Improved constraints on neutrino mixing from the T2K experiment with 3.13×10^{21} protons on target

K. Abe,[54] N. Akhlaq,[45] R. Akutsu,[60] A. Ali,[30] C. Alt,[9] C. Andreopoulos,[52,34] M. Antonova,[17] S. Aoki,[29] T. Arihara,[57] Y. Asada,[67] Y. Ashida,[30] E. T. Atkin,[19] Y. Awataguchi,[57] G. J. Barker,[64] G. Barr,[42] D. Barrow,[42] M. Batkiewicz-Kwasniak,[13] A. Beloshapkin,[24] F. Bench,[34] V. Berardi,[20] L. Berns,[56] S. Bhadra,[68] A. Blanchet,[51] A. Blondel,[51,11] S. Bolognesi,[5] T. Bonus,[66] B. Bourguille,[16] S. B. Boyd,[64] A. Bravar,[11] D. Bravo Berguño,[1] C. Bronner,[54] S. Bron,[11] A. Bubak,[50] M. Buizza Avanzini,[33] S. Cao,[18] S. L. Cartwright,[49] M. G. Catanesi,[20] A. Cervera,[17] J. Chakrani,[33] D. Cherdack,[15] G. Christodoulou,[10] M. Cicerchia,[22,*] J. Coleman,[34] G. Collazuol,[22] L. Cook,[42,26] D. Coplowe,[42] A. Cudd,[6] G. De Rosa,[21] T. Dealtry,[31] C. C. Delogu,[22] S. R. Dennis,[34] C. Densham,[52] A. Dergacheva,[24] F. Di Lodovico,[28] S. Dolan,[10] D. Douqa,[11] T. A. Doyle,[31] J. Dumarchez,[51] P. Dunne,[19] A. Eguchi,[53] L. Eklund,[12] S. Emery-Schrenk,[5] A. Ereditato,[2] A. J. Finch,[31] G. Fiorillo,[21] C. Francois,[2] M. Friend,[14,†] Y. Fujii,[14,†] R. Fukuda,[58] Y. Fukuda,[37] K. Fusshoeller,[9] C. Giganti,[51] M. Gonin,[33] E. A. G. Goodman,[12] A. Gorin,[24] M. Grassi,[22] M. Guigue,[51] D. R. Hadley,[64] P. Hamacher-Baumann,[48] D. A. Harris,[68] M. Hartz,[60,26] T. Hasegawa,[14,†] S. Hassani,[5] N. C. Hastings,[14] Y. Hayato,[54,26] A. Hiramoto,[30] M. Hogan,[7] J. Holeczek,[50] N. T. Hong Van,[18,25] T. Honjo,[41] F. Iacob,[22] A. K. Ichikawa,[63] M. Ikeda,[54] T. Ishida,[14,†] M. Ishitsuka,[58] K. Iwamoto,[53] A. Izmaylov,[24] N. Izumi,[58] M. Jakkapu,[14] B. Jamieson,[65] S. J. Jenkins,[49] C. Jesús-Valls,[16] J. J. Jiang,[39] P. Jonsson,[19] C. K. Jung,[39,‡] P. B. Jurj,[19] M. Kabirnezhad,[68] H. Kakuno,[57] J. Kameda,[54] S. P. Kasetti,[35] Y. Kataoka,[54] Y. Katayama,[67] T. Katori,[28] E. Kearns,[3,26,‡] M. Khabibullin,[24] A. Khotjantsev,[24] T. Kikawa,[30] H. Kikutani,[53] S. King,[28] J. Kisiel,[50] T. Kobata,[41] T. Kobayashi,[14,†] L. Koch,[42] A. Konaka,[60] L. L. Kormos,[31] Y. Koshio,[40,‡] A. Kostin,[24] K. Kowalik,[38] Y. Kudenko,[24,§] S. Kuribayashi,[30] R. Kurjata,[62] T. Kutter,[35] M. Kuze,[56] L. Labarga,[1] J. Lagoda,[38] M. Lamoureux,[22] D. Last,[43] M. Laveder,[22] M. Lawe,[31] S.-K. Lin,[35] R. P. Litchfield,[12] S. L. Liu,[39] A. Longhin,[22] L. Ludovici,[23] X. Lu,[42] T. Lux,[16] L. N. Machado,[21] L. Magaletti,[20] K. Mahn,[36] M. Malek,[49] S. Manly,[46] L. Maret,[11] A. D. Marino,[6] L. Marti-Magro,[1] T. Maruyama,[14,†] T. Matsubara,[14] K. Matsushita,[53] C. Mauger,[43] A. Maurel,[5] K. Mavrokoridis,[34] E. Mazzucato,[5] N. McCauley,[34] J. McElwee,[49] K. S. McFarland,[46] C. McGrew,[39] A. Mefodiev,[24] G. D. Megias,[55,54] L. Mellet,[51] M. Mezzetto,[22] A. Minamino,[67] O. Mineev,[24] S. Mine,[4] M. Miura,[54,‡] L. Molina Bueno,[17] S. Moriyama,[54,‡] Th. A. Mueller,[33] D. Munford,[15] L. Munteanu,[5] Y. Nagai,[6] T. Nakadaira,[14,†] M. Nakahata,[54,26] Y. Nakajima,[53] A. Nakamura,[40] H. Nakamura,[58] K. Nakamura,[26,14,†] Y. Nakano,[29] S. Nakayama,[54,26] T. Nakaya,[30,26] K. Nakayoshi,[14,†] C. E. R. Naseby,[19] T. V. Ngoc,[18,‖] V. Q. Nguyen,[51] K. Niewczas,[66] Y. Nishimura,[27] K. Nishizaki,[41] E. Noah,[11] T. S. Nonnenmacher,[19] F. Nova,[52] J. Nowak,[31] J. C. Nugent,[12] H. M. O'Keeffe,[31] L. O'Sullivan,[49] T. Odagawa,[30] T. Ogawa,[14] R. Okada,[40] K. Okumura,[55,26] T. Okusawa,[41] R. A. Owen,[45] Y. Oyama,[14,†] V. Palladino,[21] V. Paolone,[44] M. Pari,[22] W. C. Parker,[47] J. Parlone,[34] S. Parsa,[11] J. Pasternak,[19] M. Pavin,[60] D. Payne,[34] G. C. Penn,[34] D. Pershey,[8] L. Pickering,[36] C. Pidcott,[49] G. Pintaudi,[67] C. Pistillo,[2] B. Popov,[51,¶] K. Porwit,[50] M. Posiadala-Zezula,[61] A. Pritchard,[34] B. Quilain,[33] T. Radermacher,[48] E. Radicioni,[20] B. Radics,[9] P. N. Ratoff,[31] M. Reh,[6] C. Riccio,[39] E. Rondio,[38] S. Roth,[48] A. Rubbia,[9] A. C. Ruggeri,[21] C. A. Ruggles,[12] A. Rychter,[62] L. S. M. Lakshmi,[38] K. Sakashita,[14,†] F. Sánchez,[11] G. Santucci,[68] C. M. Schloesser,[9] K. Scholberg,[8,‡] M. Scott,[19] Y. Seiya,[41,**] T. Sekiguchi,[14,†] H. Sekiya,[54,26,‡] D. Sgalaberna,[9] A. Shaikhiev,[24] A. Shaykina,[24] M. Shiozawa,[54,26] W. Shorrock,[19] A. Shvartsman,[24,§] K. Skwarczynski,[38] M. Smy,[4] J. T. Sobczyk,[66] H. Sobel,[4,26] F. J. P. Soler,[12] Y. Sonoda,[54] R. Spina,[20] S. Suvorov,[24,51] A. Suzuki,[29] S. Y. Suzuki,[14,†] Y. Suzuki,[26] A. A. Sztuc,[19] M. Tada,[14,†] M. Tajima,[30] A. Takeda,[54] Y. Takeuchi,[29,26] H. K. Tanaka,[54,‡] Y. Tanihara,[67] M. Tani,[30] N. Teshima,[41] N. Thamm,[48] L. F. Thompson,[49] W. Toki,[7] C. Touramanis,[34] T. Towstego,[59] K. M. Tsui,[34] T. Tsukamoto,[14,†] M. Tzanov,[35] Y. Uchida,[19] M. Vagins,[26,4] S. Valder,[64] D. Vargas,[16] G. Vasseur,[5] C. Vilela,[10] W. G. S. Vinning,[64] T. Vladisavljevic,[52] T. Wachala,[13] J. Walker,[65] J. G. Walsh,[31] Y. Wang,[39] L. Wan,[3] D. Wark,[52,42] M. O. Wascko,[19] A. Weber,[52,42] R. Wendell,[30,‡] M. J. Wilking,[39] C. Wilkinson,[32] J. R. Wilson,[28] K. Wood,[39] C. Wret,[46] J. Xia,[55] Y.-h. Xu,[31] K. Yamamoto,[41,**] C. Yanagisawa,[39,††] G. Yang,[39] T. Yano,[54] K. Yasutome,[30] N. Yershov,[24] M. Yokoyama,[53,‡] T. Yoshida,[56] Y. Yoshimoto,[53] M. Yu,[68] R. Zaki,[68] A. Zalewska,[13] J. Zalipska,[38] K. Zaremba,[62] G. Zarnecki,[38] M. Ziembicki,[62] M. Zito,[51] and S. Zsoldos[28]

(T2K Collaboration)

[1]*University Autonoma Madrid, Department of Theoretical Physics, 28049 Madrid, Spain*
[2]*University of Bern, Albert Einstein Center for Fundamental Physics, Laboratory for High Energy Physics (LHEP), Bern, Switzerland*
[3]*Boston University, Department of Physics, Boston, Massachusetts, USA*
[4]*University of California, Irvine, Department of Physics and Astronomy, Irvine, California, USA*

[5]*IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[6]*University of Colorado at Boulder, Department of Physics, Boulder, Colorado, USA*
[7]*Colorado State University, Department of Physics, Fort Collins, Colorado, USA*
[8]*Duke University, Department of Physics, Durham, North Carolina, USA*
[9]*ETH Zurich, Institute for Particle Physics and Astrophysics, Zurich, Switzerland*
[10]*CERN European Organization for Nuclear Research, CH-1211 Genève 23, Switzerland*
[11]*University of Geneva, Section de Physique, DPNC, Geneva, Switzerland*
[12]*University of Glasgow, School of Physics and Astronomy, Glasgow, United Kingdom*
[13]*H. Niewodniczanski Institute of Nuclear Physics PAN, Cracow, Poland*
[14]*High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki, Japan*
[15]*University of Houston, Department of Physics, Houston, Texas, USA*
[16]*Institut de Fisica d'Altes Energies (IFAE) - The Barcelona Institute of Science and Technology,
Campus UAB, Bellaterra (Barcelona) Spain*
[17]*IFIC (CSIC and University of Valencia), Valencia, Spain*
[18]*Institute For Interdisciplinary Research in Science and Education (IFIRSE), ICISE, Quy Nhon, Vietnam*
[19]*Imperial College London, Department of Physics, London, United Kingdom*
[20]*INFN Sezione di Bari and Università e Politecnico di Bari,
Dipartimento Interuniversitario di Fisica, Bari, Italy*
[21]*INFN Sezione di Napoli and Università di Napoli, Dipartimento di Fisica, Napoli, Italy*
[22]*INFN Sezione di Padova and Università di Padova, Dipartimento di Fisica, Padova, Italy*
[23]*INFN Sezione di Roma and Università di Roma "La Sapienza," Roma, Italy*
[24]*Institute for Nuclear Research of the Russian Academy of Sciences, Moscow, Russia*
[25]*International Centre of Physics, Institute of Physics (IOP), Vietnam Academy of Science and Technology
(VAST), 10 Dao Tan, Ba Dinh, Hanoi, Vietnam*
[26]*Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo Institutes
for Advanced Study, University of Tokyo, Kashiwa, Chiba, Japan*
[27]*Keio University, Department of Physics, Kanagawa, Japan*
[28]*King's College London, Department of Physics, Strand, London WC2R 2LS, United Kingdom*
[29]*Kobe University, Kobe, Japan*
[30]*Kyoto University, Department of Physics, Kyoto, Japan*
[31]*Lancaster University, Physics Department, Lancaster, United Kingdom*
[32]*Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[33]*Ecole Polytechnique, IN2P3-CNRS, Laboratoire Leprince-Ringuet, Palaiseau, France*
[34]*University of Liverpool, Department of Physics, Liverpool, United Kingdom*
[35]*Louisiana State University, Department of Physics and Astronomy, Baton Rouge, Louisiana, USA*
[36]*Michigan State University, Department of Physics and Astronomy, East Lansing, Michigan, USA*
[37]*Miyagi University of Education, Department of Physics, Sendai, Japan*
[38]*National Centre for Nuclear Research, Warsaw, Poland*
[39]*State University of New York at Stony Brook, Department of Physics and Astronomy,
Stony Brook, New York, USA*
[40]*Okayama University, Department of Physics, Okayama, Japan*
[41]*Osaka City University, Department of Physics, Osaka, Japan*
[42]*Oxford University, Department of Physics, Oxford, United Kingdom*
[43]*University of Pennsylvania, Department of Physics and Astronomy,
Philadelphia, Pennsylvania 19104, USA*
[44]*University of Pittsburgh, Department of Physics and Astronomy, Pittsburgh, Pennsylvania, USA*
[45]*Queen Mary University of London, School of Physics and Astronomy, London, United Kingdom*
[46]*University of Rochester, Department of Physics and Astronomy, Rochester, New York, USA*
[47]*Royal Holloway University of London, Department of Physics, Egham, Surrey, United Kingdom*
[48]*RWTH Aachen University, III. Physikalisches Institut, Aachen, Germany*
[49]*University of Sheffield, Department of Physics and Astronomy, Sheffield, United Kingdom*
[50]*University of Silesia, Institute of Physics, Katowice, Poland*
[51]*Sorbonne Université, Université Paris Diderot, CNRS/IN2P3, Laboratoire de Physique Nucléaire et de
Hautes Energies (LPNHE), Paris, France*
[52]*STFC, Rutherford Appleton Laboratory, Harwell Oxford,
and Daresbury Laboratory, Warrington, United Kingdom*
[53]*University of Tokyo, Department of Physics, Tokyo, Japan*
[54]*University of Tokyo, Institute for Cosmic Ray Research, Kamioka Observatory, Kamioka, Japan*
[55]*University of Tokyo, Institute for Cosmic Ray Research,
Research Center for Cosmic Neutrinos, Kashiwa, Japan*

[56]*Tokyo Institute of Technology, Department of Physics, Tokyo, Japan*
[57]*Tokyo Metropolitan University, Department of Physics, Tokyo, Japan*
[58]*Tokyo University of Science, Faculty of Science and Technology,*
Department of Physics, Noda, Chiba, Japan
[59]*University of Toronto, Department of Physics, Toronto, Ontario, Canada*
[60]*TRIUMF, Vancouver, British Columbia, Canada*
[61]*University of Warsaw, Faculty of Physics, Warsaw, Poland*
[62]*Warsaw University of Technology, Institute of Radioelectronics*
and Multimedia Technology, Warsaw, Poland
[63]*Tohoku University, Faculty of Science, Department of Physics, Miyagi, Japan*
[64]*University of Warwick, Department of Physics, Coventry, United Kingdom*
[65]*University of Winnipeg, Department of Physics, Winnipeg, Manitoba, Canada*
[66]*Wroclaw University, Faculty of Physics and Astronomy, Wroclaw, Poland*
[67]*Yokohama National University, Department of Physics, Yokohama, Japan*
[68]*York University, Department of Physics and Astronomy, Toronto, Ontario, Canada*

The T2K experiment reports updated measurements of neutrino and antineutrino oscillations using both appearance and disappearance channels. This result comes from an exposure of $14.9(16.4) \times 10^{20}$ protons on target in neutrino (antineutrino) mode. Significant improvements have been made to the neutrino interaction model and far detector reconstruction. An extensive set of simulated data studies have also been performed to quantify the effect interaction model uncertainties have on the T2K oscillation parameter sensitivity. T2K performs multiple oscillation analyses that present both frequentist and Bayesian intervals for the Pontecorvo-Maki-Nakagawa-Sakata parameters. For fits including a constraint on $\sin^2\theta_{13}$ from reactor data and assuming normal mass ordering T2K measures $\sin^2\theta_{23} = 0.53^{+0.03}_{-0.04}$ and $\Delta m^2_{32} = (2.45 \pm 0.07) \times 10^{-3}$ $\mathrm{eV^2\,c^{-4}}$. The Bayesian analyses show a weak preference for normal mass ordering 89)% posterior probability) and the upper $\sin^2\theta_{23}$ octant (80% posterior probability), with a uniform prior probability assumed in both cases. The T2K data exclude CP conservation in neutrino oscillations at the 2σ level.

I. INTRODUCTION

The fact that neutrino flavor mixing [1] and oscillations [2] account for the apparent depletion of neutrino fluxes from natural sources is now well established by detailed observations of these sources [3–5] and verified by experiments using monitored artificial sources [6–8]. Neutrino mixing requires that at least two of the neutrino masses (m_1, m_2 and m_3) be nonzero, which in turn requires expanding upon the Standard Model. Masses require either new gauge singlets—right-handed neutrinos—or a different mass generation mechanism from other Standard Model fermions, or a combination of both. The observed pattern of neutrino masses and mixing is therefore of great interest as a window onto physics beyond the Standard Model.

A. Unanswered questions in neutrino oscillations

The generally accepted explanation of leptonic mixing and neutrino oscillation phenomena centers around the 3×3 PMNS mixing matrix, named for Pontecorvo, Maki, Nakagawa and Sakata, which describes the neutrino mass eigenstates (ν_1, ν_2 and ν_3) in terms of the weak flavor eigenstates (ν_e, ν_μ, and ν_τ). Under the assumption that neutrinos are Dirac particles, the matrix is conventionally parameterized as the product of three Tait-Bryan rotations (by θ_{ij}) and a phase transformation (by δ_{CP}), as in Eq. (6) of [9]:

*Also at INFN-Laboratori Nazionali di Legnaro, Italy.
†Also at J-PARC, Tokai, Japan.
‡Affiliated member at Kavli IPMU (WPI), the University of Tokyo, Japan.
§Also at National Research Nuclear University "MEPhI" and Moscow Institute of Physics and Technology, Moscow, Russia.
‖Also at the Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Vietnam.
¶Also at JINR, Dubna, Russia.
**Also at Nambu Yoichiro Institute of Theoretical and Experimental Physics (NITEP), Japan.
††Also at BMCC/CUNY, Science Department, New York, NY, USA.

$$U_{\text{PMNS}} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} = R(\theta_{23})U(\theta_{13}, \delta_{\text{CP}})R(\theta_{12}).$$

(1)

It is well established that all of these elements must be large, with the smallest $|U_{e3}|^2 \sim 1/45$ and the majority of elements having magnitude squared of at least $1/4$. The top row elements are well constrained by measurements of ν_e disappearance [3,8,10–12]. The $U_{\mu 3}$ element is likewise constrained by disappearance of ν_μ [4,7,13,14], but the dependence is of the approximate form $|U_{\mu 3}|^2(1 - |U_{\mu 3}|^2)$, leading to a degeneracy, commonly expressed in terms of the octant of the mixing angle θ_{23}.

Since the matrix may be complex, a wide range of values for the magnitudes of the four elements $U_{\mu 1}$, $U_{\mu 2}$, $U_{\tau 1}$ and $U_{\tau 2}$ are possible, depending on the phase parameter δ_{CP}. A purely real matrix corresponds to δ_{CP} being an integer multiple of π; any other value is manifested as violation of charge-parity (CP) symmetry in any neutrino *appearance* channel, via the Kobayashi-Maskawa mechanism [15]. The discovery of CP violation in the lepton sector is of great interest and is a major focus of current experiments [13,16]. The fact that CP violation is controlled by a single parameter means that the rate of $\bar{\nu}_e$ appearance is not independent of ν_e appearance, so studying both channels provides a test of the standard PMNS picture.

Another approximate degeneracy is in the ordering of neutrino masses. It is known that $\Delta m_{21}^2 = m_2^2 - m_1^2 > 0$ and that $|\Delta m_{31}^2| > \Delta m_{21}^2$, but the sign of Δm_{31}^2 is as yet unknown. Neutrino masses (and the corresponding eigenstates) are numbered in order of decreasing ν_e content. In the case where $m_3^2 > m_1^2$, the predominant partner of the lightest charged lepton is the lightest neutrino. As the analogous pattern is seen in the quark sector, it is known as normal ordering (NO), with inverted ordering (IO) corresponding to a light ν_3. This mass ordering (MO) degeneracy is partially lifted at higher neutrino energies by the interactions of neutrinos with matter [17,18]. This matter effect changes both the propagation eigenstates ("effective masses") of the neutrinos and their mixing with the flavor states.

These three remaining unknowns (octant, δ_{CP}, and MO) are all accessible to current-generation long-baseline neutrino experiments such as T2K, through the $\nu_\mu \to \nu_e$ appearance channel and its CP conjugate $\bar{\nu}_\mu \to \bar{\nu}_e$. Although CP violation has the most obvious significance, the general structure of the matrix may give us a window into the deeper problem of neutrino mass and a broad range of new physics. An inverted ordering would imply the lightest neutrino is relatively weakly coupled to the lightest charged lepton. Similar in character would be resolving the octant degeneracy in favor of the upper octant, as this implies ν_3 is not the predominant partner of the heaviest

charged lepton. More generally, the highly nondiagonal structure of the PMNS matrix is suggestive of an origin for neutrino masses that is separate from the electroweak physics that dominates the masses of the heavier fermions, and precision measurements of the $\overset{(-)}{\nu}_\mu \to \overset{(-)}{\nu}_e$ channel can help to determine the remaining elements.

II. THE T2K EXPERIMENT

T2K is a second-generation accelerator neutrino oscillation experiment [19], utilizing a narrow-band beam and a 295 km baseline from J-PARC in Tōkai, Ibaraki to Super-Kamiokande (commonly Super-K or just SK) in Hida, Gifu. The primary proton beam is accelerated to 30 GeV by J-PARC's main ring. In each cycle eight bunches are extracted in a single turn and directed due west at a downward angle of $-3.6°$. The intensity of the extracted beam is monitored by five current transformers that are also used to normalize the neutrino exposure between the various detectors of the experiment, while the beam profile is monitored with secondary emission monitors—most of which are removed during physics runs—and an optical transition radiation monitor. The beam power has increased over time, reaching 500 kW by the end of May 2018.

The protons impinge on a 91.4 cm graphite target to produce a secondary beam, primarily composed of pions and kaons that are focused (or defocused, according to charge) by a set of three magnetic horns [20] pulsed with a peak current of 250 kA. The focused secondary beam propagates in a 96 m helium-filled decay volume where the secondaries can decay to produce neutrinos, among other particles. A beam dump sits at the end of the decay volume, followed by a muon monitor (MUMON) which is used to check the stability of the secondary beam.

If the horn current is run in the "forward" direction, positively charged secondaries are focused, which decay to produce a beam that is primarily ν_μ. This is referred to as "forward horn current" (FHC) or neutrino mode. Alternatively the horn current can be reversed, to give a beam of primarily $\bar{\nu}_\mu$, which is referred to as "reversed horn current" (RHC) or antineutrino mode. In either case, secondary hadrons produced in the very forward direction pass through the field-free necks of the horns and contribute to a "wrong-sign" flux that is of order 1% of the intended "right-sign" component in FHC mode and order 10% in RHC mode.

The horn configuration is most effective at focusing pions with momenta between 2 and 2.5 GeV/c, resulting in a broadband neutrino flux along the beam line axis that peaks at around 1 GeV. However, the beam is directed so that its center passes roughly 4 km south of and 12 km below the Super-Kamiokande detector, equivalent to an angle $2.5°$ from the beam axis, as measured from the target. This results in a narrow-band flux at the detector that peaks more strongly at 0.6 GeV—roughly the energy of the first

oscillation maximum for this baseline. This dramatically reduces the flux of higher-energy neutrinos that can mimic ν_e signal events through neutral-current interactions.

The experiment utilizes a suite of near detectors (NDs) at a site 280 m downstream of the target to characterize the initial flux of neutrinos and their interactions. A high-mass monitoring detector, INGRID, is centered on the nominal beam axis and samples the beam out to about 1° around the beam axis, as measured from the target. This allows the intensity and direction of the neutrino beam to be monitored on a daily basis, the latter to a precision of a few centimeters.

The second near detector, ND280, is optimized for measuring interaction rates and the properties of neutrino interactions. The detector is positioned off axis at the same azimuthal angle as Super-Kamiokande to minimize the effect of any directional shift of the beam and at a distance from the beam axis to make the neutrino spectrum as similar as possible to what would be seen in SK in the absence of oscillations. Residual differences in spectra exist because the decay volume covers a significant fraction of the 280 m baseline; these are automatically accounted for in the way the ND280 data are used in the analysis. The ND280 off-axis detector provides the most relevant information on the neutrino flux and interactions but more importantly allows the analysis to form a multidimensional constraint, incorporating correlations between flux and interaction uncertainties. The ND280 additionally has a 0.2 T magnetic field to provide charge identification of detected leptons and therefore separately constrain the neutrino and antineutrino components of the beam, improving sensitivity to CP violation. This is particularly important for the RHC dataset which has a larger ($\sim$30%) background of wrong-sign events.

Neutrinos that interact in SK produce signals that can be analyzed in one of five different event categories. At T2K's beam energy, the lepton produced in a charged-current interaction will be either $\mu^{\pm}$ or $e^{\pm}$, so events are categorized as muon- or electronlike based on the pattern of Cherenkov light in the detector. SK has no magnetic field, so the sign of the horn current is used to categorize events as a proxy for neutrino versus antineutrino identification, and the kinematics give some discrimination between signal events and backgrounds. Events without a leptonlike Cherenkov ring and events where there is evidence of additional hadronic activity are not used, except for a fifth class of events which selects FHC ν_e-like events with evidence of a low-energy positron from an ejected π^+. These selections will be discussed in greater detail in Sec. X B.

III. THE OSCILLATION PROBABILITY

Following [21], and assuming three neutrino masses labeled with $i, j \in \{1, 2, 3\}$, the vacuum oscillation probability between neutrino flavor states ν_α and ν_β can be written

$$P(\nu_\alpha \to \nu_\beta) = \left| \sum_j \langle \nu_\beta | e^{-i\mathcal{H}L} | \nu_\alpha \rangle \right|^2, \quad (2)$$

where

$$\mathcal{H}L = 2\mathcal{T}_2\Delta_{21} + 2\mathcal{T}_3\Delta_{31} \quad (3)$$

and $\Delta_{ji} = \Delta m_{ji}^2 L/4E$. For states ν_α and ν_β the $\mathcal{T}_i$ take values $\mathcal{T}_i^{\alpha\beta} = U_{\alpha i}U_{\beta i}^*$ and, considering the specific case of ν_μ to ν_e oscillations, the standard parameterization can be used to derive approximate forms for the two amplitudes:

$$\mathcal{T}_3^{\mu e} = \frac{1}{2}\sin 2\theta_{13}\sin \theta_{23}e^{i\delta_{CP}} = |\mathcal{T}_3^{\mu e}|e^{i\delta_{CP}},$$

$$\mathcal{T}_2^{\mu e} \simeq \frac{1}{2}\sin 2\theta_{12}\cos \theta_{13}\cos \theta_{23}. \quad (4)$$

The appearance probability in vacuum can then be approximated as a sum of three terms:

$$P(\nu_\mu \to \nu_e) \simeq 4|\mathcal{T}_3^{\mu e}|^2\sin^2\Delta_{31} + 4|\mathcal{T}_2^{\mu e}|^2\sin^2\Delta_{21}$$
$$+ 8|\mathcal{T}_2^{\mu e}||\mathcal{T}_3^{\mu e}|\sin \Delta_{31}\sin \Delta_{21}\cos(\Delta_{32} + \delta_{CP}), \quad (5)$$

which includes

(i) a dominant "atmospheric" ($\sin^2\Delta_{31}$) term, which is independent of the CP hase;

(ii) a "solar" ($\sin^2\Delta_{21}$) term that is small at T2K's L/E; and

(iii) an "interference" term that depends on $\sin \Delta_{31}\sin \Delta_{21}\cos(\Delta_{32} + \delta_{CP})$.

Because the atmospheric term is proportional to $|U_{\mu3}|^2|U_{e3}|^2$, the ν_e appearance probability is free of the octant degeneracy seen in the ν_μ disappearance probability. In practice the ν_μ channel remains an important part of the fit, because there is a larger number of events, which gives higher precision on Δm_{3i}^2 and $\sin \theta_{23}$. The disappearance channel is also important because it is relatively insensitive to $\sin \theta_{13}$, δ_{CP}, and mass ordering, which helps to reduce the impact of degeneracies in the appearance channel.

The amplitude of the interference term is about 20% of the atmospheric term, making it possible to measure the phase δ_{CP}. This interference term is of particular physical importance as the relative sign of δ_{CP} and Δ_{ji} changes between neutrinos and antineutrinos, leading to CP violation if δ_{CP} is not an integer multiple of π. Since T2K's event spectrum peaks very close to the oscillation maximum at $\Delta_{32} \sim \pi/2$, the contribution from the interference term is a direct measure of the amount of CP violation in the neutrino sector.

The interference term also depends directly on $\text{sgn}(\Delta m_{31}^2)$, through the $\sin \Delta_{31}$ part. Up to a small difference between Δm_{31}^2 and Δm_{32}^2, this is degenerate with a

substitution $\delta_{\text{CP}} \to \pi - \delta_{\text{CP}}$. This means the same data will prefer opposite signs of $(\delta_{\text{CP}} - \pi/2)$ for normal and inverted orderings. This degeneracy is lifted if the mass ordering is assumed, or if one or other ordering is strongly preferred in the fit.

A. Matter effects

The (approximately constant) density of electrons in the Earth's crust along the T2K baseline changes [22] the propagation Hamiltonian [cf. Eq. (3)]:

$$\mathcal{H}L = 2\mathcal{T}_2\Delta_{21} + 2\mathcal{T}_3\Delta_{31} \pm \text{diag}(1, 0, 0)\sqrt{2}G_{\text{F}}n_e, \qquad (6)$$

where n_e is the number density of electrons, G_{F} is the Fermi coupling constant, and the minus sign applies for antineutrinos. The impact can be understood using T2K's beam energy of ~ 0.6 GeV and a crust density [23] of 2.6 g cm^{-3}, in which case

$$\underset{250}{\Delta m_{31}^2} \gg \underset{11.9}{2\sqrt{2}G_{\text{F}}n_e E} > \underset{7.4}{\Delta m_{21}^2}. \underset{(\times 10^{-5} \text{ eV}^2)}{} \qquad (7)$$

Because of the size ordering of terms in the Hamiltonian, the matter effect dominates the solar term but is only a small perturbation compared to the atmospheric and interference terms. In practice, because the resulting oscillation probability is dominated by the atmospheric term, T2K's sensitivity to the mass ordering comes mostly from this perturbative effect on $|U_{e3}|$ and Δm_{31}^2. In this regime the matter effect can be described with a dimensionless parameter

$$\xi(E) = \pm\frac{2\sqrt{2}G_F n_e}{\Delta m_{31}^2} E \simeq \pm 0.08 E/\text{GeV}, \qquad (8)$$

which is positive for neutrinos in the normal ordering and for antineutrinos in the inverted ordering. The resulting probability for neutrinos or antineutrinos is

$$P(\overset{(-)}{v}_\mu \to \overset{(-)}{v}_e)$$

$$\simeq 4|\mathcal{T}_3^{\mu e}|^2\frac{\sin^2([1-\xi]\Delta_{31})}{[1-\xi]^2} + 4|\mathcal{T}_2^{\mu e}|^2\frac{\sin^2(\xi\Delta_{21})}{\xi^2}$$

$$+ 8|\mathcal{T}_2^{\mu e}||\mathcal{T}_3^{\mu e}|\frac{\sin([1-\xi]\Delta_{31})}{1-\xi}\frac{\sin(\xi\Delta_{21})}{\xi}\cos\Delta_{32}\cos\delta_{\text{CP}}$$

$$\mp 8|\mathcal{T}_2^{\mu e}||\mathcal{T}_3^{\mu e}|\frac{\sin([1-\xi]\Delta_{31})}{1-\xi}\frac{\sin(\xi\Delta_{21})}{\xi}\sin\Delta_{32}\sin\delta_{\text{CP}},$$

$$(9)$$

where the CP-violating $\sin\delta_{\text{CP}}$ term takes a negative sign for neutrinos and positive sign for antineutrinos.

For T2K, the largest observable effect from propagation in matter is the $[1-\xi]^{-2}$ scaling of the atmospheric term.

This modification to the leading atmospheric term is about 5%, leading to a roughly 10% difference in the appearance probability for neutrino and antineutrinos. Since this difference is about half the amplitude of the CP-violating term, it is in general difficult to disentangle the two phenomena if they have opposite effects on the total number of events observed, or if the value of $\sin\delta_{\text{CP}}$ is close to zero. However if both phenomena enhance (or both suppress) the total number of events, then the net effect can be too large to attributable to either source alone, and there will be much less ambiguity.

B. The survival probability

In the same notation, the $\overset{(-)}{v}_\mu$ survival probability at T2K is to a good approximation given by

$$P(\overset{(-)}{v}_\mu \to \overset{(-)}{v}_\mu) \simeq 1 - 4\mathcal{T}_3^{\mu\mu}(1 - \mathcal{T}_3^{\mu\mu})\sin^2\Delta_{\text{Atm}}, \qquad (10)$$

where $\Delta_{\text{Atm}} = \Delta_{32} + \Delta_{21} \times \mathcal{T}_1^{\mu\mu}/(1 - \mathcal{T}_3^{\mu\mu})$. So although the observable survival probability is not sensitive to the mass ordering, the best-fit value of Δ_{32} is different for normal and inverted orderings. As for the oscillation amplitude, in terms of the standard parameterization

$$|\mathcal{T}_3^{\mu\mu}| = \sin^2\theta_{23}\cos^2\theta_{13}, \qquad (11)$$

so the amplitude reaches a maximum value around $\sin^2\theta_{23} = 1/(2\cos^2\theta_{13})$, and values on either side of this are degenerate.

Propagation in matter does not change the survival probability by much; matter-dependent effects are suppressed by a factor of $\mathcal{T}_3^{ee}(1 - 2\mathcal{T}_3^{\mu\mu})$ [23]. Since the matter density can be approximated as symmetric, the probability also has no dependence on $\sin\delta_{\text{CP}}$ [24]. However the relationship between Δ_{32} and the observable Δ_{Atm} does depend on $\cos\delta_{\text{CP}}$ through $\mathcal{T}_1^{\mu\mu}$, which can for some parameter combinations give rise to a correlation between the measured Δ_{32} and δ_{CP}.

IV. UPDATES SINCE THE PREVIOUS RESULTS

This analysis uses a SK dataset collected up to the end of May 2018. This corresponds to an exposure of 14.94×10^{20} protons on target (POT) in FHC mode and 16.35×10^{20} POT in RHC mode, the same as used to report indications of CP violation in [16]. A detailed breakdown is given in Table I. Compared to the previous update [25] this is a nominal increase of 1% in FHC mode, but 116% in RHC mode, which is particularly of interest for indications of $\bar{\nu}_e$ appearance, described in Sec. XV.

In parallel with the statistical increase, our event selection has been refined since it was last described in detail [26]. Event reconstruction is now based on an algorithm that matches the pattern of light observed in SK directly [27]. This makes use of more information about the event,

TABLE I.   T2K run periods and exposure used in this analysis, for ND280 and SK.

| T2K run | End date | SK POT/10^{20} | | ND280 POT/10^{20} | |
		FHC	RHC	FHC	RHC
Run 1	June 2010	0.326	⋯	⋯	⋯
Run 2	Mar 2011	1.122	⋯	0.78	⋯
Run 3	June 2012	1.599	⋯	1.56	⋯
Run 4	May 2013	3.597	⋯	3.47	⋯
Run 5	June 2014	0.244	0.512	⋯	0.43
Run 6	June 2015	0.192	3.546	⋯	3.43
Run 7	May 2016	0.484	3.498	⋯	⋯
Run 8	Apr 2017	7.169	⋯	⋯	⋯
Run 9	May 2018	0.204	8.788	⋯	⋯
Total		14.938	16.346	5.80	3.86

providing better discrimination between event categories and improving the resolution of the lepton momentum and vertex location. As a result, the fiducial volume can also be expanded, roughly equivalent to a 20% increase in statistics for the ν_e samples, as described in Sec. X. The newer reconstruction algorithm has previously been used for rejecting neutral-current π^0 events in the ν_e samples [28,29] and for all aspects of event selection and reconstruction in more recent publications [16,25].

A large fraction of the analysis development focuses on the interaction model, which incorporates constraints from a number of new external datasets and theoretical improvements. Since reported in [26], the dominant charged-current quasielastic (CCQE) models have been updated in various respects, including the handling of weak charge screening in nuclei; the handling of nucleon removal energy and its effect on lepton kinematics; and additional freedom allowed in the kinematic dependence of interactions involving correlated nucleon pairs ($2p2h$). Modeling of (and uncertainties assigned to) subdominant processes have also been improved, including coherent scattering and neutral-current interactions.

Constraints on neutrino oscillations and the associated parameters come from a combined analysis of disappearance and appearance channels across FHC and RHC configurations, using the same approach as in [16]. This paper provides a fuller description of the method and a broader range of results.

V. ANALYSIS OVERVIEW

The T2K near and far detectors have different target nuclei and are based on different particle detection techniques. The T2K oscillation analysis therefore uses parameterized models of the neutrino beam flux and the neutrino interaction cross section to propagate near detector information to predict the far detector event rate.

The neutrino flux prediction is described in Sec. VI and is parameterized as a function of neutrino energy, neutrino species and whether the beam is operating in FHC or RHC mode. The prediction and its uncertainty are generated for both the near and far detector simultaneously. This correlates the near and far detector fluxes, allowing the near detector data to constrain the far detector flux.

The T2K neutrino interaction model is described extensively in Sec. VII. The model incorporates a number of tunable parameters whose prior uncertainties and nominal values come from comparisons to electron and neutrino scattering datasets. There are a number of models that agree with existing data equally well, and it is not always possible to parameterize the differences between these models. In this case simulated data studies have been performed using these alternate models, as described in Sec. XIII.

The neutrino flux and interaction models are fit to data collected by the T2K off-axis near detector, ND280 [19]. The fit varies the parameters within both models simultaneously to best match the ND280 data, as described in Sec. XI. This results in tuned flux and interaction models with correlated uncertainties, providing a more accurate and precise prediction of the event rate at Super-Kamiokande.

At the far detector a simultaneous fit of muonlike and electronlike samples from both neutrino and antineutrino beams is used to constrain the PMNS oscillation parameters. The conventional $\{\theta_{ij}, \delta_{CP}\}$ parameterization is used, enforcing unitarity, and the effect of propagation in matter is included. Data from ν_e and $\bar{\nu}_e$ disappearance experiments are used to constrain the parameters (θ_{12} and Δm^2_{21}) [30] that T2K has little sensitivity to. Fits are performed both with and without an external constraint on θ_{13} [31]. Systematic uncertainties are treated by a numerical marginalization technique: all parameterized uncertainties are randomly sampled many times according to prior constraints, including ND280 data, and the likelihoods averaged over the ensemble. This process is described in detail in Sec. XII, with the results described in Sec. XIV.

VI. NEUTRINO FLUX MODELING

The neutrino flux prediction has been described in detail in Ref. [32]. The collision of 30 GeV protons from the
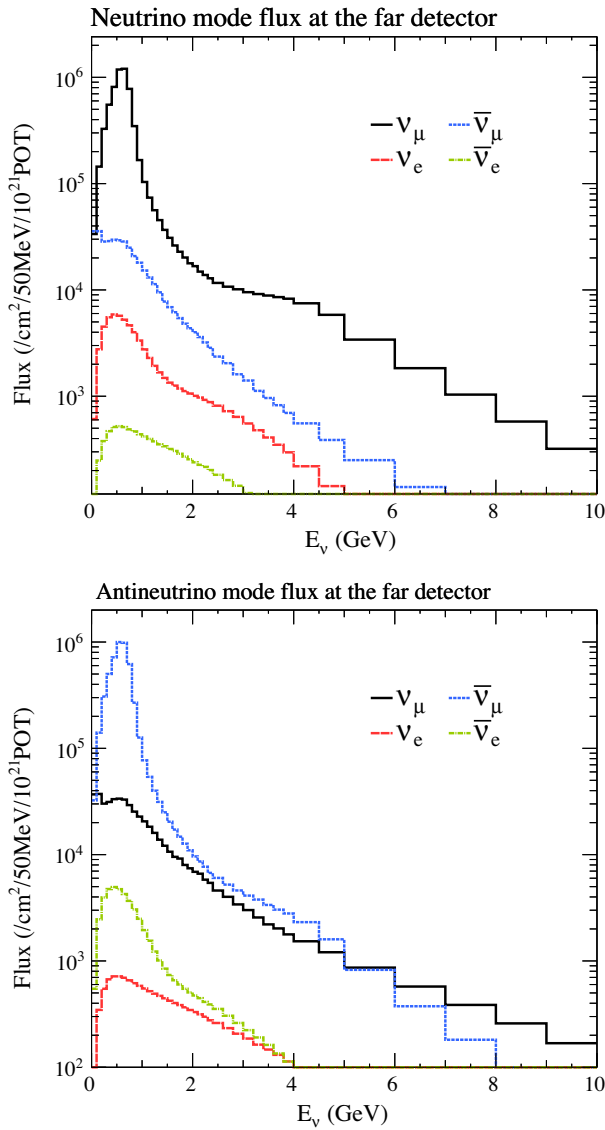
FIG. 1. The SK flux prediction for runs 1–9a with horns operating in FHC (250 kA) mode (upper) and RHC (−250 kA) mode (lower).
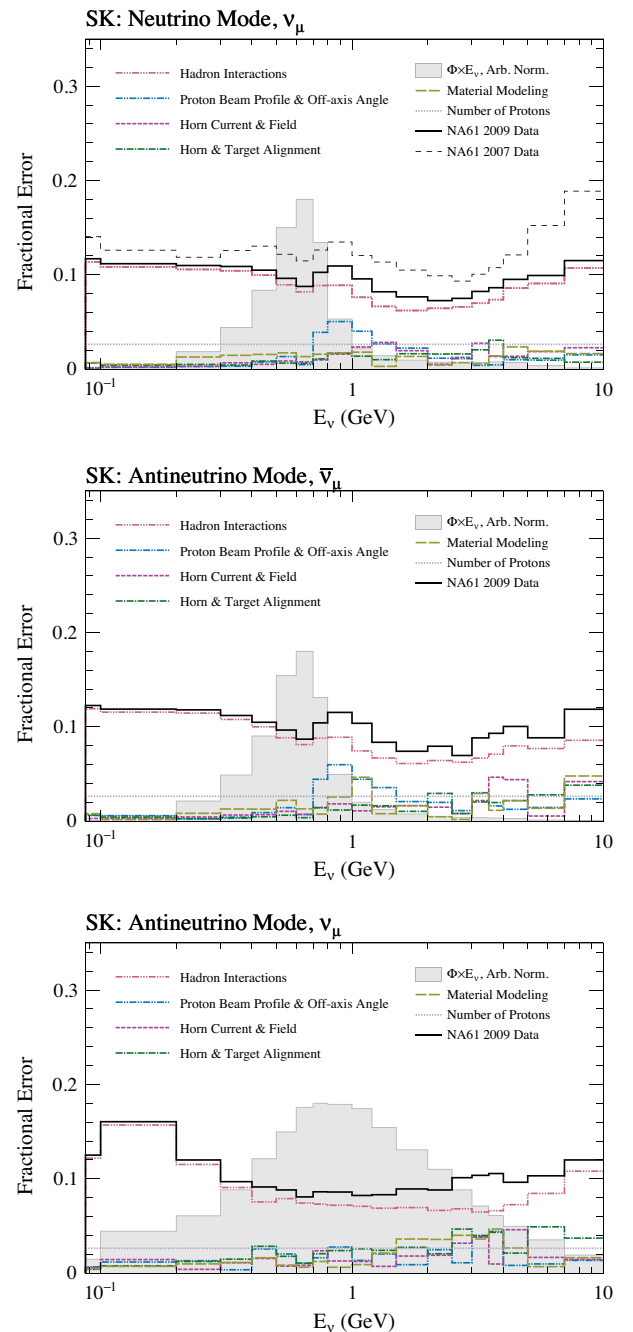

FIG. 2. The fractional systematic uncertainty on the ν_μ flux at SK in FHC mode (top), on the right-sign $\bar{\nu}_\mu$ flux at SK in RHC mode (middle), and on the wrong-sign ν_μ flux at SK in RHC mode (bottom). The solid black line shows the current total fractional uncertainty (NA61/SHINE 2009 data), while the dashed black line in the top panel shows the fractional uncertainty from an earlier flux prediction (NA61/SHINE 2007 data).

J-PARC main ring with the T2K neutrino production target is simulated using FLUKA [33–35]. The resultant secondary particles are passed to a GEANT3 [36] simulation of the magnetic focusing horns and decay volume downstream of the target. GCALOR [37,38] is used to model hadronic interactions of the secondary particles as they traverse the focusing horns and decay volume. Particles are then allowed to decay to produce neutrinos.

Data from proton beam monitors is used to tune the initial proton beam parameters in the simulation. NA61/ SHINE, a fixed-target experiment at CERN's Super Proton Synchrotron, measures particle production in nucleus and hadron collisions with a large acceptance spectrometer. This includes measurements of the collisions of 30 GeV protons with graphite. Data from the NA61/SHINE [39–42]

experiment are then used to tune the secondary particles produced from the target. Finally, the INGRID [43] on-axis near detector is used to monitor the neutrino beam direction. The uncertainties from each of these measurements are

combined with uncertainties from the beam simulation to give the final flux uncertainty.

Figure 1 shows the predicted neutrino fluxes at SK for both the FHC and RHC modes. Previous T2K flux estimates [32] used thin target hadron production data collected in the NA61/SHINE experiment in 2007 [39–41], where a 31 GeV proton beam impinged upon a graphite target with a thickness of 4% of a nuclear interaction length (the so-called *thin* target). The work presented here uses an updated flux prediction based on higher statistics thin target hadron production data collected in the NA61/SHINE experiment in 2009 [42], including the yields of π^+, π^-, K^+, K^-, K_s^0, Λ and p. Future analyses will include NA61/SHINE hadron production measurements on a replica of the T2K neutrino production target.

Figure 2 shows the fractional uncertainty on the ν_μ flux at SK in FHC mode and on the right-sign $\bar{\nu}_\mu$ and wrong-sign ν_μ fluxes at SK in RHC mode. The improvement obtained by including the 2009 NA61/SHINE thin target data is also indicated.

VII. NEUTRINO INTERACTION MODELING

Oscillation parameter values are inferred from spectra of observable quantities, herein either reconstructed charged-lepton kinematics, (p_ℓ, θ_ℓ), or reconstructed neutrino energy. The reconstructed neutrino energy is estimated from final-state charged-lepton kinematics only as

$$E_{\mathrm{QE}}^{\mathrm{Rec}}(p_\ell, \theta_\ell) = \frac{2M_{N,i}E_\ell - M_\ell^2 + M_{N,f}^2 - M_{N,i}^2}{2(M_{N,i} - E_\ell + p_\ell \cos\theta_\ell)}, \quad (12)$$

where $M_{N,i}$, $M_{N,f}$, and M_ℓ are the mass of the initial-state nucleon (an effective, off-shell mass that includes the "nucleon removal energy" is often used), final-state nucleon, and final-state charged lepton, respectively; E_ℓ, p_ℓ, and θ_ℓ are the energy, three-momentum, and angle of the final-state charged lepton, respectively. $E_{\mathrm{QE}}^{\mathrm{Rec}}$ provides a smeared but minimally biased estimate of the neutrino energy for quasielastic neutrino scattering off bound nucleons (CCQE). For other interaction channels, such as those that produce extra hadrons, $E_{\mathrm{QE}}^{\mathrm{Rec}}$ underestimates the energy of the interacting neutrino.

The procedure of inferring oscillation parameter values from observable quantities implicitly relies on an accurate understanding of the rate of background processes and a mapping between true energy and observable quantities, e.g., $E_{\mathrm{QE}}^{\mathrm{Rec}}(E_\nu)$, both of which are derived from simulation. As a result, accurate neutrino interaction modeling is critical. Event selections are trained on the simulated distribution of final-state particles and the predicted rate of various signal and background processes. The predicted rate of a number of neutrino interaction processes, which exhibit different true-to-observable mappings, is constrained by near detector data and then used to interpret the observed far detector data. This section briefly describes the neutrino interaction model, its associated uncertainties, and specific studies used to test resilience to known weaknesses of the model.

A. The base interaction model

The samples of simulated neutrino interactions used in this analysis were made with version 5.3.3 of the NEUT interaction generator [44]. NEUT simulates known neutrino interaction channels relevant for few GeV neutrinos; these channels are broadly categorized as 1p1h, 2p2h, single-pion production, and deep inelastic scattering (DIS). In addition to the "primary" interaction channels, the effect of using nuclear targets, where the struck nucleons are bound within a nuclear potential, needs to be modeled well. These effects can be separated into initial-state and final-state effects. Most updates to the interaction model since the previous analysis [26] are in the treatment of systematic uncertainties; however, a short description of the whole model is included here for completeness. As the "base" model has not changed, the interested reader is directed to Refs. [26,45] for a discussion of the motivations behind any specific model choices.

1. Initial-state nuclear effects

Nucleons bound within a nuclear potential undergo non-negligible "Fermi motion." For carbon, this means bound nucleons have a momentum of $p_{\mathrm{f}} \lesssim 217$ MeV/c, or equivalently, a Fermi energy of $E_{\mathrm{f}} \lesssim 25$ MeV. A global relativistic Fermi gas (GRFG) is used to model the initial-state nucleon momentum distribution in this analysis. Neutrino interactions with bound nucleons are largely handled under the impulse approximation, whereby a single "struck" nucleon receives a four-momentum kick while the rest of the target nucleus acts as a group of noninteracting "spectator" nucleons. This rudimentary nuclear model is a simple approximation for the correct modeling of the initial nucleon momentum distribution and nucleon removal energy; a study, presented below, accounts for the effect of this approximation.

2. 1p1h

One-particle, one-hole interactions are those where the neutrino interacts quasielastically with a single bound nucleon—the interaction is only *quasielastic* because of the bound nature of the target nucleon and, for charged-current events (CCQE), the initial-to-final-state charged-lepton and nucleon rest mass difference. Such interactions are modeled in the Llewellyn-Smith formalism [46], using the BBBA05 [47] description for the vector part of the nucleon form factors and a simple dipole form for the axial part. The NEUT model includes two additional features of note: the nucleon removal energy and in-medium modifications to the W-boson propagator via the random phase
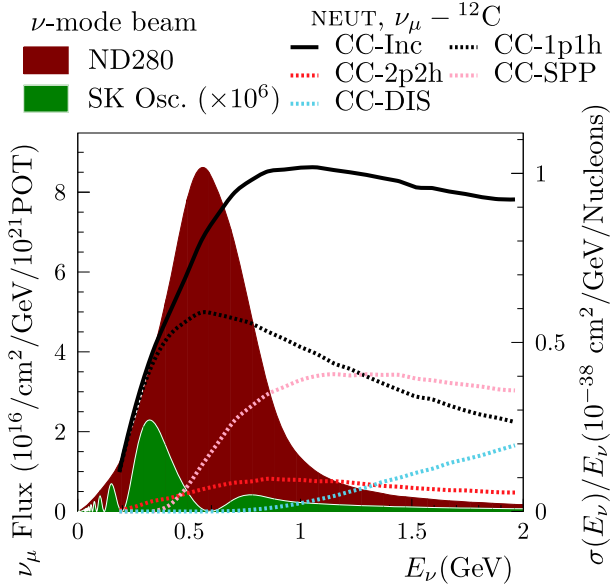
FIG. 3. The total charged-current cross section for muon neutrinos interacting with a carbon nucleus, as predicted by NEUT, overlaid on the ND280 muon neutrino flux, and an example oscillated muon neutrino flux at SK. The oscillation parameters used here are the best fit from the previous analysis [26]. The total (Inc) cross section is separated into 1p1h, 2p2h, single-pion production (SPP), and DIS channels.

approximation (RPA). Variations in the average nucleon removal energy modify the predicted kinematics of final-state particles, most importantly charged leptons. When comparing predictions based on Fermi gas nuclear models to 1p1h-like cross-section data, a suppression at low four-momentum transfer is favored relative to the free-nucleon-target calculation [48]. This is often attributed to a weak-charge screening effect as a result of the nuclear medium [49]. The effect is termed "RPA" after the random phase approximation technique used to sum up the series of contributing W-boson self-energy diagrams. Here, the distribution of four-momentum transfer is modified by the RPA calculation from Nieves *et al.* [49]. As can be seen in Fig. 3, 1p1h is the dominant interaction channel at T2K energies.

3. 2p2h

Two-particle, two-hole interactions are an inherently nuclear-target process, whereby the incoming neutrino interacts with a bound pair of nucleons, knocking both out of the nuclear potential. The Nieves *et al.* model [50] is used to predict the cross section as a function of lepton kinematics. While this process is subdominant, it produces observable final states that are indistinguishable from 1p1h interactions in the T2K detectors, but with different observed lepton kinematics as a function of neutrino energy. In the Nieves *et al.* 2p2h model, there are two distinct regions of strength in the energy and momentum
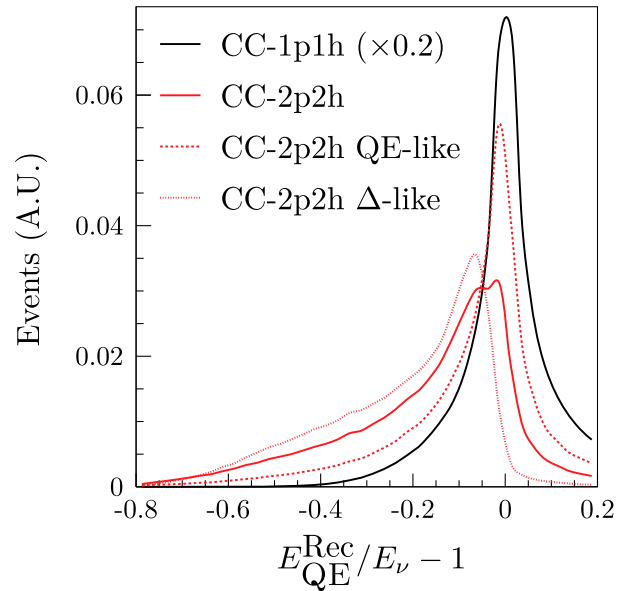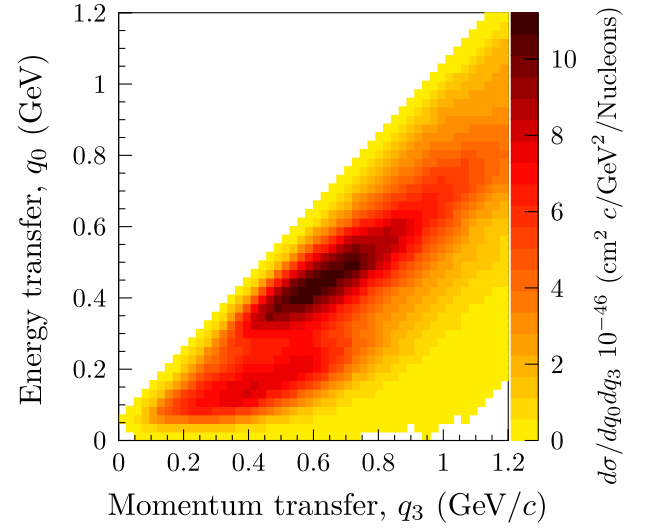
FIG. 4. Top: the energy and momentum transfer distribution for the Nieves *et al.* 2p2h model in NEUT. The two-peak structure is clear, with QE-like kinematics corresponding to the lower left peak and Delta-like kinematics to the stronger central peak. Bottom: the reconstructed energy bias at SK is shown for 1p1h and 2p2h events for an oscillated muon neutrino flux. The different reconstructed energy smearing for 2p2h events with QE-like and Delta-like interaction kinematics can be seen.

transfer space: the quasi-elastic-like (energy transfer, $q_0 \lesssim 0.3$) and Delta-like regions ($q_0 \gtrsim 0.3$). The energy-momentum transfer distribution and the corresponding $E_{\mathrm{QE}}^{\mathrm{Rec}}$ biases can be seen in Fig. 4.

4. Single-pion production

Single-pion production can be separated into three subprocesses: resonant, nonresonant, and coherent single-pion production. The resonant and nonresonant processes describe the production of a pion involving neutrino

scattering of a single nucleon, either via an intermediate baryon resonance (resonant) or not (nonresonant). These processes are modeled in the Rein-Sehgal formalism [51], with an improvement that includes the effect of the final-state charged-lepton mass [52], and updated nucleon axial form factors from Graczyk and Sobczyk [53]. The contributions from 17 baryon resonances are considered, with the $\Delta(1232)$ being dominant, and interference terms between the resonances are taken into account. The nonresonant channel augments the production of half-unit isospin final states (e.g., $\nu + n \rightarrow \ell^- + p^+ \pi^0$ and $\nu + n \rightarrow \ell^- + n\pi^+$); any interference between the resonant and nonresonant contributions is ignored. These processes are used to model final states with an invariant hadronic mass of $W \leq 2.0$ GeV/c^2. The modeling of the so-called "transition region" between single-pion production off a nucleon and shallow- and deep-inelastic scattering is an unsolved theoretical problem [54]. In the NEUT model, the region $1.3 \leq W \leq 2.0$ GeV/c^2 contains contributions from both the Rein-Sehgal single-pion model, described above, and the deep-inelastic-scattering model, described below. For higher invariant masses the deep-inelastic-scattering model is used.

The axial form factors and the strength of the nonresonant contribution in the Rein-Sehgal model were tuned to published cross-section data using the NUISANCE framework [55]. As these parameters control only the nucleon-level interaction, the central values were determined from fits to deuterium-target bubble chamber data, which are largely free from nuclear effects. Data from Argonne National Laboratory [56] (with some reanalyzed distributions taken from Ref. [57]) and Brookhaven National Laboratory [58] were used. The uncertainties determined from the fits to bubble chamber data were then inflated to approximately cover cross-section data from MiniBooNE [59] and MINERνA [60,61].

Coherent single-pion production describes the interaction of a neutrino coherently with a whole nucleus. This is a subdominant pion production process, observed at low energy for the first time by the MINERνA experiment [62], and is characterized by very little four-momentum transfer to the struck nucleus. We follow the preference of the MINERνA data and use the Berger-Sehgal model [63].

5. Deep inelastic scattering

For interactions producing hadronic systems with two or more pions and invariant hadronic masses of $W > 1.3$ GeV/c^2, the cross section is constructed from nucleon structure functions that depend on the Bjorken scaling variables x and y. The structure functions are calculated from the GRV98 [64] parton distribution functions, with modifications from Bodek and Yang [65] to account for the relatively low-momentum transfers

involved. For interactions with $1.3 < W \leq 2.0$ GeV/c^2 the hadronic state is generated by a custom multipion production model; above $W = 2.0$ GeV/c^2 PYTHIA 5.72 is used [66].

6. Final-state nuclear effects

After the *primary* neutrino interaction has been simulated, a number of additional "nuclear effects" are included. For interactions that produce a final-state proton or neutron, the Pauli exclusion principle is applied, rejecting any events that produce a nucleon below the Fermi energy. This results in a suppression at low four-momentum transfer for 1p1h events. Final-state hadrons produced at the neutrino interaction vertex are stepped through the nuclear medium in a classical cascade, in which they may interact and produce secondary particles, be absorbed, or undergo charge exchange (e.g., $\pi^+ + n \rightarrow \pi^0 + p$). Such reinteractions are often called "final-state interactions," or FSIs. Finally, after the primary interaction and hadronic cascade, the remnant nucleus can be left in an excited state that will subsequently decay. For interactions on oxygen, nuclear de-excitations that result in secondary, low-energy photons [$\mathcal{O}(1–10)$ MeV] are modeled following Ref. [67].

With the exception of 2p2h interactions, these channels and effects are also implemented for neutral-current interactions, but the details are not repeated here for brevity. The total charged-current cross sections, broken down by interaction channel, are shown in Fig. 3.

B. The uncertainty model

As the number of observed events included in the analysis grows with exposure, a robust interaction uncertainty model is required to assess the significance of the results. The uncertainty model for 1p1h and 2p2h interactions has seen recent improvements and will be discussed in detail here. For details on other, unchanged sources of interaction uncertainty see Ref. [26].

1. 1p1h

The NEUT 1p1h interaction model implements three main sources of uncertainty: the mass used in the dipole axial form factor (M_A^{QE}), the effect of RPA on the cross section as a function of four-momentum transfer, and the nucleon removal energy associated with scattering off a bound nucleons.

In this analysis, M_A^{QE} does not have a prior uncertainty and is constrained by near detector data alone. The parameterization of the uncertainty on the RPA suppression has been updated in this analysis; the previous implementation proved problematic because variations of different free parameters effected a similar response in Q^2. For this analysis, Bernstein polynomials were used to model the
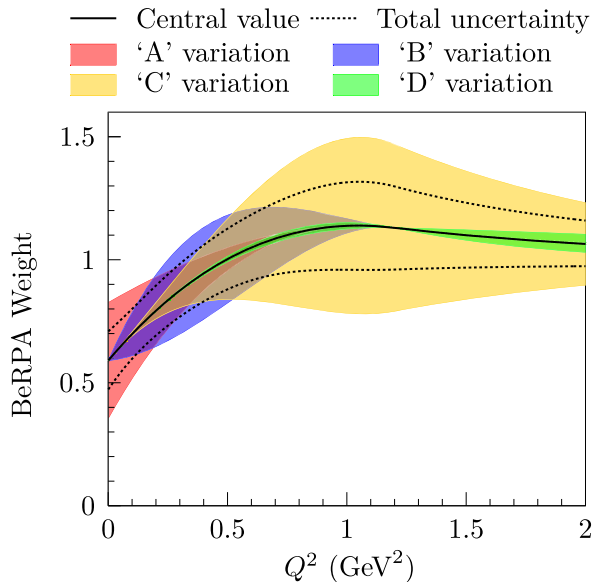
FIG. 5. The central value BeRPA suppression factor (solid line) and the total prior uncertainty (dashed lines) determined from 10^5 uncorrelated throws of the free parameters, overlaid on the corresponding envelopes generated by varying each parameter. The continuity conditions result in a perhaps unintuitive total uncertainty envelope given the individual parameter variations.

shape below some Q^2 cutoff U, above which an exponential decay form is used:

$$f(x) = \begin{cases} A(1-x')^3 + 3B(1-x')^2 x' + 3p_1(1-x')x'^2 + Cx'^3, & x < U, \\ 1 + p_2 \exp(-D(x-U)), & x > U, \end{cases}$$

where $x = Q^2$ and $x' = Q^2/U$. To ensure continuity at $Q^2 = U$, the conditions

$$p_1 = C + \frac{UD(C-1)}{3},$$
$$p_2 = C - 1$$

are enforced, leaving four free parameters, A, B, C, and D. The fifth, U, is kept fixed at 1.2 GeV^2. This parameterization has been termed "BeRPA" after the Bernstein polynomials on which it is based. The effect of varying each of the four free parameters relative to the theoretical uncertainty calculated by following Ref. [68] can be seen in Fig. 5. Together, the four free BeRPA parameters and the M_A^{QE} parameter give effective freedom over a range of Q^2. The Q^2 distribution is then largely constrained by the fit to near detector samples.

2. 2p2h

The details of the 2p2h process are highly uncertain. The total cross section, evolution with neutrino energy, and energy and momentum transfer characteristics of the process are all predicted differently by the available models (e.g., Nieves *et al.* [50], Martini *et al.* [69], SUSAv2-MEC [70], and GiBUU [71]). While data from the T2K near detector [72–74], MINERνA [75,76], and NOνA [77] favor a process with similar interaction kinematics to such a multinucleon process,[1] experimental sensitivity to this exclusive channel is weak. As a result, significant freedom is afforded to the 2p2h process in this analysis.

The uncertainty on the 2p2h process is separated into normalization and shape components. An overall 100% normalization uncertainty is separately assigned to interactions involving neutrinos and those involving antineutrinos. An additional parameter that introduces freedom in the relative normalization of 2p2h interactions with carbon and oxygen nuclei is used. Finally, a parameter that varies the relative strength of the QE-like and Delta-like components, while keeping the total cross section for 2p2h events constant, is introduced; the effect of extreme variations of this parameter is seen in Fig. 6.

C. Simulated data studies

It is strongly suspected that the described uncertainty model may not cover all differences between nature and the interaction model described above. To begin to address this, we perform fits of the model to targeted "simulated datasets" that test the robustness of the model and associated uncertainties to known missing features. In some cases the results of these studies are used to motivate additional uncertainties. This section introduces the simulated datasets that were analyzed to address specific concerns.

1. Alternative 1p1h nuclear models

The GRFG used to model the nuclear initial state is a simple model that contains no correlations between initial momentum and nucleon removal energy (NRE). Such correlations may be important for correctly modeling the observed charged lepton spectrum [78] and are seen in nuclear response measurements from electron scattering experiments. To test the robustness of the implemented uncertainty model to such details, two simulated datasets are used: the Nieves *et al.* 1p1h model and the spectral function model of Benhar *et al.* [78] (BSF). Both contain some correlation between the initial momentum and the NRE. The Nieves *et al.* model differs from the base model by implementing a local, rather than a global, Fermi gas (LFG), in which the concept of a radially dependent nuclear density profile introduces such correlations. In the BSF model, initial nucleons are chosen from a full, two-dimensional nuclear response distribution, which is constructed from $(e, e'p)$ data [78]. It should be noted that the BSF

---

[1]i.e., at fixed momentum transfer, a process that occurs between the quasielastic and pion production peaks in energy transfer.
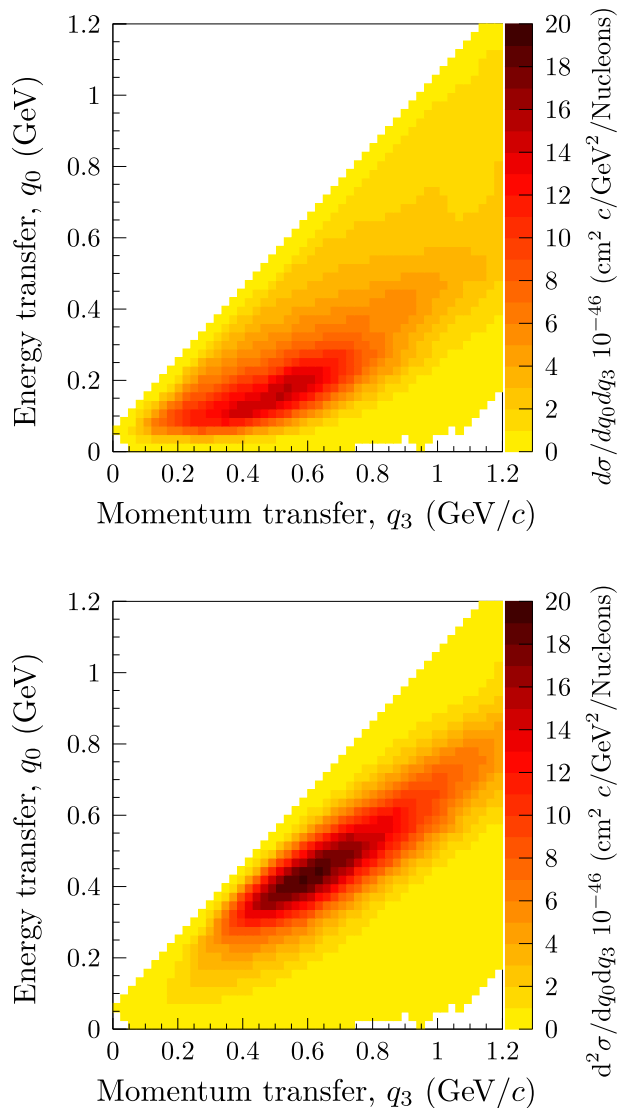
FIG. 6.   The differential cross section for the two extreme variations of the 2p2h "shape" parameter—QE-like (top) and Delta-like (bottom)—cf. Fig. 4 (top) for the central-value-predicted cross section.
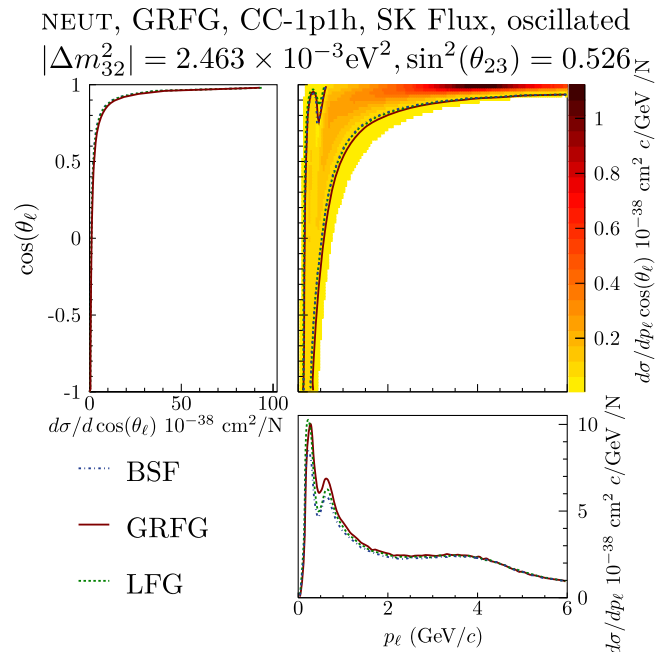

FIG. 7.   The predicted flux-averaged cross section for three different nuclear response models. The different models result in different predictions of the observed final-state lepton kinematics for the same oscillated neutrino flux. If such variations are not accounted for in the uncertainty model, extracted oscillation parameters may be biased. The contours contain the region of phase space with a differential cross section $d\sigma/dp_\ell \cos(\theta_\ell) > 0.05 \times 10^{-38}$ cm^2 c/GeV/N.

model contains no "RPA"-like, low four-momentum-transfer suppression effect. In constructing the simulated data, only the 1p1h cross section is modified. The predicted final-state muon kinematics for each model are shown in Fig. 7.

2. Nucleon removal energy

The implementation of the NRE parameter was revised for this analysis. The previous implementation relied on calculating the change in the predicted differential cross section for a variation NRE $\rightarrow$ NRE$'$, $\sigma^{\mathrm{CCQE}}(E_\nu, p_\ell, \theta_\ell, \mathrm{NRE})$; this proved problematic as variations of NRE modify the available kinematic phase space for the production of final-state muons. For some simulated interactions, $(E_\nu, p_\ell, \theta_\ell)$, and

binding-energy variation, NRE$'$, the variation weight, $w = \sigma^{\mathrm{CCQE}}(E_\nu, p_\ell, \theta_\ell, \mathrm{NRE}')/\sigma^{\mathrm{CCQE}}(E_\nu, p_\ell, \theta_\ell, \mathrm{NRE})$, can be ill defined as the denominator becomes vanishingly small. Instead, an effective implementation was used that shifts the final-state charged-lepton momentum in response to variations of NRE. The momentum shifts were calculated in bins of true neutrino energy and true final-state charged-lepton polar angle—in the NEUT implementation, variations of the binding energy effect only small changes in the final-state lepton angular distribution. An example of such a variation can be seen in Fig. 8. The prior uncertainty on the new NRE parameterization was taken as 18 MeV—this large uncertainty is motivated in part because of implementation choices in NEUT [79] and in part because of uncertainties on the analyzed electron-scattering measurements.

The NRE parameter uncertainty was not included in the near detector fit for this analysis. Instead, an extremal variation based on the results in Ref. [79] was included as a simulated dataset.

3. Martini et al. 2p2h

As previously mentioned, the modeling of 2p2h interactions is highly uncertain. We include a simulated dataset based on an alternate 2p2h calculation by Martini et al. [69]. This calculation predicts a larger inclusive 2p2h cross

NEUT, GRFG, CC-1p1h, SK Flux, oscillated

$|\Delta m^2_{32}| = 2.463 \times 10^{-3} \text{eV}^2, \sin^2(\theta_{23}) = 0.526$
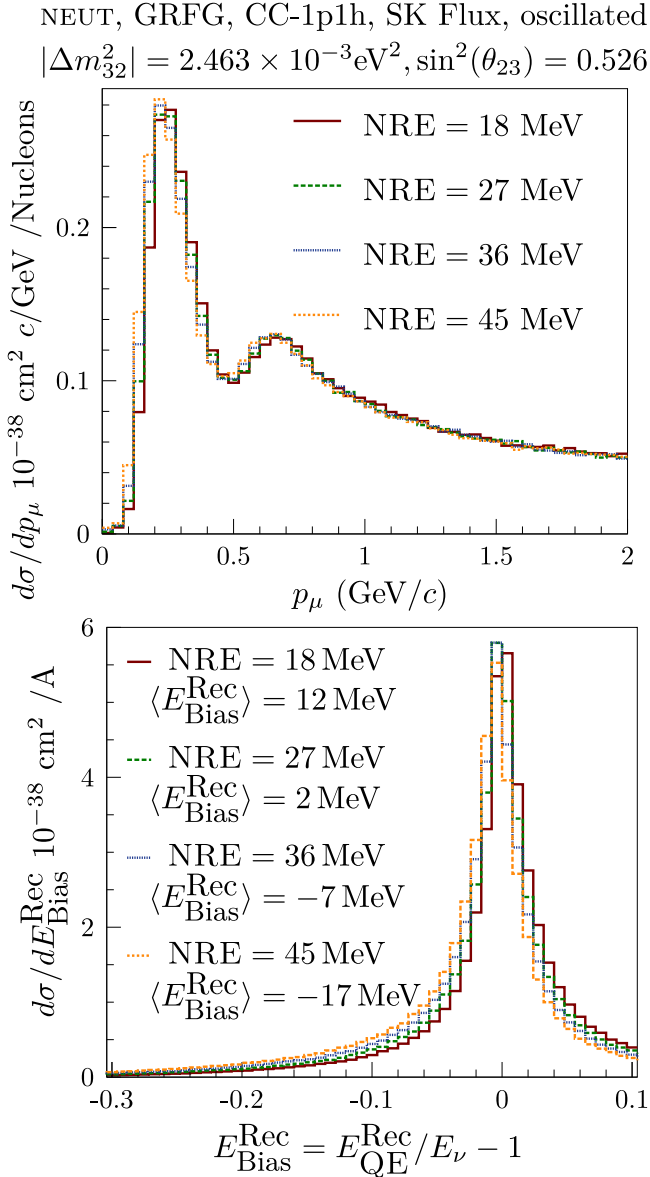

FIG. 8. Top: the predicted muon momentum spectrum for a number of different values of NRE in an oscillated SK flux. The shift toward lower momentum may be confused for a shift in $|\Delta m^2_{32}|$. Bottom: the effect of varying NRE on the reconstructed energy bias; higher values result in more energy "feed down."

section than the Nieves *et al.* calculation—importantly increasing the neutrino 2p2h strength relative to the antineutrino 2p2h strength—and is thus an instructive alternate model.

4. Kabirnezhad single-pion production

The Rein-Sehgal model accounts for interference between pion production channels that include a baryon resonance; however, interference between resonance and nonresonance channels is neglected. A new model, developed by Kabirnezhad [80], overcomes this limitation and is used to build a simulated dataset.

5. Data-driven CC $0\pi E_\nu - Q^2$ dependence

The untuned ND280 CC 0π sample prediction underestimates the data by approximately 5%. This sample is largely composed of 1p1h and 2p2h interactions but with a significant contribution from interactions that produce a pion which is then absorbed before leaving the nucleus. The 2p2h interaction can be further classified into events with and without a virtual $\Delta(1232)$ particle. Simulated datasets are created by assigning the observed CC 0π data-simulation discrepancy to either the 1p1h or 2p2h event categories. At the near detector the event category is weighted in bins of lepton momentum and angle so that the simulation matches the data. This weighting is then projected as a function of neutrino energy and Q^2 and applied to the far detector simulation to create the simulated far detector data.

6. Coulomb correction

As the final-state charged lepton leaves the nuclear potential, it undergoes a small momentum shift because of interaction with the Coulomb field of the nucleus. In addition, the Coulomb potential results in a small variation in the relative neutrino or antineutrino cross section. The effect of the Coulomb potential was not included in the base model, and thus a simulated dataset was included in which a momentum shift was applied to final-state (anti)muons and (anti)electrons, following Ref. [81], and the relative charged-current cross section for neutrinos and antineutrinos was varied by 3%.

To understand the effect of these model differences on the measured oscillation parameters the simulated datasets described here were fit assuming they were the real data collected at both the near and far detectors. The oscillation contours this produced were then compared to the expected result from the nominal T2K Monte Carlo. A change in the contours would show that the nominal interaction model and its uncertainties were unable to replicate the changes seen in the simulated data, and so additional uncertainty would be added to the analysis to take this into account. This procedure is described in detail in Sec. XIII.

VIII. NEAR DETECTOR DATA

The history of protons on target delivered to the T2K beam line until the end of May 2018 is shown in Fig. 9. Data for runs 1–9 in the muon monitors and the on-axis INGRID detector are shown in Fig. 10. The rate is stable throughout the run periods, and the horizontal and vertical beam positions are stable to less than 1 mrad throughout all of the run periods.

The off-axis near detector, ND280, is located 280 m upstream of the beam source. It consists of several subdetectors inside a 0.2 T magnet. Charged-current (CC) ν_μ and $\bar{\nu}_\mu$ neutrino interactions are selected in the tracker region, which is composed of two fine-grained detectors
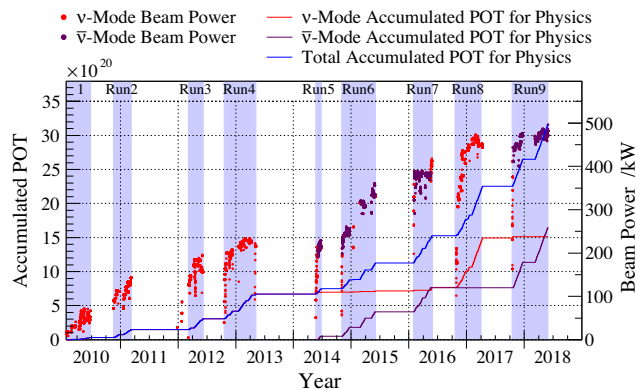
FIG. 9. The T2K data-taking periods, showing the accumulated beam protons on target delivered as a function of time.
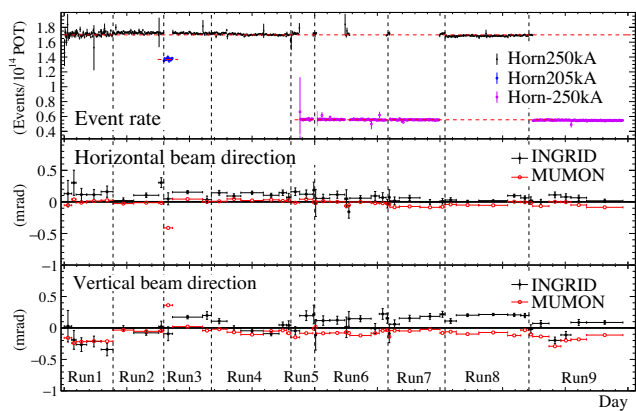

FIG. 10. Data in the muon monitors (MUMON) and INGRID on-axis detector for runs 1–9. The top row shows the event rate in both detectors, which is reduced as expected for RHC mode. The middle and bottom rows show the horizontal and vertical beam direction in both detectors.

(FGD1 and FGD2) [82] interleaving three time projection chambers (TPCs) [83]. The FGDs provide the target mass for neutrino interactions. The first FGD consists of 30 layers each composed of 192 plastic scintillator bars. The bars in each layer are oriented perpendicularly to the neutrino beam direction and to the bars in the preceding layer. Each pair of layers forms a single module providing a three-dimensional position for charged particles passing through it. The second FGD consists of alternating plastic scintillator modules and water panels. There are seven scintillator modules interspersed with six water panels, providing a water target for neutrinos to interact within. This allows effects relating to neutrino interactions on water to be isolated from those on carbon, reducing the uncertainty in extrapolating the event rate measurement from ND280 to SK. The TPCs measure both the curvature of charged particles in the magnetic field of ND280 and the energy lost by the particles as they travel through the TPC gas. The curvature of the particles provides a precise

measurement of their momentum and charge, while the energy loss allows the particle species to be identified.

Only ND280 data from runs 2–6 are used in this analysis, a smaller sample than for SK. Data quality is assessed weekly, and the total ND280 data-taking efficiency across runs 2–6 was ∼93%. The near detector analysis uses a total exposure of 5.80×10^{20} POT in FHC and 3.86×10^{20} POT in RHC, as shown in Table I.

The event selection for FHC is unchanged since the analysis described in Ref. [26]. The highest momentum, negatively charged track in each event is selected as the lepton candidate. The candidate track must start within the fiducial volume of FGD1 or FGD2 and be identified as muonlike by the TPC. This produces a selection of CC ν_μ interactions. The selected events are divided into three samples for each FGD, based on the reconstructed pion multiplicity. Positive pions are identified in three ways: a positive charged FGD-TPC track with energy loss consistent with a pion; a positively charged FGD-contained track with charge deposition consistent with a pion; or a delayed energy deposit in the FGD due to stopped $\pi^+ \to \mu^+ \to e^+$ decays. Negatively charged, minimally ionizing TPC tracks are identified as negatively charged pions. Neutral pions decay instantaneously to pairs of photons, which can then convert to electron-positron pairs. TPC tracks with charge depositions consistent with an electron are used to identify these decays.

The three FHC CC subsamples are CC 0π, which is dominated by CCQE interactions, CC $1\pi^+$, which is dominated by CC resonant single-pion production, and CC other, which is dominated by interactions producing multiple pions. The reconstructed muon momentum and angle of the selected data and simulation events in the FHC CC 0π and CC $1\pi^+$ samples are shown in Figs. 11 and 12, for both FGD1 and FGD2. The numbers of events recorded in each sample and the expectation prior to the ND280 fit are shown in Table II.

The event selection for ν_μ and $\bar{\nu}_\mu$ interactions in the RHC beam mode is unchanged since the previous analysis [26]. These selections differ from the FHC selections described above in two important ways. As a larger number of interactions are produced by wrong-sign neutrinos, selections of both ν_μ and $\bar{\nu}_\mu$ interactions are used in the RHC beam. Taking into account differences in the flux and cross section, the wrong-sign contamination is approximately 30% in the selected RHC samples compared to 4% in the FHC samples.

The selected $\bar{\nu}_\mu$ (ν_μ) CC candidate events are divided into two samples for each FGD, based on the number of reconstructed tracks crossing a TPC. These are CC 1-track, which is dominated by CCQE-like interactions, and CC N-track, which is dominated by interactions producing pions. The events are not divided according to the number of observed pions, unlike the FHC selections, due to the lower interaction rate for antineutrinos. The reconstructed
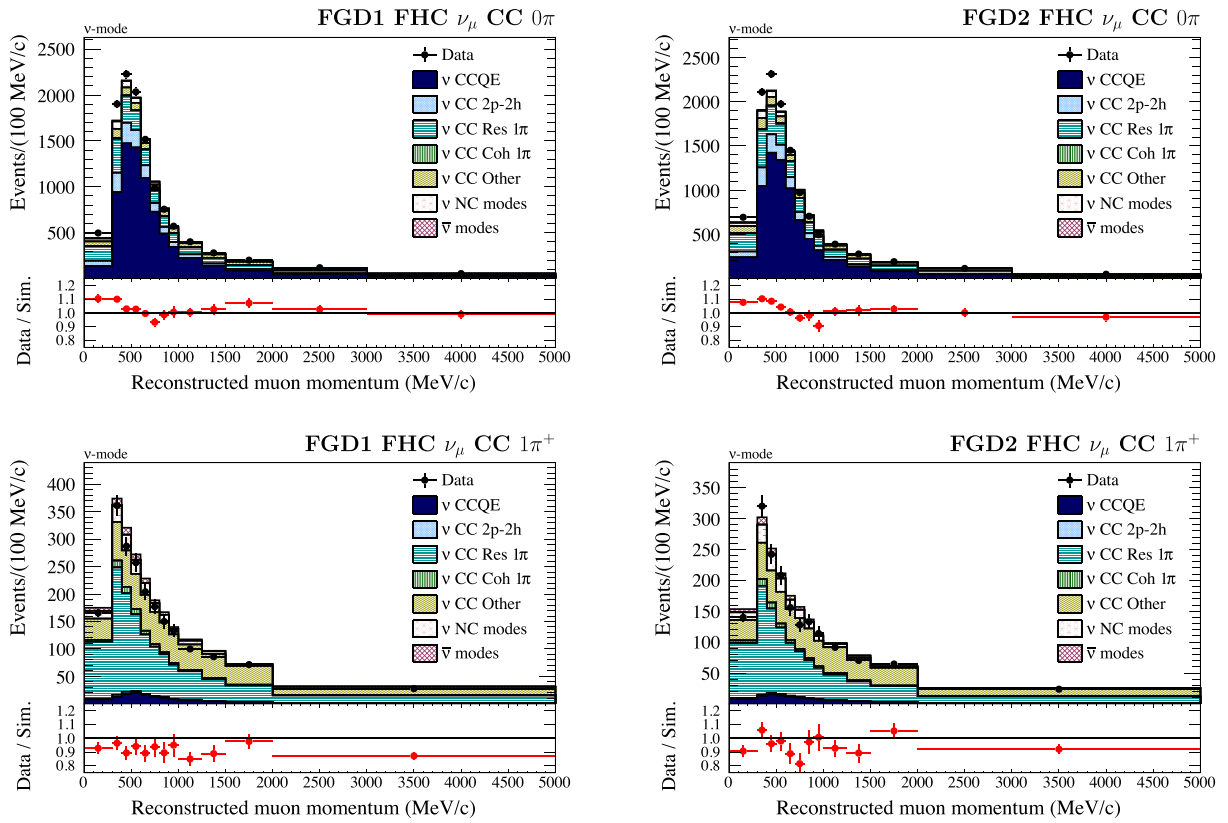
FIG. 11.   Final-state muon momentum distributions of the FHC ν_μ CC 0π (top) and ν_μ CC $1\pi^+$ (bottom) data and simulation samples in FGD1 (left) and FGD2 (right).
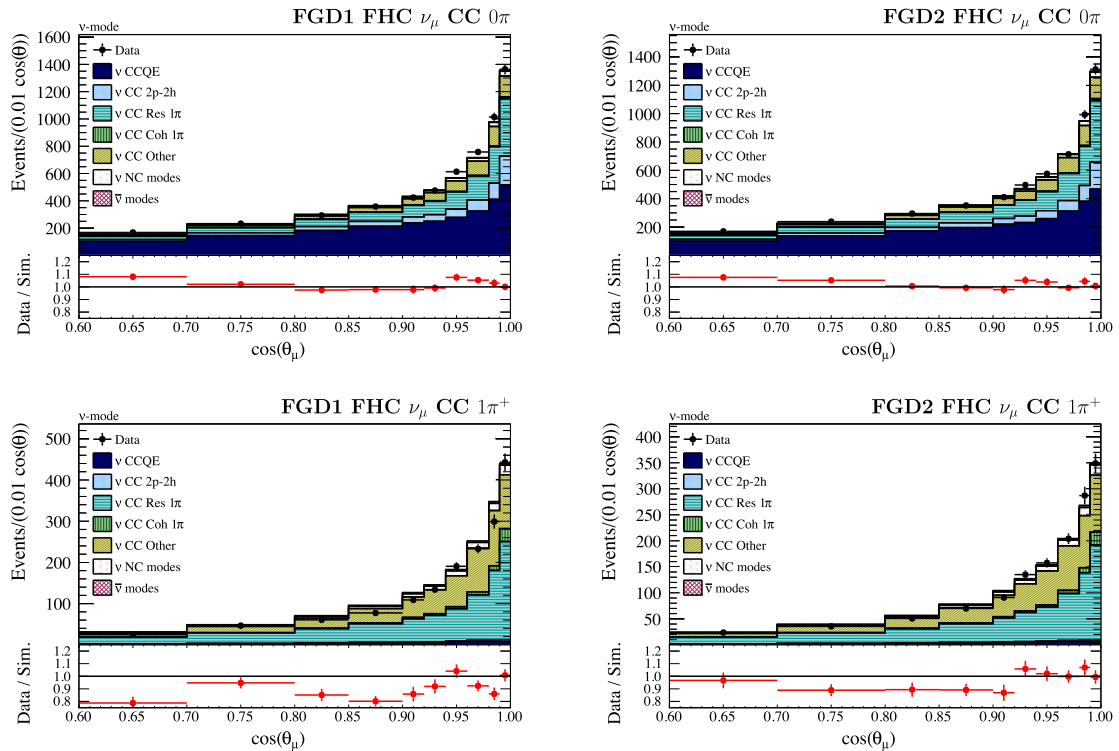

FIG. 12.   Distributions of the final-state muon angle of the FHC ν_μ CC 0π (top) and ν_μ CC $1\pi^+$ (bottom) data and simulation samples in FGD1 (left) and FGD2 (right).

TABLE II. ND280 samples, with the observed and expected numbers of events (before and after fitting at ND280).

Beam	Topology	Target	Data	Prediction	Postfit
FHC	ν_μ CC 0π	FGD1	17 136	16 724	17 122
		FGD2	17 443	16 959	17 495
	ν_μ CC $1\pi^+$	FGD1	3954	4381	4062
		FGD2	3366	3564	3416
	ν_μ CC other	FGD1	4149	3944	4096
		FGD2	4075	3571	3915
RHC	$\bar{\nu}_\mu$ CC 1-track	FGD1	3527	3588	3504
		FGD2	3732	3618	3685
	$\bar{\nu}_\mu$ CC N-track	FGD1	1054	1067	1053
		FGD2	1026	1077	1097
	ν_μ CC 1-track	FGD1	1363	1272	1353
		FGD2	1320	1263	1330
	ν_μ CC N-track	FGD1	1370	1357	1354
		FGD2	1253	1247	1263

muon momentum and angle of the selected events in these samples for FGD1 are shown in Figs. 13 and 14, respectively.

In total there are 14 ND280 event samples: six for FHC (CC 0π, $1\pi^+$ and other, for FGD1 and FGD2), four for right-sign RHC (CC 1-track and CC N-track, for FGD1 and FGD2) and four for wrong sign RHC (CC 1-track and CC N-track, for FGD1 and FGD2). The number of observed and predicted events for each sample are shown in Table II.

IX. SUPER-KAMIOKANDE DATA AND SIMULATION

The Super-Kamiokande detector [84] consists of a cylindrical tank filled with 50 kt of pure water, located in the Mozumi mine in Hida, Gifu. An overburden of 2700 meter-water-equivalent provided by Mount Ikeno suppresses the cosmic ray muon flux by 5 orders of magnitude. Photomultiplier tubes (PMTs) are supported by a 55-cm-wide steel structure, placed 2 m away from the tank walls, which divides the detector into two optically separated regions. The outer detector (OD) region, used to identify events with entering particles, is lined with reflective material and viewed by 1885 8″ PMTs. The inner detector (ID) region contains 32 kt of water and is instrumented with 11146 20″ PMTs which make up 40% of the detector's inner surface. The high density of PMTs in the ID allows for the imaging of the ringlike light patterns projected on the detector walls by particles traveling above the Cherenkov threshold in the water.
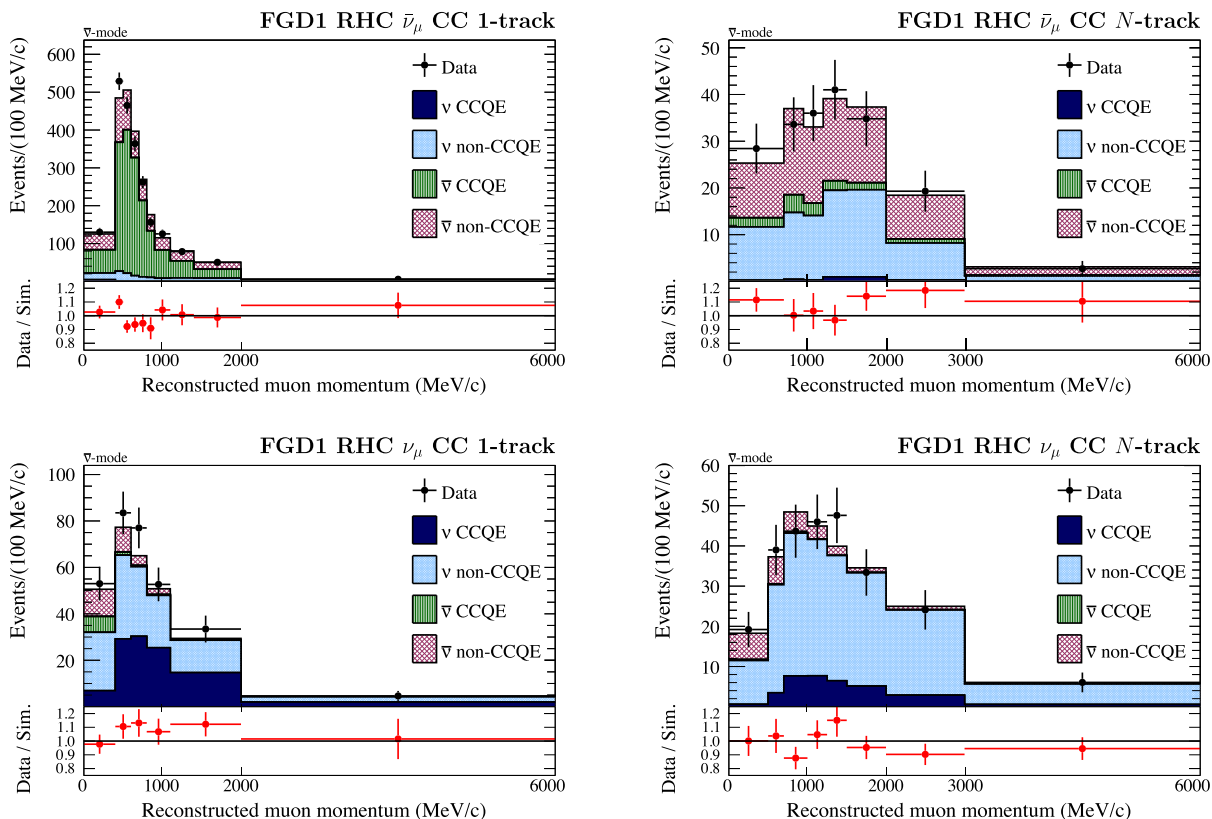

FIG. 13. Final-state muon momentum distributions for the RHC $\bar{\nu}_\mu$ (top) and ν_μ (bottom) CC 1-track (left) and CC N-track (right) FGD1 simulation samples. These distributions are before the ND280 fit.
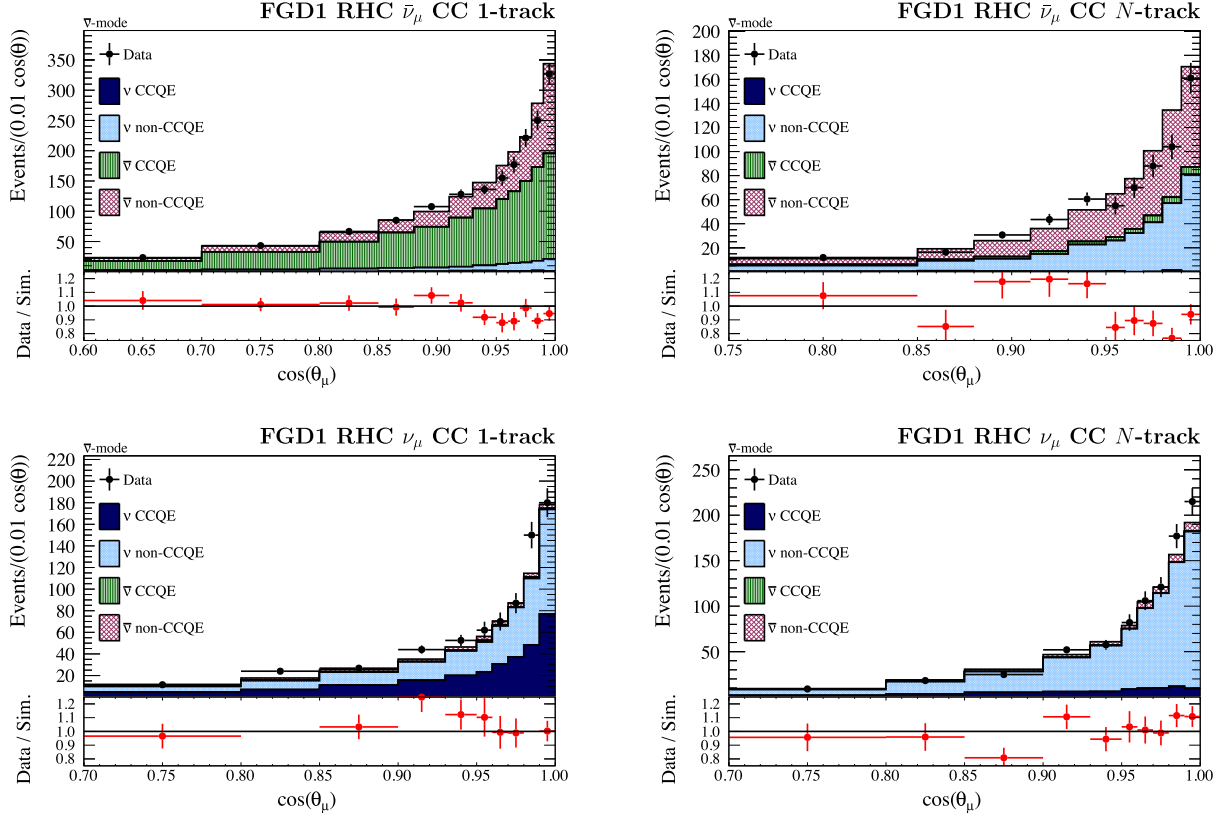
FIG. 14.   Distributions of the final-state muon angle for the RHC $\bar{\nu}_\mu$ (top) and ν_μ (bottom) CC 1-track (left) and CC *N*-track (right) FGD1 simulation samples. These distributions are before the ND280 fit.

A. Super-Kamiokande data

Pulses on PMTs exceeding a charge threshold corresponding to roughly 0.1 photoelectrons are registered as hits, all of which are processed by a software trigger system [85]. For T2K analyses, all hits occurring in the 1 ms windows centered on each beam spill arrival are written to disk. Beam spills are excluded from the analysis if they coincide with problems in the data acquisition system or the GPS system used to synchronize SK with the accelerator at J-PARC. Additionally, spills that occur within 100 μs of a beam-unrelated event are rejected to reduce the contamination of T2K data with cosmic ray muon decay electrons. The beam spill selection introduces an inefficiency of 1%, with roughly half of this being due to the preceding detector activity criterion.

For the analysis presented here, events associated with accepted spills are further required to have a reconstructed energy corresponding to an electron of at least 30 MeV and no more than 15 hits in the largest OD hit cluster. Additional criteria are used to reject spurious events that originate from spontaneous corona discharges in PMTs. Only events reconstructed in the $[-2, 10]$ μs window around the leading edge of the beam spill are used in the analysis.

Distributions of the reconstructed times for events in both 1 ms and $[-1.2, 5.6]$ μs windows around the beam

spill arrival are shown in Fig. 15. In the 1 ms window, a peak of events coincident with the beam arrival is clearly seen; after applying the OD and minimum energy criteria very few events remain outside this peak. The eight-bunch structure of the J-PARC beam is clearly seen in the narrow window event time distribution.

B. Super-Kamiokande event simulation

Events at SK are simulated using J-PARC (anti)neutrino flux predictions and the neutrino interaction generator NEUT, which implements the neutrino interaction model described in detail in Sec. VII. Particles resulting from the neutrino interactions are propagated through an SK detector model using the same GEANT3-based [36] SKDETSIM 13.90 package as in [26]. The detector model, including the optical properties of the ultrapure water and detector materials, is tuned to calibration data [86].

X. SUPER-KAMIOKANDE EVENT RECONSTRUCTION AND SELECTION

Events at SK are reconstructed with the FiTQun maximum likelihood estimation algorithm [27]. While this algorithm was initially used exclusively for NCπ^0 background suppression in the ν_e appearance channel [26,28], in recent T2K publications [16,25] FiTQun was used for all
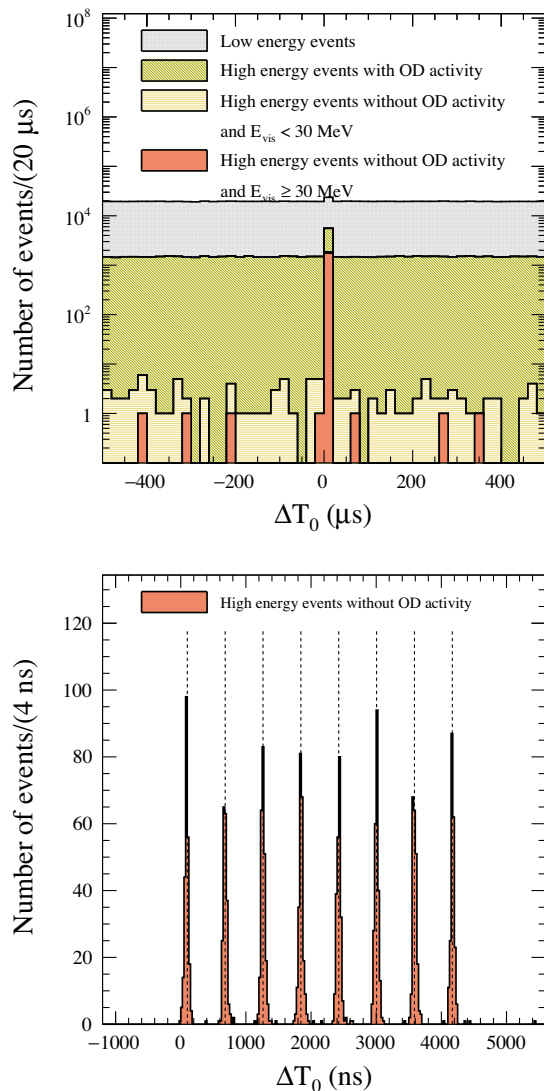
FIG. 15. Reconstructed event time distributions in the 1 ms (top) and $[-1.2, 5.6]$ μs (bottom) windows around the beam spill leading edge.

aspects of event reconstruction. Updating the reconstruction tools prompted a reoptimization of the event selection criteria, including an expansion of the fiducial volume (FV).

In this section, the reconstruction algorithm is briefly described, as well as the updated event selection criteria and the procedure for their optimization. A discussion of the systematic uncertainties related to the SK detector concludes the section.

A. Event reconstruction algorithm

The FiTQun likelihood function consists of the probability of each PMT registering a hit in a given event, and for hit PMTs, the probability density functions for the charge and time of the hit. Particles in an event are described by tracks (or track segments) parameterized by particle type,

momentum, direction and initial position. The FiTQun likelihood is a function of these track parameters and multiple tracks can be combined to form complex event hypotheses.

In an initial prefitting stage, the approximate location of the neutrino interaction is found with a simplified likelihood using only the time of the PMT hits. A residual time is calculated for each PMT hit by subtracting the Cherenkov photon time of flight, calculated using the straight-line distance from the vertex position to the PMT, from the hit time. Hits are associated to one or more clusters in residual time, with the initial cluster containing hits due to particles produced in the neutrino interaction and subsequent clusters containing hits due to products of weakly decaying prompt particles. Each hit cluster is then reconstructed separately by maximizing the likelihood function for the e, μ, π^+ and p single-track hypotheses.

For the earliest hit cluster only, multiple-track event hypotheses are also reconstructed using the results of the single-track fits as the starting point. A multitrack search algorithm is used to determine the number of tracks observed in the event. This algorithm proceeds by iteratively adding a new electronlike or π^+-like track to the event until the best-fit likelihood after adding the new track fails to improve beyond a set threshold. In the analysis described here, additional event hypotheses targeting neutral-current backgrounds are used: a π^0 hypothesis consisting of two electronlike tracks consistent with a $\pi^0 \to \gamma\gamma$ decay and a π^+ hypothesis with a single track compatible with a π^+ undergoing a hard scatter.

B. Event selection

Events are selected into samples using cuts on best-fit likelihood ratios between signal-like and backgroundlike hypotheses: $\Lambda_\beta^\alpha \overset{\text{def}}{=} \log \frac{L_\alpha}{L_\beta}$, where α and β are competing hypotheses. The cut points are typically parameterized as a function of reconstructed kinematics, such as the best-fit electron momentum or the reconstructed invariant mass obtained from the π^0 hypothesis best-fit kinematics.

Five signal-enriched SK samples are used in the analysis, each of which has a single prompt track reconstructed from a Cherenkov ring. Samples of events containing a single reconstructed μ-like ring (1R$_\mu$) and a single reconstructed e-like ring (1R$_e$) target ν_μ and ν_e CCQE interactions in both FHC and RHC beam modes. An additional e-like sample, used in FHC data only, targets CC $1\pi^+$ interactions where the π^+ is below Cherenkov threshold. The π^+ is identified by the detection of a delayed μ-decay electron following the single prompt electron which results from the CC interaction (1R$_e$ + 1 d.e.). The CCQE-like selection criteria are the same for FHC and RHC samples.

Events in all samples are required to be fully contained (FC) in the ID using the cut on OD activity described in Sec. IX above and to have only one prompt reconstructed particle identified by the multitrack iterative search
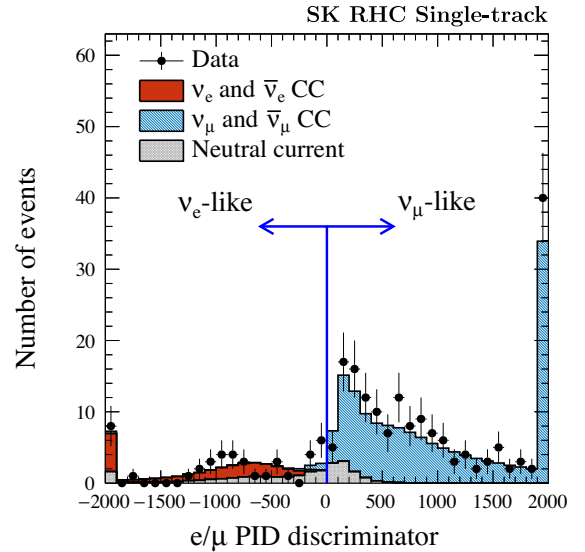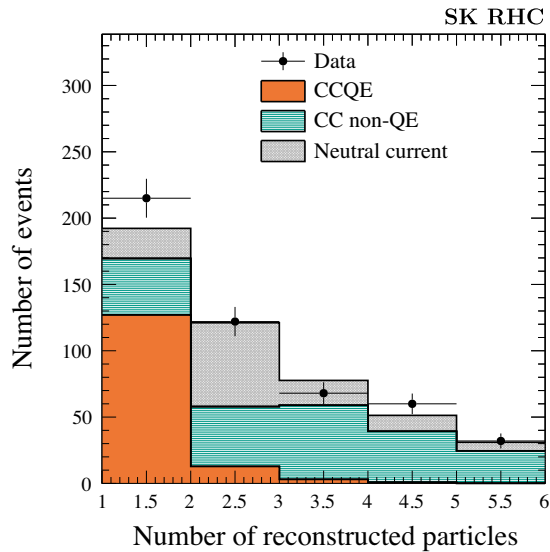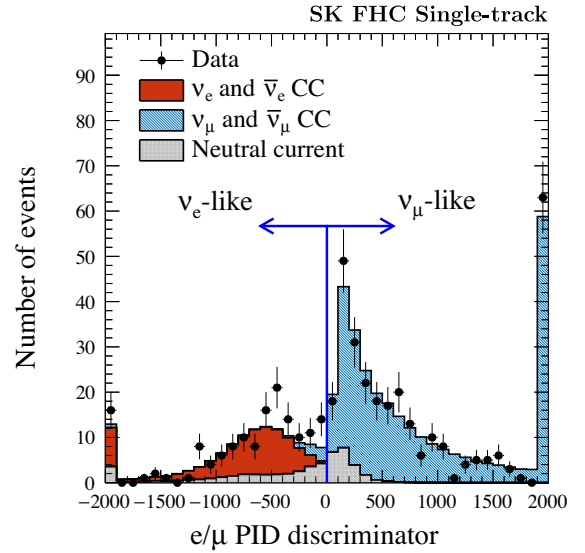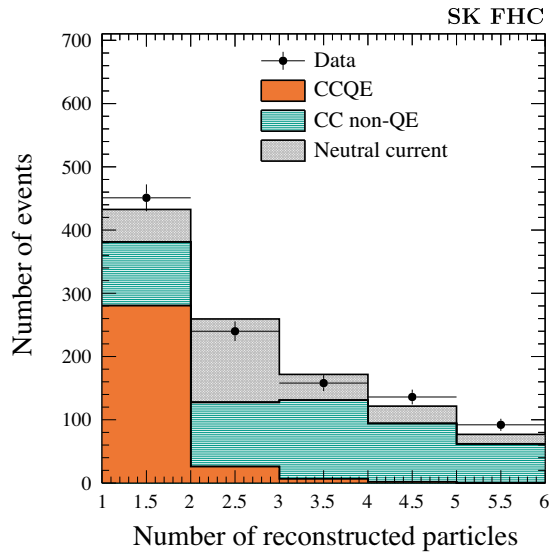
FIG. 16. Distribution of number of reconstructed particle tracks for events passing the *wall* > 80 cm and *towall* > 170 cm FV criteria in FHC (top) and RHC (bottom) data.

FIG. 17. Distribution of the e/μ PID discriminator for single-track events passing the *wall* > 80 cm and *towall* > 170 cm FV criteria in FHC (top) and RHC (bottom) data.

algorithm. Events are separated into *e*-like and *μ*-like with a criterion based on the likelihood ratio of the best-fit *e*-like to *μ*-like hypothesis (Λ_μ^e) and the reconstructed momentum for the *e* hypothesis (p_e).

The FV criteria are defined in terms of the distance from the event vertex to its closest point on the detector walls (*wall*) and the distance from the event vertex to the detector wall along the track direction (*towall*). This parameterization of the FV allows for a larger volume of the detector to be used by reducing the *wall* threshold compared to previous T2K neutrino oscillation analyses, while ensuring that Cherenkov rings projected on the detector walls illuminate a large number of PMTs with the *towall* criterion, introduced for the first time in the analysis

described here. The *wall* and *towall* criteria are chosen separately for each sample to maximize the sensitivity to θ_{23} and δ_{CP}, as described in Sec. X D. For the *μ*-like samples, a minimum *wall* of 50 cm is required, with a minimum *towall* of 250 cm. The requirements for the *e*-like samples with no decay-*e* are *wall* > 80 cm and *towall* > 170 cm, while for the sample with one decay-*e* *wall* > 50 cm and *towall* > 270 cm are required.

For both FHC and RHC FC events, the distributions of the number of reconstructed particle tracks are shown in Fig. 16. For events with a single reconstructed track, the distributions of the *e/μ* discriminator and number of identified *μ*-decay electrons are shown in Figs. 17 and 18, respectively. In these
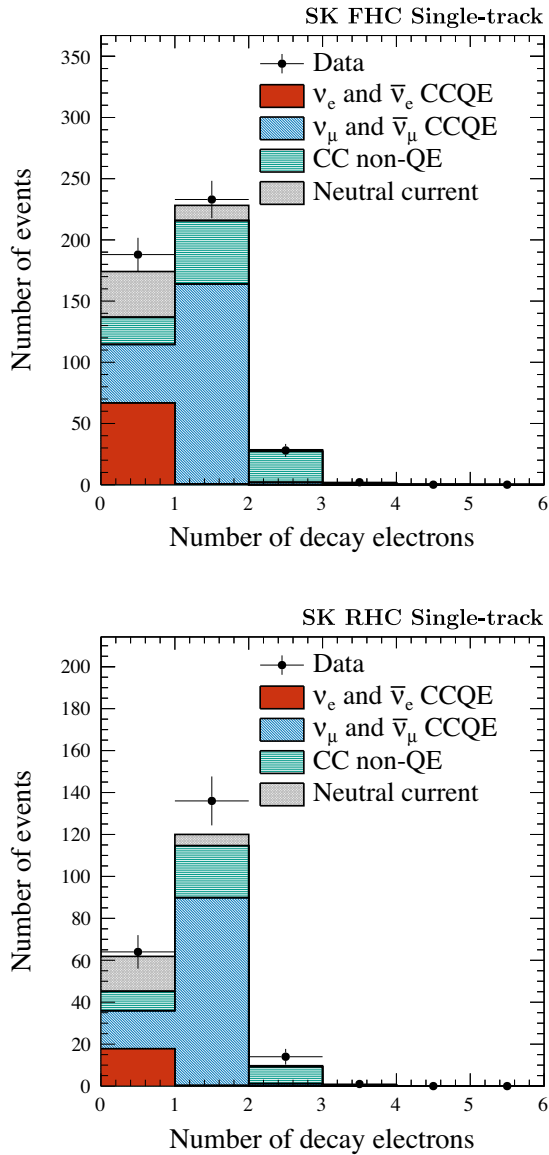
FIG. 18. Distribution of the number of identified decay electrons for single-track events passing the *wall* > 80 cm and *towall* > 170 cm FV criteria in FHC (top) and RHC (bottom) data.

figures, and throughout this section, the MC predictions are produced with the neutrino mixing parameters given in Table III, and the flux and cross-section parameters set to the best-fit value resulting from the near detector analysis described in Sec. XII.

The reconstructed momentum is required to be larger than 100 MeV/c for the e-like samples to reduce contamination from below-threshold-μ decays and larger than 200 MeV/c for the μ-like samples.

Events in the μ-like samples can have up to one reconstructed decay-e, while the e-like samples are required to have zero and one decay-e for the samples targeting CCQE and CC1π^+ interactions, respectively.

TABLE III. Reference set of values of the oscillation parameters used to evaluate the expected sensitivity, number of events and effect of systematic uncertainties. The values of $\sin^2\theta_{12}$ and Δm^2_{21} are taken from [30] and the value of $\sin^2\theta_{13}$ from [31], and all the other values are set to the most probable value found in a previous T2K measurements [29].

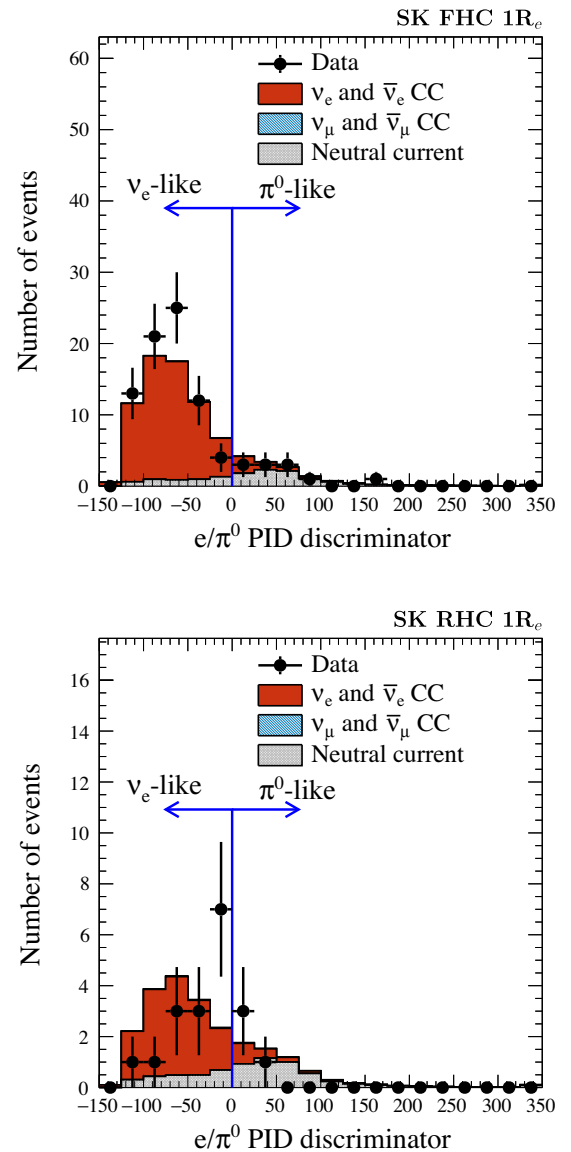| Parameter	Value
$\sin^2\theta_{23}$	0.528
$\sin^2\theta_{13}$	0.0212
$\sin^2\theta_{12}$	0.304
δ_{CP}	−1.601
Δm^2_{32} (NO)	2.509×10^{-3} eV$^2/c^4$
Δm^2_{13} (IO)	
Δm^2_{21}	7.53×10^{-5} eV$^2/c^4$
Mass ordering	Normal

FIG. 19. Distribution of the e/π^0 PID discriminator for 1R$_e$ events in FHC (top) and RHC (bottom) data.
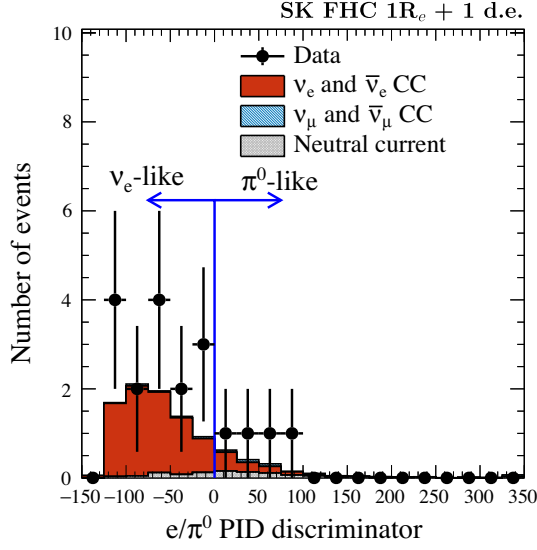
FIG. 20. Distribution of the e/π^0 PID discriminator for $1R_e +$ 1 d.e. in FHC data.

Neutral-current π production events are a background in both e-like and μ-like samples. In the former, electromagnetic showers resulting from $\pi^0 \to \gamma\gamma$ decays can mimic an electronlike event; while in the latter, charged pion detector signatures are only significantly different from those of muons through their hadronic interactions, which are not always present. Additional criteria are used to remove these backgrounds in all analysis samples.

For the μ-like samples a cut is applied on the likelihood ratio of the best-fit single-π^+ to the single-μ hypotheses ($\Lambda_\mu^{\pi^+}$) as a function of reconstructed μ momentum (p_μ). This selection criterion is only available in the FiTQun reconstruction algorithm and is deployed for the first time in the current analysis.

The NCπ^0 rejection criterion for the e-like samples, based on the likelihood ratio of the best-fit π^0 hypothesis to the e-like hypothesis ($\Lambda_e^{\pi^0}$) and the reconstructed π^0 mass ($m_{\gamma\gamma}$), is unchanged from previous analyses, where it was the only use of FiTQun reconstruction. Distributions of the neutral-current π^0 rejection discriminator for the $1R_e$ and $1R_e + 1$ d.e. samples are shown in Figs. 19 and 20, respectively. For $1R_\mu$ events, the distribution of the neutral-current π^+ discriminator is shown in Fig. 21.

Finally, e-like samples are required to have a reconstructed neutrino energy (E_{rec}) lower than 1250 MeV, as events beyond this energy are insensitive to neutrino oscillations and are more susceptible to systematic uncertainties associated with the neutral-current rejection cut.

The selection criteria described above are summarized in Table IV and the breakdown of data and MC events passing each cut is given in Tables V–VII, for the samples targeting ν_e CCQE, ν_e CC $1\pi^+$ and ν_μ CCQE interactions, respectively.
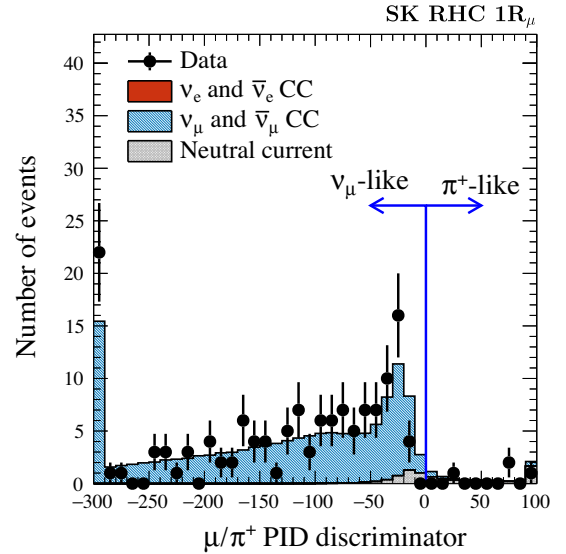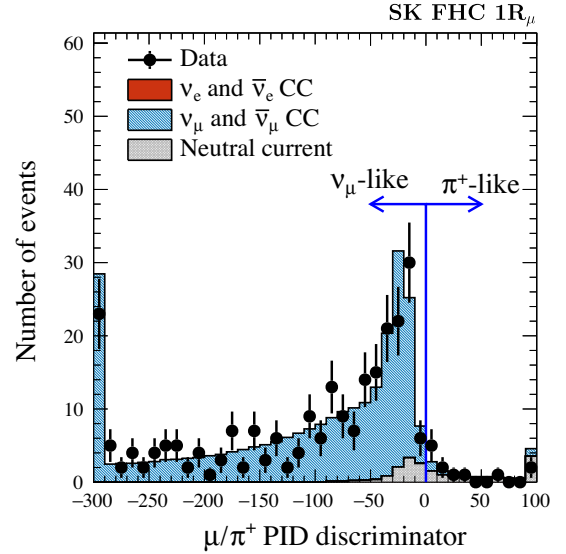

FIG. 21. Distribution of the μ/π^+ PID discriminator for $1R_\mu$ events in FHC (top) and RHC (bottom) data.

Distributions of reconstructed vertices of $1R_e$ events in FHC and RHC data are shown in Figs. 22 and 23, respectively, and for $1R_e + 1$ d.e. events in Fig. 24. Reconstructed neutrino energy distributions are shown in Figs. 25 and 26 for $1R_e$ and $1R_e + 1$ d.e. events, respectively, and in Fig. 27 for $1R_\mu$ events.

C. Optimization of selection criteria

The likelihood ratio of best-fit e and μ hypotheses gives very good separation between these classes of events, with the separation improving at higher momentum. The cut line chosen to select e-like and μ-like events accounts for the momentum dependence of the likelihood ratio and achieves misidentification rates smaller than 1% for true CCQE events across the T2K energy range.

TABLE IV. SK selection criteria for the five fully contained, single prompt track, analysis samples. The $1R_\mu$ and $1R_e$ selections are identical in both FHC and RHC beam modes, but the $1R_e + 1$ d.e. selection is only applied in FHC beam mode.

	$1R_\mu$	$1R_e$	$1R_e + 1$ d.e. (FHC mode only)
wall	>50 cm	>80 cm	>50 cm
towall	>250 cm	>170 cm	>270 cm
e/μ identification	$\Lambda_\mu^e < \frac{p_e}{5\,\text{MeV}/c}$	$\Lambda_\mu^e > \frac{p_e}{5\,\text{MeV}/c}$	$\Lambda_\mu^e > \frac{p_e}{5\,\text{MeV}/c}$
e/μ momentum	>200 MeV/c	>100 MeV/c	>100 MeV/c
Number of decay-e	≤ 1	0	1
NCπ rejection	$\Lambda_\mu^{\pi^+} < \frac{6p_\mu}{40\,\text{MeV}/c}$	$\Lambda_e^{\pi^0} < 175 - \frac{35 m_{\gamma\gamma}}{40\,\text{MeV}/c^2}$	$\Lambda_e^{\pi^0} < 175 - \frac{35 m_{\gamma\gamma}}{40\,\text{MeV}/c^2}$
$E_{\text{QE}}^{\text{Rec}}$	$\cdots$	<1250 MeV	<1250 MeV

TABLE V. Expected number of $1R_e$ signal and background events passing each selection stage, compared to the data.

	$\nu_\mu + \bar{\nu}_\mu$	$\nu_e + \bar{\nu}_e$	$\nu + \bar{\nu}$	$\nu_\mu \to \nu_e$	$\bar{\nu}_\mu \to \bar{\nu}_e$		
FHC	CC	CC	NC	CC	CC	MC total	Data
FC and FV	692.28	43.10	241.98	87.18	0.80	1065.34	1077
Single track	307.47	22.18	44.31	73.09	0.61	447.65	451
Electronlike	8.72	22.16	26.38	72.99	0.61	130.85	151
$p_e > 100$ MeV/c	3.22	22.00	18.45	71.56	0.61	115.84	131
No decay-e	0.88	18.73	15.57	64.60	0.59	100.37	108
$E_{\text{rec}} < 1250$ MeV	0.56	9.89	11.64	62.40	0.43	84.91	86
Not π^0	0.27	8.79	4.21	58.53	0.38	72.17	75
RHC							
FC and FV	311.20	21.48	122.86	5.81	10.31	471.67	497
Single track	144.49	10.88	22.61	4.13	8.81	190.93	215
Electronlike	2.81	10.88	13.83	4.13	8.80	40.44	42
$p_e > 100$ MeV/c	1.41	10.83	9.92	4.06	8.75	34.97	32
No decay-e	0.41	9.48	8.60	3.47	8.58	30.53	28
$E_{\text{rec}} < 1250$ MeV	0.28	4.27	6.77	2.91	8.13	22.37	19
Not π^0	0.13	3.70	2.40	2.65	7.37	16.26	15

Pion production in neutral-current events forms one of the main backgrounds to both μ-like and e-like selections. Furthermore, since the cross sections for these processes are not known precisely, these contributions carry significant systematic uncertainties into the signal samples. A simplified neutrino oscillation analysis framework is used to optimize the neutral-current rejection criteria taking into account systematic uncertainties and with statistics corresponding to an exposure of 7.8×10^{21} POT, evenly split between neutrino and antineutrino modes. In this simplified analysis framework, the systematic uncertainties (taking into account the near detector constraints) are propagated to

TABLE VI. Expected number of $1R_e + 1$ d.e. signal and background events passing each selection stage, compared to the data.

	$\nu_\mu + \bar{\nu}_\mu$	$\nu_e + \bar{\nu}_e$	$\nu + \bar{\nu}$	$\nu_\mu \to \nu_e$	$\bar{\nu}_\mu \to \bar{\nu}_e$		
FHC	CC	CC	NC	CC	CC	MC total	Data
FC and FV	697.81	43.87	247.50	87.30	0.81	1077.27	1085
Single track	303.60	22.24	44.39	72.83	0.61	443.67	443
Electronlike	8.46	22.22	26.98	72.74	0.61	131.01	148
$p_e > 100$ MeV/c	2.81	22.05	18.73	71.24	0.61	115.44	129
One decay-e	1.35	3.04	2.18	6.81	0.02	13.40	23
$E_{\text{rec}} < 1250$ MeV	0.46	1.10	1.01	6.27	0.01	8.85	19
Not π^0	0.16	0.93	0.38	5.64	0.01	7.13	15

TABLE VII. Expected number of $1R_\mu$ signal and background events passing each selection stage, compared to the data.

	$\nu_e + \bar{\nu}_e$	$\nu + \bar{\nu}$	$\nu_\mu + \bar{\nu}_\mu$	ν_μ	$\bar{\nu}_\mu$		
FHC	CC	NC	CC non-QE	CCQE	CCQE	MC total	Data
FC and FV	125.04	234.01	373.95	251.21	14.20	998.41	1002
Single track	92.94	43.30	63.00	220.14	12.53	431.91	429
Muonlike	0.10	17.78	58.93	215.84	12.44	305.08	285
$p_\mu > 200$ MeV/c	0.10	17.66	58.89	215.63	12.44	304.71	284
0 or 1 decay-e	0.10	17.07	37.99	213.41	12.31	280.88	255
Not π^+	0.08	8.31	36.75	210.64	12.18	267.96	243
RHC							
FC and FV	34.68	118.94	169.35	44.95	72.95	440.87	454
Single track	22.74	22.12	34.43	36.55	67.44	183.29	197
Muonlike	0.02	8.72	33.08	36.19	66.45	144.45	159
$p_\mu > 200$ MeV/c	0.02	8.63	33.07	36.18	66.40	144.30	159
0 or 1 decay-e	0.02	8.37	25.22	35.76	65.71	135.07	144
Not π^+	0.02	3.89	24.71	35.36	65.00	128.97	140

the SK prediction as a covariance matrix in reconstructed neutrino energy. As a result, the SK samples do not constrain the nuisance parameters, and no correlations between these and the neutrino mixing parameters are taken into account. The four samples targeting e-like and μ-like CCQE events in both FHC and RHC neutrino beam mode are fit simultaneously to an Asimov dataset [87] to determine the experiment's sensitivity under different neutral-current rejection cut points.

The criterion to reject NCπ^+ events in the μ-like samples, a line cut on the $\Lambda_\mu^{\pi^+}$ vs p_μ plane, is chosen to minimize the width of the $\sin^2\theta_{23}$ 1σ confidence interval. This cut, which was not available with the reconstruction techniques used in previous T2K neutrino oscillation analyses, reduces the NC contribution to the μ-like samples by a factor of 2, while selecting CCQE events with 99% efficiency.

The NCπ^0 rejection line cut in the $\Lambda_e^{\pi^0}$ vs $m_{\gamma\gamma}$ plane, applied to the e-like samples, is optimized based on the significance to exclude $\delta_{CP} = 0$. As the optimal cut line is very close to the one used in previous T2K neutrino oscillation analyses, this criterion is not updated for the analysis described here. It should be noted that the relative impact of this cut on the selected sample is significantly smaller in the analysis described here, where it reduces the NC contribution in the e-like samples by a factor of 3, compared to previous analyses, where the reduction is of a factor of 6. This is due to the excellent performance of the multitrack search algorithm, which reduces the NC background in e-like samples by a factor of 5 by correctly identifying the two γ's from π^0 decays with higher efficiency than the previously used algorithms.

Optimization metrics for both neutral-current rejection criteria are shown as a function of the cut parameters in Figs. 28 and 29, along with the chosen cut points and distributions of the signal and background events in the cut

variable planes. While this optimization is performed assuming the neutrino mixing parameters preferred by previous T2K results as given in Table III, it should be noted that in both cases the optimization metrics show large regions around the optimal points where they are consistently good. Therefore, the sensitivity of this analysis does not depend strongly on the exact value of the cut points and the experimental sensitivity is not expected to depend strongly on the choice of neutrino mixing parameters used in the optimization procedure.

D. Fiducial volume expansion

In previous T2K neutrino oscillation analyses, events at SK were required to have $wall \geq 200$ cm in order to remove entering backgrounds and ensure the quality of reconstructed quantities. The new event selection presented here, benefiting from improvements in reconstruction, provides an opportunity to reoptimize the FV criterion for the T2K analysis samples, with the objective of increasing the event yield while maintaining a high purity of signal events in the selected samples and a low impact of detector systematic uncertainty.

A two-dimensional parameterization of the FV criteria is chosen to allow for balancing two classes of effects. On one hand, the reduction of entering backgrounds and mitigation of the impact of known shortcomings of the simulated detector geometry are achieved with a $wall$ threshold, as in previous analyses of T2K data. On the other hand, important aspects of event reconstruction, such as particle identification, improve with the number of PMTs illuminated by the Cherenkov ring patterns. As this number grows with the distance to the detector walls along the particle direction of travel, $towall$, a threshold on this distance is used to select events with a high reconstruction performance.
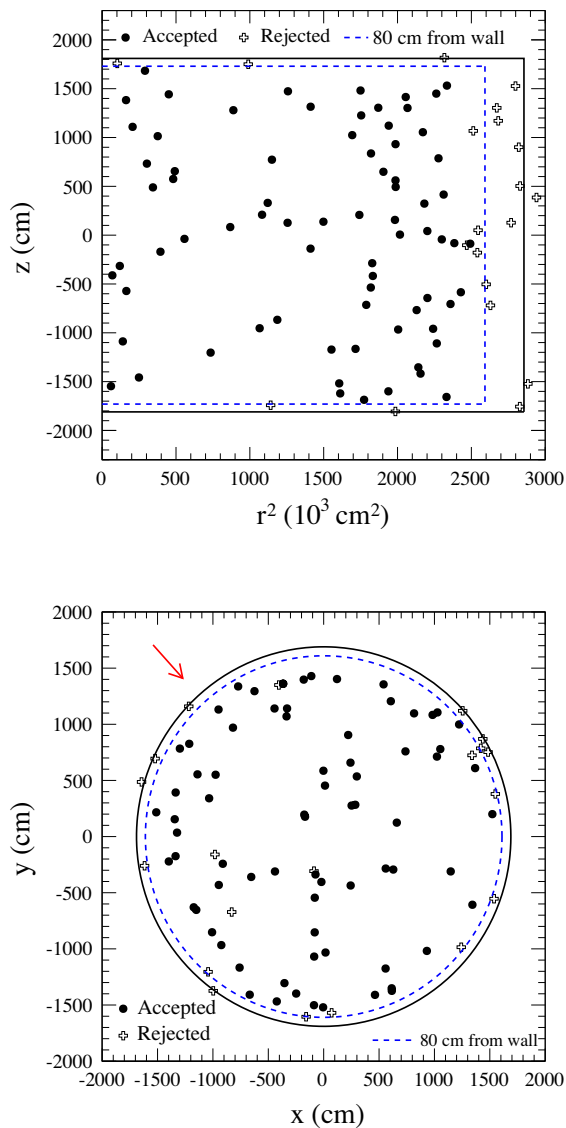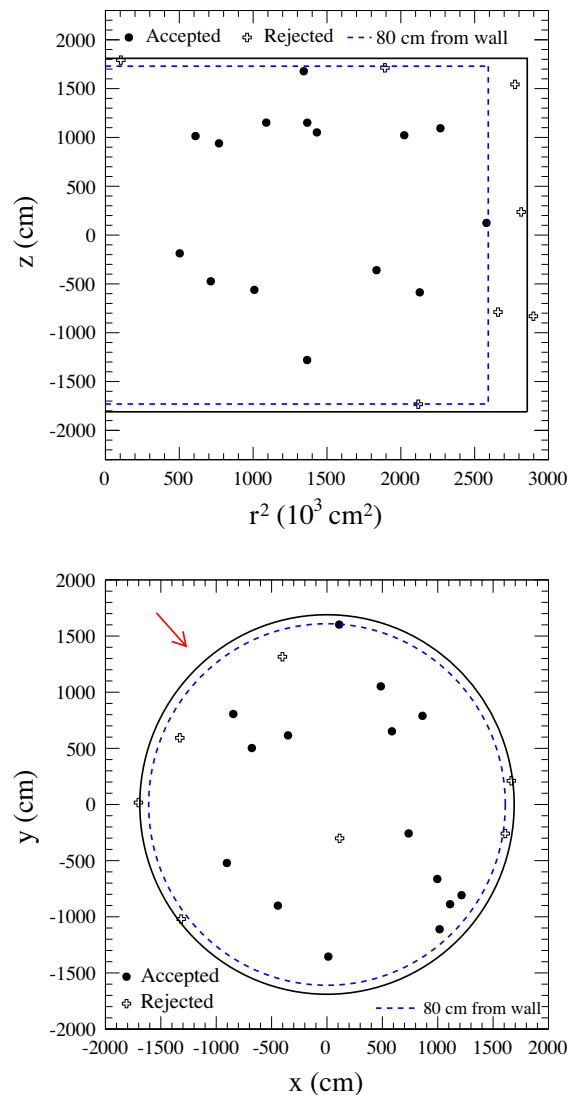
FIG. 22. Reconstructed vertices of selected FHC data 1R$_e$ events projected in z vs r^2 (top) and y vs x (bottom). The neutrino beam direction is shown as a red arrow and events which do not pass the fiducial volume criteria are shown as hollow crosses.

FIG. 23. Reconstructed vertices of selected RHC data 1R$_e$ events projected in z vs r^2 (top) and y vs x (bottom). The neutrino beam direction is shown as a red arrow and events which do not pass the fiducial volume criteria are shown as hollow crosses.

The FV criteria are optimized in a fit to SK atmospheric neutrino data, from which the systematic uncertainty associated to the particle counting and identification is also extracted.

The SK atmospheric neutrino FC data is divided into 18 samples consisting of combinations of six detector regions, defined with *wall* and *towall* as shown in Fig. 30, and three classes of events discriminated by the number (0, 1 or 2+) of detected Michel electrons. The six detector regions were chosen to isolate areas where the systematic uncertainty associated to the detector model is expected to differ, while maintaining an adequate level of statistics in the SK atmospheric neutrino data sample.

These samples are projected into three particle identification variables (Λ_μ^e, $\Lambda_e^{\pi^0}$ and $\Lambda_\mu^{\pi^+}$) and a continuous variable that discriminates single-track from multitrack events ($\Lambda_{1-\text{track}}^{n-\text{tracks}}$).

In each of the samples the MC is split into six true event topologies consisting of a single visible e, a single visible μ, a visible e with other visible particles, a visible μ with other visible particles, a single π^0, and finally events with a single visible p or π^+. For each topology, the particle identification and counting variables are linearly transformed with two nuisance parameters each:

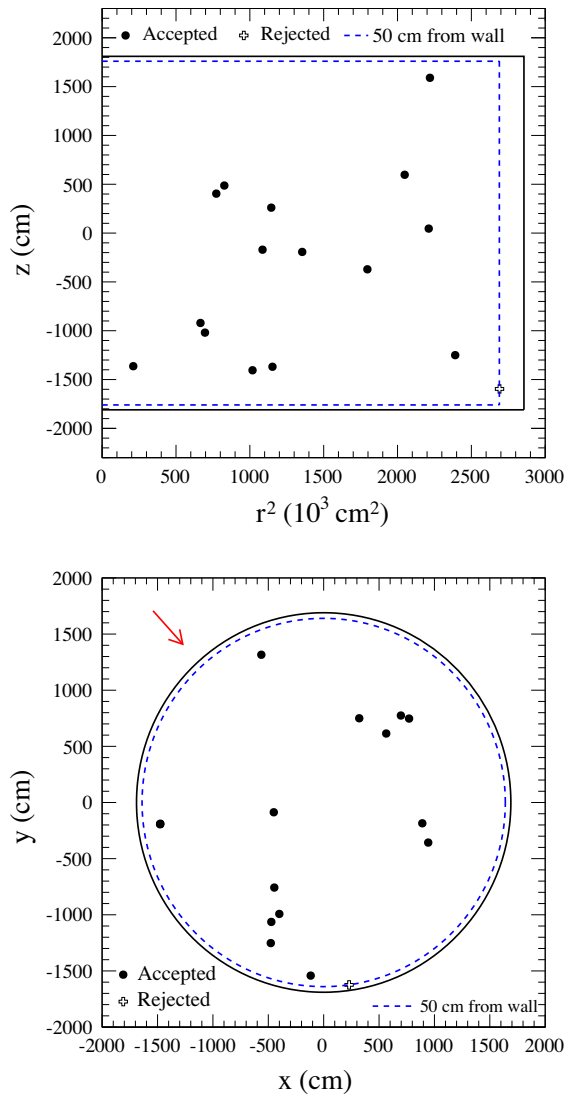$$\Lambda_\beta^{\prime\alpha} = a\Lambda_\beta^\alpha + b, \tag{13}$$

FIG. 24.   Reconstructed vertices of selected FHC data $1R_e +$ 1 d.e. events projected in z vs r^2 (top) and y vs x (bottom). The neutrino beam direction is shown as a red arrow and events which do not pass the fiducial volume criteria are shown as hollow crosses.

FIG. 25.   Reconstructed energy distribution for $1R_e$ events in FHC (top) and RHC (bottom) data.

where $\Lambda'^{\alpha}_{\beta}$ is the transformed variable and a and b are "scale" and "shift" nuisance parameters, respectively. These scale and shift parameters are estimated with a Markov Chain Monte Carlo (MCMC) method [88] that samples the Poisson likelihood for the observed data given the model that includes, in addition, parameters to capture the uncertainty on the atmospheric neutrino flux and cross sections. The SK atmospheric neutrino data used in this analysis were collected between October 2010 and May 2015. To ensure the validity of the atmospheric neutrino fit for the T2K oscillation analysis, high-statistics control sample data collected over the entire beam data period are used to monitor the detector stability.

The flux and cross-section parameterizations used in this procedure are simpler than those used in oscillation analyses and are based on Refs. [89,90]. The atmospheric neutrino flux is scaled by two parameters, one affecting events with neutrino energy lower than 1 GeV, with a prior uncertainty of 25%, and the other for events with energies higher than 1 GeV, with a prior uncertainty of 15%. The ratio of $\nu_e + \bar{\nu}_e$ to $\nu_\mu + \bar{\nu}_\mu$ events is controlled by a parameter with 5% prior uncertainty. A prior uncertainty of 20% is assigned to both CC non-QE and NC cross sections. The CCQE cross section has a more detailed parameterization, with events below 190 MeV having a prior uncertainty of 100% and events in the 190 MeV to 1 GeV range being characterized by 11 parameters which
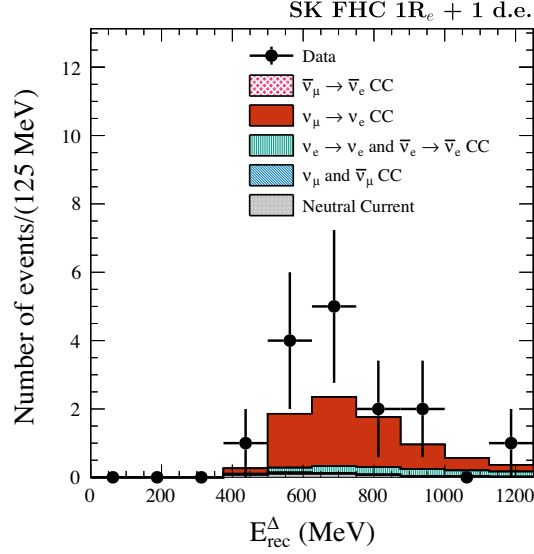
FIG. 26. Reconstructed energy distribution for $1R_e + 1$ d.e. events in FHC data. The $\Delta(1232)$ mass is used for the final-state nucleon in the reconstructed neutrino energy calculation.

scale the cross section in unevenly spaced energy bins with gradually decreasing prior uncertainties from 41% to 2.2%, with 0.6 GeV events being assigned 5.4% prior uncertainty. CCQE events with energy in the 1–2 and 2–3 GeV ranges are assigned 1.7% and 0.9% prior uncertainty, respectively. Finally, each sample used in the fit is assigned an unconstrained overall scale parameter.

The choice of fitted variables and conservative flux and cross-section prior uncertainties reduces the sensitivity of this fit to the neutrino oscillation parameters, which are fixed at the reference values given in Table III.

Examples of the posterior predictive distributions resulting from the MCMC sampling are shown in Fig. 31.

To quantify the impact of the disagreement between data and MC on the T2K samples, the shift and scale parameters are sampled from the MCMC posterior and applied to the T2K beam MC. The T2K selection criteria other than FV are applied, and a fractional uncertainty is calculated for each analysis sample in each of the detector regions, with the MC separated in five true categories: CCQE, CC non-QE, CC events with misidentified lepton flavor, neutral current, and entering backgrounds. In this procedure, the atmospheric neutrino flux and cross-section parameters used in the MCMC fit are marginalized over. The resulting fractional uncertainty on the expected number of events is taken to be the systematic uncertainty associated to the SK detector model.

With the detector systematic uncertainty estimated for each detector region, the optimal FV criteria are found by maximizing the following figure of merit (FOM), which quantifies the sensitivity of the sample with respect to changes in θ_{23} and δ_{CP} for μ-like and e-like samples, respectively:
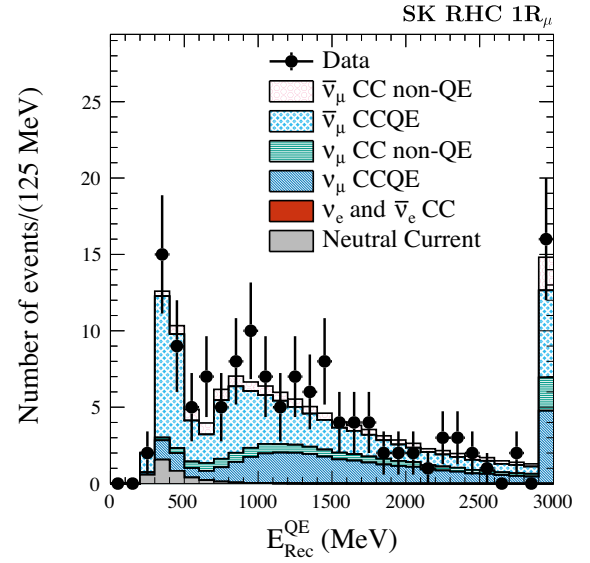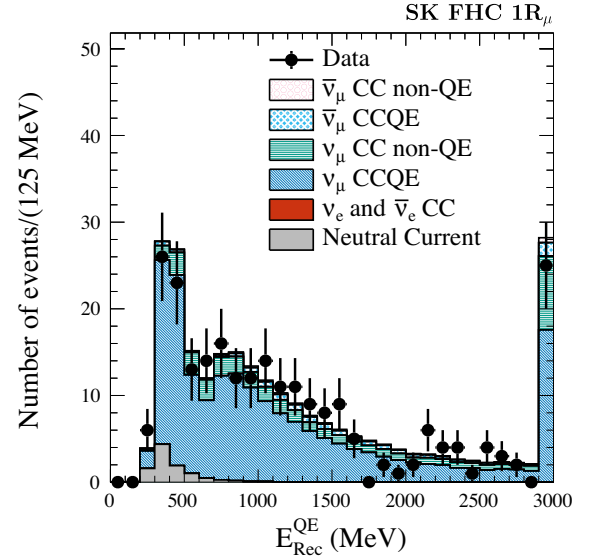

FIG. 27. Reconstructed energy distribution for $1R_\mu$ events in FHC (top) and RHC (bottom) data.

$$\text{FOM} = \frac{\left(\frac{\partial \hat{N}}{\partial \theta}\right)^2}{\hat{N} + \sigma_{\text{syst}}^2}, \qquad (14)$$

where $\hat{N}$ is the expected number of events in a given sample, θ is the parameter of interest (θ_{23} for the μ-like samples and δ_{CP} for the e-like samples) and σ_{syst} is the systematic uncertainty, including uncertainties associated to the detector and cross-section models as well as the uncertainty on the number of entering backgrounds.

A minimum requirement of at least 50 cm in *wall* and 150 cm in *towall* is chosen based on MC studies which show deterioration in momentum reconstruction beyond those regions. Different FV criteria are chosen for the $1R_\mu$, $1R_e$ and $1R_e + 1$ d.e. samples, with the same cuts applied
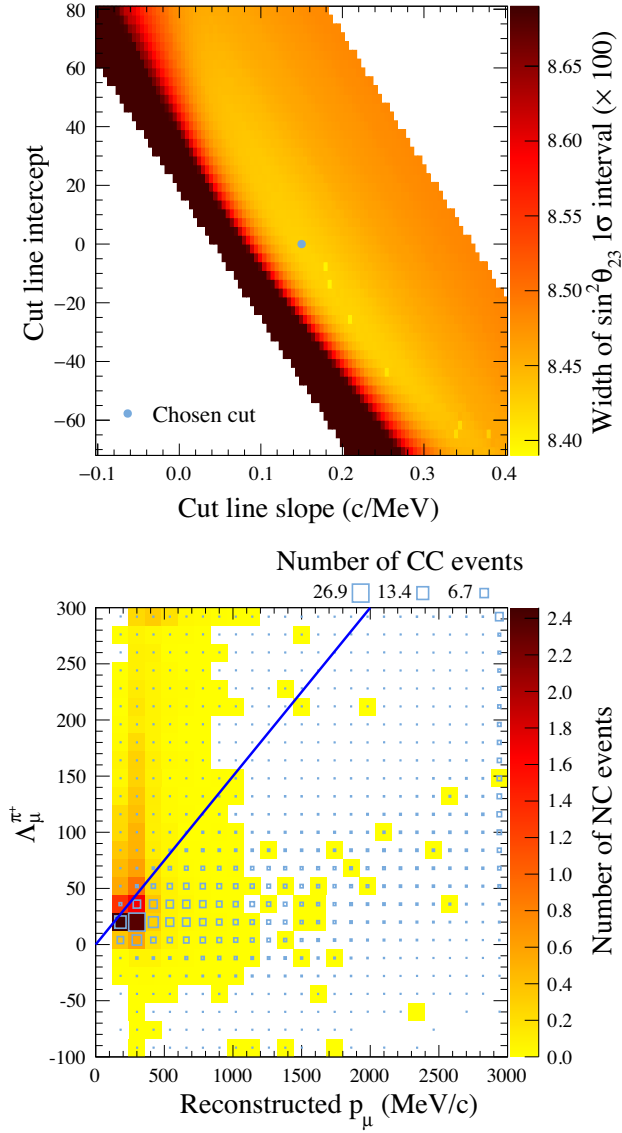
FIG. 28. The width of $\sin^2\theta_{23}$ 1σ confidence interval as a function of the slope and intercept of the NCπ^+ rejection cut line is shown on the top and the distribution of signal (box histogram, axis on top of plot) and background events (colored histogram) is shown on the bottom, along with the line below which μ-like events are selected.

FIG. 29. The sensitivity to exclude $\delta_{CP} = 0$ as a function of the slope and intercept of the NCπ^0 rejection cut line is shown on the top and the distribution of signal (box histogram, axis on top of plot) and background events (colored histogram) is shown on the bottom, along with the line below which e-like events are selected.

in equivalent selections in neutrino and antineutrino modes. The figure of merit is shown as a function of *wall* and *towall* for the three samples in Fig. 32. The optimal criteria resulting from this procedure are described in Table IV.

The combined effect of the new reconstruction algorithm, optimized neutral-current rejection cuts and expanded FV volume has a significant impact across all analysis samples. The signal acceptance in the 1R$_e$ sample increases by 20%, with the same purity as in previous analyses, while the signal events in the 1R$_e$ + 1 d.e. sample increase by 30%, with a reduction in the misidentified muon background of 70%. The 1R$_\mu$ samples have an

increase in signal efficiency of 15% and a reduction in backgrounds of 40%.

E. Systematic uncertainty

The systematic uncertainty associated with the SK event selection is propagated to the oscillation analysis fitting frameworks as a covariance matrix in bins of either reconstructed neutrino energy or reconstructed lepton momentum and broken down in true event topologies.

Systematic uncertainties on the particle count and identification are extracted from the atmospheric neutrino data with the MCMC method described above.
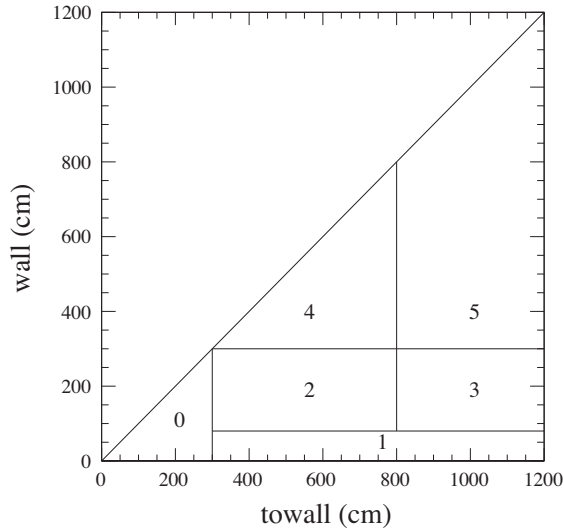
FIG. 30. Detector regions used in detector systematic uncertainty estimation.

The uncertainties on decay-e tagging and misidentification of muons as electrons are extracted from differences between the data and the MC in a control sample consisting of cosmic ray muons that stop within the ID. The cosmic ray muon events are weighted in momentum and *towall* to match the expected distribution of beam-induced muons. The uncertainty on the decay-e identification efficiency is 1% and the uncertainty on the rate of spurious decay-e tags is 0.2%. The relative uncertainty on the misidentification of muons as electrons is 30%, though the contamination of ν_μ CC events is smaller than 1% in the ν_e samples without decay electrons and around 2% in the $1R_e + 1$ d.e. sample.

Uncertainties introduced by the FV criteria are also estimated with MC to data comparisons in the cosmic ray muon sample, with both vertex and direction uncertainties taken into account. The reconstructed vertices in the cosmic ray muon sample cluster at the top or side walls of the detector, allowing for shifts in the MC relative to the data to be identified. The uncertainty on the direction is estimated
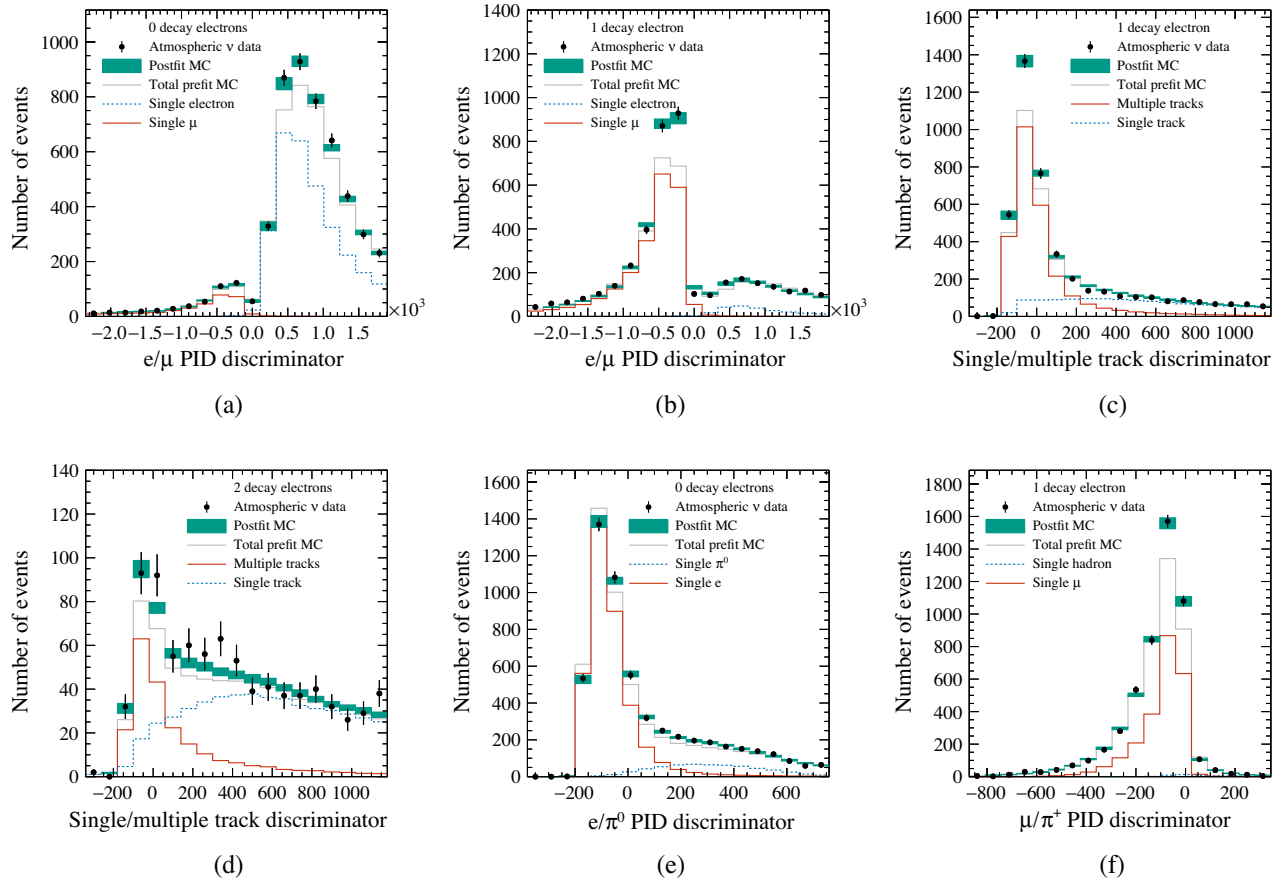

FIG. 31. Posterior predictive distributions of SK atmospheric MCMC fit to determine detector systematic uncertainties. The total nominal MC is shown in gray, with components targeted by each of the distributions shown in red and blue. The 68.27% intervals of the posterior predictive distribution is shown in green. Observed data are shown as black circles. All distributions shown are for the largest detector region, 5. Distributions of the e/μ PID discriminator for zero and one decay-e events are shown in (a) and (b). In (c) and (d), the track-counting parameter is shown for events with one and two decay electrons, respectively. The e/π^0 discriminator distribution is shown in (e) for e-like events, and the μ/π^+ distribution is shown in (f) for μ-like events.
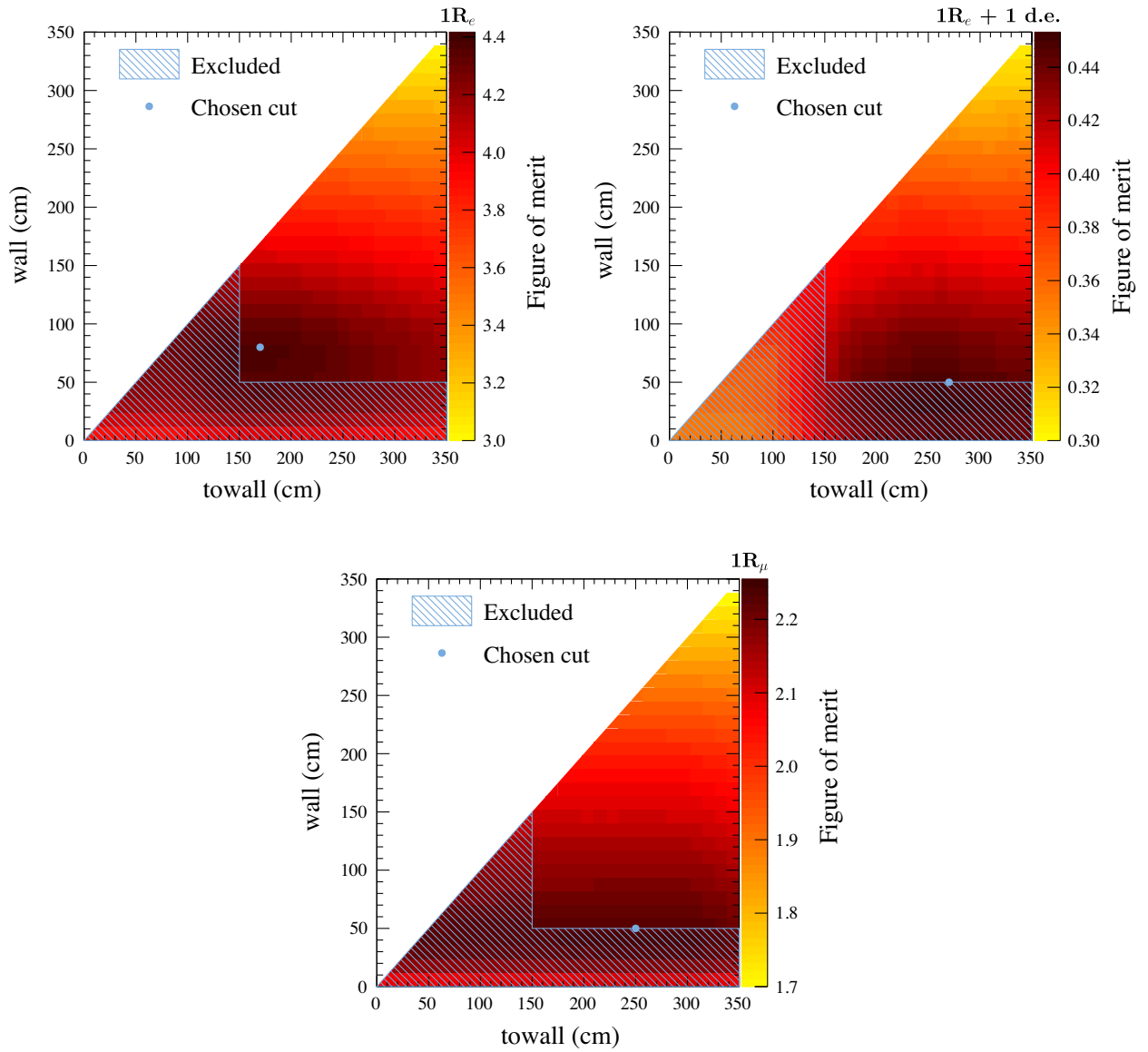
FIG. 32.   Scans of the figures of merit used to optimize the FV criteria in the *wall* vs *towall* space that defines the FV for the analysis described here. Figures of merit corresponding to the $1R_e$ (top left), $1R_e + 1$ d.e. (top right) and $1R_\mu$ (bottom) samples are shown here. The chosen cut point is indicated in the figures as is the region which is ruled out by reconstruction performance considerations.

TABLE VIII.   Super-Kamiokande detector systematic uncertainties. With the exception of the spurious decay-e tagging rate, the 1σ uncertainties are given as a fraction of the nominal value. The uncertainties on the particle identification variables given in the last row are extracted from the atmospheric neutrino fit described in the text and are not easily expressed as intervals.

| Source	Nominal value	1σ uncertainty	Sample
Decay-e tagging efficiency	90%	1.0%	Cosmic ray muon
Spurious decay-e tagging rate	0.01/100 events	0.2/100 events	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
e/μ, π^0/e, π^+/μ, single- or multitrack identification	$\cdots$	$\cdots$	Atmospheric neutrino
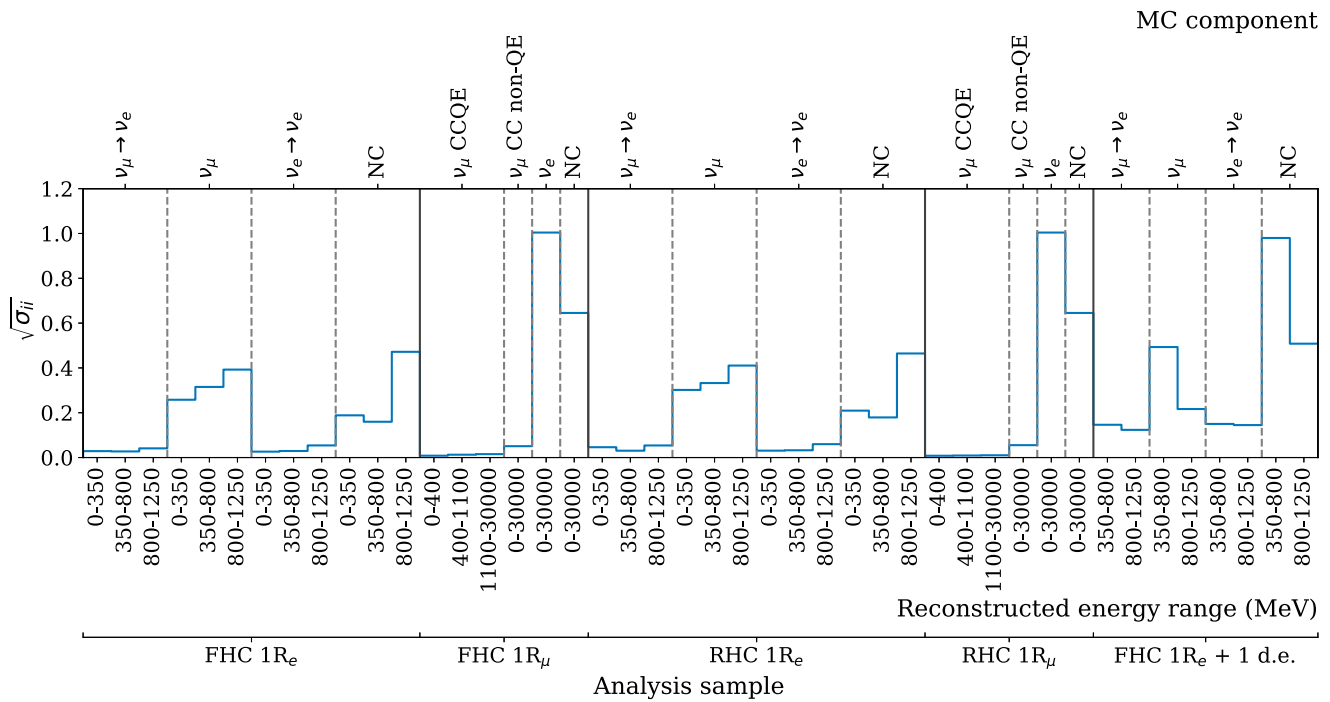
FIG. 33. Square root of the diagonal elements of the covariance matrix describing the fractional uncertainty on the number of events due to the modeling of the SK detector.
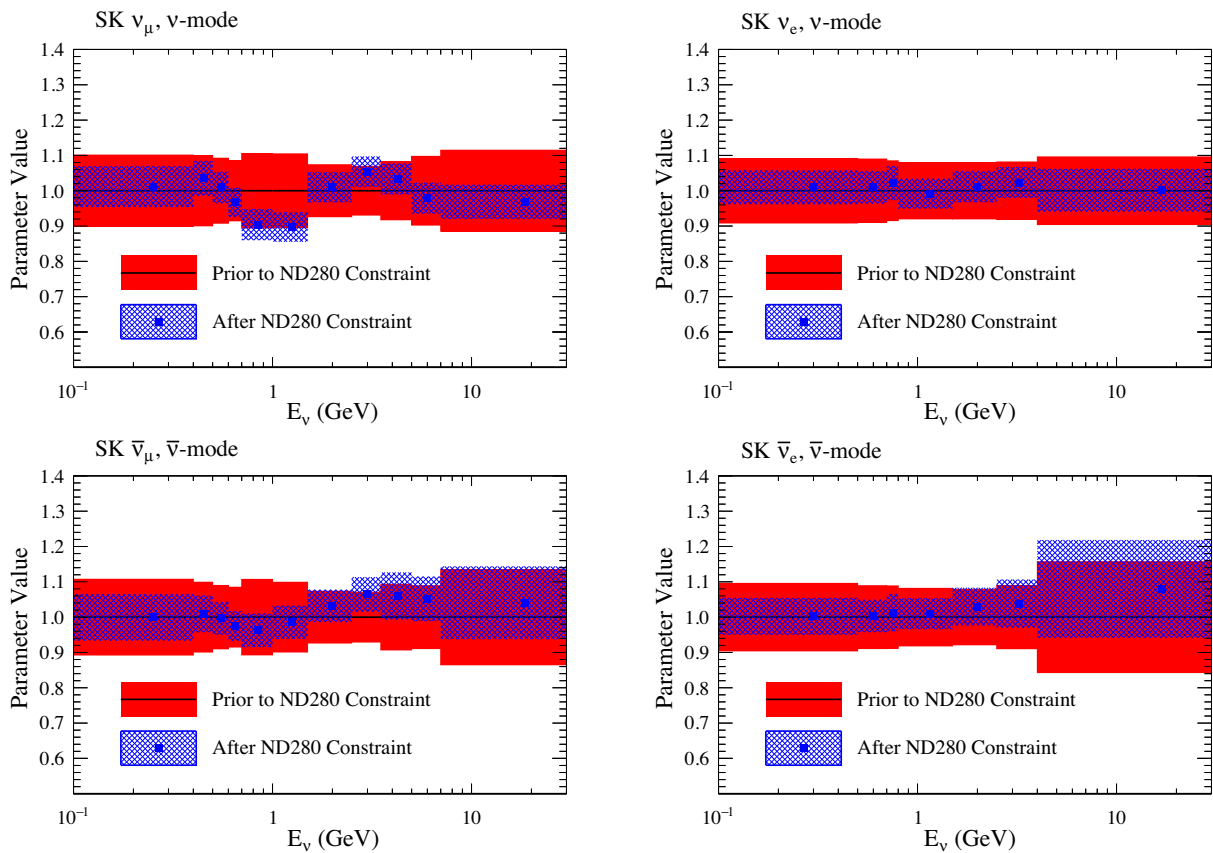

FIG. 34. The SK flux parameters for the ν_μ (top left) and ν_e (top right) neutrino species in FHC and for the $\bar{\nu}_\mu$ (bottom left) and $\bar{\nu}_e$ (bottom right) neutrino species in RHC, as a fraction of the nominal value. The bands indicate the 1σ uncertainty on the parameters before (solid, red) and after (hatched, blue) the near detector fit. The results shown here are from the MINUIT-based framework.
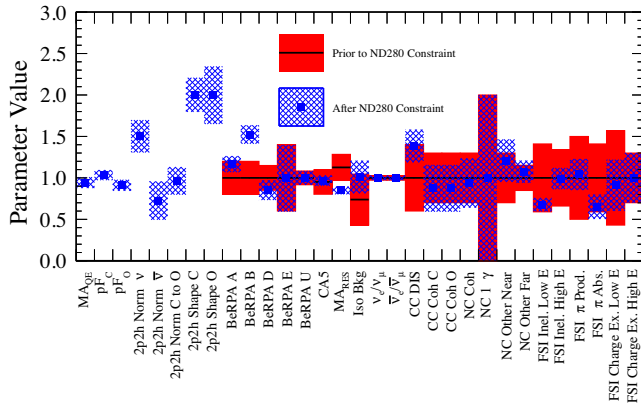
FIG. 35.   The cross-section parameters as a fraction of the nominal value, taken from the MINUIT-based analysis. The bands indicate the 1σ uncertainty on the parameters before (solid, red) and after (hatched, blue) the near detector fit.
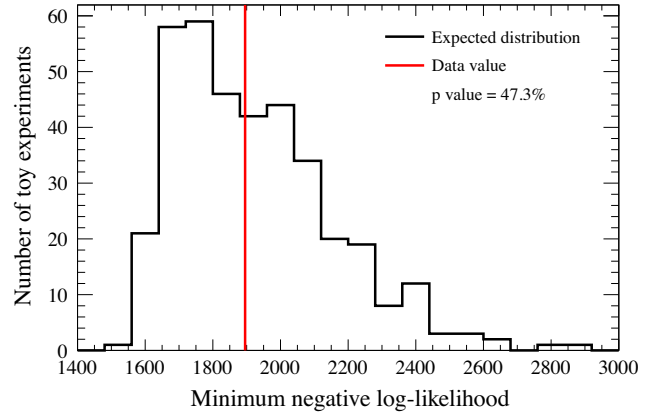

FIG. 36.   Distribution of the minimum negative log-likelihood values from fits to the mock datasets (black line), with the value from the fit to the data superimposed in red. The p value is 47.3%.

by comparing the reconstructed muon direction to the equivalent quantity estimated using the muon and subsequent decay-e vertices. The uncertainties are 2.5 cm for the vertex position and 0.24° for the direction, corresponding to a 0.3%–0.4% systematic uncertainty on the FV, depending

on the analysis sample. This uncertainty is dominated by the uncertainty on the vertex position, with the direction playing a negligible role.

The uncertainty on the π^0 rejection efficiency in $1R_e$ samples is estimated using hybrid π^0 sample constructed by
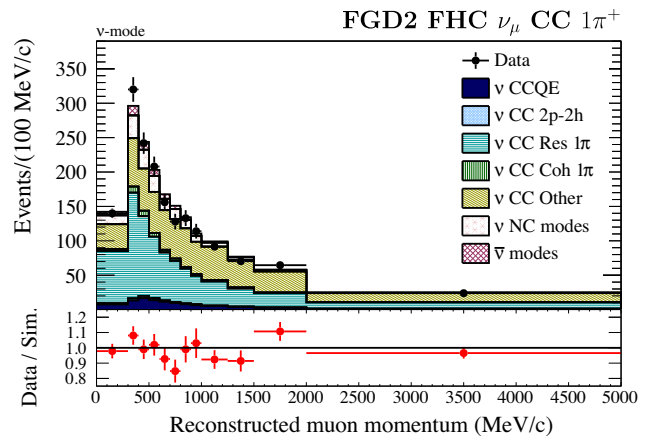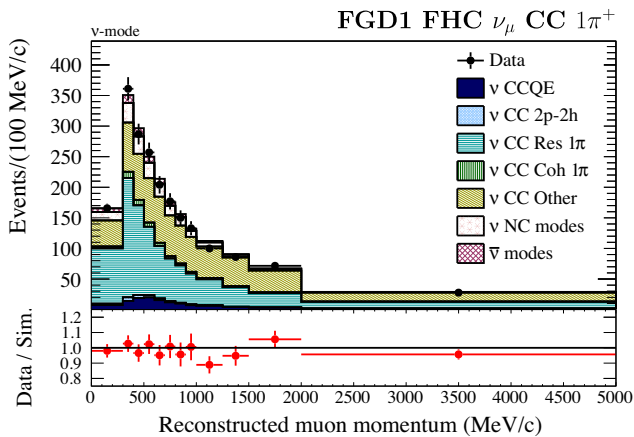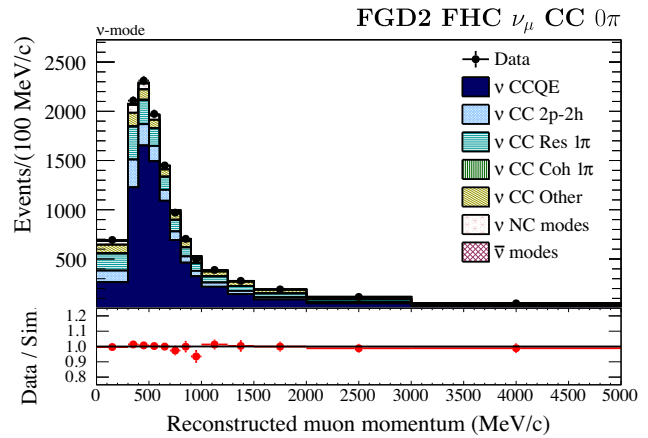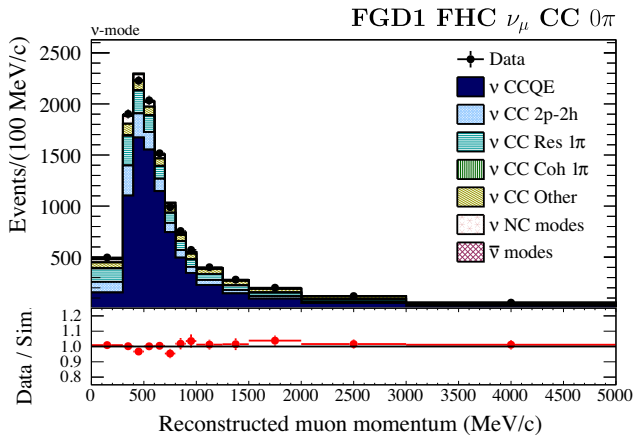

FIG. 37.   Post-ND280-fit muon momentum distributions of the FHC ν_μ CC 0π (top) and FHC ν_μ CC 1π (bottom) samples in FGD1 (left) and FGD2 (right).

superimposing an *e*-like event from the atmospheric neutrino or decay-*e* from cosmic ray muon data with a simulated γ with kinematics taken from NCπ^0 events in the MC. The procedure is performed using both the MC and the real data as the source of the event and the difference in π^0 rejection efficiency between the data-MC and MC-MC samples is taken as the systematic uncertainty, binned in reconstructed lepton momentum and angle with respect to the beam. The overall uncertainty on the π^0 rejection efficiency is 26%.

A summary of the uncertainties associated with the SK detector model is given in Table VIII.

To propagate the systematic uncertainty on the SK event selection to the neutrino oscillation analysis frameworks, a covariance matrix is computed using the beam MC and the uncertainties quoted above. The MC is weighted with the flux and cross-section parameters at their central value from the fit to the near detector data and neutrino oscillation weights using the parameters in Table III are applied. Variations of the MC are then produced with random throws of the systematic effects described above and projected into bins of true event topology and reconstructed neutrino energy or lepton momentum to produce the

covariance matrix. The diagonal elements of the covariance matrix in reconstructed neutrino energy are shown in Fig. 33.

XI. NEAR-TO-FAR EXTRAPOLATION FIT

Parameterized flux and cross-section models are used to calculate the predicted event rates at ND280 and SK. These models are fit to the high statistics near detector data to constrain the parameter uncertainties and tune their central values. The expected ND280 and SK event rate is calculated using the same neutrino interaction model and a correlated neutrino flux. This allows the ND280 data to tune the unoscillated neutrino event rate prediction for SK and reduce the uncertainty in the prediction coming from the neutrino flux and interaction model.

There are two near detector fitting frameworks used in this analysis. One fitter uses MINUIT [91] to find the parameters which maximize the likelihood, while the other uses MCMC methods to sample the parameter space. Both frameworks treat the systematics identically and apply parameter variations on an event-by-event basis in the fit. The resulting parameter values from the MINUIT-based fit are used by two of the oscillation analyses (detailed in
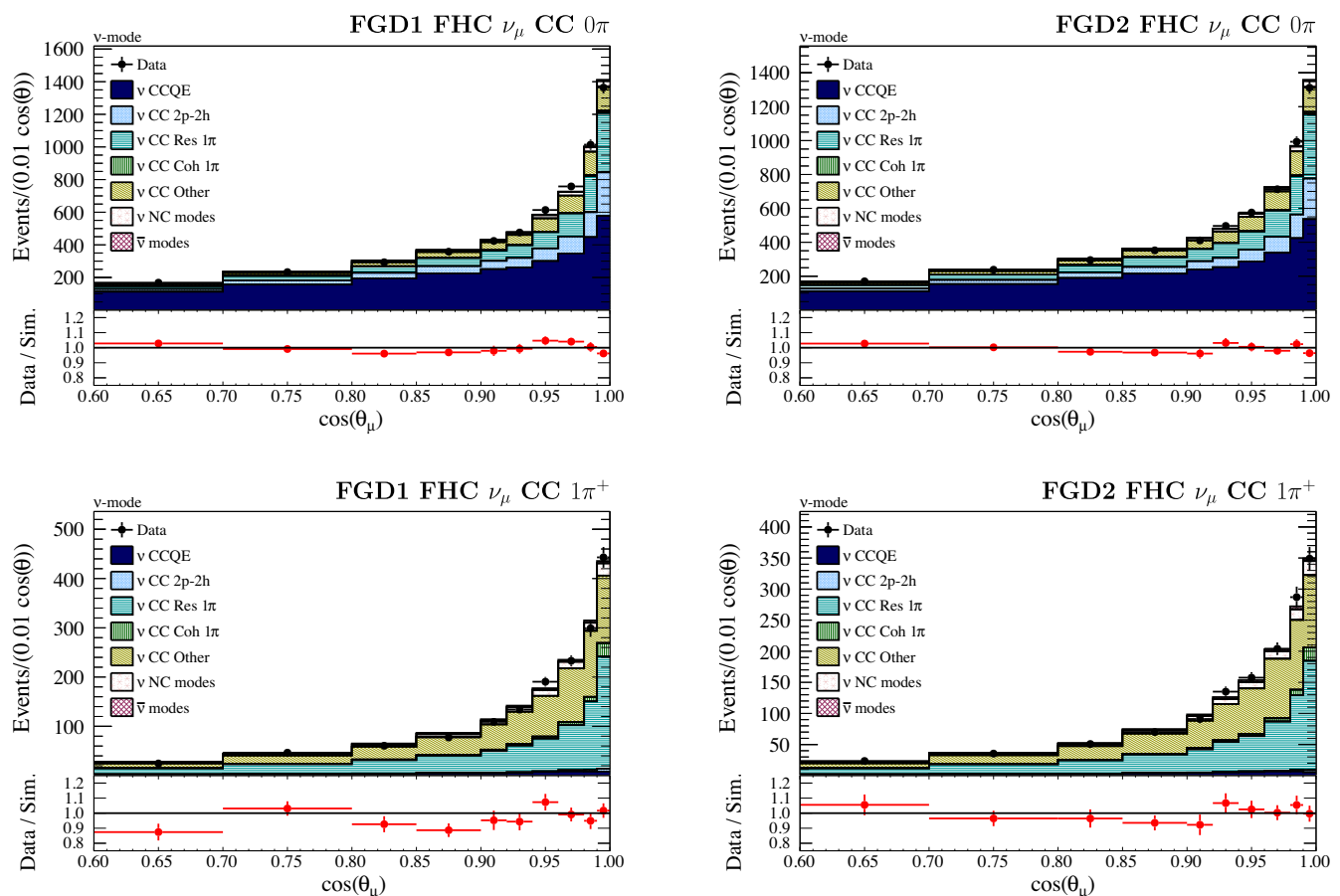

FIG. 38. Post-ND280-fit distributions of the final state muon angle of the FHC ν_μ CC 0π (top) and FHC ν_μ CC 1π (bottom) samples in FGD1 (left) and FGD2 (right).
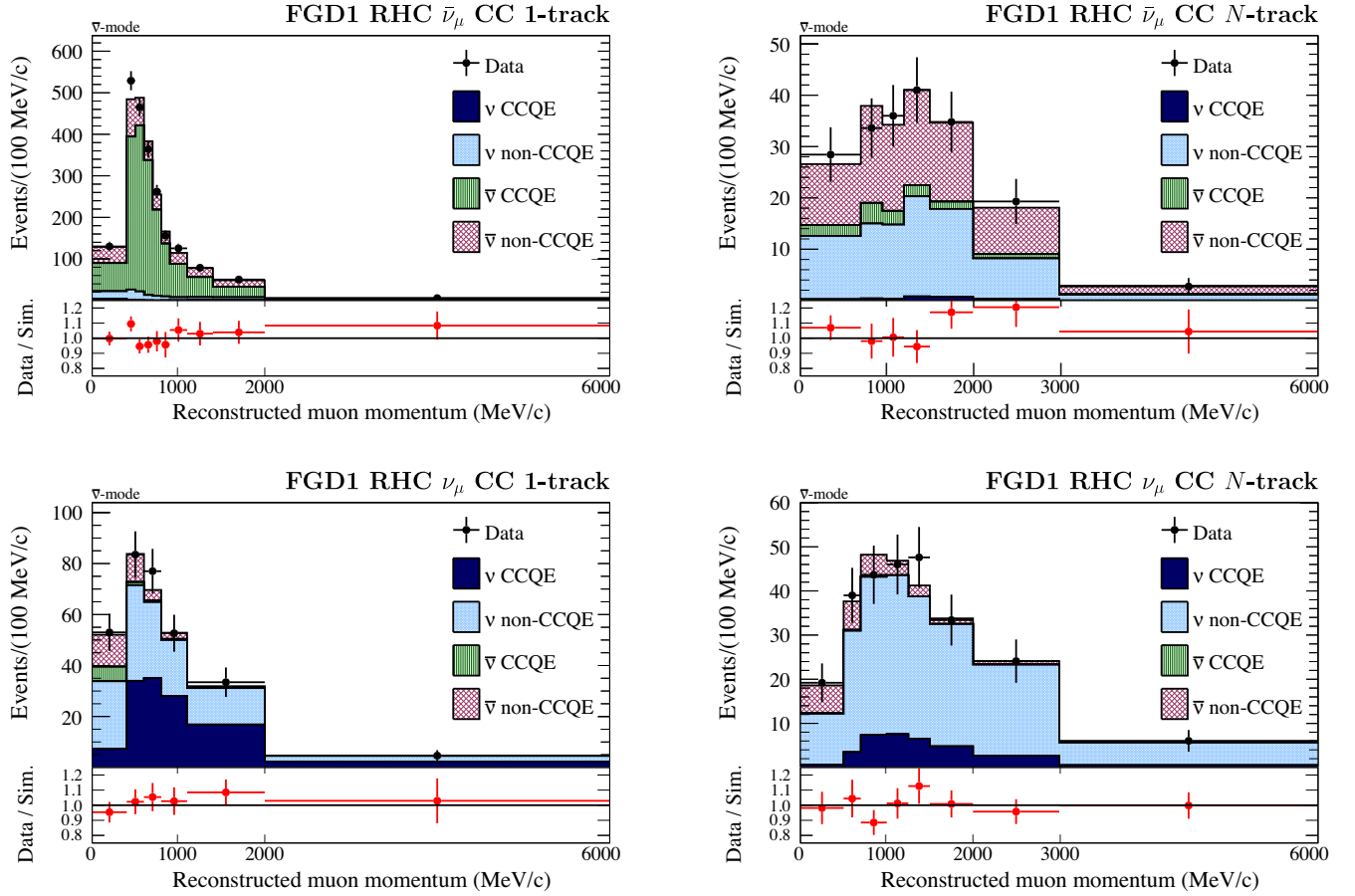
FIG. 39.   Post-ND280-fit muon momentum distributions for the RHC $\bar{\nu}_\mu$ (top) and ν_μ (bottom) CC 1-track (left) and CC *N*-track (right) FGD1 samples.

Sec. XII), with a covariance matrix describing their uncertainties. The MCMC analysis performs joint near and far detector data fits.

An additional uncertainty is included in the flux covariance to account for the fact that the near detector fit results for runs 2–6 are extrapolated to the far detector, which uses a larger dataset from runs 2–9. As in Ref. [26], additional uncertainties which affect $\overset{(-)}{\nu}_e$ events have been introduced. These account for effects which may potentially affect $\overset{(-)}{\nu}_e$ but not $\overset{(-)}{\nu}_\mu$ cross sections.

The ND280 likelihood and fitting methods are unchanged since the analysis described in Ref. [26]. The 14 event samples are binned in p_μ and $\cos\theta_\mu$, giving 1624 bins in total. The full likelihood includes a contribution from the binned χ^2 data-model comparison and a prior penalty contribution for each parameter. The only change to the fitting frameworks since the last analysis is the treatment of the Fermi surface momentum (p_F) systematics near their physical boundaries. Previously, the covariance could not be calculated when the parameter was at its limit, causing the MINUIT fit to not converge. For this analysis, the penalty contribution to the likelihood was "mirrored"

around the physical boundaries for the p_F parameters. This meant the parameters were allowed to pass beyond their physical boundaries. The mirroring was performed by setting the likelihood for values beyond the boundary to the value of the likelihood the same distance from the boundary on the other side. For example, for a physical boundary at $+1.0$, the likelihood at $+1.2$ would be the same as the likelihood at $+0.8$. This allowed the uncertainty to be calculated at the limit and the fit to converge.

The prefit and postfit SK flux and cross-section parameter values and uncertainties are shown in Figs. 34 and 35, respectively, as a fraction of the nominal values. The central values and uncertainties for all parameters are tabulated in the Appendix. There is a significant reduction in the postfit uncertainty for the majority of parameters. Those that are not constrained by the near detector fit are uncertainties that only apply to interactions with low statistics in the near detector. In the last analysis, Ref. [26], the neutrino flux increased for all samples and species, but this effect is no longer present. The difference between the nominal simulation and data is now being absorbed by the movement of other parameters, in particular the BeRPA model.
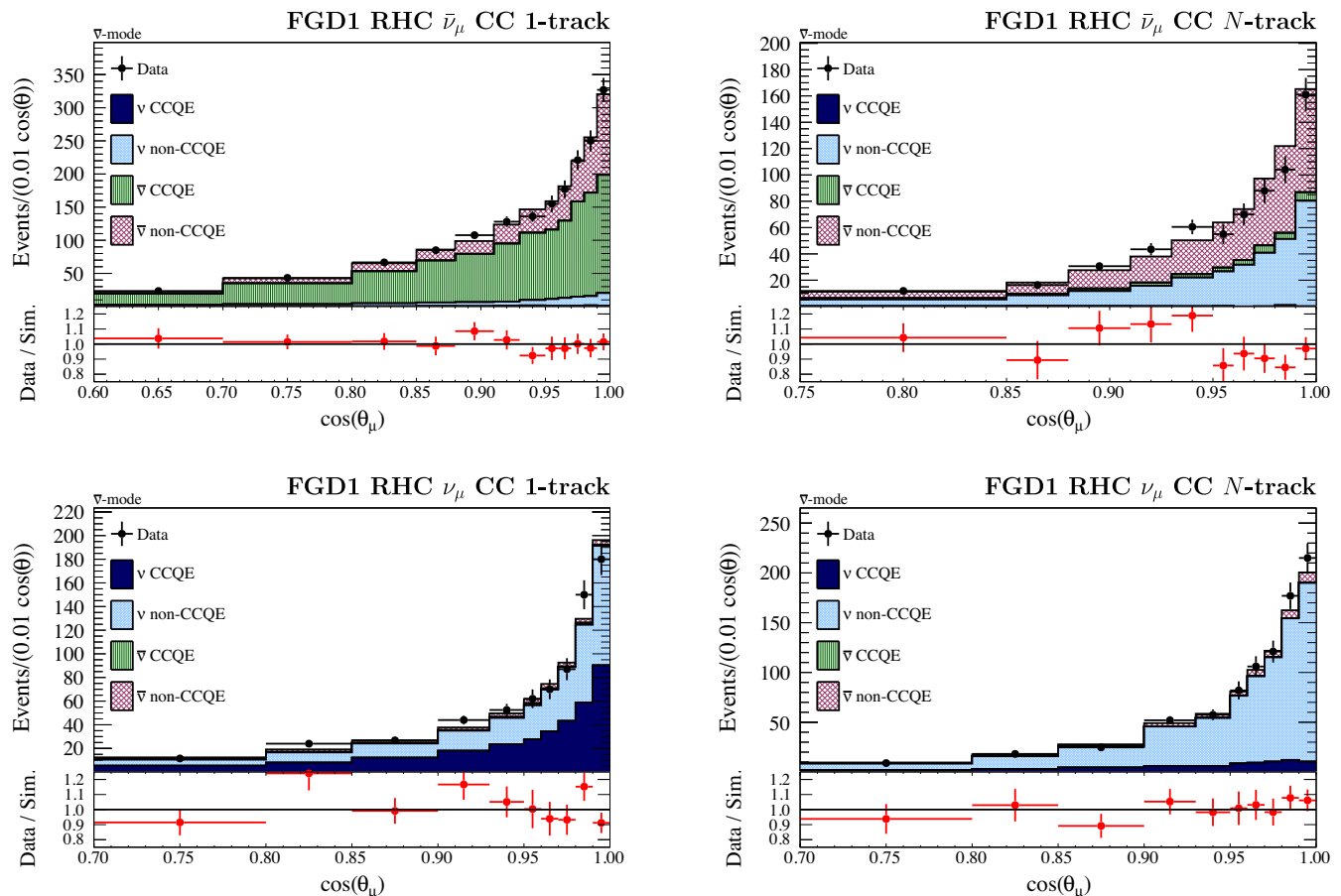
FIG. 40. Post-ND280-fit distributions of the final state muon angle for the RHC $\bar{\nu}_\mu$ (top) and ν_μ (bottom) CC 1-track (left) and CC N-track (right) FGD1 samples.

The goodness of fit for the near detector analysis was estimated by calculating the p value in the MINUIT-based framework. Toy datasets were produced for the near detector by randomly varying all systematics (neutrino flux, neutrino interaction and detector modeling uncertainties) according to their prior covariance and applying them to the nominal simulation prediction. The likelihood for each toy dataset is shown in Fig. 36, along with the likelihood from the data fit. The overall p value for the fit is 47.3%.

The postfit muon momentum and angle distributions of events are produced by applying the best-fit parameter values to the nominal simulation. These are shown broken down by interaction mode for each sample in Figs. 37–40, along with the observed distributions. There is much better agreement with the data than for the prefit distributions shown in Figs. 11–14. The numbers of postfit predicted events for all 14 samples are shown in Table II.

The best-fit values are also used to produce SK event rate predictions, shown for the FHC $1R_\mu$ and $1R_e$ SK samples in Figs. 41 and 42. The total predicted and observed event rates are shown in Table IX.

Near detector only fits with the MCMC analysis framework were used to cross-check the two analyses. The postfit

cross-section parameters for the two fits are compared in Fig. 43, showing good agreement.

XII. OSCILLATION ANALYSIS FITTERS

To produce constraints on the three-flavor PMNS oscillation parameters, the rate and kinematic distributions of all five SK event samples are analyzed simultaneously. Systematic uncertainties in the flux, interaction and detector models are accounted for using systematic parameters applied as weights to the nominal prediction as described in [26]. Confidence regions and intervals are produced from marginal likelihood distributions as a function of parameters of interest.

The predicted kinematic distributions of each SK sample are generated from the nominal SK simulation to which weights are applied for each set of oscillation and systematic parameter values. The oscillation weights correspond to the three-flavor oscillation probability, calculated with matter effects and dependent on the true neutrino energy and flavor [22]. All systematic parameters are varied to give a multiplicative weight, covering the neutrino flux, cross-section and SK detector uncertainties.
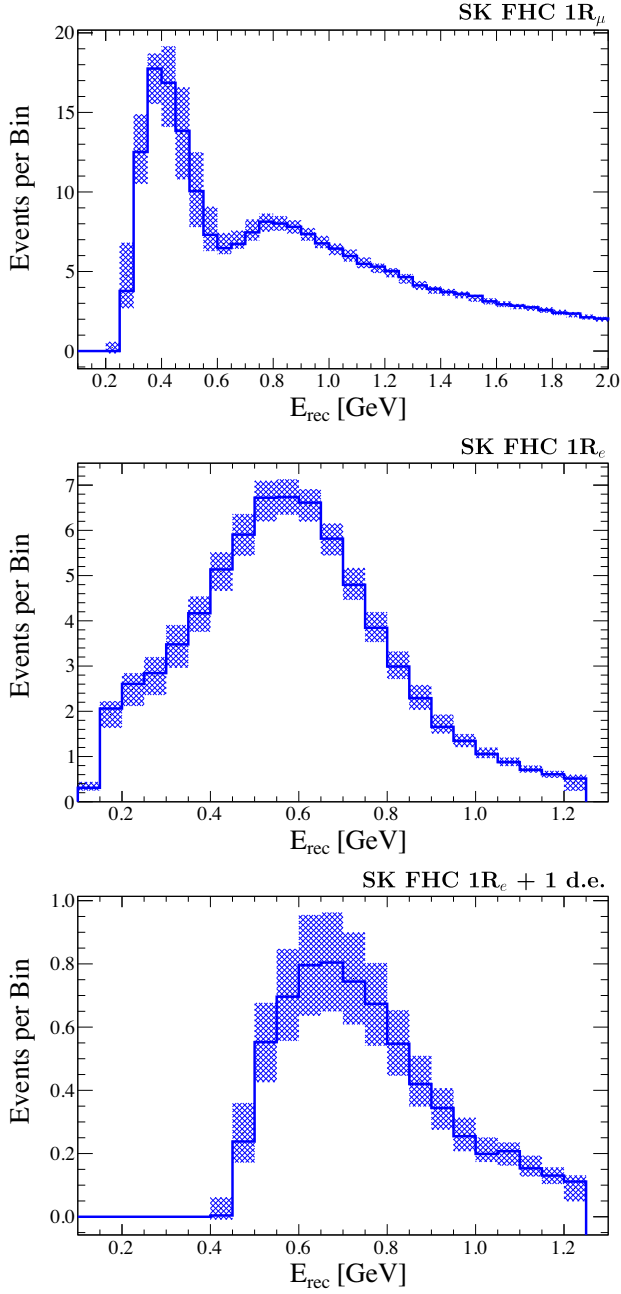
FIG. 41.   Post-ND280-fit predicted event spectrum for the FHC beam SK samples as a function of E_{QE}^{Rec}. The samples are single-ring muonlike (top), single-ring electronlike without decay electrons (middle) and single-ring electronlike with a single decay electron (bottom).

A likelihood L is calculated according to Eq. (15) as the product of the Poisson likelihood ratios for the number of events in each bin of the kinematic variables considered for each SK sample:

$$-2 \ln L(\boldsymbol{o}, \boldsymbol{f}) = 2 \sum_{s=0}^{5} \sum_{i=0}^{N_s-1} \left(n_{s,i}^{\mathrm{obs}} \cdot \ln \left(n_{s,i}^{\mathrm{obs}} / n_{s,i}^{\mathrm{exp}} \right) \right.$$
$$\left. + \left(n_{s,i}^{\mathrm{exp}} - n_{s,i}^{\mathrm{obs}} \right) \right). \tag{15}$$

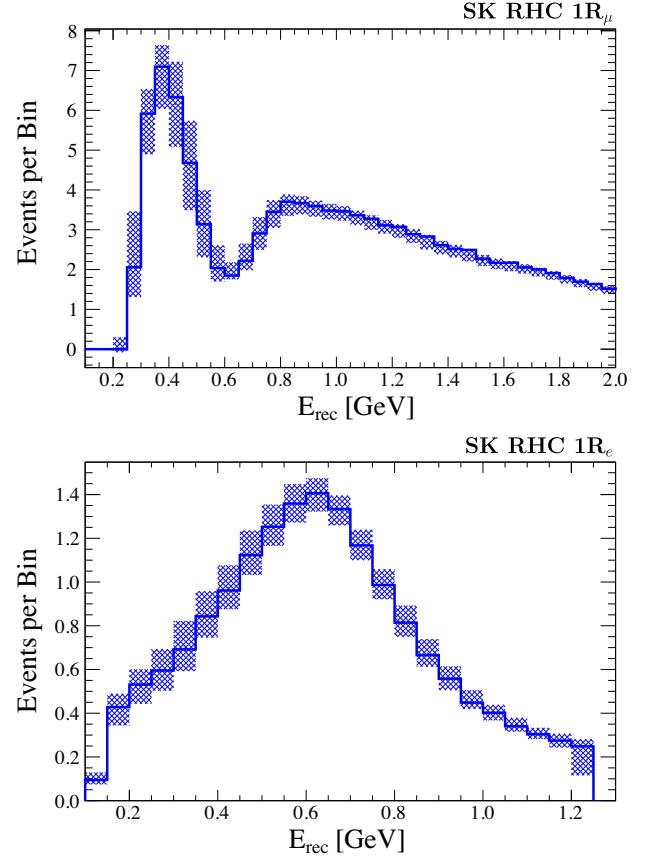FIG. 42.   Post-ND280-fit predicted event spectrum for the RHC SK samples as a function of E_{QE}^{Rec}. The samples are single-ring muonlike (top) and single-ring electronlike without decay electrons (bottom).

In Eq. (15), $\boldsymbol{o}$ is a vector of the parameters of interest, $\boldsymbol{f}$ is a vector of nuisance parameters, $n_{s,i}^{\mathrm{obs}}$ is the observed number of events in kinematic bin i of SK sample s which has a total of N_s bins, and $n_{s,i}^{\mathrm{exp}} = n_{s,i}^{\mathrm{exp}}(\boldsymbol{o}, \boldsymbol{f})$ is the corresponding expected number of events. The parameter(s) of interest correspond to one or more oscillation parameters among $\sin^2 \theta_{13}$, $\Delta m_{32/31}^2$, $\sin^2 \theta_{23}$, and δ_{CP}. The nuisance parameters correspond to the systematic parameters, and the oscillation parameters not chosen as parameters of interest in a given fit. To obtain a likelihood which only depends on

TABLE IX.   Observed and predicted event rates for SK samples, pre- and post-ND280 fit.

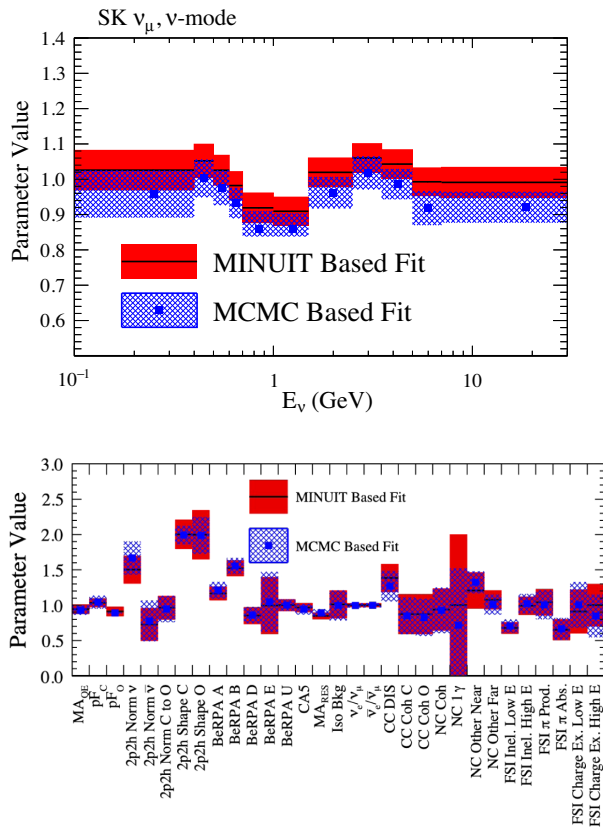| Sample	Data	Prefit	ND280 Postfit
FHC $1R_\mu$	243	250.45	272.37 ± 6.05
FHC $1R_e$	75	64.17	72.79 ± 1.52
FHC $1R_e + 1$ d.e.	15	7.80	6.87 ± 0.41
RHC $1R_\mu$	140	130.53	139.42 ± 2.57
RHC $1R_e$	15	15.73	16.77 ± 0.36

FIG. 43. Comparisons of the fitted SK FHC flux parameters (top) and cross-section parameters (bottom) between the ND280 fit (red, solid) and the MCMC analysis (blue, hashed).

the parameter(s) of interest o and a dataset x [the set of $n_{s,i}^{\mathrm{obs}}$ in Eq. (15)], a marginal likelihood L_{marg} is computed: the full likelihood, made of the product of the likelihood L defined in Eq. (15) with the prior constraint on some of the parameters $\pi(o,f)$, is numerically integrated over the nuisance parameters:

$$L_{\mathrm{marg}} = L_{\mathrm{marg}}(o;x) = \int L(o,f;x)\pi(o,f)\mathrm{d}f. \quad (16)$$

External constraints are used for some of the oscillation parameters. The solar parameters, which have limited impact on the observed event distributions at T2K, are either kept fixed to their nominal values or have a Gaussian constraint applied, depending on the analysis considered. The nominal values and uncertainties used are $\sin^2 2\theta_{12} = 0.846 \pm 0.021$ and $\Delta m_{21}^2 = (7.53 \pm 0.18) \times 10^{-5}\,\mathrm{eV}^2/c^4$ [30].

The three parameters $\sin^2 2\theta_{23}$, Δm_{23}^2 and δ_{CP} are unconstrained. T2K is sensitive to $\sin^2 \theta_{13}$, but to date, the world's most accurate measurements of this parameter come from reactor neutrino experiments [10–12]. To obtain increased sensitivity to the other oscillation parameters, the reactor average of $\sin^2 2\theta_{13} = 0.0830 \pm 0.0031$ [31] may be used as a Gaussian constraint, which will subsequently be referred to as the "reactor constraint." It is also useful to evaluate the T2K experiment's constraint of the oscillation parameters without the contribution of the reactor experiments, and fits without this constraint on $\sin^2 2\theta_{13}$ were also performed.

Three different analysis frameworks are used to perform the fit of the far detector data. They will be labeled as A, B and C and follow the general procedure described above but differ on a number of points summarized in Table X.

Kinematic information used to fit the data from the electronlike samples.—Two-dimensional distributions are used for those samples, either the combination of the momentum and angle with respect to the beam direction of the particle reconstructed as the lepton (p_{lep}, θ_{lep}) or the reconstructed energy assuming CCQE kinematics (E_{rec}) combined to this angle θ_{lep}.

Oscillation probability calculation.—The events can either be binned in true neutrino energy with an oscillation probability corresponding to the mean true neutrino energy of the bin or have individual oscillation probabilities computed for each event's true neutrino energy.

Use of the near detector data.—The near detector data are either used in a simultaneous fit with the far detector data (analysis B) or they are fit separately, using the results from the MINUIT fit described above to constrain the neutrino flux and interaction systematic uncertainties (analyses A and C). In this second case, the constraint is propagated to the far detector analysis through a covariance matrix. The two methods are expected to lead to different results for the far detector fit if the constraint on the systematic parameters obtained with the near detector data cannot be properly described by a multivariate normal distribution.

Oscillation fitting method.—Two of the analyses use a grid search method, where the likelihood in terms of the parameters of interest is computed for different fixed values of those parameters while marginalizing over the other parameters. This marginalization over the nuisance

TABLE X. Differences between the three far detector analyses.

	A	B	C
e-like sample analysis bins	(E_{rec}, θ_{lep})	(E_{rec}, θ_{lep})	(p_{lep}, θ_{lep})
$P(\nu \to \nu)$ calculation	Binned	Event by event	Binned
ND280 constraint	Covariance matrix	Simultaneous fit	Covariance matrix
Fitting method	Grid search	MCMC	Grid search

(a) FHC 1-Ring μ-like

(b) RHC 1-Ring μ-like

(c) FHC 1-Ring e-like

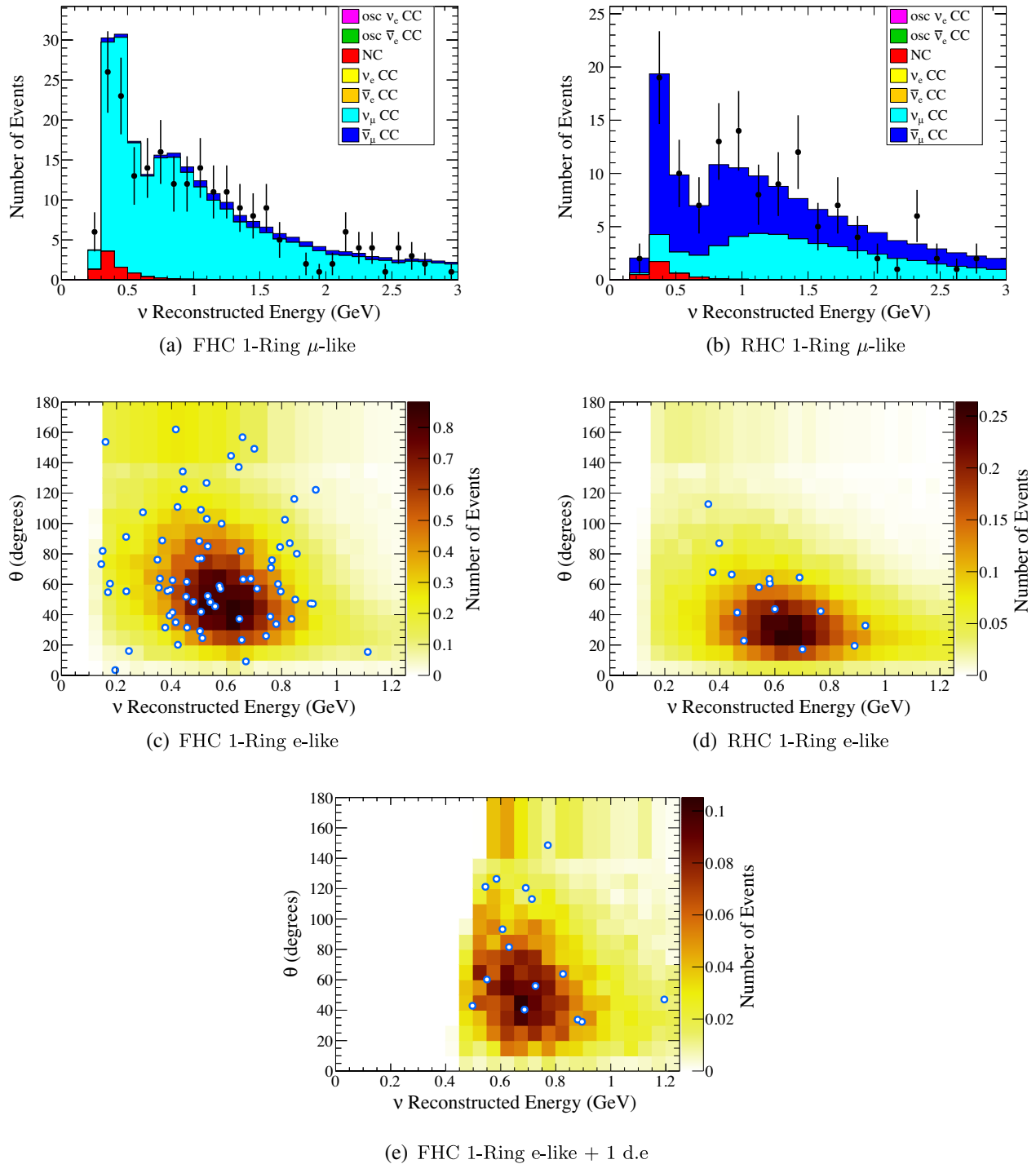

(d) RHC 1-Ring e-like

(e) FHC 1-Ring e-like + 1 d.e

FIG. 44. Observed kinematic distributions compared to the expectations generated with oscillation parameters set to the values in Table III for the different samples. The uncertainty shown around the data points in (a) and (b) accounts for statistical uncertainty only. The uncertainty range is chosen to include all points for which the measured number of data events is inside the 68% confidence interval of a Poisson distribution centered at that point.

parameters is done through numerical integration. The last analysis uses an MCMC method to sample the parameter space, where the density of the obtained samples follows the joint posterior probability density of the parameters. Those samples can then be binned to produce 1D or 2D posterior distributions for the parameter(s) of interest, effectively marginalizing over the nuisance parameters. A full description of the MCMC analysis can be found in [92].

The T2K experiment's expected median constraints on the oscillation parameters can be evaluated by fitting to an
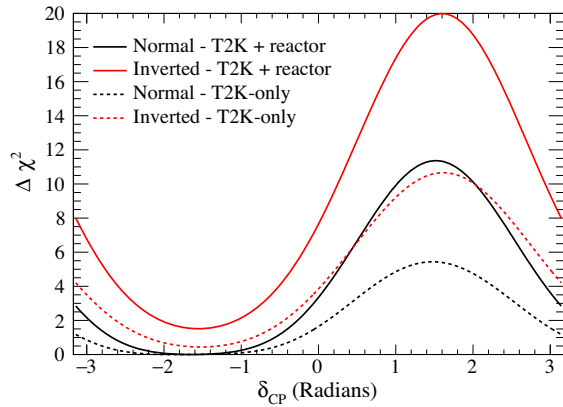
FIG. 45. The predicted $\Delta\chi^2 = -2\ln[L/L_{max}]$ function as a function of δ_{CP} with and without reactor constraint, for both mass orderings.
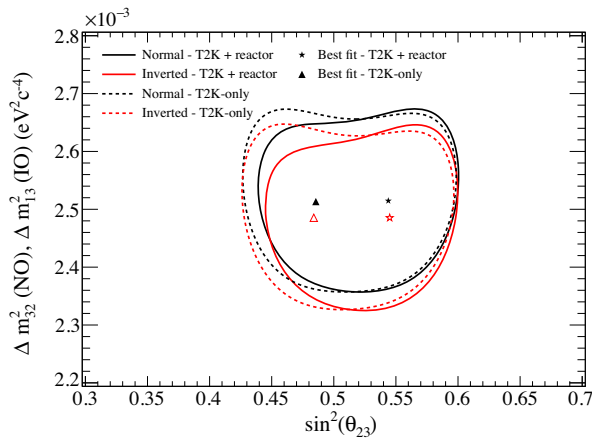

FIG. 46. Predicted 90% C.L. regions for Δm^2 vs $\sin^2\theta_{23}$ with and without the reactor constraint, for both mass orderings. Normal and inverted mass ordering contours are independent.

Asimov dataset, generated for the true values of the oscillation parameters given by Table III and nominal values of the systematic parameters. Using analysis A, the expected sensitivity to δ_{CP} is shown in Fig. 45 and for Δm^2 vs $\sin^2\theta_{23}$ in Fig. 46. The predicted and observed

TABLE XI. Predicted and observed total number of events in each Super-K sample. The oscillation parameters other than δ_{CP} were fixed to the values given in Table III and all systematic parameters were set to their nominal values.

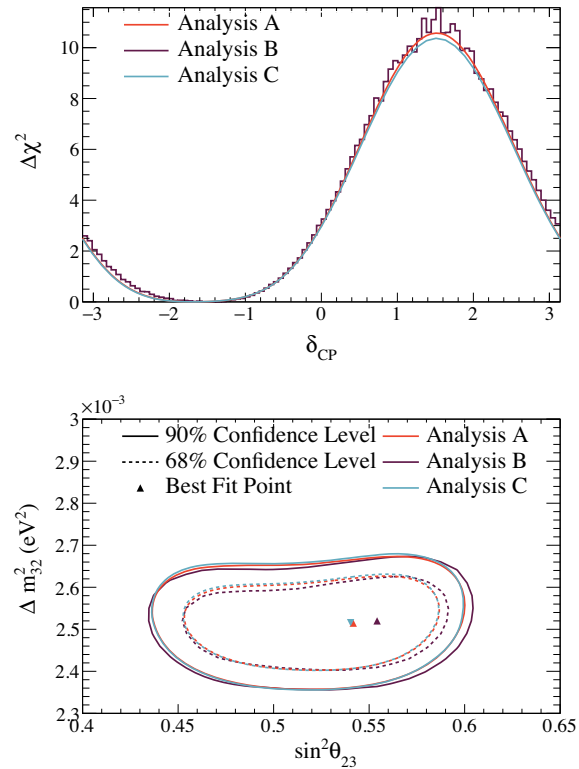| δ_{CP}	FHC μ-like	RHC μ-like	FHC e-like	RHC e-like	FHC ν_e CC 1π
$-\pi/2$	272.4	139.5	74.4	17.1	7.0
0	272.0	139.2	62.2	19.6	6.1
$\pi/2$	272.4	139.5	50.6	21.7	4.9
π	272.8	139.9	62.7	19.3	5.9
Observed	243	140	75	15	15

FIG. 47. Comparison between analyses A, B and C. The normal mass ordering is assumed and the reactor constraint is applied. Top: the predicted $\Delta\chi^2$ functions for δ_{CP}. Bottom: the 1σ (dashed lines) and 90% (solid lines) C.L. regions for Δm^2 vs $\sin^2\theta_{23}$.

event rates in each of the five SK samples are shown in Table XI at various δ_{CP} values, with other oscillation parameters fixed to the values in Table III and with all systematic parameters set to their nominal values. Similarly, the predicted and observed kinematic distributions are shown in Fig. 44, with all oscillation parameters fixed to the values in Table III.
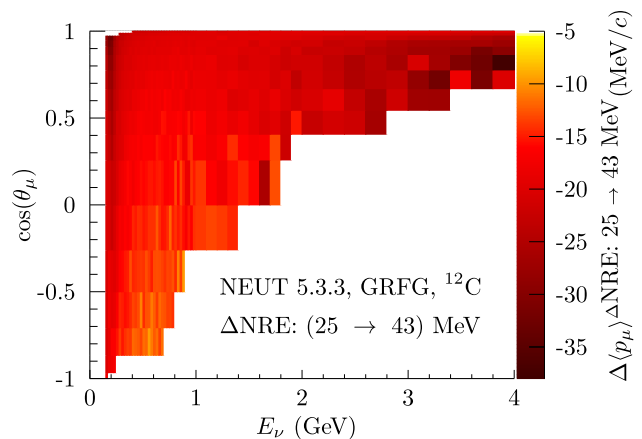

FIG. 48. The momentum shift caused by increasing the nucleon removal energy in NEUT 5.3.3 by 18 MeV.
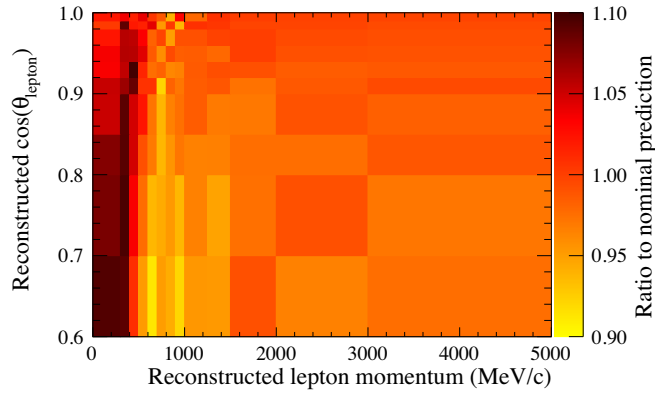
FIG. 49. The ratio of the NRE simulated data to the nominal prediction for the ND280 CC 0π FGD1 sample. The histogram axes have been truncated for clarity.

Using the differences between the three analyses, it can be checked that the results are not too sensitive to the details of how the analysis was performed. The expected sensitivities obtained by each analysis for δ_{CP} and Δm^2 vs $\sin^2 \theta_{23}$ are shown in Fig. 47. They show good agreement between each analysis; thus, the differences in the analysis methods do not signifi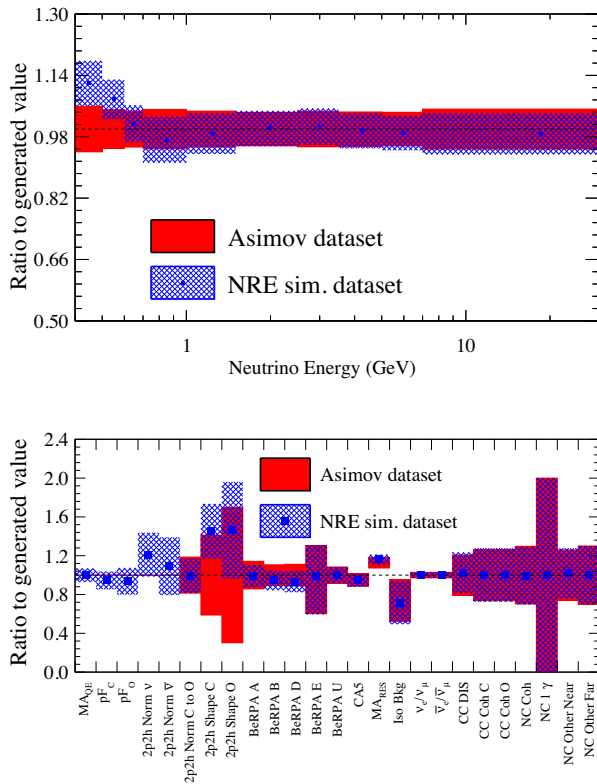cantly impact the results. Additional comparisons between the three analyses were performed for each combination of parameter(s) of interest, mass ordering, use of reactor constraint. In each of these, the results produced by the three analyses were found to be consistent.

XIII. SIMULATED DATA STUDIES

Equation (2) shows that the neutrino oscillation probability depends upon the energy of the neutrino and its path length from creation to interaction. In long-baseline accelerator-based neutrino experiments the neutrino path length is fixed. Accurately measuring the neutrino oscillation parameters therefore requires a precise understanding of the neutrino energy spectrum. All neutrino experiments use models to link the observed final states back to the initial neutrino energy. There is no set of models that describe the world's neutrino data but there are a number of models that are in comparable agreement to the world's data, as described in Ref. [54]. However, these models map true neutrino energy to reconstructed neutrino energy in different ways. This is shown in Fig. 4, where the bottom plot compares the reconstructed energy bias between

FIG. 50. Results of the fit to the near detector simulated data with an increased nucleon removal energy. The best-fit values for the ν_μ flux parameters at SK (top) and cross-section parameters (bottom) are shown for the NRE fit (blue) compared to the Asimov data set fit (red).
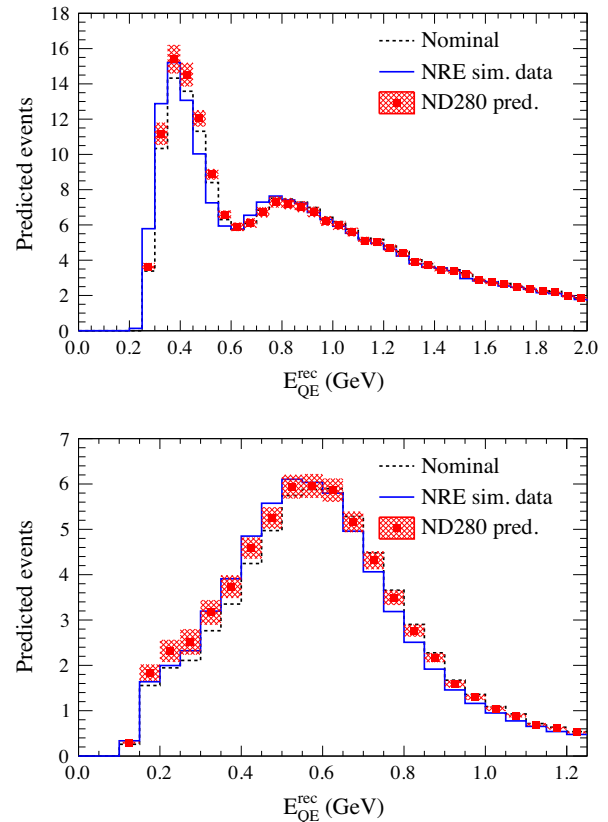
FIG. 51. Comparison of the nominal MC prediction (dashed black line), simulated data (solid blue line) and prediction from the near detector fit to the simulated data (shaded red). This is shown for the far detector muonlike (top) and electronlike (bottom) neutrino beam event samples.

quasi-elastic-like and Δ-like 2p2h interactions. The choice of interaction model therefore affects the neutrino energy distribution that experiments infer from their observed neutrino events. This in turn can affect their measurement of the neutrino oscillation parameters, as has been shown in Ref. [93].

In this analysis a comprehensive set of neutrino interaction models have been tested using simulated data studies to quantify their effect on the T2K oscillation result. The simulated data procedure is described in Ref. [26], while the model changes that are tested here are described in Sec. VII. Simulated data are created for both the near and far detectors, where a fit is performed in the same way as for the real data. The nominal flux and neutrino interaction models are fit to the simulated near detector data to produce a prediction for the far detector. The simulated far detector data are then compared to this prediction to extract oscillation parameter contours. The contours are then compared to those extracted from a fit to the Asimov dataset. If the T2K oscillation analysis is insensitive to the model change, or has the freedom to account for it correctly, then the simulated data contours and the Asimov contours should be very similar. The oscillation parameter values used for this study are shown in Table III. Other parameter sets with $\delta_{CP} = 0$ and $\sin^2 \theta_{23} = 0.45$

were also studied but showed no significant difference to the results presented here.

The likelihood distributions for each oscillation parameter are created for both the simulated data fit and the Asimov fit. Any change in the center of the 2σ confidence interval for each parameter is taken as a bias due to the change in the interaction model introduced to the simulated data. This is compared to the uncertainty on the parameter coming from the systematic uncertainties included in the analysis. If a simulated data bias is greater than 50% of the systematic uncertainty on a parameter, then an additional uncertainty is added to the analysis to account for this.

A. Simulated data study of the nucleon removal energy

As described in Sec. VII the T2K neutrino event generator, NEUT 5.3.3, implements a relativistic Fermi gas (RFG) nuclear model. To remove a nucleon from the nucleus requires energy to overcome the nuclear potential. The NRE can be measured by electron scattering experiments but is not known perfectly. Even the definition of this quantity is not simple [79]. In addition, the RFG is a very simple model of the nuclear structure. More advanced models, such as spectral function models, provide a much more detailed description of the nucleon energies within the nucleus, each of which has its own NRE.
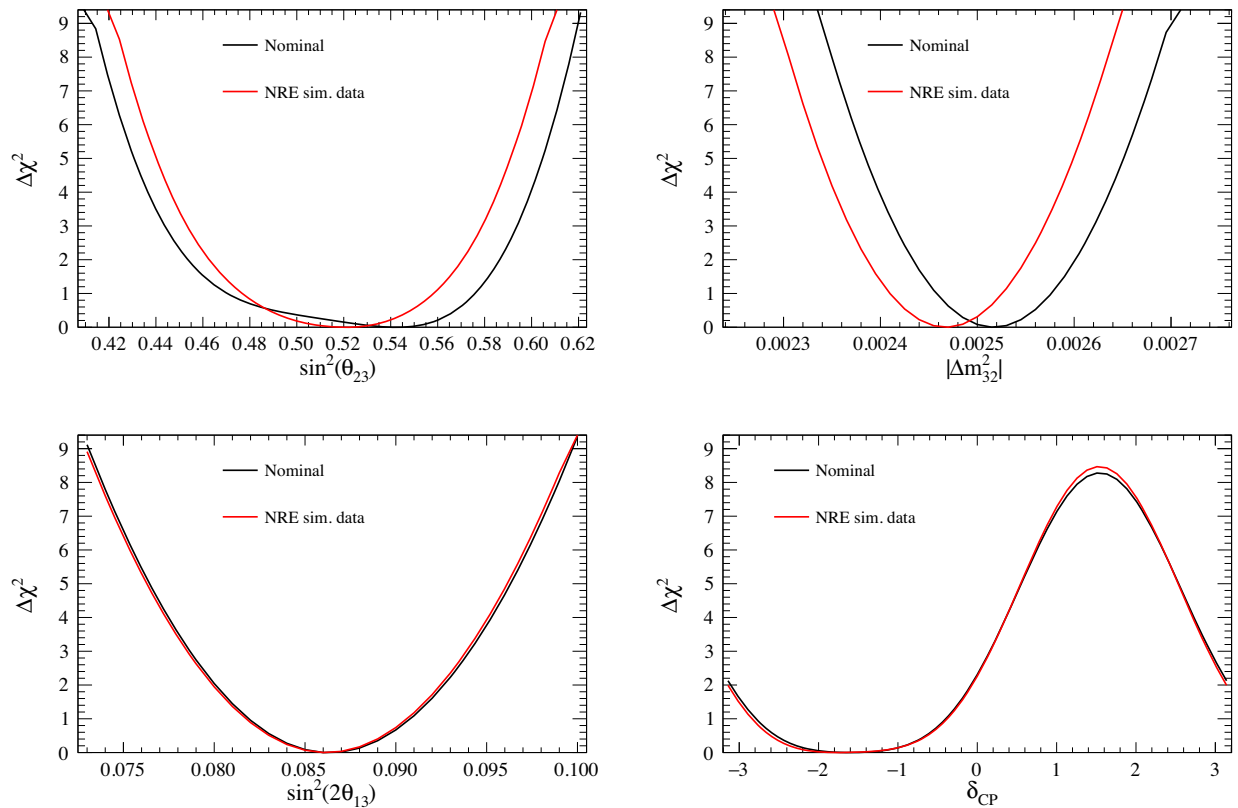

FIG. 52. Comparison of the nominal MC oscillation parameter $\Delta\chi^2$ function (black lines) to those from the simulated data fits (red lines).
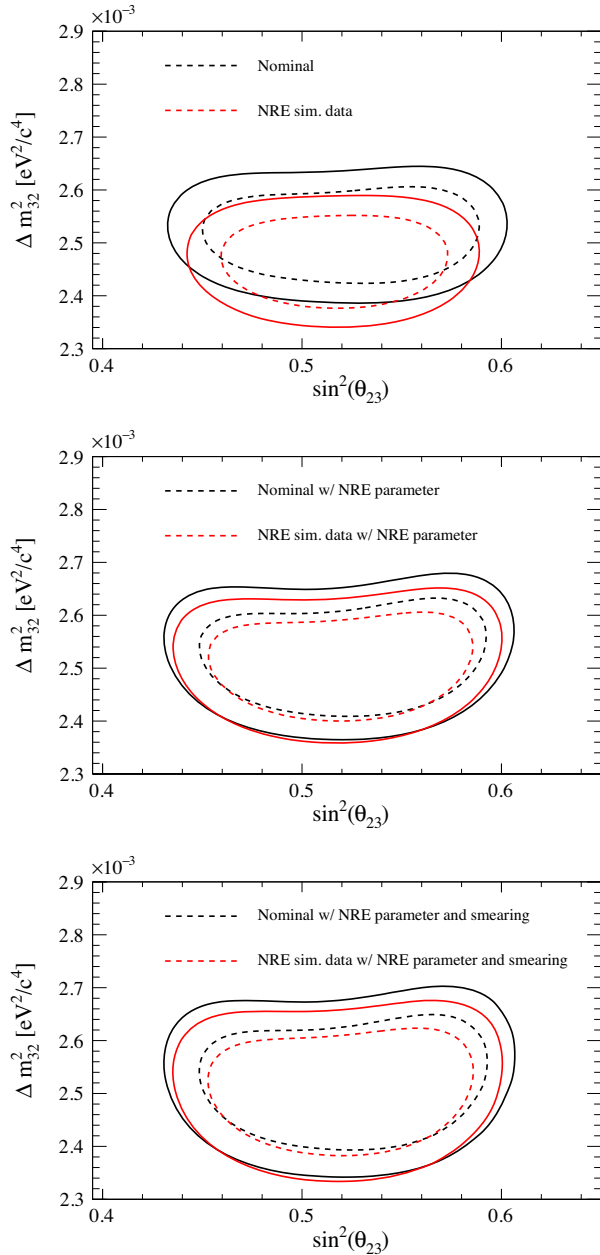
FIG. 53. Comparison of the $\sin^2\theta_{23} - \Delta m_{32}^2$ parameter contours using the nominal cross-section model (top), after the addition of the effective NRE parameter (middle) and after the additional smearing is applied to Δm_{32}^2 (bottom). In all cases the expected result, as described in the text, is shown in black and the simulated data result is in red. The solid lines represent the 90% confidence region and the dashed lines represent the 1σ confidence region.

The T2K oscillation analysis does not have a parameter to account for the NRE uncertainty directly in the fit. Therefore the effect of the NRE uncertainty on the oscillation parameter measurements is evaluated using a simulated data study. Neutrino events were generated with the RFG nucleon removal energy increased by 18 MeV from the nominal value (25 MeV for interactions on carbon

and 27 MeV for interactions on oxygen). Increasing the NRE results in a decrease in the energy available to the lepton produced in the neutrino interaction. As a consequence, the events simulated with a larger NRE produced leptons with lower momenta than the nominal MC. This momentum shift was calculated as a function of neutrino energy and lepton angle, shown in Fig. 48. The momentum shift in Fig. 48 was applied to the reconstructed momenta of the charged lepton for the full near and far detector MC. The MC was then scaled to match the POT exposure of the data, creating a simulated dataset. The ratio of the simulated data to the nominal MC is shown in Fig. 49 for the ND280 CC 0π sample, highlighting the shift in events from high-momentum bins to lower-momentum bins. A fit is then performed on this simulated data using the analysis framework described earlier in this paper. There are no parameters in the T2K cross-section model that change the lepton momentum directly, but combinations of parameters are able to mimic this effect. The result of fitting the simulated near detector data is presented in Fig. 50, which shows the best-fit values for the flux and cross-section model parameters. Figure 50 shows that the low-energy flux is increased, to increase the rate of low-momentum events. The 2p2h shape parameters control whether the 2p2h events in the MC are produced more by $\Delta(1232)$ resonances (values above 1.0) or other modes (values below 1.0). The $\Delta(1232)$ resonance produces leptons with a lower momentum than the other modes. This means that increasing the shape parameter increases the rate of 2p2h events in the low-momentum region.

The near detector fit result is used to predict the oscillated event distribution at the far detector. This is shown in Fig. 51, which compares the nominal MC to the simulated data and the near detector prediction. The near detector prediction is closer to the simulated SK data than the nominal MC simulation but does not match well below the oscillation dip. The nominal prediction is fit to the simulated SK data to extract the best-fit oscillation parameter values and their 2σ confidence intervals. These are compared to oscillation parameter values and confidence intervals produced by fitting to the Asimov dataset, as described in Sec. XII. This comparison is shown in Fig. 52.

Changing the NRE changes the shape of the simulated data, not the normalization. Since the constraint on $\sin^2\theta_{13}$ and δ_{CP} is largely driven by the normalization of the electronlike sample, it is unsurprising that these parameters are relatively unaffected. In addition $\sin^2\theta_{13}$ is strongly constrained by the reactor experiment measurements. However, both $\sin^2\theta_{23}$ and Δm_{32}^2 show a significant difference between the nominal and the simulated data results, with $\sin^2\theta_{23}$ shifting toward maximal disappearance and Δm_{32}^2 decreasing.

The changes in the oscillation parameter contours indicate that the T2K cross-section model parameterization cannot account for changes to the NRE. The near detector

TABLE XII. Oscillation parameter biases (as percentages of the total and systematic uncertainties) observed in the simulated data studies including the additional uncertainties on the NRE.

Simulated dataset	Relative to	$\sin^2\theta_{23}$	Δm^2_{32}	δ_{CP}
Martini 2p2h	Total	9.0%	16%	0.1%
	Syst	20%	22%	0.3%
Data-driven CC 0π	Total	15%	14%	4.0%
$E_\nu - Q^2$ dependence	Syst	34%	20%	17%
BSF 1p1h	Total	1.5%	22%	0.1%
	Syst	3.4%	31%	0.3%
Nieves LFG 1p1h	Total	4.0%	25%	7.0%
	Syst	8.3%	35%	20%
Nucleon	Total	5.0%	33%	0.1%
removal energy	Syst	10%	46%	0.6%
Coulomb correction	Total	1.0%	0.1%	0.1%
	Syst	2.3%	0.1%	0.3%
Kabirnezhad	Total	8.0%	34%	0.0%
single pion	Syst	20%	50%	1.0%

fit misattributes the NRE change to the flux and 2p2h model parameters. The near detector postfit prediction has a different neutrino energy distribution to the nominal MC and so produces different far detector event distributions when the neutrino oscillation probability is applied. As a result the oscillation parameters extracted from the simulated data fit no longer match those from the nominal MC analysis.

To account for this in the T2K analysis an additional uncertainty is introduced. A spline is created for each bin of the five far detector sample histograms. The value of the spline is the ratio between the simulated far detector data and the far detector prediction calculated using the near detector simulated data fit result. Spline knots are created using NREs of 18, 27 (nominal) and 45 MeV at both near and far detector. The splines are 100% correlated across all

sample bins and produce a multiplicative weight to scale the far detector prediction in each bin. The spline takes into account the change in the far detector prediction due to both the changing NRE and the misfitting at the near detector. This means that it does not provide a measurement of the true NRE but an "effective NRE" taking into account both of these effects. A prior constraint is placed on this effective NRE parameter, setting the central value to 27 MeV, with an uncertainty of ±18 MeV.

The result of including this parameter in the simulated data study is shown in Fig. 53. The new parameter increases the size of the oscillation contours while shifting the simulated data result to be in much better agreement with the expectation. There is some residual difference between the contours, particularly in the Δm^2_{32} best-fit point. This difference is included as an additional uncertainty in the analysis by smearing the Δm^2_{32} likelihood surface. The far detector oscillation fit likelihood as a function of Δm^2_{32} has a Gaussian distribution. This distribution is convolved with a Gaussian of unit area, centered at 0, with a width given by the shift in the Δm^2_{32} best-fit point between the simulated data fit and the expectation. This is equivalent to adding this shift as an uncertainty on Δm^2_{32} in quadrature with the existing uncertainties in the analysis. The result of this smearing is shown in Fig. 53.

B. Summary of simulated data studies

Table XII shows the final bias table for simulated datasets studied in the analysis, after the addition of the NRE uncertainty parameter. For the data-driven $E_\nu - Q^2$ category the largest effect from the three simulated data studies is shown. In all cases the observed bias on $\sin^2\theta_{23}$ and δ_{CP} was insignificant compared to existing systematic uncertainty on the parameter and so no additional uncertainty was introduced. Non-negligible bias was observed

TABLE XIII. Fractional uncertainty (%) on event rate by error source and sample, calculated with expected event rates generated according to the nominal oscillation parameter values from Table III. The final column is the fractional uncertainty (%) on the ratio of FHC/RHC events in the one-ring e sample. The final row, "all systematics," does not include the effects of any oscillation parameters.

Error source	One-ring μ		One-ring e			
	FHC	RHC	FHC	RHC	FHC 1 d.e.	FHC/RHC
Flux and (ND unconstrained)	14.3	11.8	15.1	12.2	12.0	1.2
cross section (ND constrained)	3.3	2.9	3.2	3.1	4.1	2.7
SK detector	2.4	2.0	2.8	3.8	13.2	1.5
SK FSI + SI + PN	2.2	2.0	3.0	2.3	11.4	1.6
Nucleon removal energy	2.4	1.7	7.1	3.7	3.0	3.6
$\sigma(\nu_e)/\sigma(\bar\nu_e)$	0.0	0.0	2.6	1.5	2.6	3.0
NC1γ	0.0	0.0	1.1	2.6	0.3	1.5
NC other	0.3	0.3	0.2	0.3	1.0	0.2
$\sin^2\theta_{23}$ and Δm^2_{21}	0.0	0.0	0.5	0.3	0.5	2.0
$\sin^2\theta_{13}$ PDG2018	0.0	0.0	2.6	2.4	2.6	1.1
All systematics	5.1	4.5	8.8	7.1	18.4	6.0

for Δm_{32}^2. The quadrature sum of the observed biases, 4.1×10^{-5} eV^2 c^{-4}, was added as an additional uncertainty on Δm_{32}^2 using the method described above. The effect of the systematic parameters on the predicted event rates on each SK event sample, including the additional NRE uncertainty, is shown in Table XIII. The effect of the prior uncertainties on the (typically marginalized) oscillation parameters $\sin^2\theta_{12}$, Δm_{21}^2 and $\sin^2\theta_{13}$ is also shown.

XIV. OSCILLATION ANALYSIS RESULTS

Following the recommendations in [94], we produce results using different statistical approaches, both frequentist and Bayesian (with analysis of sensitivity to prior) and test the frequentist properties of our Bayesian methods when possible. Using the three far detector analyses described previously, point and interval estimations are made for the parameters $\sin^2\theta_{23}$, δ_{CP} and Δm_{32}^2 (normal ordering) or Δm_{13}^2 (inverted ordering). Two types of intervals are produced: confidence intervals (with approximate coverage based on the constant $\Delta\chi^2$ method in most cases and with exact coverage using the Feldman-Cousins unified approach for δ_{CP}) and credible intervals. The mass ordering was studied using mainly Bayesian hypothesis testing, with additional frequentist checks.

A. Measurements of the parameters of the three-flavor oscillation model

1. $\Delta\chi^2$ and frequentist results

Intervals based on the constant $\Delta\chi^2$ method are produced for the different parameters using analyses A and C (analysis B can produce similar intervals for comparison purpose, although its main results are the credible intervals described in Sec. XIV A 2). As their results are in good agreement, only the results obtained with analysis A are shown in this section, unless otherwise indicated. The best-fit values and 1σ confidence intervals obtained for the different parameters in both mass ordering scenarios are summarized in Table XIV and in Table XV with and without using the results of reactor experiments to constrain $\sin^2\theta_{13}$, respectively. The global best fit was found to be for the NO, and the data show a preference for the upper octant. These preferences will be quantified in Sec. XIV C. The $\Delta\chi^2 = -2\ln(L/L_{\max})$ functions obtained for δ_{CP} with and without using the results of reactor experiments to constrain $\sin^2\theta_{13}$ are displayed in Fig. 54. The favored and disfavored values of δ_{CP} are similar between the two cases, but the constraint on δ_{CP} becomes stronger when the reactor experiments results are used. The obtained 90% confidence regions for $(\sin^2\theta_{23}, \delta_{CP})$ are displayed in Fig. 55. The largest parts of the confidence regions are located in the upper octant, especially when the constraint from reactor experiments is used, but the results are still compatible with maximal mixing. For the atmospheric parameters, the

TABLE XIV. The measured oscillation parameter best-fit and the $\pm 1\sigma$ intervals, shown for the T2K-only (without reactor constraint) fit and for normal and inverted hierarchies with respect to the hierarchy best fit. The $\pm 1\sigma$ interval corresponds to the values for which $\Delta\chi^2 \leq 1$.

	Best-fit and 1σ interval	
Parameter	NO	IO
δ_{CP}	$-2.14^{+0.90}_{-0.69}$	$-1.26^{+0.61}_{-0.69}$
$\sin^2\theta_{13}/10^{-3}$	$26.8^{+5.5}_{-4.3}$	$30.0^{+5.9}_{-5.0}$
$\sin^2\theta_{23}$	$0.512^{+0.045}_{-0.042}$	$0.500^{+0.050}_{-0.036}$
$\Delta m_{32}^2/10^{-3}$ eV^2 c^{-4}	$2.46^{+0.07}_{-0.07}$	
$|\Delta m_{13}^2|/10^{-3}$ eV^2 c^{-4}		$2.43^{+0.07}_{-0.08}$

TABLE XV. The measured oscillation parameter best-fit and the $\pm 1\sigma$ intervals, shown for the T2K+ reactor fit and for normal and inverted hierarchies with respect to the hierarchy best fit. The $\pm 1\sigma$ interval corresponds to the values for which $\Delta\chi^2 \leq 1$.

	Best-fit and 1σ interval	
Parameter	NO	IO
δ_{CP}	$-1.89^{+0.70}_{-0.58}$	$-1.38^{+0.48}_{-0.55}$
$\sin^2\theta_{23}$	$0.532^{+0.030}_{-0.037}$	$0.532^{+0.029}_{-0.035}$
$\Delta m_{32}^2/10^{-3}$ eV^2 c^{-4}	$2.45^{+0.07}_{-0.07}$	
$|\Delta m_{13}^2|/10^{-3}$ eV^2 c^{-4}		$2.43^{+0.07}_{-0.07}$

obtained normal ordering 90% confidence region for $(\sin^2\theta_{23}, \Delta m_{32}^2)$ is shown in Fig. 56 together with the measurements from other neutrino oscillation experiments. Good agreement is seen between all of the experiments.

Section XII demonstrated that analyses A, B and C have similar sensitivities (Fig. 47); their data fit results are now compared for δ_{CP} vs $\sin^2\theta_{13}$ in Fig. 57 and Δm^2 vs $\sin^2\theta_{23}$
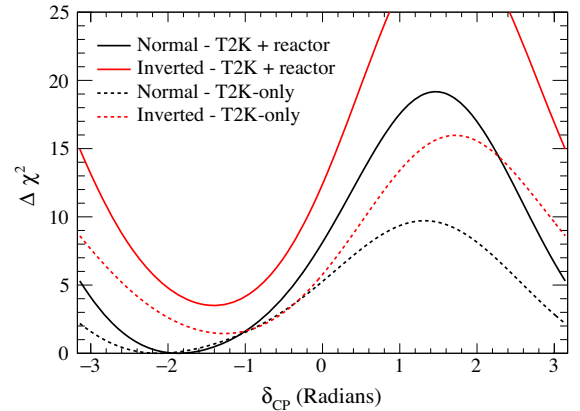

FIG. 54. The observed $\Delta\chi^2$ function of δ_{CP}, with and without the reactor constraint. The $\Delta\chi^2$ is computed with respect to the best fit over the two mass orderings, and separate best-fit points are used for the T2K-only and the T2K + reactor cases.
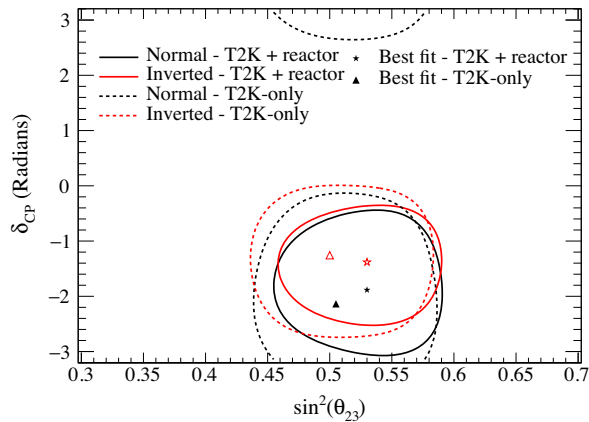
FIG. 55. The observed constant $\Delta\chi^2$ 90% confidence regions of $\sin^2\theta_{23}$ and δ_{CP} with normal and inverted mass orderings and with and without the reactor constraint. Normal and inverted mass ordering contours are independent. $\Delta\chi^2$ values are calculated independently for the functions with and without the reactor constraint.

in Fig. 58. The largest differences are observed for analysis C, in particular for the atmospheric parameters. Those differences reduce to a negligible level if analysis C is repeated with the appearance samples binned in the same E_{rec} variable as analyses A and B, indicating this is primarily an effect coming from the choice of kinematic variables.

The intervals listed in Tables XIV and XV were constructed using constant $\Delta\chi^2$ critical values. This treatment gives proper coverage when Wilks' theorem holds [95] but can result in poor coverage when this is not the case. In particular, it is not expected to give proper coverage for δ_{CP}, due to the cyclic nature of the parameter and the presence of physical boundaries at $\pm\pi/2$. In these cases intervals with exact coverage can be formed directly from the likelihood
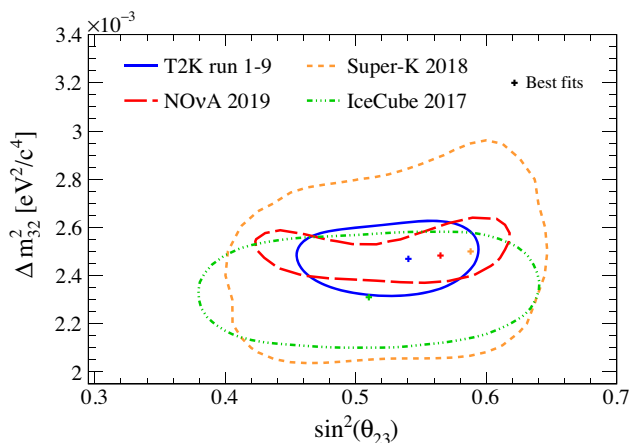

FIG. 56. The observed T2K constant $\Delta\chi^2$ 90% confidence regions (from analysis C) of $\sin^2\theta_{23}$ and Δm^2_{32} with the reactor constraint and compared to the results from Super-K [4], NOvA [13] and IceCube [14].
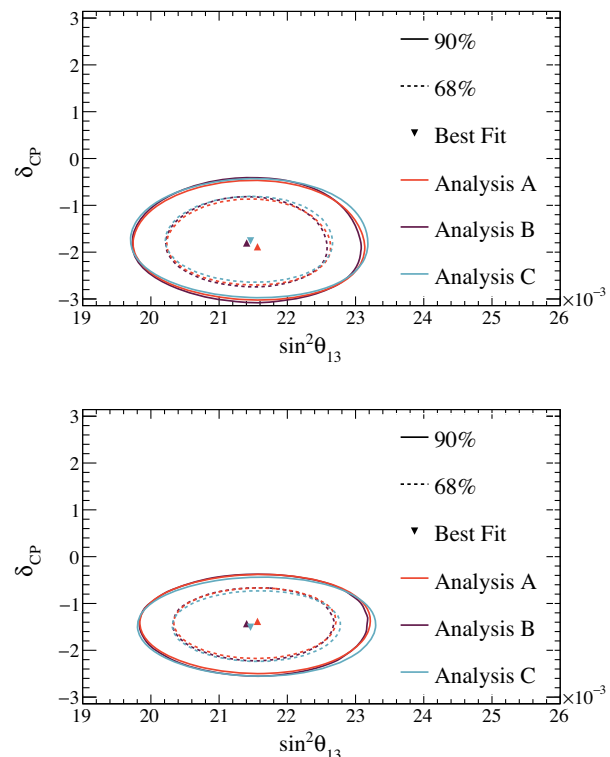

FIG. 57. Comparison between the observed 1σ (dashed lines) and 90% (solid lines) C.L. regions for δ_{CP} vs $\sin^2\theta_{13}$ produced using analyses A, B and C, assuming normal mass ordering (top) or inverted mass ordering (bottom). The reactor constraint is applied.

ratio, by computing the appropriate $\Delta\chi^2$ critical value for each value of the parameters considered, as proposed by Feldman and Cousins [96]. This method is very CPU intensive, so it is only used for the one-dimensional δ_{CP} interval.

For analyses A and C, critical values of $\Delta\chi^2$ (values of the parameter of interest for which the $\Delta\chi^2$ is lower than the critical value for a given confidence level are included in the confidence interval for this level) were calculated for multiple values of δ_{CP} as follows. Pseudoexperiments were generated at nine evenly spaced grid points of δ_{CP} in each mass ordering and at two additional points near the intersection of the 3σ critical value and $\Delta\chi^2$ curves. The systematic and other oscillation parameters are randomly varied to generate the pseudoexperiments, with different procedures for the different parameters. The systematic parameters, as well as $\sin^2 2\theta_{12}$, Δm^2_{21} and $\sin^2 2\theta_{13}$ are drawn from the prior probabilities described in Sec. XII, with $\sin^2 2\theta_{13}$ constrained by reactor data. The two oscillation parameters which do not have a prior constraint in the analysis, $\sin^2\theta_{23}$ and Δm^2_{32}, are drawn from the 2D likelihood resulting from the fit of an Asimov dataset generated at the best-fit point obtained in the T2K + reactor data fit. Each set of parameter values is used to generate predicted kinematic distributions for each sample, which
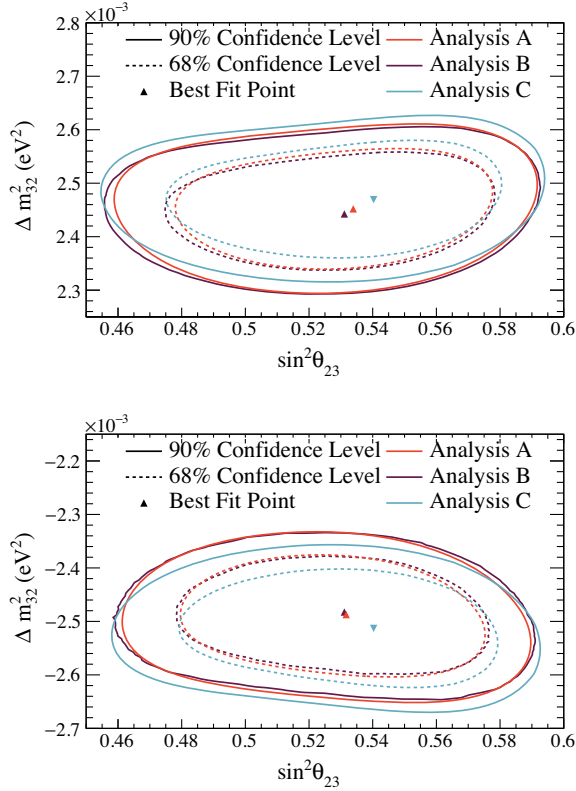
FIG. 58. Comparison between the observed 1σ (dashed lines) and 90% (solid lines) C.L. regions for Δm^2 vs $\sin^2\theta_{23}$ produced using analyses A, B and C, assuming normal mass ordering (top) or inverted mass ordering (bottom). The reactor constraint is applied.
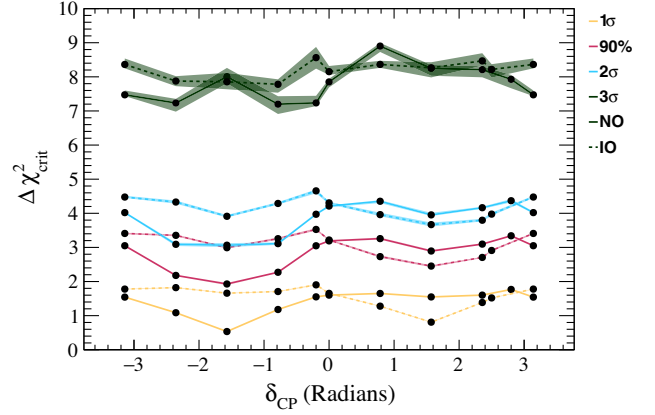

FIG. 59. Feldman-Cousins critical values for δ_{CP}. The shaded bands represent the $\pm 1\sigma$ Monte Carlo statistical uncertainty on the calculated critical values.

are then sampled assuming a Poisson probability in each reconstructed variable(s) bin to obtain a pseudoexperiment. The pseudoexperiments are fit in the same manner as the real data, and the $\Delta\chi^2$ between the true and best fit δ_{CP} and mass ordering is recorded. The Nth percentile of this distribution then forms the N% critical value for this combination of δ_{CP} and mass ordering.

The obtained critical values for δ_{CP} are displayed in Fig. 59. For all confidence levels, significant deviations from the values expected for a parabolic log-likelihood function (as assumed by the constant $\Delta\chi^2$ method) are observed, demonstrating the necessity of the method. The critical values obtained by analysis A were compared to those of analysis C and were generally found to be in good agreement, with minor differences mostly explained by the different kinematic variables used by the two analyses for appearance samples.

In order to better understand the structure of the critical values, the effects of the δ_{CP}-mass ordering and $\cos(\delta_{CP})$ degeneracies, the physical boundaries around $\delta_{CP} = \pm\pi/2$ and of the beam exposure were studied and found to have a significant effect, due to the following factors.

1. The amount by which the $\nu_\mu \to \nu_e$ and $\bar\nu_\mu \to \bar\nu_e$ oscillation probabilities can be changed by varying

δ_{CP} is limited. This creates an effect similar to physical boundaries at the values of δ_{CP} corresponding to the maximum and minimum number of events in the appearance samples. The critical values in those regions are therefore lower than those expected under the assumption of a parabolic log-likelihood.

2. Mass ordering is a discrete parameter and works as an additional degree of freedom, raising critical values, though without sufficient freedom to make the critical values behave as if the problem had two degrees of freedom.

3. The effect of physical boundaries is more visible for the 3σ critical values, since the critical values at this confidence level are usually determined by the pseudoexperiments corresponding to more extreme statistical fluctuations.

4. The critical values increase with statistics, and the values obtained have increased relative to their values in previous T2K analyses [25,26,29].

To understand this last point, critical values for the $1, 2, 3\sigma$ and 90% confidence levels were computed assuming different exposures and were found to nonlinearly increase with exposure in all cases. The leading cause was found to be the approximate degeneracy between δ_{CP} and $\pi - \delta_{CP}$ (T2K observables are mainly sensitive to $\sin\delta_{CP}$, with $\cos\delta_{CP}$ having a much smaller effect), which acts as an additional pseudo-degree-of-freedom that is negligible at low exposures but becomes more important as more data are taken. The above physical boundary effects also contribute at all exposures; however, for true δ_{CP} values away from the boundaries, the boundary effect decreases with increased exposure.

The finite number of pseudoexperiments used to compute the critical values introduces a Monte Carlo statistical uncertainty, and the number of pseudoexperiments is chosen to make the size of this uncertainty negligible for the 2σ critical values; however, the number required to reduce the 3σ critical value statistical uncertainties to this

TABLE XVI. Feldman-Cousins confidence intervals for δ_{CP} with the reactor constraint applied in both normal and inverted mass orderings.

Confidence level	Interval (radians)	
	Normal ordering	Inverted ordering
1σ	$[-2.51, -1.26]$	$\cdots$
90%	$[-2.80, -0.84]$	$\cdots$
2σ	$[-2.97, -0.63]$	$[-1.80, -0.98]$
3σ	$[-\pi, -0.03] \cup [2.87, \pi]$	$[-2.54, -0.32]$

level would be computationally intractable, so additional care is taken to ensure the validity of the 3σ confidence intervals. The confidence intervals are calculated for the different combinations of the nominal and $\pm 1\sigma$ binomial error values of the critical $\Delta\chi^2$, and the largest interval is kept. A second source of error comes from the interpolation of the critical $\Delta\chi^2$ between the points where it was computed. To minimize this effect, critical values have been calculated at two additional δ_{CP} values, each close to the boundaries of the 3σ confidence interval.

The resulting confidence intervals obtained for δ_{CP} are listed in Table XVI and the obtained 3σ intervals are displayed in Fig. 60, superimposed on to the observed $\Delta\chi^2$ function.

In both mass orderings, the two CP-conserving values of 0 and π are outside of the 2σ confidence intervals. Our data therefore exclude the conservation of CP symmetry in neutrino oscillation at the 2σ level. For the inverted mass ordering, both CP-conserving values are outside of the 3σ confidence intervals. For the normal mass ordering, $\delta_{CP} = 0$ is just outside of the 3σ confidence interval, while $\delta_{CP} = \pm\pi$ is inside. The robustness of the exclusion of $\delta_{CP} = 0$ given the model uncertainties is studied in more detail in Sec. XIV B 3, where it was found that the boundary of the 3σ interval is so close to $\delta_{CP} = 0$ that this point can move in and out of the 3σ interval due to

changes in the model. On the contrary, the exclusion of the CP-conserving values at the 2σ level was found to be robust.

2. Bayesian results

Bayesian results for the oscillation parameters were produced using analysis B. The posterior probability of δ_{CP} marginalized over both mass orderings obtained in a fit of T2K data with $\sin^2\theta_{13}$ constrained by the reactor experiments is shown in Fig. 61 with two different prior probabilities for the parameter of interest. The CP-conserving values of δ_{CP} are found to be outside of the 2σ credible interval in both cases, with the 1σ range still covering the maximal CP violation value of $\delta_{CP} = -\pi/2$. Most of the results in this section use a prior probability uniform in δ_{CP}; the alternative prior probability was tested to see the effect of the choice of prior. The alternative was chosen to be uniform in $\sin(\delta_{CP})$ as this is both the variable involving δ_{CP} to which our observables are most sensitive to and the one relevant for CP violation. The detailed comparison between the intervals obtained with the two prior probabilities can be found in Table XVII, where it can be seen that this choice affects the size of the intervals obtained but does not change the main conclusions of the analysis.

The proposal function for choosing the random steps used in analysis B's MCMC can jump between mass orderings, with a 50% chance at each step to propose a point with the opposite sign of Δm_{32}^2. It therefore produces a single posterior probability for Δm_{32}^2, covering both positive and negative values. The posterior probability for Δm_{32}^2 obtained in a fit of T2K data with constraint from the reactor experiments is shown in Fig. 62 and the credible regions for the atmospheric parameters in Fig. 63. A clear preference toward the normal ordering, which contains 89% of the total posterior density, can be seen,
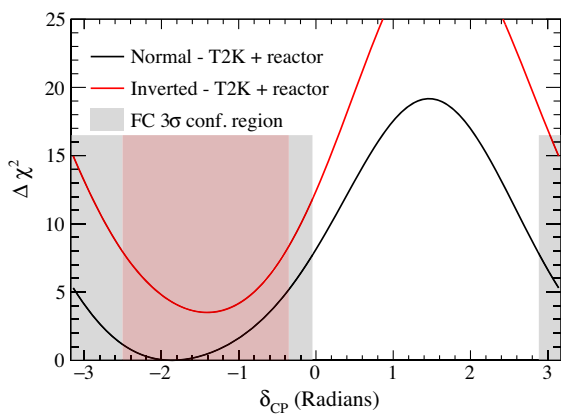

FIG. 60. The observed 3σ Feldman-Cousins (FC) confidence intervals for δ_{CP}. The $\Delta\chi^2$ is computed with respect to the best fit over the two mass orderings.
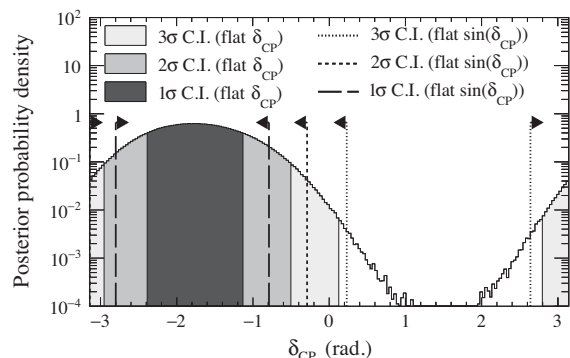

FIG. 61. Posterior probability density for δ_{CP} and credible intervals obtained using a prior uniform in δ_{CP} compared against credible intervals obtained with a prior uniform in $\sin(\delta_{CP})$. The CP-conserving values are outside of the 2σ credible intervals in both cases.

TABLE XVII. Credible intervals for δ_{CP} with uniform prior probabilities on either δ_{CP} or on $\sin(\delta_{CP})$.

Prior uniform in	δ_{CP}	$\sin \delta_{CP}$
68.26% (1σ)	$[-2.39, -1.13]$	$[-2.80, -0.79]$
95.45% (2σ)	$[-2.76, -0.72]$	$[-3.14, -0.29]$
99.73% (3σ)	$[-\pi, 0.13] \cup [2.80, \pi]$	$[-\pi, 0.23] \cup [2.65, \pi]$

with in particular the lack of an observed 1σ credible region in the $\Delta m_{32}^2 < 0$ part of the figures.

Figure 64 shows the posterior density distributions and credible regions for combinations of all the oscillation parameters, as well as their individual posterior density and credible intervals. Correlations between the estimates of the different parameters can be seen, as well as the effect of marginalizing over one of them to obtain the one-dimensional posterior distribution of another oscillation parameter. As the $\sin^2 \theta_{23}$ posterior distribution has a
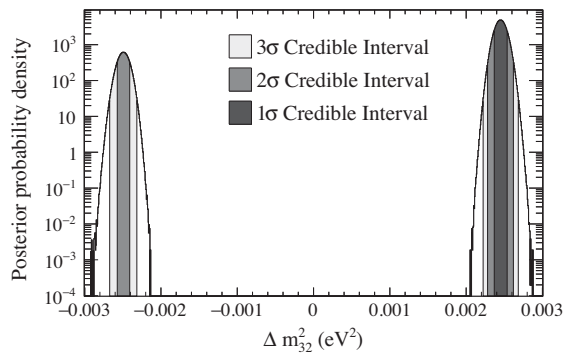

FIG. 62. Posterior probability distribution for Δm_{32}^2 covering both mass orderings with 1, 2 and 3σ Bayesian credible intervals marked. The normal ordering contains 89% of the posterior probability.
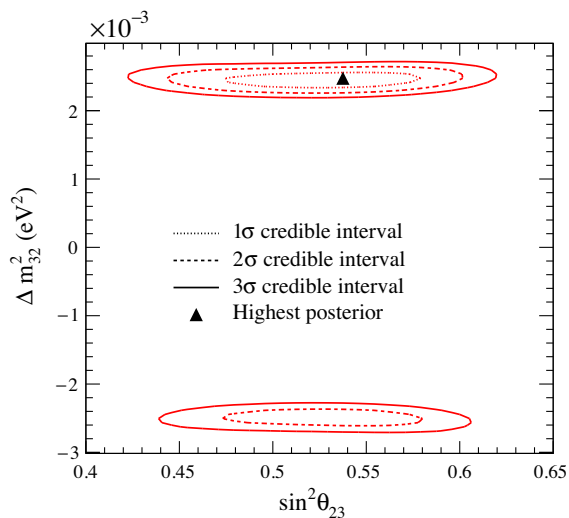

FIG. 63. Bayesian credible regions for Δm_{32}^2 against $\sin^2 \theta_{23}$ covering both mass orderings.

largely non-Gaussian shape, the marginal likelihood (for the other parameters) used in the analyses described in this paper is expected to have some differences with the profile likelihood commonly used in the neutrino community. Correlations are mostly seen between the atmospheric parameters, $\sin^2 \theta_{23}$ and Δm_{32}^2, and between Δm_{32}^2 and δ_{CP}. The correlation between the atmospheric parameters is clear from Eq. (10). For Δm_{32}^2 and $\cos(\delta_{CP})$, both parameters shift the energy of the peak (dip) in the electron (muon) neutrino oscillation probability. Combined with the energy profile of the T2K beam this energy shift produces a change in event rate for the electronlike samples and a smaller change in the muonlike samples. This change in both rate and shape introduces a correlation between the two parameters.

B. Additional checks on the validity of δ_{CP} results

The main result of the analysis described in this paper is the measurement of δ_{CP}, with in particular the exclusion of CP conservation in neutrino oscillations at the 2σ level, and the fact that some of the possible values of δ_{CP} are outside of the 3σ confidence intervals. The observed constraint on δ_{CP} is considerably stronger than that expected from sensitivity studies, as can be seen by comparing Figs. 45 and 54. A number of additional studies were done to check the likelihood and robustness of this result.

1. Probability of the δ_{CP} results

The probability of getting a constraint this strong or stronger considering systematic and statistical uncertainties was evaluated assuming normal ordering and a true value of δ_{CP} of $-\pi/2$. This case was chosen as it gave the best agreement between data and predictions in Table XI. 10 000 pseudoexperiments were generated following the same method as used in the determination of the Feldman-Cousins critical values. The pseudoexperiments were then fitted for δ_{CP}, and the resulting $\Delta\chi^2$ distributions were compared to the $\Delta\chi^2$ functions obtained for the data. Figure 65 shows the observed $\Delta\chi^2 = -2\Delta \ln L$ overlaid on the one-sided 1σ and 2σ bands. Those bands are built from the ensemble of $\Delta\chi^2$ values from the different pseudoexperiments at each value of δ_{CP} and cover the interval between zero and the 68.27% (1σ band) or 95.45% (2σ band) quantile. The observed $\Delta\chi^2$ function lies within the one-sided 2σ band for the fits done in the inverted mass ordering scenario, while falling just outside the 2σ band in the normal mass ordering scenario in the region around $\delta_{CP} = 0$. An alternative way to consider the results is to check what fraction of pseudoexperiments exclude one or both of the CP-conserving values at a specified confidence level. As shown in Table XVIII, for the normal mass ordering, $\delta_{CP} = 0$ and $\delta_{CP} = \pi$ are separately excluded at 90% confidence level in around 46%–48% of pseudoexperiments and excluded at 2σ confidence level in around 32%–34%. Both CP-conserving values are excluded at the 2σ confidence level in about 25% of the
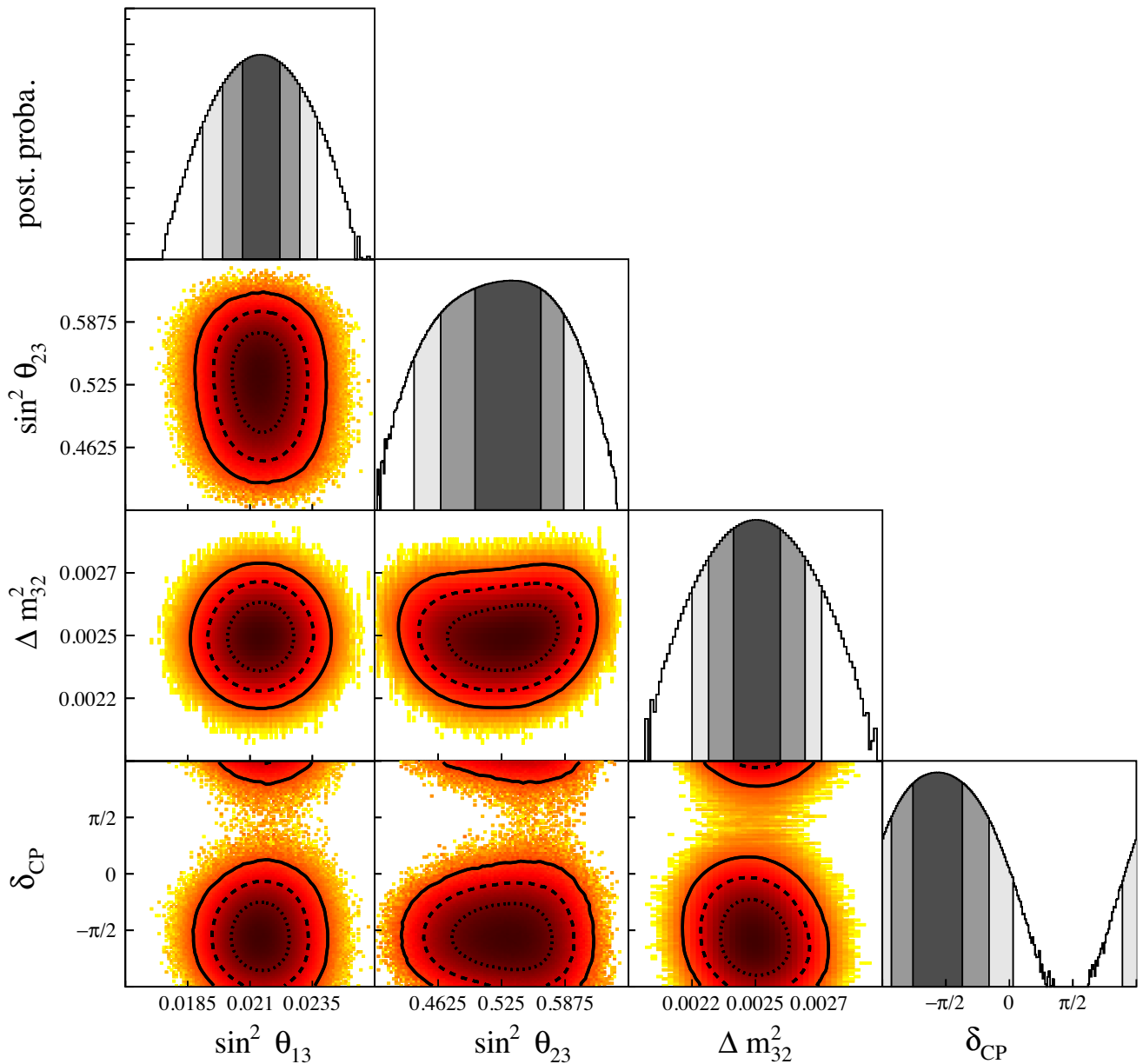
FIG. 64. Posteriors probabilities together with 1, 2 and 3σ credible intervals for all the oscillation parameters of interest and their combinations in the normal mass ordering. A logarithmic scale is used for the axis corresponding to the posterior probability density, and darker colors correspond to larger probabilities.

pseudoexperiments. Overall, a 2σ exclusion of CP conservation is not unlikely according to the model used in this analysis.

2. Contributions from individual samples

The discrepancies between predictions and observations seen in some of the samples used in the analysis were studied in more detail. In Table XI, the observed numbers of events in the far detector samples are generally in good agreement with the predicted number of events for $\delta_{CP} = -\pi/2$. Two samples show a difference with the nominal prediction: for the FHC 1R μ-like sample a small

deficit of events can be seen in the data, corresponding to 1.33 times the statistical and systematic uncertainty on the predicted number of events. For the FHC ν_e CC $1\pi^+$-like sample, the discrepancy is larger with the observed number of events being twice the predicted number. To understand the impact on the δ_{CP} result of the differences between predictions and observations from each sample, hybrid MC and data fits were performed: the data were replaced by the MC predictions for each sample in turn, as shown in Fig. 66.

It can be seen that in all cases including data, the constraint on δ_{CP} is stronger than that obtained from the MC predictions alone. For the CP-conserving value $\delta_{CP} = -\pi$, the most significant changes happen when data

FIG. 66. Change of the results of the fit of the T2K run 1–9 data for δ_{CP} when observations get replaced by predictions for the different samples. The gray dashed lines correspond to the critical values for 2σ and 3σ obtained using the Feldman-Cousins method.
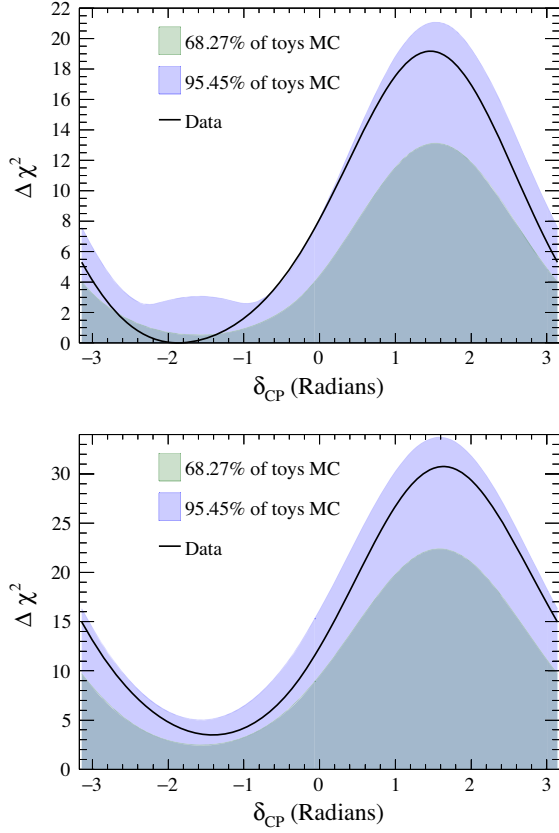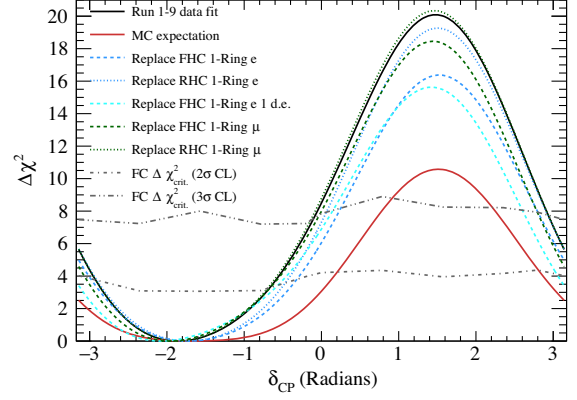
FIG. 65. The observed $\Delta\chi^2 = -2\Delta \ln L$ functions compared with one-sided distributions of $\Delta\chi^2$ values corresponding to 68.27% and 95.45% of 10 000 pseudoexperiments generated for $\delta_{CP} = -\pi/2$ and normal mass ordering. Results for the normal mass ordering (top) and inverted mass ordering (bottom) are shown.

are replaced by MC predictions for the two samples which showed differences between observations and predictions. In particular, in the case of the FHC ν_e CC $1\pi^+$-like sample, the $\Delta\chi^2$ at $\delta_{CP} = -\pi$ becomes lower than the 2σ critical value obtained with the Feldman-Cousins method, meaning that the conservation of CP symmetry is no longer excluded with 2σ significance in this case.

TABLE XVIII. Fraction of the pseudo datasets for which the $\Delta\chi^2$ at $\delta_{CP} = 0, \pi$ and both CP-conserving values are above the critical values for 90% and 2σ C.L. Pseudo datasets were generated assuming true normal ordering and a value of $\delta_{CP} = -\pi/2$.

δ_{CP}	Mass ordering scenario	90%	2σ
0		0.48	0.34
π	Normal	0.46	0.32
0 and π		0.37	0.25
0		0.83	0.74
π	Inverted	0.82	0.73
0 and π		0.78	0.69

Additional studies were therefore performed to estimate the likelihood of the observations for the FHC ν_e CC $1\pi^+$-like sample. First, the probability to obtain 15 events or more in this sample when taking into account statistical and systematic uncertainties was evaluated. This probability was found to be 2.49% for true values of the oscillation parameters corresponding to the T2K-only best fit and 1.34% for the T2K with reactor constraint best-fit point. As there are five samples in the analysis, a trial factor should be taken into account. The probability to have such an excess in at least one of the five samples (meaning an excess of events in the sample corresponding to a p value smaller or equal to the p value for the FHC ν_e CC $1\pi^+$-like sample) was found to be 11.3% for the T2K-only best fit and 5.8% for the T2K + reactor best-fit point.

p values were also calculated when taking into account not only the number of events, but also the distribution of the kinematic variables of the observed events, and were found to be equal to the rate-only p value. As can be seen in Fig. 67, the kinematic distributions of the data events are in good agreement with the prediction, and so taking into account the shape information does not increase the disagreement between data and prediction.

3. Additional interaction model checks

Section XIII described checks for possible biases coming from the choice of interaction model. Those checks were based on comparisons of MC sensitivities obtained with different interaction models. However, some model uncertainties affect primarily the FHC ν_e CC $1\pi^+$ sample, which brings only a small contribution to the δ_{CP} sensitivity compared to the single-ring e-like samples and therefore do not have a significant effect on the simulated data studies. They can nevertheless have a sizable effect on the predictions for the FHC ν_e CC $1\pi^+$ sample, and given the large excess of data events in this sample and its effect on the δ_{CP}
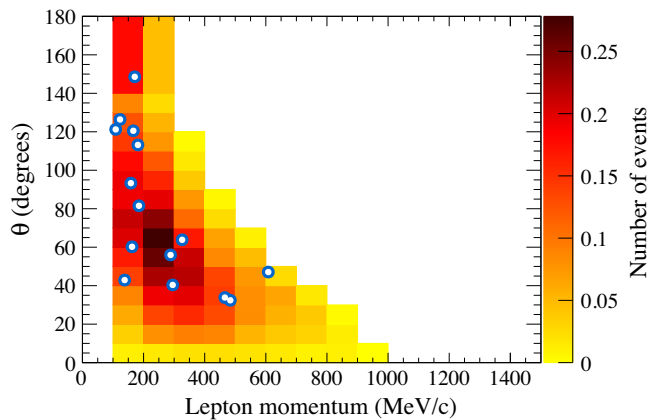
FIG. 67. Lepton momentum and angle with respect to the beam direction for the 15 events observed in the FHC ν_e CC $1\pi^+$-like sample, overlaid with the MC predictions.

measurement, the possible impact of additional model uncertainties were considered. The additional studies were done in the context of the data fit and not the MC-based sensitivity studies.

The first source of uncertainty studied is the data-MC discrepancy in the pion spectrum for the near detector CC 1π sample. Even after tuning using the results of the near detector data fit, the predicted pion momentum spectrum did not reproduce the data for the *near detector* FHC CC $1\pi^+$ sample: the data events had lower pion momentum than predicted by the model. This discrepancy could have an impact on the prediction for the far detector FHC CC $1\pi^+$ sample: charged pions can appear at SK either as rings if they have high enough momentum (>156 MeV/c) or as Michel electrons from the decay of the pion. If the pions produced in CC $1\pi^+$ interactions have lower momentum than our model predicts, a larger fraction of the CC $1\pi^+$ events at SK will enter the FHC CC $1\pi^+$ sample than our MC predicts. Studies showed that a discrepancy of the size observed at the near detector could lead to a 10% increase in the number of events in the far detector CC $1\pi^+$ sample. To evaluate the possible impact of this uncertainty on our δ_{CP} results, the fit of the run 1–9 data was redone with two different modifications:

(i) adding a 10% normalization uncertainty on the number of events in the far detector ν_e CC $1\pi^+$ sample and

(ii) increasing the number of events in the SK ν_e CC $1\pi^+$ sample by 10% in the MC predictions used to fit the data.

The second model uncertainty concerns the number of hadrons produced in deep inelastic interactions in the low invariant mass region $W < 2$ GeV. Those interactions correspond to a specific interaction mode in the NEUT generator used to produce the MC and have at least two pions produced at the interaction level. They can nevertheless enter the ν_e CC $1\pi^+$ sample if some of those pions

reinteract in the nucleus (through the final-state interactions) or are not detected. The uncertainties on the number of hadrons produced in those interactions produce an uncertainty on the fraction that will enter the ν_e CC $1\pi^+$ sample. To assess the potential impact on the δ_{CP} result, the fit of the run 1–9 data was redone using two alternative models for the hadron multiplicity model in the MC: the first (M1) is based on fits of data from deuterium bubble chamber experiments [97], while the second (M2) is based on the multiplicity part of the Andreopoulos-Gallagher-Kehayias-Yang model [98]. Only the alternative model M1 was found to significantly change the expected number of events, with the biggest effect seen for the ν_e CC $1\pi^+$ sample ($+13.5\%$). The predicted reconstructed energy spectrum was also affected for this sample, as the increase in the number of events was primarily at low reconstructed energy.

Figure 68 shows how the $\Delta\chi^2$ obtained in the data fit for δ_{CP} changed when an alternative model was used to fit the data. For each of the two model uncertainties, only the case which gave the largest effect is shown: 10% increase in the SK ν_e CC $1\pi^+$ sample MC predictions for the pion spectrum case and M1 model for the hadron multiplicity model case. Both those changes predict an increase in the number of events in the ν_e CC $1\pi^+$ sample, reducing the discrepancy with the data and therefore weakening the constraint on δ_{CP}. The magnitude of the change in $\Delta\chi^2$ is small enough that it does not change the main conclusions obtained in the fit of the data for δ_{CP}: the CP-conserving values are excluded at the 2σ level, and some values of δ_{CP} are outside of the 3σ confidence level intervals. It is large enough however to change whether $\delta_{CP} = 0$ is excluded at the 3σ confidence level or not in the results of analyses A and C.

C. Neutrino mass ordering

The question of the mass ordering was studied in a Bayesian framework, by computing posterior probabilities
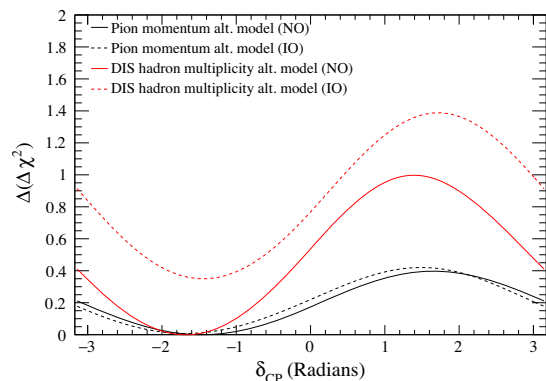

FIG. 68. Change in the $\Delta\chi^2$ function for δ_{CP} when fitting data using an alternative model for DIS hadron multiplicity (red curves) and pion momentum spectra (black curves). The $\Delta\chi^2$ was *reduced* by the value shown on the plot in each case. The difference is shown for the model change that gave the largest difference in each case.

and Bayes factors for each ordering hypothesis. Results shown in this section use the constraint from reactor experiments on $\sin^2 2\theta_{13}$ unless otherwise stated.

1. Posterior probabilities and Bayes factor

All three analyses evaluated posterior probabilities to estimate the preference for the neutrino mass orderings. Analyses B and C additionally looked at the posterior probabilities for the octant of $\sin^2 \theta_{23}$. The practical calculation of the posterior probabilities differs between the analyses, due to differences in the fitting techniques used. Analyses A and C first compute the marginal likelihood for each hypothesis and compute the posterior probability for hypothesis H_i from

$$P(H_i|N^{\text{obs}}, \boldsymbol{x}) = \frac{L_{\text{marg}}(H_i; N^{\text{obs}}, \boldsymbol{x})P(H_i)}{\sum_j L_{\text{marg}}(H_j; N^{\text{obs}}, \boldsymbol{x})P(H_j)}, \quad (17)$$

where the denominator sums over all the possible hypothesis combinations (either the two mass orderings or the four combinations of mass ordering and octant), $(N^{\text{obs}}, \boldsymbol{x})$ is the observed measurement and $P(H)$ is the prior probability, taken to be equal for all the hypotheses. The MCMC-based analysis B calculates the hypotheses' posterior probabilities by counting the number of MCMC steps in the selected hypothesis against the total number of MCMC steps. That ratio gives a fully marginalized posterior probability for a given hypothesis.

Table XIX shows the posterior probabilities for the mass orderings and the octants of $\sin^2 \theta_{23}$ obtained with analysis B. Most of the posterior probability lies in the upper octant and normal mass ordering. The obtained values of the Bayes factors, corresponding to the ratio of the marginal likelihoods of the two hypothesis, are of 8.0 for the normal over the inverted mass orderings and 3.9 for the upper over the lower octant. The commonly used Jeffreys' scale [99] classifies both results as "substantial."

The impact of the details of the analysis methods on the mass ordering results was checked by comparing the results of the different analyses. Analysis C prefers the normal mass ordering and the upper octant of $\sin^2 \theta_{23}$ with 91.1% and 80.4% posterior probability, respectively, whereas analysis A finds a posterior probability of 87.7% for the normal mass ordering. As before, the largest difference is

TABLE XIX. Model comparison posterior probabilities for normal and inverted mass orderings, and for the upper and the lower octants of $\sin^2 \theta_{23}$, from analysis B. There is a preference for the normal mass ordering and upper octant of $\sin^2 \theta_{23}$.

	$\sin^2 \theta_{23} < 0.5$	$\sin^2 \theta_{23} > 0.5$	Sum
NO ($\Delta m_{32}^2 > 0$)	0.184	0.705	0.889
IO ($\Delta m_{32}^2 < 0$)	0.021	0.090	0.111
Sum	0.205	0.795	1

seen between analysis C on one side and A and B on the other, with the main contribution being the choice of variables used for the kinematic information of the candidate events from the appearance samples.

The posterior probabilities above assumed equal prior probabilities for the different hypothesis. The effect of the choice of prior probabilities was checked by looking at how the posterior probabilities obtained in the data fit from analysis C changed as a function of the prior probabilities assumed (Fig. 69). As expected when testing discrete hypotheses using Bayesian methods, the choice of the prior probability has a significant effect on the obtained posterior probabilities. However, the obtained curves are different from $y = \pm x$, demonstrating that the data contain information about the mass ordering.

2. Frequentist properties of the Bayesian results for the mass ordering

As advocated in [100], the frequentist properties of the Bayesian mass ordering results were studied. More precisely, it was checked whether excluding a given ordering hypothesis (based on the other ordering having a posterior probability superior or equal to $\alpha\%$) was selecting the true ordering approximately $\alpha\%$ of the time. For this purpose, 20 000 pseudoexperiments were generated, both for each mass ordering hypothesis and for different true values of δ_{CP}. Then, the fraction of the pseudoexperiments which had a posterior probability higher than 95% for one of the two orderings were studied.

It was found that this fraction depended strongly on the true value of δ_{CP} assumed. Only for true values of δ_{CP} close to $-\pi/2$ could one of the two orderings be excluded a non-negligible part of the time based on this criteria. Additionally, only the normal ordering could have a posterior probability above 95% in this case, regardless of which of the two hypotheses was assumed to be true
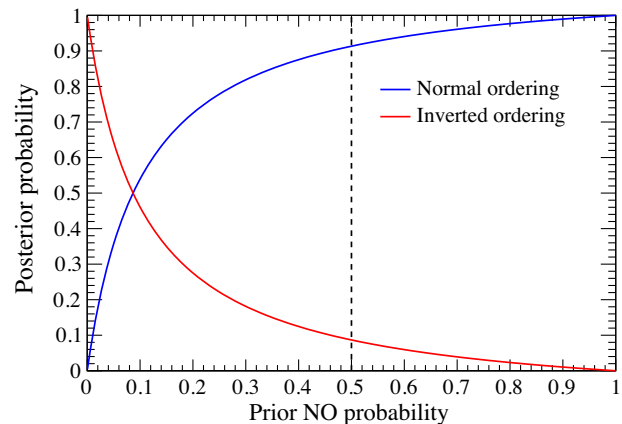

FIG. 69. Mass ordering posterior probabilities as a function of the prior probability assumed for the normal ordering, obtained in the data fit using analysis C. The dashed black line corresponds to the equal prior probability case used for the main result.
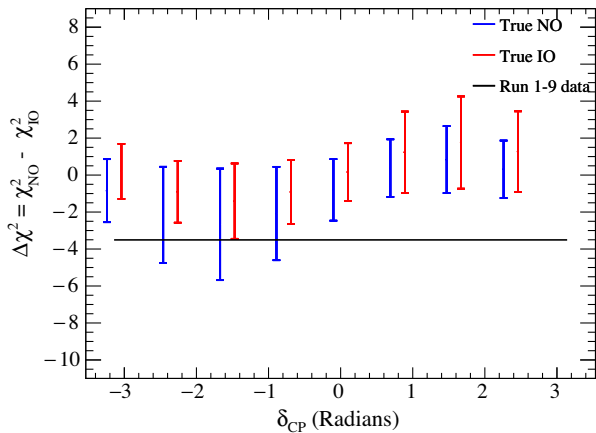
FIG. 70. Ability to distinguish between the two mass orderings as a function of the true value of δ_{CP}. The bar indicates the range of values of $\Delta\chi^2 = \chi^2_{NO} - \chi^2_{IO}$ which contains 95% of the values obtained for the pseudoexperiments generated in each case. The black line corresponds to the value obtained in the data fit by analysis A.

when generating the pseudoexperiments. For true $\delta_{CP} = -\pi/2$, the true ordering was excluded 5.67% of the time the wrong one was. This shows that using the posterior probability to select the mass ordering, although a Bayesian method, nevertheless has reasonable frequentist properties in this case.

The pattern seen highlights the degeneracy between the measurement of δ_{CP} and the determination of the mass ordering (also visible in Fig. 72 of the next section), which have similar effects on the predictions: they both change the two observables (the number of $\nu_\mu \to \nu_e$ and $\bar\nu_\mu \to \bar\nu_e$ events) in opposite directions. The sensitivity to the mass ordering as a function of δ_{CP} (Fig. 70) shows that for $\delta_{CP} < 0$ ($\delta_{CP} > 0$), only for the true ordering being normal (inverted) does one get some expected values of the test statistics outside of the range predicted for the other ordering hypothesis. One can therefore only hope to exclude with non-negligible statistical significance the inverted (normal) ordering in this case.

3. Frequentist results for the mass ordering

For completeness, frequentist results for the mass ordering were derived by computing p values for the Bayes factor. 100 000 pseudoexperiments were generated, randomizing over the nuisance parameters (including $\sin^2\theta_{23}$, δ_{CP} and Δm^2_{32}, following a similar method as for the frequentist 1D δ_{CP} results), and the value of the Bayes factor $P(NO)/P(IO)$ obtained in the data fit was compared to the values obtained for those pseudoexperiments (Fig. 71).

The fraction of the pseudoexperiments for which the Bayes factor was larger than what was observed in the data (corresponding to a result more NO-like) was found to be 4.87×10^{-3} assuming true inverted ordering (inverted ordering p value) and 6.5×10^{-2} assuming true normal
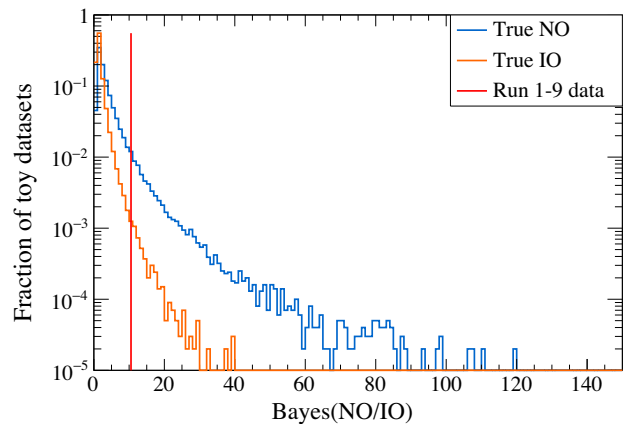

FIG. 71. Expected distributions of the Bayes factor between the mass ordering hypotheses, compared to the value obtained in the data fit.

ordering (1 minus normal ordering p value using standard definitions). Those two values are both low, and it would be misleading to claim exclusion of the inverted ordering based solely on the low p value obtained for this hypothesis. An alternative tool sometimes used in the collider community in such cases is $\mathrm{CL_s}$ [101], in which the p value obtained for one hypothesis is penalized by 1 minus the p value for the other hypothesis:

$$\mathrm{CL_s}(IO) = \frac{p_0(IO)}{1 - p_0(NO)}, \qquad (18)$$

where $p_0(IO)$ and $p_0(NO)$ are the p values for, respectively, the inverted and normal orderings. In this case, $\mathrm{CL_s}(IO) = 0.075$ is obtained. It should be noted that the more commonly used test statistic $\Delta\chi^2 = \chi^2_{NO} - \chi^2_{IO}$ is equal to $-2\ln(\mathrm{Bayes}(NO/IO))$. The p values obtained with this $\Delta\chi^2$ test statistics are therefore the same as the ones presented here for the Bayes factor.

D. Summary of results

To summarize the oscillation parameter constraints, fits to δ_{CP}, $\sin^2\theta_{13}$, $\sin^2\theta_{23}$ and Δm^2_{32} have been produced using constant $\Delta\chi^2$ critical values, and their 1σ confidence intervals are listed in Table XIV and in Table XV with and without using the results of reactor experiments to constrain $\sin^2\theta_{13}$, respectively. Additionally, δ_{CP} critical values have been calculated using the Feldman-Cousins method, and several confidence intervals are listed in Table XVI.

It is valuable to see how different values of δ_{CP}, $\sin^2\theta_{23}$ and the mass ordering affect the predicted event rates. Figure 72 shows the predicted $\bar\nu_e$ event rate vs ν_e event rate for true values of oscillation parameters where δ_{CP} is varied between CP-conserving and maximally CP-violating values, and $\sin^2\theta_{23}$ is varied around maximal mixing, for both mass orderings. Predicted event rates for a given value of $\sin^2\theta_{23}$ and mass ordering are linearly interpolated between
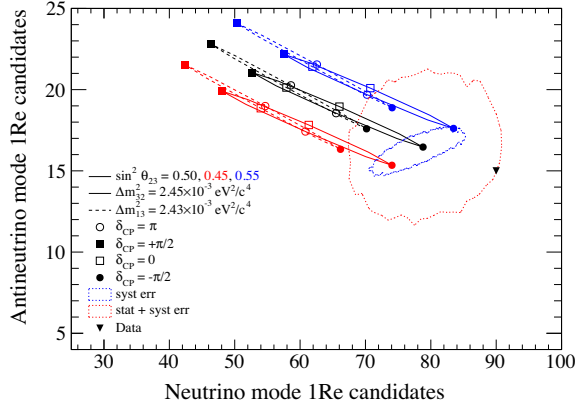
FIG. 72. Candidate RHC one-ring *e*-like event rate vs the candidate FHC one-ring *e*-like event rate (including CC 1π events) for a variety of different oscillation parameter values. Predictions are generated for the given values of δ_{CP}, $\sin^2\theta_{23}$ and mass ordering with remaining oscillation parameters fixed at the central values of the prior probability density functions defined in Sec. XIV A 1. The uncertainty regions are created assuming that $\delta_{CP} = -\pi/2$, $\sin^2\theta_{23} = 0.5$, $\Delta m_{32}^2 = 2.45 \times 10^{-3}$ eV/c^2 and normal mass ordering.

those computed for nine evenly spaced values of δ_{CP} from $-\pi$ to $+\pi$ to indicate the behavior produced by varying δ_{CP}. The observed number of FHC and RHC one-ring electron-like candidate events falls on the edge of the 1σ uncertainty region generated around true $\delta_{CP} = -\pi/2$, $\sin^2\theta_{23} = 0.5$, $\Delta m_{32}^2 = 2.45 \times 10^{-3}$ eV$^2/c^4$ and normal mass ordering and is therefore consistent with this point.

XV. $\bar{\nu}_e$ APPEARANCE ANALYSIS

This section, which differs from the main oscillation analysis reported above, evaluates the significance of the $\bar{\nu}_\mu \to \bar{\nu}_e$ oscillation under the assumption of two different hypotheses, corresponding to no $\bar{\nu}_e$ appearance and to $\bar{\nu}_e$ appearance consistent with our current knowledge of the PMNS mixing parameters. To date, the world's best measurement of $\bar{\nu}_e$ appearance has been made by the NO*v*A Collaboration [13], which reports an excess of 4.4σ over the expected background. The T2K measurement of $\bar{\nu}_e$ appearance has been made using the same analysis framework described in Sec. XII and has been performed by analyses A and C.

The $\bar{\nu}_e$ appearance analysis is performed by multiplying the $\bar{\nu}_\mu \to \bar{\nu}_e$ PMNS oscillation probability by a factor, β, i.e., $P(\bar{\nu}_\mu \to \bar{\nu}_e) = \beta \times P_{PMNS}(\bar{\nu}_\mu \to \bar{\nu}_e)$. The parameter β is set to either 0 or 1 to select a null hypothesis for two independent tests: when $\beta = 0$, the null hypothesis under consideration is that there is no $\bar{\nu}_e$ appearance, while for $\beta = 1$ the null hypothesis is that $\bar{\nu}_e$ appearance occurs according to the current best knowledge of the PMNS parameters. For each hypothesis, p values are produced from two analyses: a rate-only, whose test statistic is the

number of candidate events in the RHC one-ring *e*-like sample, and one rate + shape analysis, whose test statistic is the difference in marginal negative log-likelihood values between the $\beta = 0$ and $\beta = 1$ cases, denoted $\Delta\chi^2$, as in Eq. (19):

$$\Delta\chi^2 = \chi^2_{\text{marg}}(\beta = 0) - \chi^2_{\text{marg}}(\beta = 1). \quad (19)$$

Here the use of χ^2 is taken to be synonymous with $-2\ln L_{\text{marg}}$. Unlike in the main oscillation analysis, the likelihoods are not only marginalized over the flux, cross-section and detector parameters, but also over all oscillation parameters except β, including δ_{CP} and the mass ordering, using 2×10^5 samples of the nuisance parameter space. The number of pseudoexperiments and the number of samples used during marginalization were selected to ensure the stability of the p values.

To calculate p values, the data are compared to distributions of the test statistics from ensembles of pseudo-experiments produced under the assumption of either $\beta = 0$ or $\beta = 1$. Each pseudoexperiment is produced by randomizing nuisance parameters according to the prior probability density functions (PDFs) in Sec. XII. Additionally, a uniform PDF in the range $[-\pi, +\pi]$ is used for δ_{CP}, and a two-point PDF is used for the mass ordering with equal probability for normal and inverted ordering.

T2K data from four "control samples" (FHC single-ring *e*-like and ν_eCC1π-like and both neutrino and RHC single-ring μ-like) are used to constrain the oscillation and systematic model parameters. The impact of these four data control samples is estimated differently in the two analyses, A and C. In the latter, each of the 2×10^4 pseudoexperiments used to build the distributions of the rate-only and rate + shape test statistics is weighted by its likelihood over the four control samples, given the T2K data. Analysis A uses an alternative method, applying this constraint by using rejection sampling to select the pseudoexperiments that are most probable according to the data in the control samples.

Due to the presence of the four control samples, the marginal likelihood used to calculate $\Delta\chi^2$, the rate + shape test statistic, is also constructed differently from the main oscillation analysis. For each pseudoexperiment, the marginal likelihood is constructed based on the prediction in the RHC single-ring *e*-like sample, while its background is constrained using the four control samples. This is reflected in Eq. (20), where $L_{\bar{\nu}_e}$ denotes the likelihood of the RHC single-ring *e*-like sample compared to the pseudoexperiment E, L_c is the product of the likelihoods of the four control samples compared to the T2K data D, and $\boldsymbol{f}_j$ the set of nuisance parameters:

$$\chi^2_{\text{marg}}(\beta) = -2\ln\left[\frac{1}{n}\sum_{j=1}^{n} L_{\bar{\nu}_e}(\beta, \boldsymbol{f}_j; E) \cdot L_c(\beta, \boldsymbol{f}_j; D)\right]. \quad (20)$$
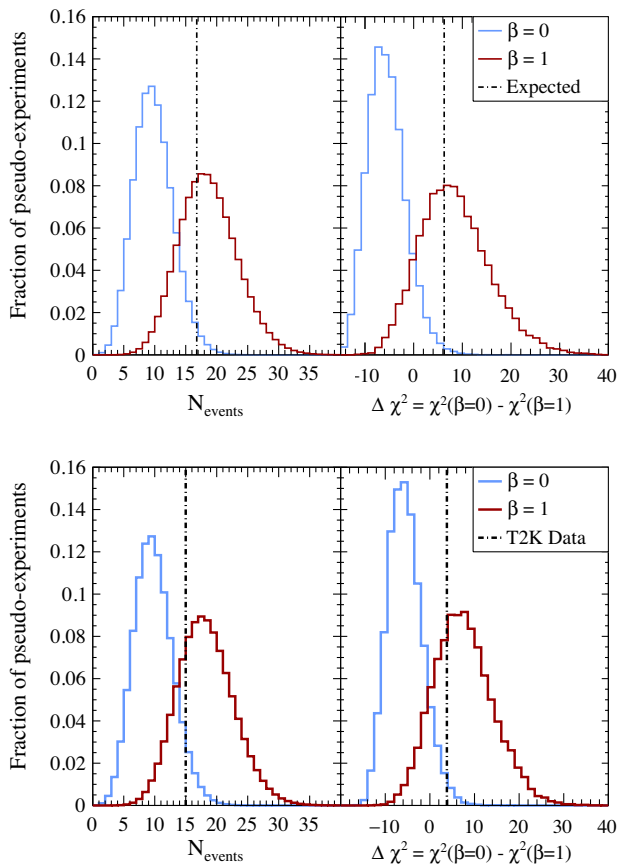
FIG. 73. Distributions of the expected (top) and observed (bottom) rate-only (left) and rate + shape (right) test statistics compared to the expected and observed test statistics. Here N_{events} denotes the number of observed events in the RHC single-ring e-like sample.

The expected and observed test statistic distributions produced using analysis A are shown in Fig. 73 and the corresponding p values are shown in Table XX. The hypothesis of no $\bar{\nu}_e$ appearance ($\beta = 0$) is excluded at the 1.9σ and 2.4σ levels, respectively, using the rate-only and rate + shape analyses. The observed p values provide a weaker exclusion of the $\beta = 0$ case than expected with Asimov dataset, in both rate-only and rate + shape analyses. This is primarily due to observing fewer events (15) in the RHC single-ring e-like sample than expected (16.8) and

TABLE XXI. Nominal and alternative model p values of the $\beta = 0$ and $\beta = 1$ hypotheses using the rate-only analysis.

	$\beta = 0$	$\beta = 1$
Model	p value (σ)	p value (σ)
Nominal	0.0686 (1.82)	0.246 (1.16)
Kabirnezhad single pion	0.0824 (1.74)	0.176 (1.35)
Nieves LFG 1p1h	0.0804 (1.75)	0.222 (1.22)
Data-driven 2p2h-Δ	0.0859 (1.72)	0.211 (1.25)

strengthened by their relatively backgroundlike spectrum in the rate + shape analysis, as shown in Fig. 44(d). Our data are consistent with the PMNS $\bar{\nu}_e$ appearance hypothesis ($\beta = 1$), with p values of 0.32 and 0.30 for the rate-only and rate + shape analyses, respectively (corresponding to a 1σ exclusion). Analyses A and C both produce test statistic distributions and p values that are in good agreement with each other, where minor differences between them were determined to result from the difference in kinematic variables used to bin the analysis templates.

It is also desirable to test the robustness of these results to various alternative choices of parts of the interaction model which have a non-negligible effect on the kinematic distributions of the RHC single-ring e-like sample. This is done using analysis C by weighting the nominal near and far detector MC (see Sec. XIII). Three simulated datasets are studied: the Kabirnezhad single-pion production model, the Nieves LFG model and the data-driven CC $0\pi E_\nu - Q^2$ simulated data, described in Sec. VII. To test their effects on the observed results of the $\bar{\nu}_e$ appearance analysis, the test statistic distributions are weighted by the ratio of the distribution produced using the alternative model to the expected distribution. Similarly, the observed test statistics are shifted by the difference between the median expected pseudoexperiment with and without the use of the alternative model. The change in the observed p values from these alternative models are shown in Tables XXI and XXII for the rate and rate + shape analyses, respectively. The observed changes are small compared to the nominal p value and the changes do not affect the conclusions, so the analyses are considered robust against alternative model choices.

TABLE XX. Expected and observed p values and significance of the $\beta = 0$ and $\beta = 1$ hypotheses using both the rate-only and rate + shape analyses.

		p value		Significance (σ)	
β	Analysis	Expected	Observed	Expected	Observed
0	Rate-only	0.019	0.059	2.36	1.89
	Rate + shape	0.006	0.016	2.76	2.40
1	Rate-only	0.379	0.321	0.88	0.99
	Rate + shape	0.409	0.300	0.83	1.04

TABLE XXII. Nominal and alternative model p values of the $\beta = 0$ and $\beta = 1$ hypotheses using the rate + shape analysis.

	$\beta = 0$	$\beta = 1$
Model	p value (σ)	p value (σ)
Nominal	0.0244 (2.25)	0.261 (1.12)
Kabirnezhad single pion	0.0227 (2.28)	0.225 (1.21)
Nieves LFG 1p1h	0.0201 (2.32)	0.277 (1.09)
Data-driven 2p2h-Δ	0.0178 (2.37)	0.301 (1.03)

XVI. CONCLUSIONS

The T2K Collaboration has analyzed the full dataset collected by the experiment between 2010 and 2018 to produce measurements of $\sin^2 \theta_{23}$, Δm^2_{32}, $\sin^2 \theta_{13}$, δ_{CP} and the mass ordering. The parameter values and uncertainties are taken from a simultaneous fit to both muonlike and electronlike event samples at Super-Kamiokande, collected from both neutrino-dominated and antineutrino-dominated beam operation. This analysis uses a new event selection at SK to increase the effective fiducial volume of the detector, resulting in a 20% increase in the electronlike sample efficiency and a 40% reduction in the background contamination in the muonlike samples.

The neutrino interaction model used for this work is improved relative to Ref. [26], incorporating in-medium effects (RPA) and 2p2h shape uncertainties in the charged-current zero-pion signal channel. A detailed set of simulated data studies were also performed to assess the robustness of the analysis to alternative choices of neutrino interaction model. These studies demonstrated that the nucleon removal energy uncertainty can have a significant effect on the allowed regions for $\sin^2 \theta_{23}$ and Δm^2_{32} and, to a lesser extent, Δm^2_{32} was also sensitive to all of the alternative models studied. This led to the addition of new systematic uncertainties to the SK samples and oscillation contours to account for these effects. The results of these simulated data studies are an important outcome of this analysis: long-baseline experiments are now entering the precision era, and changes of a few percent in the reconstructed neutrino energy can have significant effects on the oscillation parameters observed by the experiment.

The T2K oscillation parameter measurements are, however, limited by statistics. T2K will collect more data in both neutrino and antineutrino beam operation mode. More complex event topologies will be included in the analyses at both ND280 and Super-Kamiokande, and the Collaboration is working toward joint analyses with both the Super-Kamiokande and the NOvA experiments. The combination of new topologies and differing neutrino energies will lift the degeneracies between oscillation parameters present in a single experiment. This will enable the most sensitive measurements of neutrino oscillations to date and will provide a template for future analyses at the next generation of long-baseline experiments.

The data related to the measurement and results presented in this paper can be found in [102].

APPENDIX: POST-ND280-FIT PARAMETER VALUES

The nominal and post-ND280-fit values from the MINUIT-based analysis are shown in Tables XXIII and XXIV for the flux parameters and Table XXV for the cross-section parameters.

TABLE XXIII. Prefit and postfit weights for the SK FHC flux parameters.

FHC SK flux parameter (GeV)	Prefit	ND280 Postfit
SK ν_μ [0.0, 0.4]	1.00 ± 0.10	1.01 ± 0.06
SK ν_μ [0.4, 0.5]	1.00 ± 0.10	1.03 ± 0.05
SK ν_μ [0.5, 0.6]	1.00 ± 0.09	1.02 ± 0.05
SK ν_μ [0.6, 0.7]	1.00 ± 0.08	0.98 ± 0.04
SK ν_μ [0.7, 1.0]	1.00 ± 0.10	0.93 ± 0.06
SK ν_μ [1.0, 1.5]	1.00 ± 0.09	0.95 ± 0.05
SK ν_μ [1.5, 2.5]	1.00 ± 0.07	1.02 ± 0.04
SK ν_μ [2.5, 3.5]	1.00 ± 0.07	1.04 ± 0.05
SK ν_μ [3.5, 5.0]	1.00 ± 0.09	1.03 ± 0.04
SK ν_μ [5.0, 7.0]	1.00 ± 0.10	0.99 ± 0.04
SK ν_μ [7.0, ∞]	1.00 ± 0.11	0.97 ± 0.05
SK $\bar{\nu}_\mu$ [0.0, 0.7]	1.00 ± 0.10	0.98 ± 0.08
SK $\bar{\nu}_\mu$ [0.7, 1.0]	1.00 ± 0.08	0.97 ± 0.05
SK $\bar{\nu}_\mu$ [1.0, 1.5]	1.00 ± 0.08	0.98 ± 0.06
SK $\bar{\nu}_\mu$ [1.5, 2.5]	1.00 ± 0.08	1.03 ± 0.06
SK $\bar{\nu}_\mu$ [2.5, ∞]	1.00 ± 0.09	1.10 ± 0.07
SK ν_e [0.0, 0.5]	1.00 ± 0.09	1.02 ± 0.05
SK ν_e [0.5, 0.7]	1.00 ± 0.09	1.02 ± 0.04
SK ν_e [0.7, 0.8]	1.00 ± 0.08	1.02 ± 0.04
SK ν_e [0.8, 1.5]	1.00 ± 0.08	1.01 ± 0.04
SK ν_e [1.5, 2.5]	1.00 ± 0.08	1.03 ± 0.04
SK ν_e [2.5, 4.0]	1.00 ± 0.08	1.03 ± 0.04
SK ν_e [4.0, ∞]	1.00 ± 0.09	1.03 ± 0.06
SK $\bar{\nu}_e$ [0.0, 2.5]	1.00 ± 0.07	1.04 ± 0.06
SK $\bar{\nu}_e$ [2.5, ∞]	1.00 ± 0.13	1.08 ± 0.12

TABLE XXIV. Prefit and postfit weights for the SK RHC flux parameters.

RHC SK flux parameter (GeV)	Prefit	ND280 Postfit
SK ν_μ [0.0, 0.7]	1.00 ± 0.09	0.98 ± 0.07
SK ν_μ [0.7, 1.0]	1.00 ± 0.08	0.99 ± 0.05
SK ν_μ [1.0, 1.5]	1.00 ± 0.08	1.00 ± 0.05
SK ν_μ [1.5, 2.5]	1.00 ± 0.08	1.05 ± 0.05
SK ν_μ [2.5, ∞]	1.00 ± 0.08	1.04 ± 0.05
SK $\bar{\nu}_\mu$ [0.0, 0.4]	1.00 ± 0.11	1.00 ± 0.07
SK $\bar{\nu}_\mu$ [0.4, 0.5]	1.00 ± 0.10	1.01 ± 0.05
SK $\bar{\nu}_\mu$ [0.5, 0.6]	1.00 ± 0.09	0.99 ± 0.05
SK $\bar{\nu}_\mu$ [0.6, 0.7]	1.00 ± 0.08	0.97 ± 0.04
SK $\bar{\nu}_\mu$ [0.7, 1.0]	1.00 ± 0.10	0.97 ± 0.05
SK $\bar{\nu}_\mu$ [1.0, 1.5]	1.00 ± 0.09	0.99 ± 0.05
SK $\bar{\nu}_\mu$ [1.5, 2.5]	1.00 ± 0.07	1.03 ± 0.04
SK $\bar{\nu}_\mu$ [2.5, 3.5]	1.00 ± 0.07	1.06 ± 0.05
SK $\bar{\nu}_\mu$ [3.5, 5.0]	1.00 ± 0.09	1.06 ± 0.07

(Table continued)

TABLE XXIV. *(Continued)*

RHC SK flux parameter (GeV)	Prefit	ND280 Postfit
SK $\bar{\nu}_\mu$ [5.0, 7.0]	1.00 ± 0.09	1.04 ± 0.06
SK $\bar{\nu}_\mu$ [7.0, ∞]	1.00 ± 0.12	1.00 ± 0.09
SK ν_e [0.0, 2.5]	1.00 ± 0.07	1.04 ± 0.05
SK ν_e [2.5, ∞]	1.00 ± 0.08	1.04 ± 0.07
SK $\bar{\nu}_e$ [0.0, 0.5]	1.00 ± 0.10	1.01 ± 0.05
SK $\bar{\nu}_e$ [0.5, 0.7]	1.00 ± 0.09	1.00 ± 0.05
SK $\bar{\nu}_e$ [0.7, 0.8]	1.00 ± 0.09	1.00 ± 0.05
SK $\bar{\nu}_e$ [0.8, 1.5]	1.00 ± 0.08	1.01 ± 0.04
SK $\bar{\nu}_e$ [1.5, 2.5]	1.00 ± 0.08	1.04 ± 0.05
SK $\bar{\nu}_e$ [2.5, 4.0]	1.00 ± 0.09	1.04 ± 0.07
SK $\bar{\nu}_e$ [4.0, ∞]	1.00 ± 0.15	1.08 ± 0.13

TABLE XXV. Prefit and postfit values for the cross-section parameters.

Cross-section parameter	Prefit	ND280 Postfit
M_A^{QE} (GeV/c^2)	1.20 ± 0.03	1.13 ± 0.08
pF ^{12}C(MeV/c)	217 ± 13	224 ± 13
pF ^{16}O(MeV/c)	225 ± 13	205 ± 15
2p2h norm ν	1.00 ± 1.00	1.50 ± 0.20
2p2h norm $\bar{\nu}$	1.00 ± 1.00	0.73 ± 0.23
2p2h norm ^{12}C/^{16}O ratio	1.00 ± 0.20	0.96 ± 0.17
2p2h shape ^{12}C	1.00 ± 3.00	2.00 ± 0.21
2p2h shape ^{16}O	1.00 ± 3.00	2.00 ± 0.35
BeRPA A	0.59 ± 0.12	0.69 ± 0.06
BeRPA B	1.05 ± 0.21	1.60 ± 0.12
BeRPA D	1.13 ± 0.17	0.96 ± 0.13
BeRPA E	0.88 ± 0.35	0.87 ± 0.35
BeRPA U	1.20 ± 0.10	1.20 ± 0.10
C_A^5	0.96 ± 0.15	0.98 ± 0.06
M_A^{RES}(GeV/c^2)	1.07 ± 0.15	0.81 ± 0.04
$I = \frac{1}{2}$ background	0.96 ± 0.40	1.31 ± 0.26
ν_e/ν_μ	1.00 ± 0.03	1.00 ± 0.03
$\bar{\nu}_e/\bar{\nu}_\mu$	1.00 ± 0.03	1.00 ± 0.03
CC DIS	0.00 ± 0.40	0.39 ± 0.21
CC coherent ^{12}C	1.00 ± 0.30	0.87 ± 0.28
CC coherent ^{16}O	1.00 ± 0.30	0.87 ± 0.28
NC coherent	1.00 ± 0.30	0.94 ± 0.30
NC 1γ	1.00 ± 1.00	1.00 ± 1.00
NC other ND280	1.00 ± 0.30	1.21 ± 0.26
NC other SK	1.00 ± 0.30	1.00 ± 0.30
FSI inelastic low-E	0.00 ± 0.41	-0.32 ± 0.08
FSI inelastic high-E	0.00 ± 0.34	-0.01 ± 0.13
FSI pion production	0.00 ± 0.50	0.04 ± 0.19
FSI pion absorption	0.00 ± 0.41	-0.35 ± 0.15
FSI charge exch. low-E	0.00 ± 0.57	-0.09 ± 0.31
FSI charge exch. high-E	0.00 ± 0.28	0.02 ± 0.10

[1] Z. Maki, M. Nakagawa, and S. Sakata, Prog. Theor. Phys. **28**, 870 (1962).

[2] B. Pontecorvo, ZhETF **53**, 1717 (1968) [Sov. Phys. JETP **26**, 984 (1968)], http://www.jetp.ac.ru/cgi-bin/e/index/e/26/5/p984?a=list.

[3] B. Aharmim *et al.* (SNO Collaboration), Phys. Rev. C **88**, 025501 (2013).

[4] K. Abe *et al.* (Super-Kamiokande Collaboration), Phys. Rev. D **97**, 072001 (2018).

[5] K. Abe *et al.* (Super-Kamiokande Collaboration), Phys. Rev. D **94**, 052010 (2016).

[6] M. Ahn *et al.* (K2K Collaboration), Phys. Rev. D **74**, 072003 (2006).

[7] P. Adamson *et al.* (MINOS Collaboration), Phys. Rev. Lett. **112**, 191801 (2014).

[8] A. Gando *et al.* (KamLAND Collaboration), Phys. Rev. D **88**, 033001 (2013).

[9] L.-L. Chau and W.-Y. Keung, Phys. Rev. Lett. **53**, 1802 (1984).

[10] D. Adey *et al.* (Daya Bay Collaboration), Phys. Rev. Lett. **121**, 241805 (2018).

[11] G. Bak *et al.* (RENO Collaboration), Phys. Rev. Lett. **121**, 201801 (2018).

[12] Y. Abe *et al.* (Double Chooz Collaboration), J. High Energy Phys. 01 (2016) 163.

[13] M. Acero *et al.* (NOvA Collaboration), Phys. Rev. Lett. **123**, 151803 (2019).

[14] M. G. Aartsen *et al.* (IceCube Collaboration), Phys. Rev. Lett. **120**, 071801 (2018).

[15] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).

[16] K. Abe *et al.* (T2K Collaboration), Nature (London) **580**, 339 (2020).

[17] L. Wolfenstein, Phys. Rev. D **17**, 2369 (1978).

[18] S. Mikheev and A. Smirnov, Il Nuovo Cimento C **9**, 17 (1986).

[19] K. Abe *et al.* (T2K Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **659**, 106 (2011).

[20] T. Sekiguchi *et al.* (T2K Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **789**, 57 (2015).

[21] H. Nunokawa, S. J. Parke, and J. W. Valle, Prog. Part. Nucl. Phys. **60**, 338 (2008).

[22] V. Barger, K. Whisnant, S. Pakvasa, and R. J. N. Phillips, Phys. Rev. D **22**, 2718 (1980).

[23] K. Hagiwara, N. Okamura, and K.-I. Senda, J. High Energy Phys. 09 (2011) 082.

[24] H. Yokomakura, K. Kimura, and A. Takamura, Phys. Lett. B **544**, 286 (2002).

[25] K. Abe *et al.* (T2K Collaboration), Phys. Rev. Lett. **121**, 171802 (2018).

[26] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **96**, 092006 (2017); **98**, 019902(E) (2018).

[27] M. Jiang *et al.* (Super-Kamiokande Collaboration), Prog. Theor. Exp. Phys. **2019**, 053F01 (2019).

[28] K. Abe *et al.* (T2K Collaboration), Phys. Rev. Lett. **112**, 061802 (2014).

[29] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **91**, 072010 (2015).

[30] K. Olive *et al.* (Particle Data Group), Chin. Phys. C **38**, 090001 (2014).

[31] M. Tanabashi *et al.* (Particle Data Group), Phys. Rev. D **98**, 030001 (2018).

[32] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **87**, 012001 (2013); **87**, 019902(E) (2013).

[33] A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft, CERN yellow reports: Monographs, 10.2172/877507 (2005).

[34] G. Battistoni, T. Boehlen, F. Cerutti, P. W. Chin, L. S. Esposito, A. Fassò, A. Ferrari, A. Lechner, A. Empl, A. Mairani, A. Mereghetti, P. G. Ortega, J. Ranft, S. Roesler, P. R. Sala, V. Vlachoudis, and G. Smirnov, *Ann. Nucl. Energy* 82, 10 (2015), Joint International Conference on Supercomputing in Nuclear Applications and Monte Carlo 2013, SNA + MC 2013. Pluri- and Trans-disciplinarity, Towards New Modeling and Numerical Simulation Paradigms.

[35] 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, Nucl. Data Sheets **120**, 211 (2014).

[36] R. Brun, F. Bruyant, F. Carminati, S. Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban, GEANT: *Detector Description and Simulation Tool; Oct 1994*, CERN Program Library (CERN, Geneva, 1993), long Writeup W5013.

[37] C. Zeitnitz and T. Gabriel, in *Proceedings of the 3rd International Conference on Calorimetry in High-Energy Physics* (1992), pp. 394–404, https://inspirehep.net/literature/348996.

[38] A. Fasso, A. Ferrari, J. Ranft, and P. Sala, Conf. Proc. C **9309194**, 493 (1993), https://inspirehep.net/literature/367393.

[39] N. Abgrall *et al.* (NA61/SHINE Collaboration), Phys. Rev. C **84**, 034604 (2011).

[40] N. Abgrall *et al.* (NA61/SHINE Collaboration), Phys. Rev. C **85**, 035210 (2012).

[41] N. Abgrall *et al.* (NA61/SHINE Collaboration), Phys. Rev. C **89**, 025205 (2014).

[42] N. Abgrall *et al.* (NA61/SHINE Collaboration), Eur. Phys. J. C **76**, 84 (2016).

[43] K. Abe *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **694**, 211 (2012).

[44] Y. Hayato, Acta Phys. Pol. B **40**, 2477 (2009), https://inspirehep.net/literature/844435.

[45] C. Wilkinson *et al.*, Phys. Rev. D **93**, 072010 (2016).

[46] C. H. Llewellyn Smith, Phys. Rep. **3**, 261 (1972).

[47] R. Bradford, A. Bodek, H. Budd, and J. Arrington, Nucl. Phys. B, Proc. Suppl. **159**, 127 (2006).

[48] J. G. Morfin, J. Nieves, and J. T. Sobczyk, Adv. High Energy Phys. **2012**, 934597 (2012).

[49] J. Nieves, R. Gran, I. R. Simo, F. Sánchez, and M. V. Vacas, Nucl. Part. Phys. Proc. **273–275**, 1830 (2016), 37th International Conference on High Energy Physics (ICHEP).

[50] R. Gran, J. Nieves, F. Sanchez, and M. J. V. Vacas, Phys. Rev. D **88**, 113007 (2013).

[51] D. Rein and L. M. Sehgal, Ann. Phys. (N.Y.) **133**, 79 (1981).

[52] C. Berger and L. M. Sehgal, Phys. Rev. D **76**, 113004 (2007).

[53] K. M. Graczyk, J. Żmuda, and J. T. Sobczyk, Phys. Rev. D **90**, 093001 (2014).

[54] L. Alvarez-Ruso *et al.*, Prog. Part. Nucl. Phys. **100,** 1 (2018).

[55] P. Stowell, C. Wret, C. Wilkinson, L. Pickering, S. Cartwright, Y. Hayato, K. Mahn, K. McFarland, J. Sobczyk, R. Terri *et al.*, J. Instrum. **12,** P01016 (2017).

[56] G. M. Radecky *et al.*, Phys. Rev. D **25,** 1161 (1982).

[57] C. Wilkinson, P. Rodrigues, S. Cartwright, L. Thompson, and K. McFarland, Phys. Rev. D **90,** 112017 (2014).

[58] T. Kitagaki *et al.*, Phys. Rev. D **34,** 2554 (1986).

[59] A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Phys. Rev. D **83,** 052007 (2011).

[60] B. Eberly *et al.* (MINERvA Collaboration), Phys. Rev. D **92,** 092008 (2015).

[61] C. L. McGivern *et al.* (MINERvA Collaboration), Phys. Rev. D **94,** 052005 (2016).

[62] A. Mislivec *et al.* (MINERvA Collaboration), Phys. Rev. D **97,** 032014 (2018).

[63] C. Berger and L. M. Sehgal, Phys. Rev. D **79,** 053003 (2009).

[64] M. Glück, E. Reya, and A. Vogt, Eur. Phys. J. C **5,** 461 (1998).

[65] A. Bodek and U. K. Yang, arXiv:hep-ex/0308007.

[66] T. Sjöstrand, Comput. Phys. Commun. **82,** 74 (1994).

[67] H. Ejiri, Phys. Rev. C **48,** 1442 (1993).

[68] M. Valverde, J. Amaro, and J. Nieves, Phys. Lett. B **638,** 325 (2006).

[69] M. Martini, M. Ericson, G. Chanfray, and J. Marteau, Phys. Rev. C **80,** 065501 (2009).

[70] I. R. Simo, J. E. Amaro, M. B. Barbaro, A. D. Pace, J. A. Caballero, and T. W. Donnelly, J. Phys. G **44,** 065105 (2017).

[71] K. Gallmeister, U. Mosel, and J. Weil, Phys. Rev. C **94,** 035502 (2016).

[72] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **93,** 112012 (2016).

[73] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **98,** 032003 (2018).

[74] K. Abe *et al.* (T2K Collaboration), Phys. Rev. D **101,** 112001 (2020).

[75] P. A. Rodrigues *et al.* (MINERvA Collaboration), Phys. Rev. Lett. **116,** 071802 (2016).

[76] R. Gran *et al.* (MINERvA Collaboration), Phys. Rev. Lett. **120,** 221805 (2018).

[77] M. A. Acero *et al.* (NOvA Collaboration), Phys. Rev. D **98,** 032012 (2018).

[78] O. Benhar, N. Farina, H. Nakamura, M. Sakuda, and R. Seki, Phys. Rev. D **72,** 053005 (2005).

[79] A. Bodek and T. Cai, Eur. Phys. J. C **79,** 293 (2019).

[80] M. Kabirnezhad, Phys. Rev. D **97,** 013002 (2018).

[81] P. Guèye *et al.*, Phys. Rev. C **60,** 044308 (1999).

[82] P.-A. Amaudruz, M. Barbi, D. Bishop, N. Braam, D. Brook-Roberge, S. Giffin, S. Gomi, P. Gumplinger, K. Hamano, N. Hastings *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **696,** 1 (2012).

[83] N. Abgrall, B. Andrieu, P. Baron, P. Bene, V. Berardi, J. Beucher, P. Birney, F. Blaszczyk, A. Blondel, C. Bojechko *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **637,** 25 (2011).

[84] Y. Fukuda *et al.* (Super-Kamiokande Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **501,** 418 (2003).

[85] S. Yamada, Y. Hayato, Y. Obayashi, and M. Shiozawa, New online system without hardware trigger for the Super-Kamiokande experiment, in *Proceedings of the 2007 IEEE Nuclear Science Symposium Conference Record* (IEEE, New York, 2007), Vol. 1, pp. 111–114, http://dx.doi.org/10.1109/NSSMIC.2007.4436298.

[86] K. Abe *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **737,** 253 (2014).

[87] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur. Phys. J. C **71,** 1554 (2011).

[88] W. K. Hastings, Biometrika **57,** 97 (1970).

[89] K. Abe *et al.* (T2K Collaboration), Phys. Rev. Lett. **107,** 041801 (2011).

[90] R. Wendell *et al.* (Super-Kamiokande Collaboration), Phys. Rev. D **81,** 092004 (2010).

[91] F. James and M. Roos, Comput. Phys. Commun. **10,** 343 (1975).

[92] A. Sztuc, Ph.D. thesis, Imperial College London, 2020.

[93] P. Coloma, P. Huber, C.-M. Jen, and C. Mariani, Phys. Rev. D **89,** 073015 (2014).

[94] R. D. Cousins, arXiv:1807.05996.

[95] S. Wilks, Ann. Math. Stat. **9,** 60 (1938).

[96] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57,** 3873 (1998).

[97] C. Bronner and M. Hartz, J. Phys. Soc. Jpn. Conf. Proc. **12,** 010041 (2016).

[98] T. Yang, C. Andreopoulos, H. Gallagher, K. Hoffmann, and P. Kehayias, Eur. Phys. J. C **63,** 1 (2009).

[99] H. Jeffreys, *The Theory of Probability* (Oxford University Press, New York, 1998).

[100] M. Bayarri and J. Berger, Stat. Sci. **19,** 58 (2004).

[101] A. L. Read, J. Phys. G **28,** 2693 (2002).

[102] K. Abe *et al.* (T2K Collaboration), https://t2k-experiment.org/results/t2kdata-oa-2019.