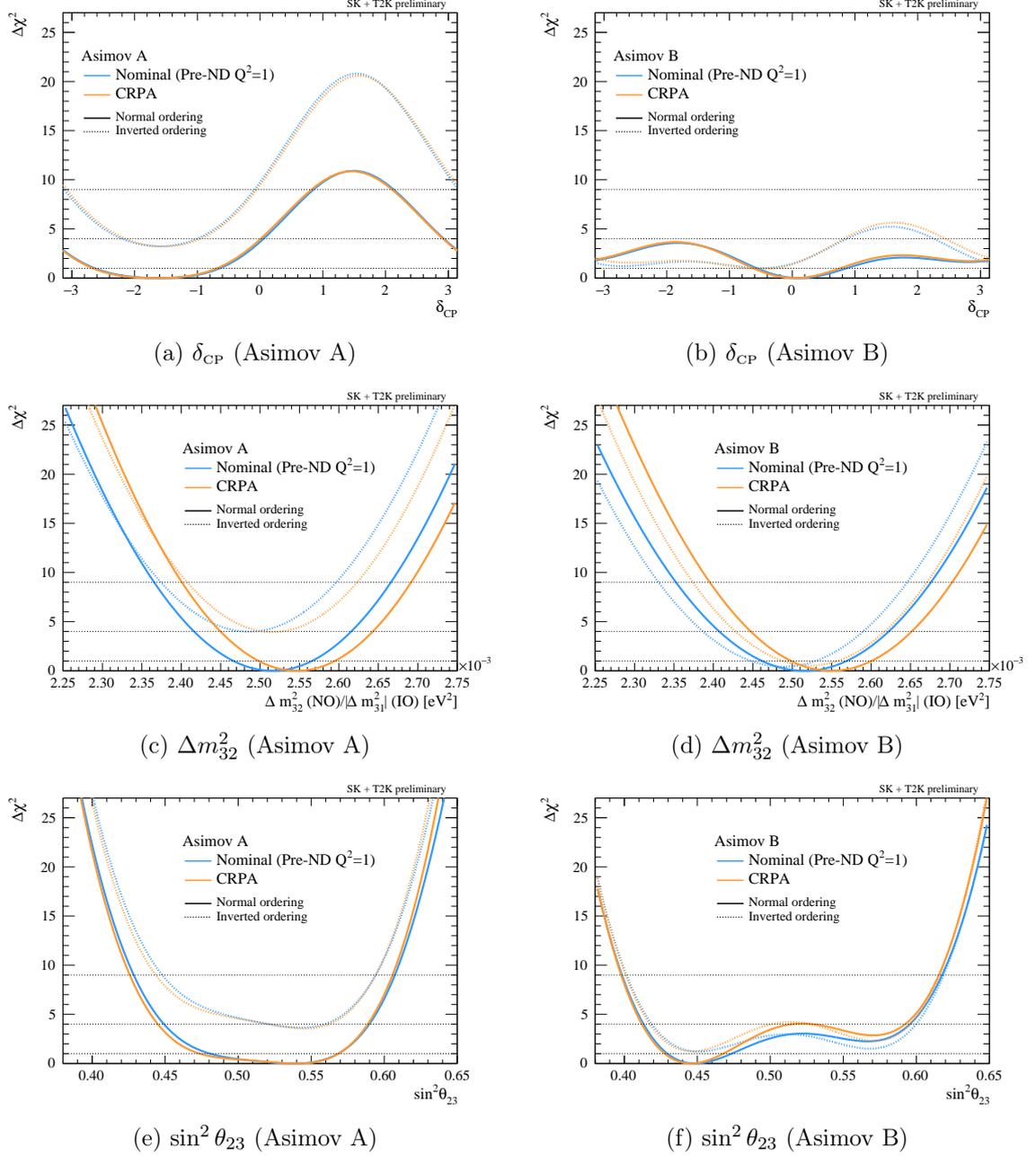(a) FHC 1Re 1de (b) Sub-GeV e-like 1 de (c) Sub-GeV μ-like 2 de
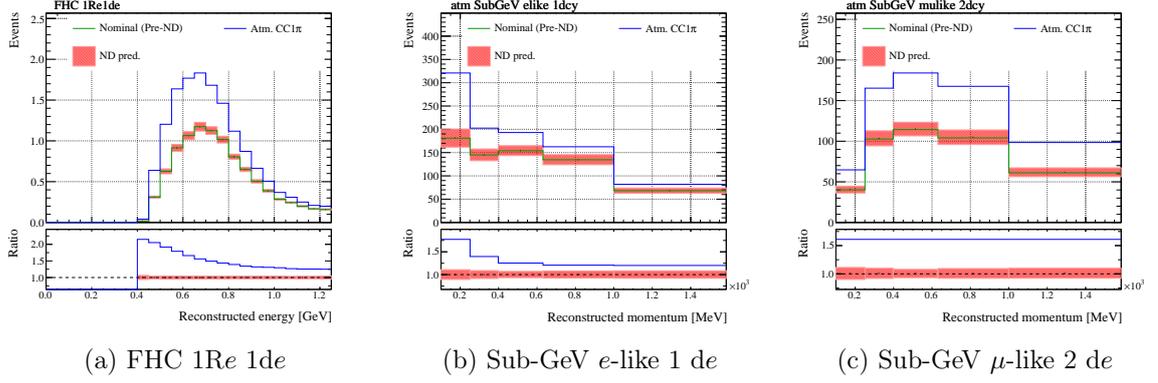
Figure 6.7. Comparison between the nominal event spectra at Pre-ND (solid green) and the atmospheric CC1π fake data event spectra (solid blue) at the Asimov A oscillation parameter set. The prediction from the T2K near detector (ND) fake data fit is shown with the red band. The bottom panel shows the ratio to the nominal spectra.

The 1D $\Delta\chi^2$ curves for $\sin^2\theta_{23}$, Δm^2_{32}, and $\delta_{\rm CP}$ are shown in Fig. 6.8, and the size of the intervals and biases are quantified in Table 6.5. The mass ordering Bayes factor for fake data fit and reference Asimov fits and their ratios are summarized in Table 6.6.

The $\delta_{\rm CP}$ contour in the Asimov B fit goes above the $\Delta\chi^2 = 4$ line and the 2σ intervals are divided into two parts. The interval size is therefore defined as the sum of two 2σ intervals. This causes a slightly smaller interval ratio when compared to the nominal, where all the $\delta_{\rm CP}$ regions $[-\pi, \pi]$ are considered as the 2σ interval in the nominal Asimov B fit. Although it breaks the criterion on the interval ratio for $\delta_{\rm CP}$, the value of the metric is very close to the criterion, and it only happens in Asimov B fit which is far from the past T2K best-fit point. Therefore, we do not require further action for $\delta_{\rm CP}$.

(a) δ_{CP} (Asimov A)

(b) δ_{CP} (Asimov B)

(c) Δm_{32}^2 (Asimov A)

(d) Δm_{32}^2 (Asimov B)

(e) $\sin^2 \theta_{23}$ (Asimov A)

(f) $\sin^2 \theta_{23}$ (Asimov B)

Figure 6.8. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND and the atmospheric CC1π fake data fit. The left column shows the results at the oscillation parameter set Asimov A and the right column shows the results at Asimov B.

Table 6.5. Summary of the bias for the atmospheric CC1π fake data study at both Asimov A and B. The values of interest are highlighted in blue and the values that break the criterion are highlighted in red.

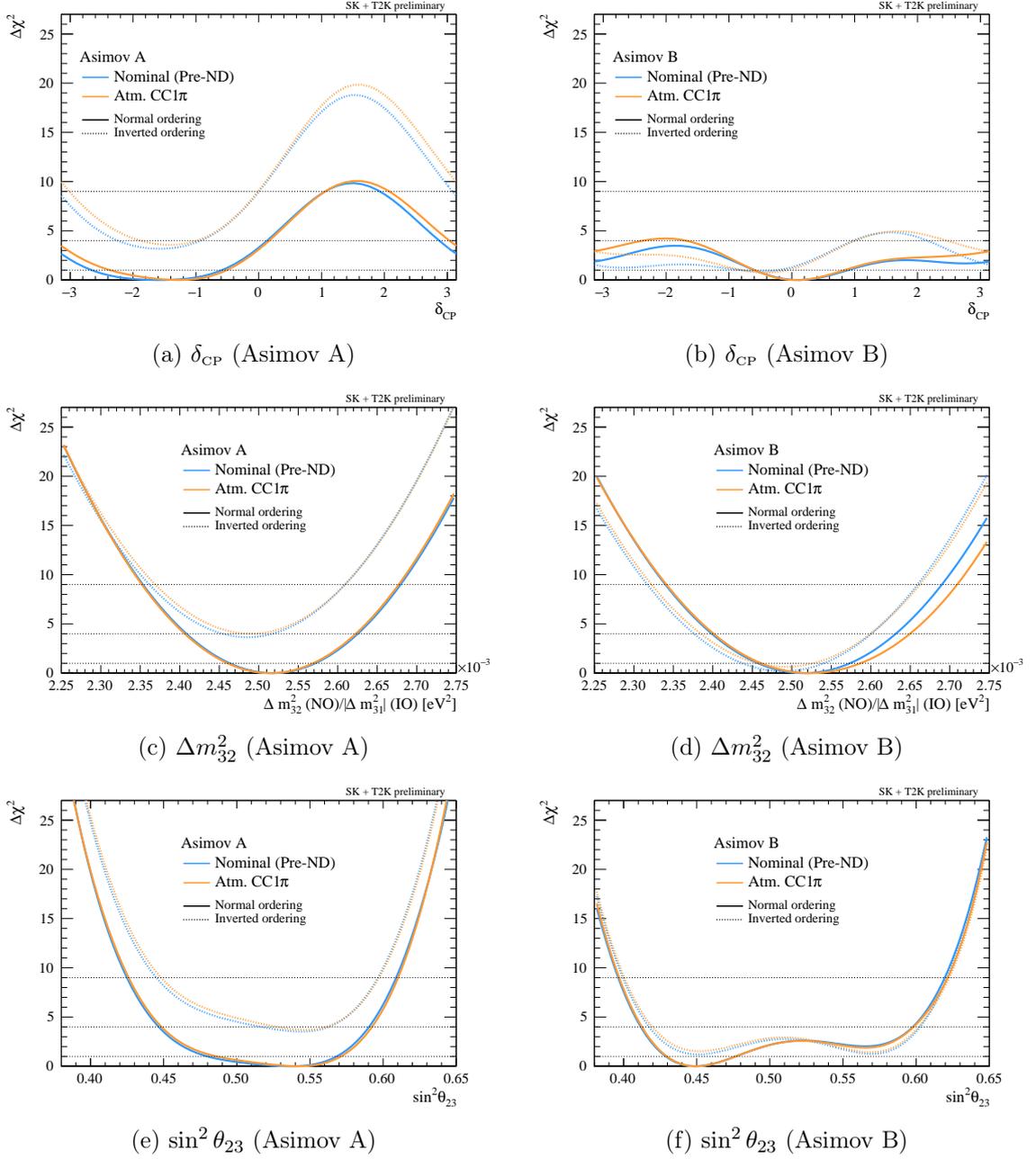|		δ_{CP}	Δm^2_{32}	$\sin^2\theta_{23}$
AsimovA	Middle of the 1σ interval	-1.381	0.002514	0.5284
	1σ interval size: 1σ	0.9453	5.39e-05	0.04272
	1σ interval ratio to Nominal	0.9316	1.003	0.9514
	Fractional change in the 1σ interval size wrt the syst interval $\left(1\sigma^{\text{FDS}}_{\text{tot.}} - 1\sigma^{\text{ref}}_{\text{tot.}}\right)/1\sigma^{\text{ref}}_{\text{syst.}}$	-11.4%	0.697%	-8.91%
	Bias in the middle of 1σ interval wrt the size of the 1$\sigma^{\text{ref}}_{\text{tot.}}$ interval	26.3%	-2.14%	8.51%
	Bias in the middle of 1σ interval wrt the size of the 1$\sigma^{\text{ref}}_{\text{syst.}}$ interval	44%	-5.57%	15.6%
	Middle of the 2σ interval	-1.5	0.002514	0.521
	2σ interval size: 2σ	1.714	0.0001079	0.07177
	2σ interval ratio to Nominal	0.9786	1.002	0.9902
	Fractional change in the 2σ interval size wrt the syst interval $\left(2\sigma^{\text{FDS}}_{\text{tot.}} - 2\sigma^{\text{ref}}_{\text{tot.}}\right)/2\sigma^{\text{ref}}_{\text{syst.}}$	-3.97%	0.6%	-1.94%
	Bias in the middle of 2σ interval wrt the size of the 2$\sigma^{\text{ref}}_{\text{tot.}}$ interval	8.29%	-0.913%	2.29%
	Bias in the middle of 2σ interval wrt the size of the 2$\sigma^{\text{ref}}_{\text{syst.}}$ interval	15.4%	-2.34%	4.54%
	Bias in the middle of 2σ interval wrt the size of the 1$\sigma^{\text{ref}}_{\text{tot.}}$ interval	14.3%	-1.83%	3.69%
	Bias in the middle of 2σ interval wrt the size of the 1$\sigma^{\text{ref}}_{\text{syst.}}$ interval	23.9%	-4.75%	6.77%
AsimovB	Middle of the 1σ interval	0.09058	0.002516	0.4531
	1σ interval size: 1σ	0.675	5.829e-05	0.02404
	1σ interval ratio to Nominal	0.9202	1.009	1.007
	Fractional change in the 1σ interval size wrt the syst interval $\left(1\sigma^{\text{FDS}}_{\text{tot.}} - 1\sigma^{\text{ref}}_{\text{tot.}}\right)/1\sigma^{\text{ref}}_{\text{syst.}}$	-14.1%	2.39%	1.35%
	Bias in the middle of 1σ interval wrt the size of the 1$\sigma^{\text{ref}}_{\text{tot.}}$ interval	-12.8%	4.01%	3.06%
	Bias in the middle of 1σ interval wrt the size of the 1$\sigma^{\text{ref}}_{\text{syst.}}$ interval	-22.7%	10.2%	5.53%
	Middle of the 2σ interval	1.135	0.002516	0.5042
	2σ interval size: 2σ	2.793	0.0001167	0.09263
	2σ interval ratio to Nominal	0.8891	1.009	0.9961
	Fractional change in the 2σ interval size wrt the syst interval $\left(2\sigma^{\text{FDS}}_{\text{tot.}} - 2\sigma^{\text{ref}}_{\text{tot.}}\right)/2\sigma^{\text{ref}}_{\text{syst.}}$	-13.7%	2.34%	-0.771%
	Bias in the middle of 2σ interval wrt the size of the 2$\sigma^{\text{ref}}_{\text{tot.}}$ interval	-	1.9%	-0.147%
	Bias in the middle of 2σ interval wrt the size of the 2$\sigma^{\text{ref}}_{\text{syst.}}$ interval	-	4.83%	-0.288%
	Bias in the middle of 2σ interval wrt the size of the 1$\sigma^{\text{ref}}_{\text{tot.}}$ interval	-	3.8%	-0.574%
	Bias in the middle of 2σ interval wrt the size of the 1$\sigma^{\text{ref}}_{\text{syst.}}$ interval	-	9.63%	-1.04%

Table 6.6. Summary of the Bayes factors for the atmospheric CC1π fake data fit and reference Asimov fit, and their ratio.

	Fake data	Reference	Ratio (fake data/reference)
Asimov A	7.99	6.75	1.18
Asimov B	1.34	1.23	1.09

6.3.3 Near detector fit extrapolation

The generated FDS spectra and ND prediction error bands are shown in Fig. 6.9. The fake data event spectra have in general more events than the nominal spectra in the atmospheric samples, but they are still within the predictions.

(a) Sub-GeV e-like 0 de (b) Sub-GeV e-like 1 de (c) Sub-GeV μ-like 0 de

(d) Sub-GeV μ-like 1 de (e) Sub-GeV μ-like 2 de (f) Sub-GeV π^0-like
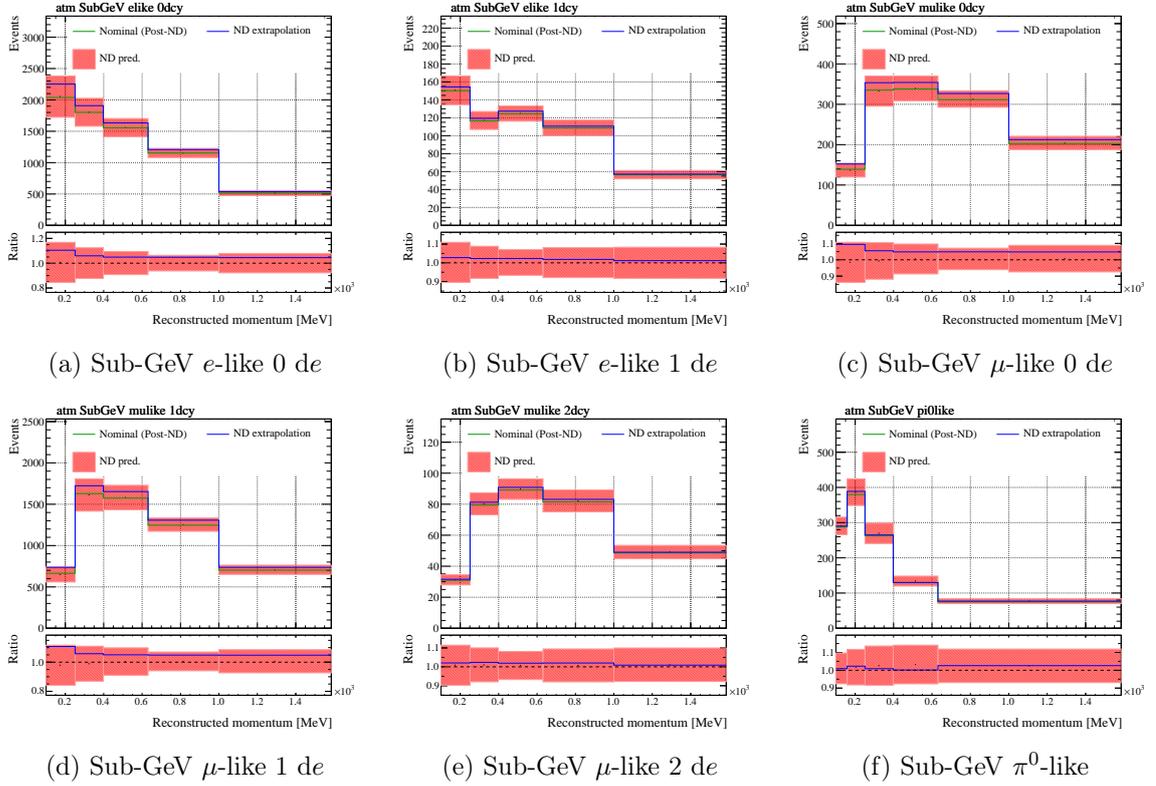
Figure 6.9. Comparison between the nominal event spectra at Post-ND (solid green) and the ND extrapolation fake data event spectra (solid blue) at the Asimov A oscillation parameter set. The prediction from the T2K near detector (ND) fake data fit is shown with the red band. The bottom panel shows the ratio to the nominal spectra.

The 1D $\Delta\chi^2$ curves for $\sin^2\theta_{23}$, Δm^2_{32}, and δ_{CP} are shown in Fig. 6.10. The mass ordering Bayes factor for fake data fit and reference Asimov fit and their ratios are summarized in Table 6.7. No significant change in the oscillation parameters is observed in the fake dat fit. Following our criteria, no action is required for this fake data study.

Table 6.7. Summary of the Bayes factors for the ND extrapolation fake data fit and reference Asimov fit, and their ratio.

	Fake data	Reference	Ratio (fake data/reference)
Asimov A	8.89	8.05	1.10
Asimov B	1.42	1.32	1.07

Figure 6.10. 1D $\Delta\chi^2$ distribution for the nominal fit at Post-ND and the ND extrapolation fake data fit. The left column shows the results at the oscillation parameter set Asimov A and the right column shows the results at Asimov B.

6.3.4   Summary

We performed 12 fake data studies to test the possible weakness of our analysis to the effects that are not taken into account in our nominal model. The results of all the fake 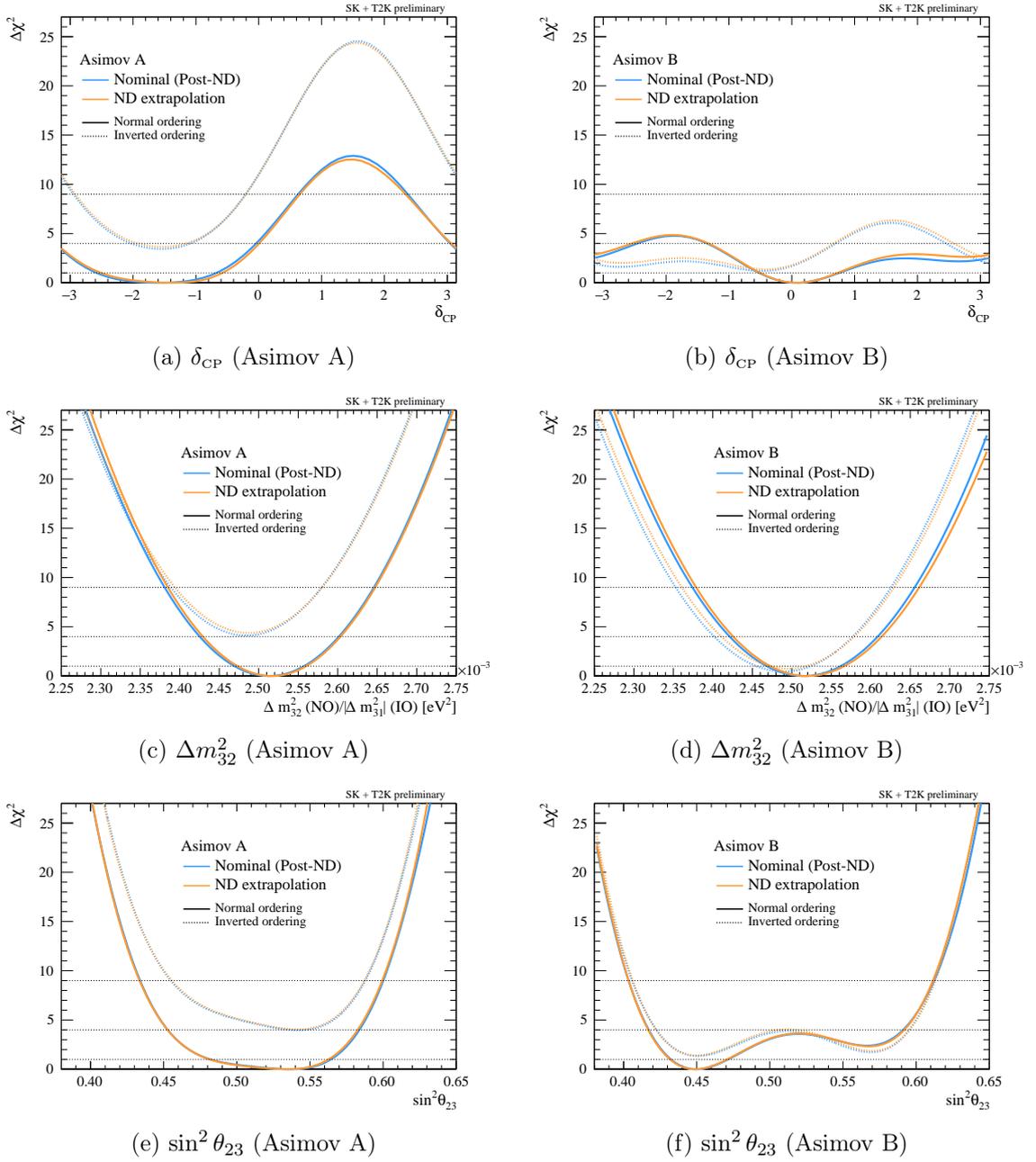data studies performed in this analysis are summarized in Tables 6.8 and 6.9. A bias was found in Δm_{32}^2 from the CRPA and pion multiplicity fake data studies. To reduce the effect of this bias in the oscillation analysis such that it is no longer significant, we will smear the likelihood curve of Δm_{32}^2. Using the result of fake data studies at Asimov A, the smearing factor is calculated from the quadrature-sum of the biases in all the fake data studies as

$$S^{\text{AsimovA}} = \sqrt{\sum_{\text{FDS}} \left(1\sigma_{\text{syst.}}^{\text{ref}} \times B_{\Delta m_{32}^2}^{\text{syst. FDS}} \right)^2} = 3.6 \times 10^{-5} \text{ eV}^2.$$

We use the results of all the fake data studies to compute the smearing factor regardless of whether each study passes the criteria or not. This approach gives a slightly conservative estimation of the fake data effects than simply using the results of the studies that broke the criteria. However, since we compute the quadrature sum of the obtained biases, the final smearing factor will be dominated by a few studies that had larger biases. The values used in the calculation of the smearing are summarized in Table 6.10. For reference, the smearing factor computed only from the fake data studies that failed the criteria is $3.3 \times 10^{-5} \text{ eV}^2$.

The ND non-QE CC0π FDS did not satisfy the interval ratio criterion for Δm_{32}^2 in Asimov B, but since the obtained metric is very close to the criterion, this effect will be covered when we apply Gaussian smearing on Δm_{32}^2. The interval criterion is not met for δ_{CP} in the Asimov B fit of the atmospheric CC1π excess FDS either. However, since this is also very close to the criterion and it only breaks the criterion in the Asimov B fit which is far from the T2K best-fit point, we do not require additional action for this fake data study. The change in the mass ordering Bayes factor is also shown to be small in all the fake data studies.

Based on these results, we conclude that our model is robust to possible out-of-model effects and we can move on to the data fit. The remaining tests of the fake data effects on the δ_{CP} and J_{CP} credible interval edges are presented in Section 7.3.6.

Table 6.8. Summary of the metrics to assess the effect of alternate models at Asimov A oscillation parameter set. The values that broke the criteria are highlighted with a red background color.

Fake data	Reference	Interval ratio $R_x^{2\sigma}$			Bias $B_x^{\text{syst.}}$			Mass ordering
		δ_{CP}	Δm_{32}^2	$\sin^2\theta_{23}$	δ_{CP}	Δm_{32}^2	$\sin^2\theta_{23}$	BF ratio
CRPA	Normal Asimov	0.991	0.969	1.02	-4.85%	154%	-10.5%	0.99
LFG	Normal Asimov	0.969	0.988	0.926	2%	38.4%	-4.76%	1.12
Removal energy	Normal Asimov	1.01	1	1	1.49%	-22.5%	3.54%	0.96
Axial form factor	Scaled Asiomv	0.998	0.995	0.978	-1.22%	7.49%	-0.944%	0.97
Martini 2p2h	Scaled Asiomv	0.972	0.989	0.982	-1.29%	-8.03%	-3.89%	1.09
Pion multiplicity	Normal Asimov	1	0.999	1.02	2.35%	-63.5%	-4.53%	0.90
Energy-dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$	Scaled Asiomv	0.988	0.994	0.992	0.04%	-5.82%	1.72%	1.08
Atmospheric CC1π	Normal Asimov	0.963	0.998	1	15.2%	-11%	8.45%	1.18
ND CC1π	Normal Asimov	1.02	1	0.964	0.501%	18.4%	-4.57%	1.01
ND Non-QE CC0π	Normal Asimov	0.948	0.904	0.995	-5.57%	-20.3%	-10.6%	1.00
ND extrapolation	Normal Asimov	1.01	0.989	0.991	5.02%	17%	-2.54%	1.10
Pion SI bug fix	Normal Asimov	0.997	1.02	0.985	2.05%	44.8%	-0.91%	1.00

Table 6.9. Summary of the metrics to assess the effect of alternate models at Asimov B oscillation parameter set. The values that broke the criteria are highlighted with a red background color.

Fake data	Reference	Interval ratio $R_x^{2\sigma}$			Bias $B_x^{\text{syst.}}$			Mass ordering
		δ_{CP}	Δm_{32}^2	$\sin^2\theta_{23}$	δ_{CP}	Δm_{32}^2	$\sin^2\theta_{23}$	BF ratio
CRPA	Normal Asimov	1	0.95	0.909	-	153%	-	1.04
LFG	Normal Asimov	1	0.985	0.945	-	32.8%	-5.97%	1.04
Removal energy	Normal Asimov	1	1.01	0.998	-	-31.2%	-4.32%	1.04
Axial form factor	Scaled Asiomv	1	1.01	1.02	-	-4.13%	9.13%	0.97
Martini 2p2h	Scaled Asiomv	1	0.993	0.976	-	-4.49%	-15.1%	0.99
Pion multiplicity	Normal Asimov	1	0.996	0.98	-	-44.6%	-23.3%	0.90
Energy-dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$	Scaled Asiomv	1	1	0.99	-	1.42%	7.06%	1.00
Atmospheric CC1π	Normal Asimov	0.891	1.07	0.999	-	43.7%	4.77%	1.09
ND CC1π	Normal Asimov	1	0.986	0.954	-	23.5%	-8.27%	1.04
ND Non-QE CC0π	Normal Asimov	1	0.896	0.937	-	-9.56%	-35.1%	0.97
ND extrapolation	Normal Asimov	0.979	1	1.01	-8.18%	25.2%	1.94%	1.07
Pion SI bug fix	Normal Asimov	1	1.01	0.99	-	49%	-5.47%	1.05

Table 6.10. Summary of the size of biases used to calculate the Δm^2_{32} smearing factor.

Fake data	Reference	$1\sigma^{\text{ref}}_{\text{syst.}}$	Bias
CRPA	Normal Asimov	1.98×10^{-5}	153.77%
LFG	Normal Asimov	1.98×10^{-5}	38.35%
Removal energy	Normal Asimov	2.14×10^{-5}	-22.52%
Axial form factor	Scaled Asiomv	2.06×10^{-5}	7.49%
Martini 2p2h	Scaled Asiomv	2.19×10^{-5}	-8.03%
Pion multiplicity	Normal Asimov	2.14×10^{-5}	-63.52%
Energy-dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$	Scaled Asiomv	2.22×10^{-5}	-5.82%
Atmospheric CC1π	Normal Asimov	2.14×10^{-5}	-11.02%
ND CC1π	Normal Asimov	2.14×10^{-5}	18.41%
ND Non-QE CC0πi	Normal Asimov	2.14×10^{-5}	-20.31%
ND extrapolation	Normal Asimov	1.79×10^{-5}	17.04%
Pion SI bug fix	Normal Asimov	1.98×10^{-5}	44.85%

Chapter 7

Oscillation analysis results

This chapter presents the results of the joint analysis between the T2K beam and SK atmospheric neutrinos. The obtained data distributions and best-fit spectra are first shown in Section 7.1. The $\Delta\chi^2$ distributions obtained from the importance sampling method are shown in Section 7.2, including the comparison to the MCMC method. Then, the main Bayesian analysis based on the MCMC method is presented in Section 7.3.

7.1 Global best-fit results

The global best-fit values are obtained by profiling all the oscillation and systematic uncertainty parameters and minimizing the test statistic $-2\log\mathcal{L}$. The best-fit values of four oscillation parameters at both the normal and inverted ordering are summarized in Table 7.1. In all three fits, the global best fit prefers the normal ordering. The offset $\Delta(-2\log\mathcal{L})$ for inverted ordering is the largest in the SK+T2K fit, which means that the power of the IO-rejection is stronger than the fits to the samples from the individual experiments. The number of events in data and best-fit predictions are summarized in Table 7.2 along with χ^2 for each sample.

It is insightful to have a look at the best-fit event spectra to understand the relationships between the prediction and the actual data. The corresponding best-fit spectra are shown in Fig. 7.1. In the T2K beam samples, the SK+T2K fit predicts more events in the ν_μ ($\bar{\nu}_\mu$) samples than the T2K-only fit while predicting fewer events in the ν_e ($\bar{\nu}_e$) samples. Since the leading term of the $\nu_\mu \to \nu_e$ oscillation probability is proportional to $\sin^2\theta_{23}$, this overall normalization difference is coming from the different θ_{23} octant preference. The difference between the SK+T2K fit and SK (+ND) fit in the SK atmospheric samples is not as apparent as in the T2K samples when projected onto one dimension. However, some differences are observed in the high-energy samples. The SK+T2K fit predicts slightly fewer upward-going events ($\cos\Theta_z < 0$) in the μ-like samples and predicts the opposite for the e-like samples. These differences are thought to be caused by the different mass ordering and δ_{CP} preferences when combined with the T2K beam data.

To illustrate the effect of alternative δ_{CP}, mass ordering, and octant choices, the event spectra of the local best fits where we fix the parameter at $\delta_{\mathrm{CP}} = 0$, inverted ordering, or upper octant are shown in Fig. 7.2. The T2K beam e-like sample prediction shows a large variation under different hypotheses. The SK atmospheric samples show the largest

Table 7.1. Best-fit oscillation parameter values for SK+T2K, T2K, and SK (+ND) fits with the reactor constraint on $\sin^2 2\theta_{13} = 0.0853 \pm 0.0027$. The errors are assigned using the fixed-$\Delta\chi^2$ method for a 1D marginal likelihood fit with fixed mass ordering (Fig. 7.3). It should be noted that the Δm^2_{32} smearing is applied only to the SK+T2K case when extracting the errors.

	SK+T2K		T2K		SK (+ND)	
Mass ordering	Normal	Inverted	Normal	Inverted	Normal	Inverted
$\sin^2(2\theta_{13})$	0.0859	0.0858	0.0854	0.0858	0.0852	0.0852
δ_{CP}	$-1.76^{+0.60}_{-0.74}$	$-1.49^{+0.50}_{-0.52}$	$-2.08^{+1.05}_{-0.52}$	$-1.40^{+0.53}_{-0.62}$	$-2.18^{+1.06}_{-1.62}$	$-2.00^{+1.13}_{-1.83}$
Δm^2_{32} (NO)/$\|\Delta m^2_{31}\|$ (IO)	$2.514^{+0.057}_{-0.060}$	$2.485^{+0.056}_{-0.061}$	$2.505^{+0.045}_{-0.054}$	$2.472^{+0.049}_{-0.049}$	$2.499^{+0.831}_{-0.197}$	$2.495^{+0.872}_{-0.214}$
$\sin^2\theta_{23}$	$0.471^{+0.104}_{-0.017}$	$0.556^{+0.021}_{-0.033}$	$0.563^{+0.023}_{-0.029}$	$0.564^{+0.022}_{-0.025}$	$0.427^{+0.054}_{-0.028}$	$0.435^{+0.082}_{-0.035}$
$-2\log\mathcal{L}$	1165.28	1168.58	586.849	588.886	575.193	577.235
$\Delta(-2\log\mathcal{L})$	0.0	3.31	0.0	2.036	0.0	2.041

impact from the mass ordering in the upward-going events in the multi-ring e-like ν_e sample.

The two-dimensional event distributions are shown in Appendix E.

Table 7.2. Summary of the event rates and sample χ^2 at the best-fit oscillation and systematic parameters at SK + T2K, T2K-only, and SK-only (+ND) fit with reactor constraint on $\sin^2\theta_{13}$. Note that UpMu samples have non-integer event numbers due to the background subtraction. The corresponding plots are shown in Fig. 7.1.

| Sample	SK+T2K		T2K		SK (+ND)		Data
	Prediction	χ^2	Prediction	χ^2	Prediction	χ^2	
FHC 1Rμ	351.10	243.83	341.91	242.28			323
RHC 1Rμ	140.33	173.54	135.21	173.50			137
FHC 1Re	88.93	77.59	95.60	76.78			95
RHC 1Re	15.49	49.17	17.09	48.73			16
FHC 1Re 1de	10.85	43.44	11.29	43.14			14
SubGeV e-like 0de	7916.50	76.07			7913.33	76.07	7899
SubGeV e-like 1de	795.44	1.63			791.67	1.67	799
SubGeV μ-like 0de	1490.37	39.35			1493.85	39.48	1489
SubGeV μ-like 1de	6558.69	51.15			6572.47	50.84	6587
SubGeV μ-like 2de	499.24	2.95			498.65	2.81	495
SubGeV pi0like	1749.32	5.56			1749.14	5.47	1768
MultiGeV e-like ν_e	181.60	46.18			181.38	46.13	167
MultiGeV e-like $\bar{\nu}_e$	1024.65	49.62			1021.35	49.31	1011
MultiGeV μ-like	1015.45	15.03			1018.65	14.70	1026
MultiRing e-like ν_e	1090.22	30.54			1089.18	30.93	1122
MultiRing e-like $\bar{\nu}_e$	1025.24	36.40			1025.54	36.50	1003
MultiRing μ-like	2693.86	39.07			2693.58	38.46	2757
MultiRing Other	1405.89	40.32			1405.47	40.30	1443
PCStop	351.28	18.40			350.89	18.24	349
PCThru	1738.45	44.07			1739.58	44.02	1728
UpStop μ	804.04	26.03			806.14	25.94	815.2
UpThruNonShower μ	3096.82	18.82			3094.44	19.33	3107.2
UpThruShower μ	538.75	9.08			538.48	9.02	537.8
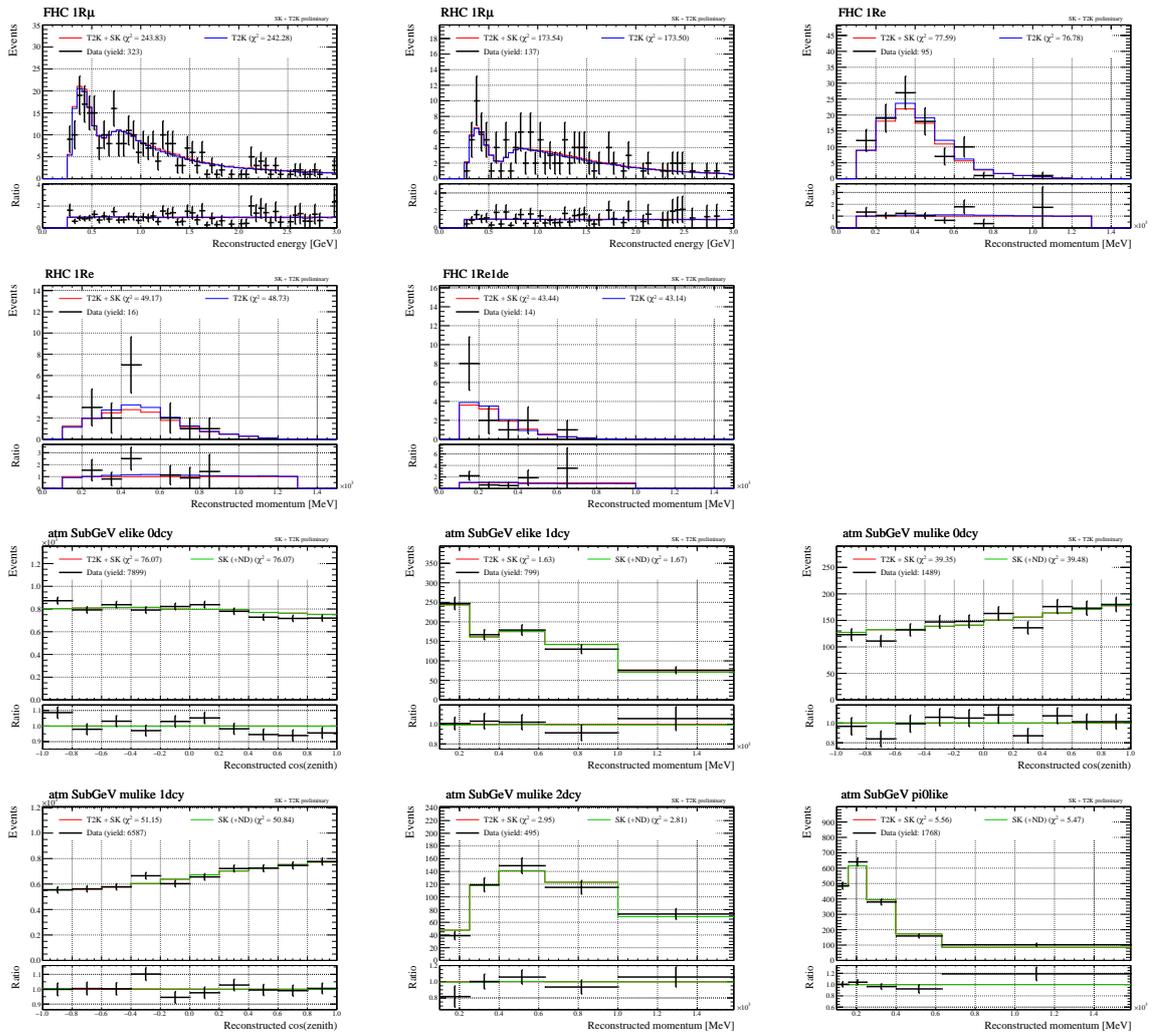
Figure 7.1. SK+T2K and T2K-only or SK-only (+ND) best-fit spectra for the T2K beam and SK atmospheric sub-GeV samples overlaid with the data. Error bars on the data show the statistical errors. The bottom panel shows the ratio to the SK+T2K best-fit spectra. The corresponding table is Table 7.2.
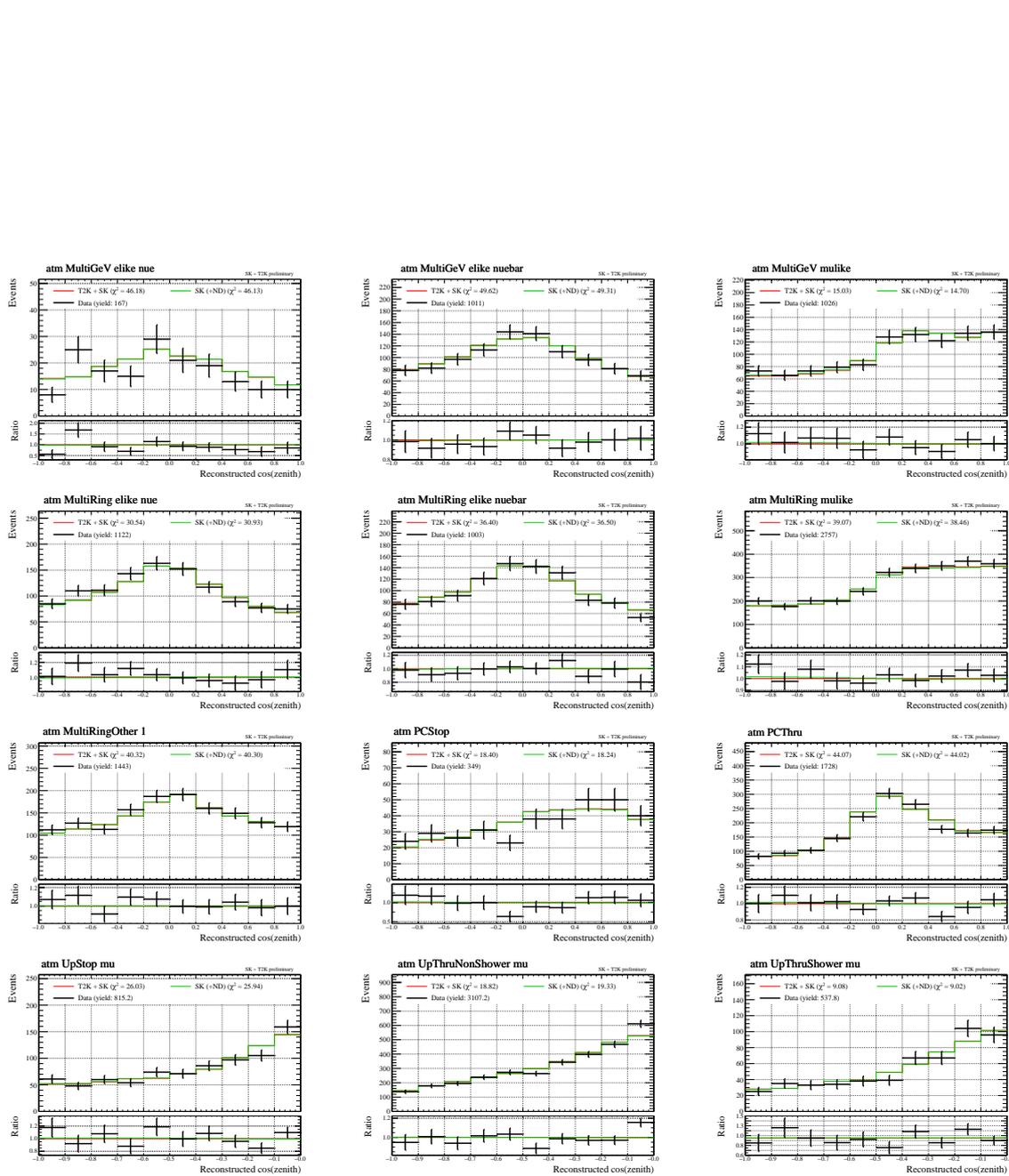
Figure 7.1. SK+T2K and SK-only (+ND) best-fit spectra for the SK atmospheric multi-GeV, PC, and UpMu samples overlaid with the data. Error bars on the data show the statistical errors. The bottom panel shows the ratio to the SK+T2K best-fit spectra. The corresponding table is Table 7.2.
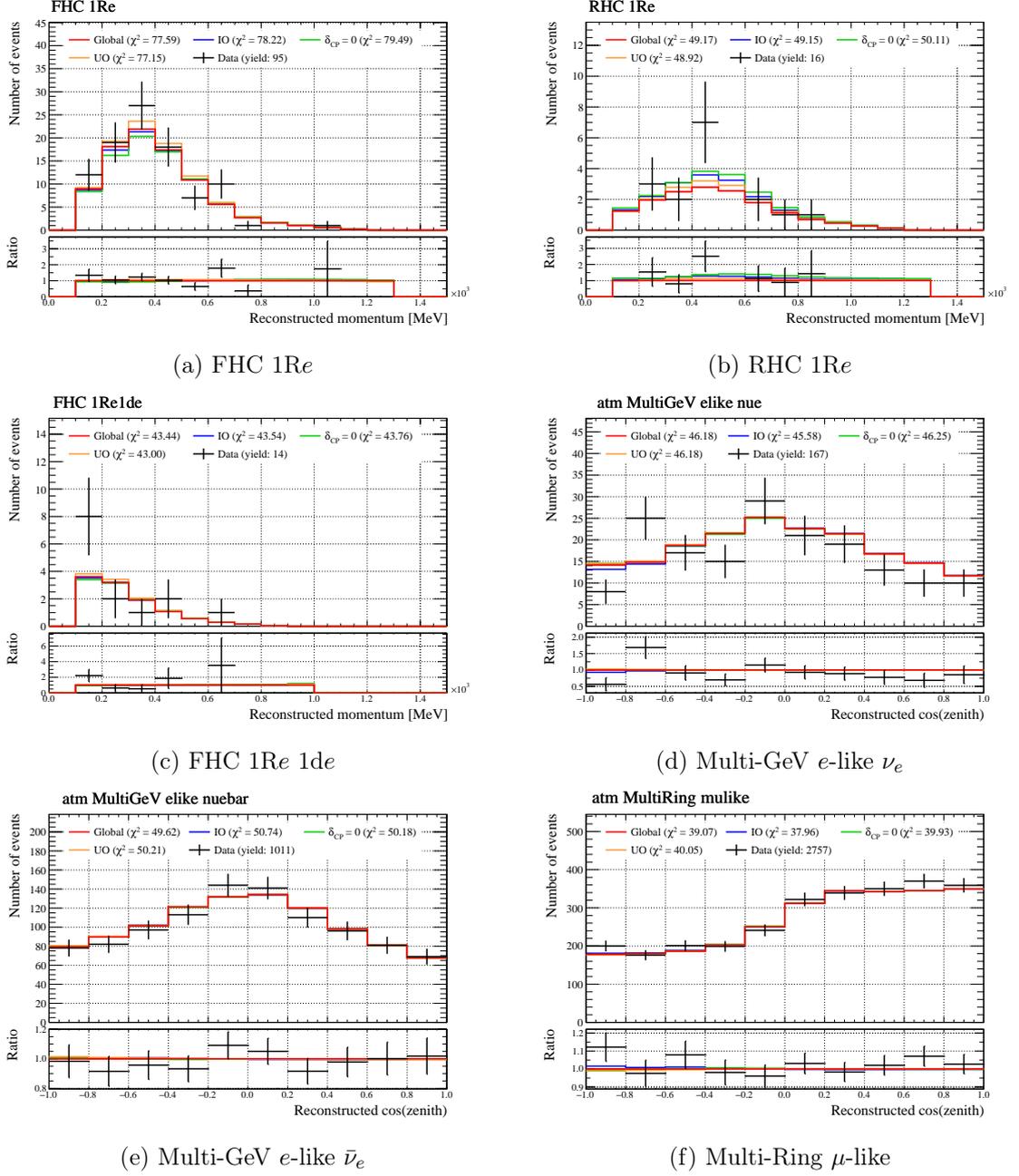
Figure 7.2. Comparison of the SK+T2K best-fit spectra under different δ_{CP} and mass ordering conditions for the selected samples. The spectra correspond to the overall best-fit, best-fit under $\delta_{\mathrm{CP}} = 0$, and best-fit under the inverted ordering are shown in red, blue, and green, respectively. The data points are overlaid with the statistical errors. The bottom panel shows the ratio to the overall best-fit spectra.

7.2 $\Delta\chi^2$ distributions

In this section, the one-dimensional and two-dimensional $\Delta\chi^2$ distributions computed using the importance sampling method are presented.

The comparison of one-dimensional $\Delta\chi^2$ curves among the SK+T2K, T2K, and SK (+ND) fits are shown in Fig. 7.3. For $\delta_{\rm CP}$, the data shows a preference for the value of $\delta_{\rm CP}$ close to the maximal mixing ($\delta_{\rm CP} \simeq -1.8$). The constraints mostly come from the T2K beam samples, but the SK atmospheric samples also contribute to the rejection of the values around $\delta_{\rm CP} \simeq [-1, 2]$. There is tension between T2K and SK for $\sin^2\theta_{23}$, where SK has a preference for the lower octant and T2K has a preference for the upper octant. Therefore, the joint fit has a very similar likelihood in both the lower and upper octants. In addition, the SK+T2K fit has a stronger preference for the normal ordering compared to the individual experiments.

The comparison of two-dimensional confidence intervals among the SK+T2K, T2K, and SK (+ND) fits are shown in Fig. 7.4. The overall constraint mostly comes from T2K in either case, but the $\delta_{\rm CP}$-$\sin^2\theta_{23}$ contour shrinks around the upper octant thanks to the constraint from SK. These two-dimensional likelihood distributions in $\delta_{\rm CP}$-$\sin^2\theta_{23}$ are used as an input to the multicanonical MCMC method.

(a) $\delta_{\rm CP}$

(b) $\Delta m^2_{32}/|\Delta m^2_{31}|$

(c) $\sin^2\theta_{23}$

(d) $\sin^2\theta_{13}$
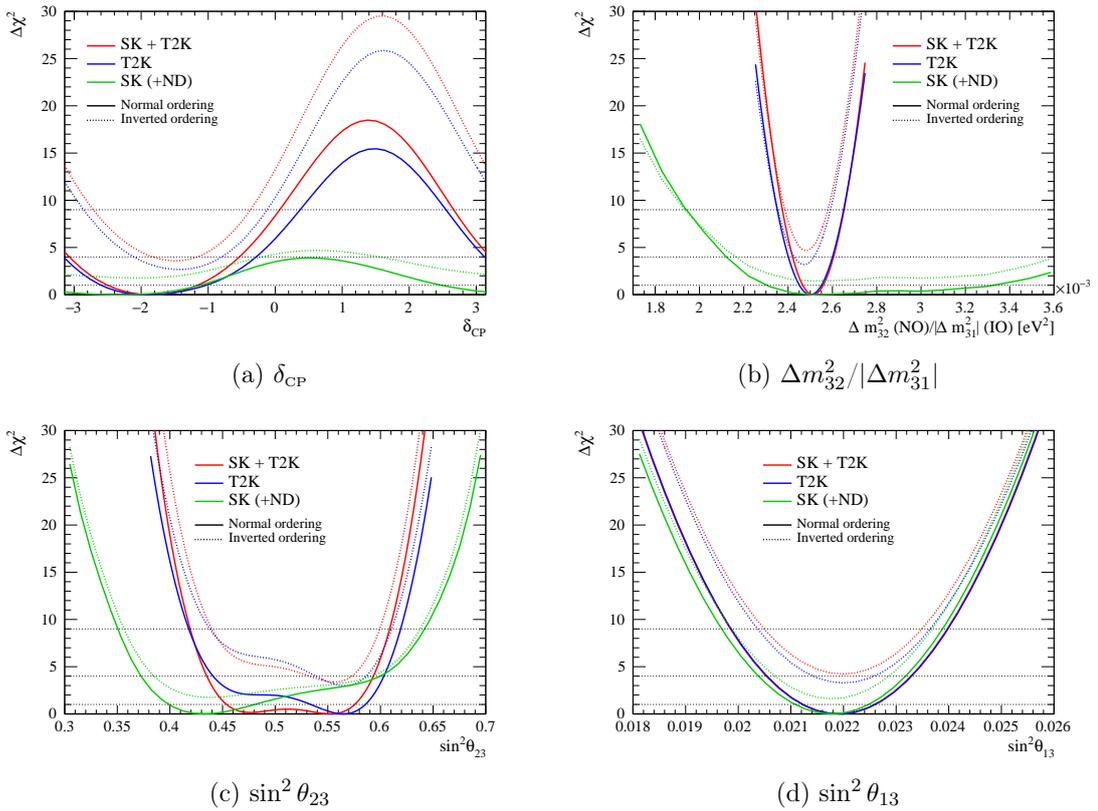
Figure 7.3. Comparison of the 1D $\Delta\chi^2$ from the SK+T2K, T2K, and SK (+ND) data fit results. The importance sampling method is used. Δm^2_{32} smearing is not applied here.
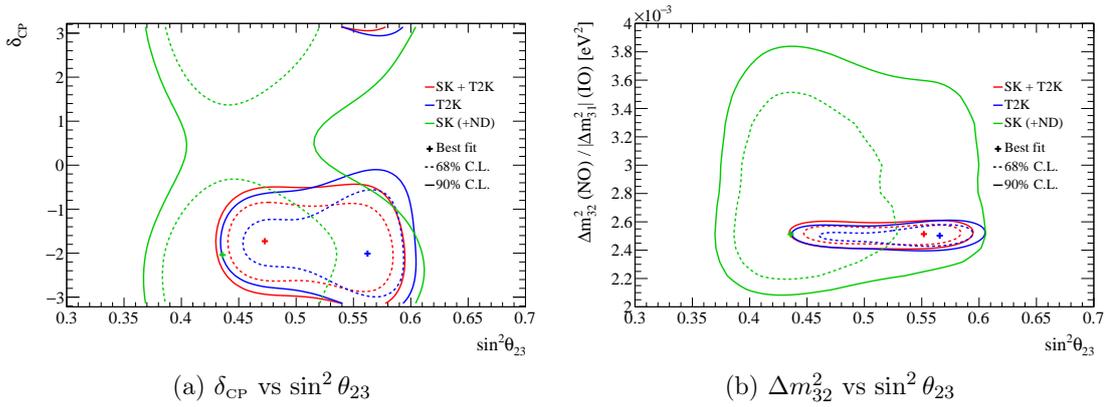
(a) δ_{CP} vs $\sin^2\theta_{23}$ (b) Δm^2_{32} vs $\sin^2\theta_{23}$

Figure 7.4. Comparison of the 2D confidence level contours in δ_{CP}-$\sin^2\theta_{23}$ and Δm^2_{32}-$\sin^2\theta_{23}$ from the SK+T2K, T2K and SK (+ND) data fit for normal mass ordering. The importance sampling method is used.

7.2.1 Comparison of the importance sampling and MCMC

The comparison of $\Delta\chi^2$ distributions obtained from the importance sampling and MCMC are shown in Fig. 7.5 with estimated errors from each method. For the MCMC, we use the results of the pilot chains with two million steps, while we use 200 million steps in the main analysis[1].

The overall agreement is good but MCMC shows a slightly stronger rejection of the inverted ordering around the maximal CP-violating value of δ_{CP} ($\sim -\pi/2$) and the maximal mixing of $\sin^2\theta_{23}$ (~ 0.5). We should be careful, however, that the importance sampling method has non-negligible errors around these regions and its results are not fully reliable. Considering the smaller errors around the most probable regions, we take the result of MCMC as our main result for Bayesian analysis in Section 7.3.

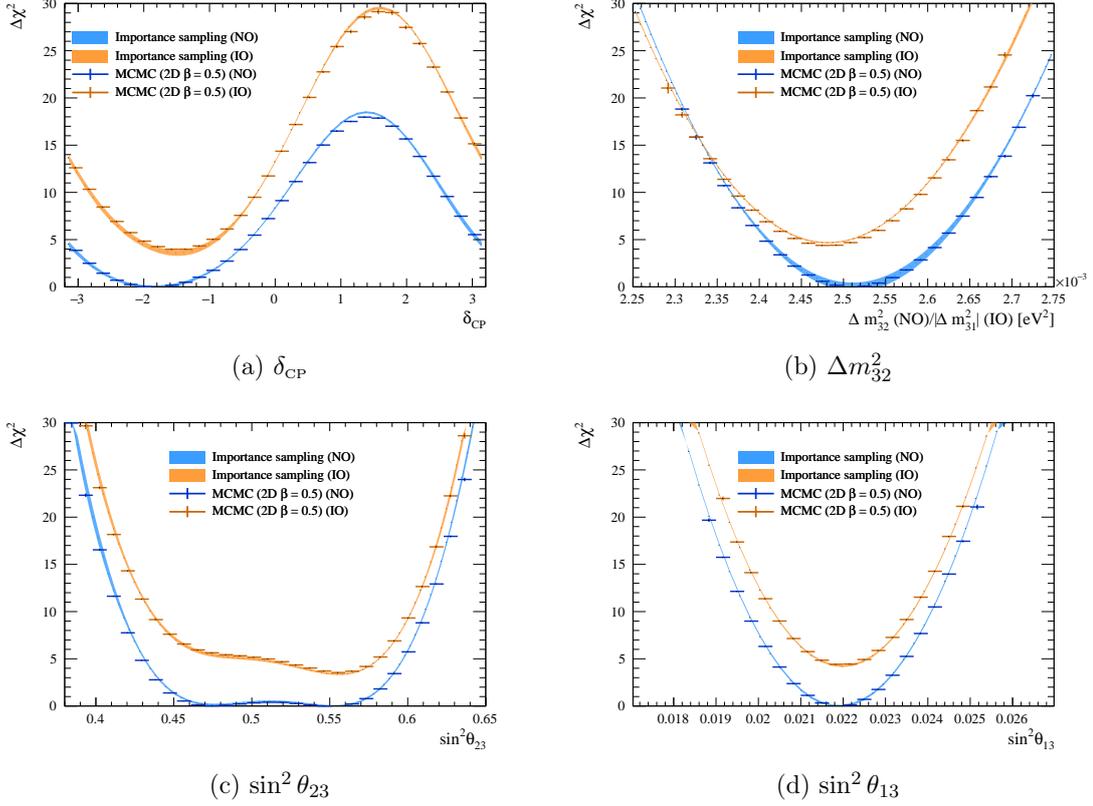
[1]See Appendix C for more details of the MCMC configurations.

Figure 7.5. Comparison of the 1D $\Delta\chi^2$ contour for each oscillation parameter computed from the importance sampling method and MCMC. Δm_{32}^2 smearing is not applied. For ease of interpretation, MCMC $\Delta\chi^2$ values are not shown for the bins that have less than five raw MCMC steps where it has a large MCMC fluctuation.

7.3   Bayesian analysis

This section presents the Bayesian analysis based on the results of 200 million MCMC steps using the 2D multicanonical method with $\beta = 0.5$. We apply the Gaussian smearing of 3.6×10^{-5} eV^2 to the obtained Δm_{32}^2 contours from the model robustness test performed in Chapter 6 unless specifically noted otherwise.

7.3.1   Posterior distribution of the CP parameters

For the test of CP conservation (violation), we use both δ_{CP} and the Jarlskog invariant J_{CP}. The posterior probabilities of these CP parameters are shown in Fig. 7.6. The credible intervals are shown for both the prior flat in δ_{CP} and the prior flat in $\sin\delta_{\mathrm{CP}}$. The corresponding most probable values and credible intervals are summarized in Tables 7.3 and 7.4.

For δ_{CP}, the CP conserving values ($\delta_{\mathrm{CP}} = 0, \pi$) are excluded from the 2σ credible intervals in both the normal and inverted ordering when we apply the flat δ_{CP} prior. However, $\delta_{\mathrm{CP}} = \pi$ is not excluded in normal ordering when the flat $\sin\delta_{\mathrm{CP}}$ prior is applied. One of the maximal CP violating values ($\delta_{\mathrm{CP}} = \pi/2$) is excluded at more than 3σ in either

case.

For J_{CP}, the CP conserving value of $J_{\mathrm{CP}} = 0$ is consistently excluded from the 2σ credible intervals under both the mass ordering hypotheses. Moreover, $J_{\mathrm{CP}} = 0$ is excluded at 3σ under the flat δ_{CP} prior when inverted ordering is assumed.

The reason we have a more conservative result in δ_{CP} is that we have a $\cos\delta_{\mathrm{CP}}$ dependence in the oscillation probability, and therefore, the result is not symmetric between $\delta_{\mathrm{CP}} = 0$ and $\delta_{\mathrm{CP}} = \pi$. Since we require both values $(\delta_{\mathrm{CP}} = 0, \pi)$ to be excluded from the credible intervals when we look at δ_{CP}, the requirement becomes more strict. On the other hand, when it comes to $\sin\delta_{\mathrm{CP}}$, the contributions from both $\delta_{\mathrm{CP}} = 0, \pi$ are marginalized over at $\sin\delta_{\mathrm{CP}} = 0$. Therefore, $\sin\delta_{\mathrm{CP}}$ (or J_{CP}) gives a more moderate conclusion on CP conservation than δ_{CP}.

Table 7.3. The most probable values and credible interval ranges for δ_{CP}, $\sin\delta_{\mathrm{CP}}$, and J_{CP} obtained from MCMC. **The flat prior in δ_{CP} is applied**. The most probable value is taken to be the point where the posterior probability density becomes the maximum by applying the cubic interpolation to the bin contents.

Normal ordering	δ_{CP}	$\sin\delta_{\mathrm{CP}}$	J_{CP}
Most probable value	-1.872	-1.000	-0.033
1σ	[-2.464, -1.205]	[-1.000, -0.776]	[-0.034, -0.026]
2σ	[-3.021, -0.556]	[-1.000, -0.261]	[-0.034, -0.008]
3σ	[-3.142, 0.085] and [2.682, 3.142]	[-1.000, 0.344]	[-0.035, 0.012]
Inverted ordering	δ_{CP}	$\sin\delta_{\mathrm{CP}}$	J_{CP}
Most probable value	-1.476	-1.000	-0.033
1σ	[-2.003, -0.976]	[-1.000, -0.870]	[-0.034, -0.029]
2σ	[-2.528, -0.506]	[-1.000, -0.523]	[-0.034, -0.017]
3σ	[-3.048, -0.023]	[-1.000, -0.052]	[-0.035, -0.002]
Both ordering	δ_{CP}	$\sin\delta_{\mathrm{CP}}$	J_{CP}
Most probable value	-1.797	-1.000	-0.033
1σ	[-2.417, -1.159]	[-1.000, -0.787]	[-0.034, -0.026]
2σ	[-2.985, -0.552]	[-1.000, -0.281]	[-0.034, -0.009]
3σ	[-3.142, 0.072] and [2.704, 3.142]	[-1.000, 0.325]	[-0.035, 0.011]
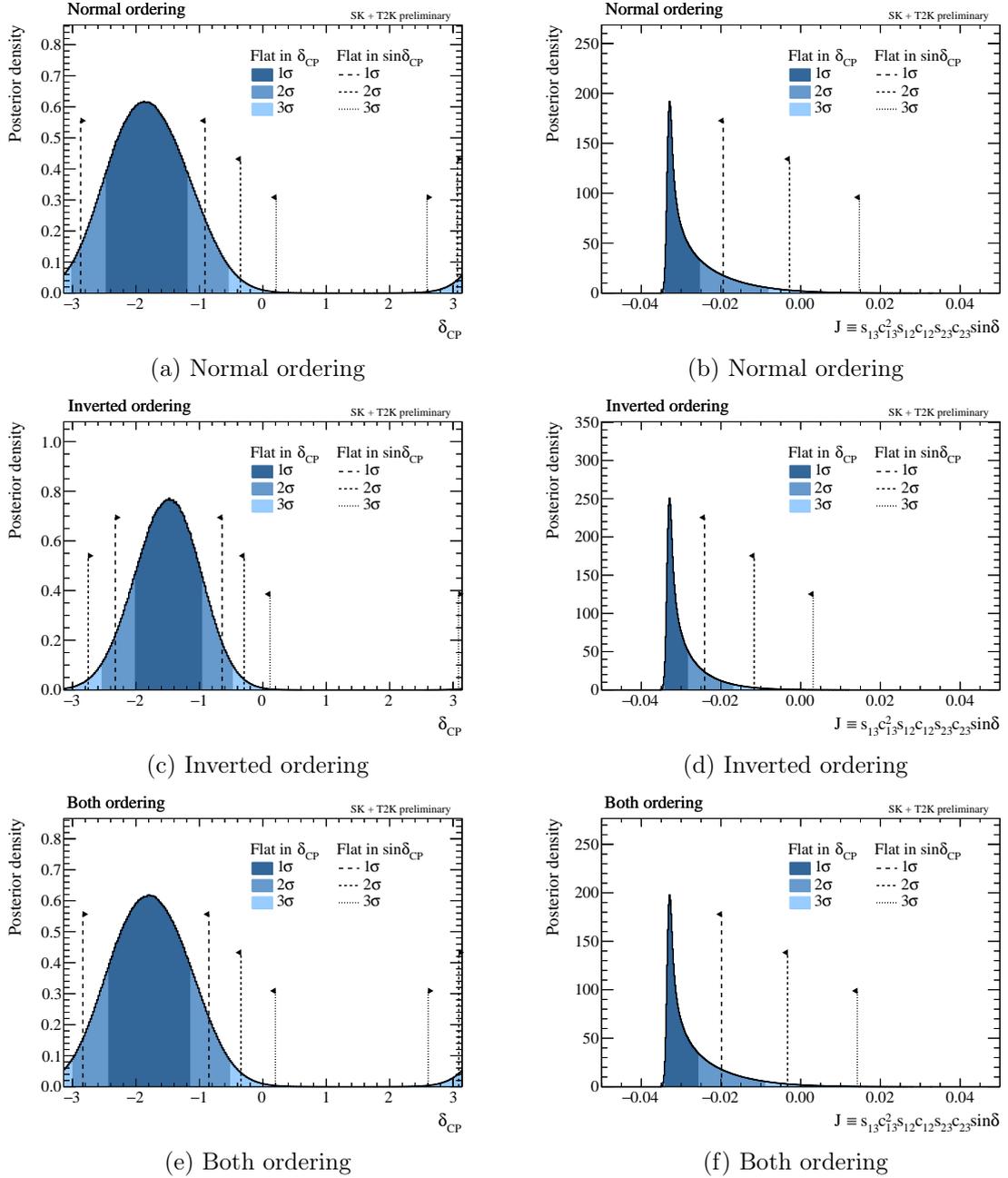
Figure 7.6. 1σ, 2σ, and 3σ credible intervals of δ_{CP} and Jarlskog invariant J_{CP} for normal and inverted ordering, and marginalized over both ordering. These are computed using the MCMC 2D multicanonical method with $\beta = 0.5$ and two different priors are applied: flat in δ_{CP} and flat in $\sin\delta_{\mathrm{CP}}$. For the δ_{CP} distributions under the flat $\sin\delta_{\mathrm{CP}}$ prior, the interval edges around $\delta_{\mathrm{CP}} = -\pi/2$ are not drawn for simplicity.

Table 7.4. The most probable values and credible interval ranges for δ_{CP}, $\sin\delta_{\mathrm{CP}}$, and J_{CP} obtained from MCMC. **The flat prior in** $\sin\delta_{\mathbf{CP}}$ **is applied**. The most probable value is taken to be the point where the posterior probability density becomes the maximum by applying the cubic interpolation to the bin contents.

Normal ordering	δ_{CP}	$\sin\delta_{\mathrm{CP}}$	J_{CP}
Most probable value	-2.315	-0.998	-0.032
1σ	[-2.873, -1.811] and [-1.182, -0.912]	[-1.000, -0.595]	[-0.033, -0.019]
2σ	[-3.142, -1.644] and [-1.491, -0.353] and [3.068, 3.142]	[-1.000, -0.090]	[-0.034, -0.003]
3σ	[-3.142, 0.210] and [2.589, 3.142]	[-1.000, 0.434]	[-0.034, 0.015]
Inverted ordering	δ_{CP}	$\sin\delta_{\mathrm{CP}}$	J_{CP}
Most probable value	-1.062	-0.998	-0.032
1σ	[-2.326, -1.811] and [-1.370, -0.642]	[-1.000, -0.737]	[-0.033, -0.024]
2σ	[-2.754, -1.625] and [-1.519, -0.295]	[-1.000, -0.359]	[-0.034, -0.012]
3σ	[-3.142, 0.113] and [3.085, 3.142]	[-1.000, 0.089]	[-0.035, 0.003]
Both ordering	δ_{CP}	$\sin\delta_{\mathrm{CP}}$	J_{CP}
Most probable value	-2.314	-0.998	-0.032
1σ	[-2.838, -1.811] and [-1.226, -0.851]	[-1.000, -0.607]	[-0.033, -0.020]
2σ	[-3.142, -1.642] and [-1.495, -0.346] and [3.089, 3.142]	[-1.000, -0.105]	[-0.034, -0.003]
3σ	[-3.142, 0.198] and [2.607, 3.142]	[-1.000, 0.422]	[-0.035, 0.014]

7.3.2 Posterior distribution of the other oscillation parameters

Figure 7.7 shows the credible intervals computed from the 2D multicanonical MCMC method for Δm^2_{32}, $\sin^2\theta_{23}$, and $\sin^2\theta_{13}$. The corresponding most probable values and credible intervals are summarized in Table 7.5.

$\sin^2\theta_{23}$ has an almost equal preference for both the upper and lower octant in normal ordering but it prefers the upper octant in inverted ordering. When marginalized over both mass ordering, it has a slightly higher posterior density in the upper octant. When marginalized over both ordering, 1σ interval of Δm^2_{32} is fully contained in the normal ordering region ($\Delta m^2_{32} > 0$).

The 2D posterior distributions are shown in Fig. 7.8 which is marginalized over both ordering. The flat δ_{CP} prior is applied for these plots.

7.3.3 Comparison of the different sample fits

The comparison of posterior distributions among the SK+T2K, T2K, and SK (+ND) MCMC fits are shown in Figs. 7.9 and 7.10. Similarly to the importance sampling results shown in Fig. 7.3, it enhances the δ_{CP} constraints when both samples are combined. Δm^2_{32} is mostly constrained from the T2K samples, but the most probable value is slightly shifted when combined. The 1σ credible interval is fully contained in the normal ordering region ($\Delta m^2_{32} > 0$) for the SK+T2K fit, while it is not in the T2K-only or SK (+ND) fits. In $\sin^2\theta_{23}$, T2K has higher probabilities in the upper octant, while SK has higher probabilities in the lower octant. Therefore, the joint fit results show almost equal posterior density for the lower and upper octant. In all cases, the joint fit has higher probabilities for the normal ordering.
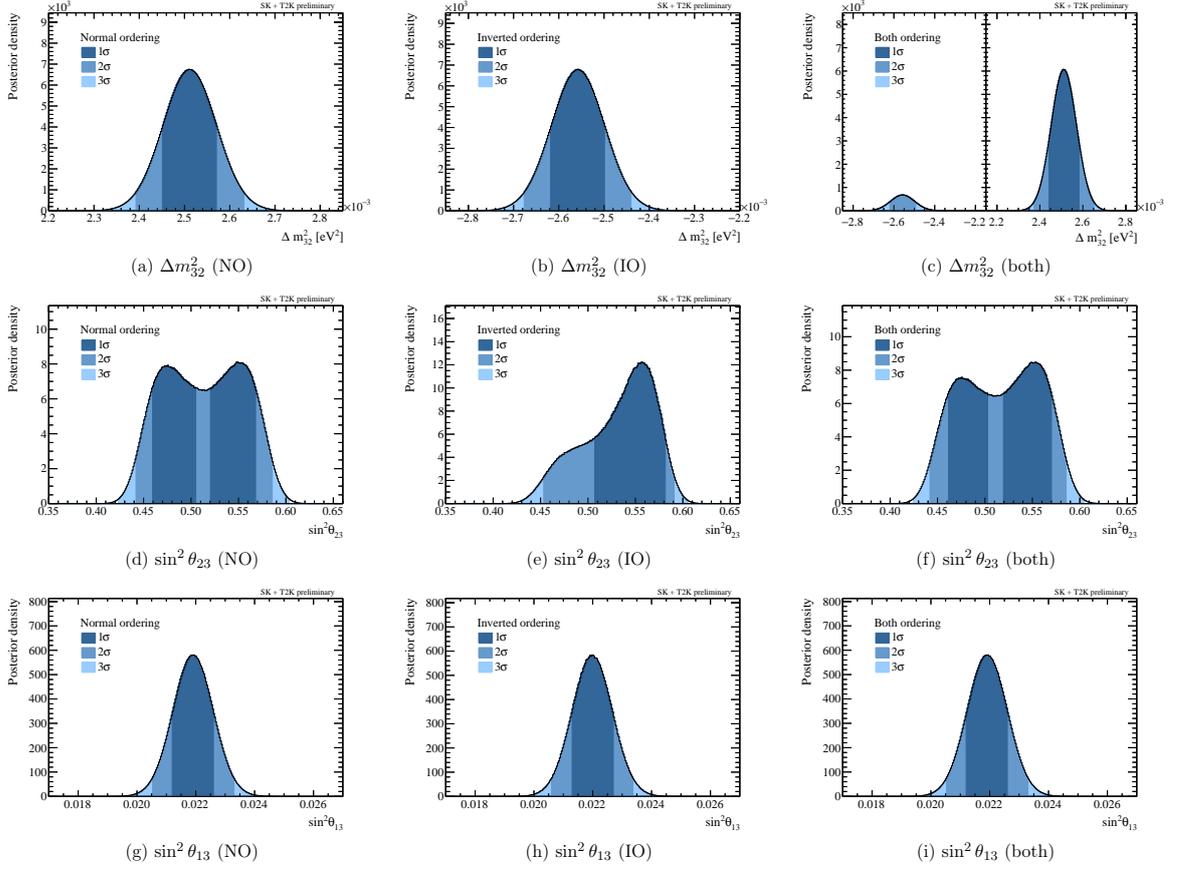
Figure 7.7. 1σ, 2σ, and 3σ credible intervals of Δm_{32}^2, $\sin^2\theta_{23}$, $\sin^2\theta_{13}$ for normal and inverted ordering, and marginalized over both ordering. These are computed using the MCMC 2D multicanonical method with $\beta = 0.5$. Δm_{32}^2 smearing of 3.6×10^{-5} eV^2 is applied. The corresponding most probable values and credible interval ranges are summarized in Table 7.5.

7.3.4 Posterior probabilities for mass ordering and $\sin^2\theta_{23}$

The posterior probabilities for the four combinations of mass ordering and $\sin^2\theta_{23}$ octants are summarized in Table 7.6. The joint SK+T2K fit favors the combination of the normal ordering and the upper octant the most, while the preference for the upper octant is less significant than the T2K-only fit.

In the standard Bayesian analysis, the Bayes factor between two hypotheses H_1 and H_2 is defined as the ratio of the marginal likelihood

$$\mathrm{BF}(H_1, H_2) = \frac{p_1(n)}{p_2(n)} = \frac{p(H_1|n)/\pi(H_1)}{p(H_2|n)/\pi(H_2)}, \tag{7.1}$$

where n denotes the data, $\pi(\cdot)$ is the prior probability, and $p(\cdot|x)$ is the posterior probability. The Bayes factor for the normal ordering over inverted ordering is computed to be $\mathbf{8.98 \pm 0.06}$. It should be noted that the error only shows the MCMC statistical errors evaluated as described in Appendix C. This Bayes factor does not have enough significance to give a certain conclusion on these hypotheses. As a reference, this value
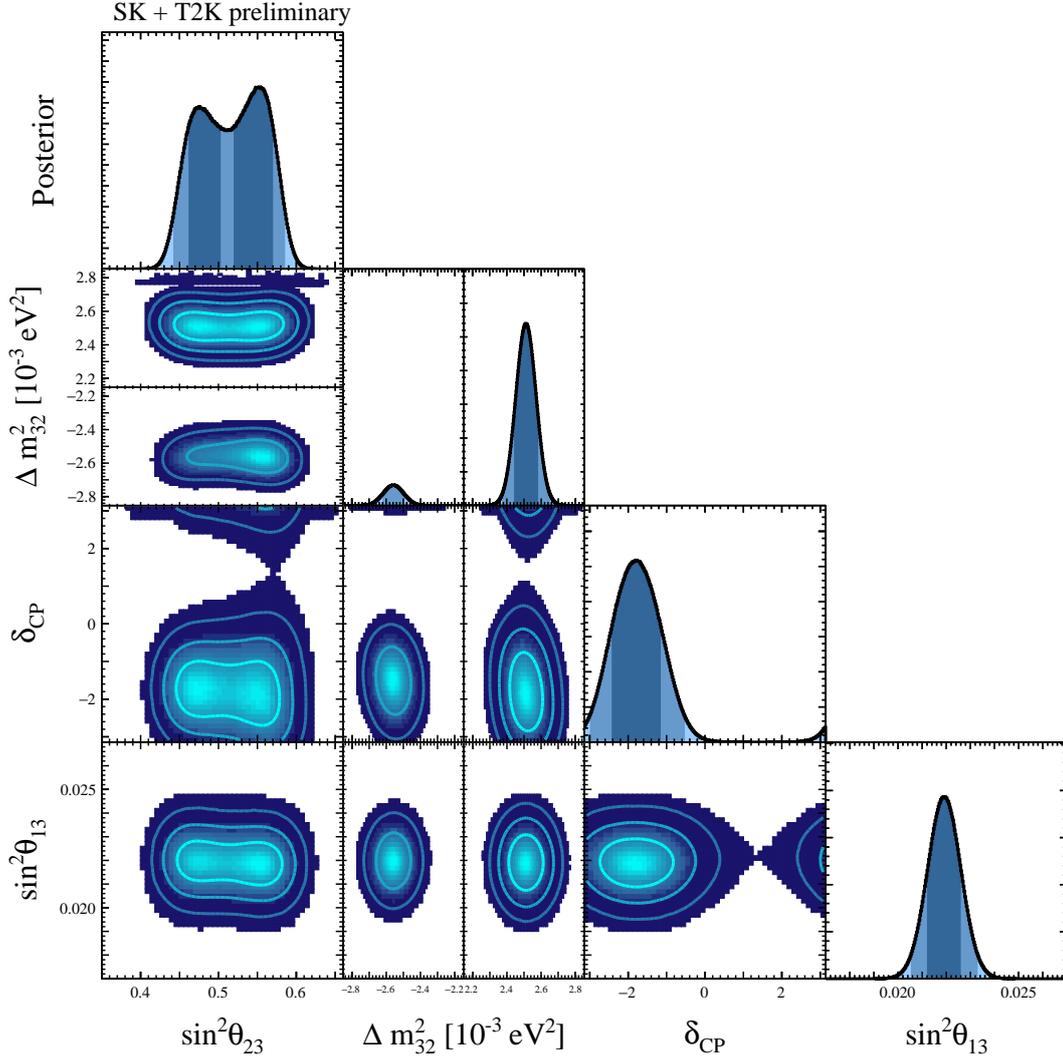
Figure 7.8. 2D posterior distributions computed with MCMC and marginalized over both ordering. 1σ, 2σ, and 3σ contours are shown together. The flat δ_{CP} prior is used and Δm^2_{32} smearing of 3.6×10^{-5} eV^2 is applied.

is classified as "substantial" evidence according to the Jeffreys' scale [167], "moderate" according to the Lee and Wagenmakers' scale [162], and "positive" according to the Kass and Raftery's scale [168].

In the traditional particle physics field, a threshold for a hypothesis being favored has been defined in terms of the confidence level in a frequentist sense. To obtain intuition of significance level in the traditional particle physics sense, we can find the Bayes factor corresponding to the posterior probability for a hypothesis is equal to the probability of obtaining a certain deviation from a Gaussian. The correspondence between these thresholds and Bayes factors are listed in Table 7.7. Therefore, the mass ordering Bayes factor we obtained from the joint analysis is equivalent to the significance level of 89.8%, which corresponds to 1.64σ.
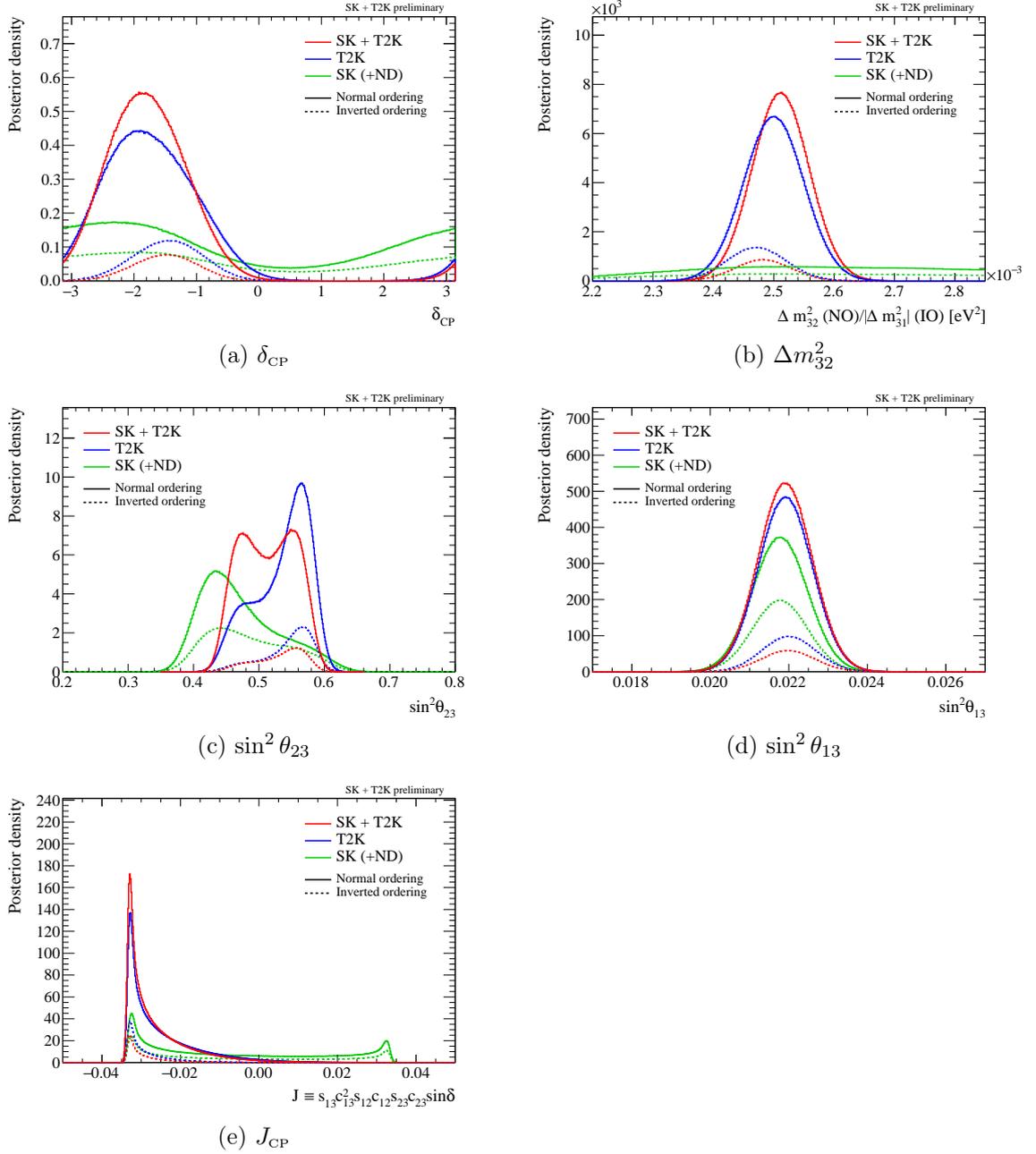
Figure 7.9. Comparison of the joint analysis SK+T2K, T2K, and SK (+ND) fit posterior probability densities obtained from MCMC. The posterior densities are normalized so that the sum of the area for normal and inverted ordering should be 1. Therefore, the ratio between the normal and inverted ordering densities shows the mass ordering preference. Δm^2_{32} **smearing is not applied here** because the smearing factor of 3.6×10^{-5} eV^2 is computed based on the SK+T2K fit to every fake data study and we do not have smearing factors for T2K and SK (+ND) fits. Therefore, the SK+T2K result differs from the final result shown in Fig. 7.7.
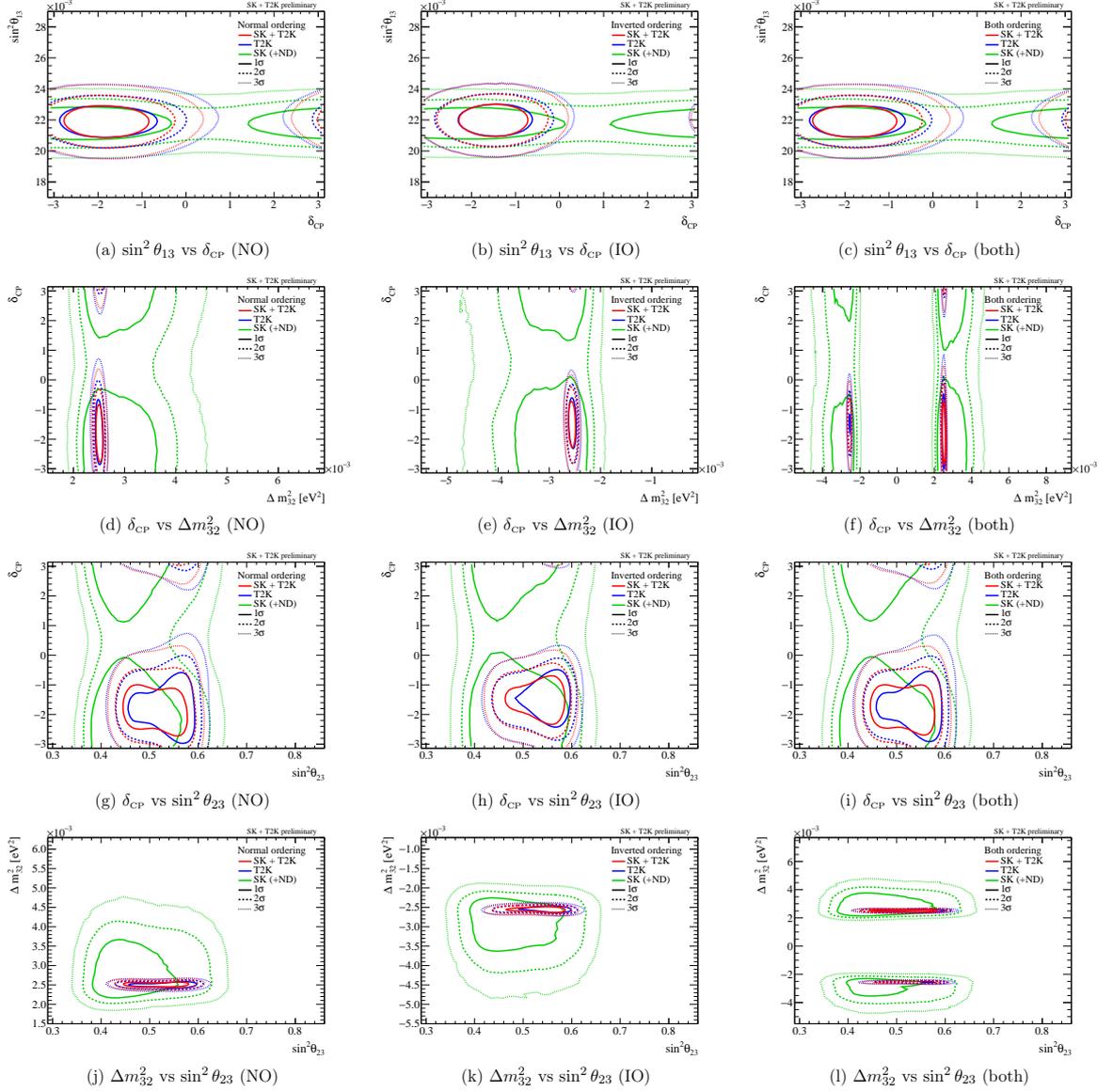
Figure 7.10. Comparison of joint analysis SK+T2K, T2K, and SK (+ND) fit 2D credible intervals obtained from MCMC. All fits are run with the reactor constraint on $\sin^2\theta_{13}$. Δm^2_{32} **smearing is not applied here** because the smearing factor of 3.6×10^{-5} eV^2 is computed based on the SK+T2K fit to every fake data study and we do not have smearing factors for T2K and SK (+ND) fits. Therefore, the SK+T2K result differs from the final result shown in Fig. 7.8.

Table 7.5. The most probable values and credible interval ranges obtained from MCMC. The flat δ_{CP} prior is used and Δm_{32}^2 smearing of 3.6×10^{-5} eV^2 is applied. The most probable value is taken to be the point where the posterior probability density becomes the maximum by applying the cubic interpolation to the bin contents. The corresponding plots are shown in Fig. 7.7.

Normal ordering	$\sin^2 \theta_{13}$	Δm_{32}^2 [10^{-3} eV^2]	$\sin^2 \theta_{23}$
Most probable value	0.0219	2.511	0.549
1σ	[0.0212, 0.0226]	[2.453, 2.570]	[0.459, 0.505] and [0.523, 0.568]
2σ	[0.0205, 0.0233]	[2.394, 2.630]	[0.442, 0.585]
3σ	[0.0199, 0.0240]	[2.335, 2.690]	[0.424, 0.603]
Inverted ordering	$\sin^2 \theta_{13}$	Δm_{32}^2 [10^{-3} eV^2]	$\sin^2 \theta_{23}$
Most probable value	0.0220	-2.558	0.557
1σ	[0.0213, 0.0227]	[-2.616, -2.499]	[0.508, 0.581]
2σ	[0.0206, 0.0234]	[-2.676, -2.440]	[0.455, 0.591]
3σ	[0.0199, 0.0240]	[-2.735, -2.382]	[0.430, 0.605]
Both ordering	$\sin^2 \theta_{13}$	Δm_{32}^2 [10^{-3} eV^2]	$\sin^2 \theta_{23}$
Most probable value	0.0219	2.510	0.553
1σ	[0.0212, 0.0226]	[2.443, 2.580]	[0.461, 0.502] and [0.520, 0.570]
2σ	[0.0206, 0.0233]	[-2.620, -2.494] and [2.373, 2.651]	[0.443, 0.585]
3σ	[0.0199, 0.0240]	[-2.701, -2.414] and [2.322, 2.703]	[0.424, 0.604]

Table 7.6. Posterior probabilities for different combinations of mass ordering and octant hypotheses for SK+T2K, joint analysis T2K-only, and SK (+ND). The Bayes factor for normal ordering (NO) over inverted ordering (IO) and the Bayes factor for upper octant (UO) over lower octant (LO) are shown together.

	SK+T2K			T2K			SK (+ND)		
	LO	UO	Line total	LO	UO	Line total	LO	UO	Line total
Normal ordering	0.367	0.533	0.900	0.190	0.642	0.832	0.468	0.186	0.654
Inverted ordering	0.022	0.078	0.100	0.025	0.142	0.168	0.214	0.132	0.346
Column total	0.389	0.611	1.000	0.215	0.785	1.000	0.682	0.318	1.000
BF(NO,IO)	8.98			4.96			1.89		
BF(UO,LO)	1.57			3.65			0.47		

Table 7.7. List of Bayes factors equating to traditional particle physics probability thresholds in the case of equal prior for the two hypotheses under test.

Threshold	$P(\theta_1|\text{Data})$	$P(\theta_2|\text{Data})$	$B(\theta_1, \theta_2)$
90%	0.9	0.1	9
2σ	0.954	0.046	20.74
99%	0.99	0.01	99
3σ	0.9973	0.0027	369.4

7.3.5   Goodness of fit

It is important to check the goodness of fit to ensure the validity of our analysis. In the context of Bayesian analysis, the posterior predictive p-value is used for this purpose. It measures how well the observed data agrees with our post-fit predictions [169].

To obtain the posterior predictive p-values, we first make many pseudo data sets by using both the oscillation and systematic uncertainty parameters sampled from the data-fit MCMC posterior distributions. Then we define the test statistic for pseudo data set i as

$$T(\boldsymbol{n}_i, \boldsymbol{\theta}_i) \equiv -2 \log \mathcal{L}_{\text{stat}}(\boldsymbol{n}_i|\boldsymbol{\lambda}_i) = 2 \sum_{j=1}^{N_{\text{bin}}} \left[\lambda_j(\boldsymbol{\theta}_i) - n_{i,j} + n_{i,j} \log \frac{n_{i,j}}{\lambda_j(\boldsymbol{\theta}_i)} \right], \qquad (7.2)$$

where the number of events $n_{i,j}$ in bin j is thrown from the Poisson distribution with the predicted event number $\lambda_j(\boldsymbol{\theta}_i)$. Finally, the posterior predictive p-value (p_{pp}) is defined as the fraction of the pseudo data sets that have larger test statistics $T(\boldsymbol{n}_i, \boldsymbol{\theta}_i)$ than the one computed using the observed data $T(\boldsymbol{n}_{\text{obs}}, \boldsymbol{\theta}_i)$:

$$p_{\text{pp}} = \frac{1}{N_{\text{pseudo}}} \sum_{i=1}^{N_{\text{pseudo}}} I\left(T(\boldsymbol{n}_i, \boldsymbol{\theta}_i) > T(\boldsymbol{n}_{\text{obs}}, \boldsymbol{\theta}_i)\right), \qquad (7.3)$$

where $I(\cdot)$ is an indicator function that returns 1 if the inequality is satisfied and returns 0 otherwise. For the sampling of the parameters, we use the posterior distribution of the MCMC without the multicanonical method ($\beta = 1.0$). We do not use the multicanonical method here because we want to equally sample the steps from all the posterior regions. In addition, to get rid of the effect of the MCMC autocorrelation, we sampled a step for every 2,000 steps from 200 million steps, which gives 100,000 steps in total. We compute the p-values using not only the binning used in the analysis (two-dimensional for most of the samples) but also a projected one-dimensional binning. This is because the analysis binning has many bins that receive few events and the p-values could be dominated by statistical fluctuation. The p-values in the 1D binning could mitigate this issue and are therefore expected to capture the systematic effects more effectively.

The computed posterior predictive p-values are summarized in Table 7.8. The distributions of the test statistics for the T2K beam, SK atmospheric, and total p-values are also shown in Fig. 7.11. The total p-values satisfy the standard criterion of $p > \alpha = 0.05$,

Table 7.8. Summary of the posterior predictive p-values for each sample.

Sample	Analysis binning		1D binning	
	p-value	Projection	p-value	Projection
FHC 1Rμ	0.695	$(E_{\mathrm{rec}}, \theta)$	0.539	(E_{rec})
RHC 1Rμ	0.962	$(E_{\mathrm{rec}}, \theta)$	0.833	(E_{rec})
FHC 1Re	0.844	(p, θ)	0.168	(E_{rec})
RHC 1Re	0.697	(p, θ)	0.658	(E_{rec})
FHC 1Re 1de	0.182	(p, θ)	0.907	(E_{rec})
SubGeV elike 0de	0.007	$(p, \cos\Theta_z)$	0.052	$(\cos\Theta_z)$
SubGeV elike 1de	0.664	(p)	0.662	(p)
SubGeV mulike 0de	0.831	$(p, \cos\Theta_z)$	0.440	$(\cos\Theta_z)$
SubGeV mulike 1de	0.361	$(p, \cos\Theta_z)$	0.447	$(\cos\Theta_z)$
SubGeV mulike 2de	0.572	(p)	0.573	(p)
SubGeV pi0like	0.221	(p)	0.220	(p)
MultiGeV elike nue	0.294	$(p, \cos\Theta_z)$	0.123	$(\cos\Theta_z)$
MultiGeV elike nuebar	0.139	$(p, \cos\Theta_z)$	0.884	$(\cos\Theta_z)$
MultiGeV mulike	0.698	$(p, \cos\Theta_z)$	0.844	$(\cos\Theta_z)$
MultiRing elike nue	0.386	$(p, \cos\Theta_z)$	0.740	$(\cos\Theta_z)$
MultiRing elike nuebar	0.186	$(p, \cos\Theta_z)$	0.608	$(\cos\Theta_z)$
MultiRingOther	0.408	$(p, \cos\Theta_z)$	0.637	$(\cos\Theta_z)$
MultiRing mulike	0.370	$(p, \cos\Theta_z)$	0.826	$(\cos\Theta_z)$
PCStop	0.497	$(p, \cos\Theta_z)$	0.398	$(\cos\Theta_z)$
PCThru	0.253	$(p, \cos\Theta_z)$	0.319	$(\cos\Theta_z)$
UpStop mu	0.590	$(p, \cos\Theta_z)$	0.360	$(\cos\Theta_z)$
UpThruNonShower mu	0.031	$(\cos\Theta_z)$	0.032	$(\cos\Theta_z)$
UpThruShower mu	0.449	$(\cos\Theta_z)$	0.448	$(\cos\Theta_z)$
T2K beam	0.944		0.763	
SK atmospheric	0.041		0.331	
Total	0.422		0.571	

which implies our fit is not a bad description of the data. The *p*-value for the total atmospheric samples does not satisfy this criterion, but since this is consistent with the value obtained in the SK-standalone analysis [78], we think this is acceptable.

We should be careful about the so-called "look elsewhere effect" when we look at

(a) T2K beam (analysis binning) (b) SK atmospheric (analysis binning) (c) Total (analysis binning)

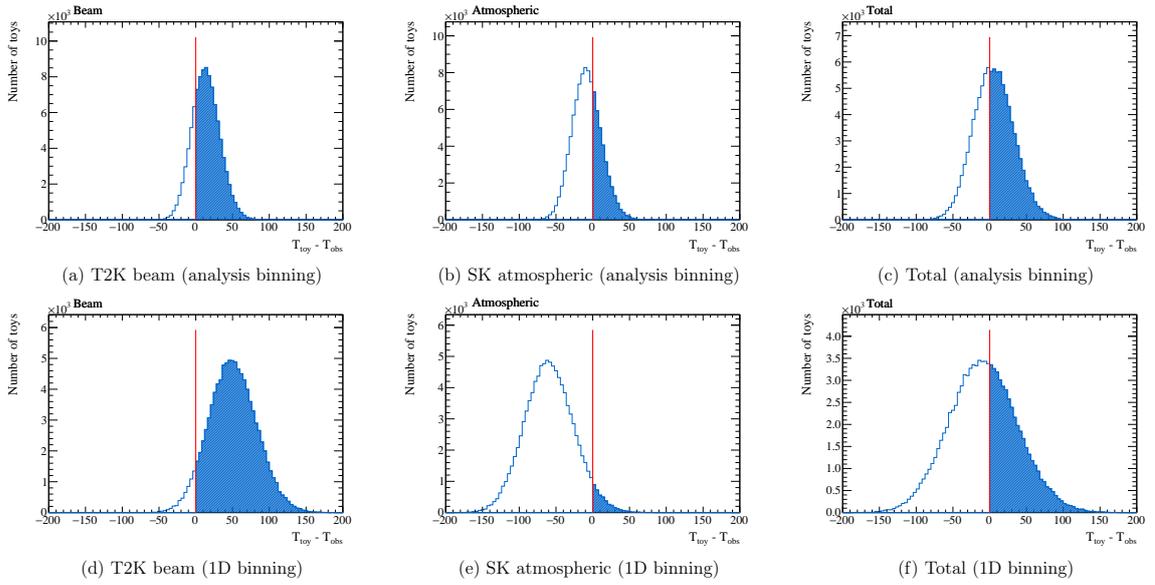(d) T2K beam (1D binning) (e) SK atmospheric (1D binning) (f) Total (1D binning)

Figure 7.11. Distributions of the test statistics used to compute the posterior predictive p-values for total T2K beam samples, SK atmospheric samples, and all samples, respectively. The top row shows the distributions with the analysis binning and the bottom row shows the distributions with the 1D binning. The shaded region shows the toys that satisfy $T(\boldsymbol{n}, \boldsymbol{\theta}) > T(\boldsymbol{n}_{\mathrm{obs}}, \boldsymbol{\theta})$.

the p-value for each sample. Since we are testing whether the deviations between the observed data and predictions are statistical fluctuations or not, it would be a too strict condition if we require all the samples to satisfy the same $\alpha = 0.05$ criterion. If we repeat the test of p-values many times, some of them can break the criterion due to pure statistical fluctuation. Therefore, we should correct our criterion to take into account this look elsewhere effect, and one of the well-known countermeasures to this kind of problem is the Holm-Bonferroni correction [170]. In the Holm-Bonferroni method, we sort the obtained p-values (let us say we have n p-values) in increasing order and apply the criterion of $p_i > \alpha/(n - i)$ for i-th p-value ($i = 0, 1, \cdots, n - 1$). The test is repeated from $i = 0$ (i.e. the smallest p-value) to $i = n - 1$ one by one, and we consider that the i-th and all the following tests pass the criteria when the i-th p-value satisfies the criterion $p_i > \alpha/(n - i)$. In this study, we have $n = 23$ tests and the smallest p-values are 0.007 (sub-GeV e-like 0de) for the analysis binning and 0.032 (UpThru non-shower μ) for the 1D binning. Consequently, these p-values satisfy the first test ($p > 0.05/23 = 0.0022$) and all the samples automatically pass the tests. In conclusion, we found no problem with these p-values and therefore conclude that the goodness of fit in this analysis is acceptable level.

The 1D posterior predictive distributions for all the samples are shown in Fig. 7.12. These distributions are drawn using the same steps sampled from the MCMC posterior distributions. The UpThru non-shower μ sample, which showed the worst p-value in the 1D binning, has a clear data/MC excess in the most horizontal bin ($\cos\Theta_z \in [-0.1, 0.0]$). This is also seen in the SK-standalone analyses [52, 36]. One possible reason is the underestimation of the background events such as cosmic muons, but the exact reason

has not been understood yet. Similarly, the atmospheric sub-GeV e-like 0de sample has a data/MC excess in the most upward-going bin ($\cos\Theta_z \in [-1.0, -0.8]$), which might be a statistical fluctuation but could affect the oscillation parameter constraints. We will have more detailed discussions in Section 8.3.2.
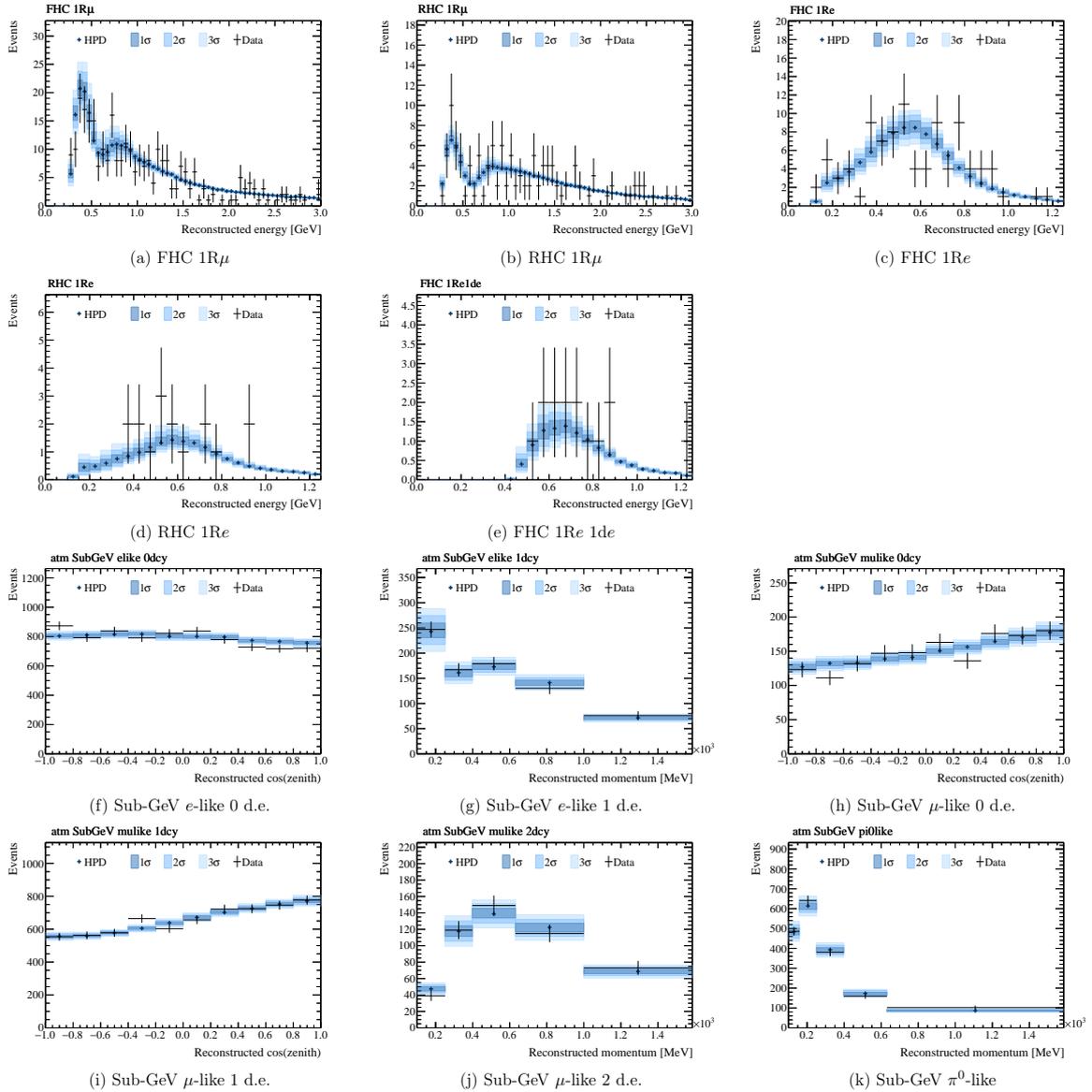

(a) FHC 1Rμ (b) RHC 1Rμ (c) FHC 1Re

(d) RHC 1Re (e) FHC 1Re 1de

(f) Sub-GeV e-like 0 d.e.   (g) Sub-GeV e-like 1 d.e.   (h) Sub-GeV μ-like 0 d.e.

(i) Sub-GeV μ-like 1 d.e.   (j) Sub-GeV μ-like 2 d.e.   (k) Sub-GeV π^0-like

Figure 7.12. Posterior predictive distributions for the T2K beam and SK atmospheric sub-GeV samples. The highest posterior density point is drawn with a cross marker and 1, 2, and 3σ credible intervals are drawn with the shaded histograms.

(l) Multi-GeV e-like ν_e (m) Multi-GeV e-like $\bar{\nu}_e$ (n) Multi-GeV μ-like

(o) Multi-Ring e-like ν_e (p) Multi-Ring e-like $\bar{\nu}_e$ (q) Multi-Ring μ-like

(r) Multi-Ring other   (s) PCStop   (t) PCThru

(u) UpStop μ (v) UpThru non-shower μ (w) UpThru shower μ
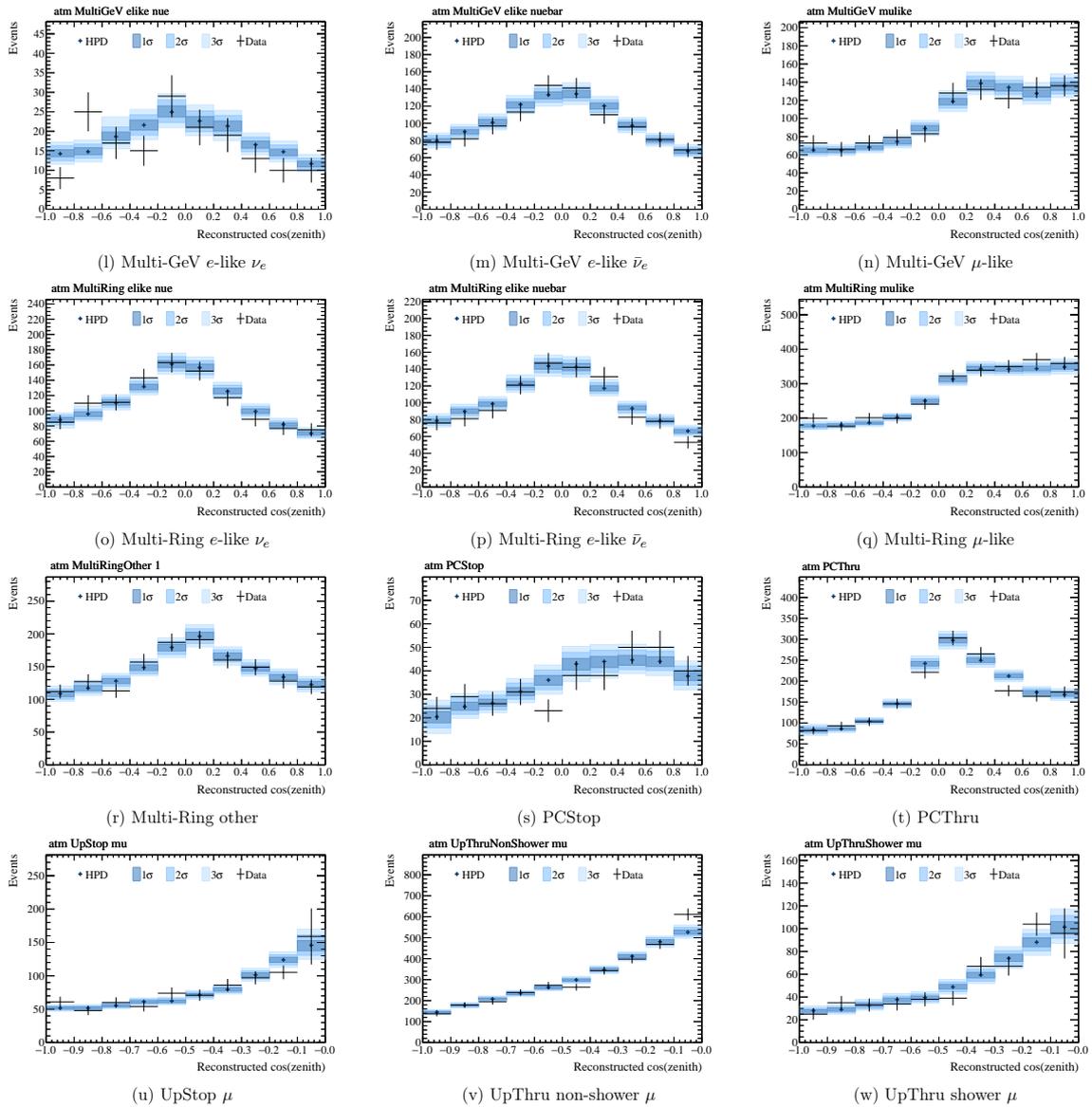
Figure 7.12. Posterior predictive distributions for the SK atmospheric muli-GeV, PC, and UpMu samples. The highest posterior density point is drawn with a cross marker and 1, 2, and 3σ credible intervals are drawn with the shaded histograms.

7.3.6   Impact of the alternative model effect

As described in Section 6.1.3, we test the effects of alternative models on the $\delta_{\rm CP}$ and $J_{\rm CP}$ credible interval edges. We tested twelve fake data studies in total, but only the atmospheric CC1π excess FDS is described here, representative of the others.

7.3.6.1   Atmospheric CC1π excess

The effects of alternative models on the $\delta_{\rm CP}$ and $J_{\rm CP}$ credible intervals are tested with both the prior flat in $\delta_{\rm CP}$ and the prior flat in $\sin \delta_{\rm CP}$. The data fit posterior distribution and the distribution shifted by taking into account the atmospheric CC1π FDS effects are shown in Figs. 7.13 and 7.14. The corresponding credible intervals of $1\sigma, 1.5\sigma, 2\sigma, 3\sigma$ are shown with the vertical dotted/dashed lines.

When we apply the shift from this fake data study, the left edges of the credible intervals of $\delta_{\rm CP}$ move closer to $\delta_{\rm CP} = -\pi$. However, the 2σ credible interval under the flat $\delta_{\rm CP}$ prior and the 1.5σ credible interval under the flat $\sin \delta_{\rm CP}$ prior still exclude the CP-conserving values of $\delta_{\rm CP}$. Since the effect was mostly seen in $\cos \delta_{\rm CP}$, the shift in $J_{\rm CP}$ was not as large as that in $\delta_{\rm CP}$. Therefore, our conclusion on the Jarlskog invariant is also not affected by this data/MC excess.
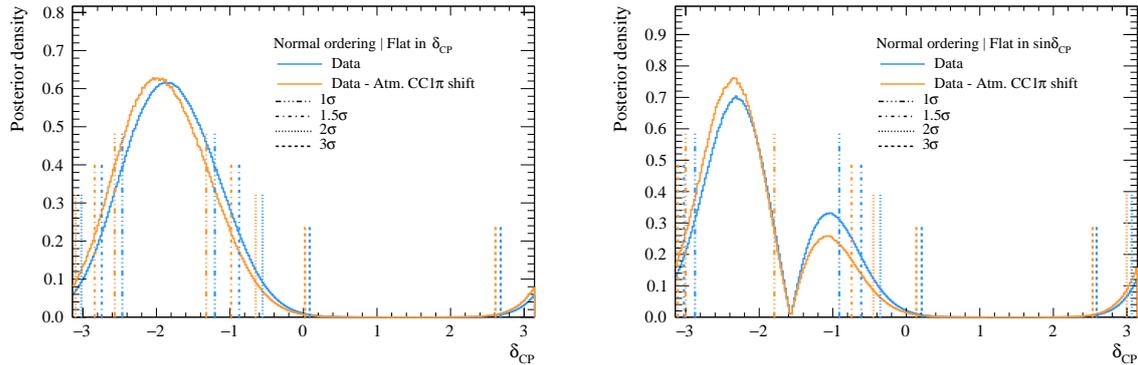
Figure 7.13. Comparison of the $\delta_{\rm CP}$ 1D posterior distribution and credible intervals from the data with and without the shift from the atmospheric CC1π fake data study. The flat prior in $\delta_{\rm CP}$ is applied on the left, and the flat prior in $\sin \delta_{\rm CP}$ is applied on the right.
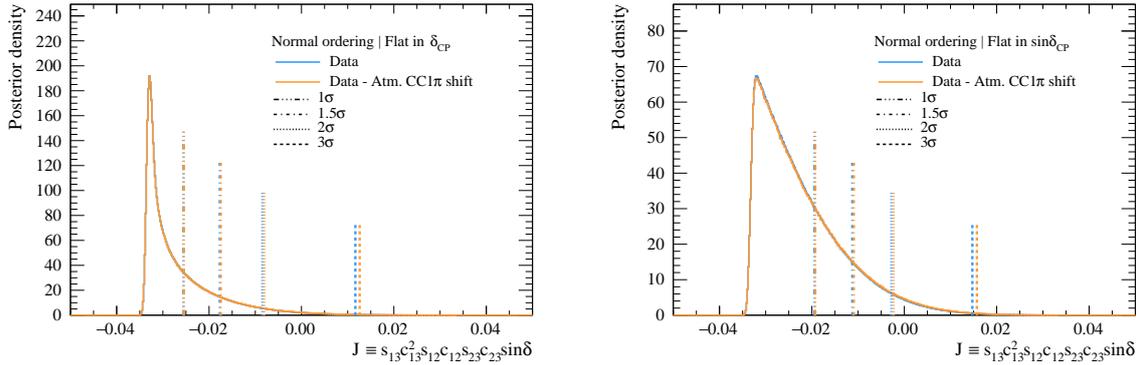
Figure 7.14. Comparison of the Jarlskog invariant 1D posterior distribution and credible intervals from the data with and without the shift from the atmospheric CC1π fake data study. The flat prior in δ_{CP} is applied on the left, and the flat prior in $\sin \delta_{\mathrm{CP}}$ is applied on the right.

7.3.6.2   Summary

Tables 7.9 to 7.12 show the shift in δ_{CP} and J_{CP} credible interval edges with the priors flat in δ_{CP} and flat in $\sin \delta_{\mathrm{CP}}$. The largest shifts were observed not from a single fake data study but from a few different fake data studies depending on the interval edges. None of the studies causes a shift over a value of interest for δ_{CP} at the 2σ level under the flat δ_{CP} prior and at the 1.5σ level under the flat $\sin \delta_{\mathrm{CP}}$ prior. The conclusion on the Jarlskog invariant (exclusion at 2σ) is not affected.

Therefore, our final conclusion, the conservation of CP symmetry is excluded at 2σ under the prior flat in δ_{CP} and at 1.5σ under the prior flat in $\sin \delta_{\mathrm{CP}}$, is robust against the possible out-of-model effects.

Table 7.9. Summary of the fake data study effects on the δ_{CP} credible intervals with the **prior flat in $\delta_{\mathbf{CP}}$**, calculated for the Asimov A oscillation parameter set. The largest positive and negative shift for each interval boundary is highlighted with the red and blue background colors, respectively.

Credible interval		1σ		1.5σ		2σ		3σ	
Data fit interval edge		$b_{\delta_{\mathrm{CP}}}^{1\sigma,\downarrow}$	$b_{\delta_{\mathrm{CP}}}^{1\sigma,\uparrow}$	$b_{\delta_{\mathrm{CP}}}^{1.5\sigma,\downarrow}$	$b_{\delta_{\mathrm{CP}}}^{1.5\sigma,\uparrow}$	$b_{\delta_{\mathrm{CP}}}^{2\sigma,\downarrow}$	$b_{\delta_{\mathrm{CP}}}^{2\sigma,\uparrow}$	$b_{\delta_{\mathrm{CP}}}^{3\sigma,\downarrow}$	$b_{\delta_{\mathrm{CP}}}^{3\sigma,\uparrow}$
		-2.464	-1.205	-2.743	-0.875	-3.021	-0.556	-3.601	0.085
Fake data	Reference	$\Delta b_{\delta_{\mathrm{CP}}}^{1\sigma,\downarrow}$	$\Delta b_{\delta_{\mathrm{CP}}}^{1\sigma,\uparrow}$	$\Delta b_{\delta_{\mathrm{CP}}}^{1.5\sigma,\downarrow}$	$\Delta b_{\delta_{\mathrm{CP}}}^{1.5\sigma,\uparrow}$	$\Delta b_{\delta_{\mathrm{CP}}}^{2\sigma,\downarrow}$	$\Delta b_{\delta_{\mathrm{CP}}}^{2\sigma,\uparrow}$	$\Delta b_{\delta_{\mathrm{CP}}}^{3\sigma,\downarrow}$	$\Delta b_{\delta_{\mathrm{CP}}}^{3\sigma,\uparrow}$
CRPA	Normal Asimov	0.020	0.043	0.012	0.052	0.014	0.055	-0.002	0.076
LFG	Normal Asimov	-0.051	-0.010	-0.054	0.001	-0.066	0.024	-0.124	0.084
Removal energy	Normal Asimov	0.000	-0.021	0.005	-0.022	0.011	-0.023	0.009	-0.011
Axial form factor	Scaled Asimov	0.002	0.001	0.005	0.002	0.013	0.002	0.009	0.015
Martini 2p2h	Scaled Asimov	0.040	0.124	0.026	0.128	0.005	0.124	-0.092	0.159
Pion multiplicity	Normal Asimov	-0.019	-0.021	-0.011	-0.015	-0.004	-0.018	-0.008	-0.015
Energy-dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$	Scaled Asimov	0.004	0.040	0.008	0.028	0.004	0.029	-0.011	0.031
Atmospheric CC1π	Normal Asimov	-0.103	-0.118	-0.096	-0.108	-0.079	-0.093	-0.071	-0.065
ND CC1π	Normal Asimov	0.035	0.024	0.031	0.005	0.025	-0.005	0.024	-0.013
ND Non-QE CC0πi	Normal Asimov	-0.007	0.036	-0.013	0.043	-0.023	0.049	-0.061	0.095
ND extrapolation	Normal Asimov	-0.012	0.032	-0.010	0.051	-0.004	0.047	-0.039	0.044
Pion SI bug fix	Normal Asimov	0.020	0.029	0.007	0.021	0.007	0.018	-0.008	0.012

Table 7.10. Summary of the fake data study effects on the $\delta_{\rm CP}$ credible intervals with the **prior flat in** $\sin \delta_{\rm CP}$, calculated for the Asimov A oscillation parameter set. The largest positive and negative shift for each interval boundary is highlighted with the red and blue background colors, respectively.

Credible interval		1σ		1.5σ		2σ		3σ	
Data fit interval edge		$b_{\delta_{\rm CP}}^{1\sigma,\downarrow}$	$b_{\delta_{\rm CP}}^{1\sigma,\uparrow}$	$b_{\delta_{\rm CP}}^{1.5\sigma,\downarrow}$	$b_{\delta_{\rm CP}}^{1.5\sigma,\uparrow}$	$b_{\delta_{\rm CP}}^{2\sigma,\downarrow}$	$b_{\delta_{\rm CP}}^{2\sigma,\uparrow}$	$b_{\delta_{\rm CP}}^{3\sigma,\downarrow}$	$b_{\delta_{\rm CP}}^{3\sigma,\uparrow}$
		-2.873	-0.912	-3.024	-0.614	-3.215	-0.353	-3.694	0.210
Fake data	Reference	$\Delta b_{\delta_{\rm CP}}^{1\sigma,\downarrow}$	$\Delta b_{\delta_{\rm CP}}^{1\sigma,\uparrow}$	$\Delta b_{\delta_{\rm CP}}^{1.5\sigma,\downarrow}$	$\Delta b_{\delta_{\rm CP}}^{1.5\sigma,\uparrow}$	$\Delta b_{\delta_{\rm CP}}^{2\sigma,\downarrow}$	$\Delta b_{\delta_{\rm CP}}^{2\sigma,\uparrow}$	$\Delta b_{\delta_{\rm CP}}^{3\sigma,\downarrow}$	$\Delta b_{\delta_{\rm CP}}^{3\sigma,\uparrow}$
CRPA	Normal Asimov	0.024	0.090	0.018	0.059	0.008	0.071	0.009	0.081
LFG	Normal Asimov	-0.059	-0.001	-0.059	0.009	-0.063	0.034	-0.102	0.075
Removal energy	Normal Asimov	-0.005	-0.022	0.001	-0.017	0.004	-0.013	0.013	-0.015
Axial form factor	Scaled Asimov	0.006	0.010	0.005	0.011	0.007	0.013	0.004	0.010
Martini 2p2h	Scaled Asimov	0.041	0.187	0.010	0.143	-0.012	0.130	-0.078	0.134
Pion multiplicity	Normal Asimov	-0.018	-0.027	-0.004	-0.028	-0.008	-0.018	-0.002	-0.020
Energy-dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$	Scaled Asimov	0.008	0.057	0.007	0.038	0.001	0.030	-0.005	0.027
Atmospheric CC1π	Normal Asimov	-0.128	-0.882	-0.083	-0.130	-0.071	-0.095	-0.056	-0.074
ND CC1π	Normal Asimov	0.029	0.021	0.021	0.003	0.017	-0.007	0.014	-0.011
ND Non-QE CC0πi	Normal Asimov	-0.007	0.062	-0.015	0.054	-0.024	0.062	-0.059	0.089
ND extrapolation	Normal Asimov	-0.000	0.071	-0.003	0.055	-0.015	0.056	-0.046	0.026
Pion SI bug fix	Normal Asimov	0.009	0.044	0.004	0.027	0.000	0.025	0.000	0.017

Table 7.11. Summary of the fake data study effects on the Jarlskog credible intervals ($\times 10^2$) with the **prior flat in $\delta_{\mathbf{CP}}$**, calculated for the Asimov A oscillation parameter set. The largest positive and negative shift for each interval boundary is highlighted with the red and blue background colors, respectively.

Credible interval		1σ	1.5σ	2σ	3σ
Data fit interval edge		$b_{J_{\mathrm{CP}}}^{1\sigma,\uparrow}$	$b_{J_{\mathrm{CP}}}^{1.5\sigma,\uparrow}$	$b_{J_{\mathrm{CP}}}^{2\sigma,\uparrow}$	$b_{J_{\mathrm{CP}}}^{3\sigma,\uparrow}$
		-2.55487	-1.76817	-0.84733	1.16596
Fake data	Reference	$\Delta b_{J_{\mathrm{CP}}}^{1\sigma,\uparrow}$	$\Delta b_{J_{\mathrm{CP}}}^{1.5\sigma,\uparrow}$	$\Delta b_{J_{\mathrm{CP}}}^{2\sigma,\uparrow}$	$\Delta b_{J_{\mathrm{CP}}}^{3\sigma,\uparrow}$
LFG	Normal Asimov	0.06214	0.12216	0.18769	0.35173
CRPA	Normal Asimov	0.02526	0.04391	0.05763	0.05263
Removal energy	Normal Asimov	-0.01572	-0.02499	-0.03246	-0.03936
Axial form factor	Scaled Asimov	0.00369	0.00539	0.00478	-0.01819
Pion multiplicity	Normal Asimov	-0.00677	-0.01109	-0.01248	-0.01094
Energy-dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$	Scaled Asimov	0.01853	0.02783	0.01695	0.03292
Martini 2p2h	Scaled Asimov	0.08223	0.12561	0.16956	0.31376
Atmospheric CC1π	Normal Asimov	0.00825	0.02197	0.04142	0.09370
ND CC1π	Normal Asimov	-0.02609	-0.04500	-0.05814	-0.09331
ND Non-QE CC0πi	Normal Asimov	0.04081	0.07889	0.12344	0.23962
ND extrapolation	Normal Asimov	0.05207	0.06842	0.06595	0.10836
Pion SI bug fix	Normal Asimov	0.00692	0.01166	0.01553	-0.00169

Table 7.12. Summary of the fake data study effects on the Jarlskog credible intervals ($\times 10^2$) with the **prior flat in** $\sin \delta_{\mathbf{CP}}$, calculated for the Asimov A oscillation parameter set. The largest positive and negative shift for each interval boundary is highlighted with the red and blue background colors, respectively.

Credible interval		1σ	1.5σ	2σ	3σ
Data fit interval edge		$b_{J_{\mathrm{CP}}}^{1\sigma,\uparrow}$	$b_{J_{\mathrm{CP}}}^{1.5\sigma,\uparrow}$	$b_{J_{\mathrm{CP}}}^{2\sigma,\uparrow}$	$b_{J_{\mathrm{CP}}}^{3\sigma,\uparrow}$
		-1.94423	-1.12605	-0.28100	1.47322
Fake data	Reference	$\Delta b_{J_{\mathrm{CP}}}^{1\sigma,\uparrow}$	$\Delta b_{J_{\mathrm{CP}}}^{1.5\sigma,\uparrow}$	$\Delta b_{J_{\mathrm{CP}}}^{2\sigma,\uparrow}$	$\Delta b_{J_{\mathrm{CP}}}^{3\sigma,\uparrow}$
CRPA	Normal Asimov	0.03353	0.04988	0.04927	0.00987
LFG	Normal Asimov	0.09193	0.14158	0.18929	0.27265
Removal energy	Normal Asimov	-0.01880	-0.02378	-0.03348	-0.03454
Axial form factor	Scaled Asimov	0.00398	0.00353	0.00201	0.00894
Martini 2p2h	Scaled Asimov	0.07254	0.11415	0.13642	0.25058
Pion multiplicity	Normal Asimov	-0.00791	-0.02123	-0.00935	-0.00479
Energy-dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$	Scaled Asimov	0.01935	0.02497	0.01559	0.03006
Atmospheric CC1π	Normal Asimov	0.01701	0.03388	0.05006	0.09828
ND CC1π	Normal Asimov	-0.03403	-0.04480	-0.04922	-0.04228
ND Non-QE CC0πi	Normal Asimov	0.06223	0.09266	0.12251	0.17649
ND extrapolation	Normal Asimov	0.05642	0.04566	0.02120	0.06343
Pion SI bug fix	Normal Asimov	0.00954	0.01201	0.01339	0.00041

7.4 Summary of the Bayesian analysis results

In this chapter, the first data fit results of the joint analysis of T2K beam and SK atmospheric neutrinos are presented. We achieved improvement in the δ_{CP} constraints and mass ordering determination, compared to the fit to the samples from individual analyses.

The 2σ credible interval for δ_{CP} excludes the CP conserving values ($\delta_{\mathrm{CP}} = 0, \pi$) under the flat δ_{CP} prior but does not exclude them with the flat prior in $\sin\delta_{\mathrm{CP}}$. For the Jarlskog invariant J_{CP}, the CP conserving value ($J_{\mathrm{CP}} = 0$) is consistently excluded from the 2σ credible intervals under both the flat prior in δ_{CP} and the flat prior in $\sin\delta_{\mathrm{CP}}$.

The conclusions on CP conservation for different parameters and different priors are summarized in Table 7.13. Although the conclusion depends on which parameter to look at (δ_{CP} or J_{CP}), we would like to adopt the most conservative statement for each prior choice as our final conclusion: **conservation of CP symmetry is excluded at 2σ under the flat prior in $\delta_{\mathbf{CP}}$ and at 1.5σ under the flat prior in $\sin\delta_{\mathbf{CP}}$.** We confirmed that these results are robust against the out-of-model effects.

Table 7.13. Summary of the conclusion on CP conservation based on the Bayesian credible intervals. "✓" means that the CP conserving values are excluded from the credible intervals while "×" means they are not. These results are robust against the out-of-model effects studied through the fake data studies.

Parameter	Prior	1σ	1.5σ	2σ	3σ
δ_{CP}	Flat in δ_{CP}	✓	✓	✓	×
	Flat in $\sin\delta_{\mathrm{CP}}$	✓	✓	×	×
J_{CP}	Flat in δ_{CP}	✓	✓	✓	×
	Flat in $\sin\delta_{\mathrm{CP}}$	✓	✓	✓	×

The Bayes factor for the normal ordering over inverted ordering is computed to be 8.98 ± 0.06 under the flat δ_{CP} prior, which corresponds to a significance level of 1.64σ. It suggests a moderate preference for normal ordering, but it is not enough to claim a discovery. The joint analysis shows almost equal preference to both θ_{23} octants.

In addition, the posterior predictive p-value of 0.422 (all samples in the analysis binning) shows no problem in the goodness of fit.

Chapter 8

Discussions

The results of our main Bayesian analysis have been presented in Chapter 7, but there are some remaining questions we want to investigate. In particular, we discuss the contribution of each sample, the constraints on the systematic uncertainties, and alternative ways to test the CP conservation (violation) hypothesis in this chapter. The studies presented in this chapter are only intended to validate and understand the results, and therefore, the results of the main analysis will remain unchanged.

8.1   Data best-fit results

8.1.1   Best-fit χ^2 breakdown by sample

Using the best-fit χ^2, one can study the contribution of each sample to the mass ordering and octant preferences. Figure 8.1 shows the χ^2 differences between the normal/inverted ordering and upper/lower octant hypotheses broken down for each sample. The positive values (in orange) indicate the preference for normal ordering and upper octant, respectively.

For the mass ordering, the largest χ^2 contributions come from the atmospheric multi-GeV samples. Many of the samples show a preference for normal ordering, but the atmospheric multi-GeV e-like ν_e and multi-ring μ-like samples show the opposite preference. For the octant, the T2K samples show a preference for the upper octant, while the atmospheric samples show a preference for the lower octant except for the multi-ring e-like ν_e.

8.1.2   Systematic uncertainty constraints

The post-fit constraints on the systematic uncertainty parameters can be studied using the profiling fit results. The best-fit values and the estimated errors of the systematic uncertainty parameters for the ND, SK (+ND), T2K, and SK+T2K fits are shown in Fig. 8.2. The errors are computed from the Hessian at the best-fit point as described in Section 5.3.1. Since the 2p2h parameters are given a flat prior and strongly pulled against the boundary, it is not easy to estimate the post-fit uncertainty for these parameters. Therefore, they are fixed at their post-fit values when estimating the uncertainties for all

(a) Mass ordering
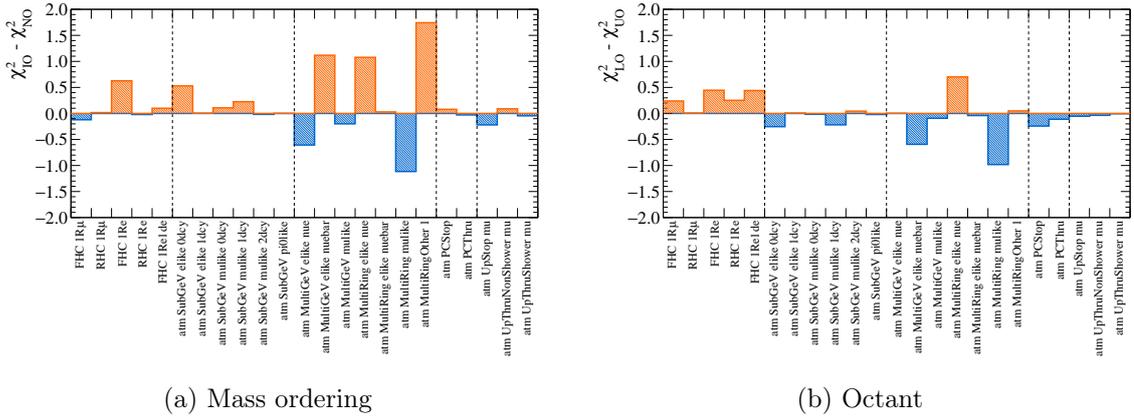
(b) Octant

Figure 8.1. Differences of the best-fit χ^2 between the normal/inverted ordering and upper/lower octant. In the mass ordering (octant) plot, the positive values (in orange) indicate a preference for normal ordering (upper octant), and the negative values (in blue) indicate a preference for inverted ordering (lower octant).

the other systematic uncertainty parameters and are drawn with 1σ error in the plots for convenience.

Flux systematics In Fig. 8.2a, the T2K flux systematics show mostly consistent results in all the cases. These parameters are well-constrained by the external experiments and the T2K near detector fit. The near detector pulls up the flux parameters at the lower energy regions and pulls down them at the higher energy regions. The opposite behavior is seen for example in the Q^2 cross-section parameters (low-Q^2 parameters are pulled down), which balances the overall event number predictions. The atmospheric flux systematics shown in Fig. 8.2b are constrained by the atmospheric samples only.

Cross-section systematics Among the cross-section systematic uncertainty parameters developed for this analysis, both of the Adler angle systematics (`Adler angle lowp`, `Adler angle highp` in Fig. 8.2d) are pulled lower and they are about 2-3 post-fit σ away from the nominal value. These Adler angle systematics are designed to make the pion momentum lower when the parameter moves to a smaller value. When the pion momentum gets below the Cherenkov threshold of $p_\pi = 157$ MeV, it cannot be detected as a ring but can produce the decay electron signal. Since both the T2K and SK sub-GeV CC1π samples are defined using the number of decay electrons, it increases the number of CC1π events when the parameter is pulled lower. Therefore, the direction of the observed shift is reasonable to cover the data excess in the CC1π samples. From the size of the errors, it turned out that these Adler angle systematics are mostly constrained by the atmospheric samples.
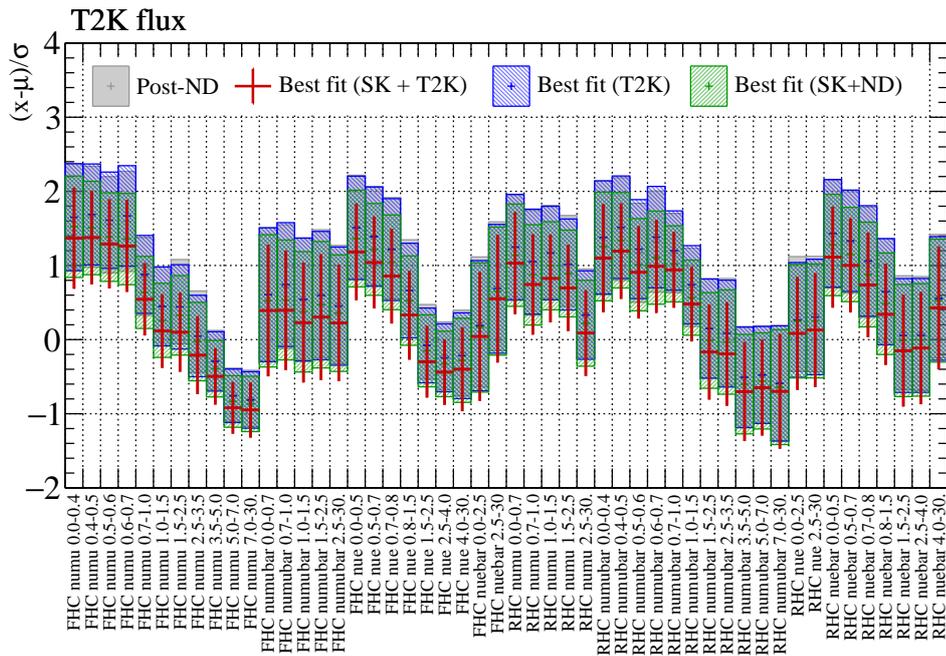
(a) T2K flux systematics
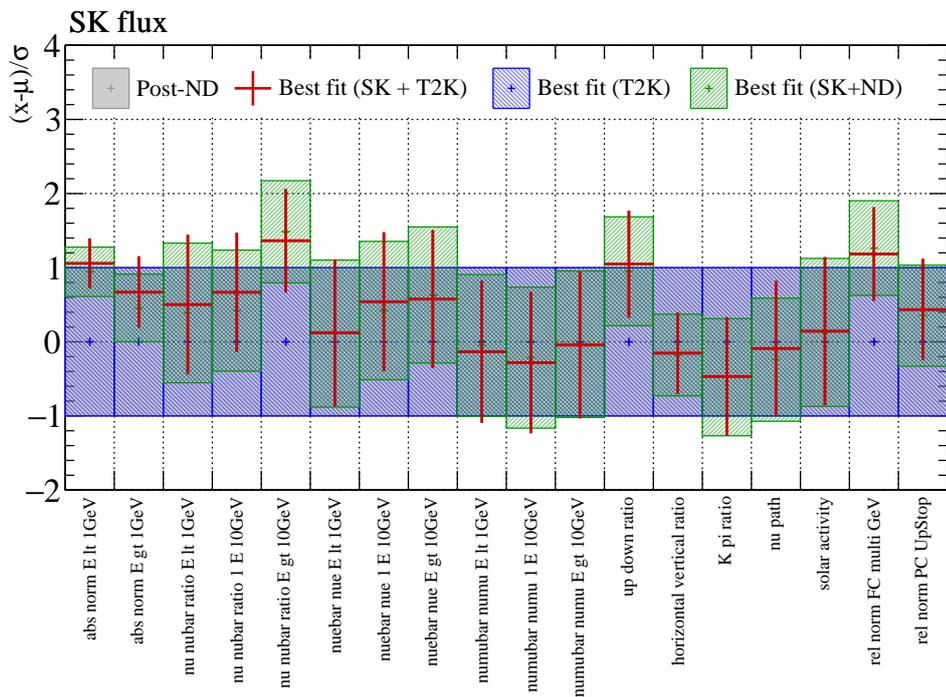
(b) SK atmospheric flux systematics

Figure 8.2. Comparison of the **flux** systematic parameter best fits among SK+T2K, T2K, and SK (+ND) fits. Parameters are normalized with their prior uncertainties.
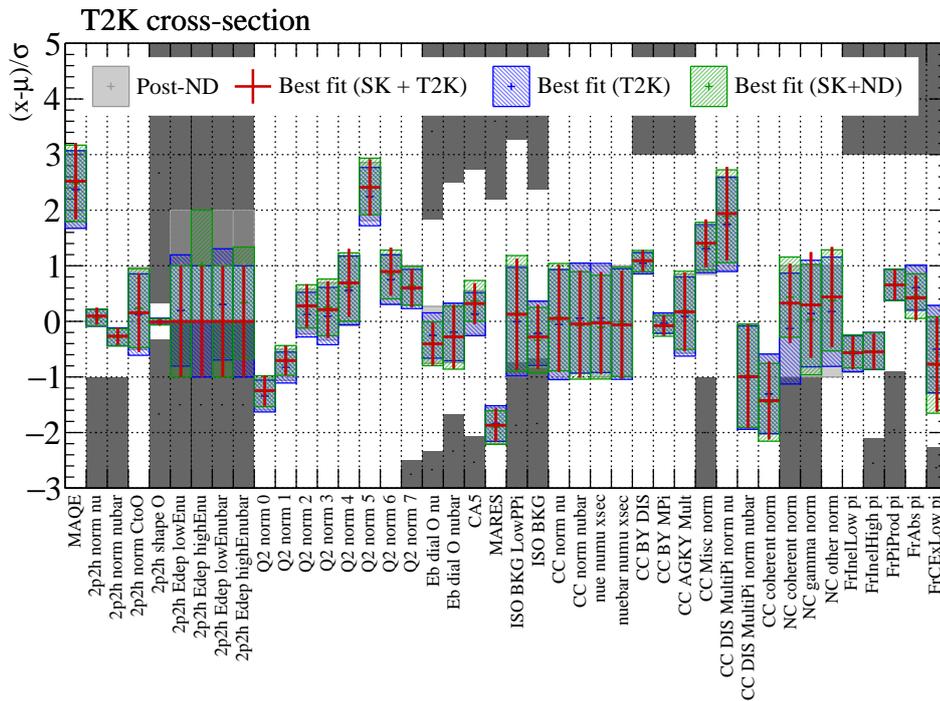
(c) T2K cross-section systematics
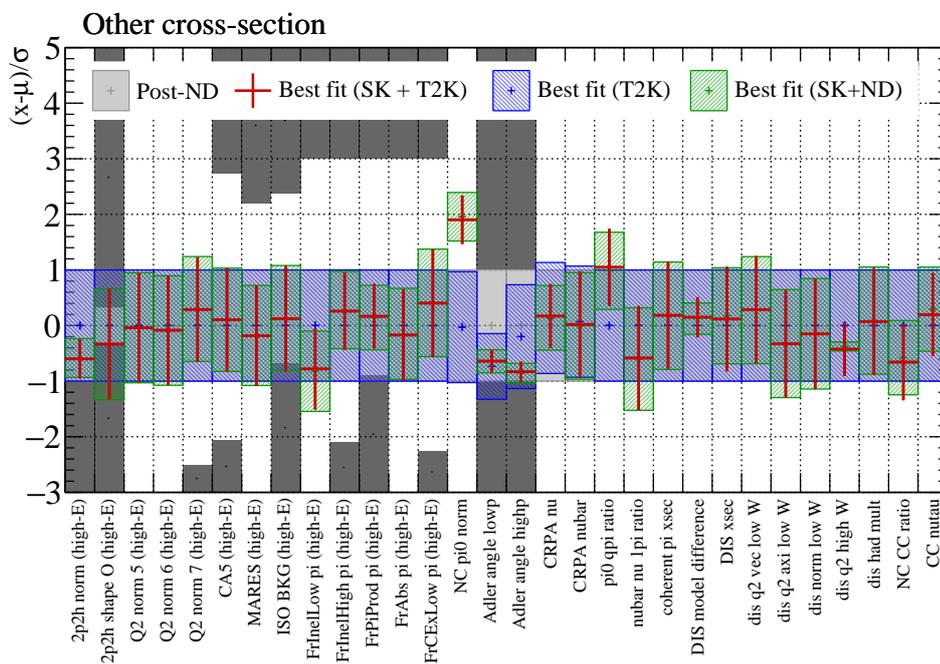
(d) Other cross-section systematics

Figure 8.2. Comparison of the **cross-section** systematic parameter best fits among SK+T2K, T2K, and SK (+ND) fits. Parameters are normalized with their prior uncertainties. The dark-shaded regions show the parameter boundaries.
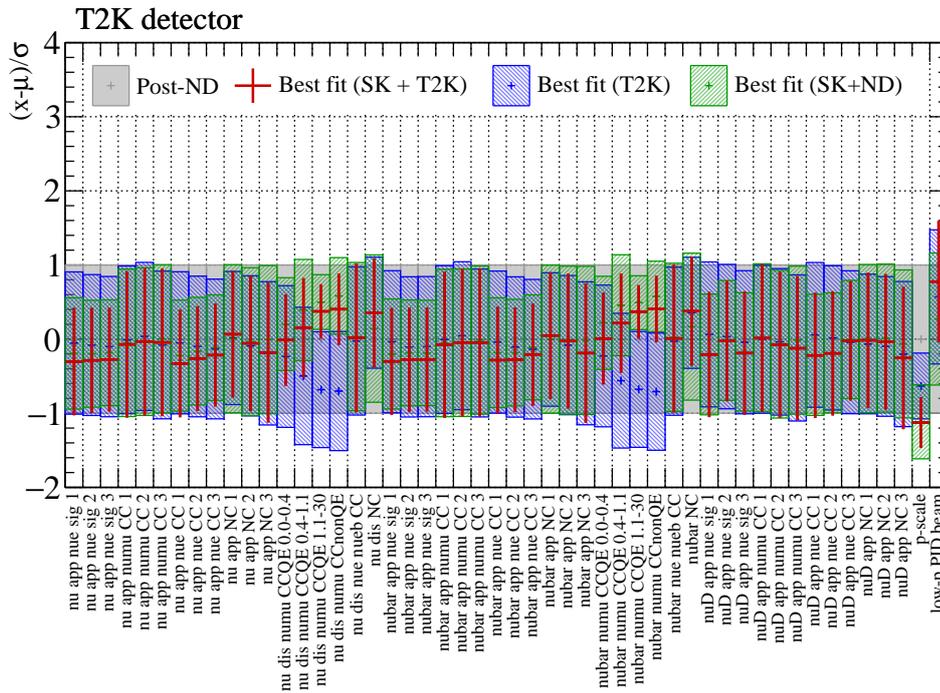
(e) T2K detector systematics
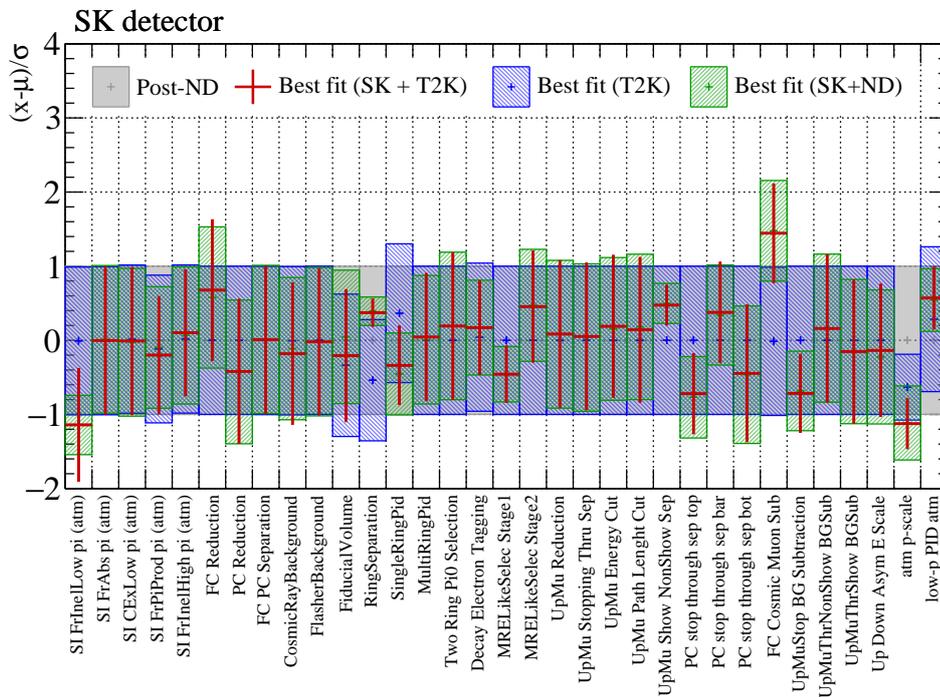
(f) SK atmospheric detector systematics

Figure 8.2. Comparison of the **detector** systematic parameter best fits among SK+T2K, T2K, and SK (+ND) fits. Parameters are normalized with their prior uncertainties.

We have two NCπ^0-related parameters: one gives 30% uncertainty on NC1π^0 resonance and the other gives 100% uncertainty on NC1π^0 coherent, which are tuned to the MiniBooNE data. Both of them are mostly constrained by the Sub-GeV π^0-like sample. The former (`NC pi0 norm` in Fig. 8.2d) is significantly pulled up in the data fit, which is nearly $+2\sigma$ ($+60\%$ stat increase). This might indicate a mis-modeling of NCπ^0 in our current nominal model, which is worth investigating for future analysis. This time we perform an additional fake data study to check whether the NC resonance normalization change could bias the $\delta_{\rm CP}$ sensitivity by changing the background estimation in the T2K e-like samples. We construct a fake data set by increasing the NC1π^0 resonance events by 60% (corresponds to $+2\sigma$ for `NC pi0 norm`) and fit the fake data set with our nominal model. The result showed a very small difference in $\delta_{\rm CP}$ as shown in Fig. 8.3. Therefore, we concluded the impact of the observed shift in `NC pi0 norm` on $\delta_{\rm CP}$ is small. This may be because the NCπ^0 background is well separated from the ν_e signals and the contribution of NCπ^0 is small enough in the e-like samples.
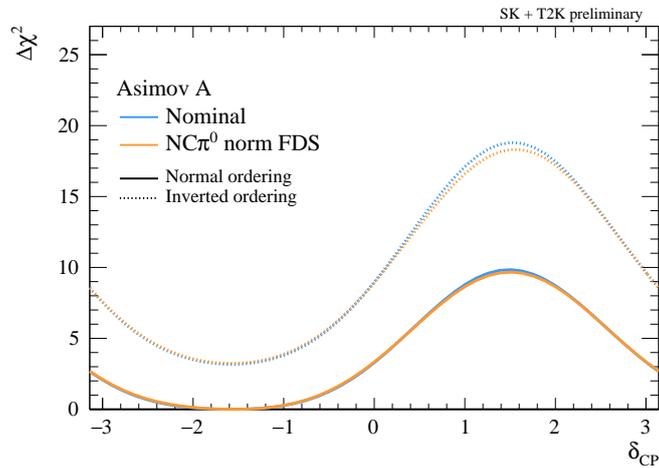
Figure 8.3. Comparison of the 1D $\Delta\chi^2$ for $\delta_{\rm CP}$ between the nominal Asimov and NCπ^0 normalization fake data fit. The oscillation parameters are set to the Asimov A parameter set and the systematic uncertainty parameters are set to the Pre-ND tuning.

Detector systematics The effects of the correlations in the detector systematic uncertainties can be seen in Fig. 8.2. Among the T2K detector systematic uncertainties, the parameters applied to the FHC/RHC $\nu_\mu \to \nu_\mu$ disappearance CCQE ([1.1, 30] GeV) and non-QE events are strongly correlated with the atmospheric ring separation systematics (Section 4.4). The ring separation is the uncertainty on the count of reconstructed rings (single-ring or multi-ring separation), and is constrained by the atmospheric multi-GeV samples as shown in Fig. 8.4. Since this parameter is strongly pulled in the atmospheric part of the fit, the correlated T2K detector systematics are also pulled up. Similarly, the FHC/RHC $\nu_\mu \to \nu_\mu$ disappearance CCQE ([0.4, 1.1] GeV) parameter is also anti-correlated with the SK atmospheric single-ring PID systematic uncertainty parameter and is pulled up in the SK+T2K fit.

The new low-momentum PID systematics (`low-p PID beam`, `low-p PID atm`) are pulled up by both the beam and atmospheric CC1π samples as expected. They are still within a post-fit 1.5σ uncertainty from the nominal.

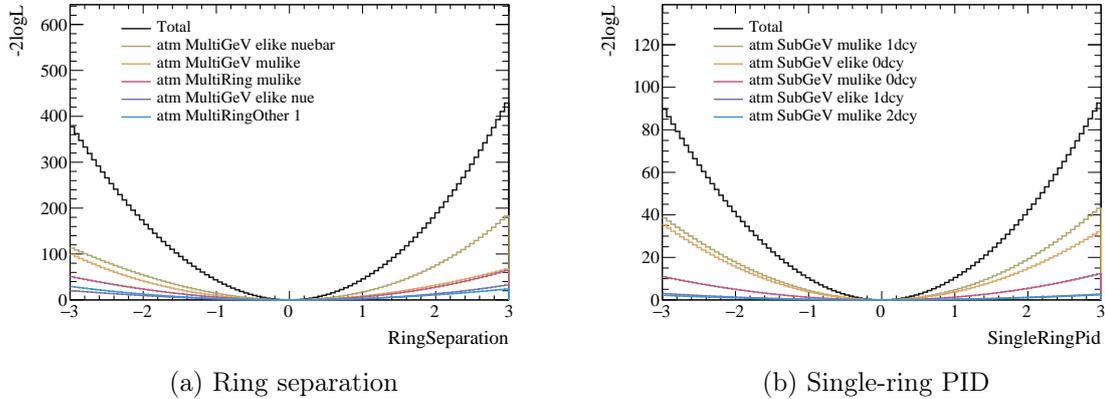(a) Ring separation

(b) Single-ring PID

Figure 8.4. Sample likelihood distributions when varying the ring separation and the single-ring PID systematic parameters within $[-3\sigma, 3\sigma]$ ranges. While one parameter is varied, the other oscillation parameters and the systematic parameters are fixed at the Asimov A parameter set and the Post-ND tuning. Only the total likelihood and likelihood of the five samples with the largest contributions are shown for each parameter.

8.2 Effect of the model changes

In this analysis, we added correlations between the beam and atmospheric detector systematics and employed new systematic uncertainties to cover the possible weakness in the CC1π model. It has been shown in Section 8.1.2 that these parameters got constrained in the data fit, and therefore, we further study the effect of these treatments. In this section, fits are performed using the importance sampling method.

8.2.1 Detector systematic correlations

First, we test the effect of the correlations between the beam and atmospheric detector systematic uncertainties. This can be done by simply performing a data fit with the correlated and uncorrelated detector systematic uncertainties. The results of these fits are shown in Fig. 8.5. There is no significant difference from this change, but the fit with the correlated detector systematics shows a slightly stronger rejection of the inverted mass ordering. The correlations of the detector systematic uncertainties do not affect the overall conclusion of this analysis.

8.2.2 Energy scale correlation

We have two energy-scale (p-scale) systematics to cover the uncertainties in the energy calibration. One is assigned to the T2K beam samples and the other is assigned to all the
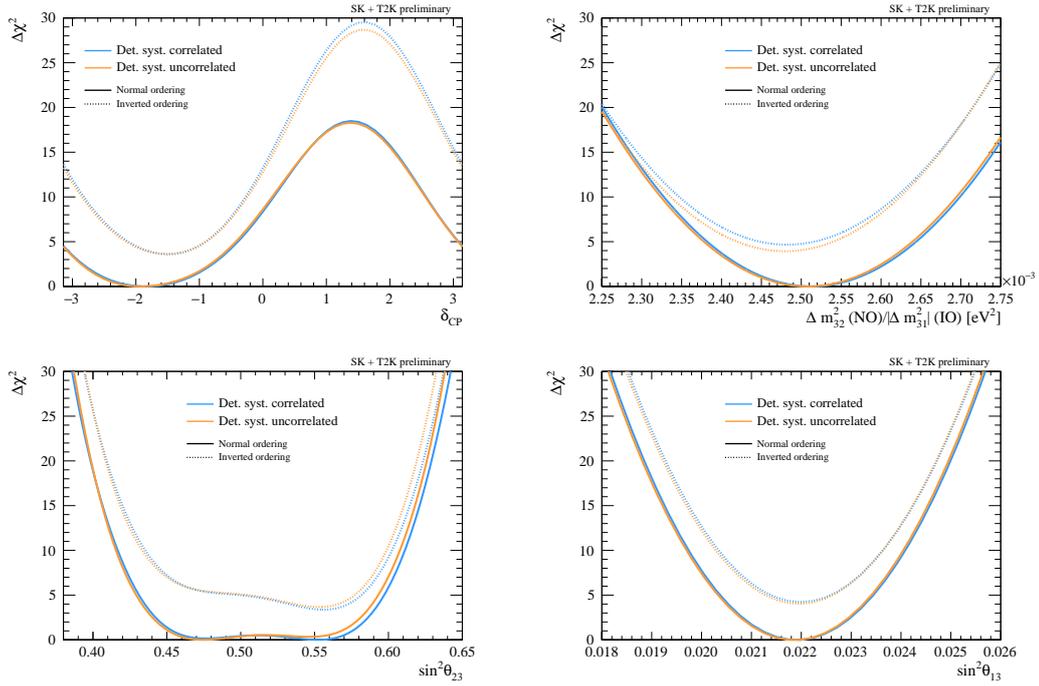
Figure 8.5. Comparison of data fit 1D $\Delta\chi^2$ contours with the correlated and uncorrelated detector systematic parameters between T2K beam and SK atmospheric samples. Both fits were performed using the importance sampling method with the reactor constraint. Δm_{32}^2 smearing is applied.

SK atmospheric samples. These systematic uncertainty parameters are implemented as scaling parameters that can directly modify the reconstructed lepton momentum independent of the event topology and energy. In this analysis, we apply full (100%) correlations between the beam and atmospheric energy-scale parameters assuming that the energy-scaling effect is common to all the samples as we use the same detector. Here we also perform the data fit with and without the energy-scale correlations to see if there could be a difference in the results.

Figure 8.6 shows the results of the data fit with the correlated and uncorrelated energy-scale parameters. When the two parameters are uncorrelated, it moves the best-fit point of Δm_{32}^2 to a smaller value. In the formulation of the oscillation probabilities, the squared mass difference is directly coupled to the neutrino energy through the $\sin^2 \frac{\Delta m^2 L}{4E}$ term. Therefore, it is natural that the change in the energy-scale parameter affects the measurement of Δm_{32}^2. As shown in Fig. 8.2, the atmospheric counterpart of the parameter is strongly pulled in the SK-only (+ND) data fit. When the beam and atmospheric energy-scale parameters are correlated, the beam energy scale can also be pulled by the atmospheric energy scale, which causes the shift in Δm_{32}^2.

Although we do not change our conclusion presented in Chapter 7, it will remain an interesting question for future analysis whether we should correlate the energy-scale systematics or not. Since we are using the same detector for the beam and atmospheric analyses, it seems to be natural to assume that the energy scale systematics are correlated. However, the energy scale uncertainties estimated using the control samples vary
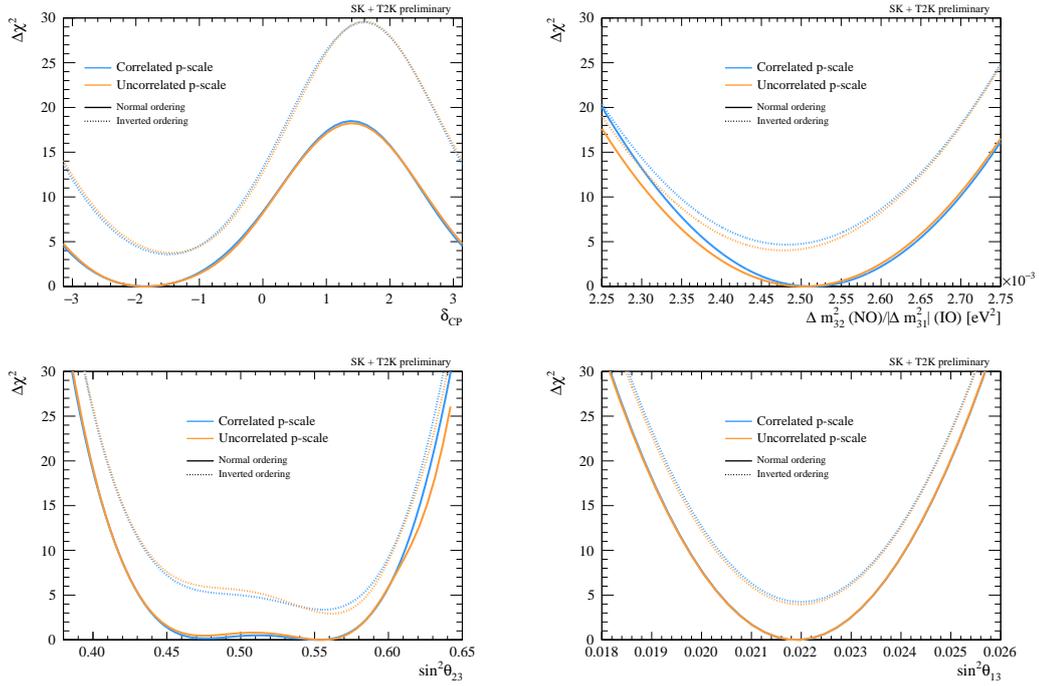
Figure 8.6. Comparison of data fit 1D $\Delta\chi^2$ contours with the correlated and uncorrelated energy-scale systematic parameters between T2K beam and SK atmospheric samples. Both fits were performed using the importance sampling method with the reactor constraint. Δm^2_{32} smearing is applied.

in different energy regions (Appendix A.1), and the atmospheric samples have wider energy ranges than T2K beam samples. Therefore, the full correlation of the atmospheric and beam energy-scale parameters could result in unexpected bias. A more sophisticated treatment of the energy-scale parameters (e.g. separating the parameters into different energy regions) will be studied in future analysis.

8.2.3 Atmospheric CC1π samples and related parameters

We also test the contribution of the atmospheric CC1π samples and the effects of the new systematics. In this analysis, we introduced the systematic uncertainties on the low-momentum PID errors and the pion momentum freedom due to the Adler angle distortion to cover the data/MC excess seen in the atmospheric CC1π samples. Here we perform additional data fit with two options to investigate the effects of these treatments:

1. Data fit with the full systematic uncertainty model but not including the atmospheric sub-GeV CC1π samples

2. Data fit using all the beam and atmospheric samples but without including the low-momentum PID and Adler angle systematic uncertainties in the fit.

Figure 8.7 shows the results of these fits along with the standard data fit. When we do not include the additional systematic uncertainties, the results are significantly affected by the

CC1π samples (denoted as "No low-p PID and Adler param." and drawn in green). On the other hand, the removal/addition of the atmospheric CC1π samples does not cause a difference when we include the additional systematics. These results imply that the extra parameters successfully reduce the biases due to the data/MC excess in the atmospheric CC1π samples. However, these parameters are ad-hoc parameters to deal with the excess and they may be revised in future analysis.
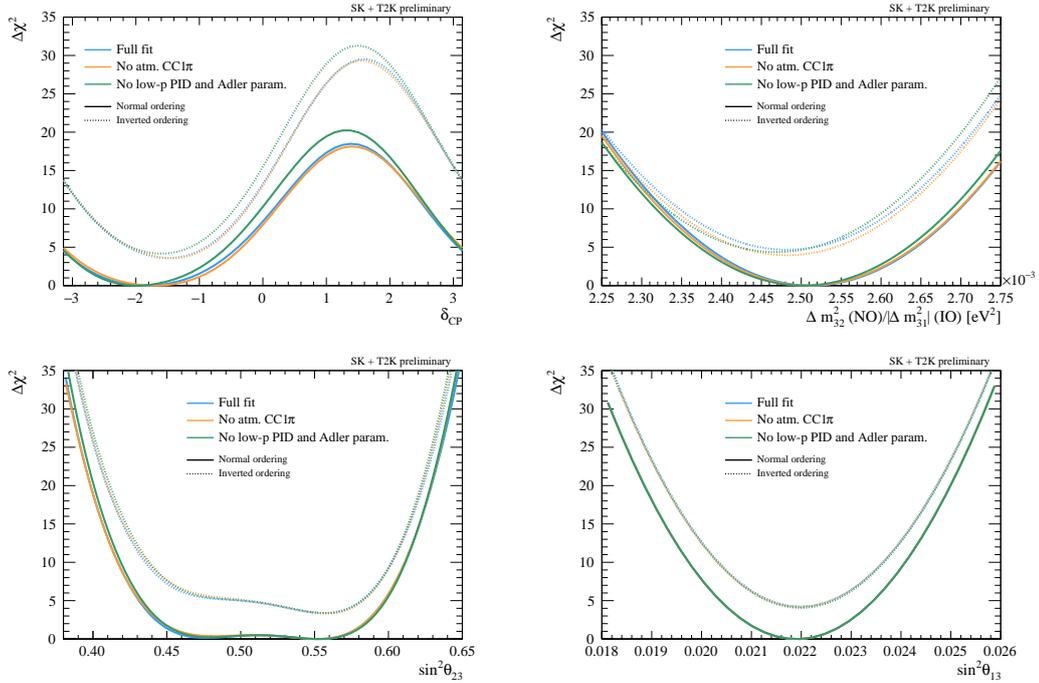

Figure 8.7. Comparison of data fit 1D $\Delta\chi^2$ distributions with several options for investigating the effect of atmospheric CC1π samples. All the fits were performed using the importance sampling method with the reactor constraint. Δm^2_{32} smearing is applied.

8.3   Comparison of the data fit and sensitivity

8.3.1   Data fit and sensitivity comparison

Additional studies are performed to understand how our data fit results behave compared to the sensitivity study. Figure 8.8 shows the comparison of the 1D $\Delta\chi^2$ distributions from the data fit and MC sensitivity studies using the importance sampling method. Two sensitivity studies are performed using the data best-fit values at the lower and upper octant summarized in Table 8.1 (called the "best-fit Asimov sensitivity").

The data fit shows tighter constraints on δ_{CP} compared to the data best-fit sensitivity results. For the other parameters, the data fit has mostly consistent constraints with the best-fit Asimov sensitivity at the upper octant. The comparison of the two best-fit Asimov sensitivity results implies that the choice of the octant has a non-negligible impact on δ_{CP} and mass ordering sensitivity. In particular, the rejection of the inverted mass ordering becomes significantly stronger when the upper octant is assumed.

Table 8.1. Summary of the oscillation parameter best-fit values obtained from the data fit at the upper and lower octant, respectively.

Parameter	Global best fit (lower octant)	Local best fit in upper octant
$\sin^2 \theta_{23}$	0.471	0.551
Δm^2_{32} [$\times 10^{-3}$ eV^2]	2.514	2.516
δ_{CP}	-1.76	-2.00
Mass ordering	Normal ordering	
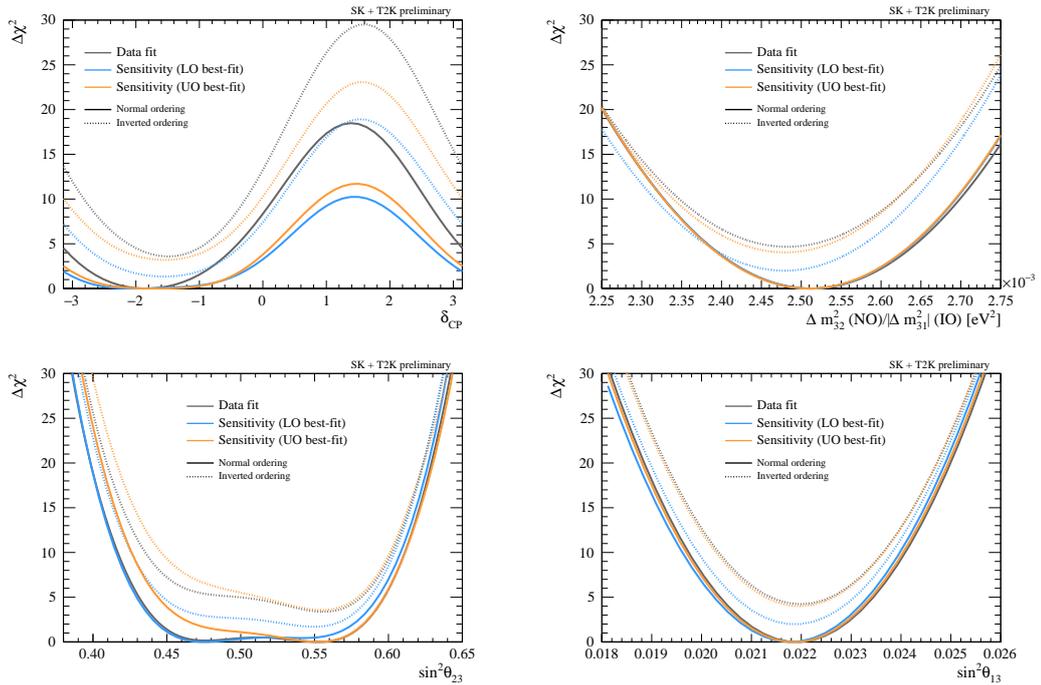
Figure 8.8. Comparison of the 1D $\Delta\chi^2$ curves of the data fit and sensitivity. The sensitivity is computed using the data best-fit parameter values at both the lower and upper octant. The importance sampling method is used for all the fits.

8.3.2 Sample contribution

Here we try to investigate which sample makes the δ_{CP} constraints tighter in the data fit. To visualize the contribution of each sample, we repeat a fit to the "hybrid" data set where one sample is taken from the best-fit Asimov data set, and the others are taken from the real data set. If the constraints of δ_{CP} change when we replace a certain sample with the MC, it means this sample contributed to the change in the data fit constraints. Fig. 8.9 shows the result of these hybrid studies for the T2K beam and SK atmospheric FC samples. As we found the octant of $\sin^2 \theta_{23}$ has a large impact on the δ_{CP} sensitivity, here we use the local best-fit values in the upper octant, which showed closer constraint to the data fit.

The largest contributions come from the SK atmospheric sub-GeV e-like 0de sample.

As we studied in Section 7.3.5, this sample had the smallest posterior probability p-value due to the data/MC excess in the most upward-going bin. As the upward-going part of the e-like samples have sensitivities to δ_{CP} through the $\nu_\mu \to \nu_e$ appearance oscillation, the better δ_{CP} constraints in the data fit may be caused by this excess. In addition, it shows a stronger IO rejection when the multi-ring μ-like sample is replaced with the best-fit Asimov, which means the multi-ring μ-like sample prefers the inverted ordering more in the data fit.

On the T2K side, the FHC 1Re sample has the largest contribution. This effect is further amplified when we use the best-fit values at the lower octant as shown in Fig. 8.10. When $\sin^2\theta_{23}$ is set to the lower octant, it reduces the MC prediction of the event rates around the oscillation maximal ($E_\nu \simeq 0.6$ GeV) as shown in Fig. 7.2, so the data/MC ratio becomes larger. This effect is therefore absorbed in δ_{CP} and makes the δ_{CP} constraints tighter in the data fit.

(a) Beam

(b) Atmospheric sub-GeV

(c) Atmospheric multi-GeV

(d) Atmospheric multi-GeV multi-ring
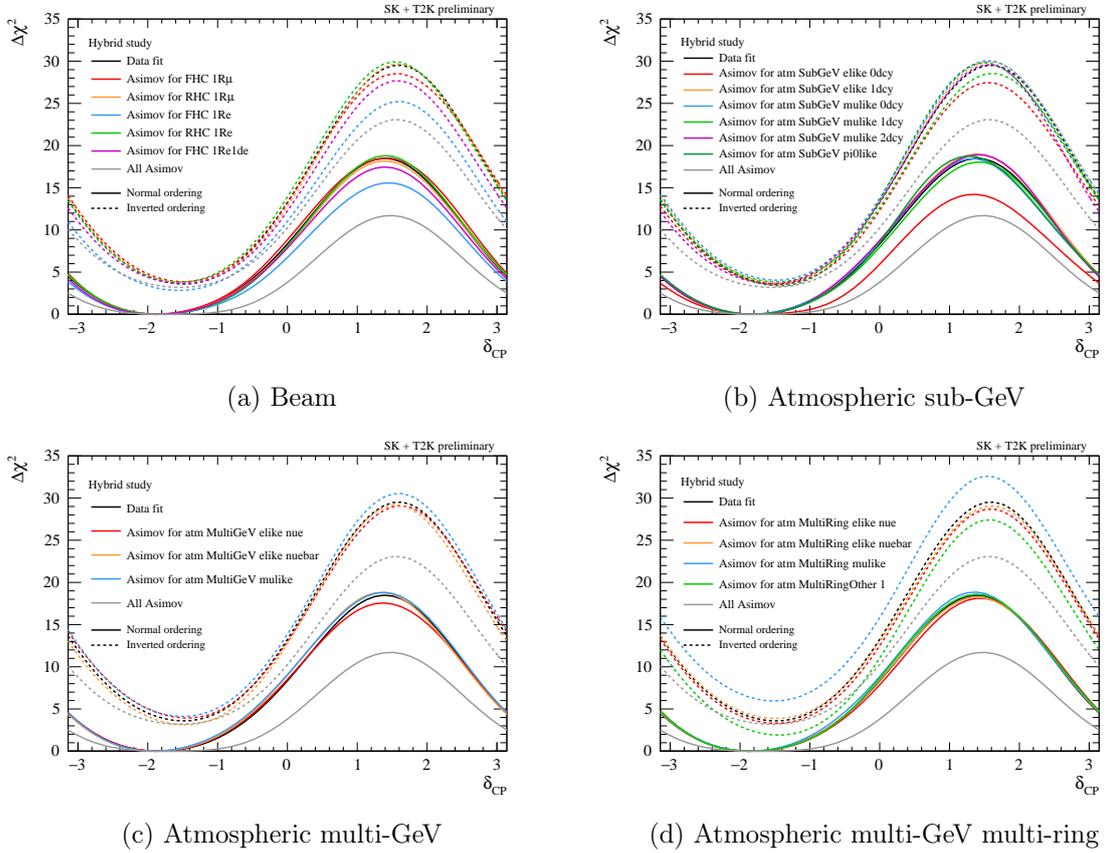
Figure 8.9. Comparison of the 1D $\Delta\chi^2$ contours where one sample is replaced with the data best-fit Asimov. The best-fit values at the upper octant are used.

(a) Beam

(b) Atmospheric sub-GeV
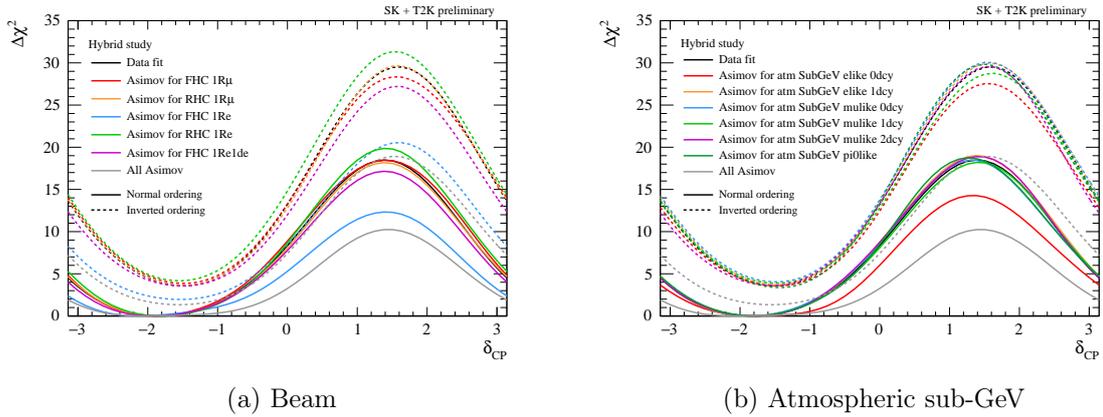
Figure 8.10. Comparison of the 1D $\Delta\chi^2$ contours where one sample is replaced with the data best-fit Asimov. The best-fit values at the lower octant are used.

8.4 Effect of the systematic uncertainties on the CP measurements

In this section, we investigate the effect of the systematic uncertainties on the δ_{CP} measurements.

8.4.1 Overall contribution of the systematic uncertainties

An overall contribution of the systematic uncertainties to the oscillation parameter constraints can be estimated by using the statistics-only uncertainty $\sigma_{\mathrm{stat.}}$ and computing

$$\sigma_{\mathrm{syst.}} = \sqrt{\sigma_{\mathrm{total}}^2 - \sigma_{\mathrm{stat.}}^2}. \tag{8.1}$$

The statistics-only uncertainty can be obtained by fixing the systematic uncertainty parameters at the best-fit values and running a fit only using the statistical term of the likelihood $\mathcal{L}_{\mathrm{stat.}}$ which is defined in Eq. (5.8).

Comparison of the $\Delta\chi^2$ distributions for the full-likelihood fit and statistics-only fit is shown in Fig. 8.11. It shows that the largest contribution to the uncertainty of these oscillation parameters comes from the statistical uncertainties, which implies increasing the statistics is very important for future analysis. However, the effects of systematic uncertainties are already non-negligible, especially in δ_{CP} and $\sin^2\theta_{23}$.

The computed statistic and systematic 2σ uncertainties on these parameters are summarized in Table 8.2 from the fixed-$\Delta\chi^2$ method. For δ_{CP}, the 2σ systematic uncertainty (1.478) is roughly $\sim 70\%$ of the statistic uncertainty (2.065), which implies that in future analysis when we get more statistics, the relative contribution of the systematic uncertainties will be more significant.

8.4.2 Correlations between δ_{CP} and systematic uncertainties

It has been shown that the systematic uncertainties already have a non-negligible impact on the δ_{CP} measurements. The next question we want to answer is which systematic
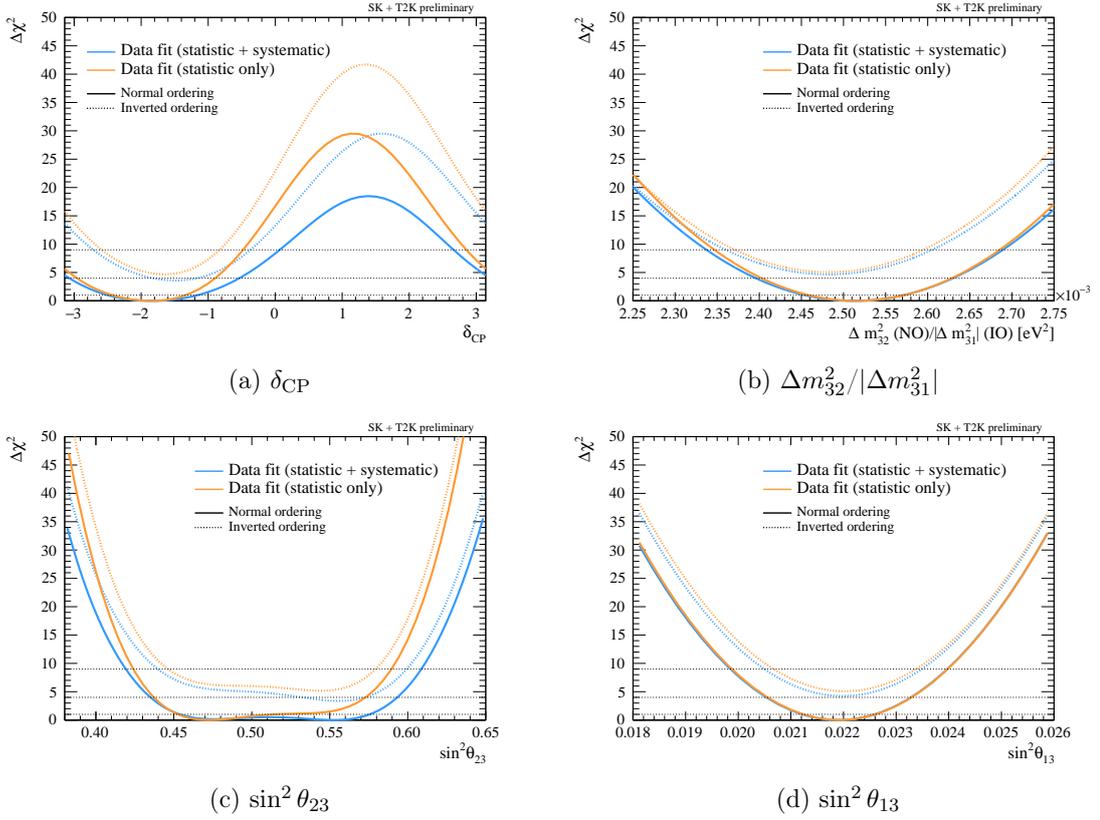
Figure 8.11. Comparison of data fit 1D $\Delta\chi^2$ distributions for the full likelihood fit and statistics-only fit. All the fits were performed using the importance sampling method with the reactor constraint. Δm_{32}^2 smearing is applied.

Table 8.2. Summary of the 2σ uncertainties on the oscillation parameters.

Normal ordering	$\sin^2\theta_{13}$ $[10^{-3}]$	δ_{CP}	Δm_{32}^2 $[10^{-3}$ eV$^2]$	$\sin^2\theta_{23}$
$2\sigma_{\mathrm{total}}$	2.752	2.540	0.234	0.158
$2\sigma_{\mathrm{stat.}}$	2.739	2.065	0.226	0.136
$2\sigma_{\mathrm{syst.}}$	0.266	1.478	0.062	0.081
Inverted	$\sin^2\theta_{13}$ $[10^{-3}]$	δ_{CP}	Δm_{32}^2 $[10^{-3}$ eV$^2]$	$\sin^2\theta_{23}$
$2\sigma_{\mathrm{total}}$	2.742	2.028	0.235	0.145
$2\sigma_{\mathrm{stat.}}$	2.732	1.694	0.225	0.135
$2\sigma_{\mathrm{syst.}}$	0.235	1.116	0.067	0.052

uncertainty has the biggest impact. To study this, we use the MCMC data fit results and see how the constraints on δ_{CP} change when we modify each systematic uncertainty parameter.

We first construct the posterior distribution for each systematic uncertainty parameter, approximate it with a Gaussian, and obtain its width σ. Then, we reweight the MCMC

steps so that the selected systematic uncertainty parameter has the posterior distribution of half-width $(\sigma/2)$. We also shift this narrowered posterior distribution to $\pm 1\sigma$. An example of the modified posterior distributions is shown in Fig. 8.12.
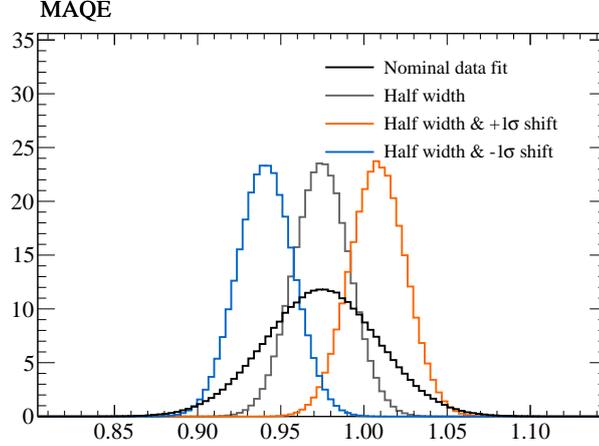

Figure 8.12. The original and modified posterior distributions for M_A^{QE} systematic uncertainty parameter.

The same reweighting is applied to the data-fit δ_{CP} posterior probability distribution to see how the constraint is affected. Figure 8.13 shows the ratio of the δ_{CP} 2σ credible interval size to the nominal data fit result when each parameter is modified as described above. Some systematics show a visible impact on δ_{CP} when they are shifted by $\pm 1\sigma$. Among the cross-section systematics, the ν_e/ν_μ and $\bar{\nu}_e/\bar{\nu}_\mu$ cross-sections, and CRPA ν systematic uncertainties make a visible shift in δ_{CP}. This is as expected because the constraints on δ_{CP} mainly come from the $\nu_\mu \to \nu_e$ and $\bar{\nu}_\mu \to \bar{\nu}_e$ appearance oscillation probabilities and the ν_e ($\bar{\nu}_e$) cross-section uncertainties are almost directly coupled to the δ_{CP} measurements through these channels.

The largest contribution among all the systematic uncertainty parameters comes from the single-ring PID in the atmospheric detector systematics. However, since these detector systematics are strongly correlated with each other, the effect of the shift cannot be simply explained by the contribution of a single parameter. For example, some T2K detector systematics are strongly correlated with this parameter, so they also change the δ_{CP} intervals through the correlation with the single-ring PID systematics.
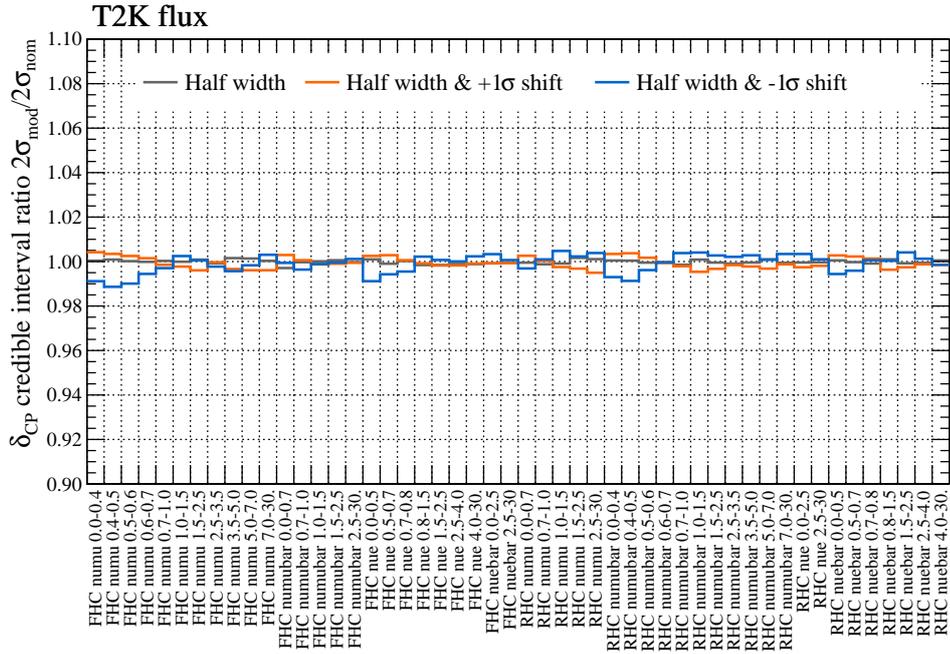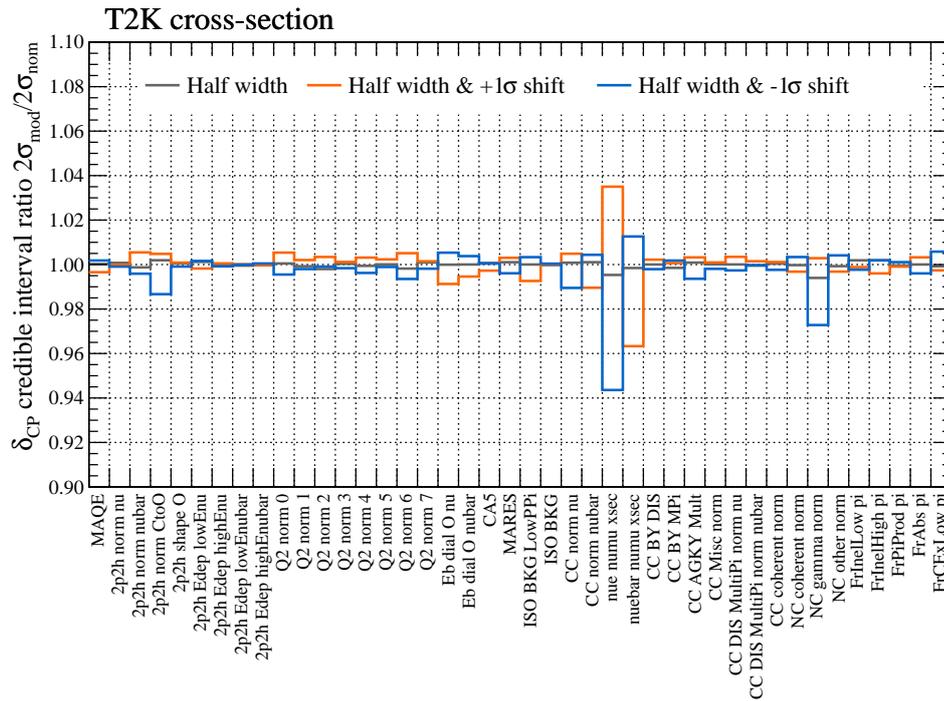
(a) T2K flux systematics
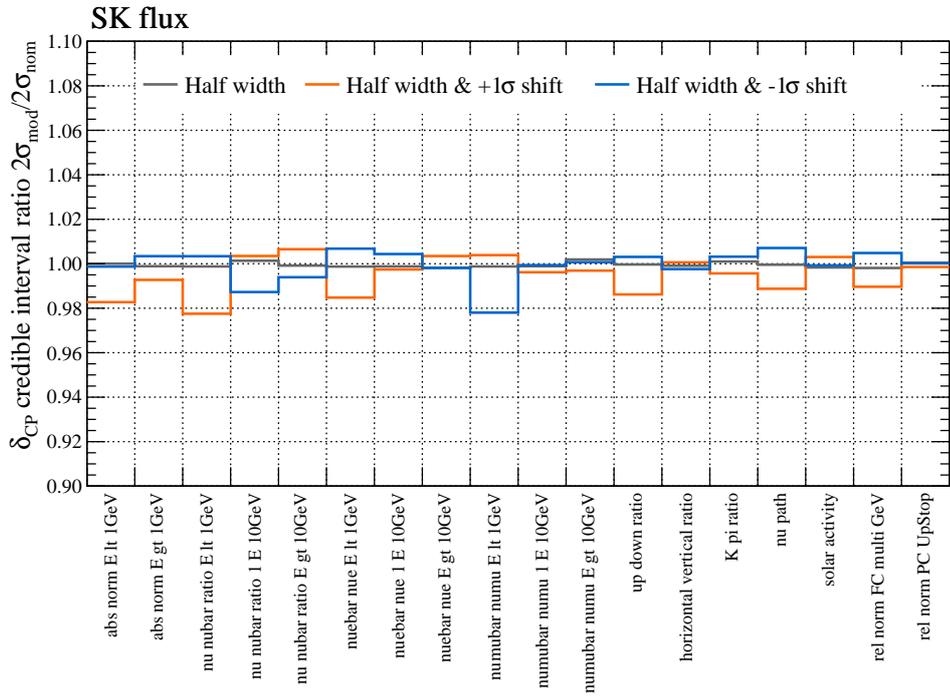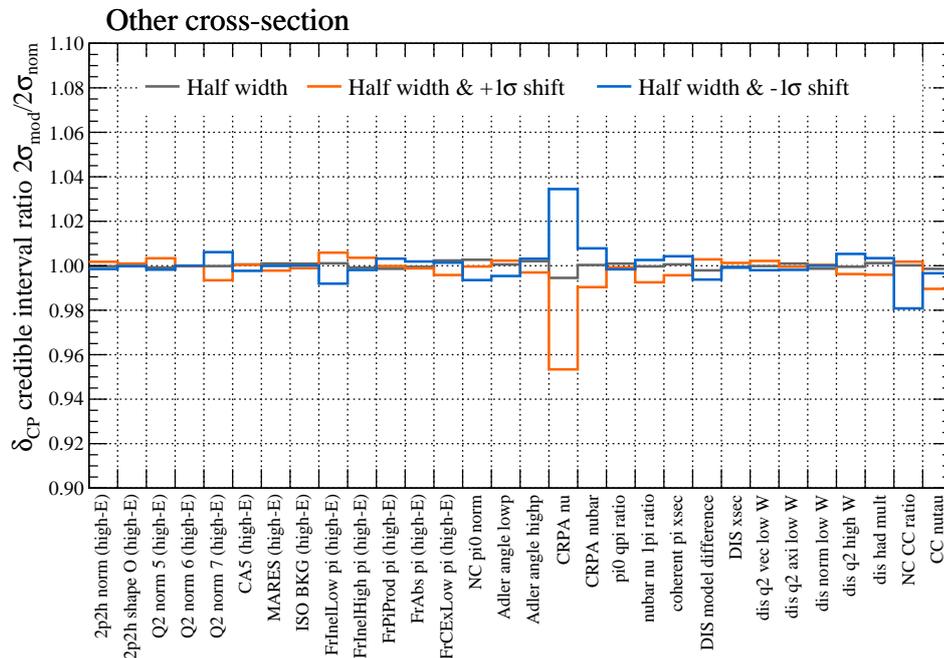

(b) SK atmospheric flux systematics

Figure 8.13. Comparison of the scaling and shifting effect of **flux** systematic parameters on the δ_{CP} 2σ credible intervals. The vertical axis shows the ratio of the 2σ credible interval size of the modified δ_{CP} posterior distribution ($2\sigma_{\mathrm{mod}}$) to that of the nominal data-fit posterior distribution ($2\sigma_{\mathrm{nom}}$).
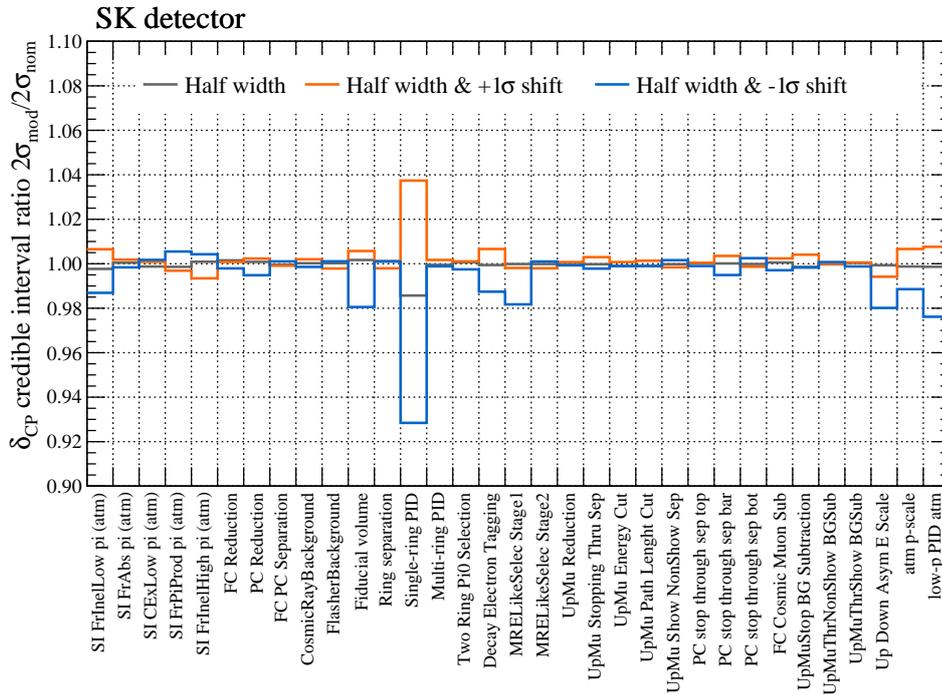
(c) T2K cross-section systematics

(d) Other cross-section systematics

Figure 8.13. Comparison of the scaling and shifting effect of **cross-section** systematic parameters on the δ_{CP} 2σ credible intervals. The vertical axis shows the ratio of the 2σ credible interval size of the modified δ_{CP} posterior distribution ($2\sigma_{\mathrm{mod}}$) to that of the nominal data-fit posterior distribution ($2\sigma_{\mathrm{nom}}$).
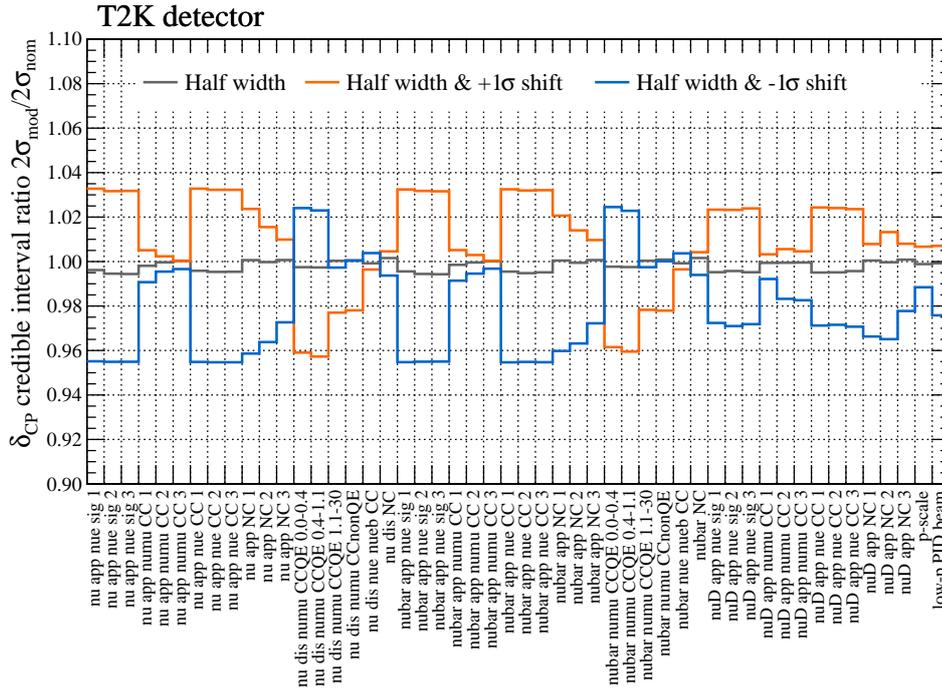
(e) T2K detector systematics

(f) SK atmospheric detector systematics

Figure 8.13. Comparison of the scaling and shifting effect of **detector** systematic parameters on the δ_{CP} 2σ credible intervals. The vertical axis shows the ratio of the 2σ credible interval size of the modified δ_{CP} posterior distribution ($2\sigma_{\mathrm{mod}}$) to that of the nominal data-fit posterior distribution ($2\sigma_{\mathrm{nom}}$).

The comparison of the δ_{CP} posterior distributions when the selected systematic uncertainty parameters are modified is shown in Fig. 8.14. The cross-section systematic uncertainties mostly affect the width of the δ_{CP} posterior probability distribution, whereas the single-ring PID also shifts the δ_{CP} distribution.
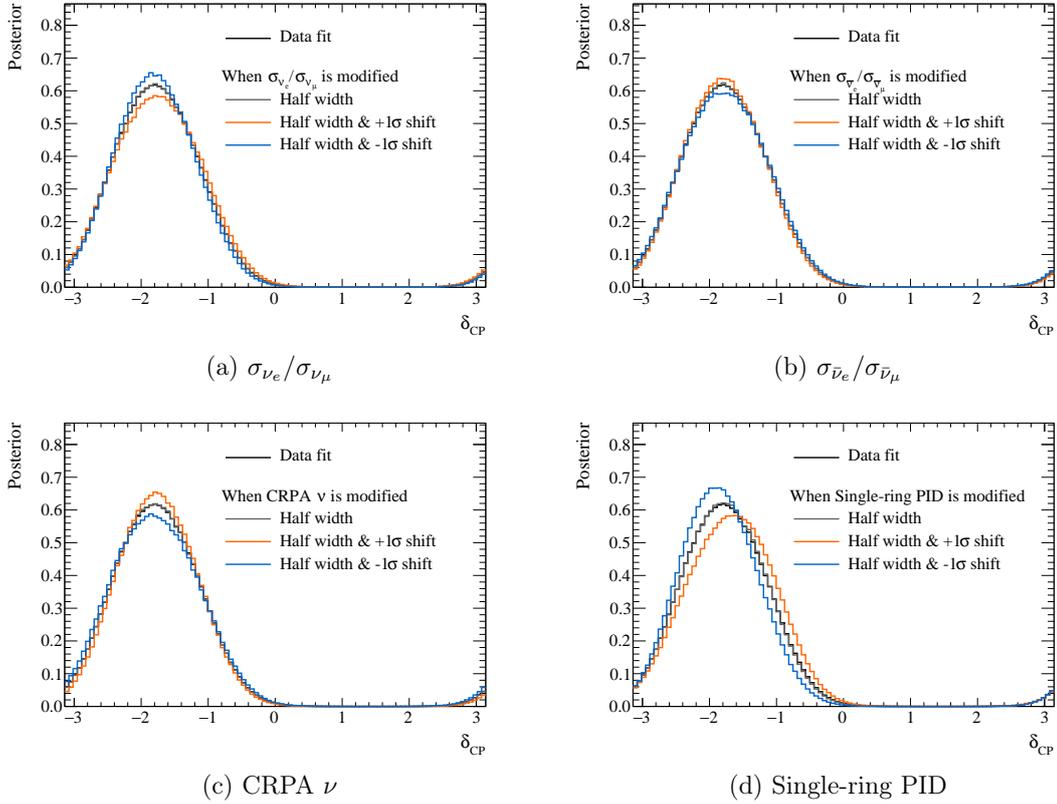

(a) $\sigma_{\nu_e}/\sigma_{\nu_\mu}$

(b) $\sigma_{\bar{\nu}_e}/\sigma_{\bar{\nu}_\mu}$

(c) CRPA ν

(d) Single-ring PID

Figure 8.14. Comparison of the δ_{CP} posterior distributions when scaling and shifting the selected systematic uncertainty parameters.

8.5 Effect of different priors

We performed a Bayesian analysis in Chapter 7 using both the flat prior in δ_{CP} and the flat prior in $\sin\delta_{\mathrm{CP}}$. However, the result of the Bayesian analysis also depends on the choice of the priors for the other parameters. In this section, some additional studies are performed to test the effect of alternative prior choices.

Mixing angle   In our main analysis, we use the flat prior in $\sin^2\theta_{23}$. However, the prior can be also chosen to be flat in θ_{23}, $\sin\theta_{23}$ or $\sin^2 2\theta_{23}$. The comparison of these different prior distributions for the mixing angle parameter is shown in Fig. 8.15. In addition, the constraints on θ_{23} mainly come from the $\nu_\mu \to \nu_\mu$ survival probability, which can also be approximated using the effective mixing ($\sin^2\theta_{\mu\mu}$) and squared mass difference ($\Delta m^2_{\mu\mu}$)

as [17]

$$P_{\nu_\mu \to \nu_\mu} \approx 1 - \sin^2 2\theta_{\mu\mu} \sin^2 \frac{\Delta m_{\mu\mu}^2 L}{4E_\nu}, \tag{8.2}$$

where $\sin^2 \theta_{\mu\mu}$ and $\Delta m_{\mu\mu}^2$ are defined as

$$\sin^2 \theta_{\mu\mu} = \cos^2 \theta_{13} \sin^2 \theta_{23}, \tag{8.3}$$

$$\Delta m_{\mu\mu}^2 = \sin^2 \theta_{12} \Delta m_{31}^2 + \cos^2 \theta_{12} \Delta m_{32}^2$$
$$+ \cos \delta_{\text{CP}} \sin \theta_{13} \sin 2\theta_{12} \tan \theta_{23} \Delta m_{21}^2. \tag{8.4}$$

It is shown that the survival probability has a dependence on $\sin^2 2\theta_{\mu\mu}$ at the leading order. Therefore, here we perform the additional comparison for the flat prior in $\sin^2 2\theta_{\mu\mu}$ as well.
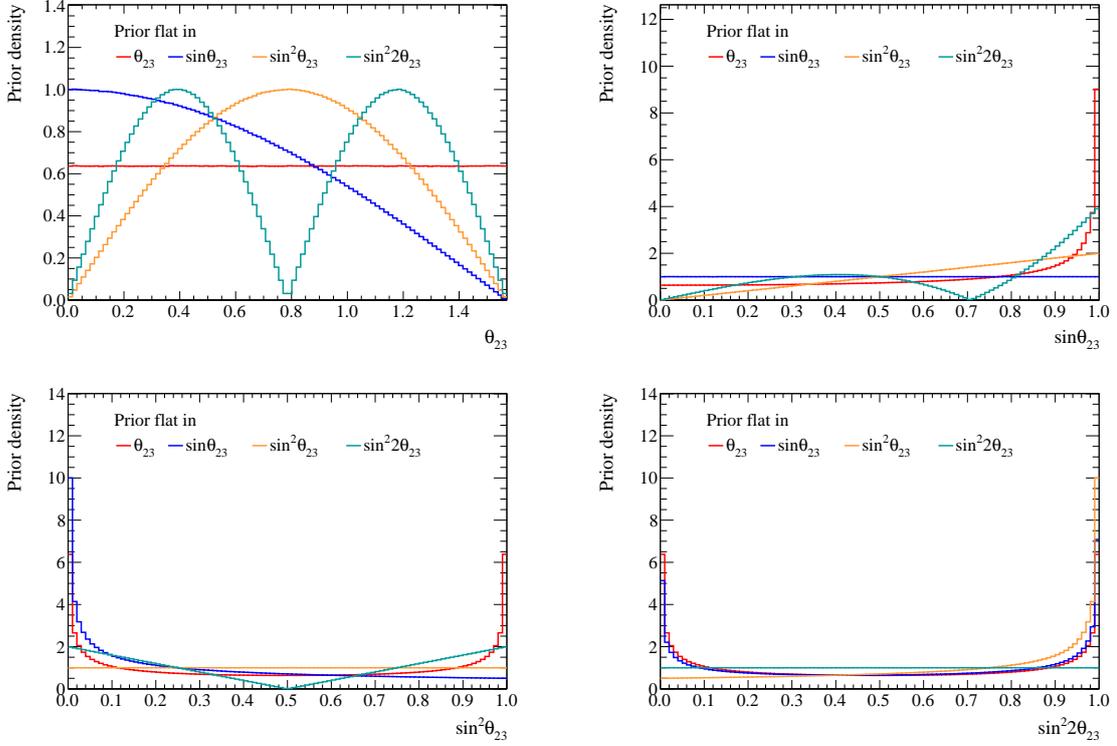
Figure 8.15. Comparison of the alternative prior distributions in the mixing angle parameter.

Squared mass difference   In the main analysis, we apply the prior constraint on $\sin^2 2\theta_{13} = 0.0853 \pm 0.0027$ from the reactor experiments. We also test the prior constraints on $(\Delta m_{32}^2, \sin^2 2\theta_{13})$ taken from the latest publication by the Daya Bay reactor neutrino experiment [171]. The reactor neutrino experiments measure the $\bar{\nu}_e \to \bar{\nu}_e$ survival probability and it can be approximated as

$$P_{\bar{\nu}_e \to \bar{\nu}_e} \approx 1 - \sin^2 2\theta_{13} \sin^2 \frac{\Delta m_{ee}^2 L}{4E_\nu}, \tag{8.5}$$

where Δm_{ee}^2 is defined as

$$\Delta m_{ee}^2 = \cos^2\theta_{12}\Delta m_{31}^2 + \sin^2\theta_{12}\Delta m_{32}^2. \tag{8.6}$$

Therefore, Daya Bay can provide good two-dimensional constraints on $(\Delta m_{32}^2, \sin^2 2\theta_{13})$ as shown in Fig. 8.16. Since it is pointed out by Nunokawa et al. that the combination of the measurements of electron and muon neutrino (antineutrino) disappearance channels has a sensitivity to determine the mass ordering [172], it is of interest to use the two-dimensional constraints from Daya Bay.

(a) Normal ordering

(b) Inverted ordering
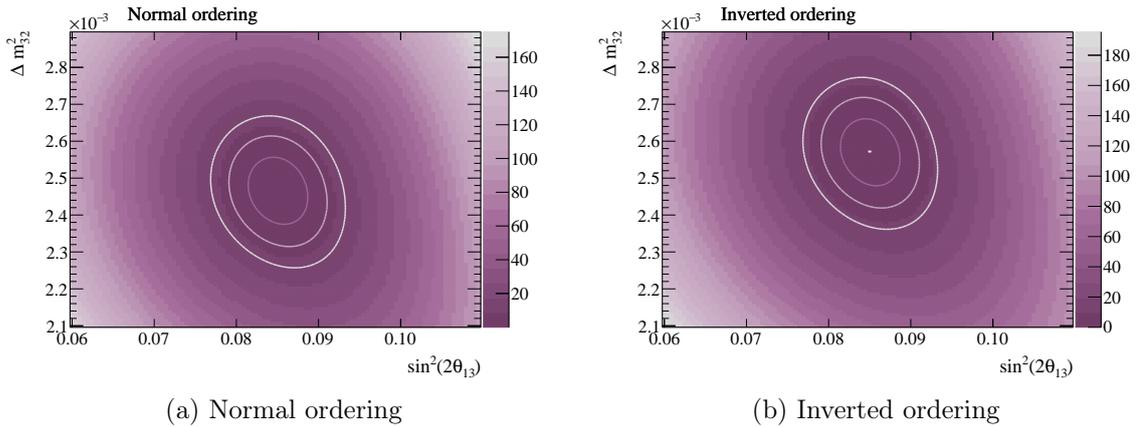
Figure 8.16. Two dimensional $\Delta\chi^2$ distributions in the Δm_{32}^2 and $\sin^2 2\theta_{13}$ space reported by the Daya Bay experiment [171]. The 1σ, 2σ, and 3σ confidence intervals based on the fixed-$\Delta\chi^2$ method are shown together.

We perform these alternative prior studies by applying additional weights to the MCMC steps used for the main analysis in Chapter 7. Denoting the parameter conversion from θ to ϕ as $\phi = f(\theta)$, the prior conversion can be written as

$$\pi'(\phi) = \left|\frac{d\phi}{d\theta}\right|\pi(\theta) = \left|\frac{df(\theta)}{d\theta}\right|\pi(\theta). \tag{8.7}$$

Therefore, each MCMC step is reweighted with $\left|\frac{df(\theta)}{d\theta}\right|$ to obtain the posterior distributions under the different prior choices.

8.5.1 Posterior distribution comparison

Fig. 8.17 shows the posterior distributions of the oscillation parameters under different prior assumptions. For the CP-related parameters (δ_{CP} and J_{CP}), the effects from the flat $\sin^2 2\theta_{\mu\mu}$ and Daya Bay priors are found to be small. Although the rejection of CP conservation is slightly weaker under the flat prior in $\sin^2 2\theta_{23}$, the choice of the priors flat in δ_{CP} and flat in $\sin\delta_{\mathrm{CP}}$ has a much larger impact.

For the squared mass differences, the Daya Bay prior has a visible impact in Δm_{32}^2 as expected, which shifts the distribution to the smaller value of Δm_{32}^2 and makes the distribution narrower. Other choices of the alternative priors ($\sin\delta_{\mathrm{CP}}$, $\sin^2 2\theta_{23}$, and $\sin^2 2\theta_{\mu\mu}$) do not have much impact on this parameter.

Some interesting features are seen in the mixing angle parameters. If we compare the flat δ_{CP} prior and the flat $\sin \delta_{\mathrm{CP}}$ prior, the latter makes the upper octant preference stronger. On the other hand, the flat $\sin^2 2\theta_{\mu\mu}$ prior makes the upper octant preference weaker compared to the flat $\sin^2 2\theta_{23}$ prior. The Bayes factors for the upper octant hypothesis over the lower octant hypothesis are computed to be $B_{\mathrm{UO/LO}}^{\sin^2 2\theta_{23}} = 2.4$ and $B_{\mathrm{UO/LO}}^{\sin^2 2\theta_{\mu\mu}} = 1.3$ under the flat $\sin^2 2\theta_{23}$ prior and the flat $\sin^2 2\theta_{\mu\mu}$ prior, respectively. In this analysis, we do not have a strong statement on the octant preference, but we will need to care about these prior effects in future analyses when we have better sensitivities to the octant.

The comparisons of the 2D posterior distributions under different prior choices are shown in Fig. 8.18. They also show similar tendencies as described above.

8.5.2   Mass ordering Bayes factor

The Bayes factor does not depend on the prior choice for the two hypotheses that are being compared because it is defined as the ratio of marginal likelihood. However, it may depend on the priors for the nuisance parameters. In the main analysis described in Chapter 7, we computed the mass ordering Bayes factor under the nominal prior choice that is flat in δ_{CP} and $\sin^2 \theta_{23}$. It was computed to be ~ 9.0 and showed a moderate preference for the normal ordering. The mass ordering Bayes factors computed at different prior choices are summarized in Table 8.3. It has a variation around 7-11, but in any case, the corresponding significance is lower than 2σ and it does not change our conclusion on the mass ordering.

Table 8.3. Mass ordering Bayes factor under the different priors.

Prior	MO Bayes factor	Significance
δ_{CP}	8.97	1.64σ
$\sin \delta_{\mathrm{CP}}$	11.10	1.74σ
$\sin^2 2\theta_{23}$	7.78	1.58σ
$\sin^2 2\theta_{\mu\mu}$	8.65	1.63σ
DayaBay	7.81	1.58σ

(a) δ_{CP}

(b) J_{CP}

(c) Δm^2_{32} (NO) / Δm^2_{31} (IO)

(d) $|\Delta m^2_{\mu\mu}|$

(e) $\sin^2\theta_{23}$

(f) $\sin^2\theta_{\mu\mu}$
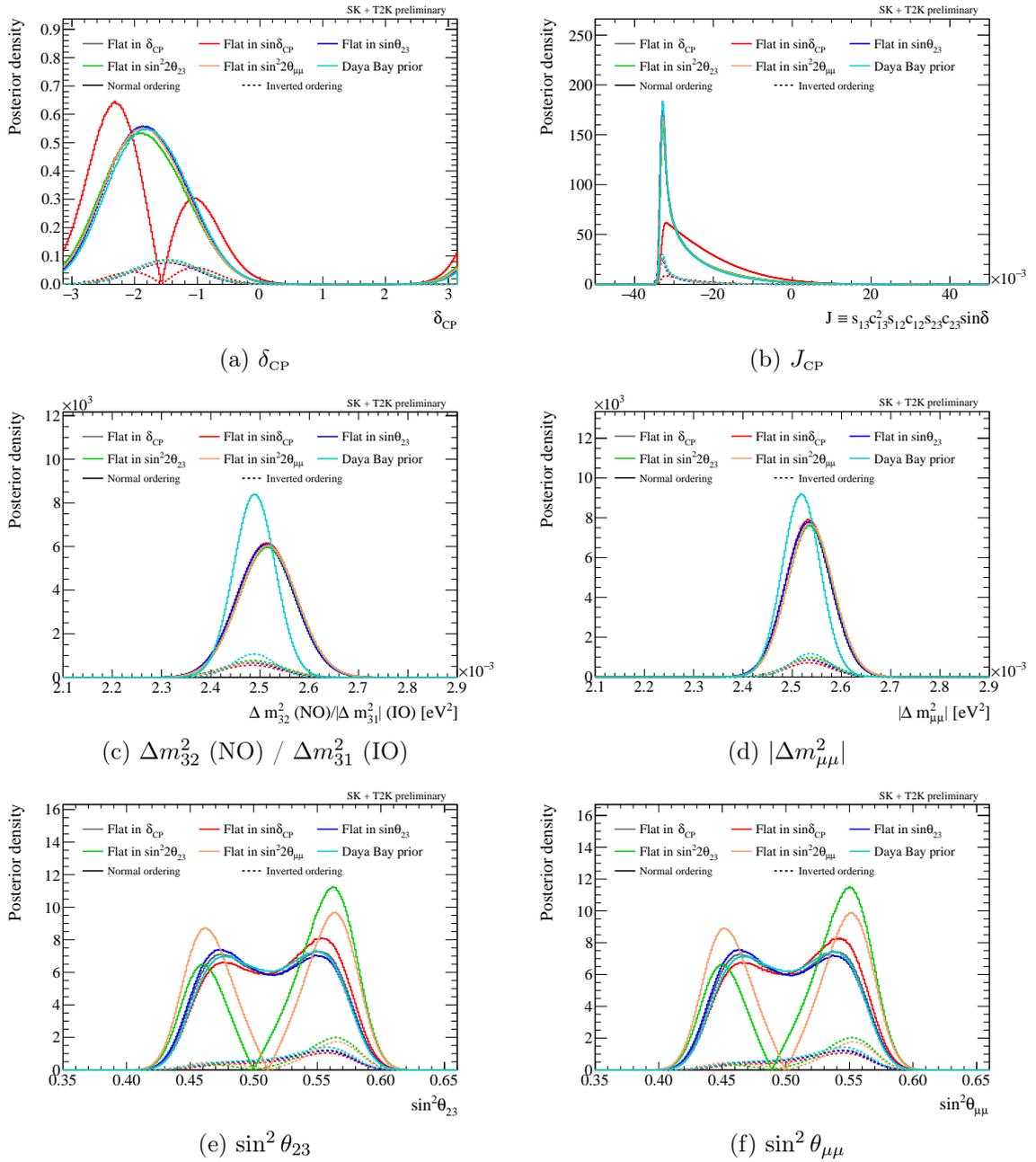
Figure 8.17. Comparison of the posterior distributions under the different priors.

(a) δ_{CP} vs Δm^2_{32} (NO)

(b) δ_{CP} vs Δm^2_{32} (IO)

(c) δ_{CP} vs $\sin^2 \theta_{23}$ (NO)

(d) δ_{CP} vs $\sin^2 \theta_{23}$ (IO)

(e) Δm^2_{32} vs $\sin^2 \theta_{23}$ (NO)

(f) Δm^2_{32} vs $\sin^2 \theta_{23}$ (IO)
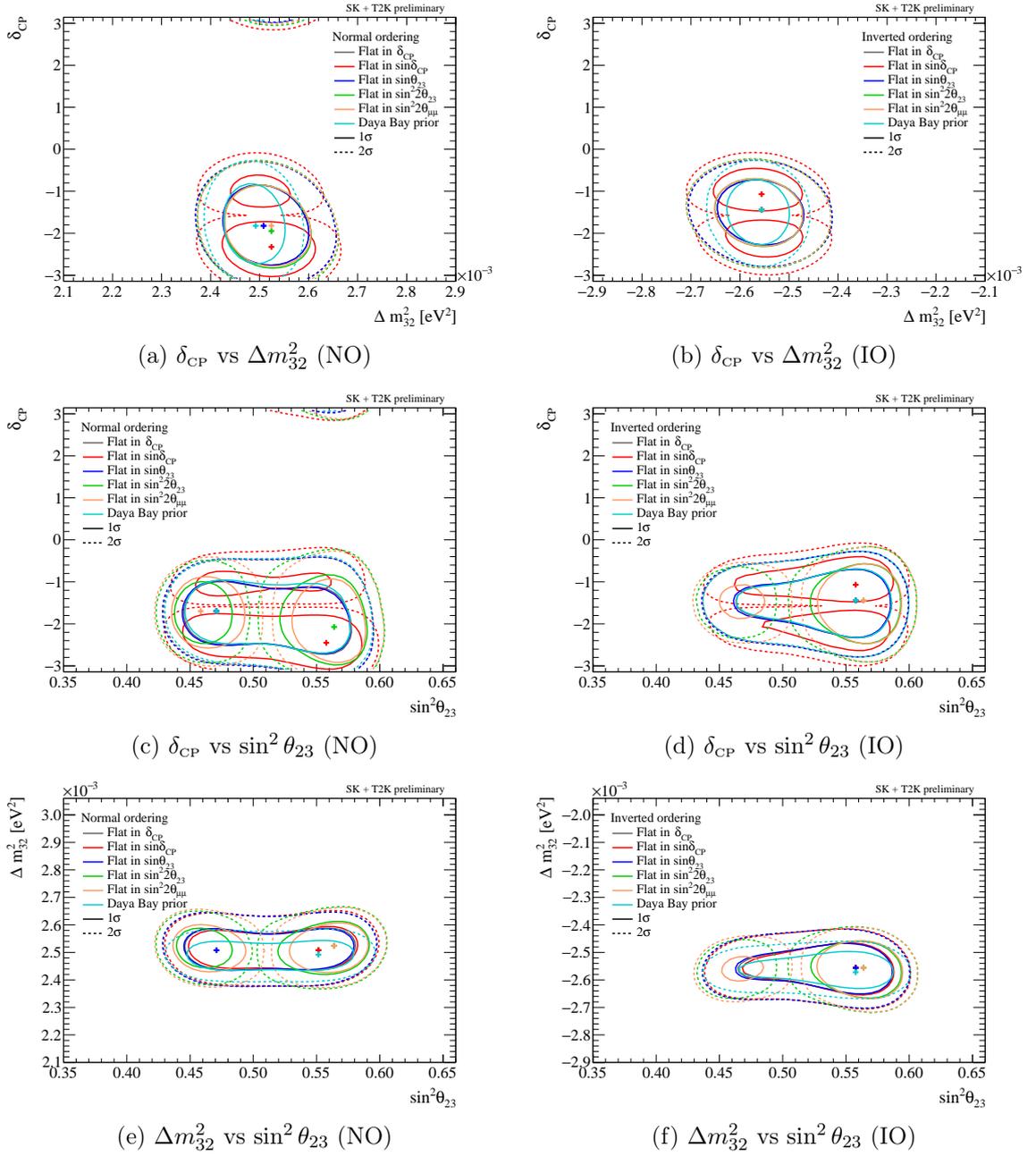
Figure 8.18. Comparison of the 2D posterior distribution under the different priors.

8.6 CP conservation and violation test

8.6.1 Savage-Dickey Bayes factor for CP violation

In general, the output of Bayesian analysis is given as a posterior distribution. Therefore, it is not straightforward to test our preference for the CP violation hypothesis ($\sin\delta_{\text{CP}} \neq 0$) over the CP conserving hypothesis which is a point-like hypothesis ($\sin\delta_{\text{CP}} = 0$). One approach is to construct credible intervals for the CP parameters (δ_{CP} or J_{CP}) and to test whether the CP-conserving values are excluded from the intervals or not, as we did in Chapter 7. An alternative approach for directly testing the point-like hypothesis within the Bayesian analysis framework is the Savage-Dickey method [173, 174].

As described in Section 7.3.4, the Bayes factor is defined as the ratio of the marginal likelihood between two hypotheses. Let us assume that the point-like hypothesis we want to test (H_0) corresponds to a particular value of parameter $\theta = \theta_0$ in the complex model for the alternative hypothesis (H_1). For example, in our CP violation test, $\sin\delta_{\text{CP}} = 0$ stands for the CP-conserving hypothesis, and any other values of $\sin\delta_{\text{CP}}$ ($\sin\delta_{\text{CP}} \neq 0$) stand for the CP-violating hypothesis. One can write the marginal likelihoods for the hypotheses H_0 and H_1 as

$$p_0(n) = \int \mathcal{L}_0(\eta|n)\pi_0(\eta)d\eta, \tag{8.8}$$

$$p_1(n) = \iint \mathcal{L}_1(\theta,\eta|n)\pi_1(\theta,\eta)d\theta d\eta, \tag{8.9}$$

where n is the data, η is the nuisance parameter(s), and $\pi(\cdot)$ is the prior. The marginal likelihood for the hypothesis H_0 can then be rewritten as

$$p_0(n) = \int \mathcal{L}_1(\eta|n, \theta = \theta_0)\pi_1(\eta|\theta = \theta_0)d\eta \tag{8.10}$$

$$= \mathcal{L}_1(\theta = \theta_0|n). \tag{8.11}$$

By applying the Bayes' theorem (Eq. (5.2)), we can write

$$p_0(n) = \frac{p_1(\theta = \theta_0|n)p_1(n)}{\pi_1(\theta = \theta_0)}. \tag{8.12}$$

Therefore, the Savage-Dickey Bayes factor, the ratio of the marginal likelihood between the hypotheses H_0 and H_1, can be computed as

$$\text{BF}_{\text{Savage-Dickey}} = \frac{p_0(n)}{p_1(n)} = \frac{p_1(\theta = \theta_0|n)}{\pi_1(\theta = \theta_0)}. \tag{8.13}$$

It means that the ratio between the posterior and prior at $\theta = \theta_0$ gives the Bayes factor for this point-like hypothesis H_0. Since we want to test the CP-violating hypothesis in our analysis, we take the inverse of this statistic to compute the Bayes factor for CP violation ($= p_1(n)/p_0(n)$).

We compute the Savage-Dickey Bayes factor by taking the ratio of the posterior and prior of the Jarlskog invariant at $J_{\text{CP}} = 0$. Fig. 8.19 shows the comparison of the Jarlskog

invariant prior and posterior distributions under normal ordering, inverted ordering, and both ordering, respectively. Since this factor directly depends on the choice of the prior, we plotted both the flat δ_{CP} prior and flat $\sin\delta_{\mathrm{CP}}$ prior. In addition, we find that the parameter range for the $\sin^2\theta_{23}$ prior has a non-negligible impact on the prior value at $J_{\mathrm{CP}} = 0$. In the fit, we allow $\sin^2\theta_{23}$ to vary freely in the range of $[0, 1]$, but here we show several options of $\sin^2\theta_{23}$ prior ranges for completeness.

(a) Flat prior in δ_{CP} (NO)

(b) Flat prior in $\sin\delta_{\mathrm{CP}}$ (NO)

(c) Flat prior in δ_{CP} (IO)

(d) Flat prior in $\sin\delta_{\mathrm{CP}}$ (IO)

(e) Flat prior in δ_{CP} (both)

(f) Flat prior in $\sin\delta_{\mathrm{CP}}$ (both)
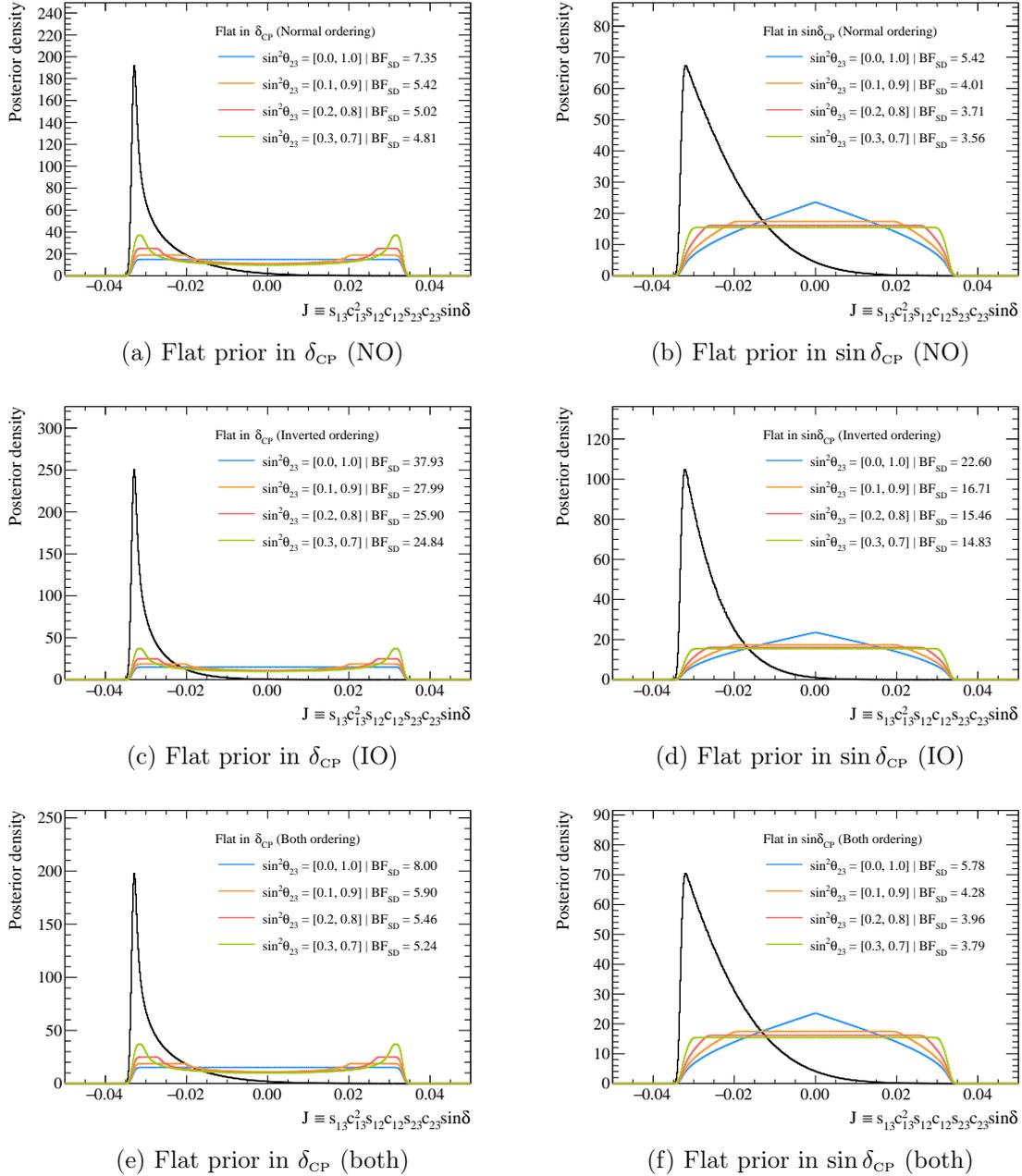
Figure 8.19. Jarlskog invariant with the flat prior in δ_{CP} (left) and the flat prior in $\sin\delta_{\mathrm{CP}}$ (right). Savage-Dickey Bayes factor is computed for CP violation ($J_{\mathrm{CP}} \neq 0$) over CP conservation ($J_{\mathrm{CP}} = 0$) under different prior ranges for $\sin^2\theta_{23}$.

The values of the Savage-Dickey Bayes factors are summarized in Table 8.4. It gives a stronger preference for CP violation when we assume the wider parameter ranges for the $\sin^2 \theta_{23}$ prior. In the inverted ordering, the Bayes factor is larger than 20 for both the flat $\delta_{\rm CP}$ and flat $\sin \delta_{\rm CP}$ priors when we assume $\sin^2 \theta_{23} \in [0,1]$, which corresponds to the significance level of $> 2\sigma$. For normal ordering, the Bayes factor is around 5-7, which suggests a weak preference for CP violation. Overall, the CP-violating hypothesis is preferred to the CP-conserving hypothesis, but it is not enough to claim the evidence of CP violation.

Table 8.4. Summary of the Savage-Dickey Bayes factors under the different prior assumptions.

Prior	Mass ordering	$\sin^2 \theta_{23}$			
		$[0.0, 1.0]$	$[0.1, 0.9]$	$[0.2, 0.8]$	$[0.3, 0.7]$
	Normal ordering	$7.35\ (1.56\sigma)$	$5.42\ (1.42\sigma)$	$5.02\ (1.38\sigma)$	$4.81\ (1.37\sigma)$
Flat in $\delta_{\rm CP}$	Inverted ordering	$37.93\ (2.23\sigma)$	$27.99\ (2.11\sigma)$	$25.90\ (2.08\sigma)$	$24.84\ (2.07\sigma)$
	Both ordering	$8.00\ (1.59\sigma)$	$5.90\ (1.46\sigma)$	$5.46\ (1.42\sigma)$	$5.24\ (1.40\sigma)$
	Normal ordering	$5.42\ (1.42\sigma)$	$4.01\ (1.28\sigma)$	$3.71\ (1.25\sigma)$	$3.56\ (1.23\sigma)$
Flat in $\sin \delta_{\rm CP}$	Inverted ordering	$22.60\ (2.03\sigma)$	$16.71\ (1.91\sigma)$	$15.46\ (1.88\sigma)$	$14.83\ (1.86\sigma)$
	Both ordering	$5.78\ (1.45\sigma)$	$4.28\ (1.31\sigma)$	$3.96\ (1.28\sigma)$	$3.79\ (1.26\sigma)$

8.6.2 Frequentist p-value for CP conservation

Although our Bayesian analysis result depends on the choice of prior, it is possible to perform a prior-independent CP conservation test using a frequentist approach. In this test, we compute a p-value of obtaining our data fit result, which prefers CP violation (CPV), under the CP conservation (CPC) hypothesis. If the p-value takes a small value (e.g. $p \leq 0.05$ is a common criterion), it implies that our data preference for CPC is not likely a statistical fluctuation but a real feature of nature.

In terms of $\delta_{\rm CP}$, CPC corresponds to the values of $\delta_{\rm CP} = 0, \pi$, while CPV corresponds to any values other than them. Therefore, χ^2 for each hypothesis can be written as

$$\chi^2[\text{CPC}] \equiv \min_{\alpha} \chi^2(\alpha), \tag{8.14}$$

$$\chi^2[\text{CPV}] \equiv \min_{\delta_{\rm CP}, \text{MO}} \chi^2(\delta_{\rm CP}, \text{MO}), \tag{8.15}$$

where α takes any of the combinations of $\delta_{\rm CP} = 0, \pi$ for normal or inverted ordering.

Here we define the p-value as the probability of obtaining a difference in the minimum χ^2 values between CPC and CPV hypotheses $\Delta\chi^2 \equiv \chi^2[\text{CPC}] - \chi^2[\text{CPV}]$ larger than that from the data fit ($\Delta\chi^2_{\text{data}}$), under a certain hypothesis. To obtain this p-value, we generate

(N) pseudo data sets and compute

$$p = \frac{1}{N} \sum_{i=1}^{N} I[\Delta\chi_i^2 > \Delta\chi_{\text{data}}^2], \qquad (8.16)$$

where $I[\cdot]$ is an indicator function that returns 1 when the argument is true and returns 0 otherwise.

For the CPC test, the pseudo data sets are generated by taking the oscillation parameters from the result of the MCMC data fit without the multicanonical method ($\beta = 1.0$), which gives proper posterior correlations among the oscillation parameters. There is one caveat that in MCMC we do not have the steps exactly at $\delta_{\text{CP}} = 0, \pi$. Therefore, we take the steps around $\sin\delta_{\text{CP}} \approx 0$ and force them to be $\delta_{\text{CP}} = 0, \pi$ when generating the true CP-conserving pseudo data sets. The densities of different combinations of $\delta_{\text{CP}} = 0, \pi$ and the mass ordering are summarized in Table 8.5. The systematic uncertainty parameters are varied according to the Gaussian from the near detector fit. The statistical variations are applied to these data sets assuming the Poisson fluctuation. A fit to the ensemble of pseudo data sets is performed using the profiling method.

Table 8.5. Densities of the generated true CP-conserving toys under different combinations of $\delta_{\text{CP}} = 0, \pi$ and the mass ordering. These densities are extracted from the data fit MCMC.

	Normal ordering	Inverted ordering
$\delta_{\text{CP}} = 0$	0.143	0.013
$\delta_{\text{CP}} = \pi$	0.837	0.006

Figure 8.20 shows the distribution of the obtained $\Delta\chi^2$ distributions from the fit to 2000 pseudo data sets. The corresponding p-value is computed to be

$$p = 0.027 \pm 0.004,$$

which means that the CP conservation hypothesis is excluded at more than 2σ confidence level. However, since this result is based on only 2000 pseudo data sets, the validity of the result is not fully guaranteed. A relatively small number of pseudo data sets used in this analysis is due to the computational limit. Unfortunately, since it takes more than 10 days to perform a fit to 2000 data sets even with 1000 CPUs, it is difficult to increase the number of pseudo data sets within the current framework. It is also pointed out that, with p-values, the significance of the result may often be overestimated due to the lack of reference to the alternative hypothesis [175]. Therefore, due to these limitations of the p-value study, we do not treat it as our conclusion in this analysis.
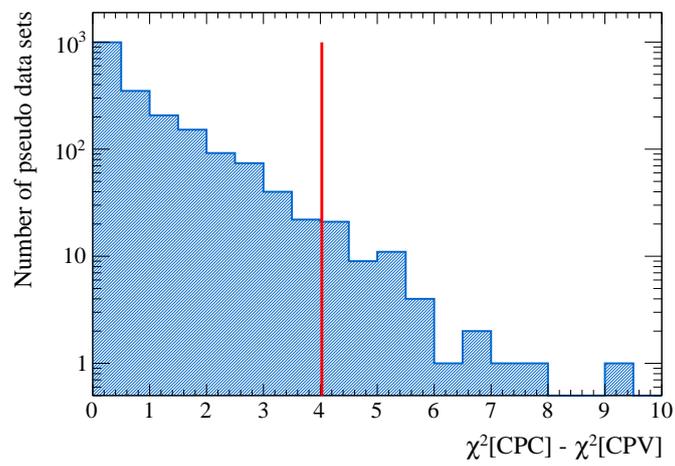
Figure 8.20. Distribution of the χ^2 difference between the CP conservation and CP violation hypothesis fit. The red vertical line shows the value we obtained from the data fit.

Chapter 9

Future sensitivity and prospects

This chapter presents the future sensitivity study assuming the data set expected to be collected in the coming years. The prospects for the future analysis are also stated.

9.1 Future sensitivity with increased statistics

In this section, we discuss the future sensitivity of the joint beam and atmospheric neutrino analysis assuming the data set that is expected to be collected by 2027, before starting the data taking of the Hyper-Kamiokande (HK) detector, which is a successor of SK [176]. In addition, we study the impact of the inclusion of SK I-III periods which was not used in this analysis.

9.1.1 Differences between the SK I-III phase and later phases

In the main analysis, we have used the SK atmospheric data taken during the SK IV phase and T2K Run 1-10 data which corresponds to the SK IV and later phases. It is valuable to add the SK atmospheric data taken during SK I-III as well to increase the statistics. However, as summarized in Table 9.1, SK was operated under very different conditions during the SK I-III periods compared to the SK IV and later. To summarize, the main differences are the following:

1. Photo coverage is roughly half (19%) during SK II compared to the other periods (40%) due to a PMT implosion accident that happened in 2001

2. An old electronics Analog Timing Module (ATM) [177] was used in SK I-III while QBEE (Section 3.2.1.4) was used in SK IV and later

3. An old reconstruction algorithm APFit [178] was used in SK I-III while fiTQun (Section 3.2.1.5) was used for SK IV and later.

In addition, in 2020, Gd was resolved into the water to obtain better neutron tagging efficiencies. The SK VI phase was operated with 0.01% mass concentration of the Gd and the current SK VII is operated with 0.03% concentration.

Due to these differences, it could be difficult to perform the joint analysis using the data taken during these periods. In the following future sensitivity study, therefore, we

study two possible scenarios with and without using the SK I-III data for the atmospheric samples.

Table 9.1. Summary of the SK data-taking phases and properties.

Phase	Dates	Livetime [days]	Photo-coverage	Electronics	Reconstruction	Composition
SK I	1996-2001	1489.19	40%	ATM	APFit	H_2O
SK II	2002-2005	798.59	19%	ATM	APFit	H_2O
SK III	2006-2008	518.08	40%	ATM	APFit	H_2O
SK IV	**2008-2018**	**3244.4**	**40%**	**QBEE**	**APFit/fiTQun**	H_2O
SK V	2019-2020	461.02	40%	QBEE	APFit/fiTQun	H_2O
SK VI	2020-2022	583.3	40%	QBEE	APFit/fiTQun	H_2O + Gd (0.01%)
SK VII	2022-	-	40%	QBEE	APFit/fiTQun	H_2O + Gd (0.03%)

9.1.2 Future sensitivity

The future sensitivity study is performed simply by scaling the MC data set to the estimated amount of data. Therefore, we do not consider any additional modifications in our analysis. In reality, we dissolve gadolinium in water from the SK VI phase and it will enhance our ability to tag the neutron events. T2K has also developed an additional multi-ring μ-like sample to increase the statistics, but these improvements are not included in the following study.

The assumed data sets for the SK atmospheric and T2K accelerator neutrinos until 2026 are summarized in Tables 9.2 and 9.3. For the SK atmospheric samples, we assume the 365 days of livetime per year with 90% efficiency $365 \times 0.9 = 328.5$ for the coming years. For T2K, we assume that we will obtain an equal amount of FHC and RHC data set by the end of 2026, which end up in 9.7×10^{21} POT in total.

Table 9.2. Assumed SK atmospheric neutrino data set for the future sensitivity study. JFY denotes the Japanese fiscal year.

SK	I-III	IV	V-VI	Future		Total
Year	1996-2008	2008-2018	2019-2022	JFY2023-2024	JFY2025-2026	
Livetime (day)	2,805.86	3244.4	1044.32	657.0	657.0	8408.58

The results of the sensitivity study are shown in Fig. 9.1 including the SK I-III data and in Fig. 9.2 without including the SK I-III data. The results show the gradually improved constraints on the oscillation parameters year by year. Based on the fixed-$\Delta\chi^2$ method, it shows a sensitivity to exclude CP conserving values of δ_{CP} ($\delta_{CP} = 0, \pi$) at a significance between 2-3σ. The sensitivities for the mass ordering Bayes factor are summarized in Table 9.4. It shows that we expect to achieve a 2σ level of significance in the normal ordering preference within a few years in both with and without SK I-III scenarios.

Table 9.3. Assumed T2K accelerator neutrino data set for the future sensitivity study. JFY denotes the Japanese fiscal year.

T2K	Run 1-10	Run 11	Future		Total
Year	2009-2020	2021	JFY2023-2024	JF2025-Y2026	
FHC ($/10^{20}$ POT)	19.664	1.764	9.959	17.00	48.387
RHC ($/10^{20}$ POT)	16.346	0	15.041	17.00	48.387
Total ($/10^{20}$ POT)	36.01	1.764	25.0	34.00	96.774
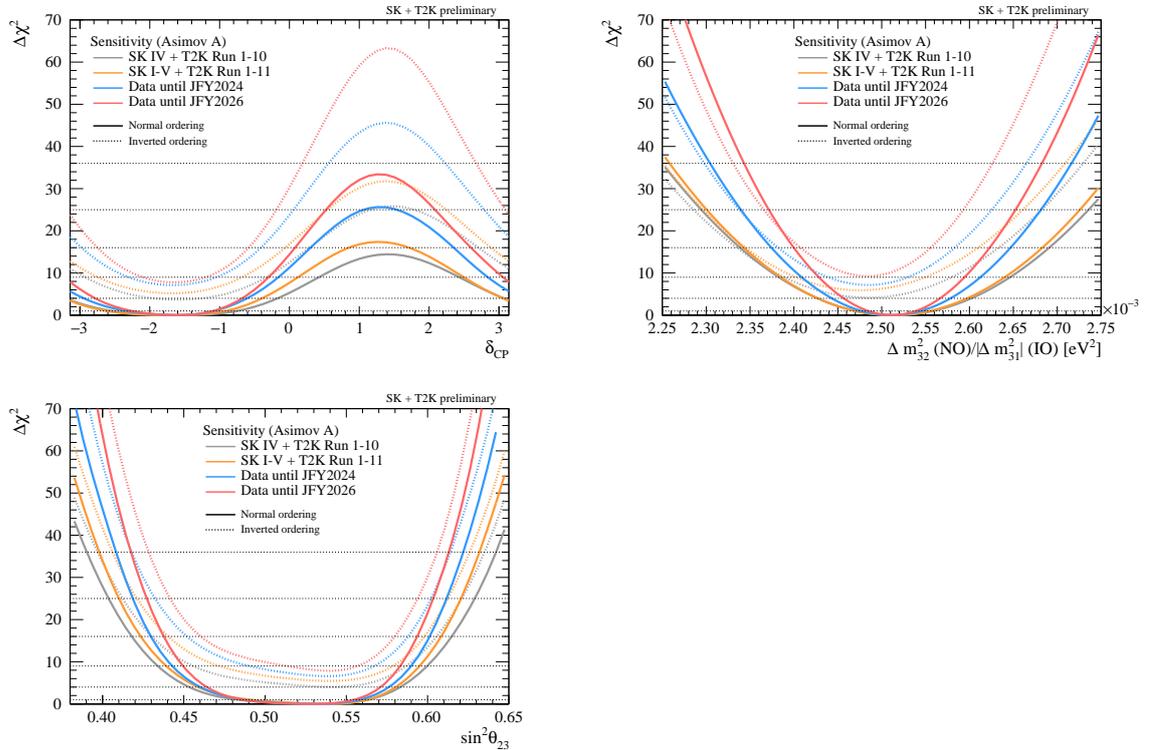
Figure 9.1. Future sensitivity with estimated data until 2026.
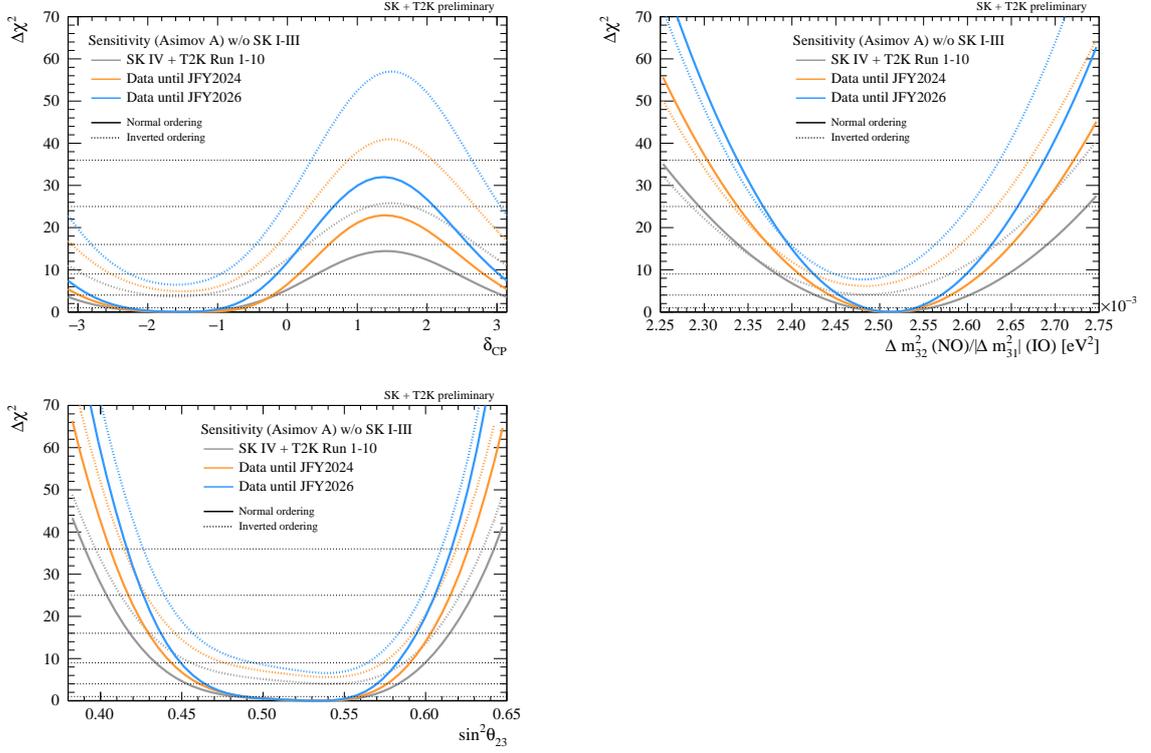
Figure 9.2. Future sensitivity with estimated data until 2026. The SK I-III phases are excluded from the estimation.

Table 9.4. Expected Bayes factor for normal ordering over inverted ordering.

Data set	MO Bayes factor	Significance
SK IV + T2K Run 1-10	8.85 ± 0.35	1.64σ
SK I-V + T2K Run1-11	19.48 ± 1.94	1.97σ
JFY2024	47.63 ± 8.13	2.32σ
JFY2026	68.47 ± 13.10	2.45σ
Data set	MO Bayes factor	Significance
SK IV + T2K Run 1-10	8.85 ± 0.35	1.64σ
SK IV-V + T2K Run1-11	10.11 ± 0.31	1.70σ
JFY2024 (w/o SK I-III)	19.97 ± 6.01	1.98σ
JFY2026 (w/o SK I-III)	39.56 ± 4.27	2.25σ

9.2 Future prospects

Possible updates for future joint analysis   The future sensitivity study presented in Section 9.1 was done by simply scaling the statistic of the data, but there are some possible updates for future joint analysis. For the SK-VI and the following SK data-taking phases, gadolinium is dissolved into water. This will improve the neutron tagging efficiency from $\simeq 20\%$ (pure water) to $\simeq 50\%$ (0.01% concentration) and to $\simeq 75\%$ (0.03% concentration), which results in a better separation of the neutrino and antineutrino events and a better resolution of the reconstructed neutrino energy. Therefore, the sensitivity to δ_{CP} and mass ordering is expected to be improved.

In the joint analysis framework used in this thesis, the systematic uncertainties of the neutrino fluxes are treated as uncorrelated due to the different origins of the SK atmospheric neutrinos and the T2K accelerator neutrinos. However, the hadron interaction processes between protons (primary cosmic ray protons or accelerator protons) and target nuclei and the propagation of the secondary hadrons can be regarded as common to some extent between SK and T2K. Although we apply the T2K near detector constraints only to the cross-section systematic uncertainties for the atmospheric samples in this analysis, it will be possible to also apply the constraints to the flux uncertainties if we can correlate the flux modeling in future analysis.

Upgrades of the T2K beamline and near detectors   T2K is currently undergoing upgrades of the accelerators and the neutrino beamline [179], and the near detectors [180]. With the upgrade of the beamline, the beam power will be increased from 500 kW to 1.3 MW, which will help us to obtain more statistics in future analysis. The upgraded ND280 is expected to reduce the neutrino interaction uncertainties through the precise measurements of outcoming hadron kinematics and improve the sensitivities to the oscillation parameters [181]. The upgrade of ND280 may also improve the capability of ν_e cross-section measurements using the intrinsic ν_e components in the neutrino beam [182], which is particularly important for improving the δ_{CP} constraints.

Hyper-Kamiokande   Hyper-Kamiokande (HK) is a 258 kton water Cherenkov detector [176]. It is being constructed and is expected to start data-taking in 2027. The HK project includes both the atmospheric and long-baseline accelerator neutrino analyses, so the joint analysis framework developed in this study will be vital for future HK analysis. The latest study by the HK Collaboration shows that we expect to achieve a 5σ-level significance of $\sin\delta_{\mathrm{CP}} = 0$ exclusion with 10 years of data taking at HK as shown in Fig. 9.3. The mass ordering determination at $> 3.8\sigma$ confidence level can also be achieved when a combined analysis of 10 years of atmospheric and accelerator neutrinos is performed.
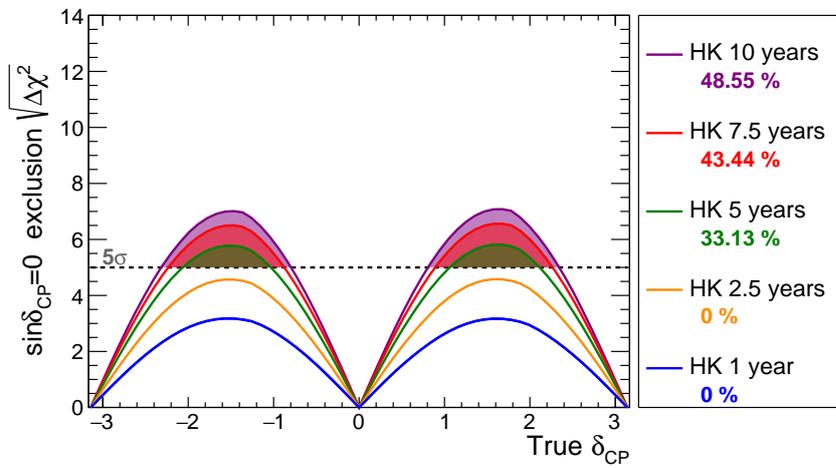
Figure 9.3. Expected power of excluding $\sin \delta_{\mathrm{CP}} = 0$ with the HK long-baseline analysis as a function of true δ_{CP} and the data-taking years. The Asimov A values are assumed for the oscillation parameters except for δ_{CP}, and the mass ordering is fixed to be normal in the analysis. The systematic uncertainty model is taken from the T2K analysis reported in Ref. [37]. The true δ_{CP} intervals that give the 5σ-level of $\sin \delta_{\mathrm{CP}} = 0$ exclusion are shown with the shaded colors and the ratios of these interval sizes to the total δ_{CP} range are summarized in the legend. Note that atmospheric neutrinos are not used in this study. The figure is made by the Hyper-Kamiokande collaboration.

Chapter 10

Conclusions

This thesis presented the first data fit results of the joint oscillation analysis between the SK atmospheric neutrinos and T2K accelerator neutrinos. The SK atmospheric neutrino analysis and the T2K accelerator neutrino analysis have complementary sensitivities to the open questions in neutrino oscillation physics: CP symmetry in the lepton sector, neutrino mass ordering, and the octant of θ_{23}. In this analysis, we established a method to analyze the atmospheric neutrinos and accelerator neutrinos in a single framework by correctly taking into account the correlations of systematic uncertainties.

The Bayesian analysis results showed improvements in the constraints of the CP-related parameters and mass ordering determination compared to the individual experiments. The CP-conserving values of δ_{CP} ($\delta_{\mathrm{CP}} = 0, \pi$) are excluded at 2σ under the flat prior in δ_{CP} and at 1.5σ under the flat prior in $\sin \delta_{\mathrm{CP}}$. The CP-conserving value of the Jarlskog invariant ($J_{\mathrm{CP}} = 0$) is excluded at 2σ under both priors. The Jarlskog invariant is a parameterization-independent measure of CP violation, and it is the first time in the world that we consistently exclude the CP-conserving value of J_{CP} from the 2σ credible intervals.

The Bayes factor for the normal ordering over inverted ordering is computed to be 8.98 ± 0.06 under the flat δ_{CP} prior, which corresponds to the significance level of 1.64σ. It suggests a moderate preference for normal ordering, but it is not enough to claim a discovery. The preference for both octants was found to be almost equal as the atmospheric and accelerator neutrino samples show opposite preferences for the octant.

A variety of detailed discussions have been presented in Chapter 8 to understand the systematic uncertainty constraints and contribution of each sample. All these discussions will be valuable inputs for future analysis. Finally, the future sensitivity study in Chapter 9 showed that we expect to achieve better constraints in the CP phase and stronger rejection of inverted ordering after a few years of data taking.

Appendix A

Details of the detector systematic uncertainties

This appendix presents the details of the detector systematic uncertainties that have not been described in Section 4.4.

A.1   SK atmospheric detector systematic uncertainties

The SK atmospheric detector systematics are parametrized with 27 parameters that correspond to the event normalization effect at each step of the selection, as described in the appendix of Ref. [52]. The sizes of some detector systematic uncertainties are defined separately for the events within $50 < d_{\mathrm{wall}} < 200$ cm and events in $d_{\mathrm{wall}} > 200$ cm to take into account the relatively larger uncertainties for the events with the vertex near the wall. Detailed descriptions of these systematic uncertainties may be found in [78, 79], but an overview is given in the following. The full list of the atmospheric detector systematic uncertainty parameters is given in Table 4.3.

Data reduction and sample separation   The first step of the selection, which classifies the observed events into FC, PC, and UpMu categories depending on the observed inner and outer detector hits, is called "data reduction". The uncertainties of these reduction processes are estimated from the data/MC comparison and are applied as a 1.3% (1.0%) normalization error fully correlated among all the FC (PC) samples. For UpMu samples, fully correlated normalization uncertainties of 0.5% and 0.3% are applied to the stopping and through-going samples, respectively. In addition, a small fraction of FC or PC events could be wrongly classified into the other sample, so an additional FC/PC sample separation uncertainty is included as a 0.02% uncertainty fully anticorrelated between FC single-ring multi-GeV μ-like events and the corresponding PC events. The PC samples have additional uncertainties on the separation of stopping and through-going events. They are implemented as three anticorrelated normalization uncertainties for the stopping/through-going events depending on the position of the hits in the outer detector (top, bottom, and barrel).

UpMu specific   There are some additional uncertainties for the UpMu samples. For the separations of the stopping/through-going events and showering/non-showering events, fully anticorrelated normalization uncertainties are applied. Two additional normalization uncertainties are applied to the UpMu stopping and through-going events due to the energy selection cut and path-length cut. Finally, the uncertainties on the cosmic muon background in the UpMu samples are included. These backgrounds come from the up-scattering of cosmic ray muons and cannot be reduced in the selection. Therefore, the estimated number of background events must be subtracted from the data when we perform the data fit. This subtraction is only applied to the bins near the horizontal direction ($-0.2 < \cos \Theta_z < 0$). The uncertainty on the estimated number of backgrounds is used as the 1σ of normalization uncertainty of these samples. More details on the UpMu sample background subtraction may be found in Ref. [183].

Fiducial volume and decay electron cut   The fiducial volume definition is also regarded as a source of uncertainty. The FC and PC events can be wrongly reconstructed inside or outside the fiducial volume due to the resolution of the vertex reconstruction. This uncertainty is defined as a 2% normalization uncertainty fully correlated between the FC and PC samples. It is not applied to the UpMu samples as the vertex fiducial volume cut is not applied to these samples. The uncertainty on the decay electron tagging is estimated by comparing the number of tagged Michel electrons in the cosmic ray muon data and MC.

Reconstruction   No control sample can cover the same energies and event topologies as the atmospheric samples, and therefore, we use the atmospheric neutrino data to evaluate the systematic uncertainties for the PID, the number of rings (ring separation), and the likelihood-based two-stage multi-GeV multi-ring e-like sample separation (MME). For each step of the selection, we construct the signal and background likelihood distributions based on the MC and apply a *shift* and *scaling* to obtain the modified likelihood as

$$\mathcal{L}_{\mathrm{mod}}^{\mathrm{MC}} = (\beta_{\mathrm{sig}}^1 \mathcal{L}_{\mathrm{sig}}^{\mathrm{MC}} + \beta_{\mathrm{sig}}^0) \oplus (\beta_{\mathrm{bkg}}^1 \mathcal{L}_{\mathrm{bkg}}^{\mathrm{MC}} + \beta_{\mathrm{bkg}}^0), \tag{A.1}$$

where β^i denotes the shifting and scaling parameters. The signal and background samples are selected for each likelihood. For example, e-like and μ-like samples are used as the signal and background for the PID likelihood. Then, we fit this modified MC likelihood to the data likelihood and obtain the errors for the four β^i parameters. Using the obtained errors, we vary the likelihood assuming the Gaussian fluctuation of each β^i around their best-fit values, and compute the fluctuation of the event rate in each sample. The size of the fluctuation is then used as the normalization errors from these systematic uncertainty sources. These systematic uncertainties therefore have correlations or anticorrelations among samples as shown in Table 4.3.

   The uncertainty on the two-ring π^0 selection is estimated by applying the ring-counting error and the PID error to both rings. The method described above is used to estimate the uncertainties on each ring, and the final uncertainty on this selection is estimated to be 1.03%.

Non-neutrino background There are uncertainties on the two non-neutrino background sources. One is the cosmic ray muon background where cosmic muons can reach the detector before decaying or stopping on the way. A normalization uncertainty is applied to the FC and PC samples. We also have an additional cosmic muon subtraction step for the events with the vertex close to the inner detector wall ($50 < d_{\text{wall}} < 200$) to remove the larger cosmic ray background found in the multi-GeV μ-like samples. An additional uncertainty on this cosmic muon subtraction step is included on top of the overall cosmic muon background normalization. The other non-neutrino background source is the PMT flasher background. The PMTs can sometimes produce flashes of light which mimics the low-energy neutrino events. A normalization uncertainty is included for this flasher background.

Energy calibration In addition to the systematic uncertainties that affect the event rate normalizations, there are also systematic uncertainties in the energy calibration. The uncertainty of the absolute energy scale is estimated using several control samples as shown in Fig. A.1. Although it has some variations in different energy regions, the total systematic error is computed to be 2.17% from the quadrature sum of the largest MC/data difference and the average of the momentum. The size of the uncertainty is the same for all the samples independent of the event topology and energy. There is one additional energy-scale uncertainty due to the up/down asymmetry of the detector light transmission, which is only applied to the FC and PC samples.
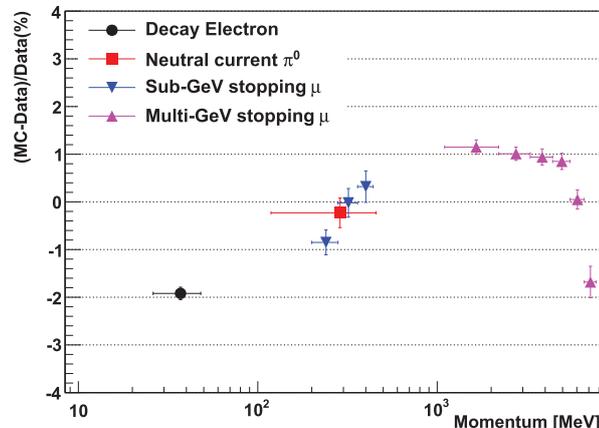
Figure A.1. Uncertainty on the energy scale estimated from the comparison of data and MC in the three calibration samples: cosmic muon decay, atmospheric neutrino NCπ^0 events, and stopping cosmic ray muons. The vertical error bars show the statistical uncertainty and the horizontal error bars show the momentum range covered by each control sample. The figure is taken from [36].

A.2 T2K detector systematic uncertainties

In the main analysis, we evaluated the T2K detector systematic uncertainties by taking into account the correlations with the atmospheric detector systematic uncertainties. The details of the uncorrelated detector systematic uncertainty can be found in [80], but briefly reviewed in the following. The detector systematics are implemented as the event rate normalization in each bin defined in Table 4.4. The size of uncorrelated detector systematic uncertainties is estimated using the atmospheric control samples, cosmic ray muon events, and the so-called hybrid π^0 sample as summarized in Table A.1.

Table A.1. Summary of the T2K detector systematic uncertainties with the values and evaluation method. The 1σ uncertainties are given as a fraction of the nominal value, except for the spurious decay$e$ tagging rate. See the text for the details of the PID and ring-counting systematic uncertainties.

Source	Nominal value	1σ	Sample
Decay-e tagging efficiency	90%	1.0%	Cosmic ray muon
Spurious decay-e tagging rate	0.0001/event	0.002/event	Cosmic ray muon
$\mu \to e$ misidentification	3.0%	30%	Cosmic ray muon
Fiducial volume acceptance	100%	0.3-0.4%	Cosmic ray muon
NCπ^0 reduction factor	41.3	26%	Hybrid π^0
PID and ring counting			Atmospheric neutrino

Atmospheric control sample A fit to the atmospheric FC sub-GeV samples is used to obtain the systematic uncertainties related to the PID (e/μ, e/π^0, μ/π^+) and ring counting likelihood cuts. We first classify the MC events into six groups based on the true event topologies: a single e, a single μ, a single e with other particles, a single μ with other particles, a single π^0, and a single p or π^+. Then, the likelihood distributions ($\mathcal{L}_{e/\mu}$, $\mathcal{L}_{e/\pi^0}$, $\mathcal{L}_{\mu/\pi^+}$, $\mathcal{L}_{\text{single/multi}}$) are constructed for each group. In the fit to the atmospheric data, similarly to the atmospheric detector systematic uncertainty evaluation, these four likelihoods are linearly transformed using the scaling (β^1) and shifting (β^0) parameters as

$$\mathcal{L}'_i = \beta^1_i \mathcal{L}_i + \beta^0_i \quad i \in [e/\mu, e/\pi^0, \mu/\pi^+, \text{single/multi}]. \tag{A.2}$$

A fit to the atmospheric data is then performed for each number of decay electrons ($n = 0, 1, 2+$). The fit is performed using the Markov-Chain Monte Carlo method where a simplified flux and cross-section parameterization is used compared to the full atmospheric analysis. The obtained constraints on the scaling and shifting parameters (β^j_i) are then applied to the T2K beam MC events and translated into a covariance matrix taking into account the correlations among β^j_i.

Cosmic ray muons The cosmic muon events stopped inside the inner detector are used to evaluate the uncertainties on the decay electron tagging efficiencies and the fiducial volume cut. The uncertainty on the Michel electron efficiency is estimated by comparing the

number of decay electrons between the stopping cosmic muon data and MC. With a correction to obtain the T2K beam muon momentum and vertex distributions, uncertainties are estimated to be 1% for the decay electron tagging and 0.002/event for the spurious decay electron tagging rate (i.e. fake decay electrons due to the coincident signal). The rate of muon misidentification as an electron is also estimated from the cosmic muon to be 30% (of the nominal misidentification rate of 3%), though the contamination of ν_μ CC events is smaller than 1% (2%) for the 1Re (1Re 1de) sample. An overall uncertainty due to the fiducial volume, which is dominated by the resolution of the reconstructed vertex, is also estimated from the data/MC comparison in the cosmic ray muon sample.

Hybrid π^0 sample   To estimate the uncertainty on the π^0 background, a so-called "hybrid sample" is used, where a single e-like event from the atmospheric data or MC is superimposed with an MC gamma ring so that the composite event kinematics mimic the π^0 event. The T2K 1Re event selection is applied to these hybrid samples, and the π^0 background uncertainty is estimated by comparing the (atmospheric MC 1Re + MC gamma) and (atmospheric data 1Re + MC gamma) results.

Other   For a few background components that are difficult to estimate the uncertainties, conservative values are assumed. These additional background uncertainty sources include the ν_μ and NC backgrounds in the e-like samples and ν_eCC background in the μ-like samples.

All the sources of the uncertainties described above are combined and summarized as a covariance matrix which represents the event rate normalization for binned energy (momentum) ranges. The binning is defined differently for each combination of the reconstructed samples and true event topologies as summarized in Table 4.4. To construct this covariance matrix, we make one million T2K beam pseudo data sets (MC) and randomly throw the parameters corresponding to each systematic source in each pseudo data set. Then, the selection is applied to these pseudo data sets to get the fluctuations in the event rates.

Appendix B

Prior choice for the CP phase in the Bayesian analysis

The Bayesian analysis depends on the choice of the prior by definition. Since the conservation (or violation) of CP symmetry is one of the most important questions we seek to answer in this analysis, we should be careful about our prior choices. In the context of the Bayesian analysis, the prior distribution for a certain parameter represents our prior knowledge (belief) of that parameter. In case we have no prior knowledge of the parameter or do not want to apply any arbitrary assumptions in advance, it is natural to choose the so-called non-informative priors[1]. There are a few choices of non-informative priors. A uniform probability distribution is the simplest choice where we assume that the prior $\pi(\theta)$ for the parameter θ is flat over the allowed parameter space:

$$\pi(\theta) = \text{const.} \tag{B.1}$$

However, this prior has a potential problem in that it is not robust against different parameterizations. For example, flat prior in θ is not flat in $\sin\theta$ anymore.

One alternative approach is to use Jeffreys' prior [184]:

$$\pi(\theta) \propto \sqrt{\det I(\theta)}, \tag{B.2}$$

where $I(\theta)$ is the Fisher information matrix. For example, Jeffreys' prior for the mean μ of the Gaussian distribution $f(x|\mu,\sigma) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{1}{2}\dfrac{(x-\mu)^2}{\sigma^2}\right)$ can be calculated as

$$\pi(\mu) \propto \sqrt{I(\mu)} = \sqrt{\text{E}\left[\left(\frac{x-\mu}{\sigma}\right)^2\right]} = 1/\sigma \propto \text{const.}, \tag{B.3}$$

which is constant for μ. Similarly, Jeffrey's prior for the Poisson distribution $f(n|\lambda) =$

---

[1] The statistical methods frequently used in the particle physics field are reviewed in Ref. [17]. The different choices of the non-informative priors in the Bayesian analysis are also described.

$\dfrac{e^{-\lambda}\lambda^n}{n!}$ can be calculated as

$$\pi(\lambda) \propto \sqrt{I(\lambda)} = \sqrt{\mathrm{E}\left[\left(\frac{n-\lambda}{\lambda}\right)^2\right]} = 1/\sqrt{\lambda}. \tag{B.4}$$

The uniform prior for $\sin\delta_{\mathrm{CP}} \in [-1,1]$ is adopted as an approximation of the Jeffreys' prior for δ_{CP}. According to Ref. [185], the neutrino oscillation probability can be written in the form of

$$P(\nu_\mu \to \nu_e) = A\cos\delta_{\mathrm{CP}} + B\sin\delta_{\mathrm{CP}} + C. \tag{B.5}$$

At the combination of the T2K neutrino energy and baseline, the δ_{CP}-independent term C has the dominant contribution and the $\sin\delta_{\mathrm{CP}}$ term has only a sub-leading contribution as shown in Fig. B.1. The likelihood is defined as the Poisson distribution of the predicted number of events λ, and λ is proportional to the oscillation probability. Therefore, the Jeffreys' prior can be written as

$$\pi(\lambda) \propto \frac{1}{\sqrt{\lambda}} \approx \frac{1}{\sqrt{B\sin\delta_{\mathrm{CP}} + C}} \approx \frac{1}{\sqrt{C}} + \mathcal{O}(\sin\delta_{\mathrm{CP}}), \tag{B.6}$$

which is approximately flat in $\sin\delta_{\mathrm{CP}}$. We should however note that this flat $\sin\delta_{\mathrm{CP}}$ prior is just a rough approximation to see the impact of an alternate prior choice to our analysis, and we do not aim to represent a complete Jeffreys' prior with it.
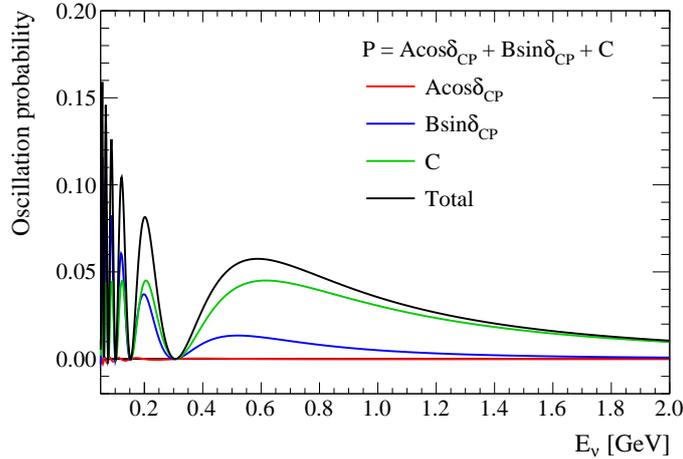
Figure B.1. $\nu_\mu \to \nu_e$ appearance oscillation probability broken down by the δ_{CP} contributions shown in Eq. (B.5). The oscillation parameter set A defined in Table 5.3 is used and the baseline of $L = 295$ km is assumed.

Appendix C

Configuration and validation of MCMC

This appendix presents the technical details of the MCMC method.

C.1 Overview

When using the MCMC method, one needs a long chain with many MCMC steps to reduce the statistical fluctuation due to a finite step size. In addition, one should monitor the obtained MCMC chains and confirm the following [186]

1. The chains have converged around the highest probability region(s) in the parameter space.

2. The steps can move between the highest probability regions.

The first requirement is expressed as that the chain has *achieved stationarity* or has *converged*. If the chain has not converged, it is not guaranteed that the obtained steps represent the posterior probability distributions we want. In that case, the obtained distribution would depend on the initial step and it returns different results depending on when we stop the chain. The second condition is known as the *mixing* of the steps. When we have several highest posterior probability regions in the parameter space and the transitions between them are suppressed due to lower probabilities at the valley, the steps would get stacked at one of these regions. This should also be avoided because the results would depend on the choice of the first step. In Section C.3, some metrics for monitoring these conditions are introduced and evaluated for our data fit MCMC chains.

In addition, since we adopted a multicanonical method in this analysis as discussed in Section 5.3.3, we should tune the parameter β. In Section C.4, autocorrelations of the MCMC chains with different values of β are computed and used to determine the value of β used in the analysis.

It is important to estimate the errors in the inferences obtained from the MCMC method. The MCMC statistical errors corrected with the integrated autocorrelation time (IAT) are discussed in Section C.5.

Finally, the implementation of the mass ordering in the MCMC method is discussed in Section C.6.

C.2 Configuration of MCMC

When utilizing the MCMC method in our Bayesian analysis, we run some short pilot chains before actually running the longer data fit chains. The results of these pilot chains are used to monitor the convergence of the MCMC chains and determine the value of multicanonical β to be used in the main chains. The configurations of these chains are summarized in Table C.1.

Although we run the main chains with both $\beta = 1.0$ and 0.5, only the results with $\beta = 0.5$ are used for obtaining the credible intervals. The chains with $\beta = 1.0$ are only used for the goodness of fit study and CP conservation test as described in Section 7.3.5 and Section 8.6.

Table C.1. Configurations of MCMC chains used for the pilot run and data fit.

Configuration	Pilot chain	Main chain
Multicanonical dimension	2D ($\sin^2\theta_{23}$,δ_{CP})	
Multicanonical β	0.0, 0.5, 1.0	0.5, 1.0
Step size	0.09	0.09
Number of chains	10	1000
Number of steps/chain	1,100,000	210,000
Number of burn-in steps	100,000	10,000
Total number of steps	10 million	200 million

C.3 Convergence diagnostics

To test the convergence of the MCMC chains, we use two metrics known as Geweke's Z statistic [187] and Gelman-Rubin's $\hat{R}$ statistic [188]. As described in Section C.1, there are two conditions to be satisfied when we use the MCMC results: stationarity (convergence) and mixing.

Geweke's Z statistic is used to test the convergence to a stationary distribution within one chain by comparing the distributions at the beginning of the chain and the end of the chain. Let us denote the MCMC steps as x_i for $i = 1, \cdots, n$. To compute Geweke's statistic, we extract the steps within two windows from each chain where one contains the first 10% steps ($n_1 = n/10$) after removing burn-in and the other contains the last 50% steps ($n_2 = n/2$). Then Geweke's Z statistic for a statistic g is defined as

$$Z = \frac{g_{n_1} - g_{n_2}}{\sqrt{S_g(0)/n_1 + S_g(0)/n_2}},\tag{C.1}$$

where

$$g_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1} g(x_i), \quad g_{n_2} = \frac{1}{n_2}\sum_{i=n-n_2+1}^{n} g(x_i),\tag{C.2}$$

and $S_g(\omega)$ is the spectral density of function g. If the chain is converged, the asymptotic limit of the statistic Z follows the standard normal distribution $Z \sim \mathcal{N}(0,1)$.

We calculate the p-value from Z for all the parameters to test whether they are converged or not. We use the criterion of $\alpha = 0.05$ with the Holm-Bonferroni correction described in Section 7.3.5. Since this criterion is very conservative in this case due to a large number of parameters ($= 228$), we also check the trace plots for oscillation parameters and some systematic parameters that show a worse p-value.

Gelman-Rubin's diagnostic is used to test the mixing of the MCMC chains. We run several independent chains with different initial steps and see whether these MCMC chains converged to the same stationary distribution. We compute the statistic $\hat{R}$ for each parameter from m independent chains with n steps as

$$\hat{R} = \frac{\hat{V}}{W} \frac{df}{df - 2}, \tag{C.3}$$

where $\hat{V}$, W, and df are defined as

$$\hat{V} = \frac{n-1}{n}W + \frac{m+1}{mn}B \tag{C.4}$$

$$B = \frac{n}{m-1}\sum_{i=1}^{m}(\bar{x}_{i.} - \bar{x}_{..})^2, \quad \bar{x}_{i.} = \frac{1}{n}\sum_{j=1}^{n}x_{ij}, \quad \bar{x}_{..} = \frac{1}{m}\sum_{i=1}^{m}\bar{x}_{i.} \tag{C.5}$$

$$W = \frac{1}{m}\sum_{i=1}^{m}s_i^2, \quad s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i.})^2 \tag{C.6}$$

$$df = \frac{2\hat{V}^2}{\hat{\mathrm{Var}}[\hat{V}]} \tag{C.7}$$

$$\hat{\mathrm{Var}}[\hat{V}] = \left(\frac{n-1}{n}\right)^2 \frac{1}{m}\frac{1}{m-1}\sum_{i=1}^{m}(s_i^2 - W)^2 + \left(\frac{m+1}{mn}\right)^2\frac{2B^2}{m-1}$$

$$+ \frac{2(m+1)(n-1)}{mn^2}\frac{n}{m}\left[\frac{1}{m-1}\sum_{i=1}^{m}(s_i^2 - W)(\bar{x}_{i.}^2 - \overline{\bar{x}_{.}^2})\right.$$

$$\left. - \frac{2\bar{x}_{..}}{m-1}\sum_{i=1}^{m}(s_i^2 - W)(\bar{x}_{i.} - \bar{x}_{..})\right], \quad \overline{\bar{x}_{.}^2} = \frac{1}{m}\sum_{i=1}^{m}\bar{x}_{i.}^2 \tag{C.8}$$

Here, the index i runs over m chains ($i = 1, \cdots, m$) and j runs over n steps in each chain ($j = 1, \cdots, n$). $\bar{x}_{i.}$ is the mean in chain i, $\bar{x}_{..}$ is the inter-chain mean, and s_i^2 is the variance in chain i. The chains are regarded as well-mixed if $|\sqrt{\hat{R}} - 1| < 0.01$ is satisfied.

Both Geweke's diagnostic and Gelman-Rubin's diagnostic are computed using the 10 pilot chains with 1 million steps after removing the first 100,000 steps as burn-in. The results are shown in Fig. C.1. For Geweke's diagnostic, all the parameters passed the Holm-Bonferroni p-value test. Gelman and Rubin's diagnostic also satisfies the criterion for all the parameters.

The trace plots of the oscillation parameters are shown in Fig. C.2. It shows good mixing and stationarity in all the parameters and no weird behavior is observed in the trace plots. Based on these results, we conclude that our MCMC chains achieved good convergence and can be used for the analysis.
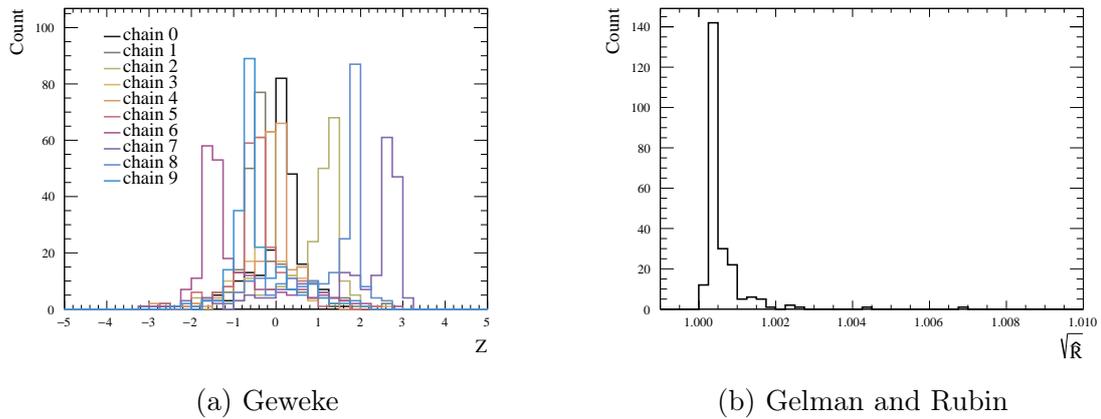
(a) Geweke   (b) Gelman and Rubin

Figure C.1. Distribution of Geweke's Z diagnostics (left) and Gelman and Rubin's $\sqrt{\hat{R}}$ statistic (right). Both diagnostics are computed for all the oscillation and systematic parameters.
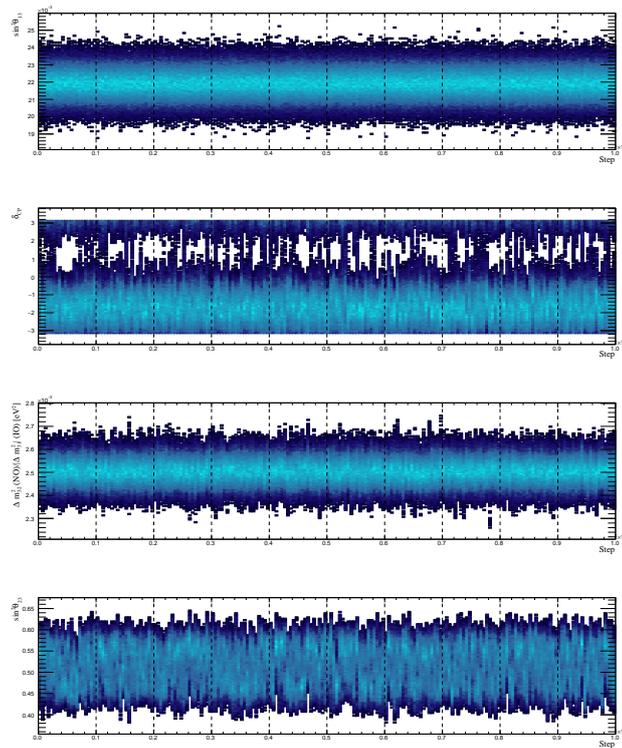

Figure C.2. Trace plot of oscillation parameters. All the pilot chains are concatenated into one trace plot after removing the burn-in period.

C.4   Autocorrelation

In the Monte Carlo method, we always have statistical errors due to a finite size of throws or steps. The MCMC steps have correlations with the consecutive steps since the update process depends on the previous step by definition. It means that an effective number

of steps is much less than the obtained number of steps. The correlation between the consecutive steps in the MCMC chains is called the *autocorrelation*.

Let us treat the MCMC steps $x(t)$ as a function of time t. The autocorrelation function at lag t, which is the correlation between the step $x(t')$ and $x(t'+t)$, is defined as

$$\chi(t) = \int \left[x(t') - \mathrm{E}[x]\right]\left[x(t'+t) - \mathrm{E}[x]\right]dt' \tag{C.9}$$

$$= \int \left[x(t')x(t'+t) - (\mathrm{E}[x])^2\right]dt', \tag{C.10}$$

where $\mathrm{E}[\cdot]$ denotes the expectation. In general, the autocorrelation is expected to fall off exponentially as a function of lag t [189]

$$\chi(t) \approx \chi(0)e^{-t/\tau}. \tag{C.11}$$

The time constant τ, which is known as the *integrated autocorrelation time* (IAT), can be estimated by integrating the autocorrelation divided by the value of autocorrelation at $t = 0$:

$$\int_0^\infty \frac{\chi(t)}{\chi(0)}dt \approx \int_0^\infty e^{-t/\tau}dt = \tau. \tag{C.12}$$

This can also be denoted as

$$\chi(t) = \mathrm{Cov}[x(t'), x(t'+t)], \tag{C.13}$$

and therefore,

$$\tau = \frac{\chi(t)}{\chi(0)} = \frac{\mathrm{Cov}[x(t'), x(t'+t)]}{\mathrm{Var}[x(t')]} = \mathrm{Cor}[x(t'), x(t'+t)], \tag{C.14}$$

where $\mathrm{Var}[\cdot]$, $\mathrm{Cov}[\cdot]$, and $\mathrm{Cor}[\cdot]$ denote the variance, covariance, and correlation, respectively.

When we get an autocorrelation of τ in our results, the effective number of steps is estimated to be $n_{\mathrm{eff}} = n/\tau$. Therefore, it is important to have a smaller value of τ to increase the effective number of MCMC steps.

The integrated autocorrelation times for the oscillation and systematic parameters under different β are shown in Fig. C.3. When we do not use the multicanonical method ($\beta = 1.0$), the autocorrelation of $\sin^2\theta_{23}$ is roughly ~ 10 times longer than the other oscillation parameters. This is due to the bimodality of $\sin^2\theta_{23}$ posterior distribution and the suppressed transition probabilities between the upper and lower octant. This issue, however, disappears when we use the multicanonical method with 2D posterior prediction $f_{\mathrm{pred.}}(\delta_{\mathrm{CP}}, \sin^2\theta_{23}, \mathrm{MO})$, which mitigates the suppression of the octant transition probability. The option $\beta = 0.0$ is an extreme example and it enforces the MCMC steps to explore a very low probability region, which makes the IAT longer than the $\beta = 0.5$ case. Therefore we do not use $\beta = 0.0$ in our main results.

A similar tendency is seen in the systematics parameters where the multicanonical method with $\beta = 0.0$ and $\beta = 0.5$ give better autocorrelations. Among the systematic

uncertainty parameters, some cross-section parameters related to DIS have the longest
IAT ($\tau \simeq 4000$ at the longest), but other systematics have IATs shorter than 1500 steps.

Based on these results, we decided to use the multicanonical $\beta = 0.5$ in our main
analysis. Since the longest IAT is around $\simeq 4000$, the main chains give roughly $\simeq 50,000$
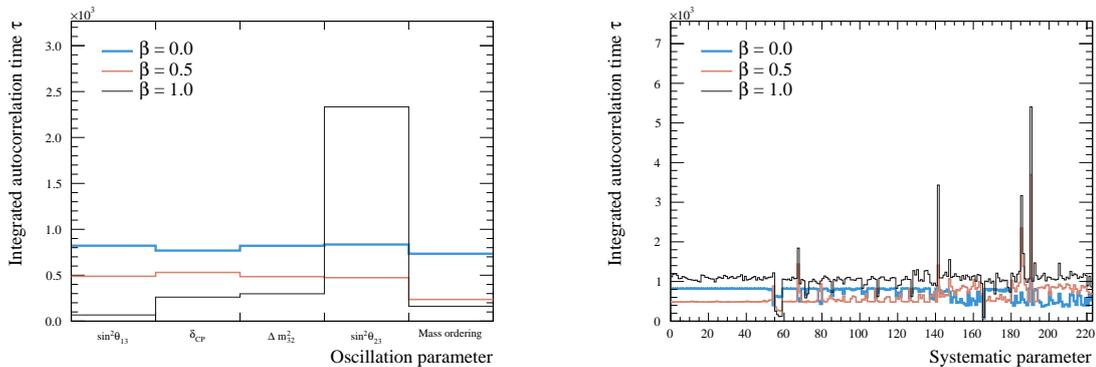effective steps when we run 200 million steps.

Figure C.3. Summary of integrated autocorrelation time of oscillation and systematics
parameters for different values of multicanonical method β. The unit of integrated auto-
correlation time is steps.

C.5 MCMC statistical errors

In this section, we present a method to estimate the MCMC statistical errors in our
results. As described in Section C.4, an effective number of steps is smaller than the
actual number of steps in MCMC. Therefore, we should take into account the integrated
autocorrelation time when estimating the errors in our measurements.

Let us think about the statistic x and its estimator $\hat{x}$. Since each step is weighted in
the multicanonical method, we will use the weighted mean of x as the estimator $\hat{x}$:

$$\hat{x} := \frac{\sum_i w_i x_i}{\sum_i w_i} = \frac{1}{n} \sum_i w'_i x_i = \mathrm{E}[w'x], \tag{C.15}$$

where we define the normalized weight w' as

$$w'_i = \frac{w_i}{\frac{1}{n} \sum_j w_j}. \tag{C.16}$$

In this case, the estimator $\hat{x}$ is equivalent to the (unweighted) mean of $w'x$. The variance

of this estimator can be computed as

$$\text{Var}\left[\hat{x}\right] = \text{E}\left[(\hat{x} - \text{E}[\hat{x}])^2\right] \tag{C.17}$$

$$= \frac{1}{n^2}\text{E}\left[\sum_i \sum_j \left(w'_i x_i - \text{E}[w'x]\right)\left(w'_j x_j - \text{E}[w'x]\right)\right] \tag{C.18}$$

$$= \frac{1}{n^2}\text{E}\left[\sum_i \sum_{k=-\infty}^{\infty} \left(w'_i x_i - \text{E}[w'x]\right)\left(w'_{i+k} x_{i+k} - \text{E}[w'x]\right)\right] \tag{C.19}$$

$$= \frac{1}{n}\text{Var}[w'x]\sum_{k=-\infty}^{\infty} \frac{\text{Cov}\left[w'x, w'_k x_k\right]}{\text{Var}[w'x]} \tag{C.20}$$

$$= \frac{1}{n}\text{Var}[w'x] \times 2\tau, \tag{C.21}$$

where τ is the integrated autocorrelation time defined in Eq. (C.14).

The error in each bin is estimated using this method in the plots shown in Section 7.2.1. Here we define the statistic x as the (weighted) number of steps that fall in bin b

$$\hat{x}_b \equiv \frac{n}{\sum_i w_i}\sum_{i=1}^{n} w_i I_b[x_i] = \sum_{i=1}^{n} w'_i I_b[x_i], \tag{C.22}$$

where $I_b[\cdot]$ is an indicator function which returns 1 when the parameter x falls in bin b and returns 0 otherwise. The errors on the mass ordering Bayes factors in Eq. (7.1) are also estimated using this method.

As a cross-check, we perform a jackknife resampling to estimate the error on the mass ordering Bayes factor. The jackknife error is defined as

$$\text{Var}_{\text{jackknife}}[\text{BF}] = \frac{m-1}{m}\sum_{i=1}^{m}(\text{BF}_i - \overline{\text{BF}}) \tag{C.23}$$

$$\overline{\text{BF}} = \frac{1}{M}\sum_{i=1}^{m}\text{BF}_i, \tag{C.24}$$

where BF_i is the mass ordering Bayes factor calculated using $m-1$ chains by removing the m-th chain. In this case, we use $m = 10$ pilot chains for the calculation. The result is 9.27 ± 0.21 and it is consistent with the error value estimated using the variance with the integrated autocorrelation time correction (9.24 ± 0.24). We should note that this is just for the validation of the method based on the pilot chains, and these values are different from our main results presented in Section 7.3.4.

C.6 Mass ordering implementation in MCMC

Potential problems of continuous mass ordering implementation In the fitter we used for the analysis, we implement the atmospheric squared mass difference as the combination of two parameters: the absolute squared mass difference $\Delta m_{32}^2/|\Delta m_{31}^2|$ and mass ordering. The neutrino mass ordering is a binary parameter that takes "normal

ordering" or "inverted ordering". However, in the MCMC method that we used for the data fit, it was implemented as a continuous parameter (let us denote as θ) in [0,1], similarly to the other continuous oscillation and systematics uncertainty parameters. We use the uniform prior in [0,1] and it was regarded as normal ordering when $\theta \in [0.0, 0.5]$ and as inverted ordering when $\theta \in [0.5, 1.0]$.

There are several potential problems with this implementation. When the random walk goes away from the boundary ($\theta = 0.5$), it suppresses the flipping probability of mass ordering. It is not physical that the flipping probability depends on the distance from the boundary. It could be regarded as an effective mass-ordering flip when the jump between the two mass-ordering states happens frequently enough. However, another problem is that we did not have a method to tune the relative step size for the mass ordering parameter. The mass ordering flipping probability should be tunable to make sure that the transition happens frequently. Finally, in the previous implementation, proposals of the parameters were not modified when the mass ordering flips. The absolute squared mass difference $\Delta m_{32}^2 / |\Delta m_{31}^2|$ is known to have different best-fit values in normal and inverted ordering. Therefore, if we only flip the mass ordering and do not adjust the value of $\Delta m_{32}^2 / |\Delta m_{31}^2|$, the transition probability would be suppressed.

Updated implementation   After finishing the main analysis, we implemented mass ordering as a binary parameter and added a tunable flipping probability parameter for normal and inverted ordering.

The integrated autocorrelation time (IAT) of MCMC chains with the updated mass ordering implementation for different combinations of flipping probability and multicanonical β are shown in Fig. C.4. The combination of flipping probability of 0.3 and multicanonical $\beta = 0.2$ gives the shortest IAT for these parameters, so we adopt this combination in the following study.
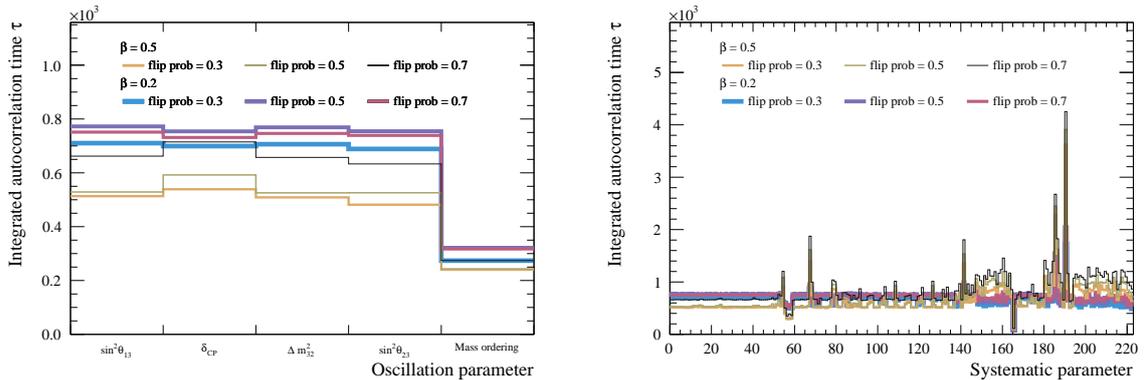

Figure C.4. Summary of integrated autocorrelation time of oscillation and systematics parameters for different values of multicanonical method β and mass ordering flip probability.

Figure C.5 shows the comparison of posterior distributions with the continuous and binary mass ordering implementation. Both implementations give very consistent results.

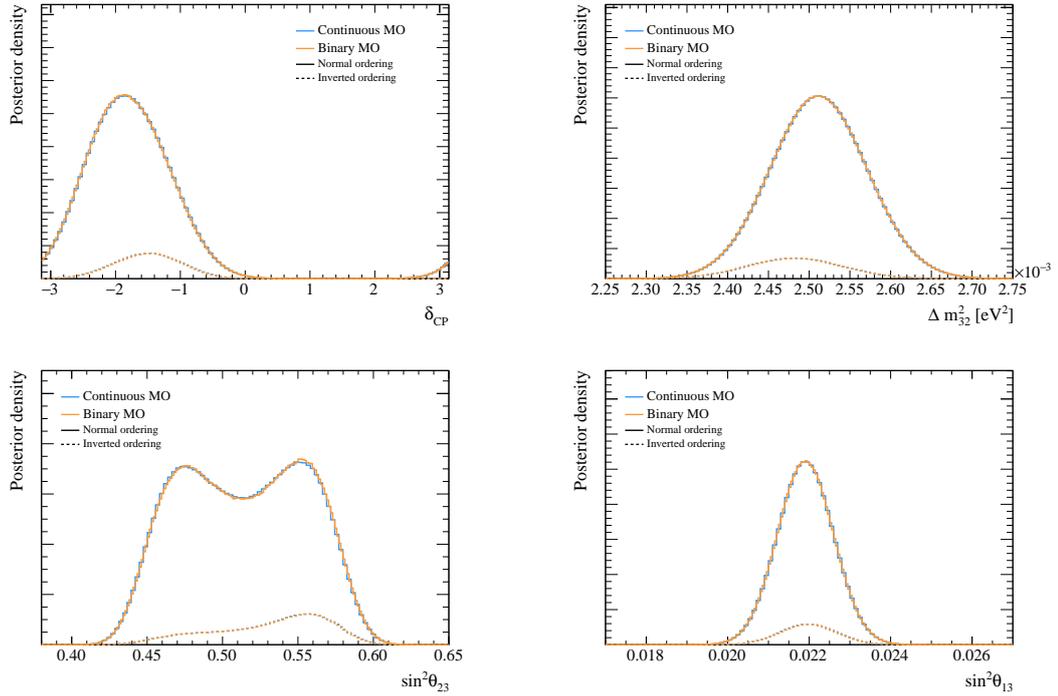The posterior probabilities for the different combinations of mass ordering and octant

Figure C.5. Comparison of the posterior distributions with the continuous and binary implementations of mass ordering.

are shown in Table C.2. It agrees with the continuous implementation results shown in Table 7.6 within the estimated size of MCMC statistical errors.

In conclusion, the updated binary implementation of mass ordering gave consistent results with the previous implementation with continuous mass ordering. Therefore, we will not update our main results and consider this result as just for validation.

Table C.2. Posterior probabilities for different combinations of mass ordering and octant hypotheses computed with the binary mass ordering implementation. The mass ordering Bayes factor is shown together.

	SK+T2K		
	Lower octant	Upper octant	Line total
Normal ordering	0.365	0.535	0.900
Inverted ordering	0.022	0.077	0.100
Column total	0.387	0.613	1.000
MO Bayes factor	9.04 ± 0.05		

Appendix D

Additional material for the robustness test of the model

This appendix presents the details of the scaled Asimov method and the rest of the model robustness tests that are not included in Chapter 6.

D.1   Scaled Asimov method

A possible problem of the fake data study described in Section 6.1 is that the change of statistics in the fake data set could lead to a change of the constraint on the oscillation parameters, even in the absence of problems in the model itself. Therefore, we introduce a method called "scaled Asimov" where we apply the same scaling to the model prediction as the one we apply to the MC when constructing the fake data set.

In this oscillation analysis, the predicted number of events λ_i at a certain reconstructed bin i is expressed as follows:

$$\lambda_i = \frac{s}{N} \sum_{k=1}^{N} I_i[x_k],$$

$$I_i[x_k] = \begin{cases} 1 & (x_k \text{ in bin } i) \\ 0 & (\text{otherwise}) \end{cases}$$

where N is the total number of events in MC, s is the overall scale factor and x_k denotes any reconstructed variables of event k by which events are binned. In the generation of the fake data sets, the predicted number of events is modified with event-by-event weights:

$$\lambda_i^{\text{FDS}} = \frac{s}{N} \sum_{k=1}^{N} I_i[x_k] \times w_k^{\text{FDS}}.$$

During the fit, we use a parameterized model:

$$\lambda_i(\boldsymbol{\theta}) = \lambda_i \times \prod_p f_{i,p}(\theta_p),$$

where the number of events in each bin is scaled by the product of the oscillation probabilities and the response functions of systematic uncertainty parameters (collectively denoted as $f(\boldsymbol{\theta})$).

For the nominal Asimov fit, we fit the data set λ_i with our nominal model $\lambda_i(\boldsymbol{\theta})$, while for fake data fit, the data set λ_i is replaced with the fake data set λ_i^{FDS}. For the scaled Asimov fit, we also use the fake data set λ_i^{FDS} but fit it with the scaled model defined as:

$$\lambda_i^{\mathrm{scaled}}(\boldsymbol{\theta}) = \lambda_i^{\mathrm{FDS}} \times \prod_p f_{i,p}(\theta_p).$$

The inputs used in the normal Asimov and scaled Asimov fits are summarized in Table D.1.

Table D.1. Inputs used in the different types of fits done for fake data study. The fit model used at the T2K ND is always the nominal model with a nominal central value.

	SK data set generation	ND data set generation	SK fit model
Nominal fit	Nominal Asimov	Nominal Asimov	Nominal
Fake data fit	Fake data	Fake data	Nominal
Scaled Asimov fit	Fake data	Nominal Asimov	Rescaled to fake data

Although the scaled Asimov method can reduce the effect of statistical change between the nominal Asimov and fake data set, it could also absorb the effects coming from the alternate models. To avoid this issue, we apply the scaled Asimov method only when the SK event spectra are well covered by the predictions from the ND fake data fit. For the determination of whether to use the scaled Asimov method or not, we first construct the fake data spectra and the ND-predicted error bands by varying the flux and ND-constrained cross-section parameters according to the Gaussian approximation. Then, we compute χ^2 for each sample as

$$\chi^2 = \sum_{i=1}^{n_{\mathrm{bin}}} \left(\frac{n_i^{\mathrm{FDS}} - n_i^{\mathrm{ND}}}{\sigma_i^{\mathrm{ND}}} \right)^2,$$

where n_i^{ND} and σ_i^{ND} are the mean and standard deviation of the error band at bin i, respectively. We use the scaled Asimov method when all the beam and sub-GeV atmospheric samples satisfy $\chi^2/\mathrm{ndf} < 1$ where ndf denotes the number of degrees of freedom (i.e. the number of bins in the sample n_{bin}). For the beam μ-like samples, we only use the bins below $E_{\mathrm{rec}} < 3$ GeV to avoid statistical fluctuation in the higher energy bins where we have much fewer events. Regarding the binning, the E_{rec} projection is used for the beam samples and the momentum projection is used for the atmospheric samples when they have momentum binning, otherwise, $\cos\Theta_z$ is used. This is because some fake data sets are produced by reweighting the MC as a function of the energy/momentum, and the E_{rec}/p binning can capture the difference between the fake data set and prediction

better than the angular binning. We should note that the error bands for the atmospheric sub-GeV samples are dominated by the flux systematics which is not constrained by the near detector fit. Therefore, we tend to overestimate the agreement of fake data sets and near detector predictions for the atmospheric samples with the definition described above. A more proper criterion will be investigated for future analysis.

In Section 6.3, the event spectra and the error bands of these predictions for the T2K and SK atmospheric sub-GeV samples are shown for each fake data set and used to decide whether we can use the scaled Asimov method or not. When we decide to use the scaled Asimov method for a certain fake data study, the metrics introduced in Section 6.1.3 are constructed by comparing the fake data fit results to the scaled Asimov fit results instead of the nominal Asimov fit results. The choice of whether to use the scaled Asimov or not for each fake data study is shown in the summary tables such as Table 6.8.

D.2   Alternative models

The list of robustness tests is summarized in Table D.2.

Table D.2. Summary of robustness test studied in the joint SK + T2K analysis.

Alternative model name	Model component	Description	Results
CRPA	Nuclear model	Section 6.2.1	Section 6.3.1
LFG	Nuclear model	Section D.2.1	Section D.3
Removal energy	Nuclear model	Section D.2.2	Section D.3
Axial form factors	CCQE	Section D.2.3	Section D.3
Martini 2p2h	2p2h	Section D.2.4	Section D.3
Pion multiplicity	CC$n\pi$	Section D.2.5	Section D.3
Energy-dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$	$\sigma_{\nu_e}/\sigma_{\nu_\mu}$	Section D.2.6	Section D.3
Atmospheric CC1π	CC1π	Section 6.2.2	Section 6.3.2
ND Non-QE CC0π	CC0π	Section D.2.7	Section D.3
ND CC1π	CC1π	Section D.2.8	Section D.3
ND extrapolation	Fit	Section 6.2.3	Section 6.3.3
Pion SI bug fix	CC1π, CC$n\pi$	Section D.2.9	Section D.3

D.2.1   Nuclear Model: LFG

Local Fermi Gas (LFG) model was proposed by Nieves et al. [115, 116] as a natural extension of the global Relativistic Fermi Gas (RFG) model. As described in Section 4.3.1.2, the Fermi momentum p_F is assumed to be constant in RFG. LFG is based on the local density approximation, and the Fermi momentum is assumed to be

$$p_F(r) = (3\pi^2 \rho(r))^{1/3}, \tag{D.1}$$

where $\rho(r)$ is a radial nucleon density function that describes the initial state of nucleons in a nucleus. As already explained in Section 4.3.1.2, LFG is known to overpredict the

CCQE cross-section at the low-momentum (energy) transfer. Therefore, it is used with the random phase approximation (RPA). In the following, the LFG+RPA model is simply referred to as LFG.

In the analysis, we use the spectral function (SF) as our nominal model for the nuclear model, but both the LFG and SF give a similar level of precision in the cross-section predictions. From the T2K ND280 cross-section measurements, it was found that LFG gives the best description of the events without a proton and SF gives the best description of the events with one or more protons [122]. Therefore, we cannot simply discard one of the models and want to test the possible biases using the alternative model.

Similarly to the CRPA fake data set generation procedure described in Section 6.2.1, the input MC at the Pre-ND ($Q^2 = 1$) tuning is reweighted to obtain the event spectra predicted by LFG. The reweighting factors are obtained separately for different neutrino types (neutrino and antineutrino), different neutrino flavors (muon and electron), and different target nucleus species (carbon and oxygen) by taking the ratio of SF and LFG cross-section predictions. Since we use the Pre-ND ($Q^2 = 1$) tuning for the generation of this fake data set, we use the nominal Asimov defined at the same Pre-ND ($Q^2 = 1$) tuning for the reference fit.

D.2.2   Removal energy

In our systematic model, we have four removal energy systematic uncertainty parameters for each combination of neutrino and antineutrino, and oxygen and carbon. Figure D.1 shows the probability density distribution for SF for oxygen as a function of the removal energy. In NEUT, the removal energy distribution is treated as the same for protons and neutrons. However, an alternative model by Bodek et al. [190] predicts different values of removal energies for protons and neutrons. The agreement for protons is relatively good, but the predictions for neutrons differ by up to 4 MeV for oxygen and up to 2 MeV for carbon. Therefore, the parameters for neutrinos (the main targets are neutrons) are assigned a nominal value of +4(2) MeV for oxygen (carbon), whereas the nominal values of 0 MeV are assigned to antineutrinos (the main targets are protons) to take into account these differences. The size of the systematic uncertainty is defined to be ±6 MeV from the uncertainties in the electron scattering measurements [191, 192] to which the model is tuned. For the far detector fit, only the two variables for the oxygen target are used.

A robustness test is performed to test whether this treatment is enough to cover the effect of having a large deviation in the true removal energy compared to our nominal value. We generate the fake data set by setting the removal energy at +15 MeV. We should note that this value is a rather extreme assumption just for the test of the model and not physical. We use the nominal Asimov defined at the Pre-ND tuning for the reference fit.

D.2.3   Axial form factors

In our nominal model, the axial form factor of the nucleon is modeled using a dipole shape form with an axial mass M_A^{QE} as shown in Eq. (4.16). The world average of the axial mass is known to be $M_A^{\mathrm{QE}} = 1.026 \pm 0.021$ GeV [101]. However, as shown in Fig. D.2, the MiniBooNE ν_μ CCQE cross-section measurement suggested a much larger
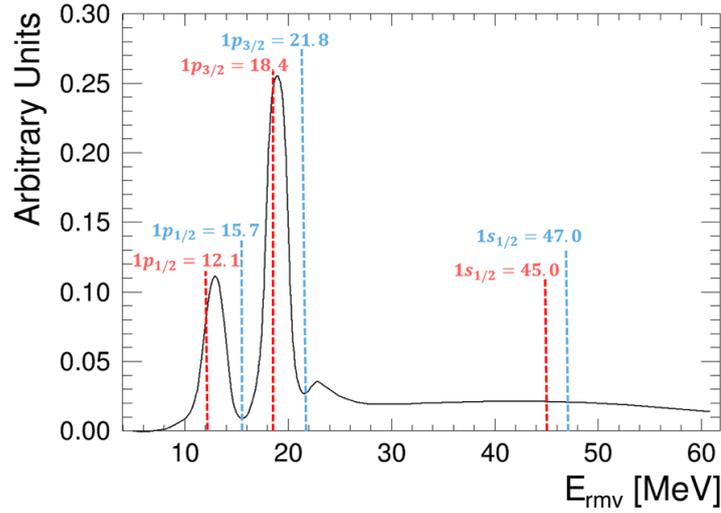
Figure D.1. Probability density distribution for the spectral function for oxygen as a function of the removal energy in `NEUT` [98] (black solid line). The predictions for the shell positions from Bodek et al. [190] are shown with the vertical dashed lines for protons (red) and neutrons (blue). The figure is taken from Ref. [37].

value of $M_A^{\mathrm{QE}} = 1.35 \pm 0.17$ GeV [193]. This anomaly suggested the existence of nucleon correlations inside a nucleus [126], which is known as the 2p2h process. However, at the same time, the assumption of the dipole form factor is not well-motivated, and some alternative parameterizations of the form factors have been proposed.

Adamuščín et al. [164] proposed the two-component form factor, in which the form factor is described as

$$F_A^{2\text{-comp}}(Q^2) = F_A(0)(1 + \gamma Q^2)^{-2} \left[1 - \alpha + \alpha \frac{m_A^2}{m_A^2 + Q^2} \right], \tag{D.2}$$

where m_A is the mass of the lowest axial meson $a_1(1260)$ with $I^G(J^{PC}) = 1^-(1^{++})$ and can be fixed at $m_A = 1.230$ GeV. Ignoring the constant term written as $F_A(0)$, Eq. (D.2) consists of the two components. The first term represents the coupling to the intrinsic structure of the nucleon with a free parameter γ, and the second term represents the contribution from the axial meson quark-antiquark cloud with a free parameter α. Although this model gives more freedom to fit data, it still underestimated the cross-section at the high-Q^2 region [196].

As a natural extension of this model, a three-component form factor is also proposed by adding an exponential term as

$$F_A^{3\text{-comp}}(Q^2) = F_A^{2\text{-comp}}(Q^2) + F_A(0) \left[\theta' C Q^2 e^{-\beta Q^2} \right], \tag{D.3}$$

$$\theta' = \mathrm{sgn}(\theta)\sqrt{|\theta|\beta}. \tag{D.4}$$

This model allows us to smoothly connect the two-component form factor and the usual dipole form factor. We should note that θ' and β were originally implemented as free
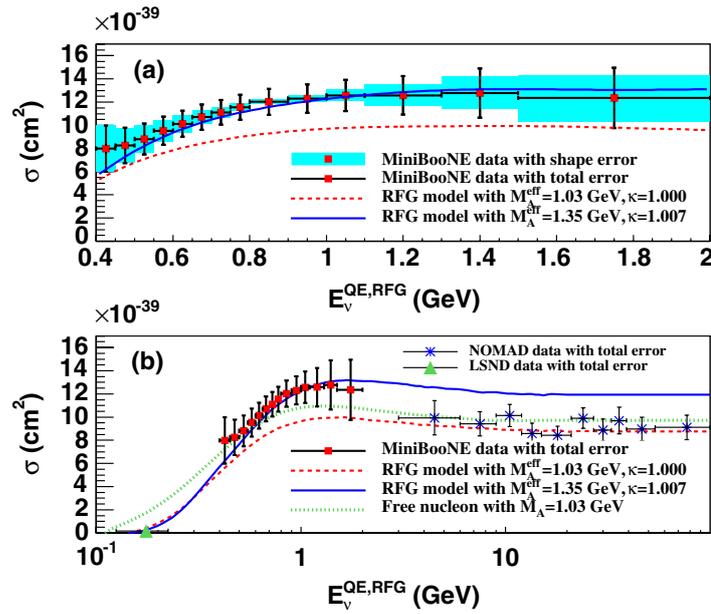
Figure D.2. MiniBooNE ν_μ CCQE cross-section on neutron as a function of neutrino energy. The top plot shows the MiniBooNE data in a narrower range, and the bottom plot shows the data in a wider range along with the data from the LSND [194] and NOMAD [195] experiments. The model predictions are shown with two values of axial mass $M_A^{\mathrm{QE}} = 1.03, 1.35$. The figure is taken from Ref. [193].

parameters but they were able to suppress the contribution of the exponential term unrealistically when β takes a large value. Therefore, instead of treating θ' itself as a free parameter, we redefine it as shown above and treat θ and β as the two free parameters.

This fake data study aims to test whether our systematic model can cover the possible variations due to the different assumptions in the form factor. For the generation of the fake data study, the weights are calculated by tuning the three-component form factor to the bubble chamber experiment data, such as ANL [197, 198], BNL [199], FNAL [200], and BEBC [201]. Then, the weight of each event is calculated as the cross-section ratio between the three-component form factor and our nominal dipole form factor. We use the nominal Asimov defined at the Pre-ND tuning for the reference fit.

D.2.4 Martini 2p2h

The 2p2h process is the most important background to the CC0π selection as it biases the reconstructed neutrino energy if one assumes the CCQE-like topology. We use the Nieves et al. model [116] as our nominal model for the 2p2h process. However, as shown in Fig. 4.13, different models give cross-section predictions that differ by a factor of 2. Therefore, we test the effect of the different model predictions using the Martini et al. model [127]. SuSAv2 [131, 129] is not used for this test as it has a less significant difference compared to the Martini et al. model.

The fake data set is generated by simply reweighting the 2p2h events to match the

predictions from the Martini et al. model. The weights are extracted by taking the ratio between the Martini and Nieves predictions as a function of the neutrino energy for neutrinos and antineutrinos as shown in Fig. D.3. When extracting the weights, a cap is applied at 3 to avoid unphysical weighting. We use the nominal Asimov defined at the Pre-ND tuning for the reference fit.
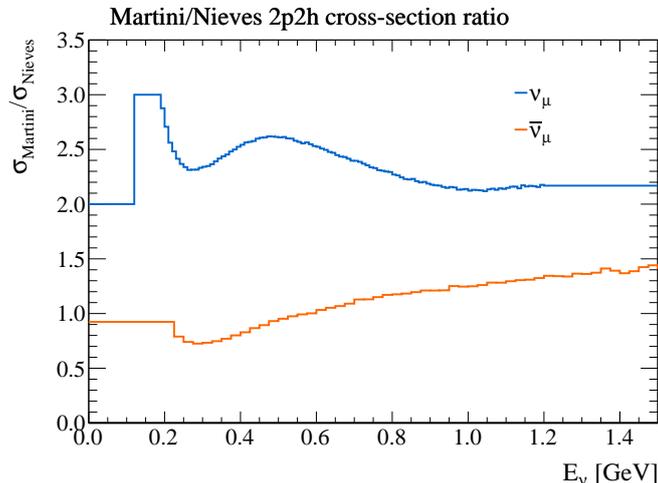

Figure D.3. Ratio between the 2p2h cross-section predictions from the Martini et al. model [127] and Nieves et al. model [115].

D.2.5 Alternative pion multiplicity model

As described in Section 4.3, the multi-pion process is simulated using the multi-pion mode for the events with $W < 2$ GeV and using DIS for $W > 2$ GeV, where W denotes the hadron invariant mass. The single-pion production process is handled by the resonance model at $W < 2$ GeV. To avoid double-counting, the pion multiplicity model is modified to simulate only the events with two or mode pions in the final states. The cross-section for the multi-pion mode is modified accordingly by taking into account the subtraction of the single-pion process.

In the multi-pion mode, the pion multiplicity model gives the number of pions produced in the interaction. Our baseline model is based on the KNO scaling [202] and tuned to the bubble chamber data. However, these measurements show very large variations in the cross-sections as studied in Ref. [203], which results in different predictions depending on the model choice. The different model prediction changes not only the number of pions but also the cross-section of the multi-pion process because the magnitude of the cross-section subtraction is affected.

The difference in the pion multiplicity is expected to have more impact on the high-energy atmospheric samples rather than T2K or low-energy atmospheric samples, as there is a significant fraction of the multi-pion contributions. To test whether the different choices of the pion multiplicity model could affect the oscillation analysis result, we use an alternative model proposed by Bronner et al. [204]. The comparison of the cross-sections from our nominal model and the alternative model is shown in Fig. D.4.
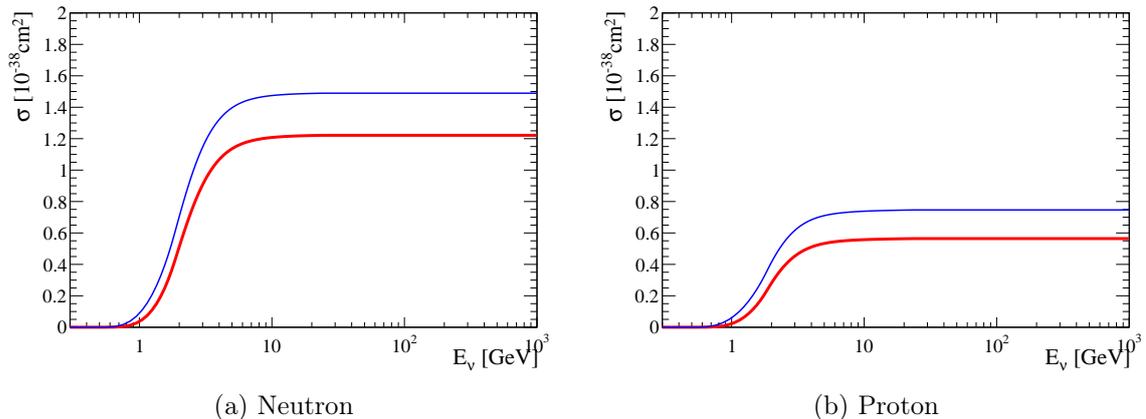
(a) Neutron

(b) Proton

Figure D.4. Comparison of the multi-pion mode cross-sections predictions from the default `NEUT` pion-multiplicity model (red) and an alternative multiplicity model by Bronner et al. (blue) [204]. The cross-sections of ν_μ on the proton and neutron targets are shown.

For the construction of the fake data set, we extract three sets of weights by comparing the nominal model and the alternative model: cross-section ratio as a function of the neutrino energy (E_ν), probability density ratio as a function of the invariant mass (W) and the neutrino energy, and probability density ratio as a function of the number of pions (n_π) and invariant mass. These weights are obtained separately for neutrinos and antineutrinos, ν_μ and ν_e, and protons and neutrons, which results in the eight combinations. Examples of these weights are shown in Fig. D.5. We use the nominal Asimov defined at the Pre-ND tuning for the reference fit.

D.2.6 Energy dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$

In the T2K analysis, the uncertainty on the ν_e ($\bar{\nu}_e$) cross-sections is known as one of the main uncertainties when searching for CP violation [6]. Since CP violation in the neutrino oscillation can be seen in the $\nu_\mu \to \nu_e$ and $\bar{\nu}_\mu \to \bar{\nu}_e$ appearance probabilities, the effects of these systematic uncertainties on the ν_e cross-sections are inevitable. However, a direct measurement of these ν_e cross-sections is extremely difficult as there is no dedicated electron neutrino beam. Some measurements have been performed using the intrinsic ν_e ($\bar{\nu}_e$) components in the ν_μ ($\bar{\nu}_\mu$) beam, such as MINERνA [205], T2K [206], and Micro-BooNE [207, 208]. These measurements still have large uncertainties due to the limited statistics and large background.

In this analysis, we use a ν_e ($\bar{\nu}_e$) cross-sections estimated from the ν_μ ($\bar{\nu}_\mu$) cross-sections with some theoretical corrections [209]. In the systematic model, two systematic uncertainty parameters are assigned separately for the $\sigma_{\nu_e}/\sigma_{\nu_\mu}$ and $\sigma_{\bar{\nu}_e}/\sigma_{\bar{\nu}_\mu}$ ratios. These parameters are applied to all the ν_e ($\bar{\nu}_e$) events in all the energy regions. In the joint analysis, we have more ν_e ($\bar{\nu}_e$) events compared to the T2K-standalone analysis. This implies that if these ν_e cross-section systematic uncertainty parameters are strongly pulled by the atmospheric samples, it would affect the constraints on δ_{CP} through the changes in the T2K sample event spectra.

(a) Cross-section weight

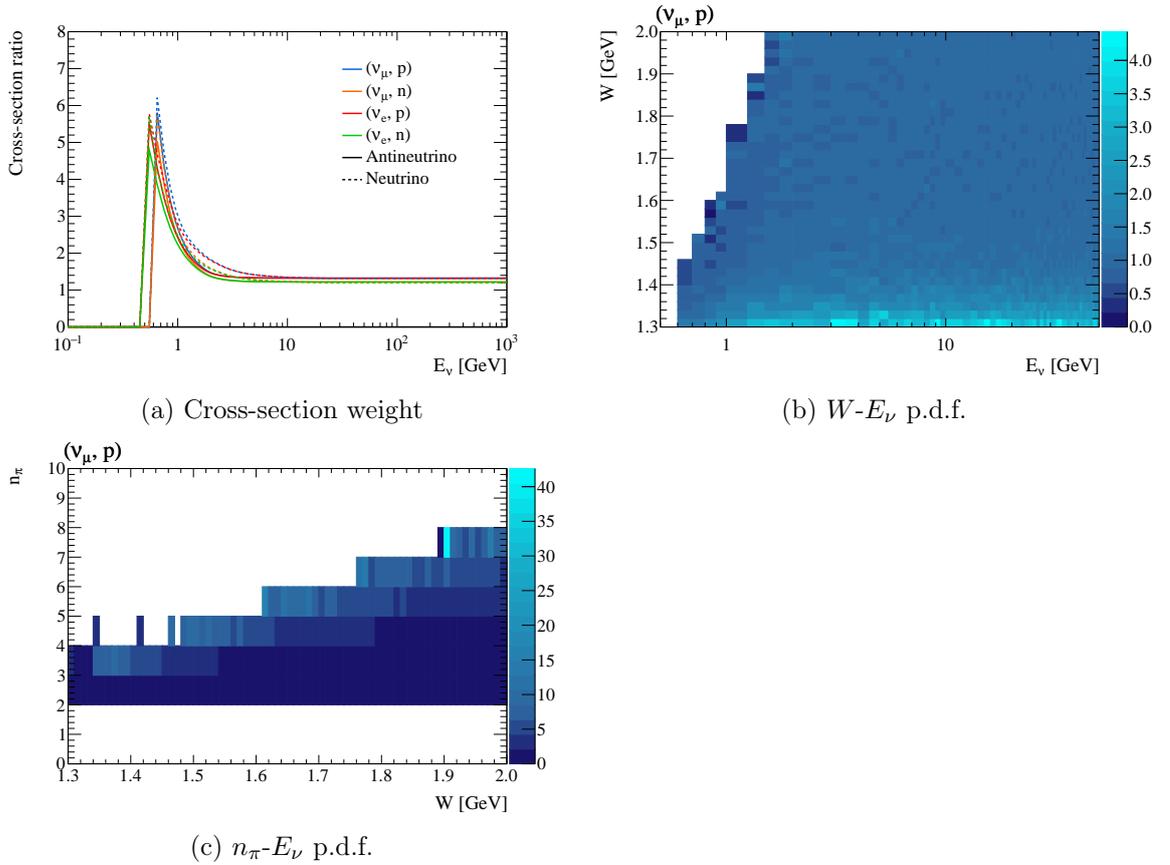(b) W-E_ν p.d.f.

(c) n_π-E_ν p.d.f.

Figure D.5. Examples of the three sets of weights used for the construction of alternative pion multiplicity fake data set. The cross-section weights are shown for all combinations of the neutrino and target types. The probability density functions for W-E_ν and n_π-E_ν are shown only for (ν_μ, p).

The results of the likelihood scan[1] for the $\sigma_{\nu_e}/\sigma_{\nu_\mu}$ and $\sigma_{\bar\nu_e}/\sigma_{\bar\nu_\mu}$ systematic uncertainty parameters are shown in Fig. D.6. Here we fix the oscillation parameters at the Asimov A parameter set and the systematic uncertainty parameters at the Post-ND tuning and vary the ν_e cross-section systematics only. It illustrates that these parameters are strongly constrained by the atmospheric sub-GeV and multi-GeV e-like samples than the T2K e-like samples. Since the T2K beam and SK atmospheric sub-GeV samples have similar energy ranges as shown in Fig. 4.15, these samples should affect ν_e cross-section systematics in the same way. However, using the same parameters for the high-energy samples can give unexpected constraints on these parameters if there is an energy-dependent effect in the ν_e cross-sections.

This fake data study is therefore aiming to test a possible bias due to using a common set of ν_e ($\bar\nu_e$) cross-section systematic uncertainty parameters in both the low-energy and high-energy samples. The fake data set is constructed by weighting the ν_e ($\bar\nu_e$) events that have true neutrino energies of $E_\nu > 1.7$ GeV with $+2.83\%$, which is the size of $+1\sigma$ uncertainty for the $\sigma_{\nu_e}/\sigma_{\nu_\mu}$ ($\sigma_{\bar\nu_e}/\sigma_{\bar\nu_\mu}$) systematic uncertainty parameters. The ν_e ($\bar\nu_e$)

---

[1]The likelihood scan is described in Section 5.5.2.3.

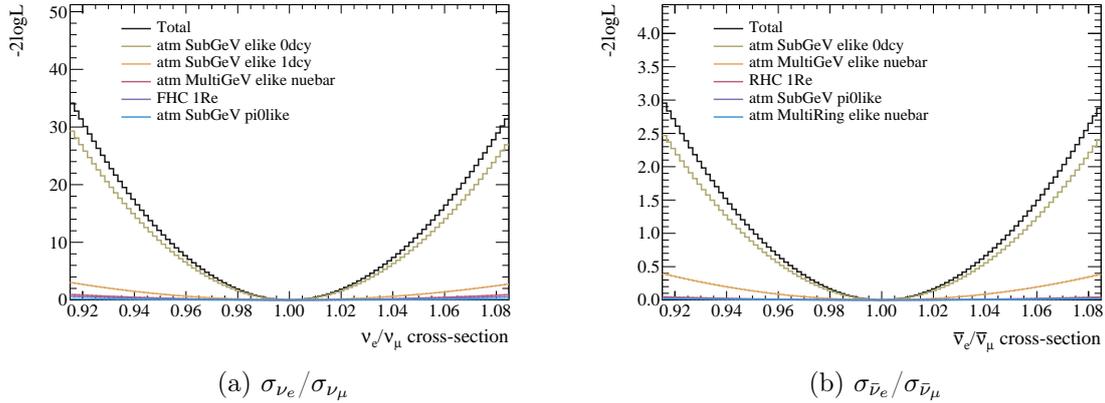(a) $\sigma_{\nu_e}/\sigma_{\nu_\mu}$ (b) $\sigma_{\bar{\nu}_e}/\sigma_{\bar{\nu}_\mu}$

Figure D.6. Sample likelihood distributions when varying the $\sigma_{\nu_e}/\sigma_{\nu_\mu}$ ($\sigma_{\bar{\nu}_e}/\sigma_{\bar{\nu}_\mu}$) systematic uncertainty parameters within $[-3\sigma, 3\sigma]$ ranges ($\sigma = 0.0283$). While one parameter is varied, the other oscillation parameters and the systematic uncertainty parameters are fixed at the Asimov A parameter set and the Post-ND tuning. Only the total likelihood and likelihood of the five samples with the largest contributions are shown for each parameter.

events below this energy are left untouched. The choice of $E_\nu = 1.7$ GeV was made to roughly separate the energy spectra of low-energy and high-energy samples in Fig. 4.15. We use the nominal Asimov defined at the Pre-ND tuning for the reference fit.

D.2.7   ND data-driven non-QE CC0π

As discussed in Section D.2.3, the ND cross-section measurements for the CC0π samples suggest that the MC underpredicts the data by approximately 10%. In our CCQE cross-section model, we assign some freedom to account for this difference mainly through the Q^2-dependent normalization parameters. However, the CC0π sample could also have contributions from the CC non-QE components such as 2p2h and pion absorption (FSI). It could be possible that the Q^2 parameters assigned to the CCQE cross-section are covering the inadequacy in the non-QE modeling and therefore the predictions for the non-QE contribution differ from the "true" nature behind the data. We test whether it could affect our oscillation analysis through a fake data study.

Since this test is mainly motivated by the data/MC discrepancies observed in the ND measurements, we define a fake data set in a data-driven way using the FGD1 CC0π data. The fake data set is generated by pushing all the data/MC inconsistencies in the ND CC0π sample to the non-QE components. This is done by fixing the Q^2 parameters to 1, which is the nominal value in SF, and applying the scaling extracted from the data/MC difference to the non-QE components.

Figure D.7 shows the data and MC comparison in the ND FGD1 CC0π sample where the Q^2 parameters are set to 1.0 in the MC. The MC (red solid line) overpredicts the data in the low-Q^2 region and underpredicts the data in the higher Q^2 region. In the fake data generations, these differences are assumed to be coming from the non-QE components, and a modification is applied to the non-QE events so that the total prediction matches
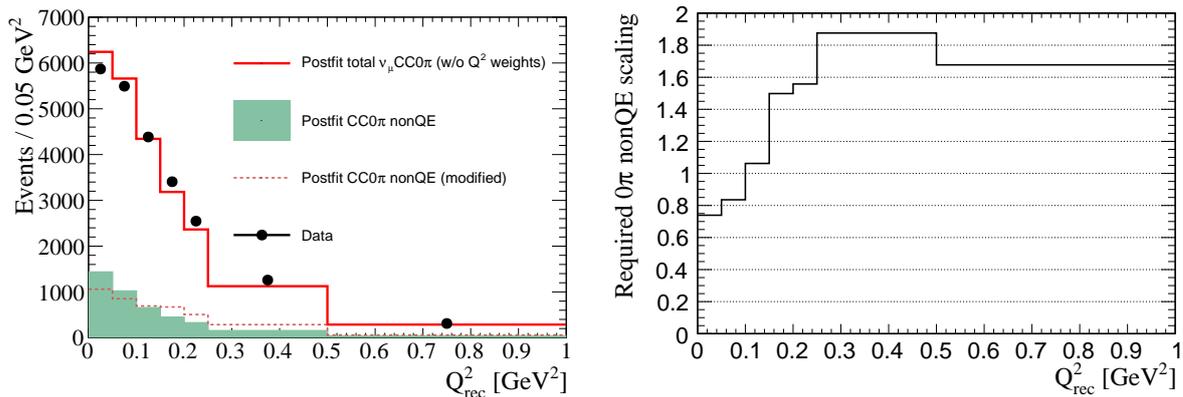
Figure D.7. The ND data and MC predictions in the FGD1 ν_μ CC0π sample as a function of the reconstructed momentum transfer $Q^2_{\rm rec}$. The red solid line and green shaded histogram show the post-fit spectra of the total and CC0π non-QE events with the Q^2 parameters set to 1.0. The red dotted line shows the modified CC0π non-QE event spectra so that the total prediction matches with the data. The right plot shows the obtained scaling of the CC0π non-QE events as a function of $Q_{\rm rec}$ extracted from the comparison shown in the left plot. The left figure is taken from Ref. [37].

with the data. The green histogram and red dotted line show the non-QE components before and after the modification. The obtained scaling is shown in the right plot in Fig. D.7.

The fake data set is generated by resetting all the Q^2 parameters to 1 and applying the obtained scaling to the true CC0π non-QE events as a function of the true momentum transfer under the QE hypothesis ($Q^2_{\rm QE}$). $Q^2_{\rm QE}$ is calculated as

$$Q^2_{\rm QE} = 2E_\nu(E_\mu - p_\mu \cos\theta_\mu) - M^2_\mu, \tag{D.5}$$

where E_μ, p_μ, θ_μ, and M_μ are the outgoing muon's energy, momentum, direction with respect to the incoming neutrino direction, and mass, respectively. The reconstructed neutrino energy is calculated using Eq. (4.10). We use the nominal Asimov defined at the Pre-ND tuning for the reference fit.

D.2.8 ND data-driven CC1π

The CC1π samples in the T2K beam and SK atmospheric analyses are built by selecting the events with one prompt lepton signal and one delayed signal from a Michel electron produced in the pion decay[2]. When a pion has momentum above the Cherenkov threshold, it can be reconstructed as a ring and classified into the multi-ring samples rather than the CC1π samples. Therefore, these CC1π samples are targeting the events with pion momentum below the Cherenkov threshold. The distribution of true pion momentum in these samples is shown in Fig. D.8.

---

[2]Since muons can also create a delayed signal, two or more delayed signals are required for the atmospheric μ-like CC1π sample. We do not have a dedicated μ-like CC1π sample for the T2K beam neutrinos in this analysis

On the other hand, in the T2K ND selection, there are multiple ways to identify pions such as the TPC PID, FGD PID, and a Michel electron signal. This may give different selection efficiencies as a function of the pion momentum compared to the selection at SK. In addition, the current ND fit does not have systematic uncertainties on the pion kinematics to constrain the CC1π model. We added the Adler angle systematic uncertainties in the SK fit, but they are not used in the ND fit. Therefore, it could be possible that the absence of the pion kinematics systematic uncertainties in the ND fit causes an unexpected bias when the ND constraints are extrapolated to the far detector.

As shown in Fig. D.9, a sizable discrepancy has been observed in the pion momentum distributions between the data and MC in the ND CC1π sample. Here the data is underpredicted by 23% for the events with a pion momentum $p_\pi^{rec} < 200$ MeV, where the chosen range roughly corresponds to the Michel electron tagging efficiency turn-on as shown in Fig. D.8.

We perform a fake data study to test whether this discrepancy in the ND CC1π sample could affect our oscillation analysis. Similarly to the CC0π non-QE fake data study described in Section D.2.7, this test is also motivated by the observed data/MC discrepancy and the fake data set is built in a data-driven way. The fake data set is generated simply by applying an overall normalization of $+23\%$ to the true CC1π events with $p_\pi^{rec} < 200$ MeV, which is then fit with our nominal systematic model. We use the nominal Asimov defined at the Pre-ND tuning for the reference fit.
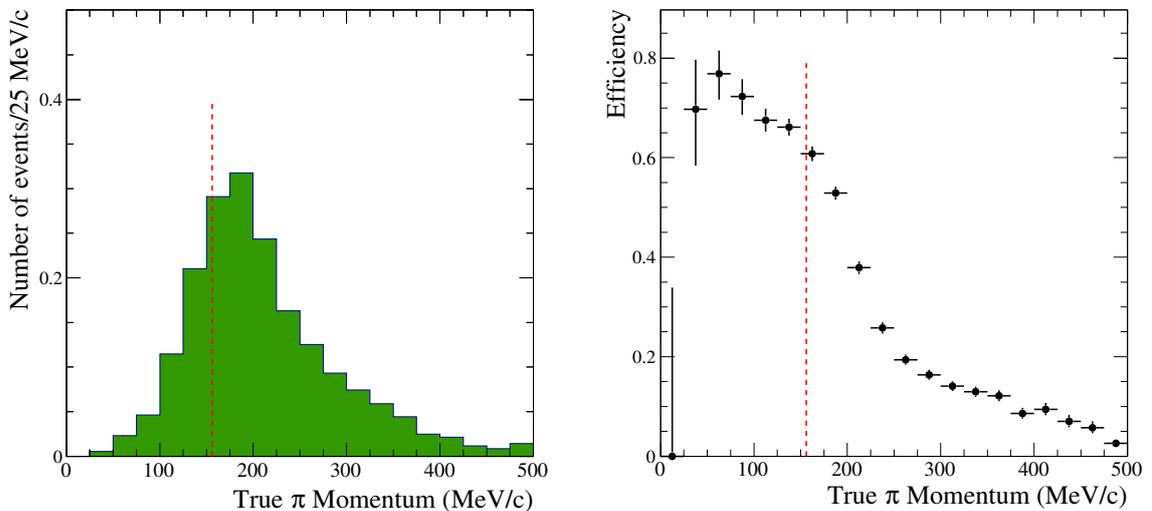

Figure D.8. The true pion momentum distribution for selected simulated signal events in the T2K CC1π^+ sample (left) and the selection efficiency of these events (right). The red dashed line indicates the Cherenkov threshold for charged pions from Table 3.1. The figures are taken from Ref. [39].
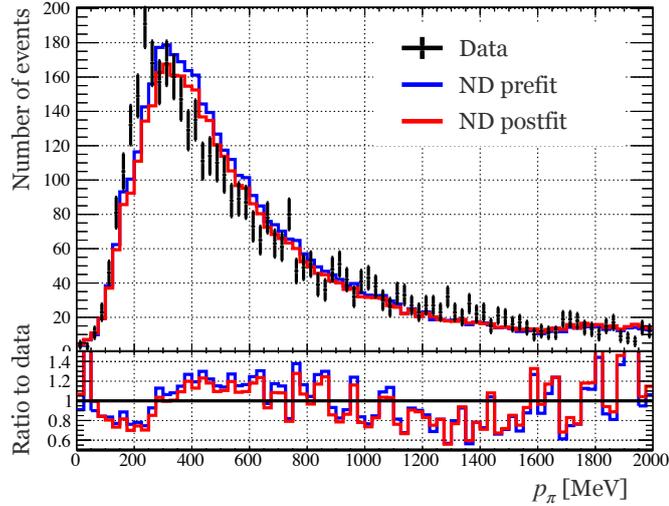
Figure D.9. The observed data and MC comparison in the ND FGD1 CC1π sample as a function of the reconstructed pion momentum where pions are identified using the TPC PID. The bottom panel shows the ratio of the prefit and postfit MC to data.

D.2.9   Pion SI bug fix

After the T2K part of the analysis was completed, two bugs were found in the input MC. One bug was that the `NEUT` cascade model was not properly applied in the MC production used for the ND fit, resulting in an incorrect cross-section for pion secondary interactions (SI), especially above $p_\pi > 1$ GeV. The second bug was that the radius of the oxygen nucleus was used in the calculation of FSI for carbon as well. It means that hadrons produced inside the carbon are propagated longer distances than expected and get more FSI effects. Although the FSI parameters on carbon are not propagated to SK as it is a water target, this bug can also have some impacts on the oxygen FSI parameters through the correlation. The second bug is expected to have a small contribution compared to the pion SI bug.

Both the T2K beam and SK atmospheric MC at SK are not affected by these bugs, but it could be possible that pion SI systematics get incorrect constraints from the ND fit and have a bias in the oscillation parameter measurements at SK. We therefore perform a fake data study to test whether this is the case. These bugs were fixed by reweighting the events in the ND MC and repeating the ND fit. In this fake data study, the SK data set is not modified. Instead, we simply use the covariance matrix from the fixed ND fit to fit the SK nominal Asimov data set and compare the results with the nomianl fit. We use the nominal Asimov defined at the Pre-ND tuning for the reference fit.

D.3   Result of the model robustness test

In the following, comparisons of the oscillation parameter $\Delta\chi^2$ distributions are shown for the robustness tests that are not presented in Chapter 6. The summary of the results

can be found in Section 6.3.4.

Nominal

The results of the nominal Asimov fit for different systematic uncertainty parameter tunings are summarized in Table D.3, Table D.4, and Table D.5. The corresponding plots are shown in Fig. 6.4.

Table D.3. Summary of the intervals for the nominal Asimov fit at the Pre-ND tuning.

		δ_{CP}	Δm^2_{32}	$\sin^2\theta_{23}$
AsimovA	Middle of the 1σ interval	-1.622	0.002517	0.5241
	1σ interval size: $1\sigma_{\mathrm{tot.}}$	1.036	5.422e-05	0.04426
	1σ stat-only interval size: $1\sigma_{\mathrm{stat.}}$	0.8142	4.983e-05	0.03771
	$1\sigma_{\mathrm{syst.}} = \sqrt{(1\sigma_{\mathrm{tot.}})^2 - (1\sigma_{\mathrm{stat.}})^2}$	0.6403	2.138e-05	0.02318
	Middle of the 2σ interval	-1.639	0.002517	0.5185
	2σ interval size: $2\sigma_{\mathrm{tot.}}$	1.773	0.0001085	0.07194
	2σ stat-only interval size: $2\sigma_{\mathrm{stat.}}$	1.48	9.953e-05	0.06269
	$2\sigma_{\mathrm{syst.}} = \sqrt{(2\sigma_{\mathrm{tot.}})^2 - (2\sigma_{\mathrm{stat.}})^2}$	0.9749	4.33e-05	0.03528
AsimovB	Middle of the 1σ interval	0.1493	0.002515	0.4527
	1σ interval size: $1\sigma_{\mathrm{tot.}}$	0.7437	5.819e-05	0.02398
	1σ stat-only interval size: $1\sigma_{\mathrm{stat.}}$	0.6086	5.334e-05	0.01996
	$1\sigma_{\mathrm{syst.}} = \sqrt{(1\sigma_{\mathrm{tot.}})^2 - (1\sigma_{\mathrm{stat.}})^2}$	0.4275	2.325e-05	0.01329
	Middle of the 2σ interval	-	0.002515	0.5047
	2σ interval size: $2\sigma_{\mathrm{tot.}}$	-	0.0001165	0.09306
	2σ stat-only interval size: $2\sigma_{\mathrm{stat.}}$	-	0.0001069	0.07996
	$2\sigma_{\mathrm{syst.}} = \sqrt{(2\sigma_{\mathrm{tot.}})^2 - (2\sigma_{\mathrm{stat.}})^2}$	-	4.637e-05	0.04761

Table D.4. Summary of the intervals for the nominal Asimov fit at the Pre-ND ($Q^2 = 1$) tuning.

		δ_{CP}	Δm^2_{32}	$\sin^2 \theta_{23}$
AsimovA	Middle of the 1σ interval	-1.638	0.002516	0.5246
	1σ interval size: $1\sigma_{\mathrm{tot.}}$	0.9976	4.979e-05	0.04391
	1σ stat-only interval size: $1\sigma_{\mathrm{stat.}}$	0.7897	4.569e-05	0.03608
	$1\sigma_{\mathrm{syst.}} = \sqrt{(1\sigma_{\mathrm{tot.}})^2 - (1\sigma_{\mathrm{stat.}})^2}$	0.6095	1.98e-05	0.02502
	Middle of the 2σ interval	-1.65	0.002516	0.5195
	2σ interval size: $2\sigma_{\mathrm{tot.}}$	1.705	9.972e-05	0.07033
	2σ stat-only interval size: $2\sigma_{\mathrm{stat.}}$	1.429	9.137e-05	0.06035
	$2\sigma_{\mathrm{syst.}} = \sqrt{(2\sigma_{\mathrm{tot.}})^2 - (2\sigma_{\mathrm{stat.}})^2}$	0.9302	3.995e-05	0.0361
AsimovB	Middle of the 1σ interval	0.1503	0.002515	0.4515
	1σ interval size: $1\sigma_{\mathrm{tot.}}$	0.7133	5.365e-05	0.02201
	1σ stat-only interval size: $1\sigma_{\mathrm{stat.}}$	0.5577	4.868e-05	0.01858
	$1\sigma_{\mathrm{syst.}} = \sqrt{(1\sigma_{\mathrm{tot.}})^2 - (1\sigma_{\mathrm{stat.}})^2}$	0.4446	2.255e-05	0.0118
	Middle of the 2σ interval	-	0.002515	0.5044
	2σ interval size: $2\sigma_{\mathrm{tot.}}$	-	0.0001075	0.09121
	2σ stat-only interval size: $2\sigma_{\mathrm{stat.}}$	-	9.76e-05	0.04751 and 0.01616
	$2\sigma_{\mathrm{syst.}} = \sqrt{(2\sigma_{\mathrm{tot.}})^2 - (2\sigma_{\mathrm{stat.}})^2}$	-	4.496e-05	-

Table D.5. Summary of the intervals for the nominal Asimov fit at the Post-ND tuning.

		δ_{CP}	Δm^2_{32}	$\sin^2 \theta_{23}$
AsimovA	Middle of the 1σ interval	-1.627	0.002513	0.5231
	1σ interval size: $1\sigma_{\text{tot.}}$	0.9541	4.402e-05	0.0413
	1σ stat-only interval size: $1\sigma_{\text{stat.}}$	0.7224	4.023e-05	0.03422
	$1\sigma_{\text{syst.}} = \sqrt{(1\sigma_{\text{tot.}})^2 - (1\sigma_{\text{stat.}})^2}$	0.6233	1.787e-05	0.02314
	Middle of the 2σ interval	-1.639	0.002513	0.5183
	2σ interval size: $2\sigma_{\text{tot.}}$	1.599	8.811e-05	0.06558
	2σ stat-only interval size: $2\sigma_{\text{stat.}}$	1.331	8.048e-05	0.05746
	$2\sigma_{\text{syst.}} = \sqrt{(2\sigma_{\text{tot.}})^2 - (2\sigma_{\text{stat.}})^2}$	0.8868	3.586e-05	0.03162
AsimovB	Middle of the 1σ interval	0.1174	0.002515	0.4519
	1σ interval size: $1\sigma_{\text{tot.}}$	0.6372	4.686e-05	0.01965
	1σ stat-only interval size: $1\sigma_{\text{stat.}}$	0.5011	4.285e-05	0.01668
	$1\sigma_{\text{syst.}} = \sqrt{(1\sigma_{\text{tot.}})^2 - (1\sigma_{\text{stat.}})^2}$	0.3936	1.897e-05	0.0104
	Middle of the 2σ interval	1.241	0.002515	0.5042
	2σ interval size: $2\sigma_{\text{tot.}}$	2.59	9.37e-05	0.08704
	2σ stat-only interval size: $2\sigma_{\text{stat.}}$	1.569	8.585e-05	0.03846 and 0.007978
	$2\sigma_{\text{syst.}} = \sqrt{(2\sigma_{\text{tot.}})^2 - (2\sigma_{\text{stat.}})^2}$	2.06	3.753e-05	-

LFG
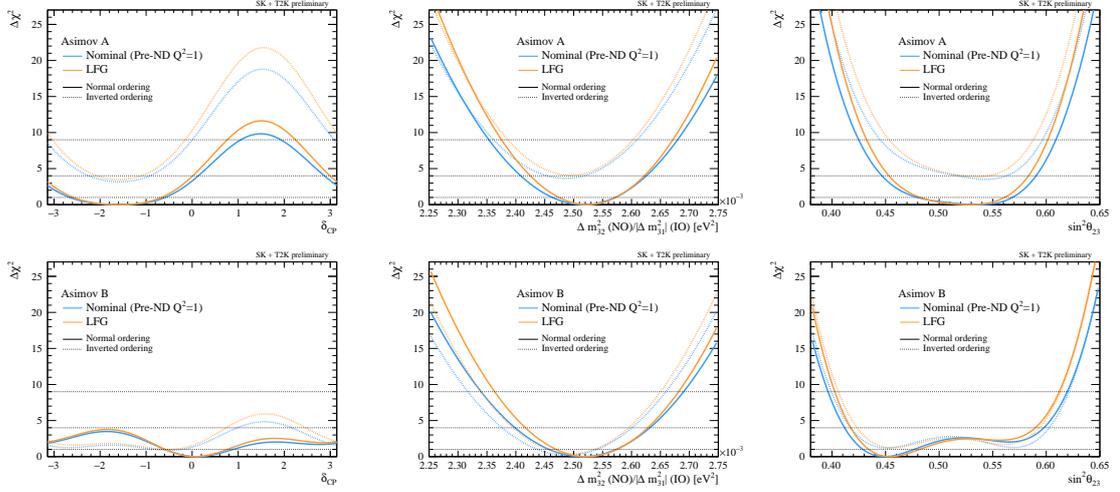
Figure D.10. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND ($Q^2 = 1$) and the LFG fake data fit. The top row shows the results at the oscillation parameter set Asimov A and the bottom row shows the results at Asimov B.

Removal energy
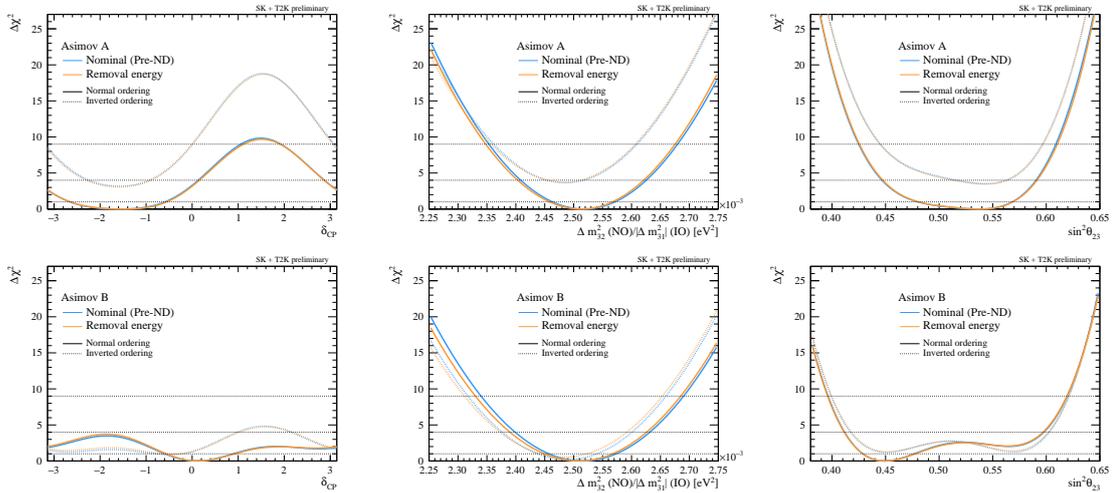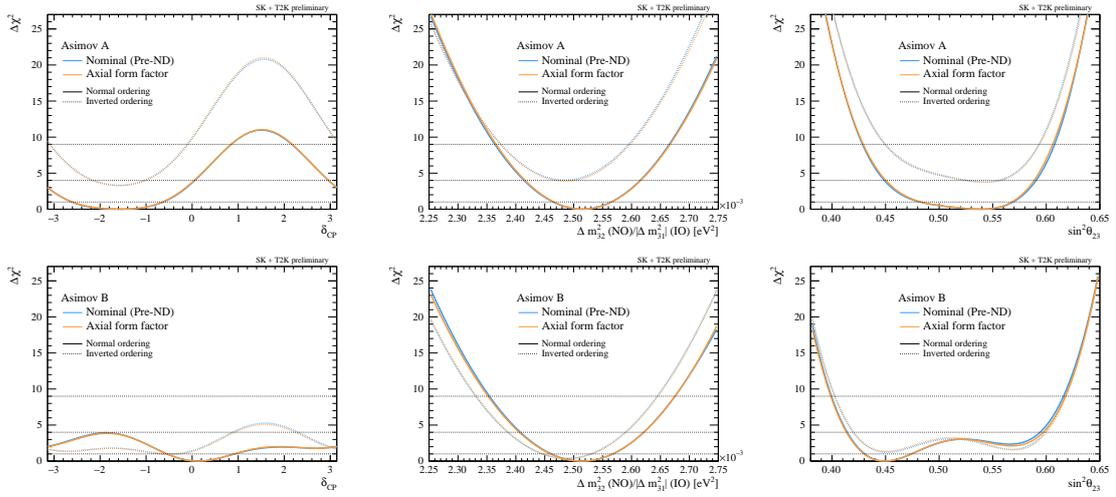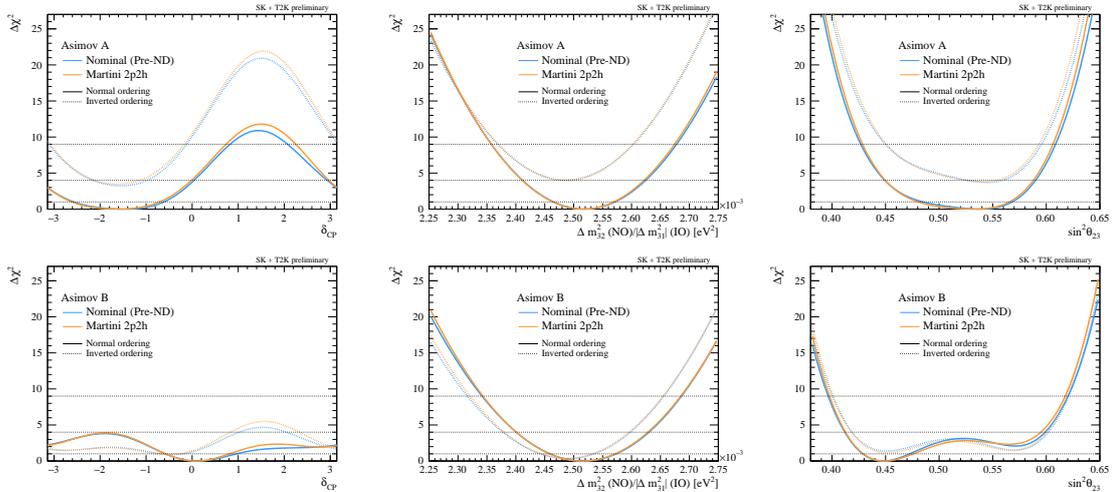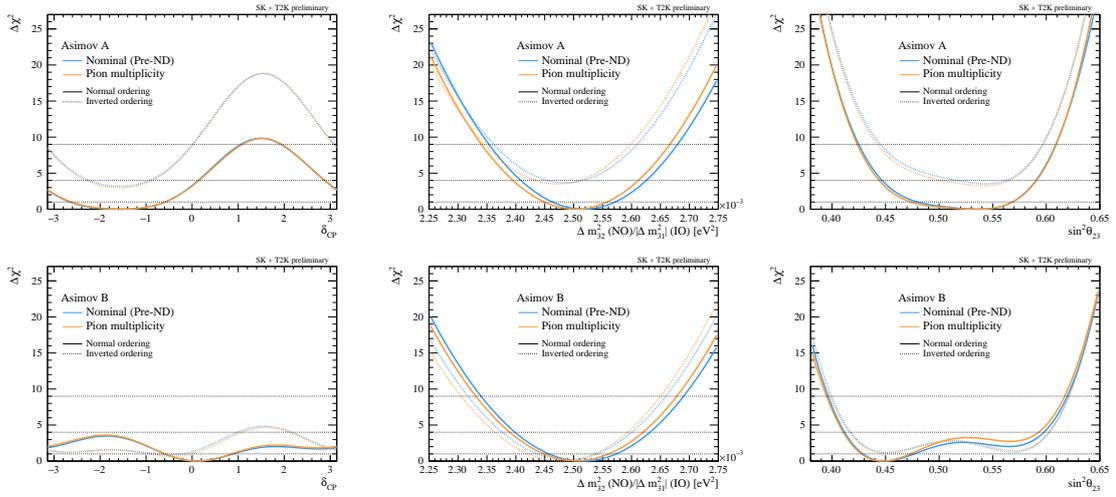

Figure D.11. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND and the removal energy fake data fit. The top row shows the results at the oscillation parameter set Asimov A and the bottom row shows the results at Asimov B.

Axial form factors

Figure D.12. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND and the axial form factor fake data fit. The top row shows the results at the oscillation parameter set Asimov A and the bottom row shows the results at Asimov B.
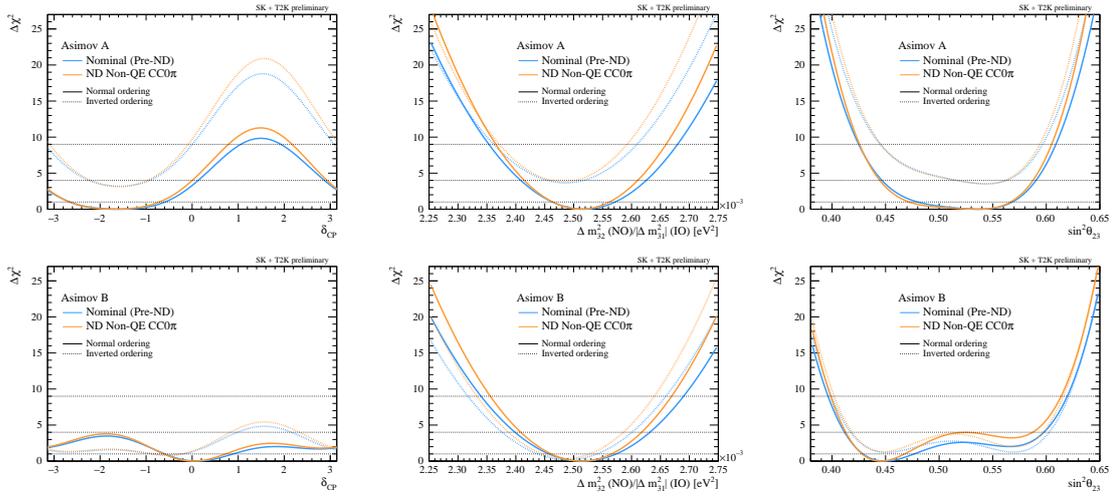
Martini 2p2h

Figure D.13. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND and the Martini 2p2h fake data fit. The top row shows the results at the oscillation parameter set Asimov A and the bottom row shows the results at Asimov B.
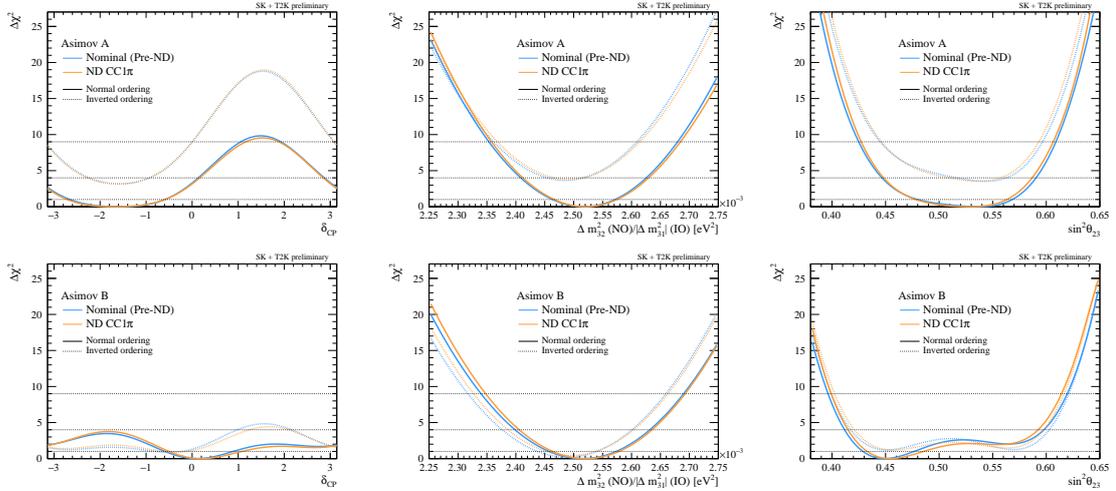
Pion multiplicity

Figure D.14. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND and the pion multiplicity fake data fit. The top row shows the results at the oscillation parameter set Asimov A and the bottom row shows the results at Asimov B.

Energy dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$
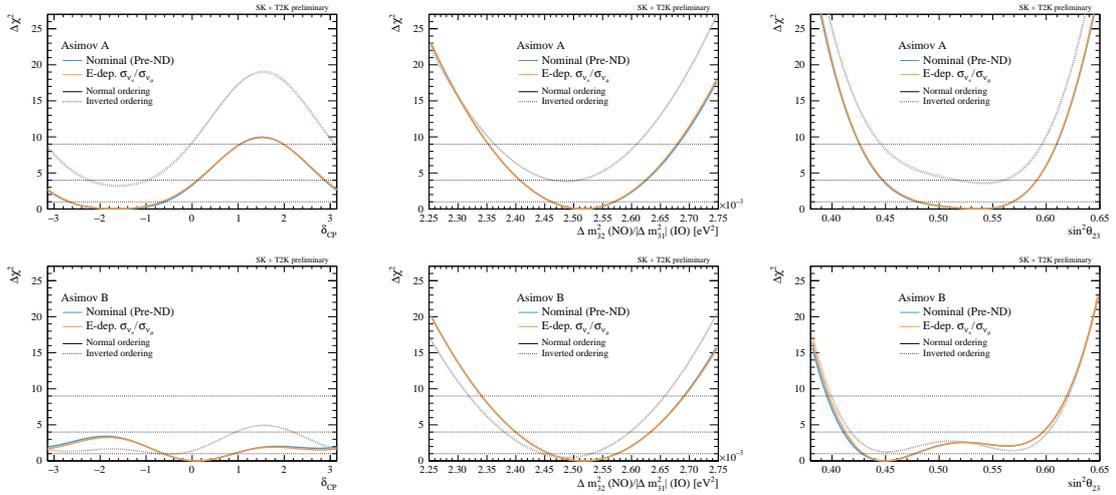
Figure D.15. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND and the energy dependent $\sigma_{\nu_e}/\sigma_{\nu_\mu}$ fake data fit. The top row shows the results at the oscillation parameter set Asimov A and the bottom row shows the results at Asimov B.
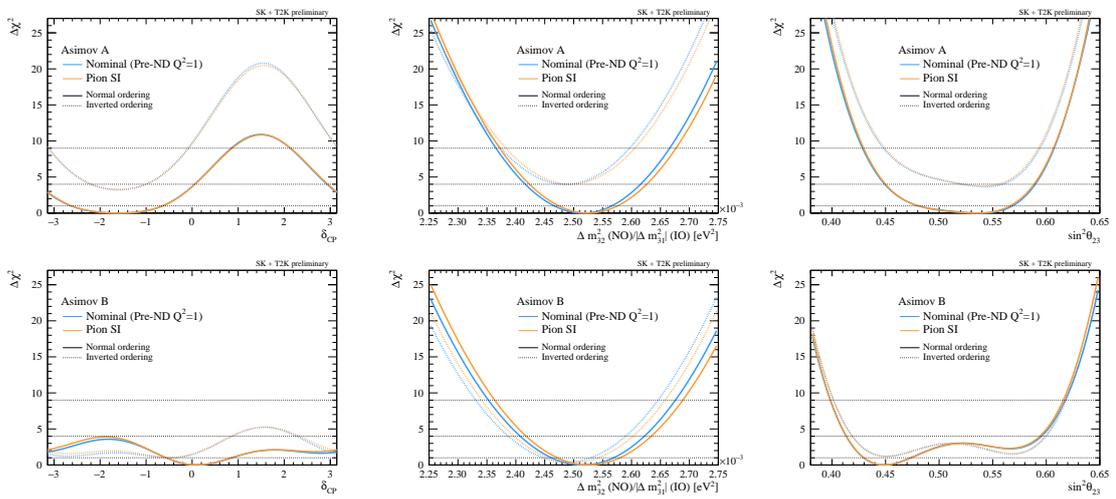
ND data-driven non-QE CC0π

Figure D.16. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND and the ND Non-QE CC0π fake data fit. The top row shows the results at the oscillation parameter set Asimov A and the bottom row shows the results at Asimov B.

ND data-driven CC1π

Figure D.17. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND and the ND CC1π fake data fit. The top row shows the results at the oscillation parameter set Asimov A and the bottom row shows the results at Asimov B.

Pion SI bug fix

Figure D.18. 1D $\Delta\chi^2$ distribution for the nominal fit at Pre-ND ($Q^2 = 1$) and the pion SI fake data fit. The top row shows the results at the oscillation parameter set Asimov A and the bottom row shows the results at Asimov B.

Appendix E

Two-dimensional event distributions

This appendix presents the two-dimensional event distributions for the samples with two-dimensional binning. The predicted event rates at the oscillation parameters and systematic uncertainty parameters at the SK+T2K best-fit values are also overlaid. The atmospheric sub-GeV e-like 1de, μ-like 2 de, and π^0-like samples have only the momentum binning, whereas the atmospheric UpMu through-going shower and non-shower samples have only the cosine zenith binning. Therefore, these samples are not shown in this appendix. The one-dimensional event distributions can be found in Section 7.1.
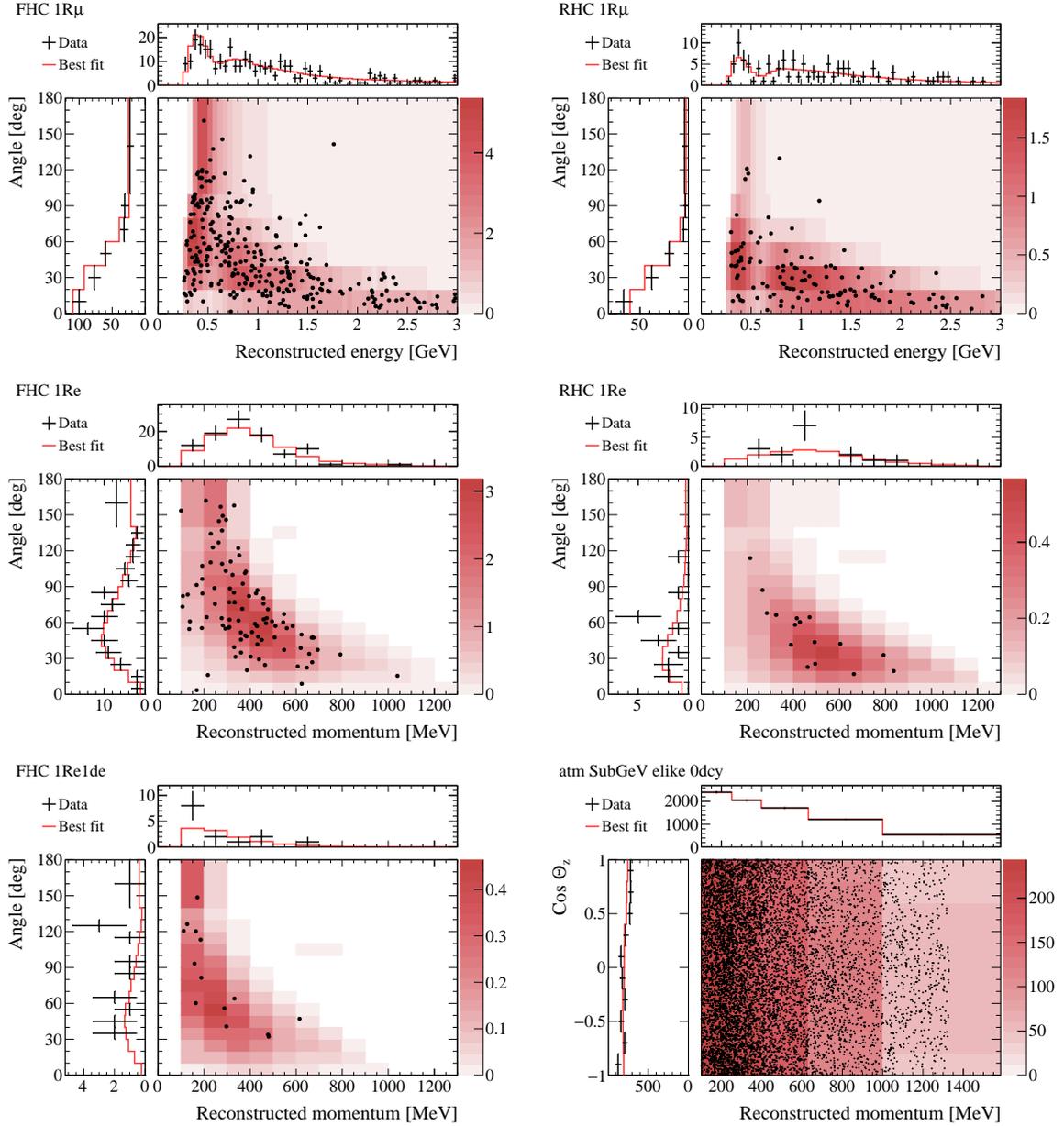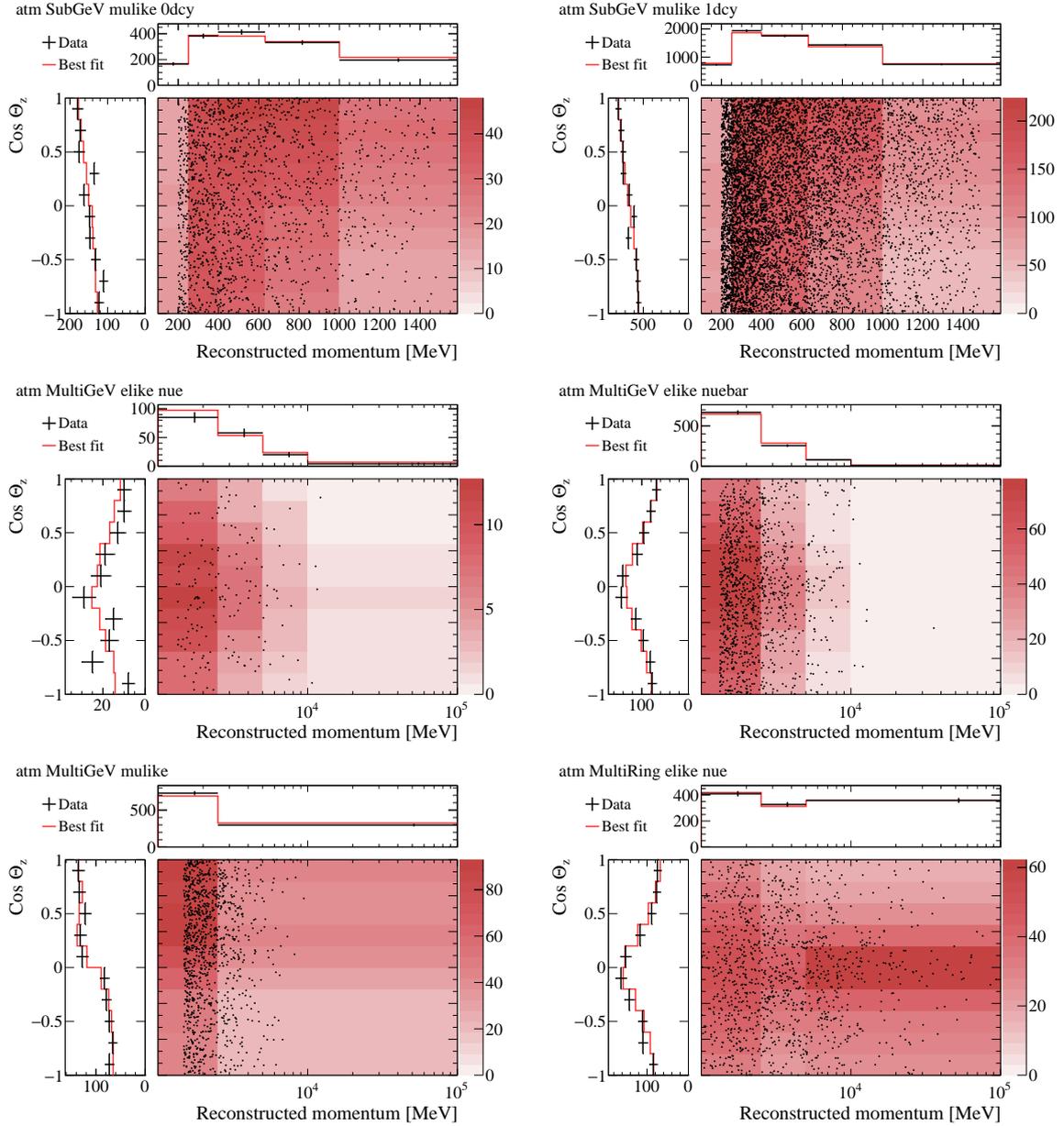
Figure E.1. The two-dimensional event distributions for the samples with two-dimensional binning. The colored background in the two-dimensional plot shows the expected number of events at the best-fit values for the oscillation and systematic uncertainty parameters. The top and left panels show the events projected onto each single dimension, and the red line is the expected number of events from the best fit.
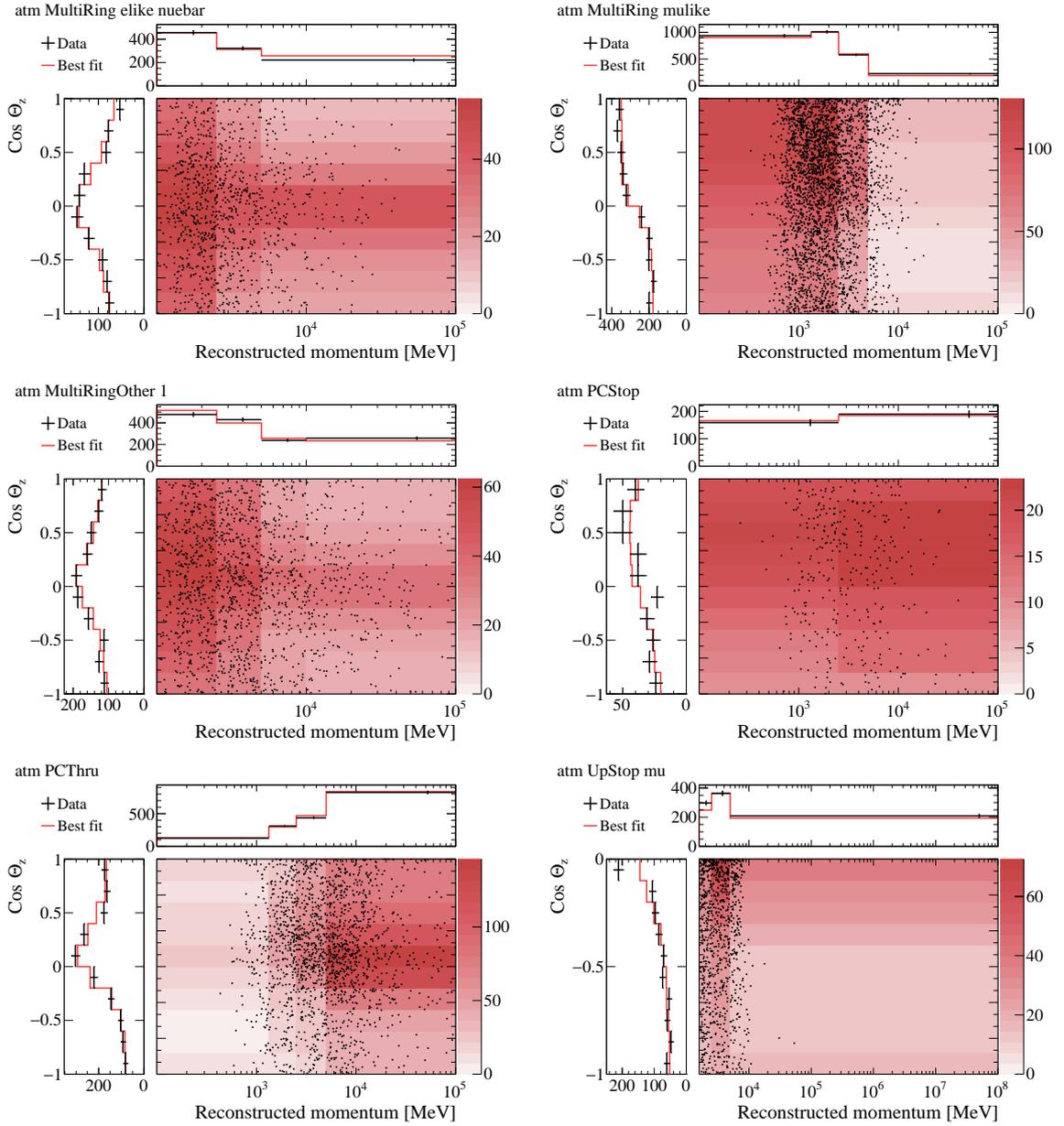
Figure E.1. The two-dimensional event distributions for the samples with two-dimensional binning. The colored background in the two-dimensional plot shows the expected number of events at the best-fit values for the oscillation and systematic uncertainty parameters. The top and left panels show the events projected onto each single dimension, and the red line is the expected number of events from the best fit.

Figure E.1. The two-dimensional event distributions for the samples with two-dimensional binning. The colored background in the two-dimensional plot shows the expected number of events at the best-fit values for the oscillation and systematic uncertainty parameters. The top and left panels show the events projected onto each single dimension, and the red line is the expected number of events from the best fit. The background subtraction (Section A.1) is not applied to the UpMu samples yet.

Appendix F

Comparison to the published results from the individual experiments

As an additional validation, the $\Delta\chi^2$ distributions obtained from the T2K-only fit and SK atmospheric-only fit within the joint analysis framework are compared with the published results from individual experiments. These studies are performed only for validation purposes, and we will not modify our main analysis results based on these studies. Moreover, since many changes have been made to the joint analysis, we do not expect a perfect agreement with the published results. When there is a difference, we try to see whether we can reproduce the results of individual analyses by reverting the changes we made for the joint analysis. We use the importance sampling method for the comparison to T2K and the profiling method for the comparison to SK because these are the methods used in each published result.

F.1 Comparison with the published T2K analysis

The T2K part of the joint analysis is mostly based on the T2K oscillation analysis reported in Ref. [37] (denoted as "published" T2K analysis hereafter). However, we have made a few changes in the analysis model to adapt it to the joint fit. The main changes are the following:

1. Use the detector systematic uncertainties re-evaluated with correlations between the T2K beam and SK atmospheric samples, which changes not only the correlation but also the size of each detector systematic error on the T2K part (Section 4.4).

2. Add additional systematic uncertainty parameters including the Adler angle systematic uncertainties and low-momentum PID systematic uncertainties to cover the data/MC excess observed in the atmospheric sub-GeV CC1π sample down-going events (Section 4.3, Section 4.4.2.3).

3. Add a new decay electron and neutron separation in the SK event selection to reduce the neutron backgrounds which produces delayed signals and are misidentified as decay electrons.

4. Update the definition of the kinematic cut in the SK event selection. It was applied before shifting the kinematics (e.g. Coulomb correction in Section 4.3.1.3) in the published T2K analysis, but in the joint analysis, the cut is applied after the modifications on the kinematics.

The comparison of the joint analysis T2K-only fit and the published T2K standalone analysis is shown in Fig. F.1. In the joint analysis, we test several options to investigate the most effective change in the oscillation analysis. The red line shows the result of the T2K-only fit with our nominal joint analysis model. The blue, green, orange, and purple lines also show the joint analysis T2K-only fit, but using the original T2K detector systematic uncertainties (before the revaluation of correlations), turning off the Adler angle and low-momentum PID systematic uncertainties, using the data and MC without the new neutron cut, and applying the original kinematic cut definitions, respectively. The T2K published result is shown with the gray line.

For δ_{CP}, the joint analysis T2K-only fit shows a slightly weaker constraint but is mostly consistent with the published T2K analysis. For Δm_{32}^2, the best-fit point shifts to a larger value but the size of the intervals stays roughly the same. Among the options we tested, the new neutron cut and the new kinematic cut especially have the largest contributions to the shift in the Δm_{32}^2 best-fit point. For $\sin^2 \theta_{23}$, the best-fit point shifts to a larger value, and the preference for the maximal mixing becomes slightly weaker in the joint analysis. The size of the confidence intervals is also slightly larger compared to the published T2K analysis.

Fig. F.2 shows the comparison between the published T2K analysis and the joint analysis T2K-only fit where all the changes tested in Fig. F.1 are simultaneously reverted to the original. It reproduces the published T2K results almost perfectly for $\sin^2 \theta_{23}$ and Δm_{32}^2. Therefore, we conclude that our analysis framework is fully validated against the published T2K standalone analysis.
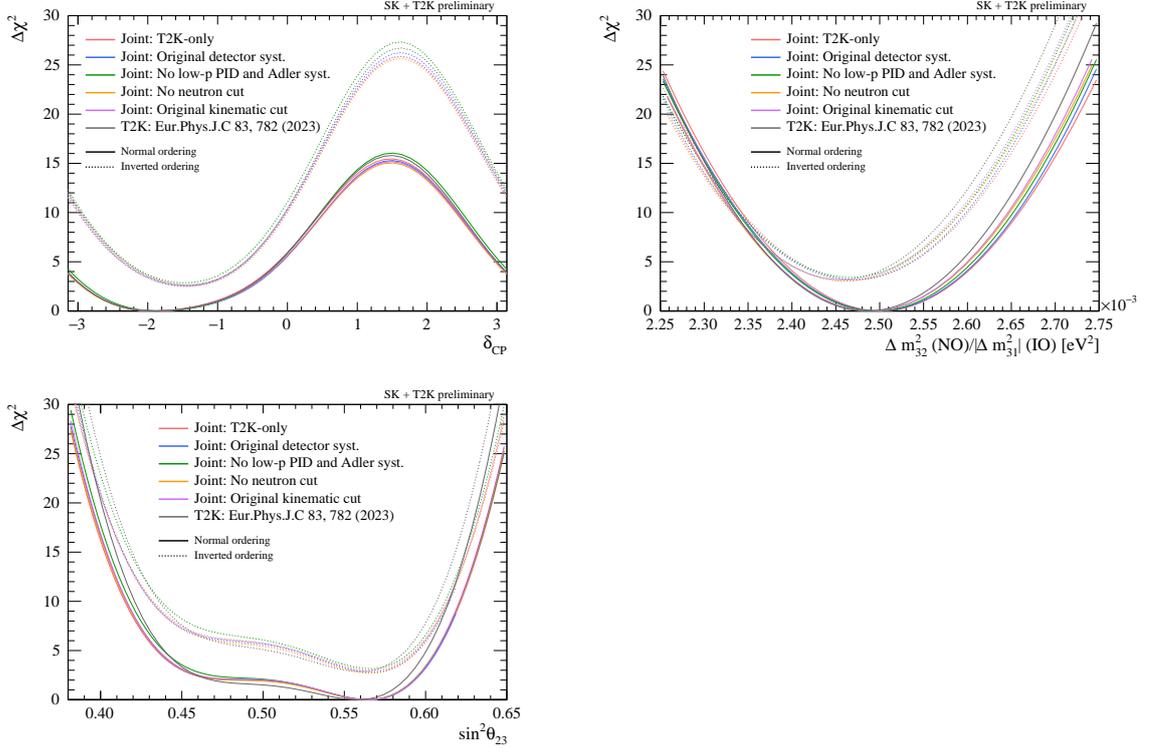
Figure F.1. Comparison of the T2K published analysis [37] and the joint analysis T2K-only fit. Several options are tested in the joint analysis to test which change has the largest impact, as explained in the text and legend. The Δm^2_{32} smearing is not applied in all the cases.
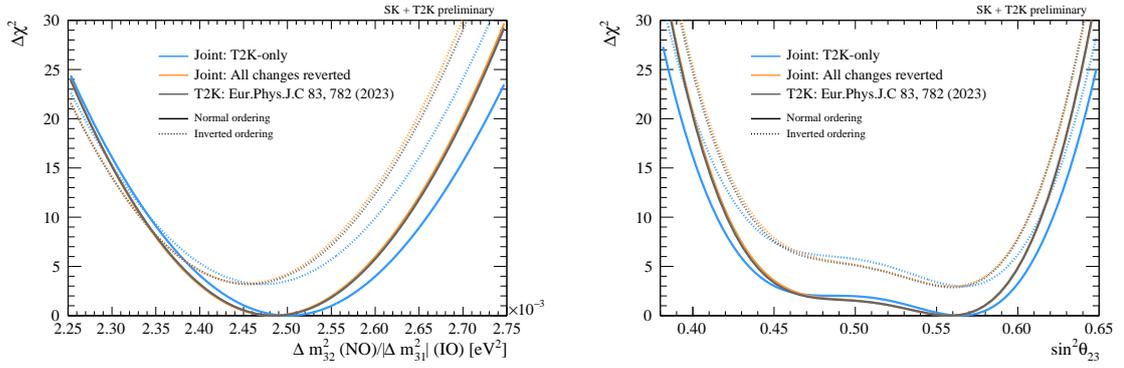

Figure F.2. Comparison of the T2K published analysis [37] and joint analysis T2K-only fit where all the changes tested in Fig. F.1 are reverted simultaneously. The Δm^2_{32} smearing is not applied in all the cases.

F.2   Comparison with the published SK analysis

The SK atmospheric part of the joint analysis is mostly based on the SK standalone analysis reported in Ref. [36], but there have been many updates in every aspect of the analysis. The differences between the joint analysis and the published SK analysis are

summarized in Table F.1.

Table F.1. Summary of the changes made for the joint analysis compared to the published SK analysis.

Analysis		SK published analysis [36]	Joint analysis
Livetime (day)		3118	3244
fiTQun version		v6	v4
Cross-section	CCQE	LFG	SF
	2p2h	Hadron tensor	Lookup table
	DIS	Old low-W model	New low-W model
Detector			Re-evaluated
APFit for PC/UpMu			Improved gain correction and ID/OD crosstalk
New Michel-e/neutron cut			Used
T2K ND			Used

Considering the changes made in joint analysis, these two analyses are very different and we do not expect the same result from both analyses. Therefore, the following studies are performed to understand the effect of each change, and we do not aim to fully reproduce the results of the published SK analysis.

The comparison of the joint analysis SK atmospheric-only fit and SK published result is shown in Fig. F.3. In the joint analysis, we test several options to investigate which change has the largest impact on the oscillation analysis results. The red line shows the result of our nominal joint analysis model where we use the low-energy cross-section model for the atmospheric sub-GeV samples with the T2K near detector constraints. The green, orange, and blue lines show the joint analysis result where we do not use the near detector constraints for the low-energy model, use the high-energy model for all the atmospheric samples, and turn off the Adler angle and low-momentum PID systematic uncertainties, respectively. The gray line shows the result of the published SK analysis. In this comparison, we use the profiling method where all the nuisance parameters are profiled instead of being marginalized. This is because the published SK analysis also profiles the nuisance parameters.

For $\delta_{\rm CP}$, the joint analysis atmospheric-only fit has a stronger constraint in normal ordering. It also shows a weaker constraint in the upper octant of $\sin^2\theta_{23}$ and for the larger values of Δm_{32}^2. Among the several options we tested, the closest result to the SK published result is obtained when we use the high-energy model for all the atmospheric samples (shown with the orange line). It shows a very good agreement in the $\delta_{\rm CP}$ $\Delta\chi^2$ distributions in normal ordering and a similar shape at the upper octant of $\sin^2\theta_{23}$. These differences could be not just because of the difference of the model itself, but also because

with a common model, sub-GeV samples (which have larger statistics) can constrain the cross-section systematic uncertainty parameters for the high-energy samples.
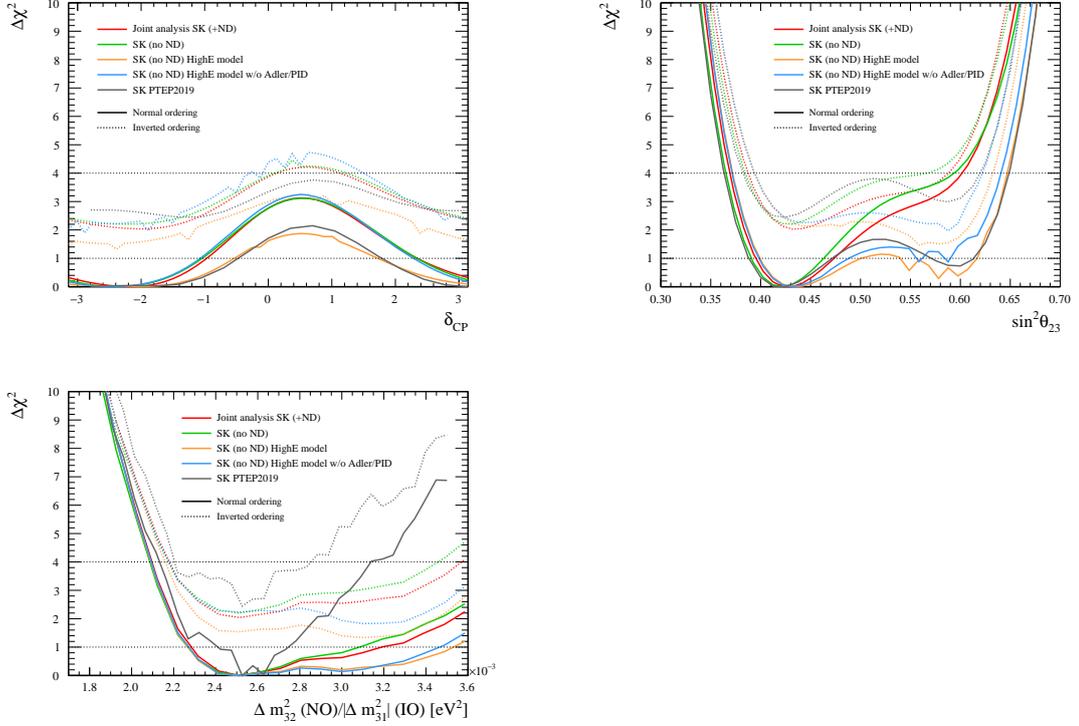

Figure F.3. Comparison of the SK published analysis [36] and the joint analysis SK-only fit. Several options are tested in the joint analysis to test which change has the largest impact, as explained in the test and legend. Here the nuisance parameters are profiled in the joint analysis SK fit to get more consistent analysis results with the SK official analysis.

Bibliography

[1] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998. (pages 9, 12, 15).

[2] Kazuo Abe et al. Observation of large CP violation in the neutral B meson system. *Phys. Rev. Lett.*, 87:091802, 2001. (page 9).

[3] Bernard Aubert et al. Observation of CP violation in the B^0 meson system. *Phys. Rev. Lett.*, 87:091801, 2001. (page 9).

[4] M. Fukugita and T. Yanagida. Baryogenesis Without Grand Unification. *Phys. Lett. B*, 174:45–47, 1986. (pages 9, 27).

[5] K. Abe et al. Observation of Electron Neutrino Appearance in a Muon Neutrino Beam. *Phys. Rev. Lett.*, 112:061802, 2014. (pages 10, 13).

[6] K. Abe et al. Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations. *Nature*, 580(7803):339–344, 2020. [Erratum: Nature 583, E16 (2020)]. (pages 10, 26, 218).

[7] J. Chadwick. The intensity distribution in the magnetic spectrum of beta particles from radium (B + C). *Verh. Phys. Gesell.*, 16:383–391, 1914. (page 11).

[8] W. Pauli. Dear radioactive ladies and gentlemen. *Phys. Today*, 31N9:27, 1978. (page 11).

[9] F. Reines and C. L. Cowan. Detection of the free neutrino. *Phys. Rev.*, 92:830–831, 1953. (page 11).

[10] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the free neutrino: A Confirmation. *Science*, 124:103–104, 1956. (page 11).

[11] G. Danby, J. M. Gaillard, Konstantin A. Goulianos, L. M. Lederman, Nari B. Mistry, M. Schwartz, and J. Steinberger. Observation of High-Energy Neutrino Reactions and the Existence of Two Kinds of Neutrinos. *Phys. Rev. Lett.*, 9:36–44, 1962. (page 11).

[12] Martin L. Perl et al. Evidence for Anomalous Lepton Production in e+ - e- Annihilation. *Phys. Rev. Lett.*, 35:1489–1492, 1975. (page 12).

[13] K. Kodama et al. Observation of tau neutrino interactions. *Phys. Lett. B*, 504:218–224, 2001. (page 12).

[14] S. Schael et al. Precision electroweak measurements on the Z resonance. *Phys. Rept.*, 427:257–454, 2006. (page 12).

[15] Raymond Davis, Jr., Don S. Harmer, and Kenneth C. Hoffman. Search for neutrinos from the sun. *Phys. Rev. Lett.*, 20:1205–1209, 1968. (page 12).

[16] R. Davis. Nobel Lecture: A half-century with solar neutrinos. *Rev. Mod. Phys.*, 75:985–994, 2003. (page 12).

[17] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022. (pages 12, 14, 15, 24, 25, 36, 178, 200).

[18] Aldo Serenelli. Alive and well: a short review about standard solar models. *Eur. Phys. J. A*, 52(4):78, 2016. (page 12).

[19] V. N. Gribov and B. Pontecorvo. Neutrino astronomy and lepton charge. *Phys. Lett. B*, 28:493, 1969. (page 12).

[20] L. Wolfenstein. Neutrino Oscillations in Matter. *Phys. Rev. D*, 17:2369–2374, 1978. (pages 12, 21).

[21] S. P. Mikheyev and A. Yu. Smirnov. Resonance Amplification of Oscillations in Matter and Spectroscopy of Solar Neutrinos. *Sov. J. Nucl. Phys.*, 42:913–917, 1985. (page 12).

[22] Q. R. Ahmad et al. Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87:071301, 2001. (page 13).

[23] Q. R. Ahmad et al. Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89:011301, 2002. (page 13).

[24] K. Eguchi et al. First results from KamLAND: Evidence for reactor anti-neutrino disappearance. *Phys. Rev. Lett.*, 90:021802, 2003. (page 13).

[25] K. Abe et al. Evidence for the Appearance of Atmospheric Tau Neutrinos in Super-Kamiokande. *Phys. Rev. Lett.*, 110(18):181802, 2013. (page 13).

[26] N. Agafonova et al. Discovery of τ Neutrino Appearance in the CNGS Neutrino Beam with the OPERA Experiment. *Phys. Rev. Lett.*, 115(12):121802, 2015. (page 13).

[27] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961. (page 13).

[28] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967. (page 13).

[29] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968. (page 13).

[30] M. Aker et al. Direct neutrino-mass measurement with sub-electronvolt sensitivity. *Nature Phys.*, 18(2):160–166, 2022. (page 16).

[31] Murray Gell-Mann, Pierre Ramond, and Richard Slansky. Complex Spinors and Unified Theories. *Conf. Proc. C*, 790927:315–321, 1979. (page 16).

[32] Osamu Sawada and Akio Sugamoto, editors. *Proceedings: Workshop on the Unified Theories and the Baryon Number in the Universe: Tsukuba, Japan, February 13-14, 1979*, Tsukuba, Japan, 1979. Natl.Lab.High Energy Phys. (page 16).

[33] Rabindra N. Mohapatra and Goran Senjanovic. Neutrino Mass and Spontaneous Parity Nonconservation. *Phys. Rev. Lett.*, 44:912, 1980. (page 16).

[34] Samoil M. Bilenky, J. Hosek, and S. T. Petcov. On Oscillations of Neutrinos with Dirac and Majorana Masses. *Phys. Lett. B*, 94:495–498, 1980. (page 16).

[35] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the unified model of elementary particles. *Prog. Theor. Phys.*, 28:870–880, 1962. (page 17).

[36] M. Jiang et al. Atmospheric Neutrino Oscillation Analysis with Improved Event Reconstruction in Super-Kamiokande IV. *PTEP*, 2019(5):053F01, 2019. (pages 19, 23, 28, 37, 44, 47, 56, 76, 149, 197, 238, 239, 240).

[37] K. Abe et al. Measurements of neutrino oscillation parameters from the T2K experiment using 3.6×10^{21} protons on target. *Eur. Phys. J. C*, 83(9):782, 2023. (pages 20, 23, 25, 28, 30, 40, 41, 48, 50, 51, 57, 58, 67, 96, 97, 98, 116, 193, 215, 221, 236, 238).

[38] C. Jarlskog. Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation. *Phys. Rev. Lett.*, 55:1039, 1985. (page 20).

[39] K. Abe et al. Measurement of neutrino and antineutrino oscillations by the T2K experiment including a new additional sample of ν_e interactions at the far detector. *Phys. Rev. D*, 96(9):092006, 2017. [Erratum: Phys.Rev.D 98, 019902 (2018)]. (pages 21, 222).

[40] Jiro Arafune, Masafumi Koike, and Joe Sato. CP violation and matter effect in long baseline neutrino oscillation experiments. *Phys. Rev. D*, 56:3093–3099, 1997. [Erratum: Phys.Rev.D 60, 119905 (1999)]. (page 22).

[41] Peter B. Denton, Megan Friend, Mark D. Messier, Hirohisa A. Tanaka, Sebastian Böser, João A. B. Coelho, Mathieu Perrin-Terrin, and Tom Stuttard. Snowmass Neutrino Frontier: NF01 Topical Group Report on Three-Flavor Neutrino Oscillations. 12 2022. (page 23).

[42] M. A. Acero et al. Improved measurement of neutrino oscillation parameters by the NOvA experiment. *Phys. Rev. D*, 106(3):032004, 2022. (pages 23, 25, 26).

[43] M. G. Aartsen et al. Measurement of Atmospheric Neutrino Oscillations at 6–56 GeV with IceCube DeepCore. *Phys. Rev. Lett.*, 120(7):071801, 2018. (pages 23, 25).

[44] K. Abe et al. Search for solar electron anti-neutrinos due to spin-flavor precession in the Sun with Super-Kamiokande-IV. *Astropart. Phys.*, 139:102702, 2022. (pages 23, 28).

[45] G. Alimonti et al. The Borexino detector at the Laboratori Nazionali del Gran Sasso. *Nucl. Instrum. Meth. A*, 600:568–593, 2009. (page 23).

[46] A. Gando et al. Reactor On-Off Antineutrino Measurement with KamLAND. *Phys. Rev. D*, 88(3):033001, 2013. (pages 23, 24).

[47] G. Bak et al. Measurement of Reactor Antineutrino Oscillation Amplitude and Frequency at RENO. *Phys. Rev. Lett.*, 121(20):201801, 2018. (pages 23, 24).

[48] C. D. Shin et al. Observation of reactor antineutrino disappearance using delayed neutron capture on hydrogen at RENO. *JHEP*, 04:029, 2020. (pages 23, 24).

[49] D. Adey et al. Measurement of the Electron Antineutrino Oscillation with 1958 Days of Operation at Daya Bay. *Phys. Rev. Lett.*, 121(24):241805, 2018. (pages 23, 24).

[50] H. de Kerret et al. Double Chooz θ_{13} measurement via total neutron capture detection. *Nature Phys.*, 16(5):558–564, 2020. (pages 23, 24).

[51] Fumihiko Suekane. *Neutrino Oscillations: A Practical Guide to Basics and Applications*, volume 898. Springer Tokyo, 2015. (page 24).

[52] K. Abe et al. Atmospheric neutrino oscillation analysis with external constraints in Super-Kamiokande I-IV. *Phys. Rev. D*, 97(7):072001, 2018. (pages 25, 28, 29, 40, 43, 69, 90, 149, 195).

[53] P. Adamson et al. Precision Constraints for Three-Flavor Neutrino Oscillations from the Full MINOS+ and MINOS Dataset. *Phys. Rev. Lett.*, 125(13):131802, 2020. (page 25).

[54] F. P. An et al. New measurement of θ_{13} via neutron capture on hydrogen at Daya Bay. *Phys. Rev. D*, 93(7):072011, 2016. (page 24).

[55] A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967. (page 26).

[56] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov. On the Anomalous Electroweak Baryon Number Nonconservation in the Early Universe. *Phys. Lett. B*, 155:36, 1985. (page 27).

[57] S. Pascoli, S. T. Petcov, and Antonio Riotto. Leptogenesis and Low Energy CP Violation in Neutrino Physics. *Nucl. Phys. B*, 774:1–52, 2007. (page 27).

[58] K. Abe et al. Diffuse supernova neutrino background search at Super-Kamiokande. *Phys. Rev. D*, 104(12):122002, 2021. (page 28).

[59] A. Takenaka et al. Search for proton decay via $p \to e^+\pi^0$ and $p \to \mu^+\pi^0$ with an enlarged fiducial volume in Super-Kamiokande I-IV. *Phys. Rev. D*, 102(11):112011, 2020. (page 28).

[60] K. Abe et al. Search for Cosmic-Ray Boosted Sub-GeV Dark Matter Using Recoil Protons at Super-Kamiokande. *Phys. Rev. Lett.*, 130(3):031802, 2023. [Erratum: Phys.Rev.Lett. 131, 159903 (2023)]. (page 28).

[61] M. Aguilar et al. Precision Measurement of the Proton Flux in Primary Cosmic Rays from Rigidity 1 GV to 1.8 TV with the Alpha Magnetic Spectrometer on the International Space Station. *Phys. Rev. Lett.*, 114:171103, 2015. (pages 31, 53).

[62] K. Abe et al. Measurements of cosmic-ray proton and helium spectra from the BESS-Polar long-duration balloon flights over Antarctica. *Astrophys. J.*, 822(2):65, 2016. (pages 31, 53).

[63] Marius Potgieter. Solar Modulation of Cosmic Rays. *Living Rev. Solar Phys.*, 10:3, 2013. (page 31).

[64] T. Futagami et al. Observation of the east - west anisotropy of the atmospheric neutrino flux. *Phys. Rev. Lett.*, 82:5194–5197, 1999. (page 31).

[65] K. Abe et al. The T2K Experiment. *Nucl. Instrum. Meth. A*, 659:106–135, 2011. (page 32).

[66] K. Matsuoka et al. Design and Performance of the Muon Monitor for the T2K Neutrino Oscillation Experiment. *Nucl. Instrum. Meth. A*, 624:591–600, 2010. (page 33).

[67] Y. Fukuda et al. The Super-Kamiokande detector. *Nucl. Instrum. Meth. A*, 501:418–462, 2003. (pages 34, 36).

[68] H. Nishino, K. Awai, Y. Hayato, S. Nakayama, K. Okumura, M. Shiozawa, A. Takeda, K. Ishikawa, A. Minegishi, and Y. Arai. High-speed charge-to-time converter ASIC for the Super-Kamiokande detector. *Nucl. Instrum. Meth. A*, 610:710–717, 2009. (page 36).

[69] M. Barranco-Luque et al. The Construction of the Central Detector for an Experiment at the CERN anti-p p Collider. *Nucl. Instrum. Meth.*, 176:175, 1980. (page 38).

[70] J. Altegoer et al. The NOMAD experiment at the CERN SPS. *Nucl. Instrum. Meth. A*, 404:96–128, 1998. (page 38).

[71] P. A. Amaudruz et al. The T2K Fine-Grained Detectors. *Nucl. Instrum. Meth. A*, 696:1–31, 2012. (page 38).

[72] N. Abgrall et al. Time Projection Chambers for the T2K Near Detectors. *Nucl. Instrum. Meth. A*, 637:25–46, 2011. (pages 39, 40).

[73] I. Giomataris, R. De Oliveira, S. Andriamonje, S. Aune, G. Charpak, P. Colas, A. Giganon, Ph. Rebourgeard, and P. Salin. Micromegas in a bulk. *Nucl. Instrum. Meth. A*, 560:405–408, 2006. (page 39).

[74] D. Allan et al. The Electromagnetic Calorimeter for the T2K Near Detector ND280. *JINST*, 8:P10019, 2013. (page 39).

[75] S. Assylbekov et al. The T2K ND280 Off-Axis Pi-Zero Detector. *Nucl. Instrum. Meth. A*, 686:48–63, 2012. (page 40).

[76] S. Aoki et al. The T2K Side Muon Range Detector (SMRD). *Nucl. Instrum. Meth. A*, 698:135–146, 2013. (page 40).

[77] R. Wendell et al. Atmospheric neutrino oscillation analysis with sub-leading effects in Super-Kamiokande I, II, and III. *Phys. Rev. D*, 81:092004, 2010. (page 43).

[78] Miao Jiang. *Study of the neutrino mass hierarchy with the atmospheric neutrino data collected in Super-Kamiokande IV*. PhD thesis, Kyoto University, Mar. 2019. (pages 44, 46, 148, 195).

[79] Thomas Wester. *Discerning the Neutrino Mass Ordering using Atmospheric Neutrinos in Super-Kamiokande I-V*. PhD thesis, Boston University, Boston U., 2023. (pages 44, 195).

[80] K. Abe et al. Improved constraints on neutrino mixing from the T2K experiment with $\mathbf{3.13 \times 10^{21}}$ protons on target. *Phys. Rev. D*, 103(11):112008, 2021. (pages 48, 198).

[81] K. Abe et al. Measurements of neutrino oscillation in appearance and disappearance channels by the T2K experiment with 6.6×10^{20} protons on target. *Phys. Rev. D*, 91(7):072010, 2015. (page 49).

[82] M. Honda, T. Kajita, K. Kasahara, and S. Midorikawa. Improvement of low energy atmospheric neutrino flux calculation using the JAM nuclear interaction model. *Phys. Rev. D*, 83:123001, 2011. (pages 53, 54).

[83] Y. Nara, N. Otuka, A. Ohnishi, K. Niita, and S. Chiba. Study of relativistic nuclear collisions at AGS energies from p + Be to Au + Au with hadronic cascade model. *Phys. Rev. C*, 61:024901, 2000. (page 54).

[84] Stefan Roesler, Ralph Engel, and Johannes Ranft. The Monte Carlo event generator DPMJET-III. In *International Conference on Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications (MC 2000)*, pages 1033–1038, 12 2000. (page 54).

[85] US Standard Atmosphere. *US standard atmosphere*. National Oceanic and Atmospheric Administration, 1976. (page 54).

[86] G. D. Barr, T. K. Gaisser, P. Lipari, Simon Robbins, and T. Stanev. A Three - dimensional calculation of atmospheric neutrinos. *Phys. Rev. D*, 70:023006, 2004. (page 54).

[87] G. Battistoni, A. Ferrari, T. Montaruli, and P. R. Sala. The FLUKA atmospheric neutrino flux calculation. *Astropart. Phys.*, 19:269–290, 2003. [Erratum: Astropart.Phys. 19, 291–294 (2003)]. (page 54).

[88] T. Sanuki, Morihiro Honda, T. Kajita, K. Kasahara, and S. Midorikawa. Study of cosmic ray interaction model based on atmospheric muons for the neutrino flux calculation. *Phys. Rev. D*, 75:043005, 2007. (page 54).

[89] T. T. Böhlen, F. Cerutti, M. P. W. Chin, A. Fassò, A. Ferrari, P. G. Ortega, A. Mairani, P. R. Sala, G. Smirnov, and V. Vlachoudis. The FLUKA Code: Developments and Challenges for High Energy and Medical Applications. *Nucl. Data Sheets*, 120:211–214, 2014. (page 55).

[90] Alfredo Ferrari, Paola R. Sala, Alberto Fasso, and Johannes Ranft. FLUKA: A multi-particle transport code (Program version 2005). 10 2005. (page 55).

[91] K. Abe et al. T2K neutrino flux prediction. *Phys. Rev. D*, 87(1):012001, 2013. [Addendum: Phys.Rev.D 87, 019902 (2013)]. (pages 55, 60).

[92] N. Abgrall et al. Measurements of $\pi^{\pm}$ differential yields from the surface of the T2K replica target for incoming 31 GeV/c protons with the NA61/SHINE spectrometer at the CERN SPS. *Eur. Phys. J. C*, 76(11):617, 2016. (pages 57, 58).

[93] N. Abgrall et al. Measurements of $\pi^{\pm}$, $K^{\pm}$ and proton double differential yields from the surface of the T2K replica target for incoming 31 GeV/c protons with the NA61/SHINE spectrometer at the CERN SPS. *Eur. Phys. J. C*, 79(2):100, 2019. (page 57).

[94] N. Abgrall et al. Measurements of $\pi^{\pm}$, $K^{\pm}$, K^0_S , Λ and proton production in proton–carbon interactions at 31 GeV/c with the NA61/SHINE spectrometer at the CERN SPS. *Eur. Phys. J. C*, 76(2):84, 2016. (page 58).

[95] Y. Hayato. NEUT. *Nucl. Phys. B Proc. Suppl.*, 112:171–176, 2002. (page 59).

[96] Yoshinari Hayato. A neutrino interaction simulation program library NEUT. *Acta Phys. Polon. B*, 40:2477–2489, 2009. (pages 59, 66, 68).

[97] M. Honda, M. Sajjad Athar, T. Kajita, K. Kasahara, and S. Midorikawa. Atmospheric neutrino flux calculation using the NRLMSISE-00 atmospheric model. *Phys. Rev. D*, 92(2):023004, 2015. (page 60).

[98] Yoshinari Hayato and Luke Pickering. The NEUT neutrino interaction simulation program library. *Eur. Phys. J. ST*, 230(24):4469–4481, 2021. (pages 60, 66, 68, 215).

[99] C. H. Llewellyn Smith. Neutrino Reactions at Accelerator Energies. *Phys. Rept.*, 3:261–379, 1972. (pages 61, 66).

[100] R. Bradford, A. Bodek, Howard Scott Budd, and J. Arrington. A New parameterization of the nucleon elastic form-factors. *Nucl. Phys. B Proc. Suppl.*, 159:127–132, 2006. (page 62).

[101] Veronique Bernard, Latifa Elouadrhiri, and Ulf-G. Meissner. Axial structure of the nucleon: Topical Review. *J. Phys. G*, 28:R1–R35, 2002. (pages 62, 214).

[102] Dieter Rein and Lalit M. Sehgal. Neutrino Excitation of Baryon Resonances and Single Pion Production. *Annals Phys.*, 133:79–153, 1981. (pages 62, 63).

[103] Krzysztof M. Graczyk and Jan T. Sobczyk. Form Factors in the Quark Resonance Model. *Phys. Rev. D*, 77:053001, 2008. [Erratum: Phys.Rev.D 79, 079903 (2009)]. (page 63).

[104] G. M. Radecky et al. Study of Single Pion Production by Weak Charged Currents in Low-energy Neutrino *d* Interactions. *Phys. Rev. D*, 25:1161–1173, 1982. [Erratum: Phys.Rev.D 26, 3297 (1982)]. (page 63).

[105] T. Kitagaki et al. Charged Current Exclusive Pion Production in Neutrino Deuterium Interactions. *Phys. Rev. D*, 34:2554–2565, 1986. (page 63).

[106] P. Allen et al. A Study of Single Meson Production in Neutrino and Anti-neutrinos Charged Current Interactions on Protons. *Nucl. Phys. B*, 264:221–242, 1986. (page 63).

[107] Dieter Rein and Lalit M. Sehgal. Coherent pi0 Production in Neutrino Reactions. *Nucl. Phys. B*, 223:29–44, 1983. (page 63).

[108] Ch. Berger and L. M. Sehgal. PCAC and coherent pion production by low energy neutrinos. *Phys. Rev. D*, 79:053003, 2009. (page 63).

[109] D. Rein. Diffractive Pion Production in Neutrino Reactions. *Nucl. Phys. B*, 278:61–77, 1986. (page 63).

[110] V. Barone, C. Pascaud, and F. Zomer. A New global analysis of deep inelastic scattering data. *Eur. Phys. J. C*, 12:243–262, 2000. (page 63).

[111] M. Glück, E. Reya, and A. Vogt. Dynamical parton distributions revisited. *Eur. Phys. J. C*, 5:461–470, 1998. (page 63).

[112] A. Bodek and U. K. Yang. Modeling neutrino and electron scattering cross-sections in the few GeV region with effective LO PDFs. *AIP Conf. Proc.*, 670(1):110–117, 2003. (page 63).

[113] A. Bodek and U. Yang. A unified model for inelastic e-N and v-N cross sections at all Q**2. *AIP Conf. Proc.*, 842(1):360–362, 2006. (page 63).

[114] R. A. Smith and E. J. Moniz. NEUTRINO REACTIONS ON NUCLEAR TAR-
GETS. *Nucl. Phys. B*, 43:605, 1972. [Erratum: Nucl.Phys.B 101, 547 (1975)].
(page 64).

[115] J. Nieves, E. Oset, and C. Garcia-Recio. A Theoretical approach to pionic atoms
and the problem of anomalies. *Nucl. Phys. A*, 554:509–553, 1993. (pages 64, 112,
113, 213, 217).

[116] J. Nieves, I. Ruiz Simo, and M. J. Vicente Vacas. Inclusive Charged–Current
Neutrino–Nucleus Reactions. *Phys. Rev. C*, 83:045501, 2011. (pages 64, 65, 113,
213, 216).

[117] P. A. Rodrigues et al. Identification of nuclear effects in neutrino-carbon interactions
at low three-momentum transfer. *Phys. Rev. Lett.*, 116:071802, 2016. [Addendum:
Phys.Rev.Lett. 121, 209902 (2018)]. (pages 64, 69, 109).

[118] J. Nieves, Jose Enrique Amaro, and M. Valverde. Inclusive quasi-elastic neutrino
reactions. *Phys. Rev. C*, 70:055503, 2004. [Erratum: Phys.Rev.C 72, 019902 (2005)].
(page 64).

[119] O. Benhar, A. Fabrocini, S. Fantoni, and I. Sick. Spectral function of finite nuclei
and scattering of GeV electrons. *Nucl. Phys. A*, 579:493–517, 1994. (pages 64, 66).

[120] A. Bodek, M. E. Christy, and B. Coopersmith. Effective Spectral Function for
Quasielastic Scattering on Nuclei. *Eur. Phys. J. C*, 74(10):3091, 2014. (page 64).

[121] Artur M. Ankowski, Omar Benhar, and Makoto Sakuda. Improving the accuracy of
neutrino energy reconstruction in charged-current quasielastic scattering off nuclear
targets. *Phys. Rev. D*, 91(3):033005, 2015. (page 65).

[122] K. Abe et al. Characterization of nuclear effects in muon-neutrino scattering on hy-
drocarbon with a measurement of final-state kinematics and correlations in charged-
current pionless interactions at T2K. *Phys. Rev. D*, 98(3):032003, 2018. (pages 65,
214).

[123] S. Dolan. Exploring Nuclear Effects in Neutrino-Nucleus Interactions Using Mea-
surements of Transverse Kinematic Imbalance from T2K and MINERvA. 10 2018.
(pages 65, 66).

[124] T. Golan, J. T. Sobczyk, and J. Zmuda. NuWro: the Wroclaw Monte Carlo Gener-
ator of Neutrino Interactions. *Nucl. Phys. B Proc. Suppl.*, 229-232:499–499, 2012.
(page 66).

[125] Luke Pickering. Examining nuclear effects in neutrino interactions with transverse
kinematic imbalance. *JPS Conf. Proc.*, 12:010032, 2016. (page 66).

[126] J. Nieves, I. Ruiz Simo, and M. J. Vicente Vacas. The nucleon axial mass and
the MiniBooNE Quasielastic Neutrino-Nucleus Scattering problem. *Phys. Lett. B*,
707:72–75, 2012. (pages 66, 67, 112, 215).

[127] M. Martini, M. Ericson, G. Chanfray, and J. Marteau. A Unified approach for nucleon knock-out, coherent and incoherent pion production in neutrino interactions with nuclei. *Phys. Rev. C*, 80:065501, 2009. (pages 65, 67, 112, 216, 217).

[128] J. Nieves, I. Ruiz Simo, and M. J. Vicente Vacas. Two Particle-Hole Excitations in Charged Current Quasielastic Antineutrino–Nucleus Scattering. *Phys. Lett. B*, 721:90–93, 2013. (page 65).

[129] 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. G*, 44(6):065105, 2017. (pages 67, 216).

[130] W.M. Alberico, M. Ericson, and A. Molinari. The role of two particle-two hole excitations in the spin-isospin nuclear response. *Annals of Physics*, 154(2):356–395, 1984. (page 67).

[131] G. D. Megias, J. E. Amaro, M. B. Barbaro, J. A. Caballero, T. W. Donnelly, and I. Ruiz Simo. Charged-current neutrino-nucleus reactions within the superscaling meson-exchange current approach. *Phys. Rev. D*, 94(9):093004, 2016. (pages 67, 113, 216).

[132] L. L. Salcedo, E. Oset, M. J. Vicente-Vacas, and C. Garcia-Recio. Computer Simulation of Inclusive Pion Nuclear Reactions. *Nucl. Phys. A*, 484:557–592, 1988. (page 68).

[133] E. Oset and L. L. Salcedo. Δ Selfenergy in Nuclear Matter. *Nucl. Phys. A*, 468:631–652, 1987. (page 68).

[134] E. S. Pinzon Guerra et al. Measurement of σ_{ABS} and σ_{CX} of π^+ on carbon by the Dual Use Experiment at TRIUMF (DUET). *Phys. Rev. C*, 95(4):045203, 2017. (page 68).

[135] Sowjanya Gollapinni. Neutrino Cross section Future. In *Prospects in Neutrino Physics*, 2 2016. (page 68).

[136] P. Gueye et al. Coulomb distortion measurements by comparing electron and positron quasielastic scattering off C-12 and Pb-208. *Phys. Rev. C*, 60:044308, 1999. (page 68).

[137] D. Ruterbories et al. Measurement of Quasielastic-Like Neutrino Scattering at $\langle E_\nu \rangle \sim 3.5$ GeV on a Hydrocarbon Target. *Phys. Rev. D*, 99(1):012004, 2019. (pages 69, 109).

[138] K. Abe et al. First combined measurement of the muon neutrino and antineutrino charged-current cross section without pions in the final state at T2K. *Phys. Rev. D*, 101(11):112001, 2020. (page 69).

[139] K. Abe et al. Simultaneous measurement of the muon neutrino charged-current cross section on oxygen and carbon without pions in the final state at T2K. *Phys. Rev. D*, 101(11):112004, 2020. (page 69).

[140] A. Capella, A. Kaidalov, C. Merino, and J. Tran Thanh Van. Structure functions and low x physics. *Phys. Lett. B*, 337:358–366, 1994. (page 71).

[141] N. Jachowicz, K. Heyde, J. Ryckebusch, and S. Rombouts. Continuum random phase approximation approach to charged current neutrino nucleus scattering. *Phys. Rev. C*, 65:025501, 2002. (pages 72, 112, 113).

[142] Junjie Xia. *Measurements of Neutrino Oscillation Parameters from A Joint Analysis of Super-K Atmospheric and T2K Accelerator Samples*. PhD thesis, The University of Tokyo, 2022. (page 72).

[143] Stephen L. Adler. Photoproduction, electroproduction and weak single pion production in the (3,3) resonance region. *Annals Phys.*, 50:189–311, 1968. (page 74).

[144] F. Sánchez. Possibility of measuring Adler angles in charged current single pion neutrino-nucleus interactions. *Phys. Rev. D*, 93(9):093015, 2016. (page 74).

[145] Lukas Berns. *Constraints on CP violation and neutrino mass ordering with a joint analysis of accelerator and atmospheric neutrino oscillations*. PhD thesis, Tokyo Institute of Technology, 2022. (pages 74, 89).

[146] Alexis A. Aguilar-Arevalo et al. Measurement of ν_μ and $\bar\nu_\mu$ induced neutral current single π^0 production cross sections on mineral oil at $E_\nu \sim \mathcal{O}(1\text{GeV})$. *Phys. Rev. D*, 81:013005, 2010. (page 74).

[147] René Brun, F. Bruyant, Federico Carminati, Simone Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban. GEANT Detector Description and Simulation Tool. 10 1994. (page 74).

[148] C. Zeitnitz and T. A. Gabriel. The GEANT - CALOR interface and benchmark calculations of ZEUS test calorimeters. *Nucl. Instrum. Meth. A*, 349:106–111, 1994. (page 74).

[149] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Annals Math. Statist.*, 9(1):60–62, 1938. (page 87).

[150] Gary J. Feldman and Robert D. Cousins. A Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57:3873–3889, 1998. (page 88).

[151] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev. D*, 98(3):030001, 2018. (page 88).

[152] A. M. Dziewonski and D. L. Anderson. Preliminary reference earth model. *Phys. Earth Planet. Interiors*, 25:297–356, 1981. (page 89).

[153] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975. (page 91).

[154] Yudi Pawitan. *In all likelihood: statistical modelling and inference using likelihood.* Oxford University Press, 2001. (page 91).

[155] Chien-Fu Jeff Wu. Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics*, 14(4):1261–1295, 1986. (page 92).

[156] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods.* Springer New York, 2004. (page 93).

[157] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953. (page 93).

[158] Bernd A. Berg. Introduction to multicanonical Monte Carlo simulations. *Fields Inst. Commun.*, 26:1, 2000. (page 95).

[159] Roger J. Barlow and Christine Beeston. Fitting using finite Monte Carlo samples. *Comput. Phys. Commun.*, 77:219–228, 1993. (page 97).

[160] J. S. Conway. Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra. In *PHYSTAT 2011*, pages 115–120, 2011. (page 97).

[161] Isaac Asimov. Franchise. In *Earth Is Room Enough*. Doubleday / SFBC, 1957. (page 99).

[162] M.D. Lee and E.J. Wagenmakers. *Bayesian Cognitive Modeling: A Practical Course.* Cambridge University Press, 2014. (pages 110, 111, 143).

[163] 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. C*, 92(2):024606, 2015. (pages 112, 113).

[164] C. Adamuscin, E. Tomasi-Gustafsson, E. Santopinto, and R. Bijker. Two-component model for the axial form factor of the nucleon. *Phys. Rev. C*, 78:035201, 2008. (pages 112, 215).

[165] R. Gonzaléz-Jiménez, G. D. Megias, M. B. Barbaro, J. A. Caballero, and T. W. Donnelly. Extensions of Superscaling from Relativistic Mean Field Theory: the SuSAv2 Model. *Phys. Rev. C*, 90(3):035501, 2014. (pages 113, 114).

[166] S. Dolan, A. Nikolakopoulos, O. Page, S. Gardiner, N. Jachowicz, and V. Pandey. Implementation of the continuum random phase approximation model in the GENIE generator and an analysis of nuclear effects in low-energy transfer neutrino interactions. *Phys. Rev. D*, 106(7):073001, 2022. (page 114).

[167] H. Jeffreys. *The Theory of Probability.* Oxford Classic Texts in the Physical Sciences. OUP Oxford, 1998. (page 143).

[168] Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995. (page 143).

[169] Andrew Gelman, Xiao-Li Meng, and Hal Stern. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, pages 733–760, 1996. (page 147).

[170] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979. (page 149).

[171] F. P. An et al. Precision Measurement of Reactor Antineutrino Oscillation at Kilometer-Scale Baselines by Daya Bay. *Phys. Rev. Lett.*, 130(16):161802, 2023. (pages 178, 179).

[172] Hiroshi Nunokawa, Stephen J. Parke, and Renata Zukanovich Funchal. Another possible way to determine the neutrino mass hierarchy. *Phys. Rev. D*, 72:013009, 2005. (page 179).

[173] James M Dickey. The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics*, pages 204–223, 1971. (page 183).

[174] Eric-Jan Wagenmakers, Tom Lodewyckx, Himanshu Kuriyal, and Raoul Grasman. Bayesian hypothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive psychology*, 60(3):158–189, 2010. (page 183).

[175] Alexander L. Read. Presentation of search results: The CL_s technique. *J. Phys. G*, 28:2693–2704, 2002. (page 186).

[176] K. Abe et al. Hyper-Kamiokande Design Report. 2 2016. (pages 188, 192).

[177] Hirokazu Ikeda, Masaki Mori, Tohru Tanimori, Ken-ichi Kihara, Yoichiro Suzuki, and Yasuhiko Haren. Front end hybrid circuit for Super-Kamiokande. *Nucl. Instrum. Meth. A*, 320:310–316, 1992. (page 188).

[178] M. Shiozawa. Reconstruction algorithms in the Super-Kamiokande large water Cherenkov detector. *Nucl. Instrum. Meth. A*, 433:240–246, 1999. (page 188).

[179] K. Abe et al. J-PARC Neutrino Beamline Upgrade Technical Design Report. 8 2019. (page 192).

[180] K. Abe et al. T2K ND280 Upgrade - Technical Design Report. 1 2019. (page 192).

[181] S. Dolan et al. Sensitivity of the upgraded T2K Near Detector to constrain neutrino and antineutrino interactions with no mesons in the final state by exploiting nucleon-lepton correlations. *Phys. Rev. D*, 105(3):032010, 2022. (page 192).

[182] Aoi Eguchi. Development of Selection Algorithms for Electron Neutrino Interaction Events with New T2K Near Detectors, 2021. (page 192).

[183] Shantanu Desai. *High energy neutrino astrophysics with Super-Kamiokande*. PhD thesis, Boston University, Massachusetts, May 2004. (page 196).

[184] Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461, 1946. (page 200).

[185] Keiichi Kimura, Akira Takamura, and Hidekazu Yokomakura. Enhancement of CP violating terms for neutrino oscillation in earth matter. *Phys. Lett. B*, 600:91–103, 2004. (page 201).

[186] P.D. Hoff. *A First Course in Bayesian Statistical Methods*. Springer Texts in Statistics. Springer New York, 2009. (page 202).

[187] John Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Staff Report 148, Federal Reserve Bank of Minneapolis, 1991. (page 203).

[188] Andrew Gelman and Donald B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457 – 472, 1992. (page 203).

[189] M.E.J. Newman and G.T. Barkema. *Monte Carlo Methods in Statistical Physics*. Clarendon Press, 1999. (page 206).

[190] Arie Bodek and Tejin Cai. Removal Energies and Final State Interaction in Lepton Nucleus Scattering. *Eur. Phys. J. C*, 79(4):293, 2019. (pages 214, 215).

[191] D. Dutta et al. A Study of the quasielastic (e,e-prime p) reaction on C-12, Fe-56 and Au-97. *Phys. Rev. C*, 68:064603, 2003. (page 214).

[192] M. Leuschner et al. Quasielastic proton knockout from O-16. *Phys. Rev. C*, 49:955–967, 1994. (page 214).

[193] A. A. Aguilar-Arevalo et al. First Measurement of the Muon Neutrino Charged Current Quasielastic Double Differential Cross Section. *Phys. Rev. D*, 81:092005, 2010. (pages 215, 216).

[194] L. B. Auerbach et al. Measurements of charged current reactions of muon neutrinos on C-12. *Phys. Rev. C*, 66:015501, 2002. (page 216).

[195] V Lyubushkin et al. A Study of quasi-elastic muon neutrino and antineutrino scattering in the NOMAD experiment. *Eur. Phys. J. C*, 63:355–381, 2009. (page 216).

[196] G. D. Megias, S. Bolognesi, M. B. Barbaro, and E. Tomasi-Gustafsson. New evaluation of the axial nucleon form factor from electron- and neutrino-scattering data and impact on neutrino-nucleus cross sections. *Phys. Rev. C*, 101(2):025501, 2020. (page 215).

[197] K. L. Miller et al. STUDY OF THE REACTION muon-neutrino D —> MU- P P(S). *Phys. Rev. D*, 26:537–542, 1982. (page 216).

[198] S. J. Barish et al. Study of Neutrino Interactions in Hydrogen and Deuterium. 1. Description of the Experiment and Study of the Reaction Neutrino d –> mu- p p(s). *Phys. Rev. D*, 16:3103, 1977. (page 216).

[199] N. J. Baker, A. M. Cnops, P. L. Connolly, S. A. Kahn, H. G. Kirk, M. J. Murtagh, R. B. Palmer, N. P. Samios, and M. Tanaka. Quasielastic Neutrino Scattering: A Measurement of the Weak Nucleon Axial Vector Form-Factor. *Phys. Rev. D*, 23:2499–2505, 1981. (page 216).

[200] T. Kitagaki et al. High-Energy Quasielastic Muon-neutrino n —> mu- p Scattering in Deuterium. *Phys. Rev. D*, 28:436–442, 1983. (page 216).

[201] D. Allasia et al. Investigation of exclusive channels in neutrino / anti-neutrino deuteron charged current interactions. *Nucl. Phys. B*, 343:285–309, 1990. (page 216).

[202] Z. Koba, Holger Bech Nielsen, and P. Olesen. Scaling of multiplicity distributions in high-energy hadron collisions. *Nucl. Phys. B*, 40:317–334, 1972. (page 217).

[203] Konstantin S. Kuzmin and Vadim A. Naumov. Mean charged multiplicities in charged-current neutrino scattering on hydrogen and deuterium. *Phys. Rev. C*, 88:065501, 2013. (page 217).

[204] Christophe Bronner and Mark Hartz. Tuning of the Charged Hadrons Multiplicities for Deep Inelastic Interactions in NEUT. *JPS Conf. Proc.*, 12:010041, 2016. (pages 217, 218).

[205] J. Wolcott et al. Measurement of electron neutrino quasielastic and quasielasticlike scattering on hydrocarbon at $\langle E_\nu \rangle$ = 3.6 GeV. *Phys. Rev. Lett.*, 116(8):081802, 2016. (page 218).

[206] K. Abe et al. Measurement of the charged-current electron (anti-)neutrino inclusive cross-sections at the T2K off-axis near detector ND280. *JHEP*, 10:114, 2020. (page 218).

[207] P. Abratenko et al. Measurement of the flux-averaged inclusive charged-current electron neutrino and antineutrino cross section on argon using the NuMI beam and the MicroBooNE detector. *Phys. Rev. D*, 104(5):052002, 2021. (page 218).

[208] P. Abratenko et al. First measurement of inclusive electron-neutrino and antineutrino charged current differential cross sections in charged lepton energy on argon in MicroBooNE. *Phys. Rev. D*, 105(5):L051102, 2022. (page 218).

[209] Melanie Day and Kevin S. McFarland. Differences in Quasi-Elastic Cross-Sections of Muon and Electron Neutrinos. *Phys. Rev. D*, 86:053003, 2012. (page 218).