First search for sterile neutrino oscillation leading to ν_μ disappearance in the Booster Neutrino Beam at ICARUS

F. Abd Alrahman,[9] P. Abratenko,[28] N. Abrego-Martinez,[5] A. Aduszkiewicz,[9] F. Akbar,[24] L. Aliaga Soplin,[27] M. Artero Pons,[20] J. Asaadi,[27] W. F. Badgett,[8] B. Baibussinov,[20] F. Battisti,[11] B. Behera,[10] V. Bellini,[12] R. Benocci,[18] J. Berger,[6] S. Bertolucci,[11] M. Betancourt,[8] A. Blanchet,[3] F. Boffelli,[21] M. Bonesini,[18] T. Boone,[6] B. Bottino,[13] A. Braggiotti,[20,*] S. J. Brice,[8] V. Brio,[12] C. Brizzolari,[18] H. S. Budd,[24] A. Campani,[13] A. Campos,[29] D. Carber,[6] M. Carneiro,[1] I. Caro Terrazas,[6] H. Carranza,[27] R. Castillo Fernández,[27] S. Centro,[20] G. Cerati,[8] A. Chatterjee,[3] D. Cherdack,[9] S. Cherubini,[16] N. Chithirasreemadam,[22] M. Cicerchia,[20] T. E. Coan,[26] A. Cocco,[19] M. R. Convery,[25] L. Cooper-Troendle,[23] S. Copello,[21] H. Da Motta,[2] M. Dallolio,[27] A. A. Dange,[27] A. de Roeck,[3] S. Di Domizio,[13] L. Di Noto,[13] D. Di Ferdinando,[11] M. Diwan,[1] S. Dolan,[3] S. Donati,[22] F. Drielsma,[25] J. Dyer,[6] A. Falcone,[18] C. Farnese,[20] A. Fava,[8] N. Gallice,[1] C. Gatto,[19] D. Gibin,[20] A. Gioiosa,[22] W. Gu,[1] A. Guglielmi,[20] G. Gurung,[3,†] K. Hassinin,[9] H. Hausner,[8] A. Heggestuen,[6] B. Howard,[30,‡] R. Howell,[24] Z. Hulcher,[25] G. Ingratta,[11] M. S. Ismail,[23] C. James,[8] W. Jang,[27] K. Jung,[24] Y.-J. Jwa,[25] L. Kashur,[6] W. Ketchum,[8] J. S. Kim,[24] D.-H. Koh,[25] J. Larkin,[24] Y. Li,[1] Z. Lin,[24] C. Mariani,[29] C. M. Marshall,[24] S. Martynenko,[1] N. Mauri,[11] K. S. McFarland,[24] D. P. Méndez,[1] A. Menegolli,[21] G. Meng,[20] O. G. Miranda,[5] A. Mogan,[6] N. Moggi,[11] E. Montagna,[11] C. Montanari,[8,§] A. Montanari,[11] M. Mooney,[6] M. Moore,[25] G. Moreno-Granados,[29] J. Mueller,[8] M. Murphy,[29] D. Naples,[23] S. Palestini,[3] M. Pallavicini,[13] V. Paolone,[23] L. Pasqualini,[11] L. Patrizii,[11] G. Petrillo,[25] C. Petta,[12] V. Pia,[11] F. Pietropaolo,[8,‖] F. Poppi,[11] M. Pozzato,[11] M. L. Pumo,[16] G. Putnam,[8] X. Qian,[1] A. Rappoldi,[21] G. L. Raselli,[21] S. Repetto,[13] F. Resnati,[3] A. M. Ricci,[22] M. Rosenberg,[28] M. Ross-Lonergan,[7] M. Rossella,[21] N. Rowe,[4] P. Roy,[29] C. Rubbia,[14] A. Ruggeri,[11] S. Saha,[23] G. Salmoria,[2] S. Samanta,[13] A. Scaramelli,[21] D. Schmitz,[4] A. Schukraft,[8] D. Senadheera,[23] S-H. Seo,[8] F. Sergiampietri,[3,¶] G. Sirri,[11] J. S. Smedley,[24] J. Smith,[1,**] M. Sotgia,[13] L. Stanco,[20] J. Stewart,[1] H. A. Tanaka,[25] M. Tenti,[11] K. Terao,[25] F. Terranova,[18] V. Togo,[11] D. Torretta,[8] M. Torti,[18] D. Totani,[6] R. Triozzi,[20] Y.-T. Tsai,[25] K. V. Tsang,[25] T. Usher,[25] F. Varanini,[20] N. Vardy,[6] S. Ventura,[20] M. Vicenzi,[1] C. Vignoli,[15] F. A. Wieler,[2] Z. Williams,[27] R. J. Wilson,[6] P. Wilson,[8] J. Wolfs,[24] T. Wongjirad,[28] A. Wood,[9] E. Worcester,[1] M. Worcester,[1] S. Yadav,[27] H. Yu,[1] J. Yu,[27] A. Zani,[17] J. Zennamo,[8] J. Zettlemoyer,[8] and S. Zucchelli[11]

(ICARUS Collaboration)

[1]*Brookhaven National Laboratory, Upton, New York 11973, USA*
[2]*CBPF, Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil*
[3]*CERN, European Organization for Nuclear Research 1211 Genève 23, Switzerland, CERN*
[4]*University of Chicago, Chicago, Illinois 60637, USA*
[5]*Centro de Investigacion y de Estudios Avanzados del IPN (Cinvestav), Mexico City*
[6]*Colorado State University, Fort Collins, Colorado 80523, USA*
[7]*Columbia University, New York, New York, 10027, USA*
[8]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
[9]*University of Houston, Houston, Texas 77204, USA*
[10]*Indian Institute of Science, Bengaluru, India*
[11]*INFN Sezione di Bologna and University, Bologna, Italy*
[12]*INFN Sezione di Catania and University, Catania, Italy*
[13]*INFN Sezione di Genova and University, Genova, Italy*
[14]*INFN GSSI, L'Aquila, Italy*
[15]*INFN LNGS, Assergi, Italy*
[16]*INFN LNS, Catania, Italy*
[17]*INFN Sezione di Milano, Milano, Italy*
[18]*INFN Sezione di Milano Bicocca and University, Milano, Italy*
[19]*INFN Sezione di Napoli, Napoli, Italy*
[20]*INFN Sezione di Padova and University, Padova, Italy*
[21]*INFN Sezione di Pavia and University, Pavia, Italy*
[22]*INFN Sezione di Pisa, Pisa, Italy*
[23]*University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*
[24]*University of Rochester, Rochester, New York 14627, USA*
[25]*SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA*
[26]*Southern Methodist University, Dallas, Texas 75275, USA*
[27]*University of Texas at Arlington, Arlington, Texas 76019, USA*

[28]*Tufts University, Medford, Massachusetts 02155, USA*
[29]*Virginia Tech, Blacksburg, Virginia 24060, USA*
[30]*York University, Toronto, Canada*

(Received 30 March 2026; accepted 9 July 2026; published 29 October 2026)

We present a search for muon neutrino disappearance in the Booster Neutrino Beam (BNB) at Fermilab using the ICARUS detector. Neutrino interactions identified as muon neutrinos interacting with argon nuclei via the charged current interaction and having only a muon and at least one proton in the final state (1μNp) have been selected from data collected in 2022–2023 (ICARUS Run 2) and compared with a simulation-based expectation. In the context of a fit to a two-neutrino approximation of the sterile $3+1$ model, including the impact of systematic uncertainty from the flux, neutrino interaction, and detector models, we find no statistically significant muon neutrino disappearance at the ICARUS baseline of 600 meters from the BNB target. Corresponding 90% CL exclusion contours in the Δm^2_{41}-$\sin^2 2\theta_{\mu\mu}$ space are presented. This is the first oscillation analysis produced by ICARUS exposed to the BNB. We note that the analysis is systematics limited due to large unconstrained uncertainties from the flux and interaction models. In future joint analyses, data from ICARUS and the Short Baseline Near Detector (SBND), exposed to the BNB at 110 meters from target, will be combined to provide significant constraint of these uncertainties, enabling a robust, world-leading two-detector analysis.

DOI: 10.1103/2vms-z2j8

I. INTRODUCTION

While three-flavor neutrino oscillation measurements are incompatible with short-baseline oscillation, a number of anomalous results [1–7] suggest the possibility of neutrino oscillation at a scale of $L/E \sim 1$ km/GeV. Definitive investigation of this possibility is one of the primary motivations of the Short-Baseline Neutrino (SBN) program 8]], consisting of three liquid argon time projection chamber [9] (LArTPC) detectors in the Booster Neutrino Beam at Fermilab. Detectors observing similar flux in the same beamline, having the same target nucleus, and having similar detector designs allows for significant constraint of systematic uncertainty when data from multiple detectors is analyzed together.

The ICARUS LArTPC, located 600 meters downstream of the Booster Neutrino Beam (BNB) target, serves as the far detector of the SBN program and has been taking physics data since 2022. Here, we present a search for muon neutrino disappearance using only the ICARUS detector [10,11], which is an important first step toward a more sensitive two-detector SBN analysis. Our results are consistent with other disappearance analyses [12–17] and demonstrate the quality of ICARUS data and the performance of reconstruction and analysis tools developed by the collaboration. As expected in a single-detector oscillation search, the sensitivity is limited by large *a priori* uncertainties in modeling the neutrino flux and neutrino interactions on argon; the addition of near detector data will provide *in situ* measurements designed to constrain uncertainty in future SBN analyses.

Data are evaluated in the context of the $3+1$ sterile neutrino model, described in Sec. II. For this analysis we consider neutrinos produced in the BNB at Fermilab and observed in the ICARUS detector, which are described in Sec. III. The details of the experimental operating conditions are described in Sec. IV. Simulated data are generated using LArSoft [18] with SBN-specific detector descriptions; the simulation is described in Sec. V.

We present analyses based on two different reconstruction frameworks: Pandora [19] and SPINE [20], which are described in Sec. V and in Appendixes A and B, respectively. These reconstruction and selection parts of the analysis are based on separate code with very different underlying algorithms, but many aspects of the analysis, such as simulation, signal processing, calibration, cosmic rejection, systematic uncertainty, and statistical methods are shared. Observed differences in the performance of the two reconstruction approaches have been used to help improve them and/or to help evaluate systematic uncertainty in the analysis. For these reasons, the results should not be treated as fully independent, but rather considered together as a single result.

*Also at Istituto di Neuroscienze, CNR, Padova, Italy.
†Also at University of Texas, Arlington.
‡Also at Fermi National Accelerator Laboratory.
§On leave of absence from INFN Pavia.
‖Also at INFN Padova and University.
¶Present address IPSI-INAF Torino, Italy.
**Also at Stony Brook University.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

Charged-current (CC) interactions of muon neutrinos on argon nuclei with a final state consisting of one muon and at least one proton, in which all final-state particles are reconstructed as contained within the detector fiducial volume, are considered as signal events. This final state is selected due to its favorable reconstruction characteristics (contained muon and proton tracks can be robustly identified, and their momenta can be estimated from range) and its strong rejection of cosmic background (the presence of a vertex naturally eliminates many cosmic muon tracks). Details of the event selection are described in Sec. VI. The impact of a comprehensive suite of systematic uncertainties, including contributions from flux, interaction, and detector modeling, is evaluated. The systematic variations included in the analysis and studies to confirm appropriate coverage of uncertainties are described in Sec. VII. A new fitting framework, PROfit, has been developed for SBN and is described in Sec. VIII and Appendix C. Results of the analyses are shown in Sec. IX.

The analysis presented here has been performed according to a blind procedure, with the reconstruction tools and selection criteria developed using only 10% of on-beam data and larger samples of non-neutrino data. Data–Monte Carlo (MC) comparisons of neutrino candidates using the full dataset were performed only after the analysis was frozen and reviewed by the collaboration. The fitting framework was developed using only simulation, with no fits performed on data until after the analysis was frozen. The unblinding fit was performed using a staged approach, as described in Appendix C.

II. SHORT-BASELINE OSCILLATION AND STERILE NEUTRINOS

Neutrino oscillation driven by a mass-squared splitting $\Delta m^2_{41} \sim 1\ \mathrm{eV}^2$, in a $3+1$ sterile neutrino extension to the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) formalism, has been proposed as an explanation for anomalous electron-neutrino-like appearance seen by LSND [1] and MiniBooNE [2]. Gallium-based detectors have also reported anomalous electron-neutrino disappearance, which could arise from a similar mass-squared splitting [3–5]. The mass-squared splittings of the PMNS three-neutrino model are known to be far smaller than $1\ \mathrm{eV}^2$, and thus oscillations at this scale would require an additional neutrino mass state in the model. As measurements from the Large Electron-Positron Collider at CERN have ruled out more than three light neutrino states coupling to the Z boson [21], the fourth flavor in the minimally extended $3+1$ model is assumed to not couple to the weak force and is called "sterile." The PMNS mixing matrix is extended to a 4×4 unitary matrix with six real valued mixing angles θ_{ij} with $1 \le i < j \le 4$ [22].

Potential appearance signals suggested by the anomalous results are in tension with other measurements from accelerator neutrinos [23,24]; most recently, results from MicroBooNE [25] and KATRIN [26] exclude at 95% confidence level the single light sterile neutrino interpretation of the LSND and MiniBooNE anomalies and most of the gallium allowed regions, respectively.

A number of other experiments have placed limits in Δm^2_{41} and $\sin^2 2\theta_{\mu\mu}$ space using ν_μ disappearance, as is done in the analysis reported here. At short baselines, the survival probability for muon neutrinos (ν_μ), of energy E_ν, traveling a distance L, is well approximated as

$$P(\nu_\mu \to \nu_\mu) \approx 1 - \sin^2 2\theta_{\mu\mu} \sin^2\left(1.267 \frac{\Delta m^2_{41} L}{E}\right), \quad (1)$$

where the PMNS parameters are expressed in an effective mixing angle

$$\sin^2(2\theta_{\mu\mu}) \equiv 4\cos^2\theta_{14}\sin^2\theta_{24}(1 - \cos^2\theta_{14}\sin^2\theta_{24}). \quad (2)$$

Most experiments have set limits on ν_μ disappearance, while IceCube reports a closed contour at 95% CL. [7]. The existing searches for ν_μ disappearance are summarized alongside the ICARUS result presented here in Sec. IX.

III. BEAM AND DETECTOR

The BNB is generated from protons accelerated to a kinetic energy of 8 GeV impinging a beryllium target. Each 1.6 μs spill of the Booster is nominally composed of $\sim 5\times10^{12}$ protons split into 81 bunches. These bunches are 2 ns wide and approximately 19 ns apart. The collisions from these protons on the BNB target produce a beam of hadrons, primarily pions with a small kaon fraction. The charged hadrons are focused by a magnetic horn into a 50 m-long decay tunnel where the hadrons decay in flight, producing neutrinos. ICARUS is located on axis of the BNB and 600 m from the BNB target. The expected flux by particle type at ICARUS is shown in Fig. 1; the flux at ICARUS is composed primarily of ν_μ (∼93%) with small backgrounds of $\bar{\nu}_\mu$ (∼6%) and $\nu_e/\bar{\nu}_e$ (< 1%). The flux has an average kinetic energy of approximately 0.8 GeV. In addition, ICARUS is exposed to the Main Injector neutrino beam (NuMI) at a 5.75° far-off-axis position, 800 m downstream of the NuMI target.

The primary ICARUS detector is a LArTPC containing a total of 760 tons of ultrapure liquid argon. ICARUS consists of two identical adjacent cryogenic modules ("east" and "west" cryostats), with each module having two TPCs separated by a common cathode. The maximum drift length is 1.5 m, corresponding to ∼1 ms at the nominal drift field of 500 V/cm. The anode structure has three readout wire planes with 3 mm pitch, spaced 3 mm apart, with wires oriented at 0° (Induction 1), +60° (Induction 2), and −60° (Collection) with respect to horizontal. With appropriate voltage biasing, the first two planes (Induction 1 and Induction 2) provide nondestructive

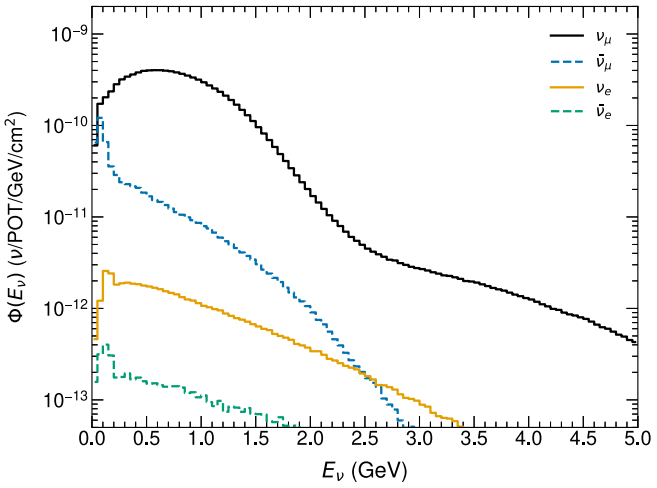


FIG. 1. Predicted BNB flux by neutrino species at ICARUS. The flux simulation was developed by MiniBooNE [27] and MicroBooNE [28] and is based on Geant4 [29], with corrections applied to the pion production cross section using a Sanford-Wang parametrization fitted to HARP [30] and E910 [31] data. Parent hadron decays to neutrinos are sampled over the front face of the ICARUS detector, where the flux is predicted to vary by less than 4%.

charge measurements, while the ionization charge is fully collected on the final (Collection) plane. Each TPC identifies neutrino interactions through the detection of ionization charge deposited in tracks and electromagnetic showers by charged particles produced in neutrino-argon interactions. The ionization charge is used to reconstruct charged particle trajectories with good calorimetry and precise (∼mm) spatial resolution. More details are available in [10,11]. Scintillation light is detected by 360 8" photomultiplier tubes (PMTs) installed behind the TPC wire planes and used for event selection and triggering [32]. A cosmic ray tagging (CRT) system [33] surrounds the cryostat with two layers of wavelength-shifting fiber embedded plastic scintillators and concrete overburden of ∼3 meters is installed above the detector to mitigate background from cosmic rays. The ICARUS trigger system [34] issues a global trigger that opens data-acquisition windows sized to capture the full charge drift and all associated scintillation light. The primary BNB "on-beam" trigger relies on coincident PMT activity inside the beam gate timed to the corresponding 1.6 μs long proton-extraction spills. Analogous "off-beam" triggers collect cosmic-ray events using gates shifted 33 ms after each beam gate. Complementary "minimum-bias" triggers are independent of scintillation activity, so provide unbiased data samples at reduced rates to support calibration and performance studies.

IV. DATA

The analysis presented here uses data collected with the BNB of Fermilab from winter 2022 to spring 2023 (ICARUS Run 2), corresponding to an exposure of $\sim 2.05 \times 10^{20}$ proton on target (POT). Data with nominal detector and beam conditions, and thus suitable for analysis, are identified using recorded run history and detector conditions, data quality metrics, and beam quality metrics.

Good runs are defined as those in which the DAQ was properly working, all detector components (TPCs, PMTs, and top and side CRT) were present in the collected data, the run had a minimum duration of 1 hour, and no other detector or DAQ anomalies were recorded that could spoil the quality of the collected data. Data quality metrics are evaluated for each run compared to the global mean for the dataset, removing runs which have a metric more than 3σ away from the global mean. The adopted metrics are chosen to reflect low-level activity in the detector, such as the average number of hits per track and the average total photoelectrons (PEs) per scintillation light burst, as well as higher-level reconstructed quantities such as the number of distinct interactions per event. Additionally, checks are placed on the wire bias voltage and cathode voltages to ensure they were within nominal ranges. The deviation from the mean for all quality metrics for each run can be seen in Fig. 2.

Good spills within good runs are identified using beam quality metrics adopted from MicroBooNE that are based on information recorded by beamline monitoring devices in the BNB. The requirements are nonzero beam intensity, nominal horn current, and a "good" value for a metric that quantifies the centering of the spill on the beam target.

After data and beam quality cuts, the total POT used for the analysis is $\sim 1.596 \times 10^{20}$ POT. The reduction of available exposure due to the data and beam quality is shown in Table I.

The POT corresponding to the analyzed datasets is extracted using spill and POT information recorded for

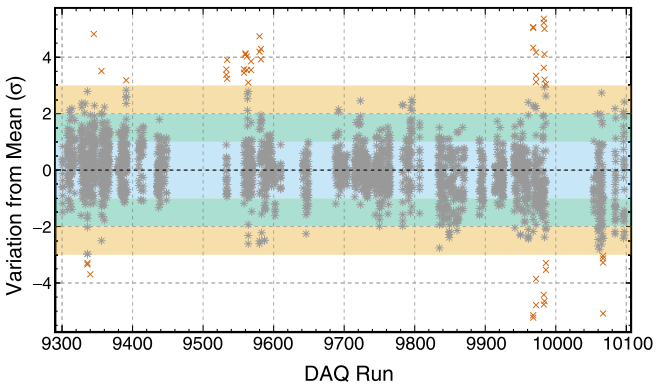


FIG. 2. Deviation of the data quality metrics from their respective means for each run that passed prefiltering. The dashed line represents no deviation from the global mean of that quantity. The blue, green, and yellow bands represent the regions 1, 2, and 3 standard deviations from the mean, respectively. Any metric marked in red falls outside out the 3σ region and the corresponding runs are flagged as poor data quality.

TABLE I. Effect of the data and beam quality conditions on the POT and spills available.

	POT		Spills	
	Tot (10^{20})	Frac	Tot (10^{7})	Frac
Total analyzed	1.957	1.000	5.232	1.000
Good runs	1.752	0.895	4.630	0.885
Beam quality	1.783	0.911	4.491	0.858
Beam quality + good runs	1.596	0.816	4.014	0.767

each triggered event and validated by comparing triggering-flash time distributions in on- and off-beam data normalized to equal exposure, as shown in the top panel of Fig. 3. A clear excess from BNB interactions is observed in on-beam triggers during the 1.6 μs neutrino beam spill, while identical cosmic-ray rates are seen outside the spill window. The normalized on- and off-beam cosmic contributions agree within statistical uncertainties throughout Run 2 data taking, as seen in the bottom panel of Fig. 3, differing only by $(0.5 \pm 0.3)\%$ when averaged over the whole Run 2.

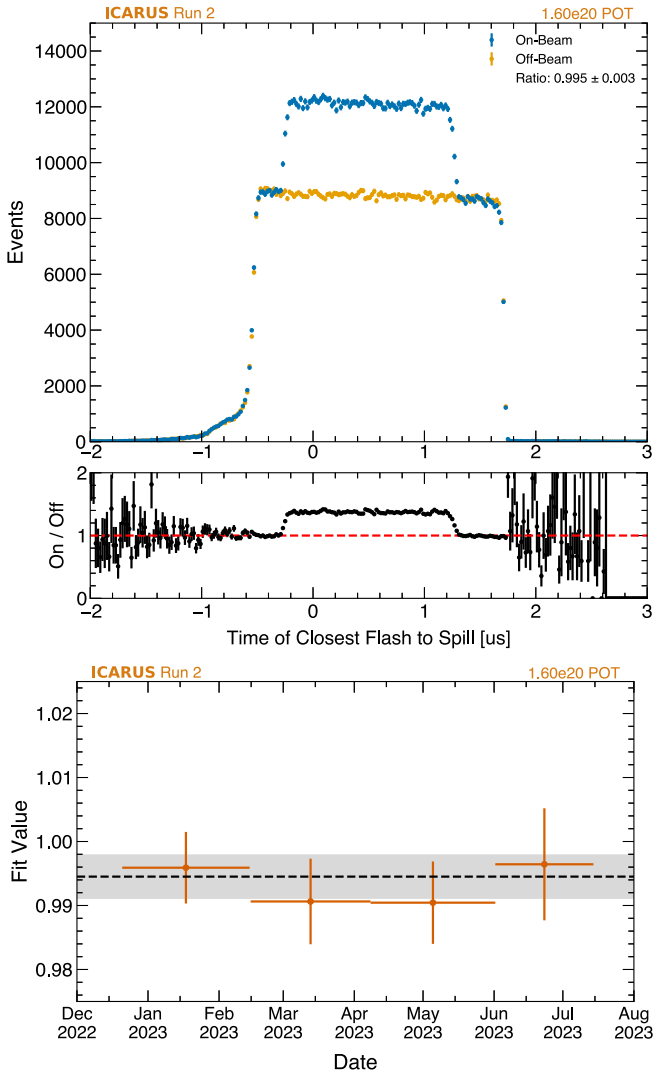


FIG. 3. (Top) Triggering flash time distribution for on-beam (blue) and off-beam (orange) Run 2 majority triggers (top), normalized to the same exposure, and the corresponding on-beam/off-beam ratio (bottom). A clear beam induced excess is observed in the $[-.3, 1.3]$ μs beam spill window. The on-beam/off-beam ratio in the region $[1.3, 1.6]$ μs is calculated and found to be consistent with 1.0 as expected. (Bottom) On-/off-beam ratio in the region $[1.3, 1.6]$ μs of the top panel for four two-month periods, with statistical uncertainties. The dotted line and red band indicate the fitted ratio and 1σ error for the full period.

V. SIMULATION AND RECONSTRUCTION

A. Simulation

A MC simulation of neutrino interactions coming from the BNB, corresponding to 1.74×10^{21} POT, i.e., ten times larger than the analyzed data, is produced. MC events include a BNB neutrino interaction in the liquid argon in the 1.6 μs of the BNB spill and cosmic ray interactions crossing the detector in a ±1 ms time window around the BNB spill. The neutrino interactions are simulated using the GENIE event generator [35,36]; the `AR23_20i_00_000` comprehensive model configuration, which is developed within the context of DUNE [37], is chosen as a reference model. Additional details of the interaction model and its uncertainties are discussed in Sec. VII B. Cosmic rays are modeled by CORSIKA [38]. MC events are processed, reconstructed, and selected using the same *ICARUSCODE* software used for data, based on the LArSoft framework [18], which is used in all LArTPC experiments.

The central value (CV) simulation includes a modeling of the response of all subdetectors (TPC, PMTs, CRT) and the corresponding generation of signals and noise. Many known effects impacting the detector response have been modeled and calibrated with a data-driven approach, as described in more detail in [39]. The effect of unmodeled or mismodeled effects are included as systematic uncertainties, as described in Sec. VII C. The main detector aspects addressed in the simulation are:

(i) TPC gain and signal shape: The wire signal gain, i.e., the signal amplitude corresponding to a known number of drift electrons, is computed using cosmic muons as described in [39]. Variations of around 1% among the average gain of different TPCs are observed and are included in the simulation.

(ii) TPC noise: The spectra and rms of both coherent (correlated across a DAQ board) and intrinsic noise components are simulated independently for each board, based on the corresponding noise observed in data.

(iii) Electron lifetime: The average values observed in data during Run 2 are 4(8) ms for the East(West) cryostat, while the simulation assumes a conservative value of 3 ms, identical for both cryostats.

(iv) Recombination model: Several models have been proposed in literature to parametrize the probability of drift electron recombination as a function of both electric field and dE/dx. The CV assumes the Modified Box model [40].
(v) Space charge: The accumulation of ions from liquid argon ionization, mainly induced by cosmic rays, generates a distortion of the TPC electric field and consequently of the reconstructed coordinates, extending up to a few millimeters for a detector on surface [41]. The simulation includes this effect.
(vi) Scintillation light modeling: A quantum efficiency (QE) of 7.3% is used in the simulation of light collected by the PMTs; this value is selected to match the average light signal amplitude between data and MC.

B. Raw signal processing and calibration

The processing of the recorded TPC wires signals follows the logic described in more detail in [11], based on the "deconvolution" of the electrostatic field and electronic response from the wire signal waveform, ideally recovering the original current of drift electrons in the TPC. Deconvolution is applied wire by wire, after removing coherent noise across the 64 channels of each front-end board, using the expected field and electronics response functions [39].

The field response, describing the drift of electrons in the vicinity of the TPC wires and the generation of signals via electrical induction or charge collection, has been tuned in a data-driven way, minimizing the difference between predicted and observed signals, as illustrated in [39]. The time intervals containing physical signals ("hits") are searched for in each wire waveform, based on a threshold algorithm. Hits are then fitted with a Gaussian, whose area is proportional to the number of drift electrons generating the signal.

The calibration of TPC wire signals converts the area of signal waveforms into an accurate measurement of the energy deposited on each wire by a physical track, through the ionization of liquid argon and electron drift [39]. This requires a quantification of detector effects impacting signal generation, such as electron recombination, attenuation of the drift electron current, diffusion, and possible nonuniformities of the TPC calorimetric response throughout the TPC volume, as described in [42] and [42]. Calibrated hits provide the local energy deposition for a track segment, typically a few millimeters in length, as measured on each wire, along with the corresponding 2D localization from the wire coordinate and drift time. Overall, the efficiency for detecting a wire signal and correctly associating it with the originating track is greater than 90% in the induction planes and exceeds 98% in the collection plane.

Each PMT waveform is processed with a threshold-based algorithm to identify fired PMTs and reconstruct the detected light. All signals above threshold within 100 ns are then clustered into a "flash," which nominally corresponds to the light produced by a single interaction, whether from cosmic rays or neutrinos, with a few ns time resolution.

CRT hits—spacetime points associated with tracks crossing the CRT—are formed by time-ordering CRT signals above the thresholds in both adjacent layers and grouping them by region, with several ns and cm resolution. CRT timing is cross-calibrated with the PMT system and is validated by studying the CRT-PMT time of flight measured with incoming cosmic muons [33].

C. Track, shower and event reconstruction

The reconstruction chain uses the calibrated TPC hits, PMT-based hits and flashes, and CRT hits from the raw-signal processing stage. Low thresholds for defining hits are chosen to maintain high hit efficiency but introduce noise-induced false signals; these are mitigated by forming 3D spacepoints from combinations of 2D hits with intersecting wires and consistent timing. Starting from these objects, the pattern recognition stage identifies the tracks and showers present in each event, hierarchically grouped into separate interactions and associated to a primary interaction vertex, which should ideally correspond to possible neutrino interactions or individual cosmic rays in the detector. In this reconstruction step, the physical properties associated to each track/shower and to the reconstructed interaction are also determined, which serve as inputs to the event selection described in the following section. The Pandora and SPINE reconstruction algorithms start from the same input, but adopt different strategies to resolve 3D objects and identify and reconstruct the interactions in the detector. Pandora first reconstructs separate 2D clusters in the different wire planes, which are then combined into 3D objects as described in Appendix A, while SPINE considers all possible combinations of 2D hits as described in Appendix B.

Pandora is a pattern recognition software package widely used in LAr-based detectors, especially LArTPCs (e.g., [43,44]). It provides a fully automated, end-to-end reconstruction of events, generating a hierarchical set of tracks and electromagnetic showers to represent final-state particles and any subsequent interactions or decays. Pandora adopts a multialgorithm approach, in which complex pattern recognition tasks are decomposed into specialized subtasks. Each algorithm operates independently within the processing chain and addresses a specific aspect in a given topology—for example, interaction vertex identification or shower 3D reconstruction. The collective output of this ensemble of algorithms progressively builds up a global representation of the underlying events, providing a robust and comprehensive reconstruction. All the 3D objects deemed to be associated with a single interaction in the detector—either a neutrino interaction or a cosmic ray—are

grouped together in a so called *slice*. A schematic of the reconstruction flow is shown in Fig. 29. More details about its application to ICARUS are present in Appendix A.

The SPINE (Scalable Particle Imaging with Neural Embeddings) [20] reconstruction is an end-to-end, machine learning (ML) based reconstruction chain. It consists of a hierarchical set of neural networks that can be trained as a cohesive whole. Two types of neural networks are employed: convolutional neural networks (CNNs) and graph neural networks (GNNs). The CNNs are used to extract point-level information from voxelized 3D images of the detector activity, while the GNNs are used to aggregate this point-level information and extract higher-level features at the particle- and interaction level. The neural networks in the SPINE reconstruction chain are trained using supervised learning techniques on a large set of simulated events independent from those used in the analysis. This training set uses generic, physics-agnostic event generators to produce neutrinolike and cosmic raylike interactions in order to cover a wide range of possible event topologies and avoid biases toward specific interaction models. The simulated events are processed through the full ICARUS simulation to ensure that the training data accurately reflects the detector response and noise characteristics. More details of the SPINE framework are provided in Appendix B.

For both reconstruction paradigms, the energy of each particle is estimated through its range, where the distance traveled by the particle is related to its momentum or kinetic energy via the continuous slowing down approximation [45]. The energy of the neutrino is reconstructed by adding the total energy of the muon (E_μ), the kinetic energy of the proton(s) (KE_p), and an empirical average removal energy of 30.9 MeV [46] for each proton above 50 MeV (RE_p),

$$E_\nu = E_\mu + KE_p + RE_p. \tag{3}$$

This neutrino energy estimator is the same in both reconstruction paradigms and is sometimes referred to as the visible energy (E_{vis}).

VI. EVENT SELECTION

A. Selection criteria

A 1μNp final-state topology is chosen to simplify reconstruction and minimize cosmic contributions. The target topology has a charged-current vertex in the fiducial volume with exactly one contained muon longer than 50 cm in length, at least one proton with deposited energy > 50 MeV, and no additional charged particles or photons. All reconstructed charged particles must be contained within the active volume, to allow for range-based measurements of particle momenta. The muon length requirement is intended to reduce the contamination from NC interactions with a pion in the final state. The proton threshold ensures that the track is long enough to be reconstructed with good efficiency and sufficiently energetic to be well separated from the muon in the final state. A veto threshold is applied to extra particles to ensure that the particles are sufficiently energetic to be visible in the final state if they are indeed present. Example event displays illustrating the target final state are shown in Fig. 4.

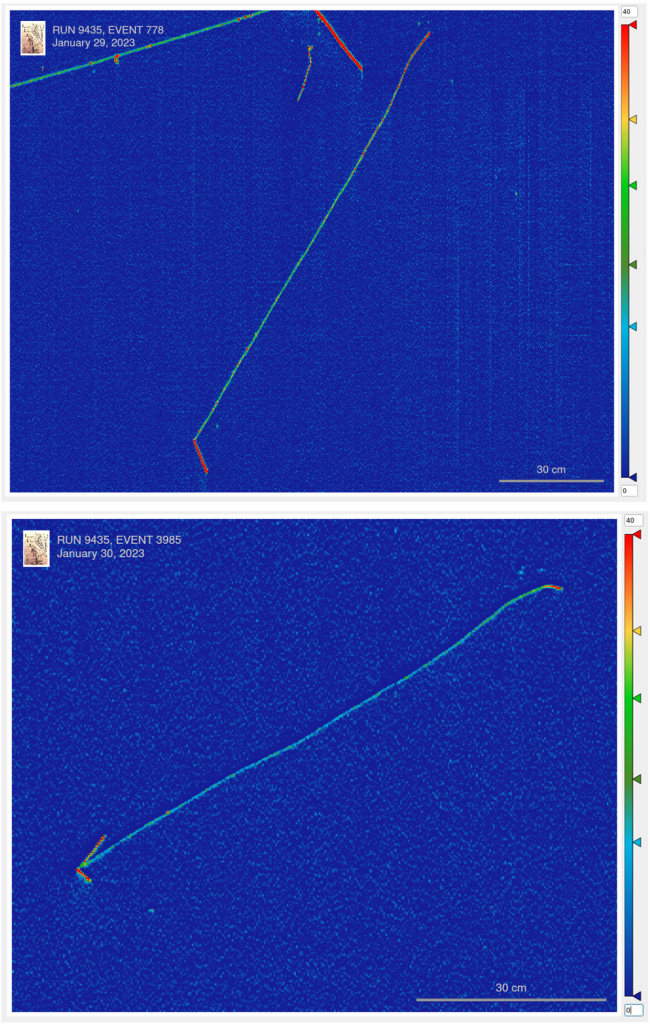


FIG. 4. Examples of 1μ1p (top) and 1μ2p (bottom) candidate events observed in Run 2 data. In the upper part of the top figure overlapping cosmic activity is also visible. The color scale indicates the energy deposition of each particle.

Event selection criteria are applied to the reconstructed events with the aim of selecting 1μNp candidates while rejecting other neutrino interactions as well as cosmic rays associated events. Some selection criteria are common between the Pandora and SPINE-based analyses while others differ due to the intrinsic nature of both algorithms. The common criteria applied are:

(i) Fiducial volume cut: requires the reconstructed event vertex to be more than 25 cm from the edge of the active volume in the horizontal (x) and vertical (y) coordinates, and more than 30 (50) cm from the

upstream (downstream) edge of the active volume in the z coordinate (along the beam axis). To remove portions of the volume in which the drift field is distorted by a hanging cable, an additional cut is applied, removing ∼0.3% of the liquid argon volume.

(ii) Containment: all particles contained within the reconstructed interactions are required to be more than 5 cm from the edge of the active volume in all directions.

(iii) Single muon in the final state: presence of a single primary muon with a length of at least 50 cm.

(iv) Proton(s) in the final state: presence of at least one primary proton depositing at least 50 MeV of energy.

(v) Pion/shower veto: the final state may not contain any charged pions or showers (electrons/photons). These particles are required to be primary particles and to have deposited at least 25 MeV of energy to be counted as final-state particles.

In Pandora, tracks and showers are identified by a boosted decision tree (BDT) [47] classifier. Particle identification (PID) variables for tracks are calculated by comparing the measured energy deposition to a MC-based template of dE/dx vs residual range for a given particle type. While perfect data/MC agreement is not achieved in the particle identification classifier distributions, the applied cuts are loose and implemented in a region of the distribution that is well modeled.

The selection based on the Pandora reconstruction includes several additional cuts for rejection of cosmic ray background. The interaction must not be tagged as "clear cosmic" by Pandora as described in Appendix A. Events are rejected in which PMT and CRT signals are in time with each other and with the beam spill, indicating a particle entering or exiting the cryostat. Compatibility between the event reconstructed in the TPC and the triggering flash in the PMT system is checked by comparing the barycenter along the z direction of the light associated with the triggering flash and the TPC charge barycenter of the neutrino candidate, also in the z axis. The event is selected if the separation between the two barycenters is less than 100 cm.

In SPINE, the selected interaction is required to be matched with an optical flash detected by the PMTs that is consistent with the time of the neutrino beam spill. This requirement provides significant background rejection of cosmic-induced interactions that are not in time with the beam, but is not expected to provide significant background rejection for cosmic-induced interactions that are in time with the beam and that trigger the detector readout. The latter might instead be targeted using the CRT system; however, given that contained events are selected in this analysis, there is very little background from in-time cosmics and so a CRT veto is not applied.

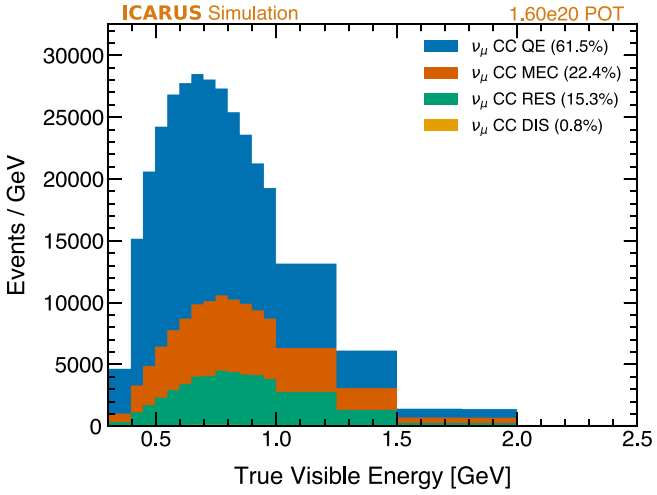


FIG. 5. True 1μNp visible energy spectrum selected by truth criteria. Visible energy is defined in Eq. (3). The true interaction type is indicated by color. The rate is scaled to the exposure of Run 2, 1.6×10^{20} POT.

B. Efficiency, purity and background estimation

The selection efficiency, purity, and background levels are estimated primarily using MC simulation. Figure 5 shows the defined 1 μNp signal spectrum, with contributions from each true interaction type. Many of the true 1 μNp events are from quasielastic interactions, but there is a significant contribution from meson exchange current and resonant pion production.

Selection efficiency is computed using MC simulation as the ratio between the number of selected events and true number of events in the target sample. Similarly, the purity is defined as selected events truly in the target sample divided by all selected candidates. As a result, some events classified as "background" are true ν_μ-CC interactions that have a different true final-state topology from the target sample. This choice does not impact the results of the analysis as the quoted efficiency and purity numbers are used only for informational purposes and not in extraction of physics results. Additionally, the true number of events in the target sample is calculated using truth information; for example, an event is determined to be truly in the fiducial volume based on the *true* vertex position, while the fiducial volume cut is applied using the *reconstructed* vertex position. Therefore, the same cut definition will behave differently in the selections using different reconstruction paradigms.

The Pandora selection achieves a 48% efficiency and 82% purity and the SPINE selection achieves 77% efficiency and 92% purity, as evaluated with MC simulation. In both selections, the primary background is from misclassified ν_μ charged-current interactions with different final-state topology from the target sample, while only 3.8% and 1.1% of the selected sample comes from NC interactions, for the Pandora and SPINE selections, respectively.

TABLE II. Summary of the cumulative efficiency values for Pandora and SPINE $1\mu Np$ selection on a cut-by-cut basis.

Selection cut–efficiency	Pandora (%)	SPINE (%)
No cut	100	100
Vertex in FV	96.8	99.4
Containment	92.4	98.9
Cosmic rejection	88.3	96.0
Muon ID	76.9	87.5
Proton ID	55.2	80.6
No pions	50.7	78.2
No showers	48.5	77.0

The CRT-PMT cosmic rejection algorithm relies only on timing information and only isolated flashes are selected, so the main inefficiency for signal comes from random CRT-PMT coincidences within the 200 ns window. In order to evaluate this contribution, a data-driven analysis is performed, showing an inefficiency of $\leq 1.0\%$. This result is cross-checked with a visual study of identified events without the application of the CRT-PMT-based cut, where compatible results are obtained.

The contribution of in-time cosmics is measured directly from data, by applying the event selection to the off-beam dataset. The in-time contribution is found to be $\sim 0.5\%$ for Pandora and $< 0.1\%$ for SPINE. The contribution from out-of-time cosmics in coincidence with neutrino events is estimated by MC and corresponds to only $\sim 0.5\%$ of the total selected events in Pandora and $< 0.1\%$ for SPINE, comparable to that from in-time cosmics. Thus the overall cosmic background accounts for only $\sim 1\%$ and $< 0.1\%$ of the total selected events for Pandora and SPINE, respectively. The cumulative efficiency and purity of the selections are shown in Tables II and III, respectively, for each analysis.

Figure 6 shows the efficiency for each of the selection criteria with respect to the above true signal definition, applied cumulatively, as a function of true neutrino energy and Fig. 7 shows the composition of the selected samples from simulation. For Pandora, the largest source of inefficiency is the proton identification and is dominated by misreconstruction of a few cm long proton tracks. The muon identification inefficiency is primarily from misreconstruction of the muon track, while the pion rejection inefficiency has a number of different sources. Note that the lower efficiency regions at low and high neutrino energy correspond to regions with low signal statistics; as such they are statistically limited even in the MC samples, as seen in the top plot of Fig. 7. For SPINE, the selection efficiency falls off at lower muon/proton kinetic energies as expected, manifesting as a dropoff in efficiency at lower total neutrino energy, as seen in the bottom plot of Fig. 7. For both selected samples, the primary contribution outside the target sample is a collection of charged-current muon neutrino interactions that either contain charged pions in the final state or are actually noncontained events reconstructed as being contained.

TABLE III. Summary of the cumulative purity values for Pandora and SPINE $1\mu Np$ selection on a cut-by-cut basis.

Selection cut—purity	Pandora (%)	SPINE (%)
No cut	< 0.1	< 0.1
Vertex in FV	0.8	0.8
Containment	1.2	6.5
Cosmic rejection	24.9	26.1
Muon ID	45.2	52.4
Proton ID	65.1	68.1
No pions	71.4	80.1
No showers	81.5	91.6

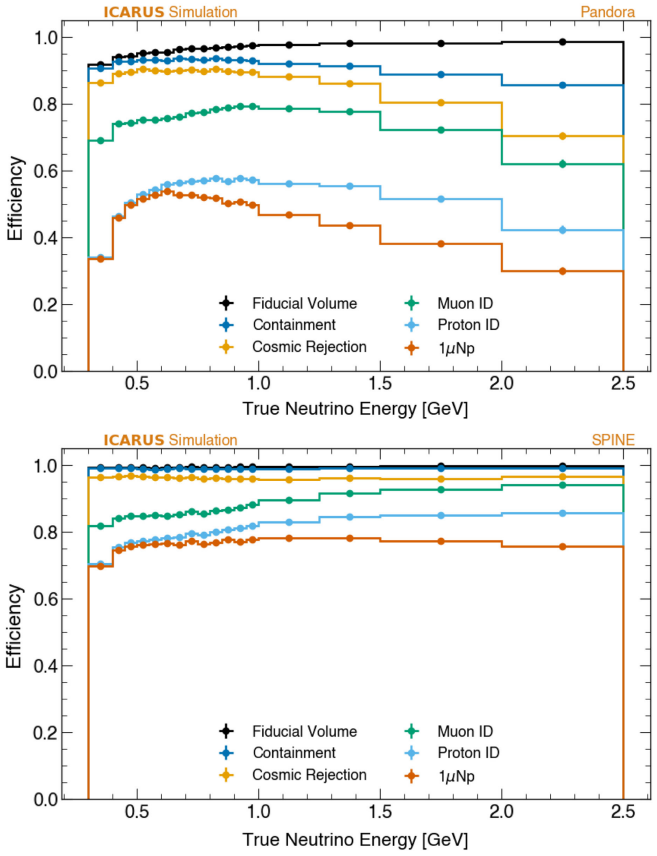


FIG. 6. Cumulative efficiency of selection criteria for Pandora (top) and SPINE (bottom) as a function of true neutrino energy applied in the order appearing in the plot, with respect to true $1\mu Np$ signal, as defined in the text. Statistical uncertainties are included but in most cases are too small to be visible in the plots.

As shown in Fig. 8, the neutrino energy is well reconstructed for signal events and, as expected, the energy resolution is quite poor for neutrino events with a misreconstructed final-state topology. The "low-side" tail of the resolution for signal events is primarily from misreconstruction of particle lengths and thus momenta. In Pandora this is primarily due to splitting of tracks, in which a single true

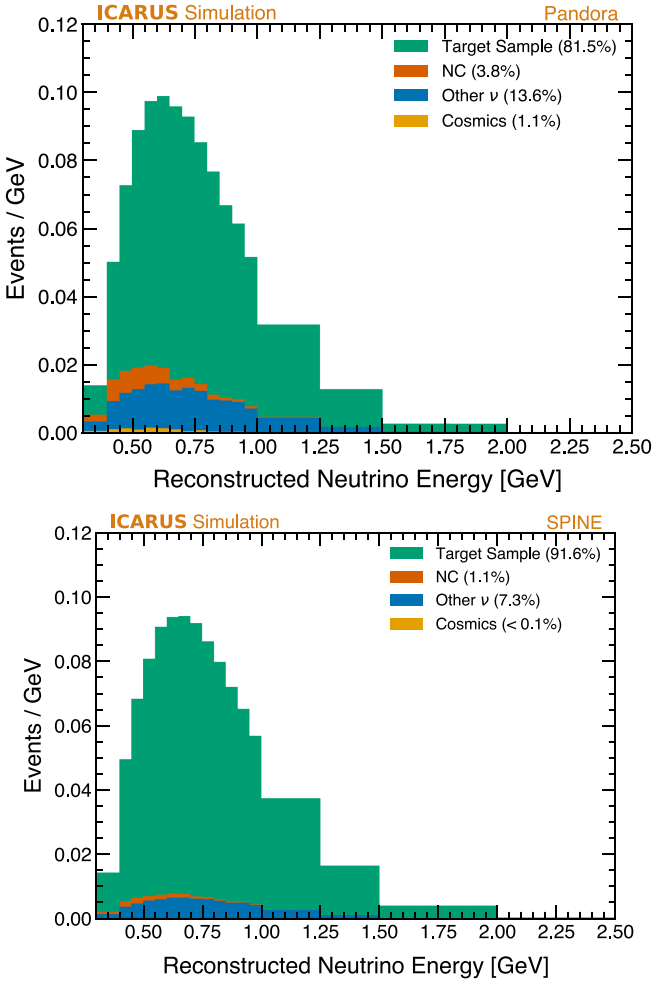


FIG. 7. Selected spectrum from simulation sample highlighting the different final-state topologies for Pandora (top) and SPINE (bottom). The "other ν" category is dominated by ν_μ CC events outside the target sample.

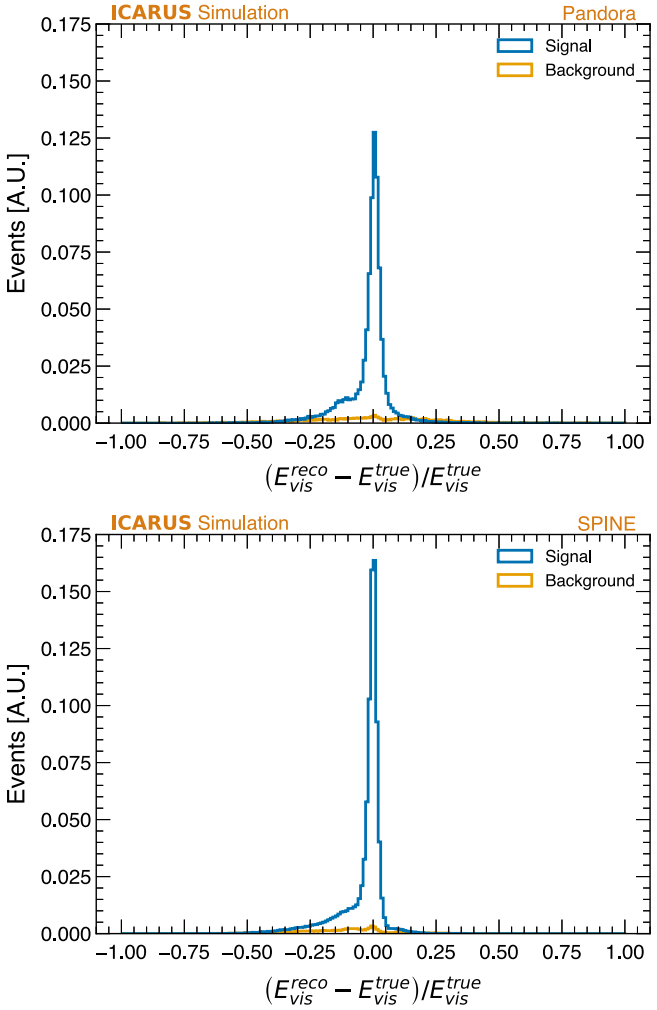


FIG. 8. Energy resolution for the selected $1\mu Np$ sample, separately for true signal and neutrino background contributions. Visible energy is defined in Eq. (3). Reconstructed visible energy is compared to true visible energy, which is calculated using the same definition with true particle momenta, for both Pandora (top) and SPINE (bottom). Comparison to true neutrino energy does not result in any significant shift of the peak but does somewhat broaden the tails as one would expect.

track is reconstructed as two separate tracks, leading to underestimate of the muon track length. Protons that are reconstructed as a separate particle after a scattering event are also expected to contribute in both SPINE and Pandora.

C. Selected events

A total of 13,914 and 19,124 events are selected from Run 2 for the Pandora and SPINE analyses, respectively. There is significant overlap between the selected samples; about two-thirds of the Pandora-selected events are also selected by SPINE, while, due largely to the differing efficiencies, only about half of the SPINE-selected events are also selected by Pandora. Figure 9 shows the reconstructed neutrino energy distribution for selected events in data and MC, for the Pandora and SPINE analyses. Figures 10, 11, and 12 show data-MC comparisons of the reconstructed transverse momentum difference between the muon and leading proton (δp_T [48]), selected muon momentum, and leading proton momentum for both analyses. The error bars on the data points represent the statistical uncertainty and the shaded band represents the 1σ range of the central value MC prediction, based on the systematics uncertainties described in Sec. VII. Small contributions to the prediction from misidentified ν_e-CC, neutral current, and cosmic events are identified. As described in the previous Section, the rate of selected cosmic events in time with the beam spill is measured using data collected in off-beam triggers. The other background rates are predicted using MC truth information.

As seen in Figs. 9–12, the reconstruction of physical quantities shows good performance for signal events. The largest data-MC discrepancy observed, in the SPINE δ_{pT} distribution in Fig. 10, has a prefit χ^2/nbins of around 2,

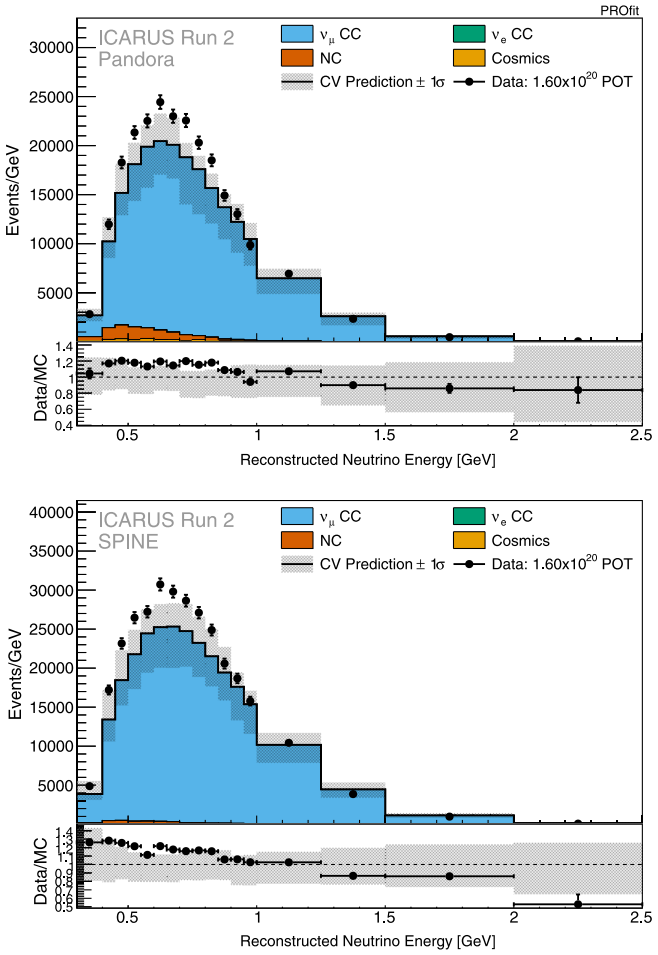


FIG. 9. Reconstructed neutrino energy distributions and data/MC ratios for selected events for Pandora (top) and SPINE (bottom). The ν_μ CC category includes both events in the target sample and misreconstructed ν_μ CC events.

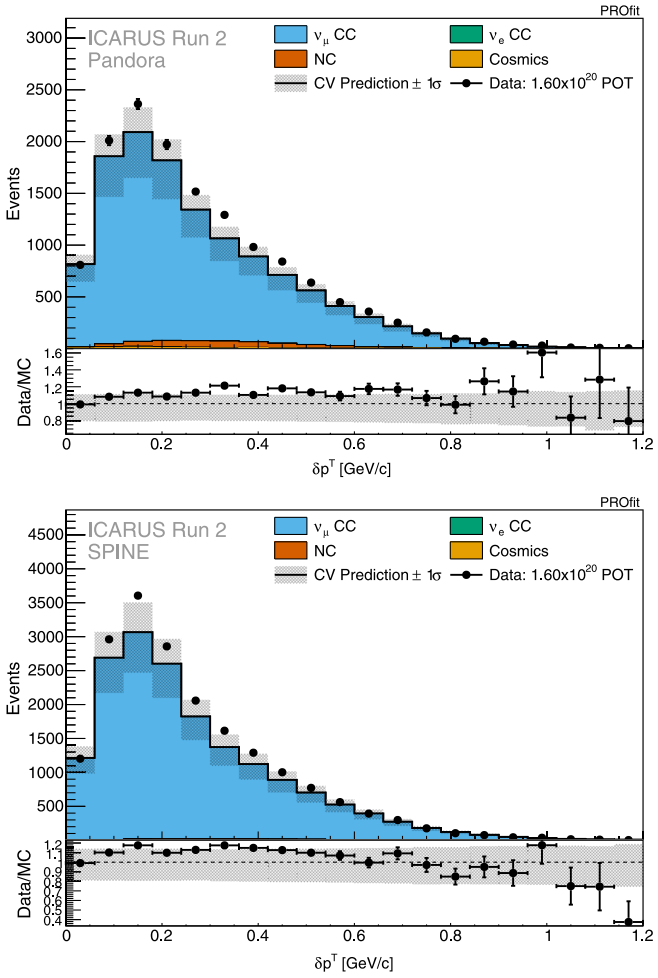


FIG. 10. Reconstructed δp_T distributions and data/MC ratios for selected events for Pandora (top) and SPINE (bottom). The ν_μ CC category includes both events in the target sample and misreconstructed ν_μ CC events.

with most distributions having χ^2/nbins near 1. In general, the data points are above the central value (CV) MC prediction; this tendency of the CV MC to underpredict the number of events is expected based on comparisons of the interaction model used in this analysis to other neutrino datasets [25,49]. Most of this known data-MC discrepancy is covered by the systematic uncertainties included in the analyses and the shapes of the distributions are generally in good agreement with simulation.

D. Validation and Pandora-SPINE comparison

A visual study campaign performed on data provides a rough evaluation of the selection purity and efficiency of both reconstruction paradigms and allows a comparison with the simulation-based efficiency and purity calculations described above. Selected events for a small sample of data (∼200 events) are visually classified as signal or background, with purity defined as the ratio between the number of visually identified $1\mu\mathrm{N}p$ candidates and the total number of selected events. The efficiency is defined as the fraction of events selected by each algorithm within a unbiased reference sample of visually selected events in the target sample. To evaluate the performance of the visual study, simulated events are also visually analyzed, and the visual classification is compared with the MC truth information. This study shows that the visual method slightly overestimates the purity, while providing a good prediction of the efficiency.

Using the visual scan method, efficiency and purity are found to be 51% and 84% for the Pandora-based selection, and 70% and 94% for the SPINE event selection. These values are in agreement with the efficiencies determined from MC, considering the small overestimation in purity and the reduced statistics available in the visual analysis. In this small subset of events, each selection accepts around half of the events identified by the other; the overlapping events are dominated by genuine $1\mu\mathrm{N}p$ events. Pandora is observed to recover some good $1\mu\mathrm{N}p$ candidates with short muon length not found by SPINE. However, Pandora is observed to have

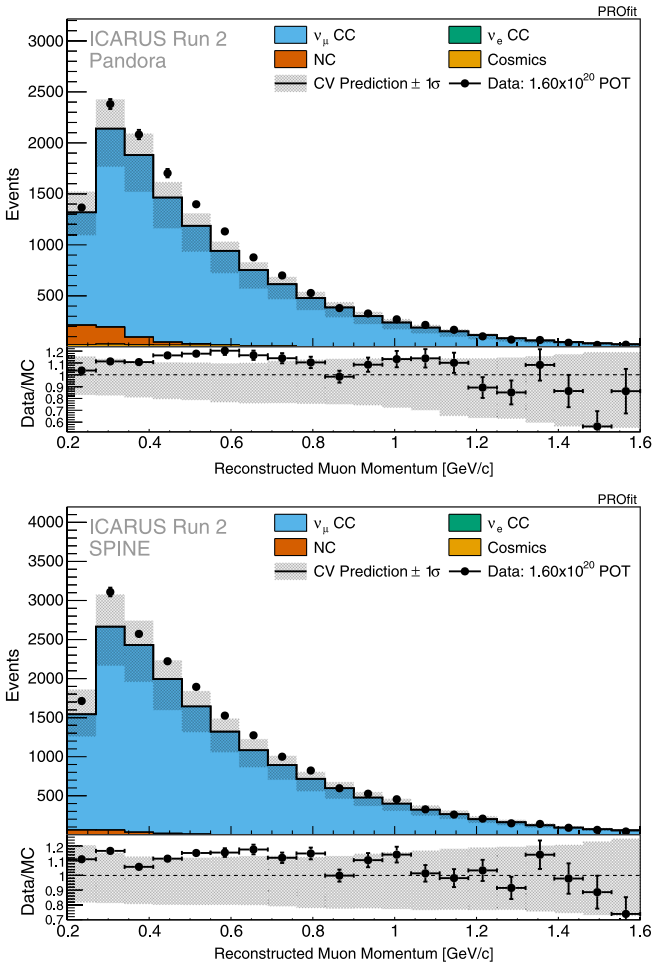


FIG. 11. Reconstructed muon momentum distributions and data/MC ratios for selected events for Pandora (top) and SPINE (bottom). The ν_μ CC category includes both events in the target sample and misreconstructed ν_μ CC events.

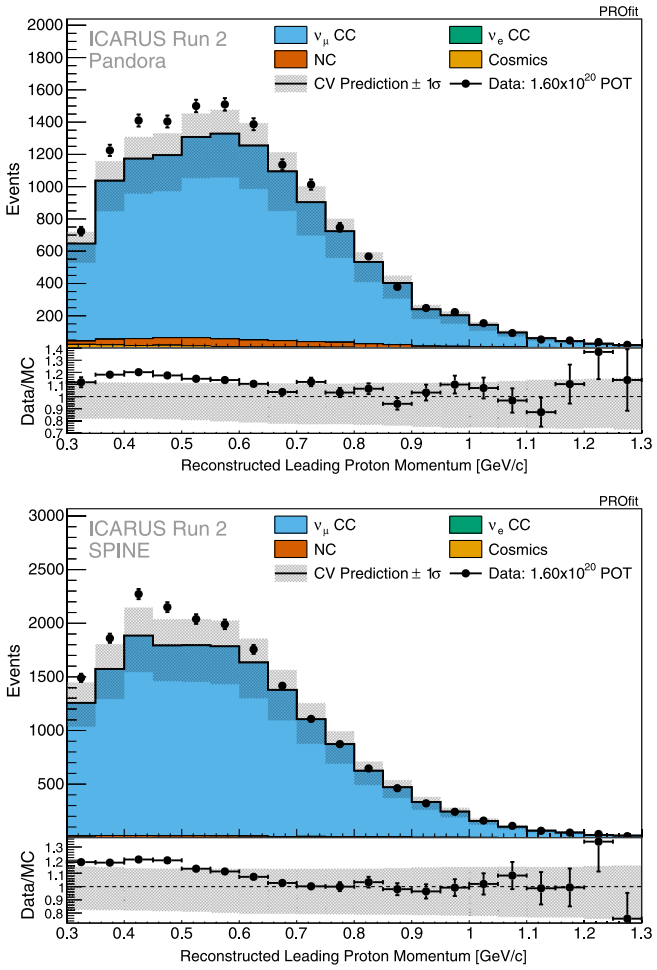


FIG. 12. Reconstructed leading proton momentum distributions and data/MC ratios for selected events for Pandora (top) and SPINE (bottom). The ν_μ CC category includes both events in the target sample and misreconstructed ν_μ CC events.

reduced identification efficiency relative to SPINE for events with low energy protons, and for tracks that are along the drift direction or parallel to the readout plane.

Additionally, data-data comparisons are made to check for any unexpected systematic effect as a function of detector position, operating conditions, and time within the run. Ratios of data with a vertex in the east vs west cryostat and under various operating conditions such as whether or not the NuMI beam was operating are inspected. For cases where differences are expected, double ratios of SPINE to Pandora data are compared, for example, events with upstream vs downstream vertices or with vertices near or far from the cathode. Event rate per POT as a function of run number is also inspected. No significant deviations from expectation are observed in any of these ratios.

VII. SYSTEMATIC UNCERTAINTY

Potential neutrino oscillations between the neutrino production point and the ICARUS detector are studied by comparing the observed spectrum at ICARUS to a prediction based primarily on MC simulation. Uncertainties in modeling the neutrino beam, neutrino interactions, and the detector lead directly to uncertainties in the predicted spectrum and thus the measured values of the oscillation parameters and so are included as systematic uncertainties in the result presented here. While the variations considered are the same for both the Pandora- and SPINE-based analyses, the impact on the predicted spectrum is expected to be different between two analyses as a result of differences in the reconstruction and selection efficiencies which lead to differing fractions of underlying event types in the selected samples. Details of the systematic uncertainties considered for the flux model, interaction model, and detector model are described in the following Subsections.

A. Flux systematics

The neutrino flux in the Booster Neutrino Beam has been studied extensively by previous experiments, including

TABLE IV. Uncertainties on the flux. The three general categories are beam focusing uncertainties, uncertainties on hadronic secondary interactions in the target (where σ^N are for secondary nucleon interactions and σ^π are for secondary pion interactions), and uncertainties from hadron production. The Be and Al secondary interaction cross sections are varied together such that there are six total secondary interaction parameters in the fit. In many rows, two values (e.g. the cross-section uncertainties for Be and Al) are provided separated by a comma.

Parameter	Prefit Unc
Horn current depth	< 18%
Horn current	±6%
$\sigma^N_{\text{inelastic}}$ (Be, Al)	±5%, ±10%
σ^N_{QE} (Be, Al)	±20%, ±45%
σ^N_{Total} (Be, Al)	±15%, ±25%
$\sigma^\pi_{\text{inelastic}}$ (Be, Al)	±10%, ±20%
σ^π_{QE} (Be, Al)	±11.2%, ±25.9%
$\sigma^\pi_{\text{Total}}$ (Be, Al)	±11.9%, ±28.7%
π^+ production (ν_μ, $\bar{\nu}_\mu$)	±11.7%, ±1.0%
π^- production (ν_μ, $\bar{\nu}_\mu$)	0%, 11.6%
K^+ production (ν_μ, $\bar{\nu}_\mu$)	0.2%, 0.1%
K^- production (ν_μ, $\bar{\nu}_\mu$)	0%, 0.4%
K^0 production (ν_μ, $\bar{\nu}_\mu$)	0%, 0.3%
POT normalization	2%

MiniBooNE and MicroBooNE. The ICARUS flux prediction and related uncertainties are based on those from MiniBooNE [27], adopting some minor modifications from MicroBooNE [28]. Uncertainties in the flux prediction arise from uncertainties in proton delivery, particle production, hadronic interactions, the horn magnetic field, and the beamline geometry. The dominant source of uncertainty for the ν_μ flux is π^+ production, while both π^+ and K^+ production are important for the ν_e component. The systematics are applied within the fitting framework (see Sec. VIII and Appendix C) as a covariance matrix generated with MC "universes" in which the individual sources of flux uncertainty are varied randomly according to their prefit uncertainties. The parameters corresponding to these universes are varied independently in the fit. Table IV provides a description of each variation included in the flux uncertainty. Additionally, there is a 2% uncertainty in the flux POT normalization, which is not included in the flux covariance matrix but rather treated as a separate flat normalization uncertainty in the fit. The fractional contributions to the uncertainty from each category of flux systematic are shown in Fig. 13. The flux correlation matrix is shown in Fig. 14. In both the analyses, lower energy events are highly correlated with each other, but not with higher (> 1 GeV) reconstructed neutrino energy bins.

B. Neutrino interaction systematics

Neutrino interactions are simulated in the analysis presented here using GENIE AR23 [35,36], with the

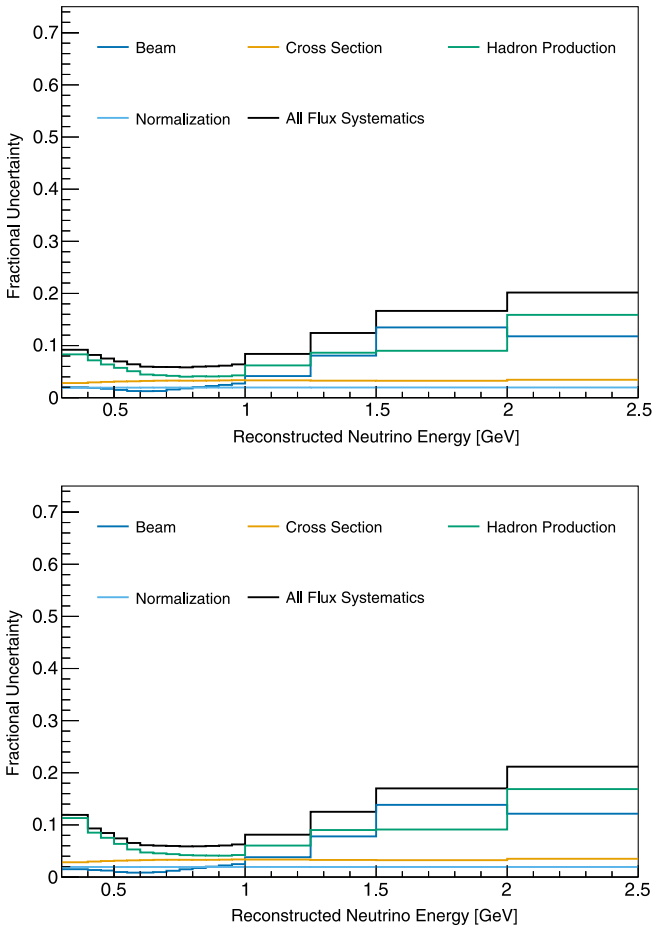


FIG. 13. Fractional size of the flux systematic uncertainty categories as a function of reconstructed neutrino energy for Pandora (top) and SPINE (bottom). The beam category includes the two horn current parameters and the cross-section category includes all the secondary interaction cross sections.

central value MC using the nominal weights for all parameters. The model uses local Fermi gas to describe the nuclear ground state (including a correlated high-momentum nucleon tail) [50,51] with the Valencia [52] 1p1h (QE) and SuSAv2 [53–56] 2p2h meson exchange current (MEC) interaction models. The 1p1h model includes the use of a z-expansion form factor fit to deuterium data [57]. The Berger-Sehgal [58–60] model of neutrino-induced coherent pion production is used for single pion production (SPP) channels. The AGKY [61] model covers multipion productions at low hadronic invariant mass ($W < 2.3$ GeV/c^2), including nonresonant backgrounds to the single-pion channels. The deep inelastic scattering (DIS) model uses the Bodek-Yang model [62] at higher W, with a linear AGKY to DIS transition at intermediate (2.3–3.0 GeV/c^2) W values. The hA2018 [63,64] model is used for final-state interactions (FSI) and includes emission of deexcitation photons for argon nuclei. Some of the model parameters are tuned to free nucleon data [65].

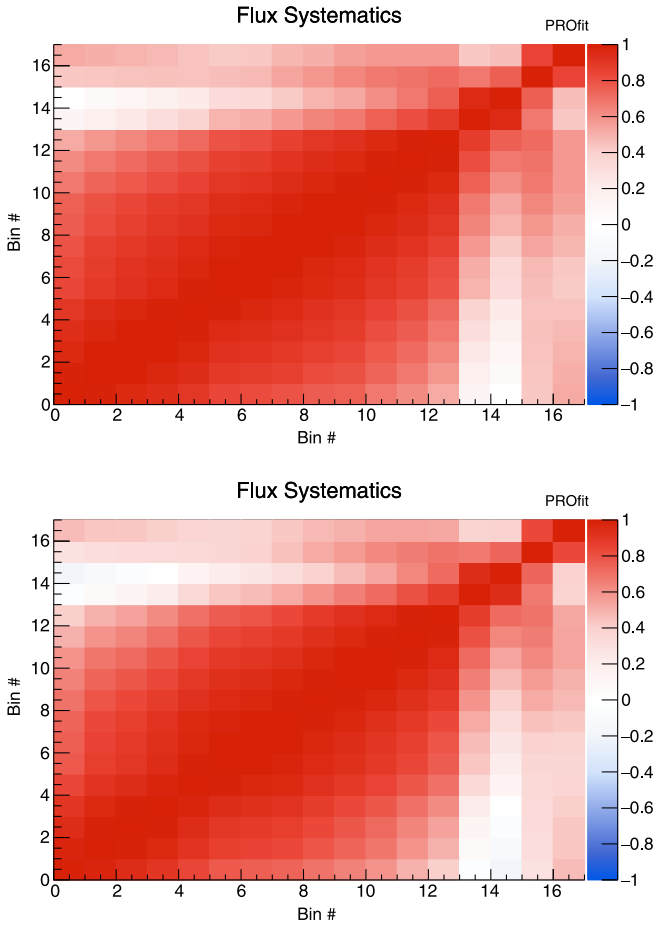


FIG. 14. Flux correlation matrix for Pandora (top) and SPINE (bottom). Axes are labeled with bin numbers using the fit binning in reconstructed neutrino energy.

Many systematic uncertainties and prior constraints on their variation are taken directly from GENIE as outlined in the GENIE Physics and User's Manual [66]. Additional uncertainties, not available in GENIE, are implemented using the NuSystematics package [67]. Reweights are calculated for each systematic, for each event, at a set of preset values: $\pm 1, 2, 3\sigma$, and splines are used to interpolate systematic weight values from these node points. Systematics are implemented in the fitter either by using random throws of parameter values to construct a covariance matrix in bins of reconstructed neutrino energy or by associating each spline with an additional nuisance parameter with its own prior uncertainty, which are then constrained in the fit. Systematics which are expected to be significant, such that we are interested in understanding which values the fit prefers, and those which are expected to be non-Gaussian, are treated using the latter method. The remaining, less significant, systematics are included in a covariance matrix. Table V provides a description of all the variations used in the analysis, including their prefit uncertainties and whether they are treated using splines or covariance in the fit.

The systematic uncertainties with the largest observed impact on the analysis are the random-phase-approximation (RPA) suppression of the quasielastic (QE) channel, the z-expansion parameters describing the QE axial form factor, the shape and normalization of charged-current meson exchanged current (CC-MEC), the axial and vector masses describing resonance production, and FSI. These uncertainties impact the predicted spectra as follows:

(i) RPA QE adjusts the strength of the low-Q^2 suppression of the QE channel implemented by the RPA correction to the base model. This is the largest impact interaction systematic for reconstructed neutrino energy, $E_{\nu,\mathrm{reco}} < 0.5$ GeV.

(ii) The z-expansion parameters control the QE form factor and are the second largest systematic for $E_{\nu\,\mathrm{reco}} < 1$ GeV.

(iii) CC-MEC normalization adjusts the normalization of the 2-particle-2-hole (2p2h)/MEC prediction. While the total contribution of the 2p2h events are relatively small compared to QE, the uncertainty on the process is large and this systematic absorbs the bulk of it. This is the largest impact interaction systematic $E_{\nu\,\mathrm{reco}} > 0.5$ GeV and has been observed to be effective at covering unrelated variations that adjust the normalization at low to medium reconstructed energy.

(iv) While the CC-MEC normalization absorbs most of the uncertainty on the 2p2h process, there is also uncertainty on the shape of the energy-momentum $(q_0 - q_3)$ distribution and the difference between the default SuSAv2 model and the alternate Valencia model is taken as an uncertainty. The result is additional uncertainty in the number of events in the $0.5 < E_{\nu\,\mathrm{reco}} < 1$ GeV region.

(v) CCRES M_A and M_v are the axial and vector masses of the associated form factors in the resonance model used. These are the second largest interaction systematics for $E_{\nu\,\mathrm{reco}} > 1$ GeV.

(vi) FSI systematics change the kinematics of the hadronic system and can impact energy smearing and move protons and pion above/below the tracking threshold. FSI systematics have roughly the same impact as the CC-MEC normalization systematic on the selected event sample, but impact the true-to-reconstructed energy smearing in important ways.

The fractional contributions to the uncertainty from each category of interaction model systematic are shown in Fig. 15. The largest contribution to the QE category is the RPA suppression parameter and the largest contribution to the MEC category is the CC-MEC normalization. The interaction model correlation matrix is shown in Fig. 16. In both analyses, there is a clear separation in correlations, with reconstructed neutrino energy bins <0.8 GeV being quite correlated with each other but not with bins >0.8 GeV and vice versa.

TABLE V. List of interaction model dials used in the analysis. For each parameter, the central value weight, the nominal $\pm 1\sigma$ uncertainties, and the implementation method within the fitter are noted. The z-expansion "B" parameters represent a new basis after the principal component analysis (PCA) performed on the original A1–A4 parameters, which are highly correlated.

Dial name	Short description	Central value	$+1\sigma$	-1σ	Implementation
ZExpB1CCQE	$B1$ Z-expansion parameter for axial-vector form factor on CCQE	See caption			Spline
ZExpB2CCQE	$B2$ Z-expansion parameter for axial-vector form factor on CCQE	See caption			Spline
ZExpB3CCQE	$B3$ Z-expansion parameter for axial-vector form factor on CCQE	See caption			Spline
ZExpB4CCQE	$B4$ Z-expansion parameter for axial-vector form factor on CCQE	See caption			Spline
RPA_CCQE	RPA suppression is turned on (off) for dial = 0 (1).	⋯	⋯	⋯	Spline
CoulombCCQE	Strength of EM potential for Coulomb corrections on CCQE	1	20%	20%	Spline
NormCC-MEC	Normalization of CC-MEC	1	50%	50%	Spline
NormNCMEC	Normalization of NC-MEC	1	50%	50%	Covariance
XSecShape_CC-MEC	dial = 1 for the reweight to Valencia 2p2h model	⋯	⋯	⋯	Spline
DecayAngMEC	dial = 1 gives an alternative distribution proportional to $\cos\theta^2$	⋯	⋯	⋯	Spline
FracPN_CC-MEC	CC-MEC proton-neutron fraction		20%	20%	Spline
MaCCRES	Axial-vector mass of the dipole form factor on CCRes	1.088962	20%	20%	Spline
MvCCRES	Vector mass of the dipole form factor on CCRes	0.840	10%	10%	Spline
MaNCRES	Axial-vector mass of the dipole form factor on NCRes	1.088962	20%	20%	Covariance
MvNCRES	Vector mass of the dipole form factor on NCRes	0.840	10%	10%	Covariance
NonRESBGvpCC1pi	Scale for nonresonance background: ν-p CC + 1π	1	50%	50%	Covariance
NonRESBGvpCC2pi	Scale for nonresonance background: ν-p CC + 2π	1	50%	50%	Covariance
NonRESBGvpNC1pi	Scale for nonresonance background: ν-p NC + 1π	1	50%	50%	Covariance
NonRESBGvpNC2pi	Scale for nonresonance background: ν-p NC + 2π	1	50%	50%	Covariance
NonRESBGvnCC1pi	Scale for nonresonance background: ν-n CC + 1π	1	50%	50%	Covariance
NonRESBGvnCC2pi	Scale for nonresonance background: ν-n CC + 2π	1	50%	50%	Covariance
NonRESBGvnNC1pi	Scale for nonresonance background: ν-n NC + 1π	1	50%	50%	Covariance
NonRESBGvnNC2pi	Scale for nonresonance background: ν-n NC + 2π	1	50%	50%	Covariance
NonRESBGvbarpCC1pi	Scale for nonresonance background: $\bar{\nu}$-p CC + 1π	1	50%	50%	Covariance
NonRESBGvbarpCC2pi	Scale for nonresonance background: $\bar{\nu}$-p CC + 2π	1	50%	50%	Covariance
NonRESBGvbarpNC1pi	Scale for nonresonance background: $\bar{\nu}$-p NC + 1π	1	50%	50%	Covariance
NonRESBGvbarpNC2pi	Scale for nonresonance background: $\bar{\nu}$-p NC + 2π	1	50%	50%	Covariance
NonRESBGvbarnCC1pi	Scale for nonresonance background: $\bar{\nu}$-n CC + 1π	1	50%	50%	Covariance
NonRESBGvbarnCC2pi	Scale for nonresonance background: $\bar{\nu}$-n CC + 2π	1	50%	50%	Covariance
NonRESBGvbarnNC1pi	Scale for nonresonance background: $\bar{\nu}$-n NC + 1π	1	50%	50%	Covariance
NonRESBGvbarnNC2pi	Scale for nonresonance background: $\bar{\nu}$-n NC + 2π	1	50%	50%	Covariance
AhtBY	A_{HT} higher twist parameter in BY model scaling ξ_W	0.538	25%	25%	Covariance
BhtBY	B_{HT} higher twist parameter in BY model scaling ξ_W	0.305	25%	25%	Covariance
CV1uBY	$C_{\nu 1u}$ u valence GRV98 PDF correction parameter in BY model	0.291	30%	30%	Covariance
CV2uBY	$C_{\nu 2u}$ u valence GRV98 PDF correction parameter in BY model	0.189	40%	40%	Covariance
NormCCCOH	Normalization of CC-COH	1	100%	100%	Covariance
NormNCCOH	Normalization of NC-COH	1	100%	100%	Covariance
MFP_pi	Scale for mean free path in the FSI of π	1	20%	20%	Covariance
FrCEx_pi	Scale for fraction of charge-exchange fate in the FSI of π	1	50%	50%	Spline
FrInel_pi	Scale for fraction of inelastic scattered fate in the FSI of π	1	40%	40%	Covariance
FrAbs_pi	Scale for fraction of absorption fate in the FSI of π	1	30%	30%	Covariance
FrPiProd_pi	Scale for fraction of pion production fate in the FSI of π	1	20%	20%	Covariance
MFP_N	Scale for mean free path in the FSI of nucleon	1	20%	20%	Covariance
FrCEx_N	Scale for fraction of charge-exchange fate in the FSI of nucleon	1	50%	50%	Spline
FrInel_N	Scale for fraction of inelastic scattered fate in the FSI of nucleon	1	40%	40%	Covariance
FrAbs_N	Scale for fraction of absorption fate in the FSI of nucleon	1	20%	20%	Covariance
FrPiProd_N	Scale for fraction of pion production fate in the FSI of nucleon	1	20%	20%	Covariance
MaNCEL	Axial-vector mass of the dipole form factor on NCEL	0.994989	25%	25%	Covariance
EtaNCEL	Strange axial-vector mass of the dipole form factor on NCEL	0.12	30%	30%	Covariance

C. Detector model systematics

Uncertainties arising from imperfections in the MC simulation of detector response are evaluated using alternative MC samples in which the detector model parameters are varied. The simulation samples used to evaluate detector modeling uncertainties include neutrino interactions uniformly distributed during the beam spill gate and cosmic rays overlapping out of time, as described in Sec. V A. Each sample contains 200,000 neutrino events occurring within the ICARUS fiducial volume. The CV simulation represents the nominal detector model, while the variation samples reflect specific modifications to individual components of the detector model relative to this nominal configuration. To minimize statistical fluctuations between the CV and variation samples, an identical set of generator-level events is used for both the CV and the variation samples. Samples are analyzed using the same reconstruction and selection tools as in the nominal analysis and ratios of the resulting selected spectra are formed between the variation and CV samples. These binned ratios

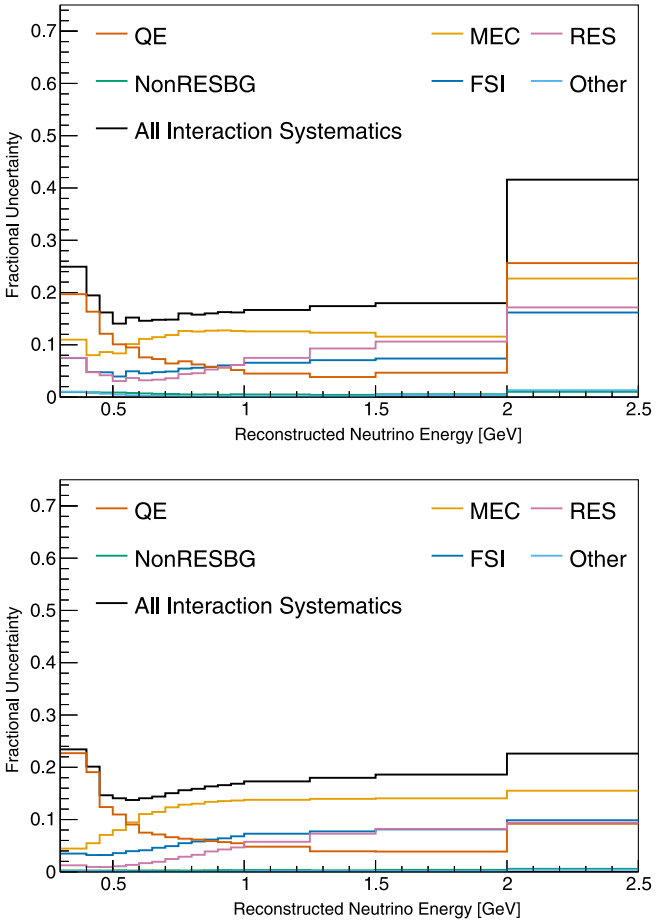


FIG. 15. Fractional size of interaction model systematic uncertainty categories as a function of reconstructed neutrino energy for Pandora (top) and SPINE (bottom).

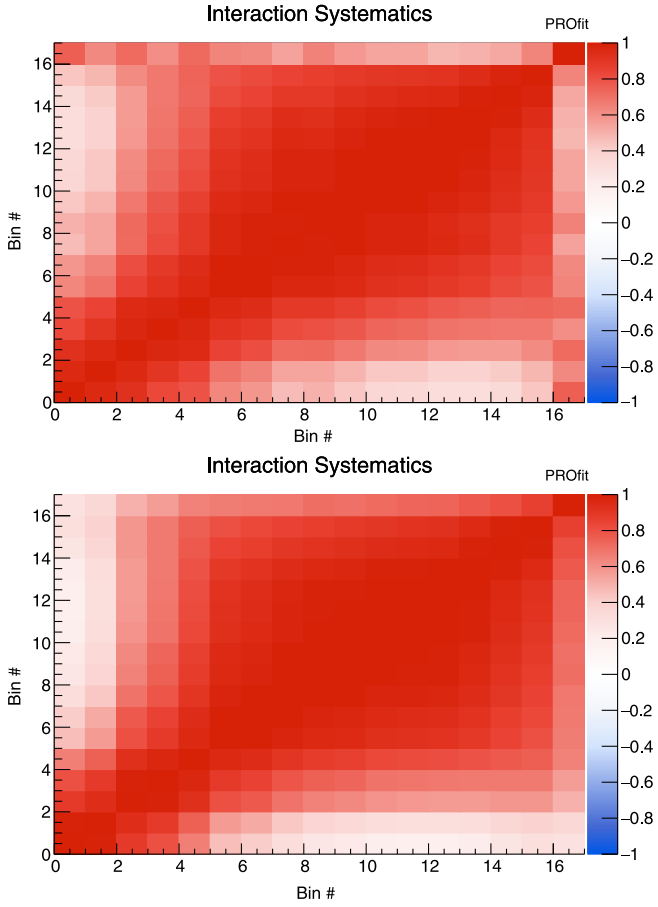


FIG. 16. Interaction model correlation matrix for Pandora (top) and SPINE (bottom). Axes are labeled with bin numbers using the fit binning in reconstructed neutrino energy.

are used to produce spline functions used by the fitting framework, with the assumption that the weight will scale according to a Gaussian distribution (i.e.: the 2σ ratio is twice the 1σ ratio). Each spline is then associated with an additional nuisance parameter with its own prior uncertainty, which is constrained in the fit, as described above. This procedure is performed separately for the Pandora and SPINE reconstruction paradigms; while the detector simulation is the same for both analyses, the very different reconstruction methods lead to significant differences in analysis sensitivity to mismodeling of a particular effect. For example, the SPINE-based analysis is less sensitive to track breaking resulting from detector effects because it has a dedicated GNN to assemble track fragments, while it may be more sensitive to mismodeling of charge scale because it makes more extensive use of charge information in reconstruction.

A number of potential detector variations were studied and ten of these were determined to be sufficiently impactful to include in the final analysis. The variations having the largest impact on the predicted spectra are:

(i) Intrinsic noise: the simulation does not include channel-to-channel noise variations which are measured to be of order 4%. The level of intrinsic noise in the TPC readout is varied by $\pm 10\%$, both to account for the impact of this unmodeled effect and to provide additional *ad hoc* freedom in the detector model.

(ii) Shaping time: the shaping time in the TPC readout electronics is varied by $\pm 15\%$ to account for the impact of imperfections in simulation of the TPC field response. An additional variation of shaping time in the second induction plane readout electronics is included to cover a mismodeling of this parameter.

(iii) Spatial nonuniformity: A detailed simulation of nonuniform detector response, believed to arise from modest incomplete transparency of the wire planes in the ICARUS detector, is implemented in this variation. In future analyses this simulation will be included in the CV MC; as this was not feasible for the analysis presented here, the impact is included as a systematic variation.

(iv) Recombination: There is an angular dependence to the recombination model that is best described using an ellipsoidal description of the β parameter in the Modified Box model that depends on the angle between the track direction and the drift field [42]. In this variation, the ellipsoidal Modified Box model replaces the original Modified Box model used in the CV MC.

(v) Free electron lifetime: The CV MC is generated with a nominal electron lifetime of 3 ms, while the measured electron lifetimes in ICARUS during Run 2 are 4 ms and 8 ms for the east and west cryostats, respectively. In this variation, the electron lifetime is set to 8 ms.

Other effects considered include variations in TPC coherent noise, variations in scintillation light yield, the impact of an unmodeled gap in the induction 1 wire plane, the impact of unmodeled cathode nonplanarity, uncertainty in trigger efficiency, and variations of fiducial volume and containment cuts.

The fractional contributions to the uncertainty from each category of detector systematic are shown in Fig. 17 and the detector systematic correlation matrix is shown in Fig. 18. As a result of the relatively small sample size used for the detector variations, some statistical fluctuation is visible in these results. The sample sizes were chosen such that systematic variations large enough to impact the analysis would be clearly visible. For the cases where fluctuations are observed, there is an impact from statistics on the energy dependence of the variation, but the magnitude of the effect is small enough to be dominated by fluctuations, such that these variations may be thought of as an upper limit on the size of the true detector modeling uncertainties and the shape fluctuations do not impact the final result. The suite of systematics is demonstrated to cover observed low-level data-MC discrepancies, as described in Sec. VII D.

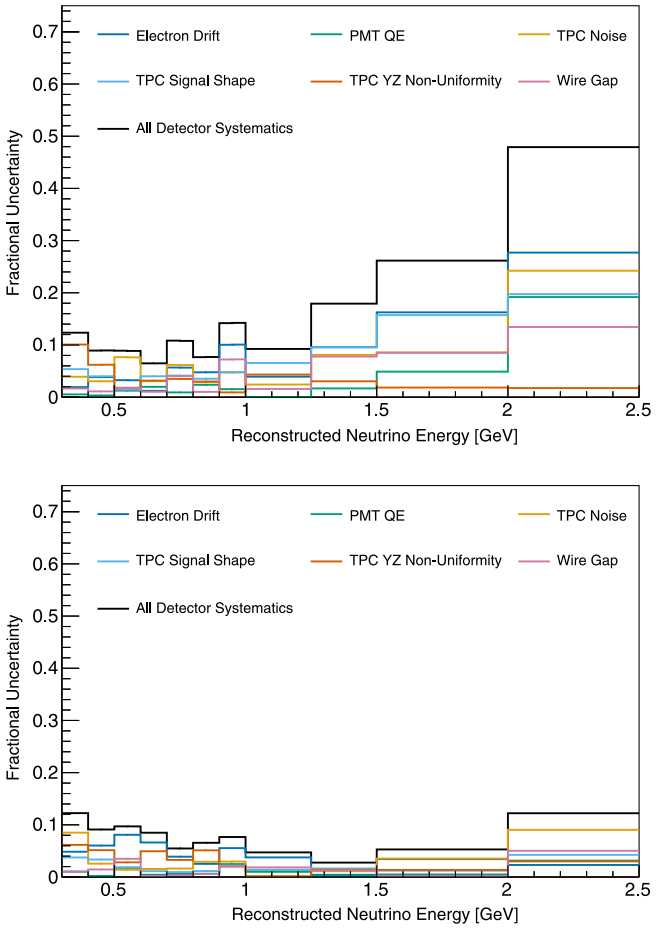


FIG. 17. Fractional size of detector model systematic uncertainty categories as a function of reconstructed neutrino energy for Pandora (top) and SPINE (bottom). The electron drift category includes electron lifetime and recombination.

D. Systematics validation

To verify that the systematic variations included in the analysis are sufficient to describe the true uncertainty in the predicted spectra, data-MC comparisons are performed on detector performance quantities that are impacted by the detector model and fitter-level studies are performed on "mock data" generated under alternative model assumptions. In the case of low-level detector quantities, which are common between the SPINE and Pandora analyses, the range of variations included as systematic uncertainties is compared to discrepancies between data and the central value MC. For the mock data studies, an evaluation is made as to whether systematic variations are sufficient to provide reasonable fits, such that none of the systematic parameters are pulled in unphysical or dramatic ways and no false oscillation signals are observed.

The efficacy of the detector variation systematics model is studied by exploiting two different low-level reconstruction variables: estimation of the wire signal (hit) identification efficiency and the charge/energy deposition (dQ/dx and dE/dx) distributions. Cosmic muon tracks crossing the cathode are considered for both studies and the analysis is performed separating the three different wire planes, to take into account their specific characteristics. The hit efficiency is evaluated as a function of the average hit pitch, which is defined as the length of the 3D track segment seen by a single wire. The green bands in Fig. 19 indicate the overall MC hit efficiency for both induction planes, and are obtained by combining all the detector systematics variations included in the analysis. Data hit efficiency is overlaid, demonstrating that the detector systematics model provides the desired coverage. The desired coverage is also observed in the collection plane, not pictured, where the hit efficiency exceeds 99% for both data and MC for almost all the studied hit pitch values.

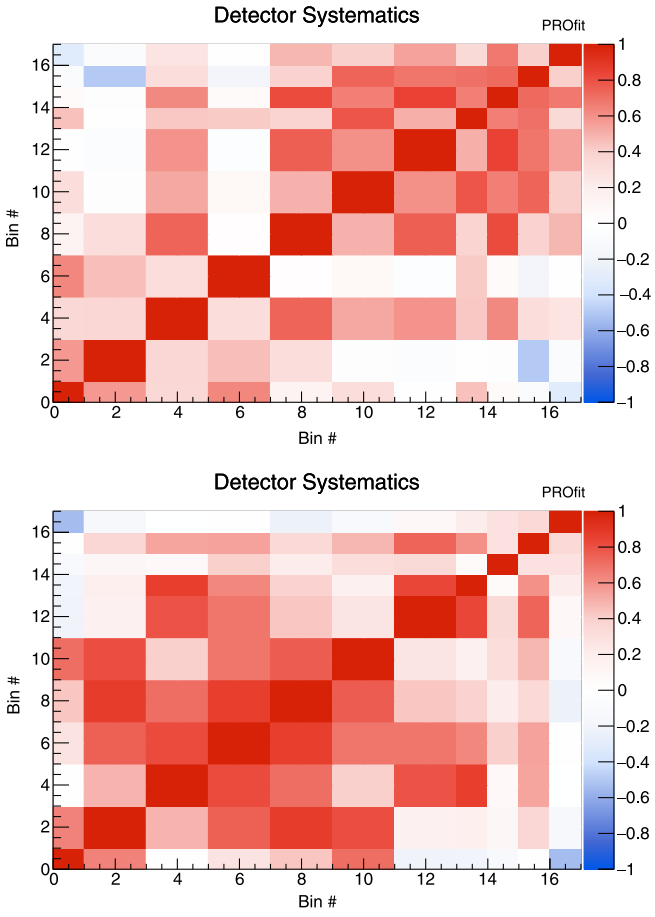


FIG. 18. Detector model correlation matrix for Pandora (top) and SPINE (bottom). Axes are labeled with bin numbers using the fit binning in reconstructed neutrino energy.

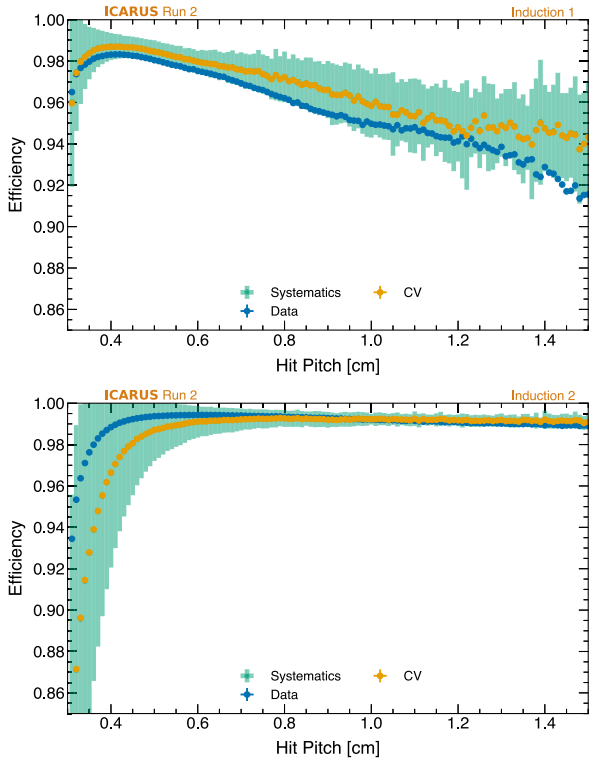


FIG. 19. Global results from hit identification efficiency studies for the Induction 1 (top) and Induction 2 (bottom) wire planes: the green band represents the hit efficiency coverage obtained by combining the detector systematic variations, to be compared with the measurements from data in blue.

In a similar way, to evaluate whether the shapes of dE/dx (for the collection wire plane only) and dQ/dx (for all the wire planes) are properly accounted for by the detector systematics, the probability density functions (PDFs) of the dQ/dx and dE/dx distributions are obtained. A comparison is made between the central value MC and the different systematics detector variations; in particular, an envelope is constructed by comparing all the PDF ratios of data/MC across the different TPCs and planes, in a bin-by-bin basis. An example is shown in Fig. 20 where the desired data/MC agreement (dashed orange line) is well contained within the green envelope across the entire dE/dx range for the WE TPC collection plane.

Uncertainty in the simulation of particle interactions within the detector volume could lead to additional uncertainty in the predicted spectra and are not explicitly included in the systematic variations described here. To verify that the impact of such variations is covered by our systematics model, a mock data study is performed in which the contribution of the impacted samples is artificially varied. Specifically, this analysis targets final states with zero pions, but truth-level studies reveal 5%–10% of selected MC events have true pions with initial energies above threshold in the final state, most of which are likely moved below threshold by reinteractions in the detector material. To study the impact of uncertainty in the reinteraction modeling, mock data is generated with the CCNπ component increased (decreased) by 100% and the NCNπ component decreased (increased) by 50%, as the two are expected to be anticorrelated. These represent extreme variations that are much larger than the actual uncertainty. Fits to this mock data do not result in any dramatic shifts of systematic parameters or any false oscillation signals, so we conclude that variations in particle interaction modeling are negligible for this analysis.

Additional mock data studies are performed to confirm that the systematic variations included in the analysis provide sufficient coverage of potential alternatives to the models used to generate the prediction. One method of generating mock data is to use alternative interaction models that are not used in the central value MC; for example, alternative models of FSIs, QE interactions, and MEC interactions are used to generate mock data, which is then fit using the central value MC for the prediction. The alternative QE model is based on a fit of the z-expansion CCQE axial form factor to lattice QCD and the alternative

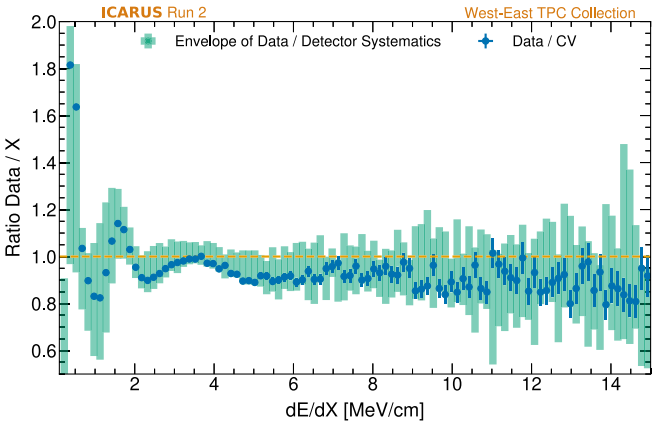


FIG. 20. Envelope of the PDF ratio of data dE/dx distributions over CV and variations, considering the final subset of detector systematics, is compared with the measurements performed in the WEST-EAST TPC collection plane.

MEC model adjusts the angular distribution of outgoing nucleons in the MEC interaction. An alternative method of generating mock data is to reweight the prediction from GENIE, in leading proton kinematic variables, to match the prediction from alternative generators NEUT [68], NuWro [69], and GiBUU [70]. In all of these mock data studies, no false oscillation signal is observed and the systematic pulls are not dramatic. Frequently the CC-MEC normalization and QE RPA parameters are pulled to compensate for a normalization effect in the mock data. As described in Sec. IX, this effect is also seen in the full data fits. Finally, combinations of systematics parameters can be purposefully constructed to generate mock data that is qualitatively similar to an oscillation signal. Even in these artificial worst-case scenarios, the $\Delta\chi^2$ between fits with oscillation parameters floating and fixed to zero do not indicate the presence of a statistically significant oscillation.

E. Systematics summary

The total uncertainty in the predicted event rate resulting from systematic variations is around 20% at the peak of the selected neutrino distribution, with the fractional contributions as a function of reconstructed neutrino energy shown in Fig. 21. The systematics coming from uncertainty in the neutrino interaction model are the largest contribution, with uncertainty from the flux and detector models having approximately equal contributions for most energies. The full correlation matrix, including all systematics variations, is shown in Fig. 22.

VIII. FITTING AND STATISTICAL ANALYSIS

The suite of necessary statistical analyses are performed using PROfit [71], a new general-purpose oscillation and fitting framework that has been developed for SBN analyses. PROfit finds the set of oscillation and systematic

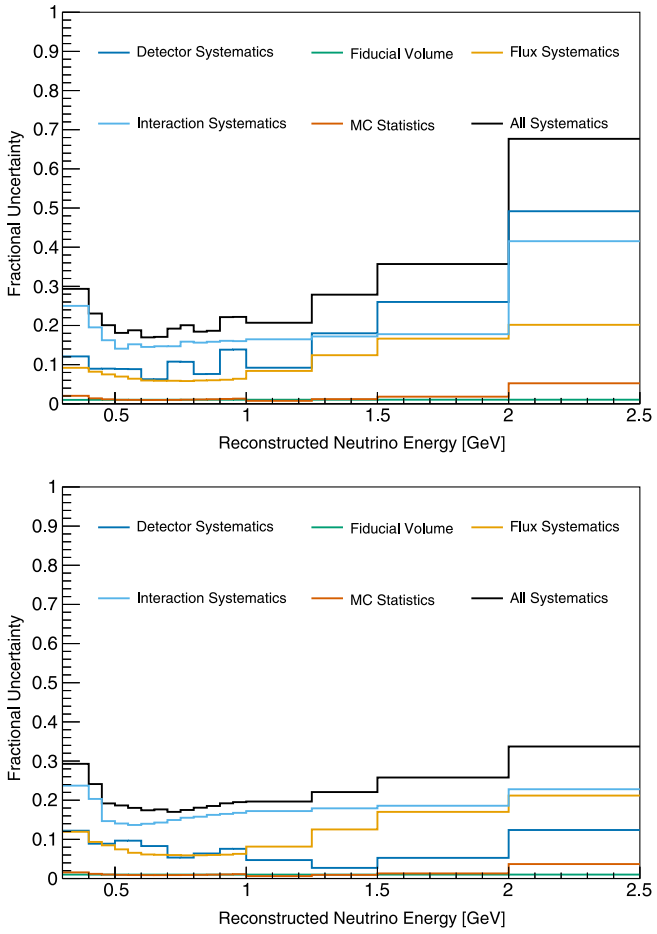


FIG. 21. Fractional size of systematic uncertainty as a function of reconstructed neutrino energy for Pandora (top) and SPINE (bottom).

parameters best describing the data, given prior uncertainties on the models used in the simulation. The analysis is performed in the context of a two-flavor approximation within the $3+1$ sterile neutrino model and results are presented in terms of Δm^2_{41} and $\sin^2 2\theta_{\mu\mu}$, with the assumption $\theta_{ee} = \theta_{\mu e} = 0$. Systematic uncertainties are incorporated as nuisance (pull-term) parameters governing the strength of correlated variations or through a covariance matrix; the choice made for each individual source of systematic uncertainty is described in Sec. VII. Minimization of the combined Neyman-Pearson (CNP) χ^2_{CNP} [72] metric is performed and confidence intervals on the oscillation parameters are constructed using the Feldman-Cousins procedure [73]. Signal unblinding follows a predetermined staged procedure in PROfit in which fits to both the SPINE and Pandora selections are performed simultaneously, with validation checks evaluated at each stage before proceeding to ensure a robust result. Details of the PROfit framework and the statistical analysis choices in this analysis are provided in Appendix C.

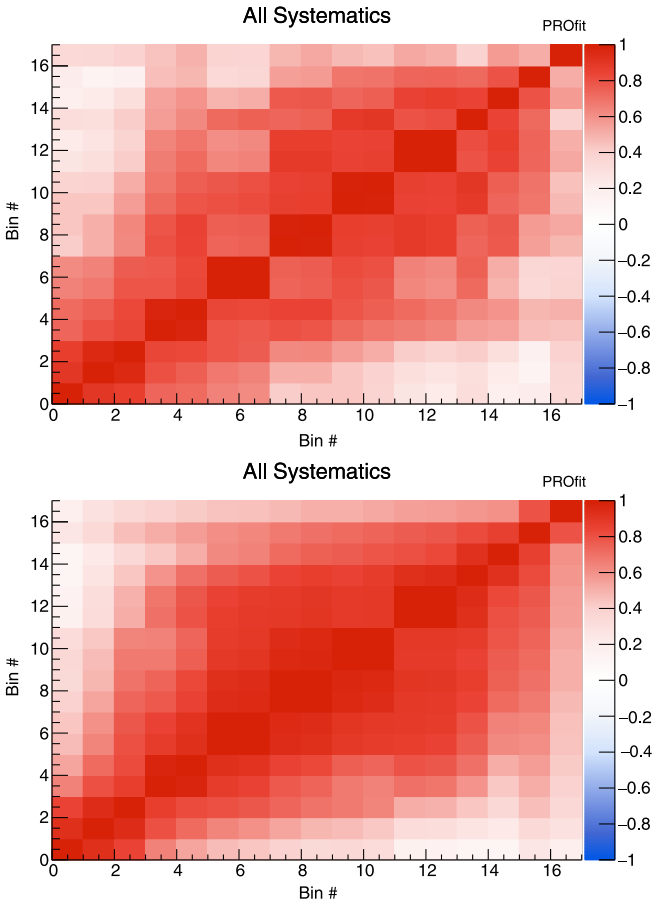


FIG. 22. Correlation matrix including all systematic uncertainties, for Pandora (top) and SPINE (bottom). Axes are labeled with bin numbers using the fit binning in reconstructed neutrino energy.

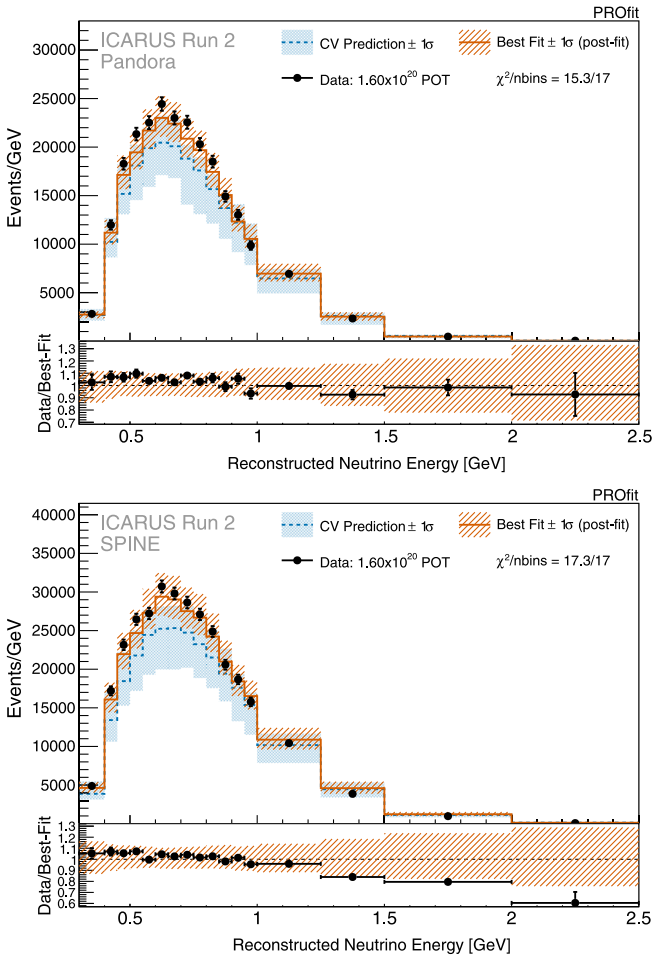


FIG. 23. Pandora (top) and SPINE (bottom) postfit spectra and data/best-fit ratios for the full fit in which oscillation parameters are allowed to float.

IX. RESULTS

No evidence for oscillation is observed in either the SPINE or Pandora fits. The best fit χ^2/ndof is 15.3/17 for Pandora and 17.3/17 for SPINE. While the oscillation is not statistically significant, the best-fit oscillation parameter values and the χ^2 values for both analyses are given in Table VI. The postfit spectra and uncertainties are shown in Fig. 23. Postfit spectra for fits to the null hypothesis, with oscillation parameters fixed at zero, are shown in Fig. 24.

TABLE VI. Fit results, including best fit oscillation parameters, the χ^2 for the nominal fit, and the $\Delta\chi^2$ with respect to a fit to the no-oscillation null hypothesis. No statistically significant oscillation is observed.

	Pandora	SPINE
$\sin^2 2\theta_{\mu\mu}$	0.07	0.24
Δm^2_{41} (eV^2)	10.2	13.5
χ^2 (17 bins)	15.3	17.3
$\Delta\chi^2$	0.4	3.2
p-value	0.910	0.421

Feldman-Cousins p values comparing the best-fit sterile neutrino hypothesis to the null (no oscillation) hypothesis are calculated as described in Appendix C. The resulting p values are 0.910 and 0.421 for the Pandora and SPINE selections, respectively, indicating no significant preference for sterile neutrino oscillations. Given the consistency with the null hypothesis, we report 90% confidence level exclusion contours.

PROfit diagnostic plots are inspected to understand the impact of systematic uncertainties in the analyses. All of the systematic parameter pulls are shown in Fig. 25. The largest systematic parameter pulls in the full fit are $\sim 1\sigma$, with the CC-MEC normalization being pulled to $+1\sigma$ in both analyses and the recombination and Induction 2 Shaping Time systematics being pulled to -1σ and $+1\sigma$, respectively, in SPINE. CC-MEC normalization has a large prior uncertainty that impacts the event rate in the bulk of our ν_μ distribution, so it is a natural way to accommodate the prefit normalization difference between data and CV MC, seen by

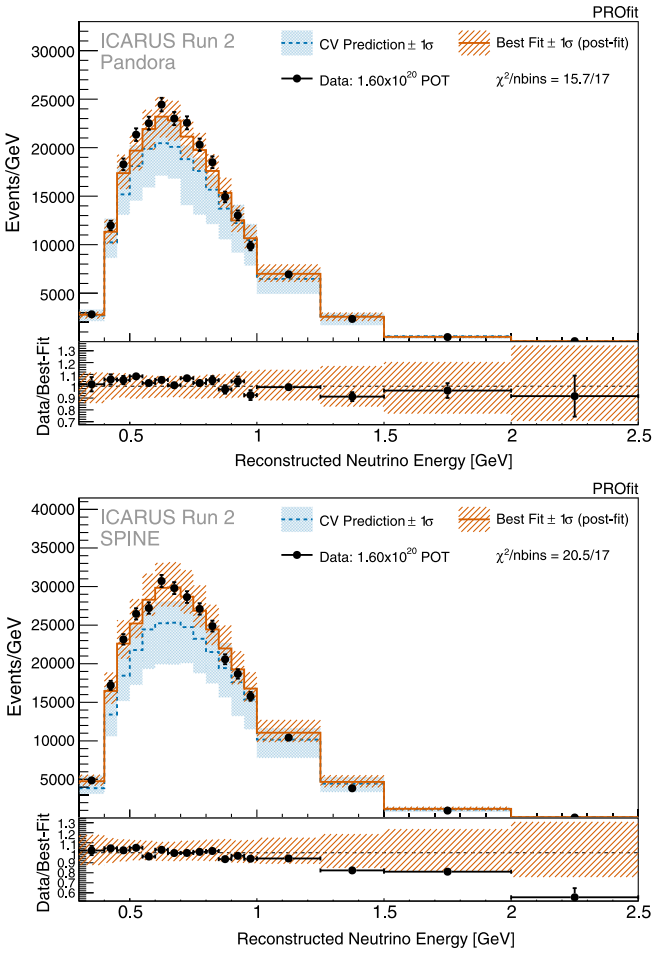


FIG. 24. Pandora (top) and SPINE (bottom) postfit spectra and data/MC ratios for a fit to the null hypothesis, where oscillation parameters are fixed to zero.

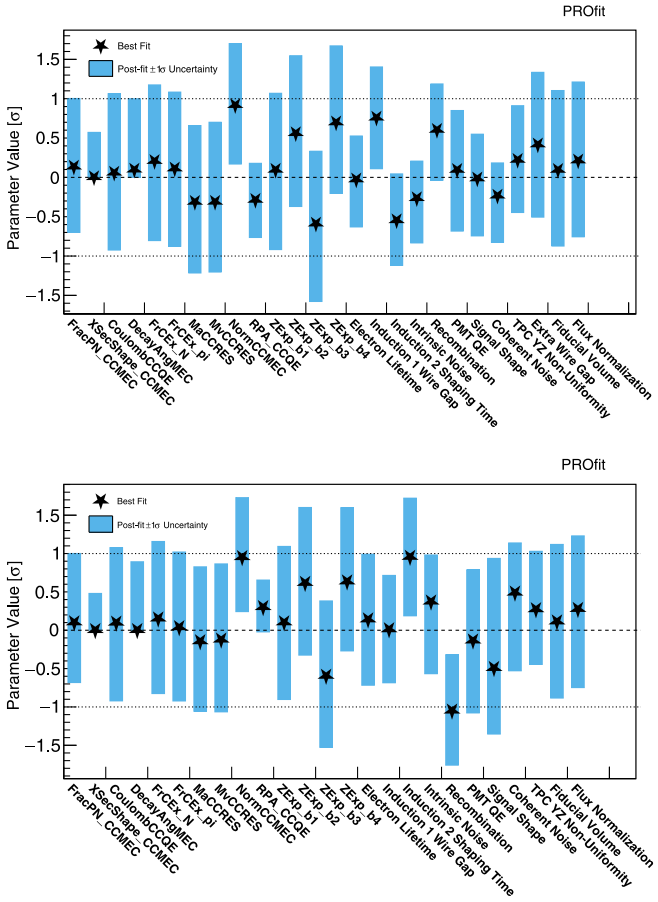


FIG. 25. The postfit nuisance parameter profiles for the Pandora (top) and SPINE (bottom) analyses. Prefit nuisance parameter values are by definition 0 (no pull), with prefit uncertainty defined as $\pm 1\sigma$ (horizontal dotted lines).

comparing the data points to the CV prediction (dashed blue line) in Fig. 23. This was observed in many mock data studies, where the CC-MEC normalization was frequently pulled to cover normalization differences. The negative pull on recombination in the SPINE analysis is not consistent with expectations from known detector parameters; the most likely explanation is that the impact of this parameter is degenerate with some other effect present in the data. The significant constraint on the RPA-CCQE parameter can be understood because this systematic has a large prior uncertainty in a specific energy range and thus can be constrained by the normalization of the data in that range.

Figure 26 shows the full correlation matrix for all parameters, including the oscillation parameters. Note that in the SPINE fit, two systematics parameters, the CC-MEC normalization and the RPA correction to the QE model, are correlated with each other and with the oscillation parameters. This is understood by noting that, as described above, these two parameters can combine to produce a normalization effect which acts oppositely to the normalization effect from a potential oscillation signal.

The final 90% CL sensitivity contours and exclusion band shown in Fig. 27 are generated using Feldman-Cousins corrected critical χ^2 values as described in the last Section. For both analyses, the data contour falls within the 2σ sensitivity band for nearly all points. The spiky behavior in the data contour as a function of Δm^2 is typical of short-baseline oscillation exclusion contours and results from the sinusoidal variation in oscillation probability. The point with least exclusion power is closest to the data best fit point and the nearby point with strong exclusion is the opposite local extremum, which is less like the data and thus more strongly excluded. This also explains why the SPINE contour has somewhat more pronounced spikes than the Pandora contour, since the SPINE best fit has a slightly higher significance for oscillation and the preferred mixing angle for the SPINE fit is significantly larger. At high values of Δm^2, the data contour falls near the $+1$ to $+2\sigma$ boundary of the sensitivity band. Oscillations in this region of parameter space would be observable at ICARUS primarily as a change in normalization; since the normalization of the data is slightly higher than that of the central value MC, while an oscillation result would result in a lower rate of

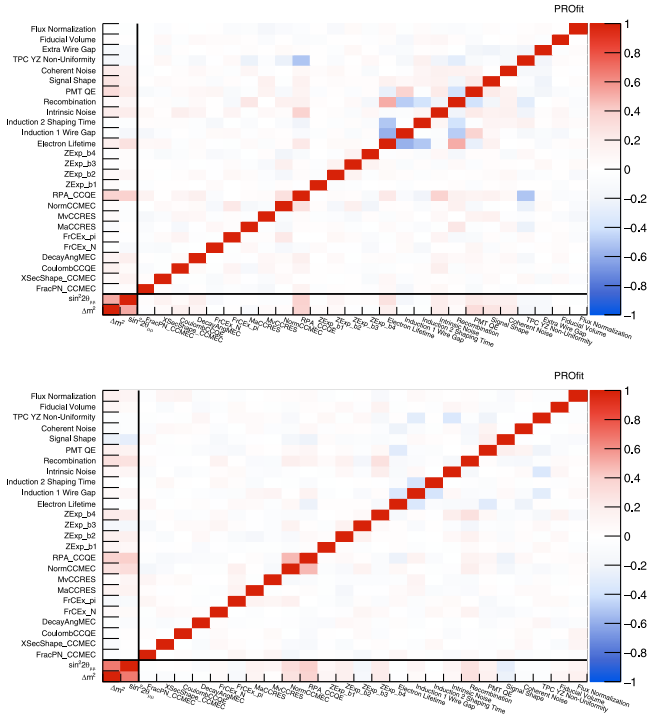


FIG. 26. The postfit correlation matrices for all parameters from Pandora (top) and SPINE (bottom) when fitting with oscillations. Highlighted on the left is the two physics parameters.

data relative to prediction, it is clear why the data exclusion is stronger than the (MC-based) median sensitivity in this region.

X. DISCUSSION AND FUTURE PROSPECTS

The results presented here are consistent with and have comparable sensitivity to previous single-detector ν_μ disappearance analyses, as seen in Fig. 28, which compares the ICARUS 90% CL exclusion contours to existing limits (and one closed contour) from other experiments. The global picture of short-baseline ν_μ disappearance remains uncertain as there is mild tension between existing exclusion contours and the IceCube closed contour. Additionally, previous anomalous results suggesting potential ν_e appearance remain unexplained, with increasingly stringent limits on the $3+1$ sterile neutrino model being set. While limits on $3+1$ oscillation parameters, such as those reported in the analysis presented here, are still the benchmark by which sensitivity is evaluated, future results from ICARUS and SBN will move toward more comprehensive and sensitive exploration of the rich theory landscape.

The ICARUS result presented here is systematics dominated, with the total systematic uncertainty of around 20% in the predicted event rate being significantly larger than the statistical uncertainty. As described in Sec. I, ICARUS serves as the far detector of the SBN program, the ultimate goal of which is to produce 5σ exclusion contours in a

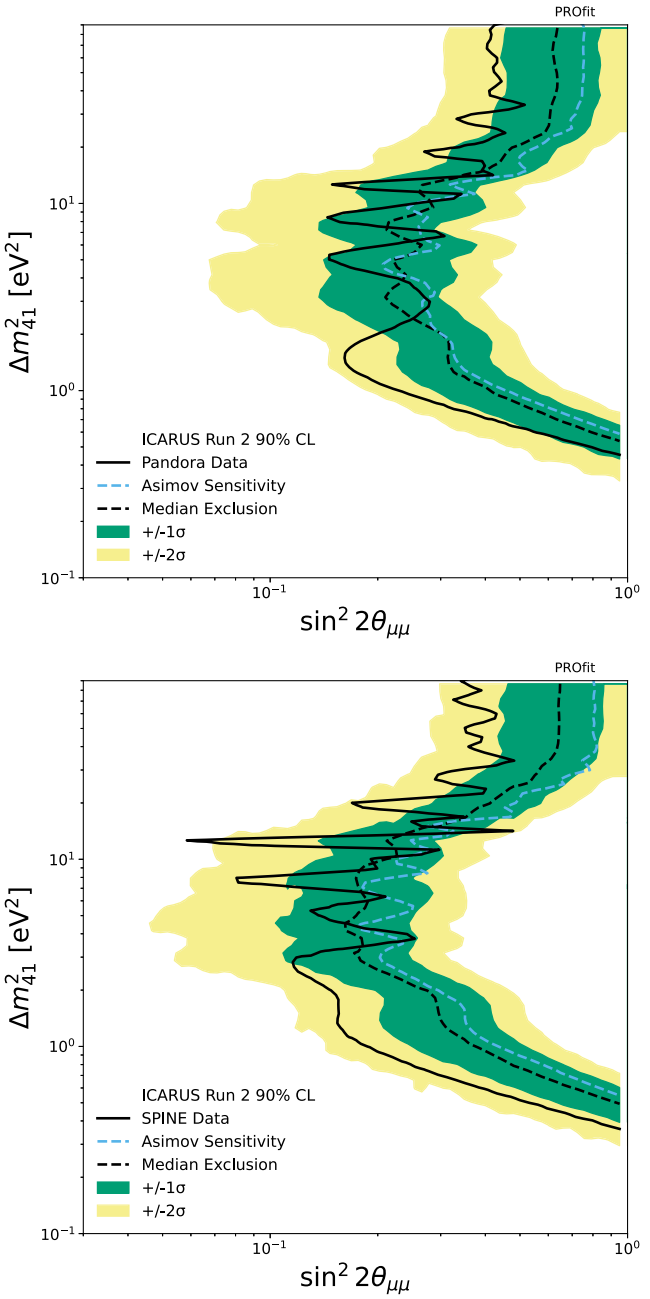


FIG. 27. Feldman-Cousins 90% CL exclusion contours for Pandora (top) and SPINE (bottom), shown alongside the sensitivity contours, with 1σ and 2σ uncertainty bands for the median exclusion sensitivity. The uncertainty bands represent the range of exclusion contours within which 68.3% and 95.4% of universes excluding oscillation fall.

combined analysis of ν_μ and ν_e disappearance and ν_e appearance in the BNB and NuMI beams, using data at different baselines from the near [Short Baseline Near Detector (SBND)] and far (ICARUS) detectors. The SBN program is designed for constraint of systematic uncertainty, with the flux and interaction model uncertainties nearly 100% correlated between the two LArTPC detectors

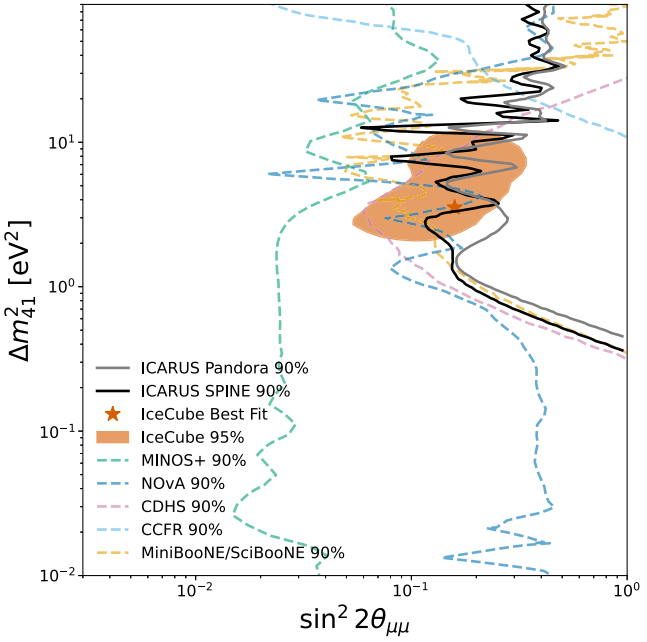


FIG. 28. 90% CL (FC corrected) exclusion contours for Pandora (gray) and SPINE (black), shown in the context of existing limits and closed contours from other experiments [12–17].

and so significantly constrained in a two-detector oscillation analysis. Furthermore, improvements to the detector simulation, many of which were developed as systematic variations in the analysis reported here, will remove many residual data-MC discrepancies in ICARUS. The leading sources of detector model uncertainty in this analysis (spatial nonuniformity, recombination, electron lifetime, TPC noise, and TPC field response) are already better modeled or taken from data in the latest versions of ICARUS simulation. Combined with the high-quality data being collected in SBND, these improvements to the detector model will produce robust and world-leading two-detector oscillation analyses.

XI. SUMMARY

ICARUS observes no evidence for ν_μ disappearance in the BNB in the Run 2 dataset, using both the Pandora and SPINE reconstruction frameworks to identify samples of $1\mu\mathrm{N}p$ interactions; 90% exclusion contours are presented for both analyses in the context of the $3+1$ sterile neutrino model. This result is consistent with previous single-detector ν_μ disappearance searches. The result is systematics dominated, as expected in the absence of near detector data to provide constraint on the large *a priori* uncertainties on the neutrino flux and interaction models. Future results will incorporate additional data and build on the tools and methods presented here to more fully constrain systematic uncertainties and more fully explore theories predicting baseline-dependent effects in SBN.

ACKNOWLEDGMENTS

This document was prepared by the ICARUS Collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by FermiForward Discovery Group, LLC, acting under Contract No. 89243024CSC000002. This work was also supported by Istituto Nazionale di Fisica Nucleare (INFN, Italy), EU Horizon 2020 Research and Innovation Program under the Marie Sklodowska-Curie Grant Agreements No. 822185, 858199, and 101003460 and Horizon Europe Program research and innovation program under the Marie Sklodowska-Curie Grant Agreement No. 101081478, the research contract per Law 240/2010, Art. 24 (3)(a), and D. G. R. 693/2023 (REF. PA:2023-20090/RER—CUP:J19J23000730002) by FSE+ 2021–2027. Furthermore the support of CERN in the detector overhauling within the Neutrino Platform framework, in the detector installation and commissioning, is acknowledged. This work was also supported by the Anusandhan National Research Foundation (ANRF, India) under the Ramanujan Fellowship (Grant No. RJF/2025/000203). The ICARUS Collaboration would like to thank the MINOS Collaboration for having provided the Side CRT panels as well as Double Chooz (University of Chicago) for the Bottom CRT panels. We also acknowledge the contribution of many SBND colleagues, in particular for the development of a number of simulation, reconstruction and analysis tools which are shared within the SBN program.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: Pandora

This Section provides a brief overview of the steps required to perform a full event reconstruction using the Pandora software framework, from raw 2D hits to the hierarchy of interactions. A detailed overview of the software, including details on the reconstruction chain and the main algorithms employed therein is provided in [19,43,44]. Here, the software infrastructure is described in general, highlighting the specific algorithm optimizations performed for ICARUS analyses.

Pandora [19] is a pattern recognition software widely used in LAr-based detectors, especially LArTPCs. The goal of this software is to provide a fully automated and complete reconstruction of events, delivering a hierarchy of tracks and electromagnetic showers to represent final-state particles and any subsequent interactions or decays.

A multialgorithm approach is adopted to perform pattern recognition on the reconstructed images for the collected events. In the processing chain, each algorithm acts

independently from the others and addresses a specific task in a particular topology, e.g., interaction vertex identification or shower 3D reconstruction. It is the ensemble of algorithms that gradually builds up a picture of the underlying events and collectively provides a robust reconstruction.

The inputs to Pandora reconstruction are the collection of *hits*, i.e. signals detected on single wires at a definite drift time, and information on the detector geometry. The starting point is a set of 3×2D images (where x coordinate represents the drift time and y the wire number), one per readout plane. Reconstruction begins with two-dimensional analysis, where separate clusters of hits are created for each structure in the input image from each readout plane. Ideally, each cluster defined at this step should contain all and only the hits belonging to a single reconstructed particle. Clustering is done in stages, of which the first one privileges purity over completeness and puts together only the hits that are likely to belong to the same reconstructed particle. A series of topological algorithms is later used to make cluster merging and splitting decisions on the overall event topology (as opposed to individual clusters in isolation) and refine the 2D clustering. Next, clusters of 2D hits from the three readout planes that represent individual tracklike particles are identified. Finally, for each input 2D hit in a reconstructed particle, a 3D hit is created, with different mechanisms depending on the cluster topology [43]. An analytic χ^2 minimization is used to match hit coordinates in one direction (say x) with the remaining ones (y and z).

Two multialgorithm reconstruction paths are employed to provide hierarchy reconstruction:

(i) *Pandora Cosmic*, which is optimized for the reconstruction of cosmic ray (CR) muons and their delta rays. This path is strongly track oriented, since the expected primary particles are muons with showers assumed to be their daughter delta rays. The interaction vertex is given by the high-y coordinate of the muon candidate track. A detailed description of the four main stages subsequently run as part of this reconstruction step is provided in [43].

(ii) *Pandora Nu*, is the alternative reconstruction path and is optimized for the reconstruction of neutrino interactions, starting from the interaction vertex and including all the particles emerging from that. The primary parent is the invisible neutrino and all the reconstructed particles are its daughters. Further details on this path, including ICARUS dedicated optimization studies, are included in the following text.

The event processing chain, schematically represented in Fig. 29, relies on both paths.

First, an initial pass of *Pandora Cosmic* reconstruction is run in order to extract a list of unambiguous candidate cosmic muons, tagged as *Clear Cosmic* interactions, and discard the hits associated with them from further steps. Pandora identifies an interaction as a clear cosmic for being too far out-of-time with the beam, for example if it clearly extends well into the extra readout window pre- and post beam (plus drift time), or if it appears to be fully through-going across the detector volume.

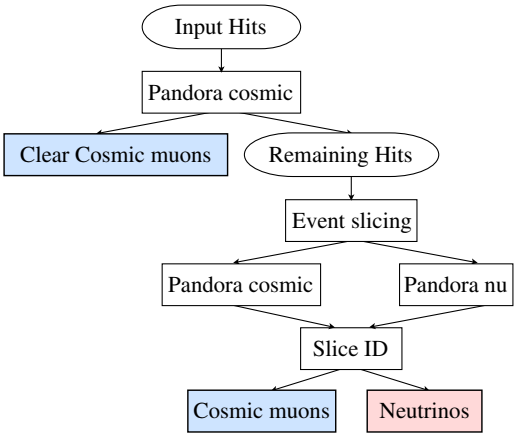


FIG. 29. Schematic view of the processing chain for the Pandora event reconstruction. Inputs to Pandora are indicated by elliptical boxes. The outputs of the reconstruction chain, that are saved and used in later steps of the processing chain, are indicated by colored boxes.

The standard rejection of cosmics from Pandora (Cut Mode = nominal) corresponds to (i) $\cos\theta > -0.6$, with θ being the angle with respect to the vertical, (ii) the containment along y, and (iii) a time cut requiring particles to be *in-time* with respect to the beam. With respect to the latter condition, particles are considered *in-time* if all their hits are within $\sim$1 ms drift-time interval as given by the trigger. In addition, for a cathode crossing particle entering the detector, Pandora will stitch together the two pieces in the adjacent TPCs, extracting the crossing time that has to be compatible with the trigger.

However, muons from ν interactions that are almost vertical can be tagged as CR muons due to the cut on $\cos\theta$. After dedicated studies comparing the efficiency of background rejection with the capability to retain neutrino candidates, it was decided to loosen the timing requirements for the cosmic ray tagging and remove the condition on the vertical angle θ (CUT mode = cautious). With such a choice the tagging of cosmic rays is still effective, with virtually no loss of neutrino events.

The remaining cosmic-removed hit collection is used for *event slicing*. This step is aimed at collecting and separating 3D hits belonging to single neutrino interactions from cosmic ray remnants, i.e., residual background contaminations. All the 3D clusters deemed to be associated with a given *slice*, representing a single interaction in the detector, are grouped together. Ideally each slice should contain all and only the hits belonging to a single particle hierarchy. As part of this sequence and prior to the slicing

step itself, some of the reconstruction steps included in the *Pandora Nu* reconstruction pathway are performed as they are preparatory to an effective event slicing. Among them are 2D reconstruction, 3D track and shower reconstruction, 3D clusters recovery and 3D hit reconstruction.

The cosmic-filtered sets of 2D hits on the readout planes associated with a given slice are the input of the dedicated reconstruction performed in the next steps, in which each slice is processed separately using both *Pandora Nu* reconstruction and *Pandora Cosmic* reconstruction. This second pass of *Pandora Cosmic* is identical to the former and is aimed at providing a proper reconstruction to leftover cosmic-ray contaminations.

A key step that distinguishes the *Pandora Nu* reconstruction path from the *Pandora Cosmic* path is *3D interaction vertex reconstruction*. Here, the lists of 2D clusters from different readout planes are used to identify the neutrino interaction vertex, ensuring a correct identification and reconstruction of individual *primary* particles as unique (reconstructed) particle flow particles (PFPs). This is done in three steps. First, all the cluster end points are used to create up to four candidate vertices for each cluster pairing between different readout planes and then, candidates not sitting on a hit or a registered detector gap are removed from the list. Finally, using a BDT, the surviving candidates are assigned a score and the one with the best score is selected as the interaction vertex and saved for later steps.

The global score is evaluated combining three estimators: an energy, an asymmetry and a beam score. The energy score is extracted by comparing the energy of each cluster in each view with the vertex projection onto the same view and is aimed at assessing consistency between vertex candidates and reconstructed clusters on all the views. The larger the score, the more the vertex candidate is suppressed. This reflects the fact that primary particles, i.e., neutrino daughters, should point back to the true neutrino interaction vertex, while secondary particles may not, but should be less energetic. The asymmetry parameter is obtained by counting the number of hits upstream and downstream of the candidate vertex position and suppresses candidates incorrectly located along a single cluster as opposed to candidates close to cluster endpoints. The beam score uses the beam direction to favor vertex candidates with low values of z, i.e. the beam direction coordinate.

The variables used to perform the classification and identify the most suitable vertex candidate include both properties of the reconstructed event and features of the candidate vertices. The former variables include the total number of hits, clusters and vertex candidates associated to the event, the fraction of hits likely originating from a shower based on cluster length and the associated reconstructed particle, an estimate of the energy and of the area of the event, extracted from the total clusters span along x and z directions and an estimate of how longitudinal the event is with respect to the beam direction. The vertex variables include all the quantities required to extract the score estimators mentioned above, i.e., a measurement of the transverse energy extracted comparing the displacement of all clusters in all views with respect to the candidate vertex projection and estimates of asymmetry parameters evaluated globally, locally and assuming the presence of a shower. In order to improve vertex selection for ICARUS, this BDT is trained on a MC sample of BNB $\nu_\mu + \nu_e$ interactions in the ICARUS detector.

The subsequent steps are 3D track and shower reconstruction. The former is described in detail above, since it is part of both neutrino and cosmic reconstruction paths, while *3D shower reconstruction* algorithms are specific to *Pandora Nu*. To select 2D shower candidate clusters, a set of geometrical and topological cuts are used to distinguish 2D tracklike clusters from 2D showerlike clusters. Then, in each view, clusters representing candidate shower branches are merged with the appropriate candidate shower spines, for example, long showerlike clusters close to the interaction vertex, to provide high-completeness 2D shower clusters. Later, a rank-three tensor is diagonalized to properly match 2D information from different readout planes and build 3D unambiguous shower particles. The goal is to collect the clusters representing the same shower in each readout plane: predictions on the spatial extent of the shower from combinations of clusters on two planes are compared with the envelope of matching clusters in the third plane. Once 3D showers have been identified, 3D track reconstruction is repeated in order to recover potential track pieces dissolved to try to build showers. Finally, an attempt to recover groups of clusters that failed to satisfy the quality cuts for particle creation is done, lowering the thresholds used to match information coming from different planes.

Before building up a hierarchy of reconstructed interactions, 3D tracklike and showerlike particle reconstruction is refined. As usual, the first step is handled by a set of 2D algorithms and then 3D tools are employed. In ICARUS, three 2D and three 3D particle-refinement cluster mop up algorithms are subsequently run with the aim of merging unassociated clusters with assigned particle clusters, especially to increase the shower reconstruction completeness. A description of the algorithms is given in [43]. Ideally, at this point only a negligible number of 2D hits should remain unassigned to reconstructed objects.

The final step in the *Pandora Nu* reconstruction is *particle hierarchy reconstruction*. The procedure starts with the creation of a neutrino particle, then a 3D interaction vertex is assigned to the ν candidate and all the associated particles are added as primary daughters of the neutrino. Daughter particles of the primaries, i.e. secondaries, are later added to the existing ν daughters. Each particle in the final neutrino

interaction is assigned a 3D vertex position which is equal to the neutrino interaction vertex for primary particles and the point of closest approach to their parents for secondaries.

The last neutrino algorithm of the chain handles the classification of *tracklike and showerlike reconstructed particles*. Using a BDT, clusters are classified as tracklike or showerlike based on a set of 13 reconstructed quantities, including *geometric* quantities and *calorimetric* features that rely on the measured charge as a proxy for energy. The BDT training is based on a sample of simulated $\nu_\mu + \nu_e$ BNB interactions containing equal numbers of tracklike and showerlike particles. The BDT classification efficiency for MC is ∼80% and consistent results are obtained with validation checks on NuMI MC events and data. The primary daughter particle with the highest number of hits dictates whether the neutrino candidate is identified as ν_μ or ν_e: the former if it is tracklike, the latter if it is showerlike.

Once both *Pandora Nu* and *Pandora Cosmic* reconstruction paths are completed, a choice of the most appropriate reconstruction for each slice in the event, called the slice classification stage, takes place. This decision is again made by means of a BDT. The variables used to define the *neutrino score* rely on both reconstruction outcomes. The list of BDT features extracted under the neutrino hypothesis includes the number of reconstructed neutrino daughter particles and the number of 3D hits in the slice, the y coordinate of the ν vertex, the average direction of primary particles weighted by the number of space points associated to them, and the output of a PCA on the space points reconstructed within a certain distance from the vertex. The variables computed under the cosmic ray assumption are in reference to the longest reconstructed track in the slice and include its direction projected along y, its vertical deflection, and both number and fraction of hits of the slice in this track. This is motivated by the fact that cosmic ray muons usually enter from the top of the detector and generate simpler, typically tracklike, topologies relative to neutrinos from the beam. The *neutrino score* could be used to classify the slice as a *neutrino candidate* or a *cosmic muon*, as is done e.g., in ProtoDUNE [44]. In SBN, ICARUS and SBND have adopted a conservative approach in which the neutrinolike reconstruction is always saved and, with the exception of BDT failures, the ν hypothesis is always selected.

While PID is not part of the Pandora framework, a few details on the method employed to perform it are provided here. The particle identification method used in ICARUS condenses spatial and calorimetric information extracted from reconstructed particles into a "χ^2 score". The underlying assumption is that, if the incident particle stops in the TPC active volume, the energy loss dE/dx as a function of the residual range rr, defined as the distance of a given energy deposition within a certain track from the endpoint of the track itself, is a powerful tool to distinguish charged particles of different species. In the χ^2 score definition, the measured dE/dx for a given reconstructed track is compared, on a hit-by-hit basis, with the average response predicted under different particle hypotheses from the Bethe formula, including protons, charged kaons, charged pions, and muons. The total χ^2, for a given readout plane, is the sum of the contribution of all the hits, except for the first and last in the selected track. This sum is divided by the total number of hits reconstructed in the same wire plane to extract the PID score,

$$\mathrm{PID} = \chi^2/\mathrm{ndof} = \sum_{\mathrm{hit}} \left[\frac{dE/dx_{\mathrm{meas}} - dE/dx_{\mathrm{theory}}}{\sigma_{dE/dx}} \right]^2 / n_{\mathrm{hits}} \tag{A1}$$

where $\sigma_{dE/dx}$ is an estimate of the resolution on the measured dE/dx per hit (∼3%). This technique was originally developed by ArgoNeut [40,74].

While a PID score can be extracted for all the wire planes, for the analysis presented here, only the collection plane is used. In addition, based on dedicated studies, unphysical hits with dE/dx signals below 0.5 MeV/cm or above 100 MeV/cm are excluded from the χ^2 calculation, thereby increasing the purity of the sample without impacting its efficiency. Furthermore, only the last 25 cm of the track is used to compute the χ^2, unless the track is shorter. It is important to note that this strategy relies on the characteristic Bragg peak of each particle; thus if a particle scatters, instead of stopping, the identification power is lost. The discrimination power is also undermined when the reconstruction fails to identify the entire track, losing some hits, especially close to the end point.

A dedicated visual study of events from a single run from Run 2, not included in the analysis presented here, was performed to select a sample of ν_μ CC interactions from BNB. These events, combined with MC simulations, were used to validate Pandora-based event reconstruction, identify the most frequent pathologies—particularly in vertex identification and track reconstruction for both muons and protons—and to explore and test possible improvements. This was also an essential step to tune our selection algorithms. Figure 30 shows the comparison between visual scan and Pandora-based reconstruction of the interaction vertex and the end point of the muon candidate along the beam direction, indicating that, for the bulk of the distribution, the automatic reconstruction and the manual measurement agree to within ∼7 mm, with a possible bias of a few millimeters.

APPENDIX B: SPINE

The end-to-end, ML-based reconstruction chain used in this analysis is referred to as SPINE (Scalable Particle Imaging with Neural Embeddings) [20]. SPINE consists of a hierarchical set of neural networks that are optimizable end to end. Two types of neural networks are employed:

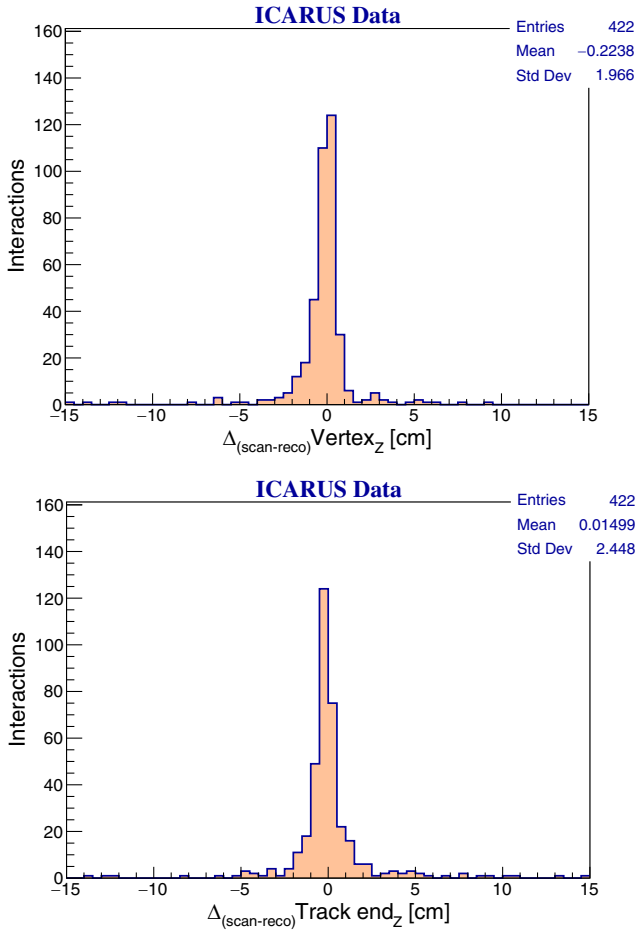


FIG. 30. Visual scan and Pandora-based automatic reconstructions are here compared for a sample of 422 candidate ν_μ CC interactions from run 9435. In the top plot, a zoom in the range $(-15, 15)$ cm in the difference between the vertex z coordinate obtained from the visual scan and automatic reconstruction is drawn. As a result 385/422 interactions are visible, meaning that the remaining 37 interactions lie on the tails of the distribution. In the bottom plot, a zoom in the range $(-15, 15)$ cm in the difference between the scan and reconstructed end of the muon track z coordinate is shown. As a result 371/422 interactions are visible.

CNNs and GNNs. The full chain of SPINE is shown schematically in Fig. 31 and each stage is summarized in Table VII. This Section reviews each stage of the SPINE reconstruction chain in order, starting from the voxel-level tasks and building up to the high-level particle- and interaction-level classifications.

1. Network training

The training of neural networks requires a labeled dataset, where the true output is known. In the context of LArTPC data, the labeled dataset is provided by the simulation of the detector response to a set of known particles. In order to avoid potential biases due to assumptions of particular physics models governing cosmic ray and neutrino generation, the reconstruction chain is trained using a set of simulated events produced by two physics-agnostic generators:

(i) Multiparticle vertex (MPV): this generator produces a set of particles (between 1 and some specified number) originating from a common vertex. These particles include electrons, positrons, photons, muons, antimuons, both negatively charged and positively charged pions, and protons.

(ii) Multiparticle rain (MPR): this generator produces a set of particles (between one and some specified number) that are isolated from one another. These particles include electrons, positrons, muons, antimuons, and protons.

The MPV generator is meant to emulate a pseudo-neutrino interaction with multiple particles emitted from a common vertex, whereas the MPR generator produces interactions that are similar to those produced by cosmic rays. For each particle, the generators are configured with a range of kinetic energies that can be assigned as well as a maximum particle multiplicity. The kinetic energy is drawn from a uniform distribution within the configurable range.

2. Tomographic reconstruction

The SPINE reconstruction chain operates on a 3D set of points representing the charge depositions in the detector. Due to the detector's readout being intrinsically 2D, some

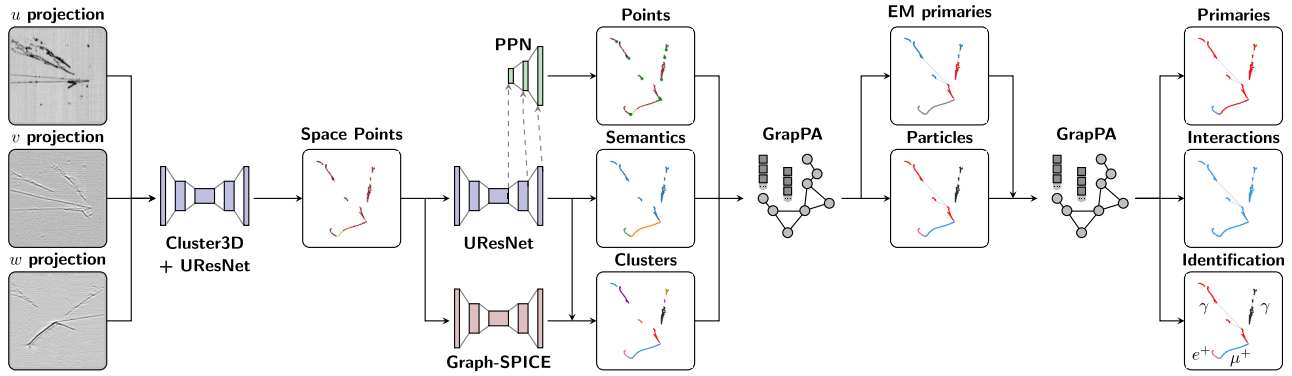


FIG. 31. Schematic architecture of the end-to-end ML-based reconstruction chain (SPINE) used in the ICARUS experiment.

TABLE VII. Summary of the neural networks used in the full ML (SPINE) event reconstruction chain in the ICARUS experiment.

Stage	Type	Description
UResNet Deghost	CNN	Classification of space points as reconstruction artifacts or real charge depositions
UResNet	CNN	Semantic segmentation (voxel-level classification of activity)
PPN	CNN	Prediction of start/end points of showers/tracks
Graph-SPICE	CNN	Coarse clustering of space points into particle fragments
GrapPA-Shower	GNN	Clustering of shower fragments into complete showers
GrapPA-Track	GNN	Clustering of track fragments into complete tracks
GrapPA-Interaction	GNN	Clustering of particles into complete interactions with PID and primary designation

additional processing is required to reconstruct the input 3D image. The process of reconstructing a higher-dimensional image from a set of lower-dimensional projections is generally known as tomographic reconstruction, and is a concept that is widely used in medical imaging. In the context of LArTPC data, the 3D space points are reconstructed from the hits found in each wire plane of the detector.

As described in Sec. V C, the set of hits selected by the Cluster3D algorithm are used in input to the tomographic reconstruction, tuned to be highly efficient at the cost of a higher number of incorrect combinations. These incorrect combinations are known as tomographic artifacts, or "ghost points," and they are especially numerous for tracks parallel to the wire planes where the number of valid combinations of 2D hits is large. Figure 32 shows an example of a 2D event display of a neutrino interaction in the ICARUS detector and the corresponding 3D space points reconstructed by the Cluster3D algorithm. Of particular note is the presence of the ghost points in the 3D space points, which pose a challenge to visually identifying the true activity in the image. The first stage of the point classification task is to identify and remove these ghost points.

3. Point classification

a. Ghost point removal

As a first step, SPINE reconstruction is tasked with the removal of the tomographic artifacts, or ghost points, from Cluster3D ("deghosting"). The neural network used for this implements the U-ResNet architecture, which was first popularized due to its success in biomedical imaging [75], to perform a binary classification of the space points, or semantic segmentation. Modifications were made to adapt U-ResNet to the sparse format of LArTPC data in contrast to regular images where each pixel contains information [76]. After deghosting, the reconstructed charge from Cluster3D is redistributed to enforce conservation of the overall charge. This is done by counting the number of times a given 2D hit is associated with a 3D space point and distributing the charge equally. The charge assigned to each space point is the average of the charge measured on each plane that contributed to the hit. An example image showing the 3D space points after the deghosting step is shown in Fig. 33.

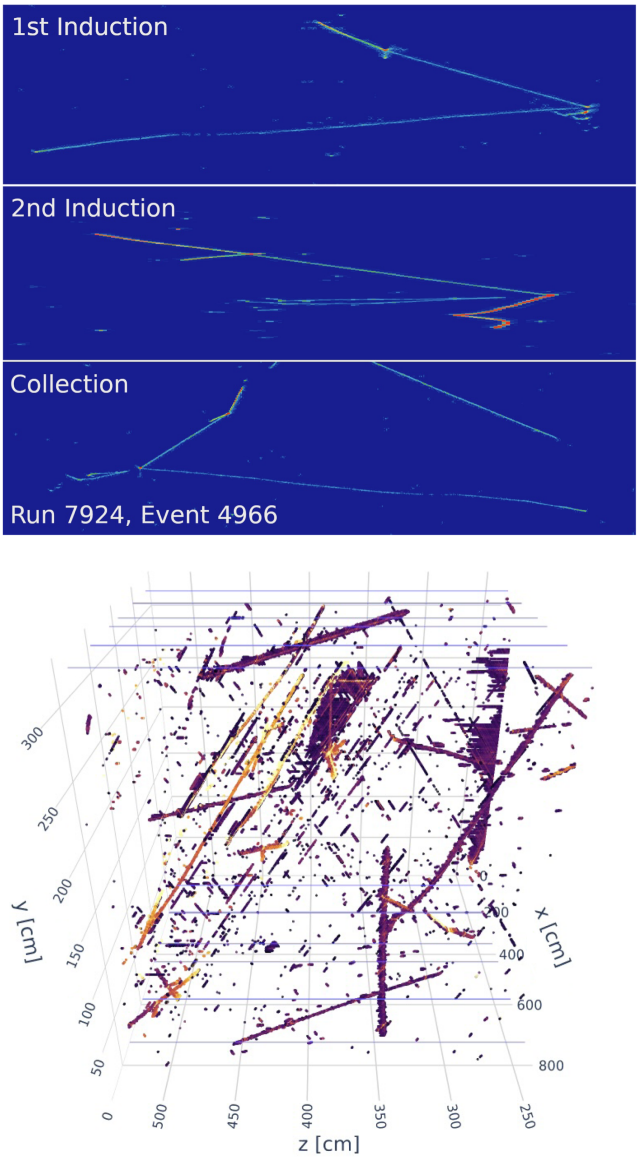


FIG. 32. Example of a 2D event display of the same neutrino interaction in the ICARUS detector (left) and the corresponding 3D space points reconstructed by the Cluster3D algorithm (right). The image shown is a data event from Run1.

b. Semantic segmentation and point proposal network

The next step in the point classification task is to classify the deghosted 3D space points into categories based on the type of parent activity that created them. The neural network used for this is a U-ResNet architecture, this time with five semantic classes of space points: Michel electron, delta ray, electromagnetic shower, track, and low-energy deposition. Each point is assigned a score for each class, and the class with the highest score is taken as the prediction. The semantic type of the space points is used in later stages of the reconstruction.

The point proposal network (PPN) is a neural network that is tasked to propose the 3D space points that are the initial and terminal points of tracklike objects and the initial points of showerlike objects [77]. The PPN shares the same encoding backbone as the semantic segmentation network as shown schematically in Fig. 31. The points of interest proposed by this network are valuable for high-level physics analyses (e.g., particle start/end points and interaction vertex finding) and are used in the clustering tasks in the later stages of the reconstruction. Figure 34 shows the 3D space points colored by semantic class and the points proposed by the PPN.

c. Point clustering

The final step in the point classification task is to cluster the 3D space points in a set of points belonging to the same activity, i.e., track and shower fragments. This network, named Graph-SPICE, also uses a U-ResNet architecture and is tasked with learning an embedding of the space points in a higher-dimensional space where the points belonging to the same activity are close together. In this embedded space, the points belonging to the same semantic class are used to build a k-nearest neighbors graph, which connects the points that are close together. Each edge is given a score based on the feature vectors of the two points it connects and the edge is activated if the score is above a threshold. Space points are clustered into fragments by finding the connected components of the graph. The output of this stage is a set of point clusters, each representing a track or shower fragment.

4. Formation of particles and interactions

a. Particle instance clustering

At this point in the SPINE reconstruction chain, groups of space points corresponding to particle fragments have been identified. The next step is to cluster these groups into particles. The neural network used for this task is a GNN that is tasked with a binary (on/off) edge classification, where an activated edge indicates that the two instances belong to the same particle. Two distinct GNNs are used for particle instance clustering: GrapPA-Track and GrapPA-Shower for tracks and showers, respectively. It is worth noting that these two GNNs have identical architecture and only differ in the tasks for which they have been trained.

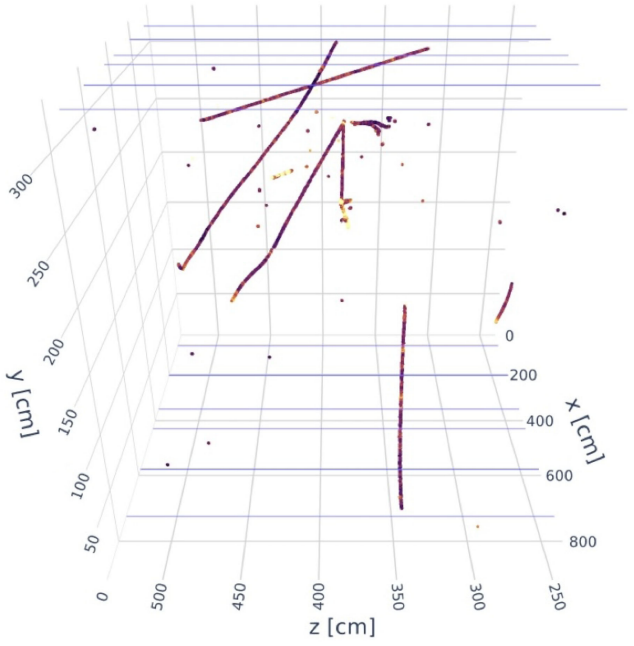


FIG. 33. Example of the 3D space points after the deghosting step. The image shown is a data event from Run1 and the same as in Fig. 32.

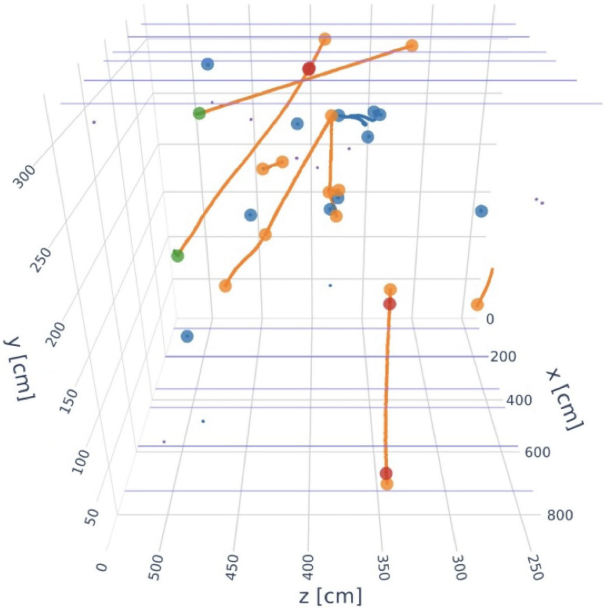


FIG. 34. Example showing the 3D space points after the semantic segmentation and PPN stage. The space points are colored by the semantic type assigned by this stage (notably orange represents tracks, blue represents showers, green represents Michel electrons, and red represents delta rays). Also shown are the points of interest proposed by the PPN colored by the semantic class. The image shown is a data event from Run1 and the same as in Figs. 32 and 33.

The details of the architecture of these GNNs are described in [78].

The node features in each GNN are mostly geometric quantities such as the particle fragment direction vector and the most-likely PPN point. Edge features are also geometry-based and include the distance between the two fragments. The initial state of the graph has activated edges for all node pairs up to a maximal distance. This limitation is motivated by both the computational cost associated with fully connecting an image in a detector as large as ICARUS and by physics concerns: beyond a certain length scale there cannot be legitimate connections. The result of these two GNNs is a set of complete particles of both the track and shower categories. Small gaps in tracks, for example due to track-breaking at the cathode, are expected to be resolved in this stage. The shower-clustering step also identifies the primary shower fragment that is used to identify the shower start point, a critical feature for the interaction-clustering step.

b. Interaction clustering

The final stage of the event reconstruction is to cluster the particles into interactions, defined as a group of particles originating from the same vertex. Particles that are directly connected to the primary vertex are considered primary particles, meaning they are immediate products of the interaction, whereas other particles are labeled as secondaries. The clustering of particles into interactions is easily cast as an edge-classification task, where an activated edge indicates two particles belong to the same interaction. In addition to the interaction clustering, the network is also tasked with predicting for each node the particle type (photon, electron, muon, pion, or proton) and the primary/secondary label. The neural network used for this task is a GNN and is described extensively in [78]. The architecture of this GNN differs from GrapPA-Track and GrapPA-Shower only in that it has more hidden features, which are useful for the extra tasks of PID and primary ID classification. The output of this stage is a full list of interactions in the event, each with a list of primary and secondary particles and their types.

5. Optical flash association

The tool used for associating optical flashes to neutrino interaction candidates reconstructed with SPINE is a package originally developed for MicroBooNE called OpT0Finder. The OpT0Finder algorithm is not available at the moment for the Pandora reconstruction. The event analysis based on Pandora exploits a simpler TPC-PMT matching as described in Sec. VI. OpT0Finder employs a likelihood-based approach to match the optical flash to charge activity associated with an interaction in the event. For each interaction, as defined by the upstream ML reconstruction, the algorithm calculates an allowed range of x offsets that are allowed by the geometry. This reflects our inability to know beforehand the exact time of the interaction. For each of these allowable positions [79], the algorithm calculates a hypothesis flash based on the expected light yield of the interaction, the distribution of the ionization charge in the interaction, and the position of the PMTs in the detector. The hypothesis flash and the observed flash each have a likelihood computed using a joint probability distribution of the photoelectron count in each PMT given the x position of the flash. Maximizing the likelihood between the two allows the assignment of a score to each potential interaction-flash pair.

A matching algorithm then selects the best pairings based on the scores while simultaneously skipping pairs with an already matched flash or interaction. The result of this association process is the assignment of a time stamp to each interaction which was successfully matched to a flash. This time stamp can be used to precisely tag an interaction as being in time or out of time with respect to the beam spill, and therefore can be leveraged as a tool for cosmic rejection.

APPENDIX C: PROFIT

The PROfit framework supports a broad class of oscillation models. For the ν_μ disappearance analysis presented here, we adopt the short-baseline effective two-flavor approximation within the full $3+1$ sterile neutrino model. Thus no matter effects are included and the oscillation probabilities are parametrized in terms of a single mass-squared splitting Δm^2_{41} and three effective mixing angles defined by the elements of the extended PMNS matrix,

$$\begin{aligned}
\sin^2 2\theta_{\mu\mu} &= 4|U_{\mu 4}|^2(1-|U_{\mu 4}|^2),\\
\sin^2 2\theta_{ee} &= 4|U_{e4}|^2(1-|U_{e4}|^2),\\
\sin^2 2\theta_{\mu e} &= 4|U_{e4}|^2|U_{\mu 4}|^2.
\end{aligned} \tag{C1}$$

Only two of the three effective mixing angles are independent. The corresponding survival and transition probabilities take the form,

$$\begin{aligned}
P(\nu_\mu \to \nu_\mu) &= 1-\sin^2 2\theta_{\mu\mu}\sin^2\left(\frac{\Delta m^2_{41}L}{4E}\right),\\
P(\nu_e \to \nu_e) &= 1-\sin^2 2\theta_{ee}\sin^2\left(\frac{\Delta m^2_{41}L}{4E}\right),\\
P(\nu_\mu \to \nu_e) &= \sin^2 2\theta_{\mu e}\sin^2\left(\frac{\Delta m^2_{41}L}{4E}\right).
\end{aligned} \tag{C2}$$

For this ν_μ disappearance-only analysis, we set $|U_{e4}|^2 = 0$ (equivalent to $\theta_{ee} = \theta_{\mu e} = 0$), such that only $P(\nu_\mu \to \nu_\mu)$ is nontrivial. As such there are two distinct degrees of freedom corresponding to the physics parameters of our model, Δm^2_{41} and $\sin^2\theta_{\mu\mu}$.

Charged-current ν_μ interactions are weighted by $P(\nu_\mu \to \nu_\mu)$, charged-current ν_e interactions remain unoscillated, and neutral-current interactions are oscillated using $P(\nu_\mu \to \nu_\mu)$. The latter choice introduces a small approximation given the 0.5% intrinsic ν_e component of the beam and the subdominant NC background in the selected sample (see Fig. 9). Future analyses incorporating full $3+1$, ν_e appearance or explicit NC oscillation channels will expand to the full suite of oscillation probabilities including NC oscillation terms dependent on $|U_{\tau 4}|^2$.

PROfit reads analysis ntuples containing one entry per reconstructed neutrino interaction candidate for both data and simulation. While the framework accommodates an arbitrary number of detectors and oscillation channels, this analysis uses only the ICARUS detector and a single ν_μ disappearance channel. This channel is decomposed into five subchannels based on true interaction type: ν_μ CC, ν_e CC, NC, in-time cosmics (cosmic rays coincident with a neutrino interaction), and out-of-time cosmics (estimated from off-beam data). The subchannels are kept separate when applying oscillation weights and systematic shifts, then collapsed into a single spectrum for the χ^2 calculation.

While oscillation probabilities can be applied event by event, this becomes computationally prohibitive for high-statistics simulation samples. Instead, a binned approach is adopted: oscillation weights are calculated in 150 equally spaced bins of true L/E spanning 0–3 km/GeV. This granularity was validated against full event-by-event reweighting and found to produce negligible differences for this single-detector ICARUS analysis.

The treatment of systematic uncertainties is central to any oscillation analysis. PROfit supports two distinct approaches: covariance matrix systematics and nuisance parameter (or pull term) systematics.

In the nuisance parameter formalism, the effect of N_{nuis} systematic variations on each bin is calculated explicitly. This approach offers greater interpretability, best-fit values and posterior uncertainties for each nuisance parameter are directly extractable, at the cost of increased computational complexity from the N_{nuis} additional fit dimensions (λ^i), each with their own prior central value (λ^i_{CV}) and prior uncertainty (σ^i_λ). The bin-by-bin response to each nuisance parameter is stored as a spline, one for each systematic and each bin, interpolating between precalculated -3σ, -2σ, -1σ, 0, $+1\sigma$, $+2\sigma$, and $+3\sigma$ weight variations.

In the covariance matrix formalism, uncertainties and correlations on a central-value prediction of N_{bins} bins are encoded in a symmetric positive-semidefinite matrix M_{syst}. A single covariance matrix can capture the combined effect of many systematic sources without introducing additional fit parameters, appearing only once in the test statistic and enabling efficient evaluation. The drawback is that systematics are treated as symmetric and Gaussian, and their individual contributions are less transparent in postfit results. The covariance matrix is constructed from N_U random universe throws as $C_{ij} = (1/N_U)\sum_u (h_i^u - h_i^{\text{CV}})(h_j^u - h_j^{\text{CV}})$, where h_i^u is the content of bin i in universe u and h_i^{CV} is the central-value prediction. N_U is 1,000 for flux systematics, and 100 for those interaction systematics treated using covariance.

PROfit allows analyses to operate in either a pure covariance or pure nuisance parameter mode, or in a hybrid configuration where selected systematics are incorporated via covariance matrices while others are treated as explicit nuisance parameters, leveraging the strengths of both approaches. In all cases, systematic effects enter through the test statistic minimized in the fit. This analysis employs the CNP χ^2_{CNP} [72] as the metric minimized. This metric encodes the effects of both the covariance and nuisance parameters,

$$\chi^2(\Delta m^2_{41}, \sin^2 2\theta_{\mu\mu}, \vec{\lambda}) = \Delta^T M^{-1}_{\text{CNP}} \Delta + \sum_{k=1}^{N_{\text{nuis}}} \left(\frac{\lambda^k - \lambda^k_{\text{CV}}}{\sigma^k_\lambda} \right)^2, \tag{C3}$$

where $\Delta \equiv O - P(\Delta m^2_{41}, \sin^2 2\theta_{\mu\mu}, \vec{\lambda})$ is the residual between observed (O) and predicted spectra (P), and M_{CNP} is the combined statistical and systematic covariance matrix $M^{\text{CNP}}_{\text{stat}} + M_{\text{syst}}$. We note $M^{\text{CNP}}_{\text{stat}}$ is a function of both physics and nuisance parameters itself through the prediction, $P(\Delta m^2_{41}, \sin^2 2\theta_{\mu\mu}, \vec{\lambda})$, and defined as

$$M^{\text{CNP}}_{\text{stat}}(\Delta m^2_{41}, \sin^2 2\theta_{\mu\mu}, \vec{\lambda})(i,j) \equiv \frac{3\delta_{ij}}{1/O^i + 2/P^i}. \tag{C4}$$

All 13 flux uncertainties, a number of interaction model uncertainties (as laid out in Table V), and the effects of finite statistics of the simulation samples are treated via the covariance matrix approach. Since oscillation fits are performed in bins of reconstructed neutrino energy (Fig. 23), all covariance matrices are constructed in this same binning. Fractional covariance matrices are stored and converted to absolute covariances at the time of χ^2 evaluation, since oscillation weights and nuisance parameter shifts modify the predicted spectrum.

The remaining interaction model, detector uncertainties and normalization uncertainties are treated as nuisance parameters. For each systematic, spectra corresponding to $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ variations are used to construct splines describing the bin-by-bin response as a continuous function of the nuisance parameter. Some variations represent discrete alternative model choices rather than continuous parameters; in these cases the nuisance parameter is bounded to prevent extrapolation beyond $\pm 1\sigma$. Normalization uncertainties are also implemented as nuisance parameters, allowing the fit to report preferred values and uncertainties for each.

PROfit employs a four-stage minimization procedure. First, Latin hypercube sampling of 5,000 points across the full parameter space (physics and nuisance parameters) identifies regions of low χ^2. A particle swarm optimization [80] then refines the most promising candidates, with 25 particles and 250 iterations converging rapidly on local minima. The best 15 candidates from both the particle swarm and Latin hypercube are subsequently passed to the L-BFGS-B gradient-based minimizer [81] to locate precise minima. Finally, to guard against convergence to a local rather than global minimum, PROfit exploits the periodic structure of the oscillation probability: for a given L/E, the transformation $\Delta m^2_{41} \to \Delta m^2_{41} + 4n\pi E/L$ leaves $\sin^2(\Delta m^2_{41} L/4E)$ invariant, for fixed L/E. A dedicated scan around harmonics of the current best-fit Δm^2_{41} is therefore performed, which is particularly important for identifying fast oscillations at large Δm^2_{41}.

An example fit to Asimov MC (the MC prediction with no statistical fluctuations) with an injected oscillation signal is shown in Fig. 35, including pre- and postfit spectra, uncertainties, and nuisance parameter values. The fit successfully recovers the injected signal.

Prefit uncertainty bands are obtained by generating an ensemble of pseudoexperiments in which nuisance parameters are sampled from their prior distributions and correlated fluctuations are drawn from the systematic covariance matrix; 5000 pseudoexperiments are used. Postfit uncertainty bands must account for potentially correlated shifts in the nuisance parameters at the best-fit point. To capture these correlations, an adaptive Markov Chain Monte Carlo variant of the Metropolis Hastings method [82] is used to sample the nuisance parameter space in the vicinity of the best fit, with 25,000 burn-in steps followed by 20,000 sampling steps. In both cases, the displayed error bars represent the central 68% interval of the resulting pseudoexperiment distribution in each bin.

Postfit uncertainties on individual nuisance parameters are obtained separately by profiling over all other parameters while scanning each parameter of interest, mapping the χ^2 surface in the vicinity of the best fit.

After the best fit parameters values are identified, confidence intervals on the oscillation parameters are constructed using the Feldman-Cousins (FC) procedure [73], the standard frequentist approach in the neutrino community. This method employs a likelihood-ratio ordering to construct confidence belts via a Neyman construction at each point in parameter space. The approach guarantees correct frequentist coverage by construction and naturally handles physical boundaries on parameters, providing a unified treatment that transitions smoothly between upper limits (when data are consistent with the null hypothesis) and two-sided allowed regions (when a signal is observed). At each grid point in the $(\Delta m^2_{41}, \sin^2 2\theta_{\mu\mu})$ plane, an ensemble of pseudoexperiments is generated by sampling nuisance parameters from their prior distributions, drawing correlated fluctuations from the systematic covariance matrix, and applying Poisson fluctuations to produce simulated datasets. The nuisance parameters are sampled using their prior uncertainties according to the Highland-Cousins approach [83].

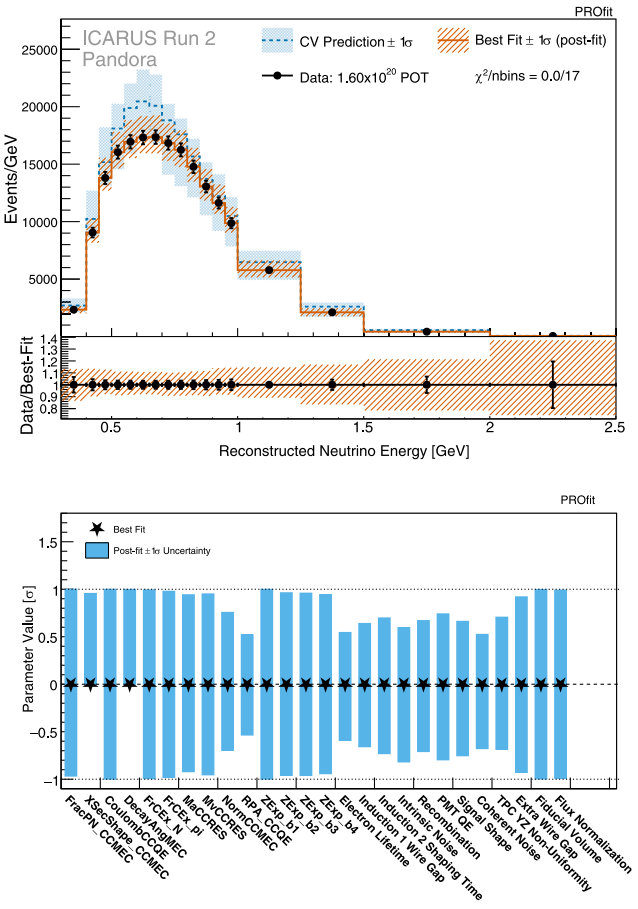


FIG. 35. Example PROfit output for an Asimov dataset with injected signal parameters $\Delta m^2_{41} = 4$ eV^2 and $\sin^2 2\theta_{\mu\mu} = 0.25$. (Top) Postfit spectrum compared to the prefit central value prediction and data. (Bottom) Nuisance parameter best-fit values and 1σ uncertainties obtained via profiling. Prefit nuisance parameter values are by definition 0 (no pull), with prefit uncertainty defined as $\pm 1\sigma$ (horizontal dotted lines).

At each point in the $(\Delta m^2_{41}, \sin^2 2\theta_{\mu\mu})$ grid, the test statistic is the likelihood ratio expressed as

$$\Delta\chi^2 = \chi^2(\Delta m^2_{41}, \sin^2 2\theta_{\mu\mu}, \hat{\vec{\lambda}}) - \chi^2(\widehat{\Delta m^2_{41}}, \widehat{\sin^2 2\theta_{\mu\mu}}, \hat{\vec{\lambda}}'), \tag{C5}$$

where the first term is the χ^2 minimized over nuisance parameters $\vec{\lambda}$ with the oscillation parameters fixed at the grid point, and the second term is the global best fit with all parameters, including physics parameters, free. For each pseudoexperiment generated at a given grid point, this $\Delta\chi^2$ is computed, building an expected distribution of the test

statistic under the hypothesis that the true parameters lie at that grid point.

When analyzing data, the observed $\Delta\chi^2_{\text{data}}$ is compared to this ensemble: the p value is the fraction of pseudoexperiments with $\Delta\chi^2$ exceeding the observed value. A grid point is excluded at confidence level $1-\alpha$ if fewer than a fraction α of pseudoexperiments yield a larger $\Delta\chi^2$ than the data. Repeating this procedure across the full parameter grid produces Feldman-Cousins-corrected exclusion contours or, in the presence of a signal, allowed regions with correct frequentist coverage. For this analysis, between 1,500 and 6,500 pseudouniverses were sampled across grids for both Pandora and SPINE analyses, with the higher numbers being sampled closer to the resulting confidence interval boundaries to improve the statistical precision closer to final result.

In addition to the Feldman-Cousins confidence intervals constructed from data, the expected sensitivity of the analysis is quantified in two complementary ways. First, the null hypothesis Asimov dataset is analyzed as a representative pseudoexperiment, yielding the Asimov sensitivity contour. Second, an ensemble of 5,000 pseudoexperiments are generated under the null hypothesis (no sterile neutrino oscillations) and Feldman-Cousins confidence intervals are constructed for each. From the ensemble of pseudoexperiments resulting in exclusion contours, we extract the median expected exclusion sensitivity as well as $\pm1\sigma$ and $\pm2\sigma$ bands, representing the expected spread of exclusion limits if the null hypothesis is indeed true.

A complete presentation of results therefore includes three elements: the observed exclusion contour from data, the Asimov sensitivity, and the distribution of expected exclusions under the null hypothesis. The Asimov contour provides a single representative sensitivity, while the ensemble bands indicate the range of statistical fluctuations expected in the measurement. Comparison of the observed contour to both of these expectations is essential for assessing whether any deviation from the null hypothesis is statistically significant or consistent with expected fluctuations.

Signal unblinding follows a predetermined staged procedure in PROfit in which fits to both the SPINE and Pandora selections are performed simultaneously. Validation checks are evaluated at each stage before proceeding to ensure a robust result. The unblinding stages are:

(1) Fitter validation: Verify that the fit executed without errors, the number of empty bins in data and simulation is acceptable, the covariance matrix is positive semidefinite, and the best-fit χ^2 and parameter values are physical.
(2) Goodness of fit: Evaluate the best-fit χ^2 and corresponding p value, requiring $p > 0.05$ to proceed. No best-fit values themselves are known at this point.
(3) Blinded pull terms: Inspect the distribution of nuisance parameter pulls without identifying individual parameters. Pulls exceeding 3σ would halt the procedure; pulls exceeding 2σ prompt further investigation before proceeding. In mock data studies, pulls rarely exceed 1σ.
(4) Unblinded pull terms: Inspect the best-fit values, 1σ uncertainties, and profiles for each nuisance parameter. Check and study any unexpected postfit correlations.
(5) Oscillation parameters: Examine the best-fit values of Δm^2_{41} and $\sin^2 2\theta_{\mu\mu}$, and evaluate the Feldman-Cousins p value testing the no-oscillation hypothesis.
(6) Confidence intervals: Construct the Feldman-Cousins confidence region. If the p value from the previous step exceeds 0.05, a 90% exclusion limit is drawn; otherwise, an allowed region is reported.
(7) Comparison to expectation: Compare the observed result to the Asimov sensitivity and the ensemble-derived $\pm1\sigma$ and $\pm2\sigma$ bands.

Stages 1–3 were first executed on a small 10% subset of the available data as an initial validation. Following successful completion, the full procedure was repeated on the complete Run 2 dataset.

[1] A. Aguilar *et al.*, Evidence for neutrino oscillations from the observation of $\bar{\nu}_e$ appearance in a $\bar{\nu}_\mu$ beam, Phys. Rev. D **64**, 112007 (2001).
[2] A. A. Aguilar-Arevalo *et al.*, Improved search for $\bar{\nu}_\mu \to \bar{\nu}_e$ oscillations in the MiniBooNE experiment, Phys. Rev. Lett. **110**, 161801 (2013).
[3] F. Kaether, W. Hampel, G. Heusser, J. Kiko, and T. Kirsten, Reanalysis of the GALLEX solar neutrino flux and source experiments, Phys. Lett. B **685**, 47 (2010).
[4] J. N. Abdurashitov *et al.*, Measurement of the solar neutrino capture rate with gallium metal, III: Results for the 2002–2007 data-taking period, Phys. Rev. C **80**, 015807 (2009).
[5] V. V. Barinov *et al.*, Search for electron-neutrino transitions to sterile states in the BEST experiment, Phys. Rev. C **105**, 065502 (2022).
[6] A. P. Serebrov *et al.*, Search for sterile neutrinos with the Neutrino-4 experiment and measurement results, Phys. Rev. D **104**, 032003 (2021).

[7] R. Abbasi *et al.*, Search for an eV-scale sterile neutrino using improved high-energy ν_μ event reconstruction in IceCube, Phys. Rev. Lett. **133**, 201804 (2024).

[8] R. Acciarri *et al.*, A proposal for a three detector Short-Baseline Neutrino oscillation program in the fermilab booster neutrino beam, arXiv:1503.01520.

[9] C. Rubbia, The liquid argon time projection chamber: A new concept for neutrino detectors, Report Nos. CERN-EP-INT-77-08, CERN-EP-77-08, 1977, https://inspirehep.net/literature/857394.

[10] S. Amerio *et al.*, Design, construction and tests of the ICARUS T600 detector, Nucl. Instrum. Methods Phys. Res., Sect. A **527**, 329 (2004).

[11] P. Abratenko *et al.*, ICARUS at the Fermilab Short-Baseline Neutrino program: initial operation, Eur. Phys. J. C **83**, 467 (2023).

[12] R. Abbasi *et al.*, Methods and stability tests associated with the sterile neutrino search using improved high-energy ν_μ event reconstruction in IceCube, Phys. Rev. D **110**, 092009 (2024).

[13] P. Adamson *et al.*, Search for sterile neutrinos in MINOS and MINOS+ using a two-detector fit, Phys. Rev. Lett. **122**, 091803 (2019).

[14] M. A. Acero *et al.*, Dual-baseline search for active-to-sterile neutrino oscillations in NOvA, Phys. Rev. Lett. **134**, 081804 (2025).

[15] G. Cheng *et al.*, Dual baseline search for muon antineutrino disappearance at $0.1\ \mathrm{eV}^2 < \Delta m^2 < 100\ \mathrm{eV}^2$, Phys. Rev. D **86**, 052009 (2012).

[16] F. Dydak *et al.*, A search for muon-neutrino ν_μ oscillations in the Δm^2 range $0.3 - 90\ \mathrm{eV}^2$, Phys. Lett. B **134**, 281 (1984).

[17] I. E. Stockdale *et al.*, Limits on muon neutrino oscillations in the mass range $55 < \Delta m^2 < 800\ \mathrm{eV}^2$, Phys. Rev. Lett. **52**, 1384 (1984).

[18] E. L. Snider and G. Petrillo, LArSoft: Toolkit for simulation, reconstruction and analysis of liquid argon TPC neutrino detectors, J. Phys. Conf. Ser. **898**, 042057 (2017).

[19] J. S. Marshall and M. A. Thomson, The Pandora Software development kit for pattern recognition, Eur. Phys. J. C **75**, 439 (2015).

[20] F. Drielsma, K. Terao, L. Domine, and D. H. Koh, Scalable, end-to-end, deep-learning-based data reconstruction chain for particle imaging detectors, arXiv:2102.01033.

[21] S. Schael *et al.*, Precision electroweak measurements on the Z resonance, Phys. Rep. **427**, 257 (2006).

[22] Joachim Kopp, Pedro A. N. Machado, Michele Maltoni, and Thomas Schwetz, Sterile neutrino oscillations: The global picture, J. High Energy Phys. 05 (2013) 050.

[23] Carlo Giunti and T. Lasserre, eV-scale sterile neutrinos, Annu. Rev. Nucl. Part. Sci. **69**, 163 (2019).

[24] J. M. Hardin, I. Martinez-Soler, A. Diaz, M. Jin, N. W. Kamp, C. A, Argüelles, J. M. Conrad, and M. H. Shaevitz, New Clues about light sterile neutrinos: Preference for models with damping effects in global fits, J. High Energy Phys. 09 (2023) 058.

[25] P. Abratenko *et al.*, Search for light sterile neutrinos with two neutrino beams at MicroBooNE, Nature (London) **648**, 64 (2025).

[26] Himal Acharya *et al.*, Sterile-neutrino search based on 259 days of KATRIN data, Nature (London) **648**, 70 (2025).

[27] A. A. Aguilar-Arevalo *et al.*, The neutrino flux prediction at MiniBooNE, Phys. Rev. D **79**, 072002 (2009).

[28] MicroBooNE Collaboration, Booster neutrino flux prediction at MicroBooNE (2018), https://www.osti.gov/servlets/purl/1573216.

[29] S. Agostinelli, J. Allison, K. Amako *et al.*, Geant4—a simulation toolkit, Nucl. Instrum. Methods Phys. Res., Sect. A **506**, 250 (2003).

[30] David W. Schmitz, A measurement of hadron production cross sections for the simulation of accelerator neutrino beams and a search for ν_μ to ν_e Oscillations in the δm^2 about equals $1 - \mathrm{eV}^2$ region, Ph.D. thesis, Columbia U., 2008.

[31] I. Chemakin *et al.*, Pion production by protons on a thin beryllium target at 6.4-Ge, 12.3-GeV/c, and 17.5-GeV/c incident proton momenta, Phys. Rev. C **77**, 015209 (2008); **77**, 049903(E) (2008).

[32] B. Ali-Mohammadzadeh *et al.*, Design and implementation of the new scintillation light detection system of ICARUS T600, J. Instrum. **15**, T10007 (2020).

[33] A. Aduszkiewicz *et al.*, Design and implementation of the cosmic ray tagger system for the ICARUS detector at FNAL, J. Instrum. **20**, T04002 (2025).

[34] F. Abd Alrahman *et al.*, Operation of the trigger system for the ICARUS detector at Fermilab, J. Instrum. **20**, P10044 (2025).

[35] C. Andreopoulos *et al.*, The GENIE neutrino Monte Carlo generator, Nucl. Instrum. Methods Phys. Res., Sect. A **614**, 87 (2010).

[36] GENIE releases, https://genie-mc.github.io/releases.html,

[37] Babak Abi *et al.*, Deep Underground Neutrino Experiment (DUNE), Far detector technical design report, volume I introduction to DUNE, J. Instrum. **15**, T08008 (2020).

[38] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw, CORSIKA: A Monte Carlo code to simulate extensive air showers, Report No. FZKA-6019, 1998.

[39] P. Abratenko *et al.*, Calibration and simulation of ionization signal and electronics noise in the ICARUS liquid argon time projection chamber, J. Instrum. **20**, P01032 (2025).

[40] R. Acciarri *et al.*, A study of electron recombination using highly ionizing particles in the ArgoNeuT liquid argon TPC, J. Instrum. **8**, P08005 (2013).

[41] M. Antonello *et al.*, Study of space charge in the ICARUS T600 detector, J. Instrum. **15**, P07001 (2020).

[42] P. Abratenko *et al.* (ICARUS Collaboration), Angular dependent measurement of electron-ion recombination in liquid argon for ionization calorimetry in the ICARUS liquid argon time projection chamber, J. Instrum. **20**, P01033 (2025).

[43] R. Acciarri *et al.*, The Pandora multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector, Eur. Phys. J. C **78**, 82 (2018).

[44] Adam Abed Abud *et al.*, Reconstruction of interactions in the ProtoDUNE-SP detector with Pandora, Eur. Phys. J. C **83**, 618 (2023).

[45] S. Navas (Particle Data Group), Review of particle physics, Phys. Rev. D **110**, 030001 (2024).
[46] A. Bodek and T. Cai, Removal energies and final state interaction in lepton nucleus scattering, Eur. Phys. J. C **79**, 293 (2019).
[47] Y. Coadou, *Boosted Decision Trees* (World Scientific Publishing Co. Pte. Ltd., Singapore, 2022), Chap. 2, pages 9–58.
[48] X.-G. Lu, L. Pickering, S. Dolan, G. Barr, D. Coplowe, Y. Uchida, D. Wark, M. O. Wascko, A. Weber, and T. Yuan, Measurement of nuclear effects in neutrino interactions with minimal dependence on neutrino energy, Phys. Rev. C **94**, 015503 (2016).
[49] Wissal Filali, Laura Munteanu, and Stephen Dolan, Benchmarking neutrino interaction models via a comparative analysis of kinematic imbalance measurements from the T2K, MicroBooNE, and MINERvA experiments, Phys. Rev. D **111**, 032009 (2025).
[50] O. Hen, G. A. Miller, E. Piasetzky, and L. B. Weinstein, Nucleon-nucleon correlations, short-lived excitations, and the quarks within, Rev. Mod. Phys. **89**, 045002 (2017).
[51] L. B. Weinstein, E. Piasetzky, D. W. Higinbotham, J. Gomez, O. Hen, and R. Shneor, Short range correlations and the EMC effect, Phys. Rev. Lett. **106**, 052301 (2011).
[52] J. Nieves, I. Ruiz Simo, and M. J. Vicente Vacas, Inclusive charged–current neutrino–nucleus reactions, Phys. Rev. C **83**, 045501 (2011).
[53] R. Gonzaléz-Jiménez, G. D. Megias, M. B. Barbaro, J. A. Caballero, and T. W. Donnelly, Extensions of superscaling from relativistic mean field theory: The SuSAv2 model, Phys. Rev. C **90**, 035501 (2014).
[54] G. D. Megias, J. E. Amaro, M. B. Barbaro, J. A. Caballero, T. W. Donnelly, and I. Ruiz Simo, Charged-current neutrino-nucleus reactions within the superscaling meson-exchange current approach, Phys. Rev. D **94**, 093004 (2016).
[55] I. Ruiz Simo, J. E. Amaro, M. B. Barbaro, A. De Pace, J. A. Caballero, and T. W. Donnelly, Relativistic model of 2p-2h meson exchange currents in (anti)neutrino scattering, J. Phys. G **44**, 065105 (2017).
[56] I. Ruiz Simo, J. E. Amaro, M. B. Barbaro, A. De Pace, J. A. Caballero, G. D, Megias, and T. W. Donnelly, Emission of neutron–proton and proton–proton pairs in neutrino scattering, Phys. Lett. B **762**, 124 (2016).
[57] Aaron S. Meyer, Minerba Betancourt, Richard Gran, and Richard J. Hill, Deuterium target data for precision neutrino-nucleus cross sections, Phys. Rev. D **93**, 113015 (2016).
[58] Ch. Berger and L. M. Sehgal, Lepton mass effects in single pion production by neutrinos, Phys. Rev. D **76**, 113004 (2007).
[59] Krzysztof M. Graczyk and Jan T. Sobczyk, Form factors in the quark resonance model, Phys. Rev. D **77**, 053001 (2008); **79**, 079903(E) (2009).
[60] Ch. Berger and L. M. Sehgal, Partially conserved axial vector current and coherent pion production by low energy neutrinos, Phys. Rev. D **79**, 053003 (2009).
[61] T. Yang, C. Andreopoulos, H. Gallagher, K. Hoffmann, and P. Kehayias, A hadronization model for few-GeV neutrino interactions, Eur. Phys. J. C **63**, 1 (2009).
[62] A Bodek and U. K. Yang, Higher twist, xi(omega) scaling, and effective LO PDFs for lepton scattering in the few GeV region, J. Phys. G **29**, 1899 (2003).
[63] S. A. Dytman and A. S. Meyer, Final state interactions in GENIE, AIP Conf. Proc. **1405**, 213 (2011).
[64] Steven Dytman, Yoshinari Hayato, Roland Raboanary, J. T. Sobczyk, Julia Tena Vidal, and Narisoa Vololoniaina, Comparison of validation methods of simulations for final state interactions in hadron production experiments, Phys. Rev. D **104**, 053006 (2021).
[65] Júlia Tena-Vidal *et al.*, Hadronization model tuning in GENIE v3, Phys. Rev. D **105**, 012009 (2022).
[66] Genie manual, https://www.overleaf.com/read/rqmbqzzvsvmb#5ab475.
[67] NuSystematics software, https://github.com/NuSystematics/nusystematics.
[68] Yoshinari Hayato and Luke Pickering, The NEUT neutrino interaction simulation program library, Eur. Phys. J. Special Topics **230**, 4469 (2021).
[69] Tomasz Golan, Cezary Juszczak, and Jan T. Sobczyk, Final state interactions effects in neutrino-nucleus interactions, Phys. Rev. C **86**, 015505 (2012).
[70] O. Buss, T. Gaitanos, K. Gallmeister, H. van Hees, M. Kaskulov, O. Lalakulich, A. B. Larionov, T. Leitner, J. Weil, and U. Mosel, Transport-theoretical description of nuclear reactions, Phys. Rep. **512**, 1 (2012).
[71] PROfit, https://github.com/markrosslonergan/Elephant_Vanishes, 2025.
[72] Xiangpan Ji, Wenqiang Gu, Xin Qian, Hanyu Wei, and Chao Zhang, Combined Neyman–Pearson chi-square: An improved approximation to the Poisson-likelihood chi-square, Nucl. Instrum. Methods Phys. Res., Sect. A **961**, 163677 (2020).
[73] Gary J. Feldman and Robert D. Cousins, Unified approach to the classical statistical analysis of small signals, Phys. Rev. D **57**, 3873 (1998).
[74] C. Anderson *et al.*, The ArgoNeuT detector in the NuMI low-energy beam line at Fermilab, J. Instrum. **7**, P10019 (2012).
[75] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation* (Springer International Publishing, Switzerland, 2015), pp. 234–241.
[76] Laura Dominé and Kazuhiro Terao, Scalable deep convolutional neural networks for sparse, locally dense liquid argon time projection chamber data, Phys. Rev. D **102**, 012005 (2020).
[77] Laura Dominé, Pierre Côte de Soux, François Drielsma, Dae Heun Koh, Ran Itay, Qing Lin, Kazuhiro Terao, Ka Vang Tsang, and Tracy L. Usher, Point proposal network for reconstructing 3d particle endpoints with subpixel precision in liquid argon time projection chambers, Phys. Rev. D **104**, 032004 (2021).
[78] François Drielsma, Qing Lin, Pierre Côte de Soux, Laura Dominé, Ran Itay, Dae Heun Koh, Bradley J. Nelson, Kazuhiro Terao, Ka Vang Tsang, and Tracy L, Usher, Clustering of electromagnetic showers and particle interactions with graph neural networks in liquid argon time projection chambers, Phys. Rev. D **104**, 072004 (2021).

[79] This is not actually a brute-force algorithm, but rather an optimization algorithm that profiles across the drift direction.
[80] J. Kennedy and R. Eberhart, Particle swarm optimization, in *Proceedings of ICNN'95—International Conference on Neural Networks* (1995), Vol. 4, pp. 1942–1948.
[81] Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. **16**, 1190 (1995).
[82] Heikki Haario, Eero Saksman, and Johanna Tamminen, An adaptive metropolis algorithm, Bernoulli **7**, 223 (2001).
[83] Robert D. Cousins and Virgil L. Highland, Incorporating systematic uncertainties into an upper limit, Nucl. Instrum. Methods Phys. Res., Sect. A **320**, 331 (1992).