

Event generator tuning as a robustness test

Jean Wolfs^{*} and Chris M. Marshall[†]

Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA



(Received 29 September 2025; accepted 10 November 2025; published 26 November 2025)

Neutrino oscillation experiments use Monte Carlo event generators to predict neutrino-nucleus interactions. Cross section uncertainties are typically implemented by varying the parameters of the model(s) used in the generator. We study the performance of two commonly used model configurations of the GENIE generator (G18_10a_02_11a and AR23_0i_00_000) and their uncertainties by tuning parameters to cross section data, and then comparing the resulting tuned prediction to a suite of other measurements from T2K, MicroBooNE, and MINERvA. This reveals whether the model can simultaneously describe several datasets, as well as whether the uncertainties are adequately robust. We find that G18 and especially AR23 are reasonable in predicting lower-energy measurements from T2K and MicroBooNE, but unable to describe MINERvA data, and discuss the implications for short-baseline oscillation searches. We attempt to replicate a tuning procedure developed by MicroBooNE using several different measurements, and find substantially different results depending on which measurement is used, and that the MicroBooNE tune does not agree with other measurements. We conclude that the SBN experiment should *not* tune its generator to external data.

DOI: [10.1103/9dxd-871x](https://doi.org/10.1103/9dxd-871x)

I. INTRODUCTION

The discovery of neutrino oscillations [1,2] implies that neutrinos have mass and that there is nontrivial mixing between the neutrino flavor states and the neutrino mass states. In an approximation with only two neutrino flavors, α and β , and two neutrino masses, m_1 and m_2 , the vacuum probability of neutrino oscillations is,

$$P_{\alpha \rightarrow \beta}(E_\nu) = \sin^2(2\theta) \sin^2\left(\frac{\Delta m^2 L}{4E_\nu}\right) \quad (1)$$

where θ is the mixing angle, $\Delta m^2 = (m_2^2 - m_1^2)$ is the mass splitting, L is the baseline, and E_ν is the neutrino energy.

With three flavors (ν_e , ν_μ , ν_τ), there are three mixing angles (θ_{12} , θ_{23} and θ_{13}) and three mass splittings (Δm_{21}^2 , Δm_{32}^2 and Δm_{31}^2), only two of which are independent because $\Delta m_{32}^2 = \Delta m_{31}^2 - \Delta m_{21}^2$. Given the measurements of $\Delta m_{21}^2 = (7.53 \pm 0.18) \times 10^{-5} \text{ eV}^2$ and $\Delta m_{32}^2 = (2.455 \pm 0.028) \times 10^{-3} \text{ eV}^2$ [3], Eq. (1) is maximized

and neutrino oscillations are most probable when $L/E_\nu \simeq 16,000 \text{ km/GeV}$ and $L/E_\nu \simeq 500 \text{ km/GeV}$.

Accelerator-based neutrino oscillation experiments measure Eq. (1) indirectly by measuring the interaction rate for some neutrino flavor as a function of reconstructed neutrino energy,

$$\frac{dN_\beta}{dE_{\text{rec}}} = \int \Phi_\alpha(E_\nu) P_{\alpha \rightarrow \beta}(E_\nu) \sigma_\beta(E_\nu) T_\beta(E_\nu, E_{\text{rec}}) dE_\nu \quad (2)$$

where E_{rec} is the reconstructed neutrino energy, Φ_α is the unoscillated neutrino flux, $P_{\alpha \rightarrow \beta}$ is the oscillation probability from Eq. (1), σ_β is the cross section for ν_β , and T_β is the detector response. A measurement of Eq. (2) is called an appearance measurement when $\alpha \neq \beta$ and a disappearance measurement when $\alpha = \beta$.

Of the variables on the right-hand side of Eq. (2), E_{rec} is measured in the experiment, while Φ_α , σ_β , and T_β are estimated using models, often informed by external data. The neutrino flux Φ_α is typically estimated using a simulation of the neutrino beam. In the case of the Booster Neutrino Beam (BNB) experiments at Fermilab, a Geant4-based software package is used to predict the flux [4,5]. The systematic uncertainty in the BNB flux estimation is dominated by uncertainties in the rate and energy of secondary mesons produced in proton interactions with the target, although uncertainties in beam focusing are also relevant.

The second estimated variable in Eq. (2), the cross section σ_β , is calculated by a neutrino event generator.

^{*}Contact author: jwolfs@ur.rochester.edu

[†]Contact author: chris.marshall@rochester.edu

Modern experiments use GENIE [6] or NEUT [7–9] in their primary analyses, but may use GiBUU [10] or NuWro [11] in other studies. Each generator uses a series of models to simulate the neutrino interaction itself, as well as the kinematics of the initial nuclear state and final state interactions (FSI) between outgoing particles and the residual nucleus. Finally, the detector response function, T_β , relates true neutrino energy to a detector observable, typically reconstructed neutrino energy. In addition to detector effects, T_β is sensitive to exclusive neutrino cross section estimates because the detector response can vary for different final states.

Since the mid-1990s, several experiments have made measurements of Eq. (1) that are inconsistent with three-flavor predictions. The LSND experiment measured a 3.8σ excess of $\bar{\nu}_e$ events in a reactor-based $\bar{\nu}_\mu$ beam [12]. Then, MiniBooNE measured a 4.5σ excess of ν_e and a 2.8σ excess of $\bar{\nu}_e$ events in accelerator-based ν_μ and $\bar{\nu}_\mu$ beams, respectively. Together, LSND and MiniBooNE’s measurements are known as the “short baseline anomaly.” Meanwhile, the GALLEX and SAGE experiments measured the rate of neutrino capture on gallium using the radioactive sources ^{51}Cr and ^{37}Ar and found a combined deficit of $3\sigma\nu_e$ events [13,14]. This result is known as the “gallium anomaly.”

The short baseline and gallium anomalies could be explained by the existence of a fourth neutrino mass state with mass splittings $\Delta m_{43}^2 \approx \Delta m_{42}^2 \approx \Delta m_{41}^2 \sim \mathcal{O}(1 \text{ eV}^2)$. As measurements of the Z boson width strongly disfavor the existence of more than three weakly interacting neutrinos [15], this fourth neutrino would be a “sterile neutrino” that could only interact via the gravitational force.

Experiments like LSND and MiniBooNE that use accelerator-based neutrino beams can be categorized by the ratio of their baseline to their energy, L/E_ν in Eq. (1). Long-baseline (LBL) experiments have L/E_ν close to the

oscillation maximum at $L/E_\nu \simeq 500 \text{ km/GeV}$, and so are designed to measure standard oscillations due to mixing with the third mass state. On the other hand, short baseline (SBL) experiments have L/E_ν close to the oscillation minimum at $L/E_\nu \simeq 1 \text{ km/GeV}$, and so are designed to search for nonstandard oscillations mediated by heavy sterile neutrinos.

LBL experiments typically operate two detectors: a near detector (ND) close to the neutrino source to measure the initial, unoscillated beam and a far detector (FD) at a distance L from the neutrino source to measure oscillations. Most SBL experiments are conducted with only a single detector because it is difficult to position a detector close enough to the neutrino source that rapid, sterile oscillations would be truly negligible. As a consequence, single-detector SBL experiments must estimate Φ_α , σ_β , and T_β without the aid of ND constraints, leading to larger systematic uncertainties in their measurements of Eq. (2).

Table I summarizes a number of past, present, and future accelerator-based neutrino experiments by listing the approximate range of L/E_ν and Δm^2 to which they are sensitive in both the ND and the FD. The energy ranges in the table correspond roughly to where the event rate is at least 10% of the peak. The baseline is taken to be the distance to the production target; for near detectors this is a rough approximation for L/E_ν because line source effects are not negligible. Figure 1 plots the ranges of L/E_ν and Δm^2 from Table I for those detectors with $\Delta m^2 > 10^{-2} \text{ eV}^2$. The upper range of Δm^2 corresponds to where the experiment could observe spectral distortions due to a sterile neutrino; experiments could also observe averaged effects due to rapid oscillations at higher Δm^2 .

Two of the experiments from Table I were designed to test the short baseline anomaly. MicroBooNE was an 85-ton liquid argon time projection chamber (LArTPC) positioned at $L = 470 \text{ m}$ from the BNB target that operated from 2015 to 2021. MicroBooNE searched for ν_e

TABLE I. Noncomprehensive list of LBL and SBL accelerator-based neutrino experiments and the ranges of L/E_ν and Δm^2 that they probe. For Δm^2 within the range quoted, oscillations due to that mass splitting would produce a spectral effect as a function of neutrino energy within the detector. For Δm^2 above the quoted upper bound, the detector would see an averaged oscillation effect across its neutrino energy spectrum.

	First data	FD target	E_ν (GeV)	L_{ND} (km)	L_{ND}/E_ν (km/GeV)	$\Delta m_{ND}^2 (\text{eV}^2)$	L_{FD} (km)	L_{FD}/E_ν (km/GeV)	$\Delta m_{FD}^2 (\text{eV}^2)$
K2K [16]	1999	H ₂ O	0.3–2.5	0.3	0.12–1.2	1.0–10	250	100–1000	$(1.2\text{--}10) \times 10^{-3}$
MiniBooNE [17]	2002	CH ₂	0.2–1.4	0.54	0.4–2.7	0.5–3.2			
MINOS [18]	2005	Fe	1–14	1.04	0.07–1.0	1.2–17	735	53–735	$(1.7\text{--}20) \times 10^{-3}$
T2K [19]	2010	H ₂ O	0.3–1.4	0.28	0.20–0.93	1.3–6.2	295	211–983	$(1.3\text{--}5.9) \times 10^{-3}$
NOvA [20]	2014	C, Cl, H	1.0–3.0	1.0	0.33–1.0	1.2–3.7	810	270–810	$(1.5\text{--}4.6) \times 10^{-3}$
MicroBooNE [21]	2015	Ar	0.2–1.4	0.47	0.3–2.4	0.5–3.7			
SBN [22]	2021	Ar	0.2–1.4	0.11	0.079–0.55	2.2–16	0.6	0.43–3.0	0.41–2.9
Hyper-K [23]	2027	H ₂ O	0.3–1.4	0.28	0.20–0.93	1.3–6.2	295	211–983	$(1.3\text{--}5.9) \times 10^{-3}$
DUNE [24]	2030	Ar	0.5–5.0	0.57	0.11–1.1	1.1–11	1285	257–2570	$(4.8\text{--}48) \times 10^{-3}$

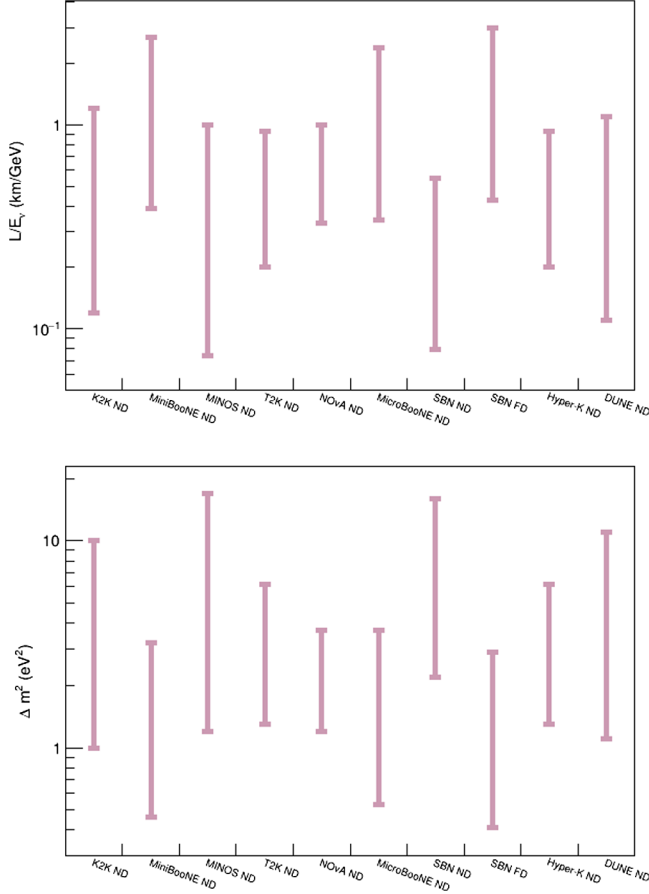


FIG. 1. Sensitivity ranges of the accelerator-based neutrino experiments from Table I to L/E_ν (left) and Δm^2 (right). The LBL far detectors from Table I are excluded, as they all have sensitivity in the few $\times 10^{-3}$ eV² region.

appearance, but found no excess of events like LSND and MiniBooNE [25–29]. The follow-up Short-Baseline Neutrino (SBN) experiment will further test the short baseline anomaly [30]. The SBN ND, called the Short-Baseline Near Detector (SBND), is a 112-ton LArTPC positioned at $L = 110$ m from the BNB target and the SBN FD, ICARUS, is a 476-ton LArTPC positioned at $L = 600$ m from the BNB target. SBN is the first two-detector SBL experiment.

The interaction cross section σ_β is a critical input to MicroBooNE and SBN’s oscillation searches. Its estimation using an event generator such as GENIE depends on numerous uncertain parameters, giving the estimation an *a priori* uncertainty at the 10%–20% level. In two-detector experiments, this uncertainty can be reduced by fitting to the ND, where oscillations are negligible [31,32].

Another plausible way to reduce the cross section uncertainty is to tune the parameters of the event generator to external cross section data. The GENIE Collaboration fits its own interaction models to data collected over a wide phase space and makes these tunes publicly available [33]. Some experiments use these official GENIE tunes in their

analyses, while others fit the interaction model themselves to data from their phase subspace. MicroBooNE, for example, tuned its GENIE model to a T2K cross section that was made using a similar neutrino flux as MicroBooNE [34,35].

We expand on the approach of MicroBooNE [34] by tuning two commonly used model configurations to five different cross section measurements, and evaluate the performance by comparing the tuned model predictions to *other* data. This provides a way to test the robustness of the model and its uncertainty parametrization. If a model has the correct physics, one would expect tunes of that model to describe all of the available data with the same parameter values. If an interaction model is incomplete but its uncertainties are robust, one would expect tunes of that model to reasonably describe available data, but perhaps with different best-fit parameter values. We also attempt to apply the specific procedure used by MicroBooNE to other measurements, in order to study how the model used in MicroBooNE is dependent on which specific measurement it is tuned to. The study details are given in Sec. II, and results are presented in Sec. III.

II. ANALYSIS METHOD

Two commonly used configurations of the GENIE version 3.4.0 interaction model are fit to a variety of cross section measurements. The G18_10a_02_11a configuration (henceforth referred to as G18) is used by several neutrino experiments, including MicroBooNE [33,34]. The AR23_0i_00_000 configuration (henceforth AR23) was developed by the DUNE collaboration for use in its oscillation sensitivity analyses, and is also used by SBN. The physics models used in these two comprehensive model configurations are outlined in Table II.

The two configurations use the same resonance, inelastic, hadronization, and FSI models, but different ground state models. AR23 uses a modified local Fermi gas (LFG) model which populates a broader region of missing energy and momentum space. Both G18 and AR23 use the Valencia charged-current quasielastic (CCQE) cross section model [36,37], although G18 uses the historical dipole approximation for the axial form factor, while AR23 uses a z expansion, which is expected to be more realistic at high Q^2 [38]. G18 also uses the Valencia model for the charged current 2-particle, 2-hole (2p2h) cross section [39], which only predicts three-momentum transfer below 1.2 GeV; AR23 uses the fully relativistic SuSAv2 2p2h model [40], which does not have a three-momentum transfer ceiling.

Alternative neutrino event generators like NEUT [7–9] and NuWro [11] also support multiple model configurations with physics models that are not dissimilar to the ones in G18 and AR23. For example, the NEUT configuration chosen by T2K for its oscillation analyses [31] uses the “Benhar spectral function” ground state model [41] that is qualitatively similar to AR23’s, the same 2p2h model as

TABLE II. Physics models used in the GENIE G18 and AR23 model configurations.

	G18_10a	AR23_0i
Ground state	LFG	Custom spectral-function-like LFG
Quasielastic	Valencia using dipole expansion	Valencia using z expansion
2p2h	Valencia	SuSAv2
Resonance	Berger-Sehgal 2020	Berger-Sehgal 2020
Inelastic	Bodek-Yang 2020	Bodek-Yang 2020
Hadronization	AGKY 2020	AGKY 2020
FSI	hA18	hA18

G18, and a resonance model that is the precursor to the one in both G18 and AR23. The GiBUU [10] is unlike the others in that it relies on transport theory [10].

MicroBooNE observed that G18 underpredicted MiniBooNE's measurement of the cross section for charged-current scattering with zero final-state pions ($CC0\pi$) taken in the same beam [34]. Although the parameters of the nominal GENIE interaction model had already been tuned to make G18, MicroBooNE performed

an additional tune of the relevant G18 parameters to a $CC0\pi$ cross section measurement made in T2K's ND, with very similar flux. The result is an improved prediction of the MiniBooNE data [34,35].

The nominal G18 configuration, which underpredicts MiniBooNE $CC0\pi$ cross section data, also underpredicts MicroBooNE's measurement of the cross section for charged-current scattering with one final-state proton and zero final-state mesons ($CC1\mu1p$). Figure 2 shows that the

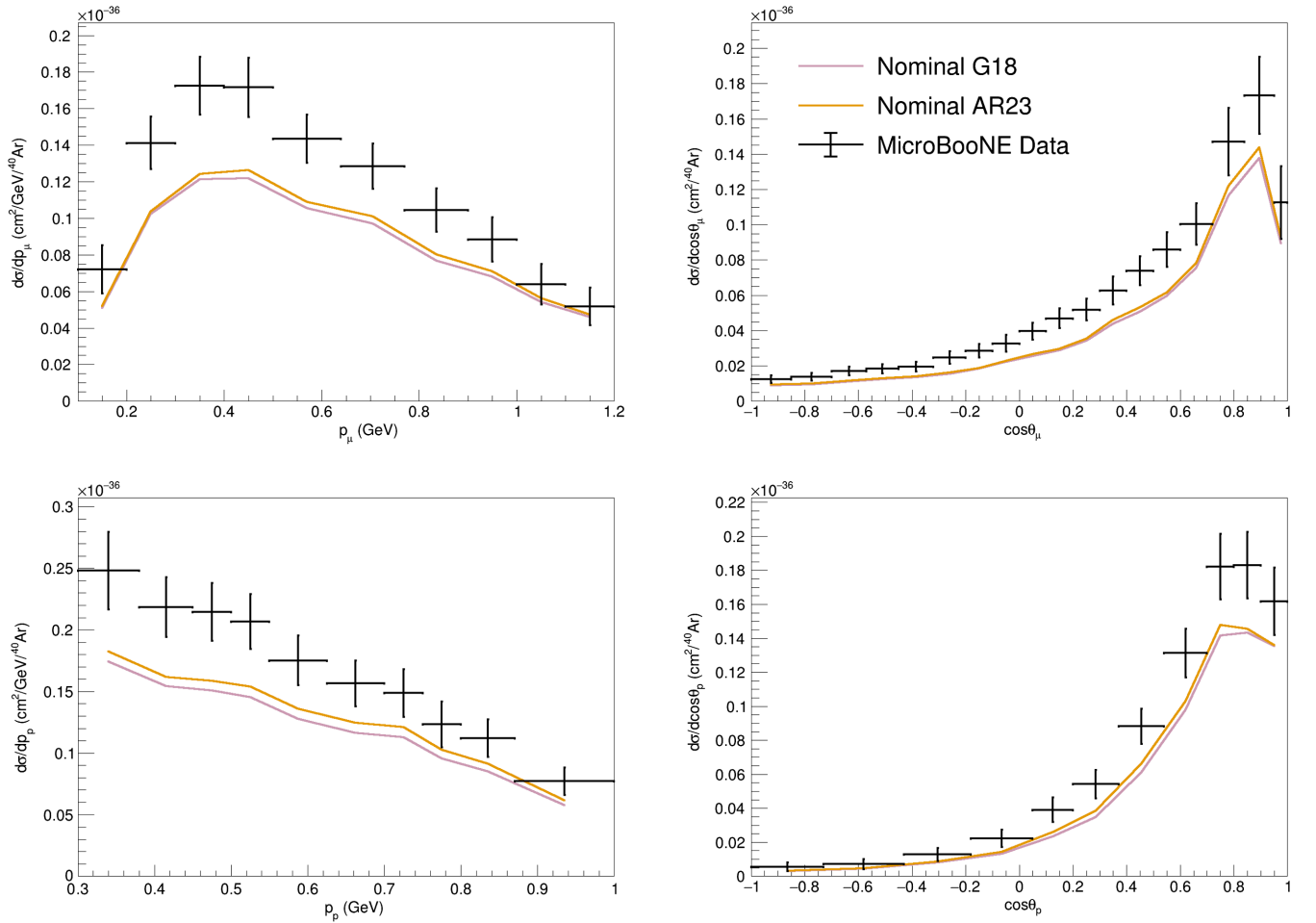


FIG. 2. MicroBooNE $CC1\mu1p$ single-differential cross section data in muon momentum (top left), muon angle (top right), proton momentum (bottom left), and proton angle (bottom right) plotted with the nominal predictions of the data made by the G18 and AR23 model configurations.

TABLE III. A summary of the cross section measurements used in this study.

	Beam	Target	Signal	σ dimensions	σ variables
T2K 2016 [35]	ν_μ	CH	CC0 π	2D	$(\cos \theta_\mu, p_\mu)$
T2K 2020 [42]	ν_μ	CH	CC0 π	2D	$(\cos \theta_\mu, p_\mu)$
MicroBooNE [43]	ν_μ	Ar	CC1 μ 1p	1D	$p_\mu, \cos \theta_\mu$
MINERvA [46]	ν_μ	CH	CC0 π	2D	$(p_{\mu,L}, p_{\mu,T})$
MiniBooNE [44]	ν_μ	CH ₂	CC0 π	2D	$(\cos \theta_\mu, T_\mu)$

nominal AR23 configuration underpredicts the MicroBooNE cross section. We study how either configuration's prediction of the MicroBooNE cross section changes after tuning to cross section measurements made by the T2K, MicroBooNE, and MINERvA experiments [35,42–44].

The measurement used in the MicroBooNE tune was a T2K 2016 CC0 π double-differential cross section in muon kinematic variables [35]. MicroBooNE selected this T2K measurement over measurements made by the ArgoNeuT [45], MINERvA [46], and MiniBooNE [44] experiments because ArgoNeuT had low statistics, MINERvA used a higher neutrino energy range, and MiniBooNE used the same neutrino source, which would introduce correlations in the flux between MiniBooNE and MicroBooNE.

The T2K 2016 measurement is reported as a function of muon momentum (p_μ) and the cosine of the muon angle with respect to the direction of the neutrino beam ($\cos \theta_\mu$). In [35], the cross section is calculated twice using different analysis methods. “Analysis I” uses a binned likelihood fit to extract the cross section, while “Analysis II” uses dAgostini unfolding [47]. Another difference between the two analysis methods is that Analysis II uses a restricted phase space where $p_\mu > 0.2$ GeV and $\cos \theta_\mu > 0.6$. MicroBooNE selected the Analysis I cross section for their tune because of its broader phase space.

T2K published an updated CC0 π cross section in 2020 [42]. While the 2016 measurement selected only forward-going muons originating from the upstream fine grained detector (FGD1) that is pure CH, the 2020 measurement selected forward-going, high-angle and backward-going muons originating from either FGD1 or FGD2, giving the 2020 measurement a smaller statistical uncertainty and broader acceptance in muon angle compared to the 2016 measurement. Although FGD2 is composed of alternative layers of CH and H₂O, the T2K 2020 cross section used here selects FGD2 neutrino interactions in only the CH layers. We tune G18 and AR23 to the T2K 2016 measurement to check that we can replicate MicroBooNE's result, then to the T2K 2020 measurement to study if anything changes.

We also tune G18 and AR23 to MicroBooNE data. Of course it would not have made sense for MicroBooNE to tune to its own data, nor is it sensible for SBN to tune to data collected in the same beam. However, these

MicroBooNE tunes serve as a consistency check for our fitting methods. Specifically, single-differential cross sections for a CC1 μ 1p signal definition are used, four of which are plotted in Fig. 2. The cross sections as a function of p_μ and θ_μ are fit because they allow for the closest comparison to the double-differential T2K measurements, which are also reported as a function of lepton kinematics. These two results are not independent because they use the same signal selection. In particular the normalization of the two MicroBooNE measurements is expected to be fully correlated.

Finally, the two GENIE configurations are tuned to a MINERvA CC0 π measurement of the muon neutrino cross section on hydrocarbon from the “low-energy” tune of the NUMI beam (Fermilab's second neutrino beam) with a peak flux of 3.5 GeV [46]. This double-differential cross section is reported as a function of the longitudinal and transverse muon momentum ($p_{||}$ and p_T , respectively). Table III summarizes the five ν_μ cross section measurements used in this study.

The MINERvA data are at notably higher neutrino energy compared to T2K and MicroBooNE, providing a test of the neutrino energy dependence of the cross section models and the robustness of the uncertainty parametrizations at simultaneously describing data at different energies. The higher-energy data also has a different mix of underlying interaction processes, and is predicted to have a larger contribution from pion production where the pion is absorbed.

For G18, the fit parameters used in this study are the same as those used by MicroBooNE because they have the greatest impact on the CC0 π cross section [34]: MaCCQE, NormCCMEC, RPA_CCQE, and XSecShape_CCMEC. The MaCCQE parameter adjusts the value of the axial mass in the dipole parametrization of the CCQE axial-vector form factor. The NormCCMEC parameter adjusts the normalization of the cross section of charged-current (CC) 2p2h interactions, which are also referred to as meson exchange currents (MEC). The RPA_CCQE parameter adjusts the strength of the suppression of the CCQE cross section at low Q^2 due to nucleon-nucleon long range correlations. The XSecShape_CCMEC parameter adjusts the shape of the 2p2h cross section between the Valencia model [48] and the empirical model [49].

Due to the differences between G18 and AR23 summarized in Table II, the fit parameters in AR23 are different. Six parameters explain nearly the entire variation in the CC0 π cross section at these energies: RPA_CCQE, NormCCMEC, ZExp_PCA_b1, b2, b3, and b4. These RPA_CCQE and NormCCMEC parameters have the same function in AR23 as in G18. The ZExp_PCA_b1, b2, b3, and b4 parameters adjust the coefficients in the z-expansion parametrization of the CCQE axial-vector form factor. Although the four coefficients are necessarily correlated, the corresponding fit parameters have been transferred to an uncorrelated basis by means of principal component analysis.

The fitting for this study is performed with the NUISANCE software package [50], which uses a gradient descent algorithm from the MINUIT package [51] to minimize the χ^2 test statistic with respect to the chosen parameters. The χ^2 is calculated in two different ways. The “full χ^2 ,” which uses the covariance matrix from the data and adds a penalty term for pulls on the model parameters, is given by Eq. (3),

$$\chi^2_{\text{full}} = \sum_i^{N_d} \sum_j^{N_d} (D_i - P_i)(\mathbf{C}^{-1})_{ij}(D_j - P_j) + \sum_i^{N_p} \sigma_i^2 \quad (3)$$

where N_d is the number of data points, D_i is the i th data point, P_i is the prediction of the i th data point, $\mathbf{C}$ is the covariance matrix describing the data uncertainty such that $(\mathbf{C}^{-1})_{ij}$ is the ij th element of its inverse, N_p is the number of fit parameters, and σ_i is the i th best-fit parameter value, expressed in units of one sigma uncertainty.

An alternative test statistic is the “diagonal χ^2 ” used by MicroBooNE [34], which uses only the diagonal elements of the covariance matrix and does not add a penalty term like in χ^2_{full} ,

$$\chi^2_{\text{diag}} = \sum_i^{N_d} \frac{(D_i - P_i)^2}{C_{ii}}. \quad (4)$$

Although the standard practice in neutrino physics is to use χ^2_{full} as the test statistic, MicroBooNE opted to use χ^2_{diag} when they tuned G18, a fitting method that will henceforth be referred to as the “diagonal G18” method. MicroBooNE explained their nonstandard choice as the solution to the problem that fitting T2K 2016 with the full covariance matrix resulted in a significant reduction in the total cross section prediction or Peelle’s pertinent puzzle (PPP) [52]. PPP states that the weighted mean of experimental data with systematic uncertainties may fall outside the range of the data. In MicroBooNE’s case, PPP caused a fit that visibly agreed with the data to be disfavored compared to a fit with a lower cross section that visibly disagrees with the data but has a lower test statistic due to strong correlations in the data uncertainties.

MicroBooNE found that disregarding the off-diagonal elements of the covariance matrix associated with the T2K 2016 cross section avoided PPP [34]. In this work, we replicate MicroBooNE’s diagonal G18 tune and compare

the results to those obtained using the full covariance matrix or a “full G18” tune. We also look at the results of tuning AR23 using the full covariance matrix (referred to as the “full AR23” method). For consistency with MicroBooNE, the penalty term from Eq. (3) is not included in our diagonal G18 method. However, we do use a penalty term in the full G18 and full AR23 methods because it is the conventional way to prevent fit parameters from running far away from their nominal values.

III. RESULTS AND DISCUSSION

Using the diagonal method, G18 is fit to T2K 2016 I, T2K 2020, MINERvA, MicroBooNE p_μ , and MicroBooNE

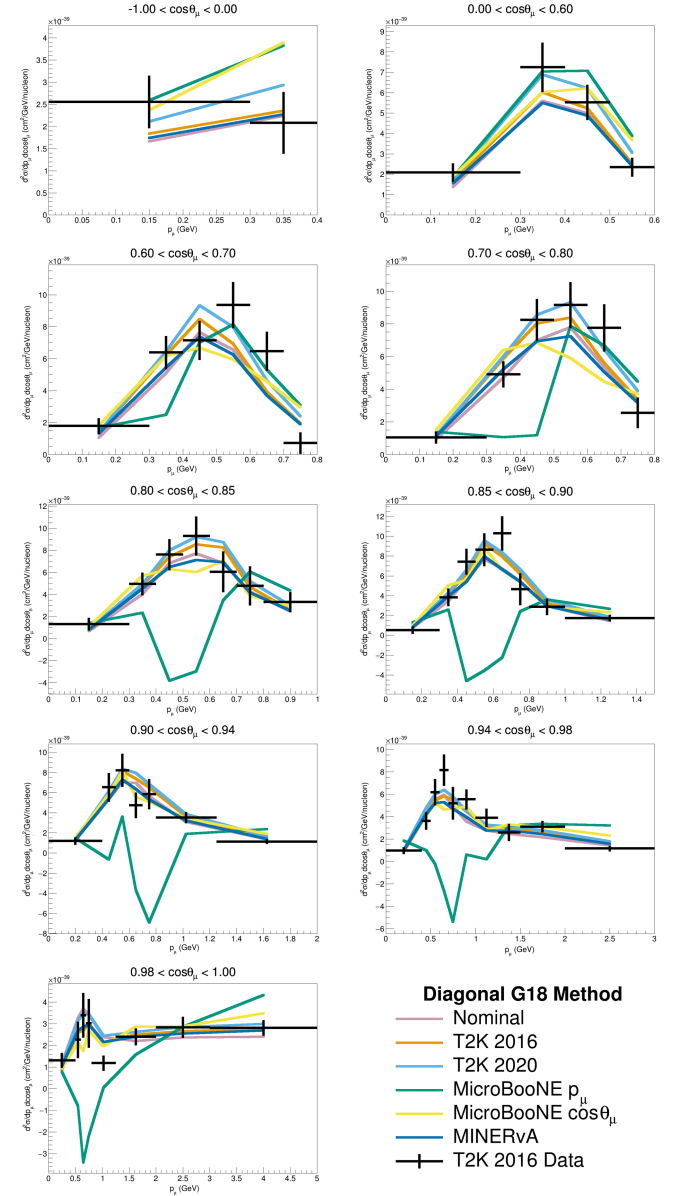


FIG. 3. T2K 2016 data is compared to predictions from the nominal G18 model configuration, as well as five different tunes obtained by fitting G18 to the data given in the legend using the “diagonal” method.

$\cos\theta_\mu$. Figure 3 shows how these five tunes perform compared to the nominal G18 model when predicting the T2K 2016 data. The orange curve in the figure is the result of tuning the G18 model to the data shown. The other curves come from applying the tuning method of Ref. [34] to other measurements, including MicroBooNE's own data, then predicting T2K 2016 using the tuned parameters. The green curve in Fig. 3 is essentially the inverse of Ref. [34]: the model is tuned to MicroBooNE data and the result is then compared to T2K. This yields an unphysical result; the tuned parameter values from the MicroBooNE p_μ fit predict unphysical, negative cross sections in several regions of the T2K parameter space. This is the result of a poor fit to the MicroBooNE p_μ measurement using the diagonal G18 method.

MicroBooNE tuned G18 to T2K 2016 in order to improve its prediction of the MiniBooNE CCQE-like ($\text{CC}0\pi$) measurement of the muon neutrino cross section on hydrocarbon [44]. Figure 4 shows the nominal predictions as well as the results of tuning G18 to each of the five measurements using the diagonal G18 method, and the full covariance with both G18 and AR23; two bins of muon angle are shown, one corresponding to forward angles and the other to wide angles, as a representation of the full 2D dataset. Diagonal G18 reproduces the MicroBooNE result in Ref. [34], namely that tuning to T2K 2016 increases the cross section and describes the MiniBooNE data well. However, applying the same method to the MicroBooNE p_μ data gives an unphysical result. Including the full covariance matrix, the G18 tunes do a visibly worse job of predicting the MiniBooNE data.

Figure 5 shows how the five diagonal G18 tunes perform when predicting all five datasets. The top row holds the nominal, untuned predictions, while each subsequent row holds the predictions of a different diagonal G18 tune, and each column contains a different cross section measurement. Printed in each cell is the χ^2_{diag} of Eq. (4) for the prediction indicated by the column using the tune indicated by the row. Five of the cells show the actual minimum χ^2_{diag} calculated during the tuning process, while the other cells show the goodness of fit of a prediction to one measurement after a fit to another measurement. The two MicroBooNE single-differential cross sections share a common signal sample and thus are not independent, so tuning to one and predicting another falls somewhere in between. The normalizations of these two measurements are correlated, and there are also correlations between the muon energy and angle.

The number of degrees of freedom in χ^2_{diag} varies significantly between columns because the number of bins (printed below the column label) varies significantly between cross section measurements. As a visual aid in interpreting the level of agreement, a p value is computed assuming that χ^2_{diag} follows a χ^2 distribution. For nominal predictions, the number of degrees of freedom is equal to the number of bins in the data, while the number of bins minus four is used for tuned predictions in order to account for the four free parameters in the G18 tuning procedure. The cells of Fig. 5 are colored according to this p value such that blueish cells represent reasonable agreement and reddish cells represent poor agreement between the data and the nominal or tuned model. A high p value indicates

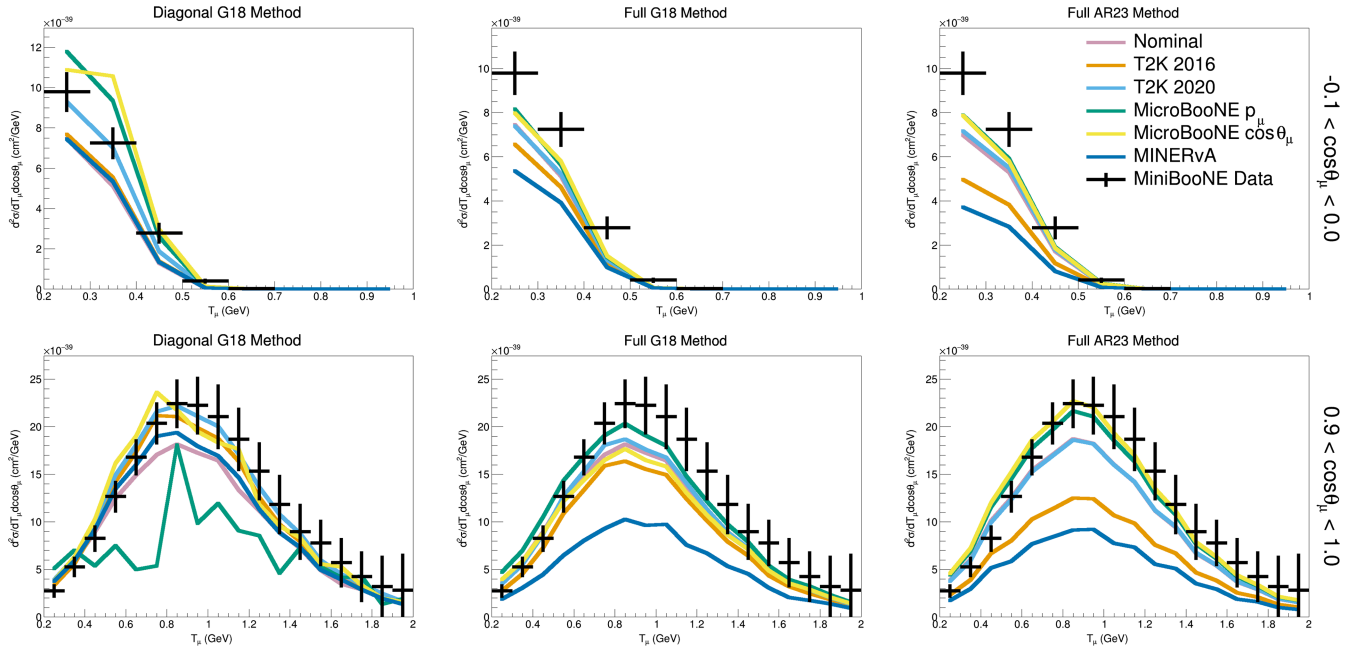


FIG. 4. MiniBooNE data in the $-0.1 < \cos\theta_\mu < 0.0$ bin (top) and $0.9 < \cos\theta_\mu < 1.0$ bin (bottom) is compared to predictions from the nominal G18 and AR23 models and their tunes to the data in the legend using the method in the plot title.

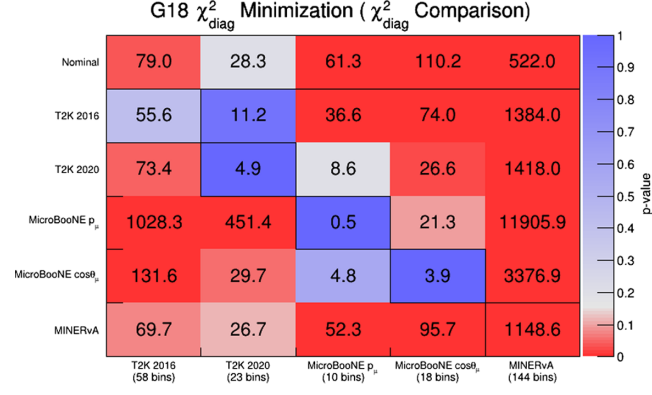


FIG. 5. χ^2_{diag} from Eq. (4) obtained by comparing the G18 model configuration to five different measurements, indicated by each column. The top row is the nominal prediction, subsequent rows are the best-fit tune to the measurement indicated by the row, using the diagonal method. The cell color is a p value assuming χ^2_{diag} follows a χ^2 distribution.

that the tuned model is able to describe the data accurately, but not necessarily that the free parameters are close to their nominal values.

Figure 6 has the same format as Fig. 5 and uses the same diagonal G18 tunes, but shows the χ^2_{full} of Eq. (3), which uses the entire covariance matrix and adds a penalty term for pulls on the model parameters. In this case where goodness of fit is assessed using χ^2_{full} , none of the diagonal G18 tuned models are an improvement over the nominal model. This suggests that the range of cross section shapes allowed by the systematic uncertainties on the data is not represented well by any diagonal G18 tune, and that the best fit obtained by minimizing χ^2_{diag} does not actually describe the data very well when the correlations in the data systematics are taken into account.

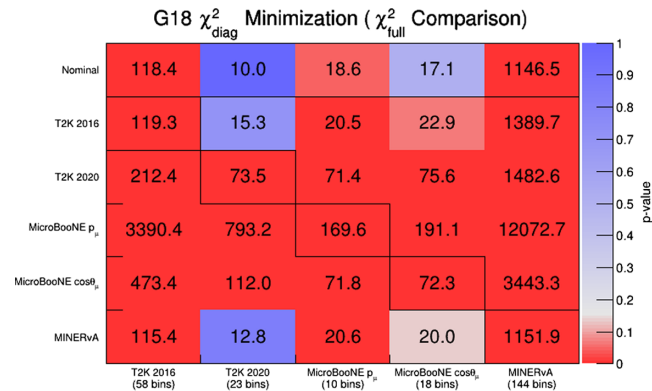


FIG. 6. χ^2_{full} from Eq. (3) obtained by comparing the G18 model configuration to five different measurements, indicated by each column. The top row is the nominal prediction, subsequent rows are the best-fit tune to the measurement indicated by the row, using the diagonal method. The cell color is a p value assuming χ^2_{full} follows a χ^2 distribution.

The parameter pulls used in each of the rows of Figs. 5 and 6 are plotted in Fig. 7 along with their postfit uncertainties. The unit of the parameter pull is the *a priori* 1σ uncertainty. The diagonal G18 tune to T2K 2016 produces reasonable results, as was observed by MicroBooNE. However, the tunes to the other four measurements result in poor fits with pull values greater than three standard deviations or unreasonably small pull errors. In particular, the tune to MicroBooNE p_μ pulls three of the four parameters outside of the prior $\pm 3\sigma$ range, causing the unphysical results in Fig. 3.

Figure 7 also shows the parameter pulls used in the five full G18 tunes. With the data covariance matrix included, the tunes to T2K 2020 and both MicroBooNE measurements produce reasonable results, while the tunes to T2K 2016 and MINERvA result in poor fits. As before, each set of full G18 parameter pulls is used to predict the five cross

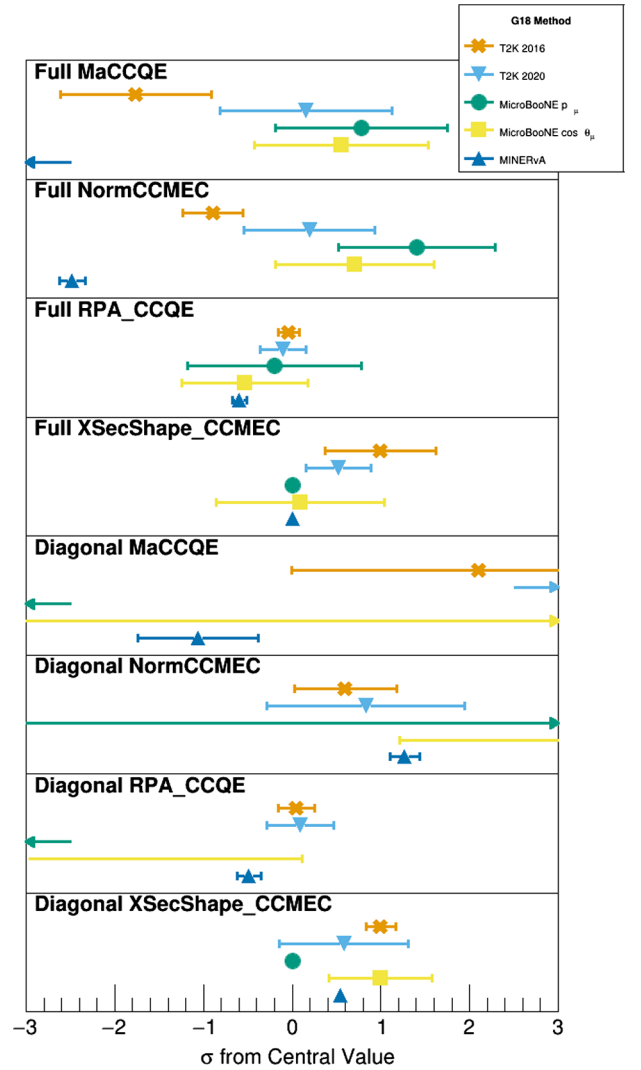


FIG. 7. Postfit parameter values and errors (in units of 1σ uncertainty) obtained from the full G18 and diagonal G18 fitting methods and the indicated cross section measurements.

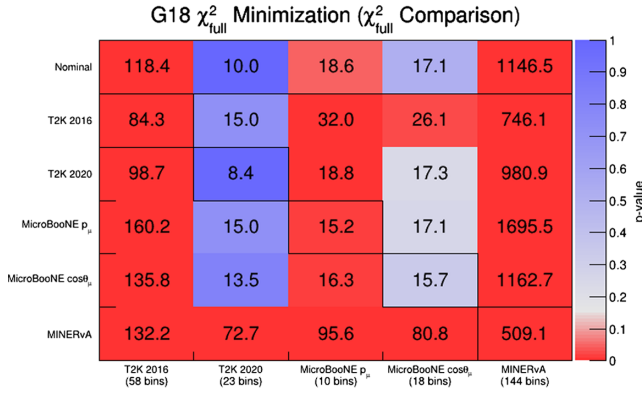


FIG. 8. χ^2_{full} from Eq. (3) obtained by comparing the G18 model configuration to five different measurements, indicated by each column. The top row is the nominal prediction, subsequent rows are the best-fit tune to the measurement indicated by the row, using the full covariance matrix of the data. The cell color is a p value assuming χ^2_{full} follows a χ^2 distribution.

section measurements, and the p values of these predictions are plotted in Fig. 8. The fit to T2K 2016 is notably poor, while the model describes T2K 2020 and both MicroBooNE measurements reasonably well out of the box, and especially with parameters tuned to any of the data except MINERvA.

Figure 9 shows the parameter pulls for fits using AR23, and Fig. 10 gives the fit quality. The AR23 configuration does slightly better in describing the MicroBooNE momentum spectrum compared to G18. It is in good agreement with T2K 2020 but not T2K 2016, and fits to the 2016 data yield a poor prediction of every other measurement. The parameter values for the fits to T2K 2020 and MicroBooNE are broadly consistent, indicating that AR23 is able to simultaneously describe lepton kinematics in this energy regime on two different nuclei.

Neither model configuration is able to describe the MINERvA data, which can be seen by the high χ^2 in the bottom row of Figs. 8 and 10. Figure 11 plots the MINERvA data with the nominal G18 and AR23 predictions as p_T distributions in bins of p_L . Because the MINERvA neutrino beam is peaked around 3.5 GeV, most of the cross section falls in the first few bins of p_L . In this region, there is disagreement between data and MC at moderate p_T .

Tuning either model configuration to the MINERvA data decreases the total cross section, giving an apparently worse fit, albeit with a smaller χ^2 . This is qualitatively similar to what MicroBooNE observed in fitting T2K 2016 and attributed to Peelle’s pertinent puzzle (PPP) [34]. The PPP visible in Fig. 11 is a sign that the model parameters used in the low-energy T2K 2020 and MicroBooNE tunes do not have sufficient freedom to describe the high-energy MINERvA cross section and its (strongly correlated) systematic uncertainties.

Compared to T2K 2020 and MicroBooNE, the MINERvA cross section has a different mix of interaction

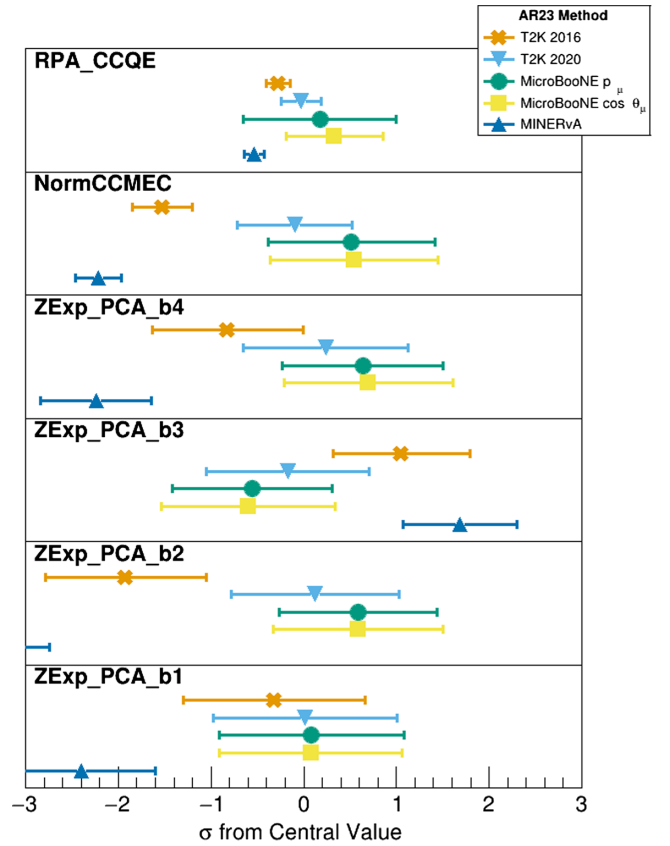


FIG. 9. Postfit parameter values and errors (in units of 1σ uncertainty) obtained by fitting AR23 using the full covariance matrix to the indicated cross section measurements.

processes with a much higher contribution from pion production followed by absorption. It might be expected that additional parameters are required in order to model these contributions. As a test, fits are performed with four

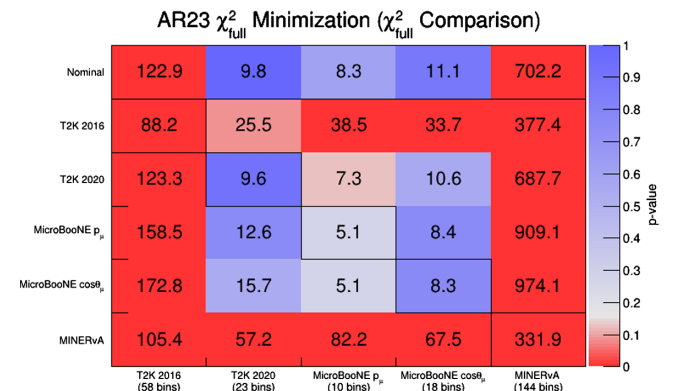


FIG. 10. χ^2_{full} from Eq. (3) obtained by comparing the AR23 model configuration to five different measurements, indicated by each column. The top row is the nominal prediction, subsequent rows are the best-fit tune to the measurement indicated by the row, using the diagonal method. The cell color is a p value assuming χ^2_{full} follows a χ^2 distribution.

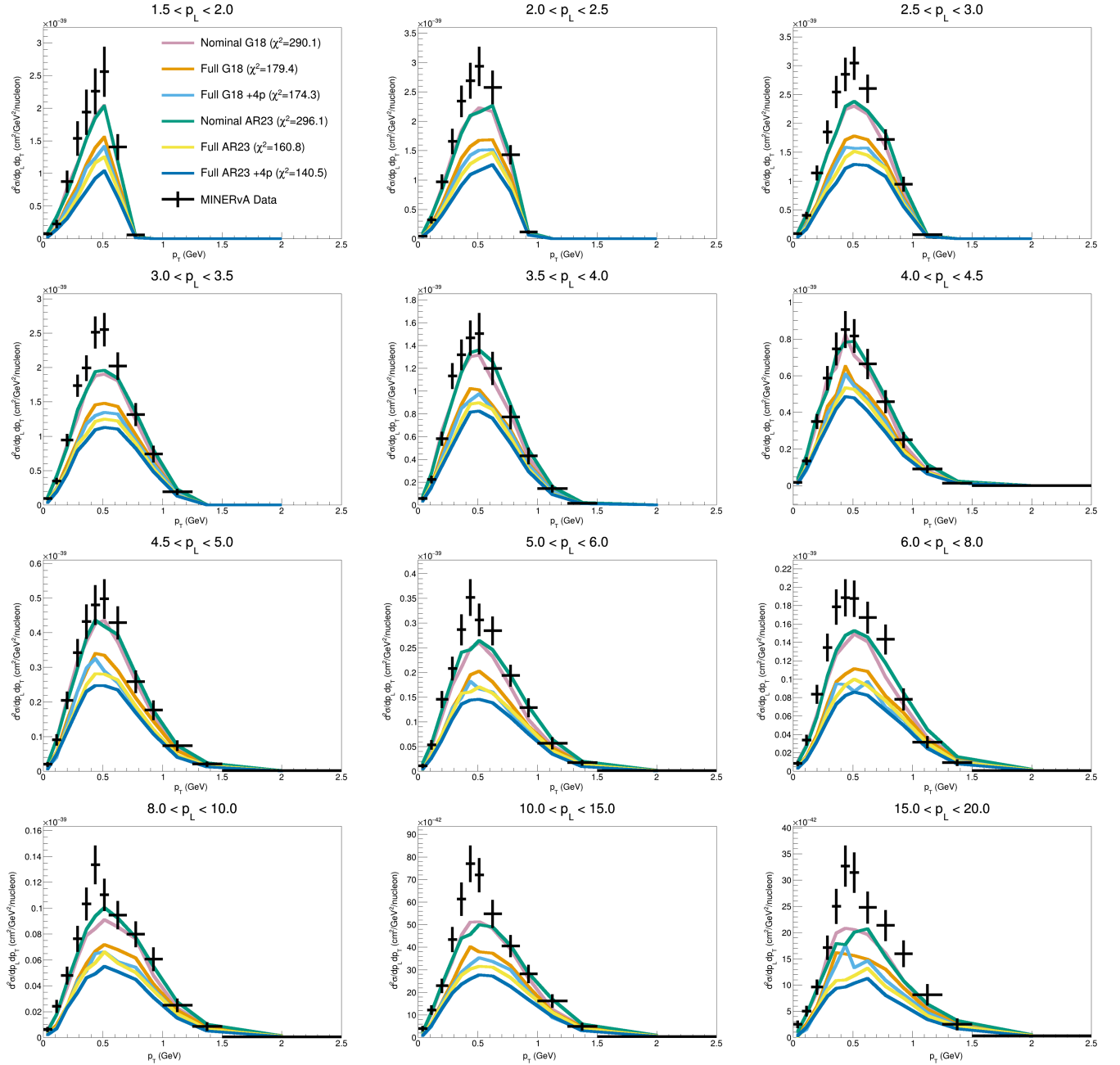


FIG. 11. MINERvA data is compared to predictions from the G18 and AR23 models, showing the nominal predictions, the result of tuning to the data with the full covariance matrix, and the tune with two additional pion production parameters.

additional parameters that govern pion production: the vector and axial masses M_{VCCRES} and M_{ACCRES} , the fraction of pions inelastically scattered $\text{FrInel}_{\text{pi}}$, and the fraction of pions absorbed FrAbs_{pi} . However, the extra freedom does not improve the result in Fig. 11.

IV. TUNING IN LBL AND SBL EXPERIMENTS

The neutrino flux appears in the denominator of the formula used to extract cross sections. For a differential measurement with respect to some quantity X , which is

unfolded to “true” quantities, the cross section in a bin i can be written as

$$\frac{d\sigma}{dX_i} = \frac{\sum_j U_{ij}(N_j - N_j^{\text{bkg}})}{\epsilon_i T \Phi \Delta X_i}, \quad (5)$$

where N_j is the number of selected events in a reconstructed bin j , N_j^{bkg} is the predicted background in that bin, U_{ij} is an element of the unfolding matrix that relates true to reconstructed X , ϵ_i is the selection efficiency, T is the

number of targets, Φ is the neutrino flux, and ΔX_i is the bin width.

In all cross section data ever reported, the flux Φ in the denominator of Eq. (5) is an estimate of the unoscillated flux of the neutrino source assuming the probability of short-baseline oscillations is zero. For an LBL experiment, tuning the interaction model to a flux-integrated SBL cross section is consistent with an oscillation analysis that neglects short-baseline oscillations. The ND data is then used to constrain the flux and cross section uncertainties with the same assumption, namely that oscillations are not present in the ND.

For an SBL experiment, however, tuning the interaction model to external data is problematic, because available cross section data is analyzed assuming there are no short-baseline oscillations. If a sterile oscillation were present in an SBL experiment, then any cross section measurement with an overlap in L/E_ν (see Table I) would be dividing by the incorrect flux due to that oscillation.

Consider a hypothetical where an experiment has a perfect interaction model, and 10% of active ν_μ oscillate to sterile neutrinos with sufficiently high Δm^2 that the oscillations are extremely rapid and only average effects are observed for L/E_ν of order 0.1–10 km/GeV. If a perfect interaction model is tuned to external data, the flux assumed in that external measurement will be 10% higher than the actual active ν_μ flux at the detector, so the extracted cross section will be 10% too low. This discrepancy would be attributed to cross section effects, and tuned away. The tuned model will then agree perfectly with the SBL data and the experiment will publish an exclusion on SBL oscillations.

Alternatively, an SBL experiment could tune its model to external data from a different range of L/E_ν than that to which the SBL experiment is sensitive. This is limited by the availability of data, and the typically broad range of L/E_ν that the data span. In this scenario, the SBL experiment should only publish limits in the range of Δm^2 where the effect of oscillations on the cross section data is negligible.

A more subtle risk is the use of pretuned interaction models in SBL searches. GENIE, for example, has numerous built-in comprehensive model configuration tunes [33]. These tunes are extremely convenient for LBL experiments like NOvA [53]. However, the data used in these tunes is analyzed with the assumption of no sterile oscillations, so their use in SBL searches could potentially mask the sterile oscillation signal in the same way as above.

In the specific case of SBN, the two-detector SBL search at Fermilab, any tuning of the interaction model to external SBL cross section data would be an unnecessary harm. Unlike its predecessors (all one-detector SBL searches), SBN will be able to resolve discrepancies between its interaction model and its data internally utilizing its near detector SBND, which sits at a different range of L/E_ν than

its far detector ICARUS, in an analysis that will explicitly control for sterile oscillations in SBND's data.

The best course of action for SBN is to choose a base model that is as free as possible from the influence of external SBL data. Instead of tuning that model, which would in principle improve its agreement with data, SBN should check that there is sufficient freedom in the model's uncertainties to describe all of the available external data, for example by following the methods presented in this work. The consequence of not tuning may be large *a priori* cross section uncertainties, but these can be constrained with data from SBND. Unlike external cross section data, the SBND constraint can be implemented in a joint fit with ICARUS data, such that the presence of SBL oscillations can be handled correctly.

Although this Section has focused on the risks of tuning in SBL experiments, LBL experiments should also beware. The observation that tunes to T2K 2020 yield excellent descriptions of that dataset, but poor descriptions of the MINERvA data set, which has the same target nucleus and nearly the same signal definition, should caution DUNE against overrelying on cross section data from SBND. While precise measurements from SBND will obviously be valuable for informing model selection for DUNE analyses, it remains critical for DUNE to rely primarily on measurement from its own near detector.

V. CONCLUSIONS

- (i) The T2K 2020 and MicroBooNE data are mutually well described by the AR23 configuration and its uncertainties. The G18 configuration also does a reasonable job of simultaneously describing T2K 2020 and MicroBooNE, but with some tension in the MicroBooNE muon momentum.
- (ii) The T2K 2016 measurement is inconsistent with the others, including the T2K 2020 measurement that uses some of the same data.
- (iii) Neither G18 nor AR23 are able to describe the MINERvA data. This suggests that these models are adequate at low energy but fall short at higher energy, or fail to model the additional processes that are present in MINERvA but not at lower energies.
- (iv) The result of the MicroBooNE tuning procedure is strongly dependent on the specific data chosen, and would be substantially different if the more recent T2K 2020 data were used. The procedure yields unrealistic results when applied to other measurements, including MicroBooNE's own data.
- (v) Short-baseline oscillation experiments should be extremely careful about tuning to cross section data. The two-detector SBN should *not* tune to data, and should use a model configuration that is not implicitly tuned to data, like AR23.

ACKNOWLEDGMENTS

We thank Jaesung Kim for assistance with installing software. We thank Jaesung Kim and Faiza Akbar for feedback on a draft of this manuscript. This material is based on work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Awards No. DE-SC0008475 and No. DE-SC0024431; by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program; and by the University of Rochester. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under Contract No. DESC0014664.

This report was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof,

nor any of their employees, makes any warranty, express or implied, for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they are owned by a third party and the terms of use prevent public distribution. The data are available from the authors upon reasonable request.

-
- [1] Y. Fukuda *et al.* (Super-Kamiokande Collaboration), Evidence for oscillation of atmospheric neutrinos, *Phys. Rev. Lett.* **81**, 1562 (1998).
 - [2] Q. R. Ahmad *et al.* (SNO Collaboration), Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory, *Phys. Rev. Lett.* **89**, 011301 (2002).
 - [3] S. Navas *et al.* (Particle Data Group), Review of particle physics, *Phys. Rev. D* **110**, 030001 (2024).
 - [4] S. Agostinelli *et al.*, Geant4—A simulation toolkit, *Nucl. Instrum. Methods Phys. Res., Sect. A* **506**, 250 (2003).
 - [5] A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), Neutrino flux prediction at MiniBooNE, *Phys. Rev. D* **79**, 072002 (2009).
 - [6] C. Andreopoulos *et al.*, The GENIE neutrino Monte Carlo generator, *Nucl. Instrum. Methods Phys. Res., Sect. A* **614**, 87 (2010).
 - [7] N. M. Coyle, S. W. Li, and P. A. N. Machado, Neutrino-nucleus cross section impacts on neutrino oscillation measurements, *Phys. Rev. D* **111**, 093010 (2025).
 - [8] Y. Hayato, A neutrino interaction simulation program library NEUT, *Acta Phys. Pol. B* **40**, 2477 (2009).
 - [9] L. Alvarez-Ruso, L. S. Geng, S. Hirenzaki, and M. J. V. Vacas, Charged current neutrino-induced coherent pion production, *Phys. Rev. C* **75**, 055501 (2007).
 - [10] 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).
 - [11] H. Prasad *et al.*, Developments in NuWro Monte Carlo generator, [arXiv:2501.11470](https://arxiv.org/abs/2501.11470).
 - [12] A. Aguilar *et al.* (LSND Collaboration), Evidence for neutrino oscillations from the observation of $\bar{\nu}_e$ appearance in a $\bar{\nu}_\mu$ beam, *Phys. Rev. D* **64**, 112007 (2001).
 - [13] J. N. Abdurashitov *et al.* (The SAGE Collaboration), Measurement of the response of a gallium metal solar neutrino experiment to neutrinos from a ^{51}Cr source, *Phys. Rev. C* **59**, 2246 (1999).
 - [14] J. N. Abdurashitov *et al.*, Measurement of the response of a Ga solar neutrino experiment to neutrinos from a ^{37}Ar source, *Phys. Rev. C* **73**, 045805 (2006).
 - [15] S. Schael *et al.* (ALEPH, DELPHI, L3, OPAL, SLD Collaborations, LEP Electroweak Working Group, SLD Electroweak Group, and SLD Heavy Flavour Group), Precision electroweak measurements on the Z resonance, *Phys. Rep.* **427**, 257 (2006).
 - [16] M. H. Ahn *et al.* (K2K Collaboration), Measurement of neutrino oscillation by the K2K experiment, *Phys. Rev. D* **74**, 072003 (2006).
 - [17] A. Aguilar-Arevalo *et al.*, The MiniBooNE detector, *Nucl. Instrum. Methods Phys. Res., Sect. A* **599**, 2846 (2009).
 - [18] D. Michael *et al.*, The magnetized steel and scintillator calorimeters of the MINOS experiment, *Nucl. Instrum. Methods Phys. Res., Sect. A* **596**, 190228 (2008).
 - [19] K. Abe *et al.* (T2K Collaboration), The T2K experiment, *Nucl. Instrum. Methods Phys. Res., Sect. A* **659**, 106 (2011).
 - [20] P. Adamson *et al.*, Constraints on oscillation parameters from ν_e appearance and ν_μ disappearance in NOvA, *Phys. Rev. Lett.* **118**, 231801 (2017).
 - [21] R. Acciarri *et al.*, Design and construction of the Micro-BooNE detector, *J. Instrum.* **12**, P02017 (2017).

- [22] P. A. Machado, O. Palamara, and D. W. Schmitz, The short-baseline neutrino program at Fermilab, *Annu. Rev. Nucl. Part. Sci.* **69**, 363 (2019).
- [23] K. Abe *et al.* (Hyper-Kamiokande Proto-Collaboration), Hyper-Kamiokande design report, [arXiv:1805.04163](https://arxiv.org/abs/1805.04163).
- [24] B. Abi *et al.*, Deep Underground Neutrino Experiment (DUNE), Far detector technical design report, volume I: Introduction to DUNE, *J. Instrum.* **15**, T08008 (2020).
- [25] P. Abratenko *et al.* (MicroBooNE Collaboration), Search for an excess of electron neutrino interactions in MicroBooNE using multiple final-state topologies, *Phys. Rev. Lett.* **128**, 241801 (2022).
- [26] P. Abratenko *et al.*, Search for an anomalous excess of inclusive charged-current ν_e interactions in the MicroBooNE experiment using wire-cell reconstruction, *Phys. Rev. D* **105**, 112005 (2022).
- [27] MicroBooNE Collaboration, Search for an anomalous excess of charged-current ν_e interactions without pions in the final state with the MicroBooNE experiment, *Phys. Rev. D* **105**, 112004 (2022).
- [28] P. Abratenko *et al.*, Search for an anomalous excess of charged-current quasielastic ν_e interactions with the MicroBooNE experiment using deep-learning-based reconstruction, *Phys. Rev. D* **105**, 112003 (2022).
- [29] P. Abratenko *et al.*, Search for neutrino-induced neutral-current Δ radiative decay in MicroBooNE and a first test of the MiniBooNE low energy excess under a single-photon hypothesis, *Phys. Rev. Lett.* **128**, 111801 (2022).
- [30] R. Acciarri *et al.* (ICARUS-WA104, LAr1-ND, and MicroBooNE Collaborations), A proposal for a three detector short-baseline neutrino oscillation program in the Fermilab Booster Neutrino Beam (2015).
- [31] K. Abe *et al.*, Measurements of neutrino oscillation parameters from the T2K experiment using 3.6×10^{21} protons on target, *Eur. Phys. J. C* **83**, 782 (2023).
- [32] T. Nosek (NOvA Collaboration), Systematic uncertainties of the NOvA neutrino oscillation analysis, *Proc. Sci. ICHEP2020* (2021) 210.
- [33] J. Tena-Vidal *et al.*, Neutrino-nucleon cross-section model tuning in GENIE v3, *Phys. Rev. D* **104**, 072009 (2021).
- [34] P. Abratenko *et al.*, New $CC0\pi$ GENIE model tune for MicroBooNE, *Phys. Rev. D* **105**, 072001 (2022).
- [35] K. Abe *et al.* (T2K Collaboration), Measurement of double-differential muon neutrino charged-current interactions on C_8H_8 without pions in the final state using the T2K off-axis beam, *Phys. Rev. D* **93**, 112012 (2016).
- [36] J. Nieves, J. E. Amaro, and M. Valverde, Inclusive quasielastic charged-current neutrino-nucleus reactions, *Phys. Rev. C* **70**, 055503 (2004).
- [37] J. Nieves, J. E. Amaro, and M. Valverde, Erratum: Inclusive quasielastic charged-current neutrino-nucleus reactions [Phys. Rev. C 70, 055503 (2004)], *Phys. Rev. C* **72**, 019902 (2005).
- [38] A. S. Meyer, M. Betancourt, R. Gran, and R. J. Hill, Deuterium target data for precision neutrino-nucleus cross sections, *Phys. Rev. D* **93**, 113015 (2016).
- [39] J. Nieves, I. R. Simo, and M. J. V. Vacas, Inclusive charged-current neutrino-nucleus reactions, *Phys. Rev. C* **83**, 045501 (2011).
- [40] S. Dolan, G. D. Megias, and S. Bolognesi, Implementation of the SuSAv2-meson exchange current 1p1h and 2p2h models in cu and analysis of nuclear effects in T2K measurements, *Phys. Rev. D* **101**, 033003 (2020).
- [41] O. Benhar, A. Fabrocini, S. Fantoni, and I. Sick, Spectral function of finite nuclei and scattering of GeV electrons, *Nucl. Phys.* **A579**, 493 (1994).
- [42] K. Abe *et al.* (T2K Collaboration), Simultaneous measurement of the muon neutrino charged-current cross section on oxygen and carbon without pions in the final state at T2K, *Phys. Rev. D* **101**, 112004 (2020).
- [43] P. Abratenko *et al.* (MicroBooNE Collaboration), Multi-differential cross section measurements of ν_μ -argon quasielasticlike reactions with the MicroBooNE detector, *Phys. Rev. D* **108**, 053002 (2023).
- [44] A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), First measurement of the muon neutrino charged current quasielastic double differential cross section, *Phys. Rev. D* **81**, 092005 (2010).
- [45] O. Palamara (ArgoNeuT Collaboration), Exclusive muon neutrino charged current pion-less topologies. ArgoNeuT results and future prospects in LAr TPC detectors, *J. Phys. Soc. Jpn. Conf. Proc.* **12**, 010017 (2016).
- [46] D. Ruterbories *et al.* (MINERvA Collaboration), Measurement of quasielastic-like neutrino scattering at $E_\nu \sim 3.5$ GeV on a hydrocarbon target, *Phys. Rev. D* **99**, 012004 (2019).
- [47] G. D'Agostini, A multidimensional unfolding method based on Bayes' theorem, *Nucl. Instrum. Methods Phys. Res., Sect. A* **362**, 487 (1995).
- [48] R. Gran, J. Nieves, F. Sanchez, and M. J. Vicente Vacas, Neutrino-nucleus quasi-elastic and 2p2h interactions up to 10 GeV, *Phys. Rev. D* **88**, 113007 (2013).
- [49] T. Katori, Meson Exchange Current (MEC) models in neutrino interaction generators, *AIP Conf. Proc.* **1663**, 030001 (2015).
- [50] P. Stowell *et al.*, NUISANCE: A neutrino cross-section generator tuning and comparison framework, *J. Instrum.* **12**, P01016 (2017).
- [51] F. James and M. Roos, MINUIT—A system for function minimization and analysis of the parameter errors and correlations, *Comput. Phys. Commun.* **10**, 343 (1975).
- [52] K. M. Hanson, T. Kawano, and P. Talou, Probabilistic interpretation of Peelle's pertinent puzzle and its resolution, *AIP Conf. Proc.* **769**, 304 (2005).
- [53] M. A. Acero *et al.* (NOvA Collaboration and R. Group), Adjusting neutrino interaction models and evaluating uncertainties using NOvA near detector data, *Eur. Phys. J. C* **80**, 1119 (2020).