Treatment of flux shape uncertainties in unfolded, flux-averaged neutrino cross-section measurements

Lukas Koch*
University of Oxford, Particle Physics Department, DWB, Keble Road, Oxford OX1 3RH, United Kingdom

Stephen Dolan
CERN, 1211 Geneva 23, Switzerland

(Received 10 September 2020; accepted 1 December 2020; published 24 December 2020)

The exact way of treating flux shape uncertainties in unfolded, flux-averaged neutrino cross-section measurements can lead to subtle issues when comparing the results to model predictions. There is a difference between reporting a cross section in the (unknown) real flux, and reporting a cross section that was extrapolated from the real flux to a well-defined, fixed reference flux. Many current analyses do the former, while the results are compared to model predictions as if they were the latter. This leads to the flux shape uncertainty being ignored at least partially, potentially leading to wrong physics conclusions. A somewhat qualitative study of two results from T2K and MINERvA as examples suggests that the size of the effect is subdominant, but non-negligible for those measurements. This paper describes how the issue arises and details possible methods for treating the flux shape uncertainties correctly.

DOI: 10.1103/PhysRevD.102.113012

I. INTRODUCTION

Most modern neutrino cross-section measurements are reported as flux-averaged cross sections, since the incident neutrino energy is not known on an event-by-event basis. The average cross section σ is reported as the number of expected events N per total incident neutrino flux Φ and number of targets T:

$$\sigma = \frac{N}{T\Phi}. \tag{1}$$

This reflects the capabilities of the detectors as closely as possible and avoids making any assumptions about the neutrino energy dependence of the cross sections one is trying to measure. This also applies for differential measurements, where N only counts the number of events in certain kinematic and/or topological bins.

The down side of this approach is that each cross section is specific for its neutrino beam. Even ignoring all smearing, efficiency, and acceptance issues, the average cross section will change when the neutrino energy spectrum changes, since the nonaveraged cross sections depend nontrivially on the neutrino energy. It is thus usually not possible to naively compare the results of different experiments with one another. Instead one must use the theoretical energy dependence to calculate expected cross sections from interaction models for each flux separately. This is all well understood and the process of generating the expected cross sections for multiple fluxes/experiments[1] has been made easier by frameworks such as NUISANCE [1].

The issue gets slightly more confusing when considering flux (and especially flux shape) uncertainties. If one is not careful, it is easy to confuse two very similar ways of measuring and reporting the flux-averaged cross section:

(1) reporting the average cross section in the *real* neutrino flux;
(2) reporting the average cross section in a *reference* neutrino flux.

These two approaches lead to different uncertainty estimations and different rules when comparing the result to theoretical predictions, *even if* the reported central values for both methods are identical.[2]

The first case is the conceptually easier one. One takes the data and a detector model, and then undoes the detector effects assuming a variety of detector, cross section, and flux parameters. For each set of parameters one calculates a

*Corresponding author.
lukas.koch@physics.ox.ac.uk

[1]This includes such annoyances as sensitive phase space constraints.

[2]For example, this would be the case when the chosen reference flux is the best fit value from a fit to the data.

flux-averaged cross section according to Eq. (1). The spread of results yields the reported uncertainty of the measurement.

The second case is slightly more complicated, as it contains an additional step. After calculating the flux-averaged cross section in the real detector, one needs to translate that result to the expected result in a *fixed* reference flux. That is, the variation of the assumed real flux is used to vary the extrapolation from the data to the reference cross section.

The difference between the two approaches is subtle, and both approaches yield correct results for what they are. Problems arise when the result of one approach is treated like one of the other.

This becomes apparent when comparing the first kind of result with a theoretical prediction. The predicted flux-averaged cross section is usually only calculated for a single flux, and it is assumed that all flux uncertainties are contained within the covariance matrix of the published result, i.e., a second-approach result. When the covariance matrix describes the uncertainties of the flux-average cross section in the *real* flux, i.e., a first-approach result, this assumption is not correct. This leads to the flux shape uncertainty being ignored at least partially, and possibly wrong conclusions to be drawn about the compatibility of a cross-section model with the data. The following examples will illustrate this issue.

A. Example One

Imagine a very well thought-out experiment that measures the total cross section of some neutrino interaction mode. Let us assume that it is so well designed that efficiencies are perfectly flat for all relevant events, backgrounds are negligible, and systematic detector and model uncertainties play no role. Also, the experiment has run for a very long time, so statistical errors are not an issue either. This means the only uncertain term in Eq. (1) is the total neutrino flux Φ. An uncertainty in the total flux will be reflected in an equivalent relative uncertainty in the cross-section measurement. Let us assume a total flux uncertainty of 5%. Flux *shape* uncertainties on the other hand will not be reflected in the result, as they do not affect the total flux Φ. For example, an uncertainty in the flux energy scale α modifies the assumed differential neutrino flux by proportionally shifting all energies up or down:

$$\frac{\mathrm{d}\Phi}{\mathrm{d}E_\nu}(E_\nu, \alpha) = \frac{\mathrm{d}\Phi}{\mathrm{d}E_\nu}(E_\nu/\alpha, 1)/\alpha. \tag{2}$$

It does not change the total integrated flux though, so if there was e.g., a 15% uncertainty on the flux energy scale, it would not affect the cross-section result.

Now imagine a model that predicts a cross section that is proportional to the neutrino energy. The usual way of comparing the model to the measurement would be to use the nominal neutrino flux to calculate the total cross section. If the model predicts a 15% lower cross section at the nominal flux, it could be ruled out at the three-sigma level. This does not consider the flux shape uncertainty though. A variation of the neutrino flux energy scale by 15% could easily explain the discrepancy, and would reduce the significance to below one standard deviation.

B. Example Two

Let us now complicate things a bit by introducing a flux shape dependence in the experimental result. The flux shape can enter Eq. (1) via the enumerator. The expectation value of number of true signal events N is rarely directly accessible, but a function of the number of recorded events N_{rec}, the estimated background contribution N_{BG} and the assumed efficiency of signal event recording ϵ:

$$N = \frac{N_{\mathrm{rec}} - N_{\mathrm{BG}}}{\epsilon}. \tag{3}$$

We are again assuming statistical uncertainties are negligible.

Both the efficiency and the expected background can depend on the assumed flux shape. The background is straightforward, as the effect of background processes depends on the flux of neutrinos at energies that contribute to them. The efficiency can depend on the flux if the detector performance depends on event properties that change within the analysis bin, and which in turn depend on the neutrino energy. It is good practice to minimize efficiency uncertainties by choosing fine analysis bins, and by applying phase-space constraints to the signal definition. The former is limited by the available statistics of the data and the detector resolution, while the latter adds the excluded events to the background prediction, so the flux dependence now enters via that route.[3]

For this example, let us ignore the efficiency and concentrate on the background prediction. Imagine a background process with a cross section *antiproportional* to the neutrino energy. This means the "measured" number of true signal events will depend on the assumed neutrino flux energy scale. The total uncertainty of the result will be larger than 5%, depending on the background contribution in the recorded events. Our test model still predicts 15% too few events at the nominal flux. With the included flux shape errors and thus increased total uncertainty, this is now less than a three-sigma effect. When considering the flux shape uncertainty in the model prediction—still at 15%—the significance of the difference again goes down below even one sigma.

This ignores the fact that the uncertainties of measurement and prediction are now correlated, though. An assumed increase in real neutrino energy would increase the model prediction for a better fit to the data, but it would

[3]The flux dependence of the background prediction can of course be reduced by using data driven background subtraction methods, like the use of control regions.

also *decrease* the background contamination, increasing the measured number of signal events and thus the cross-section result. Note that the direction and size of the correlation effect depends on the assumed background model. In extreme cases, a correct treatment of the uncertainty could even *increase* the significance of the model to data discrepancy compared to ignoring the flux shape uncertainty of the prediction.

II. CHALLENGES IN THE FIRST APPROACH

The first approach—reporting a cross section averaged over the *real* flux—is easier to measure in a model-independent way. Depending on the complicating factors of a real-world experiment and the employed unfolding algorithms, no assumptions about the neutrino energy dependence of the cross section need to be made (see e.g., Sec. I A).

It does however make it challenging to compare the result with theoretical predictions. Since the covariance matrices that go with such a result do not cover all effects of the flux on the model prediction, it is necessary to calculate the model prediction uncertainty resulting from the uncovered flux errors. This means that the flux uncertainties must be propagated through any model in order for it to be correctly compared with the result.

When doing so, another issue arises though. The experimental results will have a flux contribution to their covariance matrix. When the model predictions also gain a flux error, those should be correlated to calculate correct goodness of fit scores. Otherwise the correlations can lead to overcoverage or undercoverage.

In some cases—like the example in Sec. I A—the flux uncertainty can be cleanly broken up into a part that affects the measured result, e.g., the normalization, and a part that affects the model prediction, e.g., the shape. In those cases it should be possible to have uncorrelated flux errors in data and model prediction. In general this is not the case though, and the flux uncertainty cannot be split into an experimental and a theoretical part (see e.g., Sec. I B). In those cases the correlations need to be taken into account and the process of comparing a model to the data becomes a more involved statistical issue.

Let us assume a cross-section measurement reports the point estimates for a vector of cross-section values $\hat{\boldsymbol{x}}$, a vector of flux parameters $\hat{\boldsymbol{\phi}}$, as well as the covariance matrix S. In general, S describes the uncertainties in the measurement and the correlations between cross-section bins and the flux parameters. If the flux and cross-section uncertainties are uncorrelated, this will simply mean that the respective off-diagonal elements of S will be 0.[4] Furthermore let us assume a cross-section model that can, given a set of flux parameter values, produce a prediction $\boldsymbol{m}(\boldsymbol{\phi})$ for the values of $\boldsymbol{x}$.

For an ideal likelihood ratio test of the model, we would like to calculate the log likelihood ratio,

$$-2\lambda = -2 \sup_{\boldsymbol{\phi}} [\ln(L(\boldsymbol{m}(\boldsymbol{\phi}), \boldsymbol{\phi}))], \tag{4}$$

where sup is the supremum function, maximizing over $\boldsymbol{\phi}$, and

$$-2\ln(L(\boldsymbol{x}, \boldsymbol{\phi})) = (\boldsymbol{x}^T - \hat{\boldsymbol{x}}^T \quad \boldsymbol{\phi}^T - \hat{\boldsymbol{\phi}}^T) S^{-1} \begin{pmatrix} \boldsymbol{x} - \hat{\boldsymbol{x}} \\ \boldsymbol{\phi} - \hat{\boldsymbol{\phi}} \end{pmatrix} \tag{5}$$

describes the likelihood surface of the parameter space. If the model is true, -2λ will be χ^2_k-distributed with k equal to the number of bins in $\boldsymbol{x}$. The usual quantiles of χ^2_k can then be used to judge the goodness of fit of the model to the data.

Unfortunately, depending on the complexity of calculating $\boldsymbol{m}(\boldsymbol{\phi})$, it might not be feasible to maximize λ over the flux parameter space. In that case one can use the following likelihood ratio as approximation:

$$\begin{aligned} -2\lambda' &= -2\ln(L'(\boldsymbol{m}(\hat{\boldsymbol{\phi}}))) \\ &= (\boldsymbol{m}^T(\hat{\boldsymbol{\phi}}) - \hat{\boldsymbol{x}}^T) S'^{-1} (\boldsymbol{m}(\hat{\boldsymbol{\phi}}) - \hat{\boldsymbol{x}}). \end{aligned} \tag{6}$$

Here S' is *not* just the cross-section part of S. It is a new covariance matrix that describes the expected variation of the distribution of $(\boldsymbol{m}^T(\hat{\boldsymbol{\phi}}) - \hat{\boldsymbol{x}}^T)$ and thus $-2\lambda'$, assuming that the model is correct. It can be calculated from S, $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{\phi}}$, and $\boldsymbol{m}(\boldsymbol{\phi})$ by generating random samples of $\boldsymbol{x}$ and $\boldsymbol{\phi}$ and then calculating

$$\boldsymbol{\Delta} = (\boldsymbol{m}(\boldsymbol{\phi}) - \boldsymbol{x}) \tag{7}$$

for each sample. The sample covariance of $\boldsymbol{\Delta}$ can then be used as an estimator for S'. If there is no correlation between the flux and cross-section parameters, this procedure is equivalent to calculating the covariance of $\boldsymbol{m}(\boldsymbol{\phi})$ only and adding it to the cross-section part of S.

Note that we are ignoring the sample mean of $\boldsymbol{\Delta}$. We calculate a distribution of differences between data and prediction at the best-fit point of the data and then apply the resulting variance to the model prediction. This approach makes two assumptions:

(i) The covariance matrix S, describing the likelihood surface of the cross-section and flux parameters, can be interpreted as the covariance of the expected spread of maximum likelihood estimators (MLEs).
(ii) The covariance of the MLEs is constant in the relevant parameter range, while the expectation value depends on the true parameter value.

[4]Conversely, if an experiment only provides two separate covariance matrices for the cross-section result and the flux parameters, one can only treat them as uncorrelated.

These assumptions should hold well enough for all measurements that make a Gaussian error approximation—as is implied by reporting the uncertainty as a covariance matrix. The second assumption might break if the model prediction $\boldsymbol{m}(\boldsymbol{\phi})$ and the best fit data result $\hat{\boldsymbol{x}}$ are very different. But this should not change the physics conclusions of a model comparison.[5] If all assumptions hold and the model is actually true, the test statistic $-2\lambda'$ should again be χ^2_k distributed and the usual critical values apply.

III. CHALLENGES IN THE SECOND APPROACH

The second approach—reporting a cross section averaged over a *reference* flux—makes it easier to compare the results to model predictions. The covariance matrix already includes by construction the uncertainties of extrapolating the data from the unknown real neutrino flux to the reference flux. When testing a model, one only needs to calculate the flux-averaged cross section in that reference flux.

It is however more difficult to produce a result in this manner, because the extrapolation to a reference flux is not trivial. It necessarily requires a cross-section model to do so, since only a model can predict the connection between true neutrino energy and measured variables,[6] e.g., neutrino kinematics, particle multiplicities, etc.

So the extrapolation itself—and with it the cross-section result—will be subject to cross-section model uncertainties. This somewhat undermines the aim of producing a model-independent measurement. If the tested model was not covered by the assumed model-uncertainties—i.e., if it requires a different propagation of flux shape uncertainties compared to the considered models—the method of comparing the model with the data at only the reference flux breaks down again.

To illustrate this, consider the example in Sec. I A. If the model predicting a cross section proportional to the neutrino energy was used to extrapolate to the reference flux, the total uncertainty on the result would be about 16%. Now consider a second model, which predicts a cross section proportional to the square of the neutrino energy and overpredicts the measured value by 50%. With the reported uncertainty, this would be considered a three-sigma effect, but the squared neutrino energy dependence means that the 15% energy scale uncertainty should translate to a 30% rate uncertainty. Again the model is more compatible with the data than the covariance matrix implies.

This suggests that one should be conservative and cover as many possible flux propagation models as possible. Unfortunately this also degrades the power of the measurement. Imagine a third model that predicts a cross section that is constant over all neutrino energies, and overestimates the data of the example in Sec. I A by 15%. With the assumed flux shape uncertainties, this would not even be a one-sigma difference. But when considering that the model does not predict any shape dependence, it should be refuted at the three-sigma level from the 5% flux normalization uncertainty.

Overall, the feasibility of the second approach will depend on the size of the flux shape uncertainty and how well constrained the neutrino-energy dependence of the measured cross section is. If the effect is large, and the energy dependence is uncertain, it might be better to opt for a first-approach measurement instead.

IV. SIZE OF THE EFFECT

Whichever approach is taken, the effect we consider here comes from the flux shape dependence of the cross-section prediction. In order to assess the relevance of this effect for modern cross-section measurements, the spread of flux-averaged cross section predictions are evaluated for a fixed model propagated through an ensemble of flux predictions. These flux predictions are constructed from T2K [2] and MINERvA [3] nominal fluxes and accompanying covariance matrices.[7] Whilst the flux prediction is varied in both shape and normalization, the latter plays no role in constructing a flux-averaged cross section from a model prediction. The model used is provided by GENIE [4] (Version 3.00.06, tune G18_10b_00_000). The specific cross sections that are predicted are MINERvA's CCQE-like measurement from Ref. [5] and T2K's CC0π Analysis I in Ref. [6]. Both of these cross sections are double differential in outgoing muon kinematics. MINERvA measures the momentum broken down into its transverse (p_T) and longitudinal ($p_{||}$) components, defined with respect to the incoming neutrino direction, whilst T2K measures the magnitude of the momentum (p_μ) alongside the cosine of the angle between the muon and neutrino [$\cos(\theta_\mu)$]. Both measurements also follow the first approach, i.e., no attempt was made to extrapolate the measured cross section to a particular reference flux shape. The model is compared to the cross sections using the NUISANCE framework [1].

The sampled ensembles of flux predictions are shown in Fig. 1. The standard deviation of the mean neutrino energies across the ensemble std($\langle E_\nu \rangle$) is ~5.5 MeV for T2K and ~6.5 MeV for MINERvA. However, it should be noted that this energy scale uncertainty is not fully representative of the flux shape uncertainties. The spread

[5]At some point it is no longer important whether the data to model agreement is bad or terrible.

[6]Measuring a cross section in terms of neutrino energy directly is of course an exception to this. But in that case this whole paper becomes a moot point.

[7]The MINERvA flux extends up to 100 GeV but the numerical precision of the provided covariance matrix makes the complete matrix noninvertible. Here we consider only the first 35 GeV (containing ~99.4% of the flux). It should also be noted that the T2K covariance is provided in much coarser bins than the accompanying flux prediction.

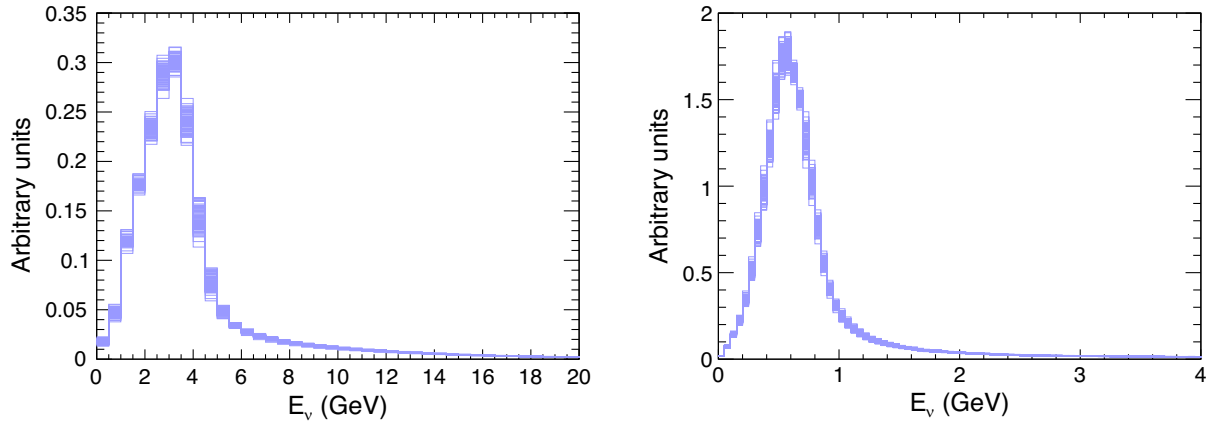

FIG. 1. Histograms showing the ensembles of flux predictions produced for MINERvA (left) and T2K (right) which have each been renormalized to have the same integral. Note that the actual flux predictions extend up to 35 GeV for MINERvA and 10 GeV for T2K but for readability the long tails are not shown.

of the cross-section predictions is shown in Fig. 2 alongside the cross-section measurements from the experiments. It also shows a bin-by-bin comparison of the spread of the cross-section predictions compared to the uncertainty on the measurements. Ignoring all correlations, this gives a qualitative indication of how much of the error budget of the measurement is consumed by the flux shape prediction uncertainty, despite not being correctly included in the error propagation of the results. That is, this is a rough measure for the "missing" uncertainty when treating these first-approach measurements like second-approach ones.

Although the spread in the prediction of the total cross section integrated across all bins is about 1% for both MINERvA and T2K, the spread in individual bins of the differential cross section is much larger. This is especially the case for bins corresponding to forward going muons with large momenta which are produced only by interactions of neutrino in the tail end of the flux predictions, where the flux uncertainty is largest. In these bins, the spread of the cross-section predictions is sometimes comparable to the entire uncertainty in the measured cross section. This said, it should of course be noted that the uncertainty on each cross-section result is characterized by a full covariance matrix describing the correlations between bins rather than just the error bars. It is therefore difficult to quantitatively compare the size of the model prediction variations with the variations allowed by the cross-section uncertainty from only Fig. 2.

A more quantitative analysis of the relevance of the flux shape uncertainty can be performed by assessing how the χ^2 statistic, calculated using the full covariance matrix to describe the uncertainty on the cross-section measurements, changes across the ensemble of flux predictions. This χ^2 is defined as

$$\chi^2 = (\boldsymbol{x}_{\mathrm{MC},i} - \boldsymbol{x}_{\mathrm{data}})^T S_{\mathrm{data}}^{-1} (\boldsymbol{x}_{\mathrm{MC},i} - \boldsymbol{x}_{\mathrm{data}}), \tag{8}$$

where S_{data} is the covariance matrix as reported by the experiments,[8] $\boldsymbol{x}_{\mathrm{MC},i}$ is the predicted cross section in the ith flux shape variation and $\boldsymbol{x}_{\mathrm{data}}$ is the cross section measured by the experiments. Figure 3 shows that this χ^2 is much larger than the number of cross-section bins and so it is therefore hard to interpret beyond stating that all of the predictions from the ensemble of fluxes are in very poor agreement with the measurements. This would be the case for the majority of currently available cross-section models.

As discussed in Sec. II, if potential correlations between the cross-section uncertainty and the spread of predictions from the ensemble of fluxes are neglected, an approximate combined covariance matrix can be formed by adding the covariance of the ensemble of predictions to the one cross-section measurement covariance. This $\chi^2 = -2\lambda'$—calculated using the combined covariance—is also shown in Fig. 3. Whilst it is clear there is a notable shift to lower χ^2 when the combined uncertainty is considered, the model remains absolutely disfavored by the measurements.

An easier to interpret assessment of the relevance of the flux shape uncertainty compared to the rest of the cross-section uncertainty can be constructed by considering a scenario in which the T2K and MINERvA analyses had measured exactly the GENIE prediction with the nominal flux. For simplicity this scenario assumes that the covariance of the data remains unchanged which, whilst unrealistic, is a suitable approximation for this illustrative example. In this scenario, the spread of the χ^2, calculated as in Eq. (8) but where $\boldsymbol{x}_{\mathrm{data}}$ is replaced by the GENIE model prediction for the nominal flux, should indicate more

[8]Note that 12 of the 156 cross-section bins in the MINERvA data release contain zero measured cross section with zero uncertainty to reflect the kinematic acceptance of the detector. The resultant covariance matrix is therefore noninvertible. To calculate a χ^2 statistic we followed the same approach as MINERvA and formed a pseudoinverse using a singular value decomposition approach.

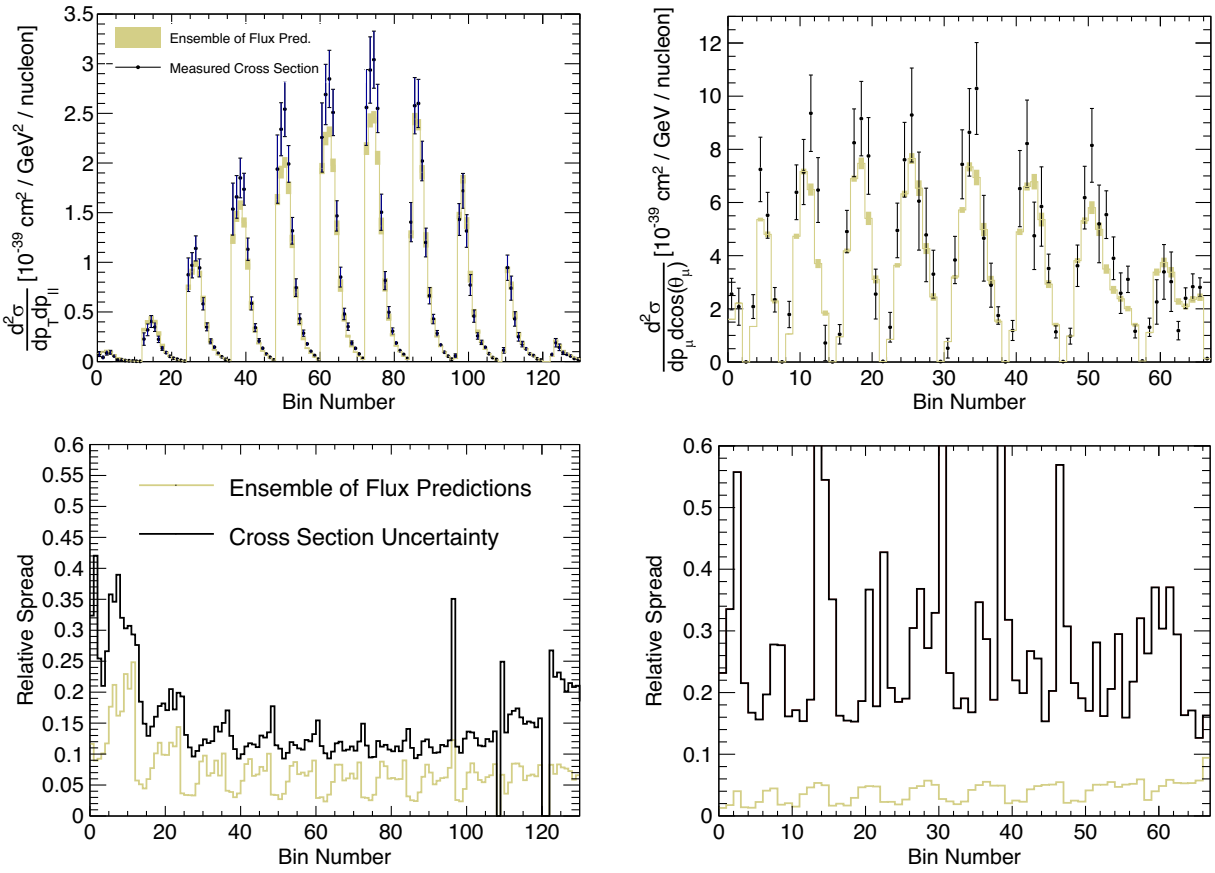

FIG. 2. Top: the differential cross sections measured by MINERvA (left) and T2K (right) are shown alongside the GENIE model prediction with an error band taken from the spread (standard deviation) of the predictions generated with the ensemble of flux predictions. Bottom: the relative uncertainty (i.e., 0.2 means 20% uncertainty) on the cross section measured by MINERvA (left) and T2K (right) is compared to the spread in the model predictions. The x axis in each plot is simply a bin number. For MINERvA the bins are ordered in increasing $p_{||}$ in slices of increasing p_T. For T2K the bins are ordered in increasing p_μ in slices of increasing $\cos(\theta_\mu)$. The very forward high momentum bins discussed in the text correspond to around bin 10 for MINERvA and the final few bins for T2K. The MINERvA analysis has additional high-p_T bins which are not shown for readability since they contribute a negligible proportion of the cross section.

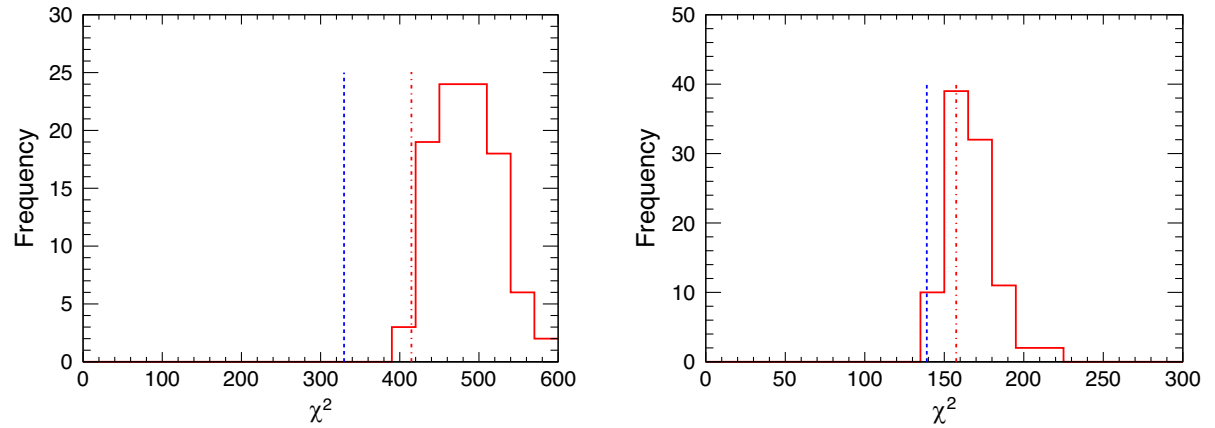

FIG. 3. χ^2 comparison. The "original" χ^2 calculated using the GENIE prediction using the nominal flux and the original covariance matrix is shown by the dot-dashed red line for MINERvA (left, 144 nonzero cross-section bins) and T2K (right, 67 cross-section bins). The distribution indicated by the solid red line shows the same but with a different model prediction for each varied flux prediction. The dashed blue line shows the χ^2 from the comparison of the nominal prediction with the data using the modified covariance matrix, corresponding to the $-2\lambda'$ statistic introduced in Sec. II.

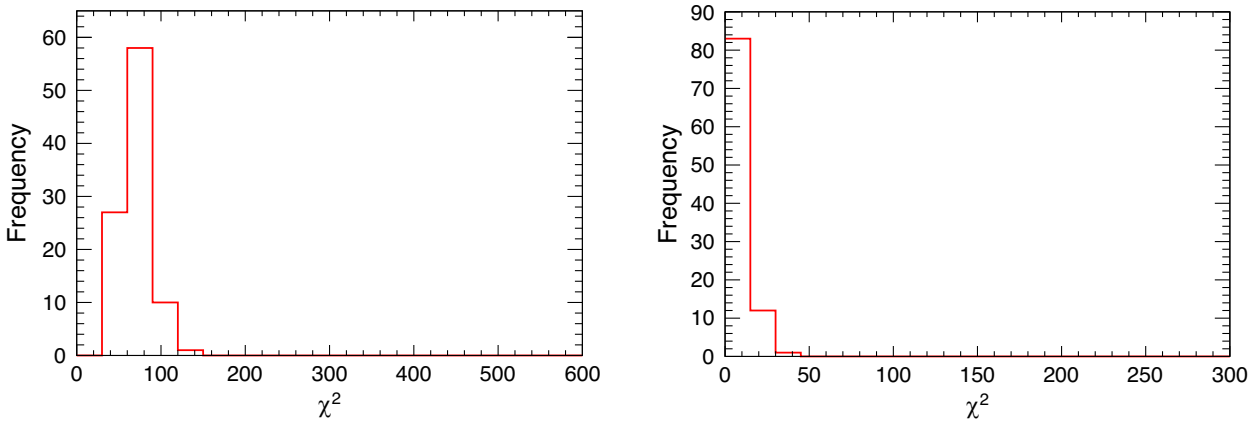

FIG. 4. The χ^2 calculated for each of the flux-shape-varied model predictions compared to the nominal flux GENIE prediction of the MINERvA (left, 144 nonzero cross-section bins) and T2K (right, 67 cross-section bins) cross-section measurements. The χ^2 are calculated using the original covariance matrices provided by T2K and MINERvA.

directly the relative importance of the flux shape uncertainty. This is shown in Fig. 4. The fact that this χ^2 is almost always less than the number of cross-section bins shows that the uncertainty from the flux shape is in a sense "covered" by the other uncertainties on the cross section. However, the change in the χ^2 is clearly large enough such that conclusions regarding the suitability of slightly altered model predictions could be significantly changed if the flux shape variations were not to be considered. Note that this study also continues to neglect potential correlations between the ensemble the flux predictions and the cross-section covariance. This should not affect the qualitative conclusions by much though, since the flux shape contribution to the uncertainty of the cross-section results is relatively small. Overall the relative size of the flux shape uncertainty appears non-negligible but subdominant in the analyses considered. A more quantitative statement would require more in-depth studies in the experiments, following the recipes from Sec. V.

V. RECIPES

The following sections will describe how to implement the two approaches within certain unfolding and error propagation schemes. This is not intended as an endorsement of these schemes, nor can these instructions be blindly applied to different algorithms. They should, however, provide examples of how to approach the issue in general.

A. Template fitter

This section concerns the template fitting approach of unfolding as e.g., described for the "STV" analysis in [7]. In short:

(i) Interaction models and detector simulations are used to create a prediction function of reconstructed events N_{reco} as a function of multiple parameters, e.g., by reweighting.

(ii) A fitting algorithm is used to determine which combination of parameter values is or is not compatible with the real data. This information could be encoded in a best-fit point plus a covariance matrix of the parameters, for example.

(iii) Sets of compatible parameter values are used to calculate the desired cross sections. The spread of compatible value sets translates to a spread of cross-section results, which e.g., can be parametrized as another covariance matrix.

The choice of parameters to vary and how to implement that variation is a complicated topic and outside the scope of this paper. Let us just split the parameters into parameters that affect the assumed neutrino flux, the flux parameters $\boldsymbol{\phi}$, and parameters that do not $\boldsymbol{\theta}$. The latter will usually contain parameters for the detector response, the (background) cross-section models, and the primary aim of the measurement: the template weight parameters for the measured cross section,

$$\sigma = \frac{N(\boldsymbol{\theta}, \boldsymbol{\phi})}{T(\boldsymbol{\theta})\Phi(\boldsymbol{\phi})}. \tag{9}$$

Here N is the predicted number of *true* events.[9] Depending on the fitting algorithm this is either a direct output of the fit, or it can be calculated from the number of predicted *reconstructed* events via a predicted[10] efficiency ϵ:

$$N(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{N_{\text{reco}}(\boldsymbol{\theta}, \boldsymbol{\phi})}{\epsilon(\boldsymbol{\theta}, \boldsymbol{\phi})}. \tag{10}$$

[9]For simplicity's sake, we will only consider total cross sections here. All arguments apply equally to differential measurements though. In fact, the calculations should be identical modulo a division by the areas of the analysis bins.

[10]Ensuring that the efficiency does not depend too much on any assumed cross-section or flux model is another can of worms that shall remain unopened in this paper.

Like this, the method produces a first-approach result. The propagated sets of parameter values correspond to possible *real* experimental conditions, and thus the calculated cross sections do the same. It is easy to imagine that under ideal experimental conditions, the variation of $N(\boldsymbol{\theta}, \boldsymbol{\phi})$ will be small as the fitter ensures that the allowed parameter values fit the number of recorded events, which is a constant number. In this case, the flux uncertainty will affect the cross-section result solely via the total flux $\Phi(\boldsymbol{\phi})$. Any shape uncertainty would have only a small effect via N, or no effect at all.

At this point it is also fairly easy to calculate a covariance matrix that correlates the cross-section result with the flux parameters. Instead of only calculating the cross-section values for each throw of parameters, one also treats the thrown flux parameters themselves as part of the result and calculates a covariance matrix over the combined cross-section values and flux parameter values.

To extract a second-approach result from the fitter, one needs to extrapolate the results of a set of parameter values to a reference flux. We can specify the reference flux using a fixed set of flux parameters: $\boldsymbol{\phi}'$. Let us call the extrapolation of expected events at that flux N'. The calculated cross section then becomes

$$\sigma' = \frac{N'(\boldsymbol{\theta}, \boldsymbol{\phi})}{T(\boldsymbol{\theta})\Phi(\boldsymbol{\phi}')}. \tag{11}$$

It might seem counterintuitive at first to fix the flux normalization to the $\boldsymbol{\phi}'$ parameter values in this formula, when the aim is to include *more* of the flux uncertainty in the result. This is not a contradiction though, since the flux uncertainty now enters the result via the extrapolation in N'.

The exact procedure will again depend on the flux parameters, but let us assume the parameters are—or can be converted to—event weights binned in neutrino energy. In this case, N' can be expressed as

$$N'(\boldsymbol{\theta}, \boldsymbol{\phi}) = \sum_i N_i(\boldsymbol{\theta}, \boldsymbol{\phi}) \frac{\phi_i'}{\phi_i}, \tag{12}$$

with the neutrino energy bin index i, and the number of true events in each bin N_i. In this form it is clear how the flux uncertainty enters the result. Under ideal conditions, the fitter output will ensure that $\sum_i N_i(\boldsymbol{\theta}, \boldsymbol{\phi})$ is compatible with the number of observed events by correlating the flux weights and the other parameters, i.e., N_i will not vary much under allowed parameter variations. Then the ratio of the varied weights $\boldsymbol{\phi}$ and the constant reference flux weights $\boldsymbol{\phi}'$ modifies that number according to the flux uncertainty.

We can further separate the flux weights from the true event prediction:

$$N_i(\boldsymbol{\theta}, \boldsymbol{\phi}) = N_i(\boldsymbol{\theta})\phi_i, \tag{13}$$

where $N_i(\boldsymbol{\theta})$ is the number of predicted true events in the ith energy bin in an unweighted, nominal[11] flux. So Eq. (12) becomes

$$N'(\boldsymbol{\theta}, \boldsymbol{\phi}) = \sum_i N_i(\boldsymbol{\theta})\phi_i \frac{\phi_i'}{\phi_i} = N(\boldsymbol{\theta}, \boldsymbol{\phi}'). \tag{14}$$

This means the extrapolation to the reference flux happens by simply ignoring the thrown flux weights and calculating the cross section for the reference flux:

$$\sigma' = \frac{N(\boldsymbol{\theta}, \boldsymbol{\phi}')}{T(\boldsymbol{\theta})\Phi(\boldsymbol{\phi}')}. \tag{15}$$

It might seem, on first glance, like there is no flux error at all propagated any more, since the only flux parameters in the equation are the *constant* reference flux parameters $\boldsymbol{\phi}'$. The uncertainty does however enter indirectly via the correlations of the other parameters $\boldsymbol{\theta}$ with the flux.

This also holds true when considering the efficiency correction as a separate step. Equation (10) can be expanded to

$$N(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{N_{\text{reco}}(\boldsymbol{\theta}, \boldsymbol{\phi})}{\epsilon(\boldsymbol{\theta}, \boldsymbol{\phi})} = \sum_i \frac{N_{\text{reco,i}}(\boldsymbol{\theta})\phi_i}{\epsilon_i(\boldsymbol{\theta})}, \tag{16}$$

where $N_{\text{reco},i}$ is the expected reconstructed number of events coming from the true energy bin i. In this case, Eq. (14) becomes

$$N'(\boldsymbol{\theta}, \boldsymbol{\phi}) = \sum_i \frac{N_{\text{reco,i}}(\boldsymbol{\theta})\phi_i}{\epsilon_i(\boldsymbol{\theta})} \frac{\phi_i'}{\phi_i} = \frac{N_{\text{reco}}(\boldsymbol{\theta}, \boldsymbol{\phi}')}{\epsilon(\boldsymbol{\theta}, \boldsymbol{\phi}')}, \tag{17}$$

and the cross section

$$\sigma' = \frac{N_{\text{reco}}(\boldsymbol{\theta}, \boldsymbol{\phi}')}{\epsilon(\boldsymbol{\theta}, \boldsymbol{\phi}')T(\boldsymbol{\theta})\Phi(\boldsymbol{\phi}')}. \tag{18}$$

It seems strange that we would use the efficiency of the reference flux to correct our real data, but in actuality we are efficiency correcting the data we *would have seen* in the reference flux $N_{\text{reco}}(\boldsymbol{\theta}, \boldsymbol{\phi}')$, not our real data.

Lastly, we need to decide which flux to use as the reference flux. Since this can be done *after* the parameter fit, the most logical choice is probably to use the best fit point of the flux parameters:

$$\boldsymbol{\phi}' = \hat{\boldsymbol{\phi}}. \tag{19}$$

The reason for this is not just that this is the best estimate for the real flux, but it is also the point in the parameter space where any approximations and linearizations done by

[11]This is not necessarily the same as the reference flux!

the fitter, like e.g., the error treatment as a covariance matrix, are most valid. Alternatively it should also be valid to use the experiment's nominal, or "design" flux. This would have the benefit of easier comparisons of multiple measurements done in the same neutrino beam. On the other hand, that should probably be done with a first-approach measurement anyway.

B. Multiverse unfolding

This section will deal with the "classical" unfolding approach. The general procedure is as follows:

(i) Interaction models and detector simulations are used to create a detector response matrix (also called "smearing matrix") that describes how likely an event in a certain true kinematic bin is to be reconstructed in another reconstructed kinematic bin.
(ii) An unfolding algorithm is used with the recorded data and the matrix to reconstruct an estimator for the true number of events in each true kinematics bin.
(iii) The smearing matrix is varied according to the systematic uncertainties of the detector and interaction models. Each varied matrix leads to a different unfolding and thus a different cross-section result. The different results can be used to build a covariance matrix.

A more in-depth example of this is described e.g., in [5,8].

Let us again split the parameters into flux parameters and nonflux parameters. The equation to calculate the cross section looks the same as in Eq. (9):

$$\sigma = \frac{N|_{\boldsymbol{\theta},\boldsymbol{\phi}}}{T(\boldsymbol{\theta})\Phi(\boldsymbol{\phi})}. \tag{20}$$

In this case however, $N|_{\boldsymbol{\theta},\boldsymbol{\phi}}$ is the unfolded (and efficiency corrected) result, which implicitly depends on the recorded data.[12] This corresponds to a first-approach measurement. Under ideal circumstances, the flux parameters have no effect on the unfolding procedure, and the flux uncertainty enters mainly via the integrated flux in the denominator. If the flux shape does influence the unfolding algorithm in some way, the correlations of flux and cross sections have to be taken into account in the model comparisons. For this—just like in the fitter case—it should be easy to create a covariance matrix that correlates the flux parameters with the cross-section values by treating the flux parameters as part of the result rather than just an input.

To turn this into a second-approach measurement, one again needs to find an extrapolation function $N'(\boldsymbol{\theta},\boldsymbol{\phi})$. Unlike in the fitter case, this presents a problem though: The unfolding algorithm will only provide the total number of events N, but not the relative contribution from the different neutrino energy bins N_i. To get this information from the unfolding process, one would need to explicitly do the unfolding of the neutrino energy as well. This is often deliberately avoided because of insufficient statistics, limited detector capabilities, and the general impossibility of measuring the neutrino energy without assuming some sort of interaction model.

Let us assume that it is possible to modify the measurement in a way to include this information though. In this case, the extrapolation function will look the same as in Eq. (12):

$$N'(\boldsymbol{\theta},\boldsymbol{\phi}) = \sum_i N_i|_{\boldsymbol{\theta},\boldsymbol{\phi}} \frac{\phi_i'}{\phi_i}. \tag{21}$$

This time it is not possible to further simplify this though, since the flux weights are not a direct multiplicative factor in the unfolding function.[13] The second-approach cross section then becomes

$$\sigma' = \frac{N'(\boldsymbol{\theta},\boldsymbol{\phi})}{T(\boldsymbol{\theta})\Phi(\boldsymbol{\phi}')} = \frac{\sum_i N_i|_{\boldsymbol{\theta},\boldsymbol{\phi}} \frac{\phi_i'}{\phi_i}}{T(\boldsymbol{\theta})\Phi(\boldsymbol{\phi}')}. \tag{22}$$

One might wonder, why not just report a differential cross section over the neutrino energy in this case? A flux-averaged measurement might still be preferable under certain circumstances, since any model and detector uncertainties regarding the neutrino energy reconstruction only enter the result via the flux error propagation.

If the unfolded result is not available in neutrino energy bins even as an intermediary step, the only way to extrapolate the result to a different flux is to assume a certain energy spectrum in the result. The obvious choice is to use the Monte Carlo data that is used in the unfolding process:

$$N'(\boldsymbol{\theta},\boldsymbol{\phi}) = \frac{N|_{\boldsymbol{\theta},\boldsymbol{\phi}}}{N_{\mathrm{MC}}(\boldsymbol{\theta},\boldsymbol{\phi})} \sum_i N_{\mathrm{MC},i}(\boldsymbol{\theta},\boldsymbol{\phi}) \frac{\phi_i'}{\phi_i} \tag{23}$$

$$= \frac{N|_{\boldsymbol{\theta},\boldsymbol{\phi}}}{N_{\mathrm{MC}}(\boldsymbol{\theta},\boldsymbol{\phi})} \sum_i N_{\mathrm{MC},i}(\boldsymbol{\theta})\phi_i \frac{\phi_i'}{\phi_i} \tag{24}$$

$$= \frac{N|_{\boldsymbol{\theta},\boldsymbol{\phi}}}{N_{\mathrm{MC}}(\boldsymbol{\theta},\boldsymbol{\phi})} \sum_i N_{\mathrm{MC},i}(\boldsymbol{\theta})\phi_i' \tag{25}$$

$$= \frac{N|_{\boldsymbol{\theta},\boldsymbol{\phi}}}{N_{\mathrm{MC}}(\boldsymbol{\theta},\boldsymbol{\phi})} \sum_i N_{\mathrm{MC},i}(\boldsymbol{\theta},\boldsymbol{\phi}') \tag{26}$$

$$= N|_{\boldsymbol{\theta},\boldsymbol{\phi}} \frac{N_{\mathrm{MC}}(\boldsymbol{\theta},\boldsymbol{\phi}')}{N_{\mathrm{MC}}(\boldsymbol{\theta},\boldsymbol{\phi})}. \tag{27}$$

[12] We again ignore any indices or bin widths associated with a differential measurement.

[13] As mentioned before, ideally the effect of the flux parameters on the unfolding should be negligible. A constant flux weight of 200% over all neutrino energy bins does not change the detector response matrix. Any effect of the flux shape should also be suppressed by the analysis design.

Here $N_{\mathrm{MC}(,i)}(\boldsymbol{\theta},\boldsymbol{\phi})$ is the number of true events (in the ith neutrino energy bin) in the simulated data assuming the given set of nuisance parameters.

So the second-approach cross section becomes

$$\sigma' = \frac{N|_{\boldsymbol{\theta},\boldsymbol{\phi}}}{T(\boldsymbol{\theta})\Phi(\boldsymbol{\phi}')}\frac{N_{\mathrm{MC}}(\boldsymbol{\theta},\boldsymbol{\phi}')}{N_{\mathrm{MC}}(\boldsymbol{\theta},\boldsymbol{\phi})}. \tag{28}$$

In both cases there is no "best fit" set of flux parameters available, so the logical choice for the reference flux is the central value of the parameter throws, i.e., the nominal flux:

$$\boldsymbol{\phi}' = \langle\boldsymbol{\phi}\rangle. \tag{29}$$

C. Hybrid measurement

Hybrid approaches somewhere in between the classical multiverse unfolding and template fitting have also been used, e.g., in [6]. Here the unfolding is repeated multiple times under differing flux and other systematic assumptions like in the multiverse approach, but the unfolding in each case is done by doing a template fit. The fit might or might not include the freedom to vary the thrown parameters within varied constraints.

If the fits are *not* free to vary the flux parameters, this setup is functionally identical to the classical unfolding case, as far as the flux uncertainty propagation is concerned. The extrapolation can be simplified though if the fits all use the same parametrization for the predicted number of true events (just with different set points):

$$N|_{\boldsymbol{\theta},\boldsymbol{\phi}} = N(\hat{\boldsymbol{\theta}}|_{\boldsymbol{\theta},\boldsymbol{\phi}},\boldsymbol{\phi}) = \sum_i N_i(\hat{\boldsymbol{\theta}}|_{\boldsymbol{\theta},\boldsymbol{\phi}})\phi_i, \tag{30}$$

$$N'(\boldsymbol{\theta},\boldsymbol{\phi}) = \sum_i N_i(\hat{\boldsymbol{\theta}}|_{\boldsymbol{\theta},\boldsymbol{\phi}})\phi_i\frac{\phi_i'}{\phi_i} = N(\hat{\boldsymbol{\theta}}|_{\boldsymbol{\theta},\boldsymbol{\phi}},\boldsymbol{\phi}'). \tag{31}$$

Here $\hat{\boldsymbol{\theta}}|_{\boldsymbol{\theta},\boldsymbol{\phi}}$ is the best-fit result of the fit under the assumptions of the thrown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. Note that the extrapolation function is *not* equal to the unfolding result at the reference flux, as the flux assumptions of the fit are still the ones for the thrown parameters.

This method of error propagation is *not* suitable if the flux parameters are free to vary in the fit. Depending on the relative constraint from the prior assumption and the fitted data, the best-fit estimate of the parameters will vary much less than the data constraint allows.[14] Thus, using only the variation of the best fit point as a measure for the uncertainty will lead to undercoverage. Instead, the post-fit covariance matrix of the fitted parameters needs to be taken into account like in the regular template fitter case. Since the covariance matrix will be different for every fit under different thrown systematic assumptions, a new set of post-fit parameters will have to be drawn for each. The cross-section uncertainty can then be deduced from the spread of all of these results.

VI. CONCLUSIONS

We have shown the difference between reporting a flux-averaged cross-section measurement in the real flux—a first-approach measurement—and reporting a flux-averaged cross-section measurement in a reference flux—a second-approach measurement. The difference between the two is subtle, and even if the central values of the two are identical, the resulting covariances can be very different.[15] When used carelessly, this can lead to drawing the wrong conclusions from model comparisons to the data.

It has been shown qualitatively that in the case of the two exemplary CCQE-like measurements of T2K and MINERvA and the evaluated Genie model, the flux shape uncertainty seems to be a subdominant but non-negligible contribution to the total uncertainty, which is currently not fully taken into account. It is expected that in future cross-section measurements the statistical uncertainties will decrease as more data becomes available and so the relevance of the flux shape uncertainty will grow. For these future measurements it is possible that neglecting the flux shape uncertainty could lead to incorrect physics conclusions. A rigorous quantitative determination of the size of the effect would require dedicated studies by the experimental collaborations.

First-approach measurements are somewhat simpler to implement and it is possible to perform them without assuming anything about the cross-section model, at least in principle. They are, however, more difficult to compare with model predictions. If the flux shape uncertainty has no influence on the result, it is possible to vary the model predictions within those shape uncertainties and treat the resulting model uncertainty as an additional covariance on the cross-section result. If the flux shape uncertainty on the other hand *does* have an impact on the result, the model uncertainties would need to be correlated with the reported data uncertainties.

Second-approach measurements are much simpler to compare to models. Since they report the cross section and its uncertainty for a single, well-defined flux, the models will only need to generate a prediction at that one flux. The down side to this is that it is necessary to make assumptions about the neutrino energy dependence of the cross section to extrapolate from possible real fluxes to the reference flux. The impact of this can be minimized by choosing the

[14] If the prior constraint is much weaker than the data constraint, changing the prior's central value will not affect the χ^2 surface of the parameter. The fit will always return virtually the same result.

[15] This depends on the size of the flux shape uncertainties compared to all other uncertainties in the measurement.

reference flux carefully, e.g., using the best-fit result as reference. In that case, it should only introduce additional model dependence in the uncertainty propagation, not in the central value of the result. To be conservative, the model uncertainties will have to cover many possible neutrino energy dependencies. This means the result will lose some discrimination power when comparing it to a model with a single explicit energy dependence.

In summary, a first-approach measurement with a correlated flux uncertainty propagation in the model would yield the better discrimination power between the two approaches, but it requires extra effort at the time of model comparison. The result alone is not the whole story. A second-approach measurement is easy to compare to models. The covariance of the result is all there is to it. Unfortunately one loses some discrimination power due to the need of covering many potentially different energy dependencies in a single result. When treating a first-approach result like a second-approach result by only comparing a model at a single flux, flux shape errors are not correctly taken into account and wrong physics conclusions could be drawn. The effect in the evaluated example analyses of T2K and MINERvA seems to be not dominant, but not negligible either.

ACKNOWLEDGMENTS

We would like to thank Marco Del Tutto for the very useful discussions about forward-folding analysis techniques, which in a twist of fate first lead to the (re)discovery of the issue presented in this paper. This work was first presented and discussed within the T2K cross-section working group. We therefore are very grateful to all members of the group, but are especially thankful for the useful discussions with Callum Wilkinson and Luke Pickering (your initial scepticism drove us further), Kendall Mahn and Sara Bolognesi. In the process of writing this paper we even realized that Sara had first raised this issue to T2K more than five years ago! Unfortunately she was ahead of her time and the issue did not receive the attention it deserves. Further thanks goes to Andrew Cudd, who helped ensure that we understood how the fitter works. We also state our appreciation of the T2K neutrino beam line working group for providing us with a flux covariance matrix that we could use in this paper. We further thank the MINERvA collaboration for having made the tools needed to run the studies in Sec. IV publicly available and especially thank Dan Ruterbories for fruitful discussions. This work was supported by a grant from the Science and Technology Facilities Council.

[1] P. Stowell, C. Wret, C. Wilkinson, L. Pickering, S. Cartwright, Y. Hayato, K. Mahn, K. McFarland, J. Sobczyk, R. Terri, L. Thompson, M. Wascko, and Y. Uchida, NUISANCE: A neutrino cross-section generator tuning and comparison framework, J. Instrum. **12**, P01016 (2017).

[2] K. Abe *et al.*, T2K neutrino flux prediction, Phys. Rev. D **87**, 012001 (2013); **87**, 019902(A) (2013).

[3] L. Aliaga *et al.*, Neutrino flux predictions for the NuMI beam, Phys. Rev. D **94**, 092005 (2016); **95**, 039903(A) (2017).

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

[5] D. Ruterbories *et al.*, Measurement of quasielastic-like neutrino scattering at $\langle E_\nu \rangle \sim 3.5$ GeV on a hydrocarbon target, Phys. Rev. D **99**, 012004 (2019).

[6] K. Abe *et al.*, 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).

[7] K. Abe *et al.*, Characterization of nuclear effects in muon-neutrino scattering on hydrocarbon with a measurement of final-state kinematics and correlations in charged-current pionless interactions at T2K, Phys. Rev. D **98**, 032003 (2018).

[8] C. Patrick *et al.*, Measurement of the muon antineutrino double-differential cross section for quasielastic-like scattering on hydrocarbon at $E_\nu \sim 3.5$ GeV, Phys. Rev. D **97**, 052002 (2018).