Data-driven Model Validation For Neutrino Cross Section Measurements

Based on a new paper: arXiv:2411.03280

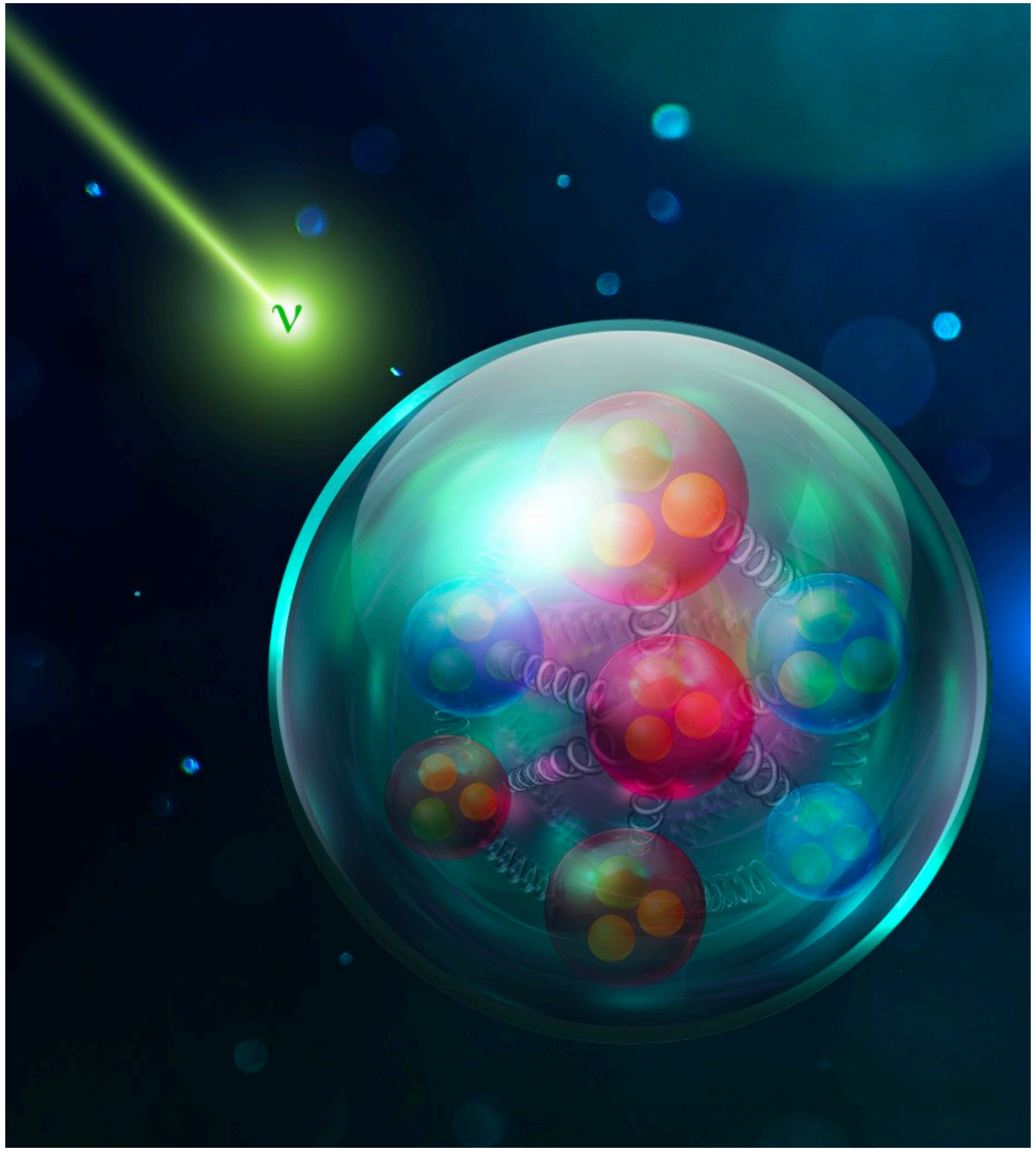
Lee Hagaman (University of Chicago)
On behalf of the MicroBooNE Collaboration

November 7, 2024
NuSTEC Cross Experiment Working Group Seminar

Run 3493 Event 27435

Why We Measure Cross Sections

- Oscillation experiments use neutrino interaction rates to infer a neutrino flux change, which requires a precise understanding of how neutrino interactions depend on true neutrino energy
  - DUNE and Hyper-K will need significantly better modeling than NOvA and T2K have needed so far
- Also, many BSM searches have neutrino backgrounds which must be correctly modeled
- We don't have a solid theoretical understanding of every interaction process, since O(GeV) neutrino interactions involve non-perturbative many-body QCD calculations
- In the absence of a full rigorous calculation of all processes, we must measure these interactions, using these measurements to improve our empirical nuclear models



https://www.jlab.org/news/releases/electrons-set-stage-neutrino-experiments-0

How We Measure Cross Sections

Protons on target **Number of targets** **Beam flux** **Cross section** **Detector response** **Selection efficiency** **Background**

$$M(T_{\text{rec}}) = POT \cdot T \cdot \int \phi\left(E_\nu\right) \cdot \sigma\left(E_\nu, T_{\text{true}}\right) \cdot D\left(T_{\text{true}} \rightarrow T_{\text{rec}}\right) \cdot \varepsilon\left(E_\nu, T_{\text{true}}\right) \cdot dE_\nu + B\left(T_{\text{rec}}\right)$$

Equation relating cross section to measured counts

How We Measure Cross Sections

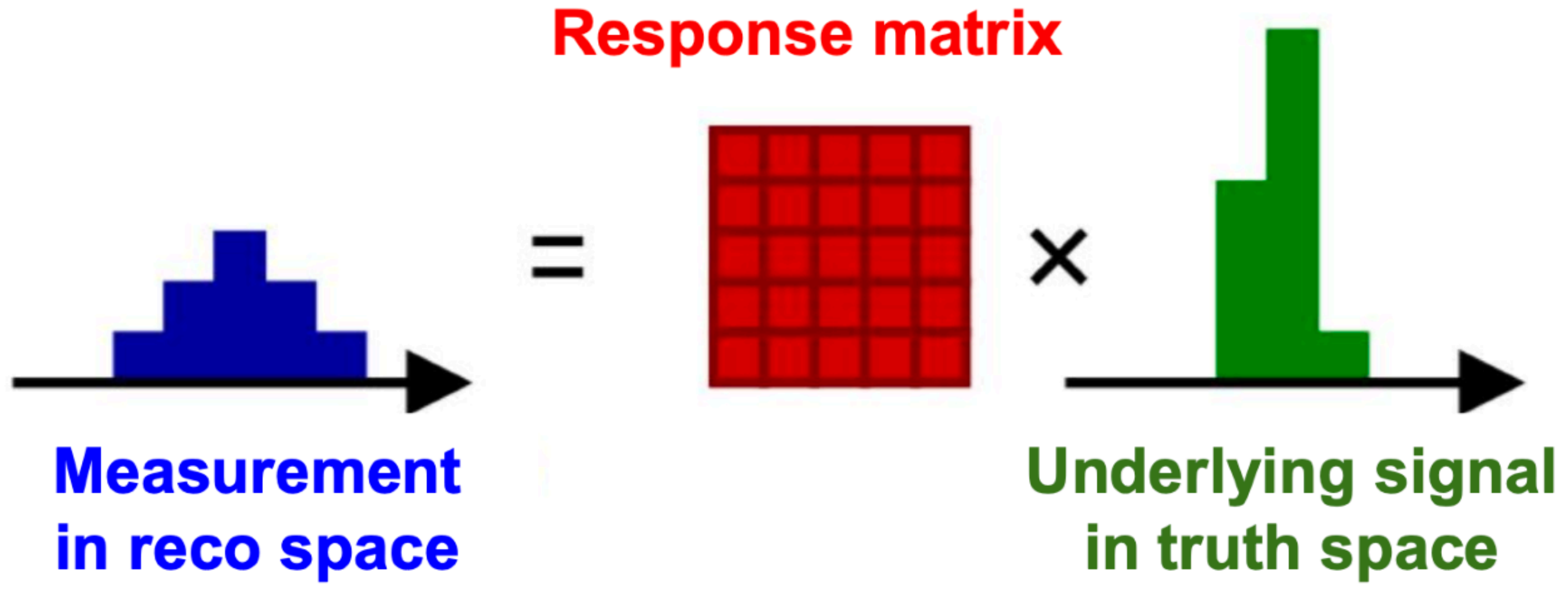
Protons on target | Number of targets | Beam flux | Cross section | Detector response | Selection efficiency | Background

$$M(T_{\text{rec}}) = POT \cdot T \cdot \int \phi\left(E_\nu\right) \cdot \sigma\left(E_\nu, T_{\text{true}}\right) \cdot D\left(T_{\text{true}} \rightarrow T_{\text{rec}}\right) \cdot \varepsilon\left(E_\nu, T_{\text{true}}\right) \cdot dE_\nu + B\left(T_{\text{rec}}\right)$$

Equation relating cross section to measured counts

$$M_i = R_{i,j} \cdot S_j + B_i$$

Writing it as a matrix equation



i: T_{reco} bin index

j: T_{true} bin index

How We Measure Cross Sections

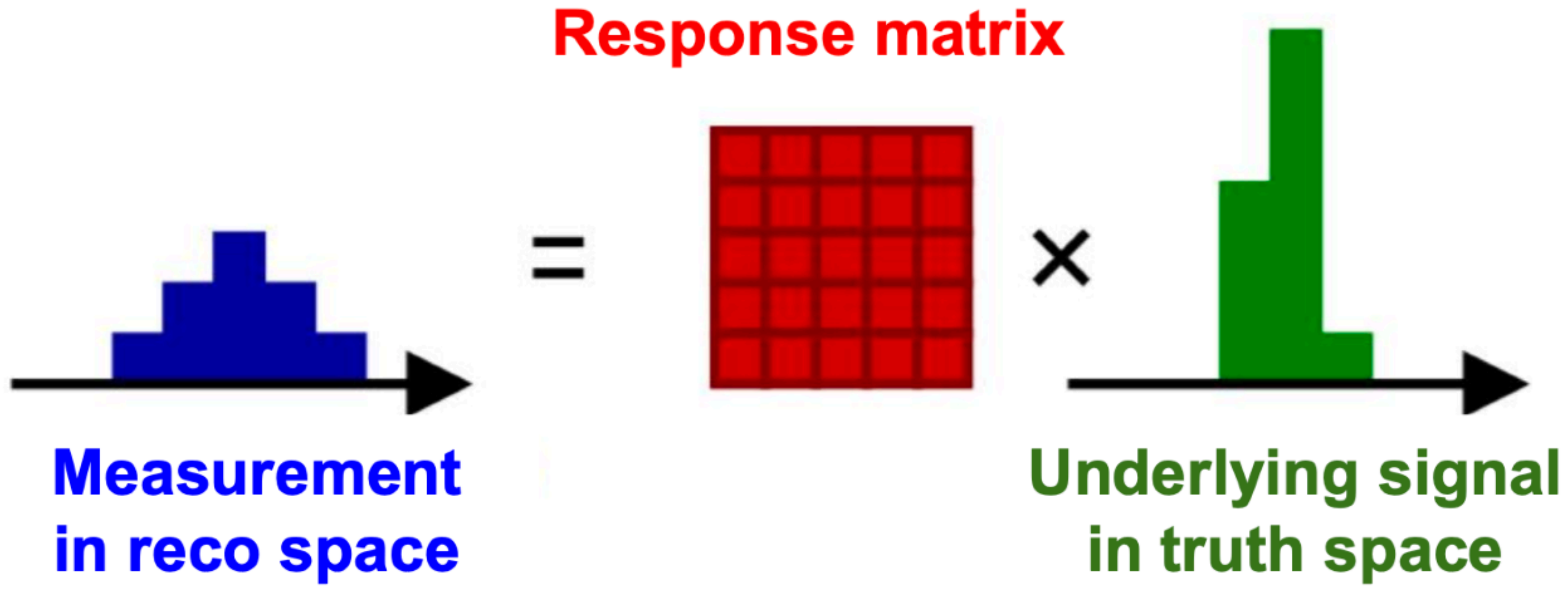
Protons on target | Number of targets | Beam flux | Cross section | Detector response | Selection efficiency | Background

$$M(T_{\mathrm{rec}}) = POT \cdot T \cdot \int \phi\left(E_\nu\right) \cdot \sigma\left(E_\nu, T_{\mathrm{true}}\right) \cdot D\left(T_{\mathrm{true}} \rightarrow T_{\mathrm{rec}}\right) \cdot \varepsilon\left(E_\nu, T_{\mathrm{true}}\right) \cdot dE_\nu + B\left(T_{\mathrm{rec}}\right)$$

Equation relating cross section to measured counts

$$M_i = R_{i,j} \cdot S_j + B_i$$

Writing it as a matrix equation



$$S_j = R_{i,j}^{-1} \cdot \left(M_i - B_i\right)$$

Solving the equation by inverting the total response matrix, usually with some regularization to ensure a smooth result

i: T_{reco} bin index
j: T_{true} bin index

Model Dependence In Cross Section Extractions

$$M(T_{\mathrm{rec}}) = \underbrace{POT}_{\text{Protons on target}} \cdot \underbrace{T}_{\text{Number of targets}} \cdot \int \underbrace{\phi\left(E_\nu\right)}_{\text{Beam flux}} \cdot \underbrace{\boxed{\sigma\left(E_\nu, T_{\mathrm{true}}\right)}}_{\text{Cross section}} \cdot \underbrace{\boxed{D\left(T_{\mathrm{true}} \to T_{\mathrm{rec}}\right)}}_{\text{Detector response}} \cdot \underbrace{\boxed{\varepsilon\left(E_\nu, T_{\mathrm{true}}\right)}}_{\text{Selection efficiency}} \cdot dE_\nu + \underbrace{\boxed{B\left(T_{\mathrm{rec}}\right)}}_{\text{Background}}$$

We directly solve for this quantity, so we don't need to assume any model for it.

But these quantities depend on an assumed cross section model!

- As an example, say that our signal is inclusive ν_μCC events, and we are measuring the kinetic energy of the muon, E_μ

Model Dependence In Cross Section Extractions, Example: ν_μCC E_μ

$$M(E_{\mu,\mathrm{rec}}) = \overset{\text{Protons on target}}{POT} \cdot \overset{\text{Number of targets}}{T} \cdot \int \overset{\text{Beam flux}}{\phi\left(E_\nu\right)} \cdot \overset{\text{Cross section}}{\sigma(E_\nu, E_{\mu,\mathrm{true}})} \cdot \boxed{\overset{\text{Detector response}}{D(E_{\mu,\mathrm{true}} \to E_{\mu,\mathrm{rec}})}} \cdot \overset{\text{Selection efficiency}}{\varepsilon(E_\nu, E_{\mu,\mathrm{true}})} \cdot dE_\nu + \overset{\text{Background}}{B(E_{\mu,\mathrm{rec}})}$$

- The detector response $D(E_{\mu,\mathrm{true}} \to E_{\mu,\mathrm{rec}})$ could depend on the cross section as a function of muon angle

  - Forward muons may be contained in the detector more often, leading to better energy resolution

Model Dependence In Cross Section Extractions, Example: ν_μCC E_μ

$$M(E_{\mu,\mathrm{rec}}) = \underbrace{POT}_{\text{Protons on target}} \cdot \underbrace{T}_{\text{Number of targets}} \cdot \int \underbrace{\phi\left(E_\nu\right)}_{\text{Beam flux}} \cdot \underbrace{\sigma(E_\nu, E_{\mu,\mathrm{true}})}_{\text{Cross section}} \cdot \underbrace{D(E_{\mu,\mathrm{true}} \rightarrow E_{\mu,\mathrm{rec}})}_{\text{Detector response}} \cdot \underbrace{\boxed{\varepsilon(E_\nu, E_{\mu,\mathrm{true}})}}_{\text{Selection efficiency}} \cdot dE_\nu + \underbrace{B(E_{\mu,\mathrm{rec}})}_{\text{Background}}$$

- The selection efficiency $\varepsilon(E_\nu, T_{\mathrm{true}})$ could depend on the cross section as a function of proton energy and multiplicity

- ν_μCC events with observed protons may be easier to distinguish from cosmic muons, leading to higher efficiencies

Model Dependence In Cross Section Extractions, Example: ν_μCC E_μ

$$M(E_{\mu,\mathrm{rec}}) = \underbrace{POT}_{\text{Protons on target}} \cdot \underbrace{T}_{\text{Number of targets}} \cdot \int \underbrace{\phi\left(E_\nu\right)}_{\text{Beam flux}} \cdot \underbrace{\sigma(E_\nu, E_{\mu,\mathrm{true}})}_{\text{Cross section}} \cdot \underbrace{D(E_{\mu,\mathrm{true}} \rightarrow E_{\mu,\mathrm{rec}})}_{\text{Detector response}} \cdot \underbrace{\varepsilon(E_\nu, E_{\mu,\mathrm{true}})}_{\text{Selection efficiency}} \cdot dE_\nu + \underbrace{\boxed{B(E_{\mu,\mathrm{rec}})}}_{\text{Background}}$$

- The background $B(E_{\mu,\mathrm{rec}})$ could depend on the cross section for NC $1\pi^-$ production
  - NC $1\pi^-$ events can mimic the ν_μCC topology, so a larger cross section could lead to a larger background

Model Dependence In Cross Section Extractions

Protons on target | Number of targets | Beam flux | Cross section | Detector response | Selection efficiency | Background

$$M(T_{\mathrm{rec}}) = POT \cdot T \cdot \int \phi\left(E_\nu\right) \cdot \sigma\left(E_\nu, T_{\mathrm{true}}\right) \cdot \boxed{D\left(T_{\mathrm{true}} \rightarrow T_{\mathrm{rec}}\right)} \cdot \boxed{\varepsilon\left(E_\nu, T_{\mathrm{true}}\right)} \cdot dE_\nu + \boxed{B\left(T_{\mathrm{rec}}\right)}$$

- We use systematic uncertainties to hopefully make these model dependent quantities reflect reality

- But if our models aren't accurate enough and our systematic uncertainties aren't big enough, this could lead to the extraction of biased cross section results

E_ν Dependence

- One potential source of model dependence is difficulty in modeling cross sections as a function of E_ν
- E_ν is not directly observable at O(GeV) energies, since there will be energy that we can't construct in the hadronic system, $E_{\mathrm{had}}^{\mathrm{invis}}$ (for example, exiting neutrons)
- When thinking about E_ν dependence, we have to think about flux uncertainties

Integrating Over Beam Flux

Protons on target — **Number of targets** — **Beam flux** — **Cross section** — **Detector response** — **Selection efficiency** — **Background**

$$M(T_{\text{rec}}) = \underbrace{POT}_{\text{Protons on target}} \cdot \underbrace{T}_{\text{Number of targets}} \cdot \boxed{\int \underbrace{\phi\left(E_\nu\right)}_{\text{Beam flux}}} \cdot \underbrace{\sigma\left(E_\nu, T_{\text{true}}\right)}_{\text{Cross section}} \cdot \underbrace{D\left(T_{\text{true}} \rightarrow T_{\text{rec}}\right)}_{\text{Detector response}} \cdot \underbrace{\varepsilon\left(E_\nu, T_{\text{true}}\right)}_{\text{Selection efficiency}} \boxed{\cdot\, dE_\nu} + \underbrace{B\left(T_{\text{rec}}\right)}_{\text{Background}}$$

- In a cross section measurement, we integrate over the neutrino flux
  - (Unless we measuring an E_ν-dependent cross section, more on that later)
- When calculating a flux averaged cross section, you should integrate over the real neutrino flux $\phi_{\text{true}}(E_\nu)$
  - But we only know $\phi_{\text{nominal}}(E_\nu)$, the central value flux prediction from our modeling
- How can we deal with this difference?

Where Do We Put Flux Uncertainties?

$$\sigma = \frac{N}{T\Phi}$$

$N = \dfrac{N_{\text{measured}} - N_{\text{background}}}{\varepsilon}$: number of signal events

Φ: integrated flux, $\Phi = \int \phi(E_\nu)\, dE_\nu$

T: number of target nuclei

Lukas Koch and Stephen Dolan
[Phys. Rev. D 102, 113012 (2020)]

“Real Flux” Extraction

$$\sigma = \frac{N}{T\boxed{\Phi}}$$

Include flux uncertainties

- Place flux uncertainties on Φ
  - Considers the difference between Φ_{nominal} and Φ_{real}
- N is calculated at the real beam flux, even though that’s unknown
  - Minimal flux uncertainties on N
  - $N_{\text{background}}$ and ε are still affected by flux uncertainties

“Nominal Flux” Extraction

$$\sigma = \frac{\boxed{N}}{T\Phi}$$

Include flux uncertainties

- Place flux uncertainties on N to account for the different expected signal between the real flux and the nominal flux
  - Considers the difference between $\phi_{\text{nominal}}(E_\nu)$ and $\phi_{\text{real}}(E_\nu)$
- No uncertainties on Φ

Lukas Koch and Stephen Dolan
[Phys. Rev. D 102, 113012 (2020)]

“Real Flux” Extraction

$$\sigma = \frac{N}{T\boxed{\Phi}}$$

Include flux uncertainties

- Model dependence:
  - We don’t need to trust modeling of how the measured counts depend of flux variations, $\Delta\phi_{\text{real}}(E_\nu) \rightarrow \Delta N_{\text{measured}}$
  - We do need to trust modeling of how the efficiency and background changes for different fluxes, $\Delta\phi_{\text{real}}(E_\nu) \rightarrow \Delta\varepsilon, \Delta N_{\text{background}}$

“Nominal Flux” Extraction

$$\sigma = \frac{\boxed{N}}{T\Phi}$$

Include flux uncertainties

- Model dependence:
  - Need to trust modeling of how every part of the signal measurement depends of flux variations, $\Delta\phi_{\text{real}}(E_\nu) \rightarrow \Delta N_{\text{measured}}, \varepsilon, \Delta N_{\text{background}}$

Lukas Koch and Stephen Dolan
[Phys. Rev. D 102, 113012 (2020)]

“Real Flux” Extraction

$$\sigma = \frac{N}{T\boxed{\Phi}}$$

Include flux uncertainties

- Generator comparisons:
  - It can be difficult to compare a real-flux-averaged cross section with nominal-flux-averaged generator predictions

“Nominal Flux” Extraction

$$\sigma = \frac{\boxed{N}}{T\Phi}$$

Include flux uncertainties

- Generator comparisons:
  - We have a measurement with a well-defined flux, so it is easy to compare with generator predictions

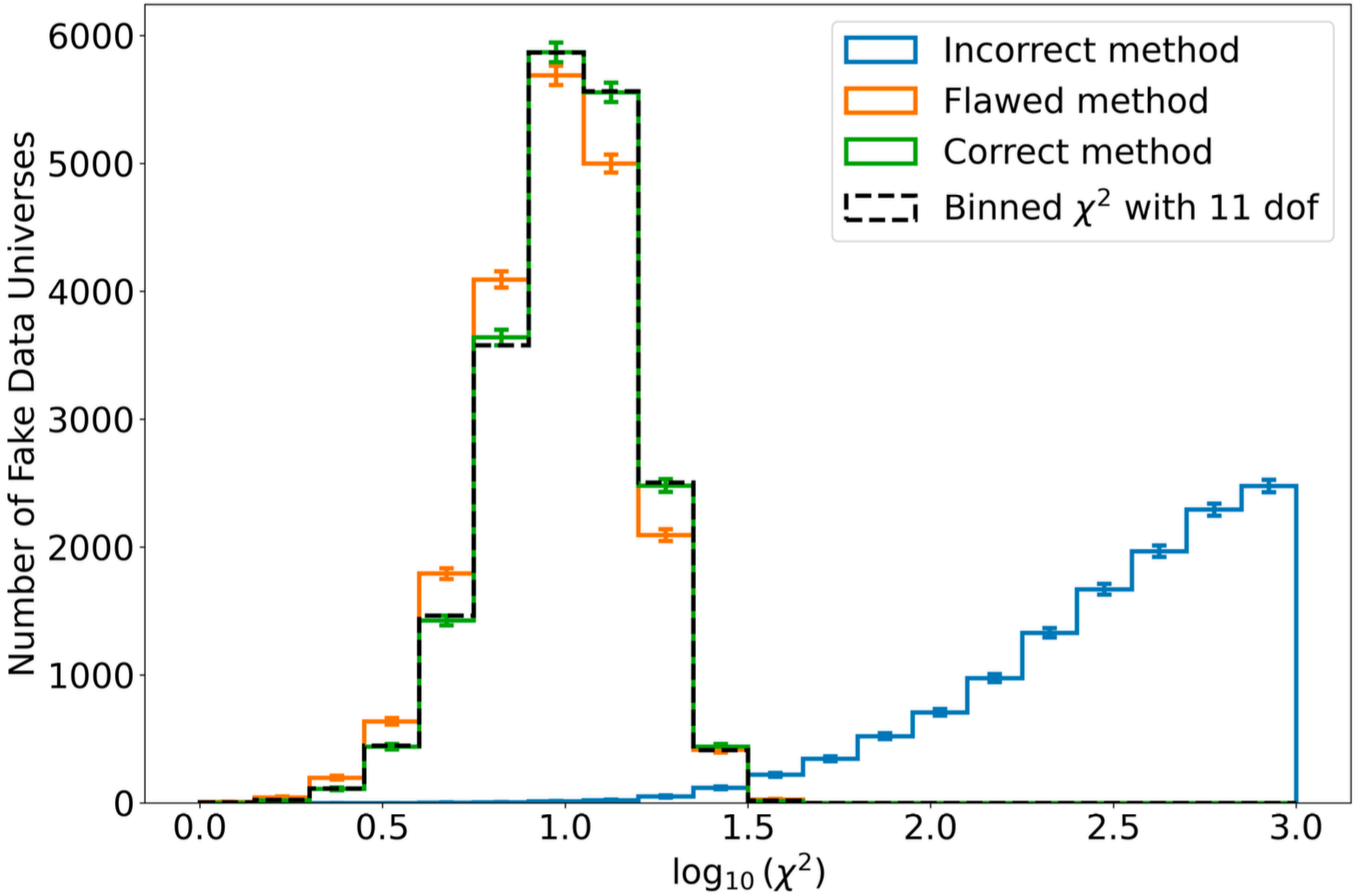
Lukas Koch and Stephen Dolan
[Phys. Rev. D 102, 113012 (2020)]

Toy Example, Illustrating "Real Flux" Challenges

- Toy example: MicroBooNE ν_μCC inclusive measurement of $d\sigma/dE_\mu$, with the same binning as [Phys. Rev. Lett. 128, 151801 (2022)]
- To focus on flux uncertainties, we assume:
  - Perfect energy resolution
  - 100% efficiency
  - Zero background
  - No cross section uncertainties, the MicroBooNE tune GENIE model is exactly correct
- We consider many flux variations, sampled from the BNB flux model at MicroBooNE
  - For each variation, we calculate χ^2 values by comparing our extracted 11 bin differential cross section with the true cross section

Toy Example, Illustrating “Real Flux” Challenges

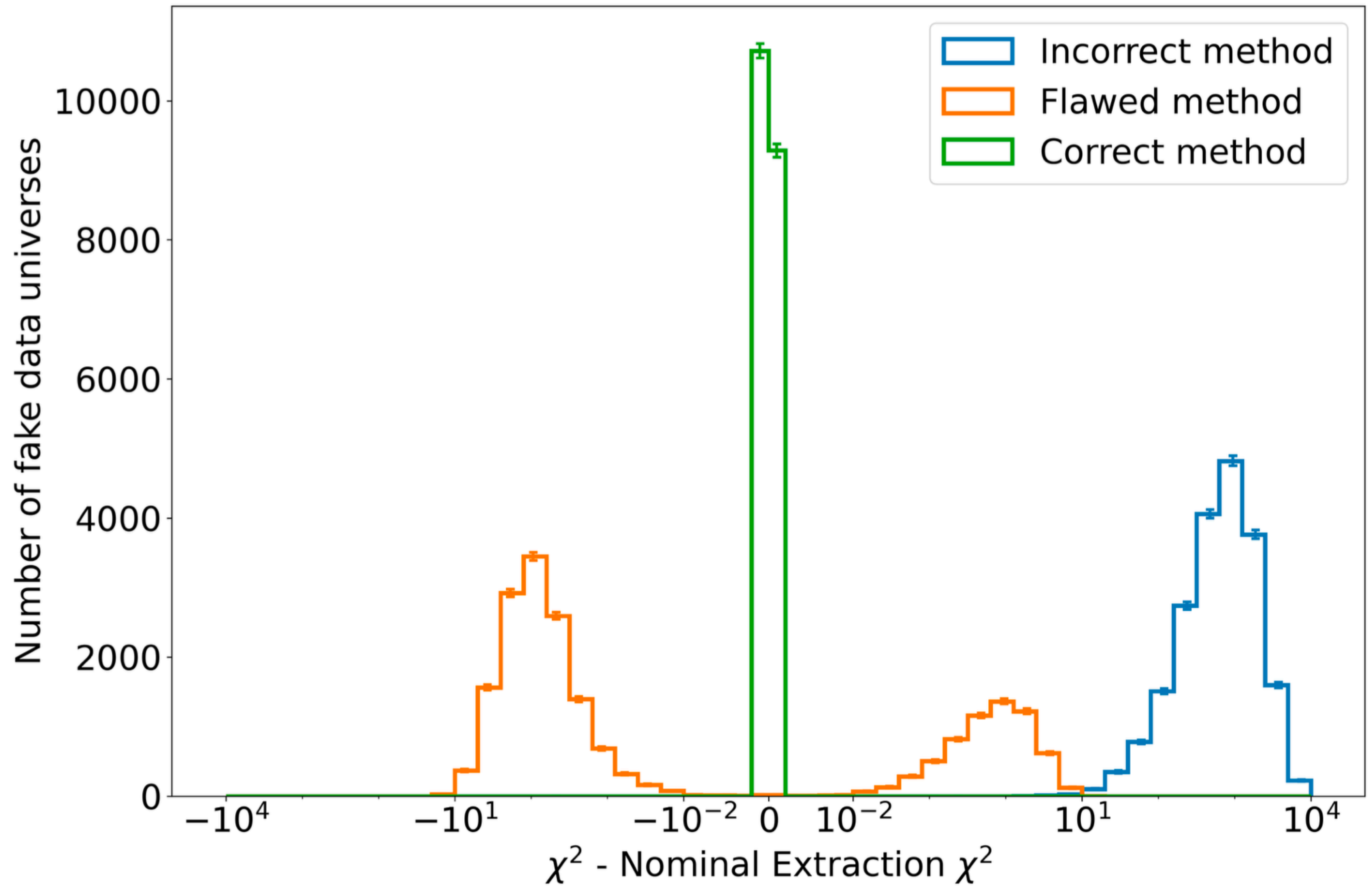
- Three methods:
- “Incorrect”: we don’t include any flux uncertainties on the nominal-flux-averaged true cross section
  - Comparing a “real flux” cross section with a “nominal flux” prediction, leads to very large χ^2 values
- “Flawed”: we include flux uncertainties on the nominal-flux-averaged true cross section, but we don’t consider correlations between this flux uncertainty and the uncertainty on the integrated flux uncertainty Φ used in the extraction
  - Leads to a double-counting of flux uncertainty, shifting the χ^2 values away from a χ^2 probability distribution
- “Correct”: we include flux uncertainties on the nominal-flux-averaged true cross section and their correlations with the extracted integrated flux uncertainty Φ



arXiv:2411.03280

Toy Example, Illustrating “Real Flux” Challenges

- We also compare these χ^2 values with a “nominal flux” extracted χ^2 value, which doesn’t need any flux uncertainties on the true cross section
- The “correct method” and “nominal flux method” are identical
- The “incorrect method” always dramatically overestimates χ^2
- The “flawed method” most often overestimates χ^2, but can underestimate χ^2 as well



arXiv:2411.03280

Toy Example, Illustrating "Real Flux" Challenges

- So, to compare a "real flux" measurement with a generator prediction, we should consider how the extracted cross section is correlated with flux spectrum uncertainty variations

- This could potentially be related to challenges encountered in tuning generator parameters to cross section measurements

  - "Peelle's Pertinent Puzzle": Tuning generators to cross section measurements can result in large χ^2 values and a best fit that's visibly far from the data points

University of Pittsburgh
Generator Studies Workshop

Phys. Rev. D 105, 072001 (2022)
Phys. Rev. C 102, 015502 (2020)
Phys. Rev. D 106, 112001 (2022)

Nominal Flux Challenge

- Because of this difficulty in comparing with generator predictions, MicroBooNE has used the “nominal flux” approach for recent cross sections

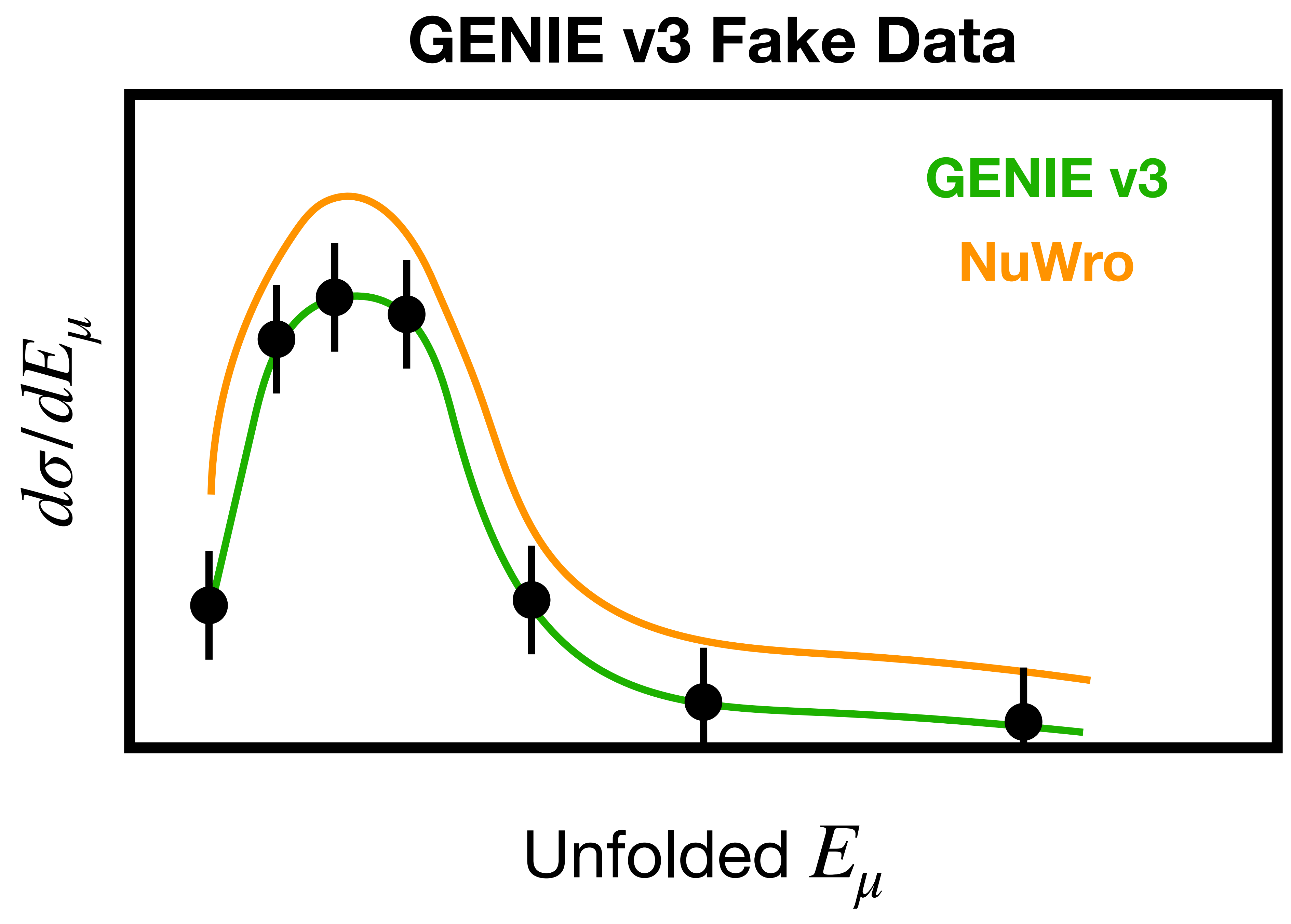
- In order to use the “nominal flux” approach, we have to trust modeling of $\sigma(E_\nu, T_{\mathrm{rec}})$ to some extent

  - This is also true in the “real flux” approach to a lesser extent

- This further increases the importance of modeling $\sigma(E_\nu, T_{\mathrm{rec}})$ sufficiently well before extracting

Fake-Data Closure Testing

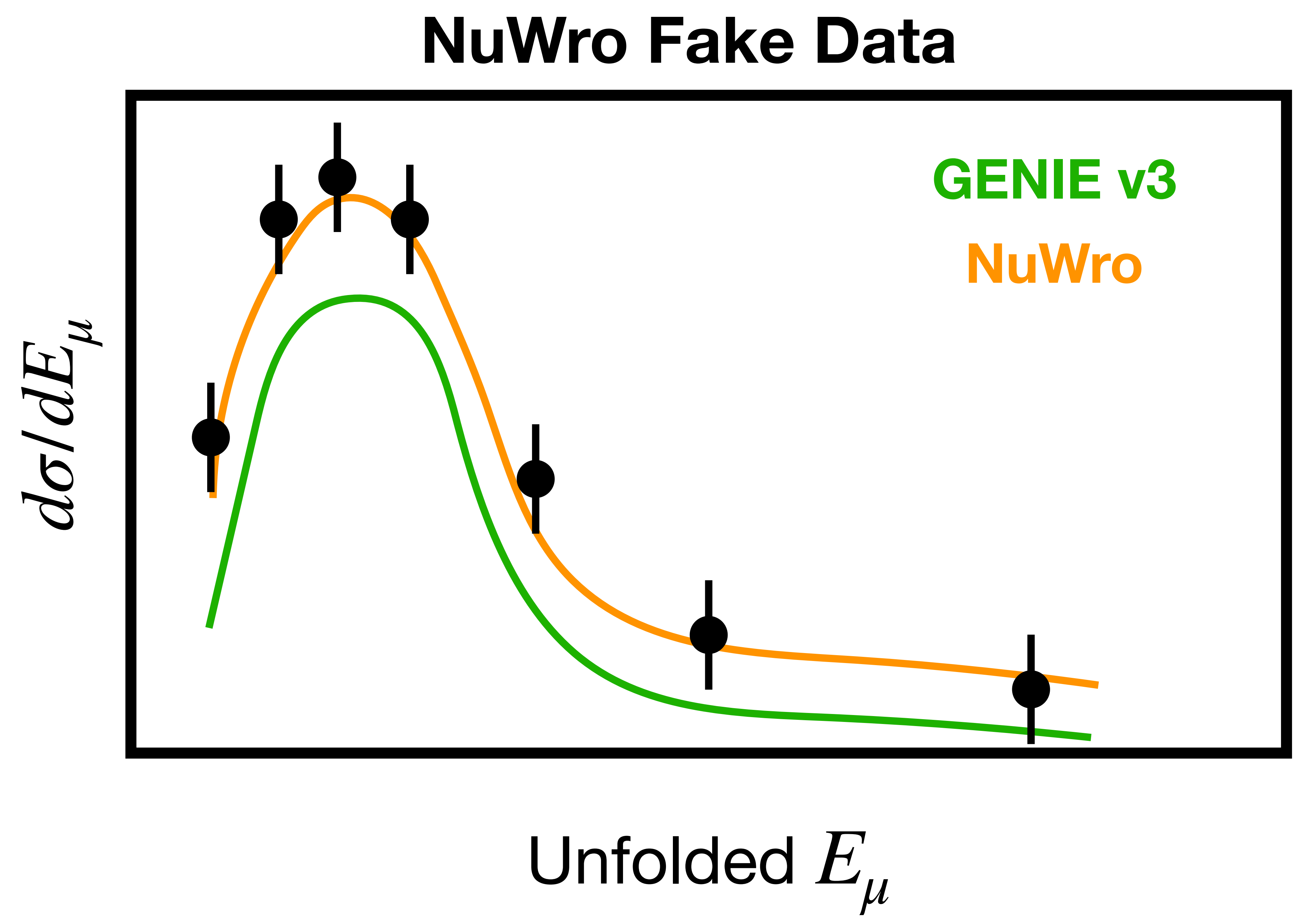
- GENIE v3 for detector response, background, efficiency, and a GENIE v3 fake data set
- Will always get exactly the correct cross section result
  - Not very interesting, usually only done to test for technical bugs



Made up points and curves, just for illustration!

Fake-Data Closure Testing

- GENIE v3 for modeling detector response, background, efficiency, and a NuWro fake data set
- The extraction will differ from the truth if NuWro and GENIE give significantly different predictions for the detector response, background, and efficiency
- If this difference is within uncertainties:
  - We have validated that our extraction doesn't seem too model dependent, safe to extract the cross section
- If this difference is outside uncertainties:
  - We have identified significant model dependence, and we should add uncertainties to reduce this before extracting
  - Example: MicroBooNE $1\mu1p$ TKI cross sections, [Phys. Rev. D 108, 053002 (2023)](). After a NuWro fake data test failed, we added additional cross section uncertainty according to the difference between GENIE and NuWro.

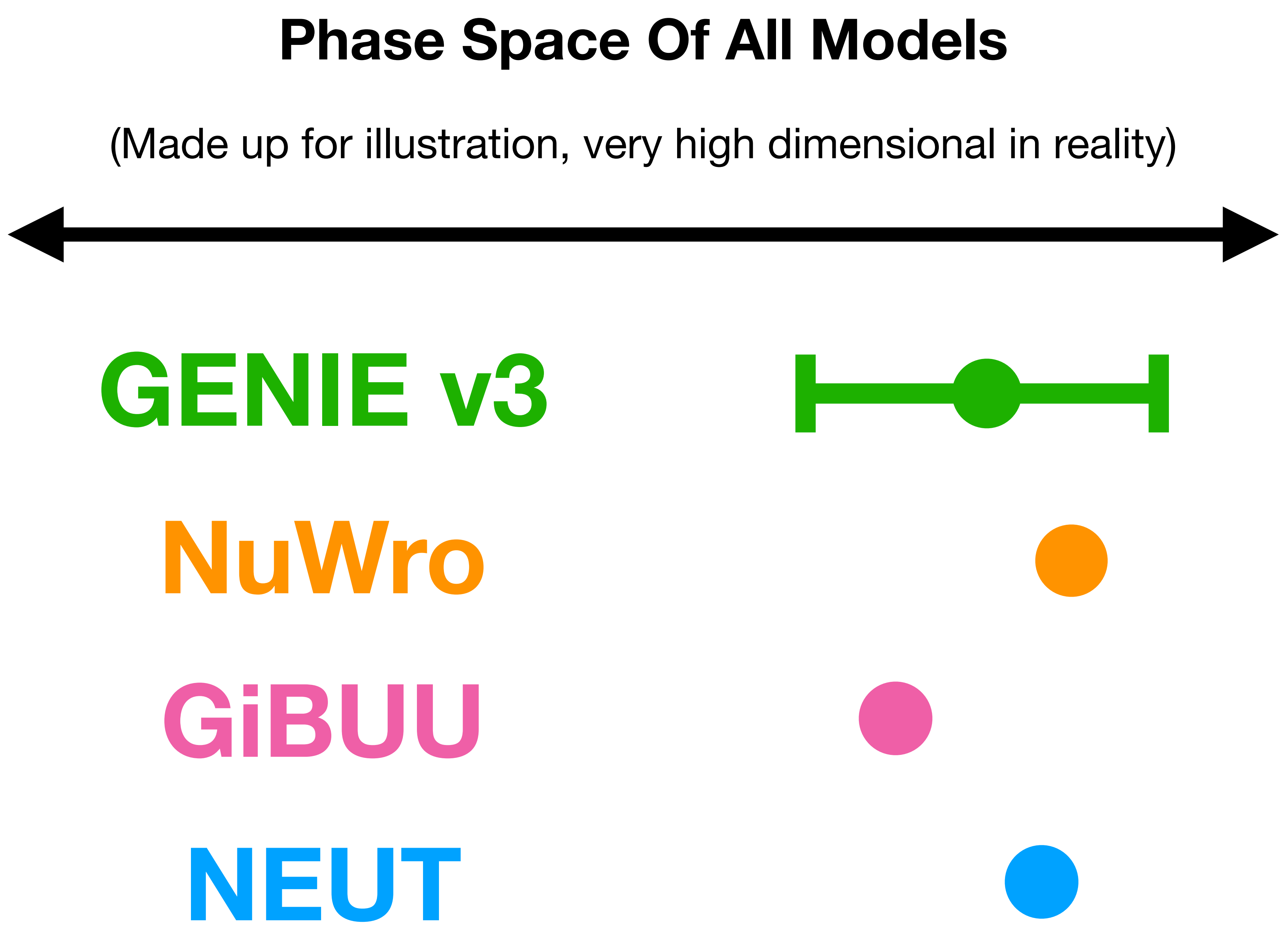


Made up points and curves, just for illustration!

Fake-Data Closure Testing

- To summarize:

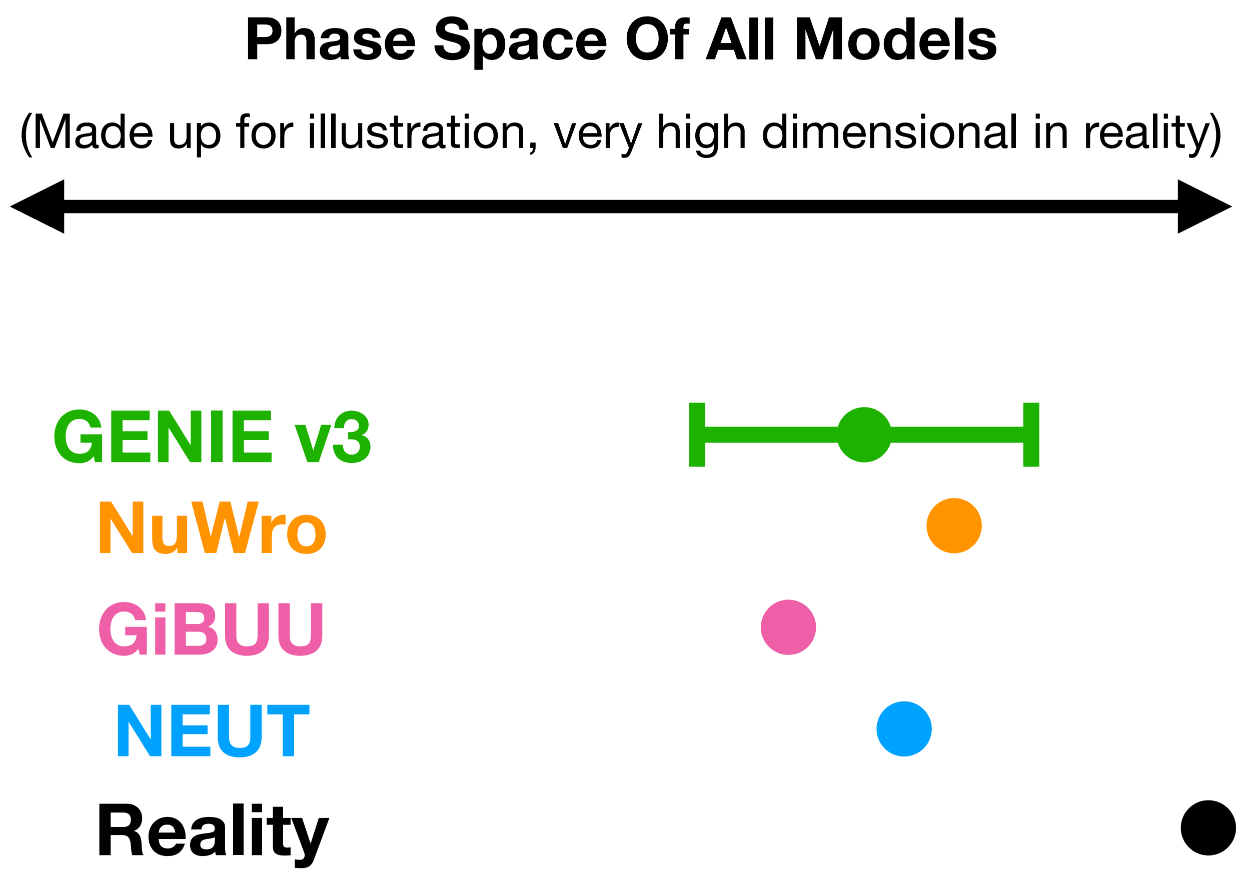
- Fake-Data closure tests try to ensure that in the relevant phase space affecting the detector response, efficiency, and background, the uncertainties of the model cover other cross section models



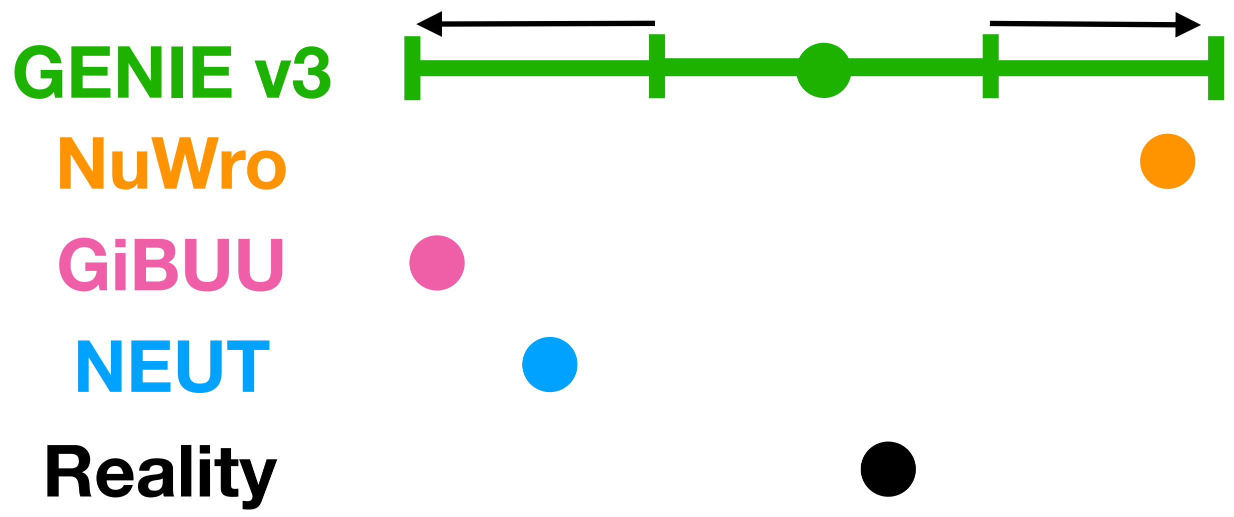
- This one error bar isn't literal, in reality we have to do statistical tests to get some idea of whether two models are within uncertainties in the relevant phase space

Fake-Data Closure Testing

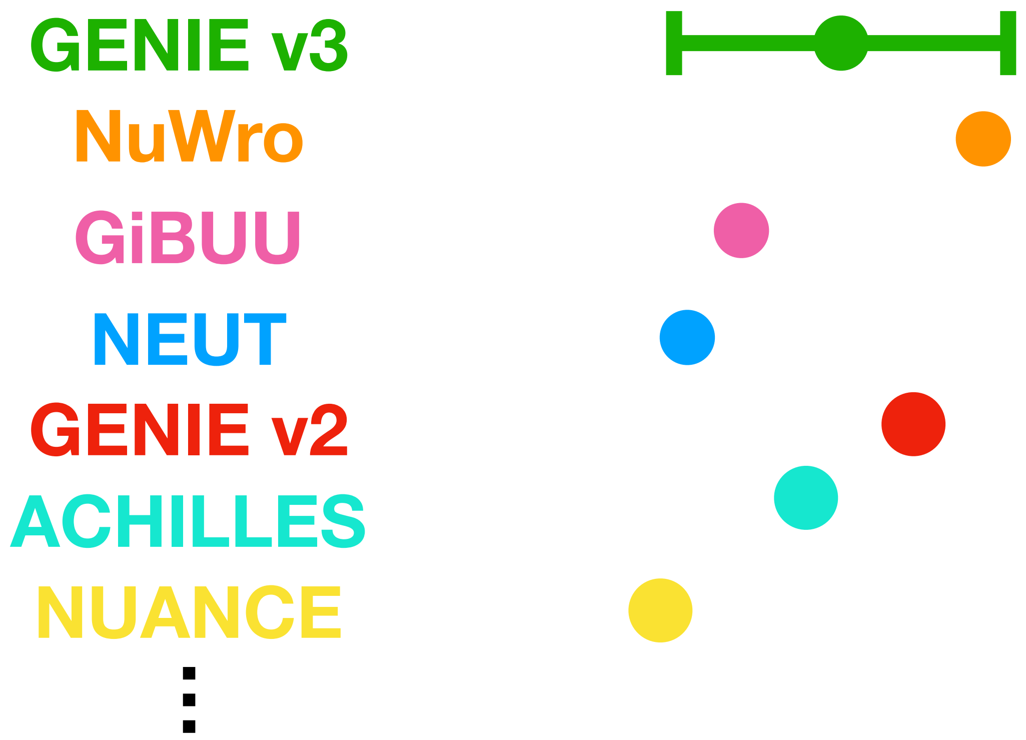
- This procedure has some limitations
  - 1: What if the spread of models included in fake data tests do not describe the real cross section?



  - 2: What if the spread of models included in fake data tests is very large, and makes you expand to very large uncertainties, even when your original model is good?

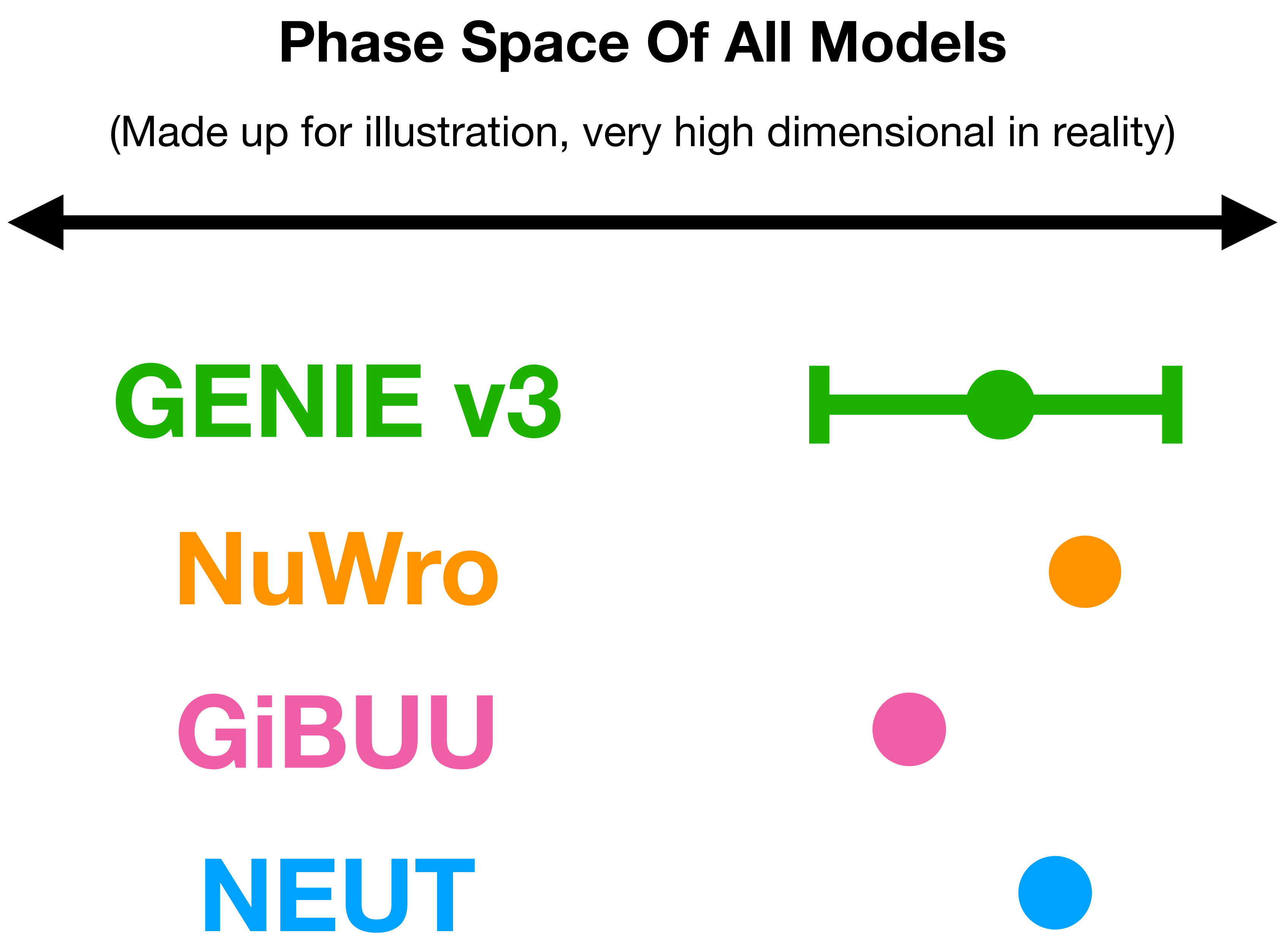


  - 3: How do you know when to stop? There are many generators, and many configurations and tunes, is testing just one or a few alternate generators enough?



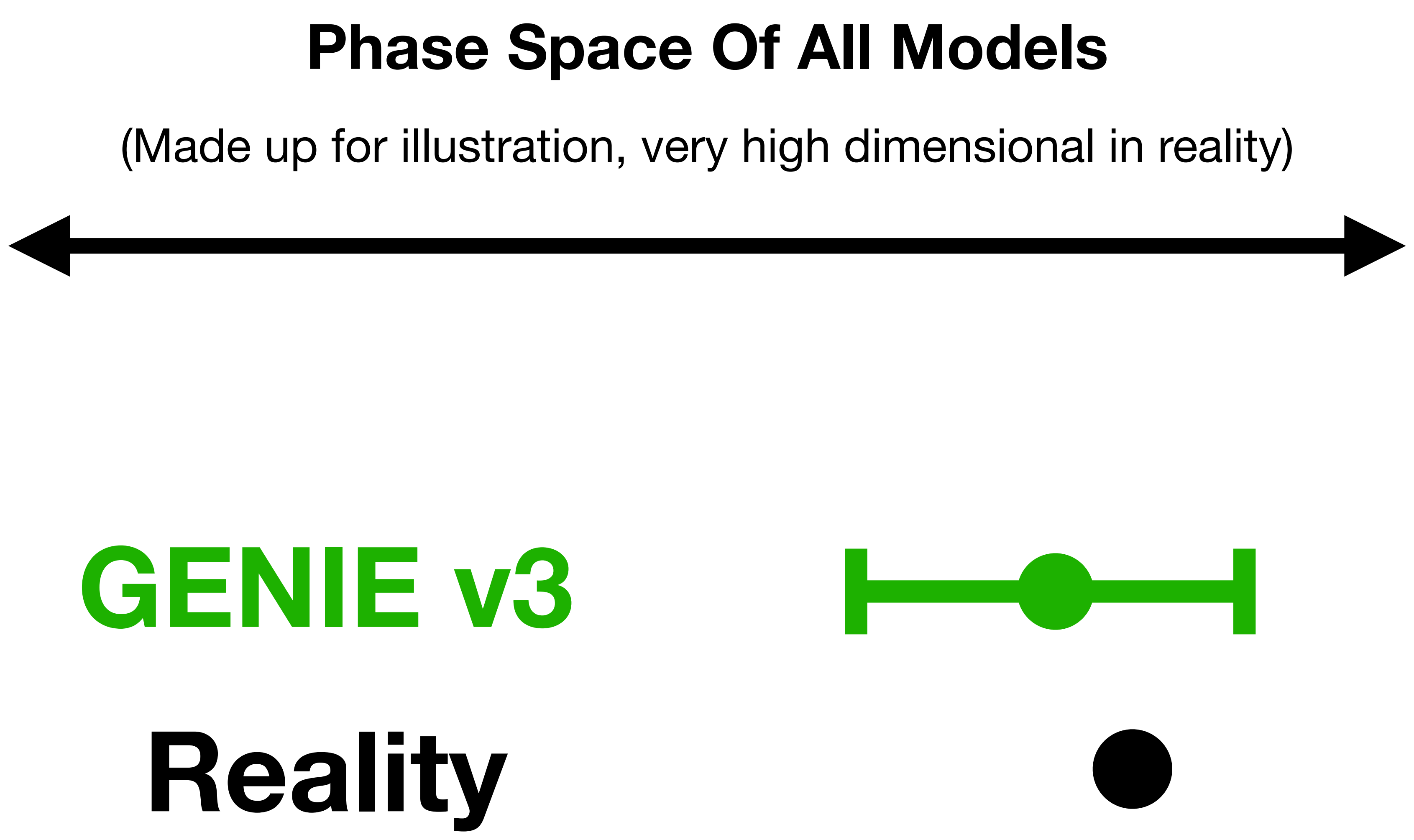
Fake-Data Closure Testing

- In traditional fake-data closure tests, we try to ensure that the cross section model used for the extraction is consistent with other cross section models in the relevant phase space



Data-driven Model Validation

- In data-driven model validation, we try to ensure that the cross section model used for the extraction is consistent with real data in the relevant phase space



χ^2/ndf Tests

- To test the consistency between data and our cross section model, we use goodness of fit tests

- First obvious thing to do: a χ^2/ndf test

    - Calculate χ^2 in the reconstructed space using the measured counts M, predicted counts P, and the covariance matrix of uncertainties V

$$\chi^2 = (M - P)^T \cdot V^{-1} \cdot (M - P)$$

- Ideally, do this not only in the variable being extracted, but also in several other variables, which could be more sensitive to physics that would affect the efficiency, detector response, and background prediction

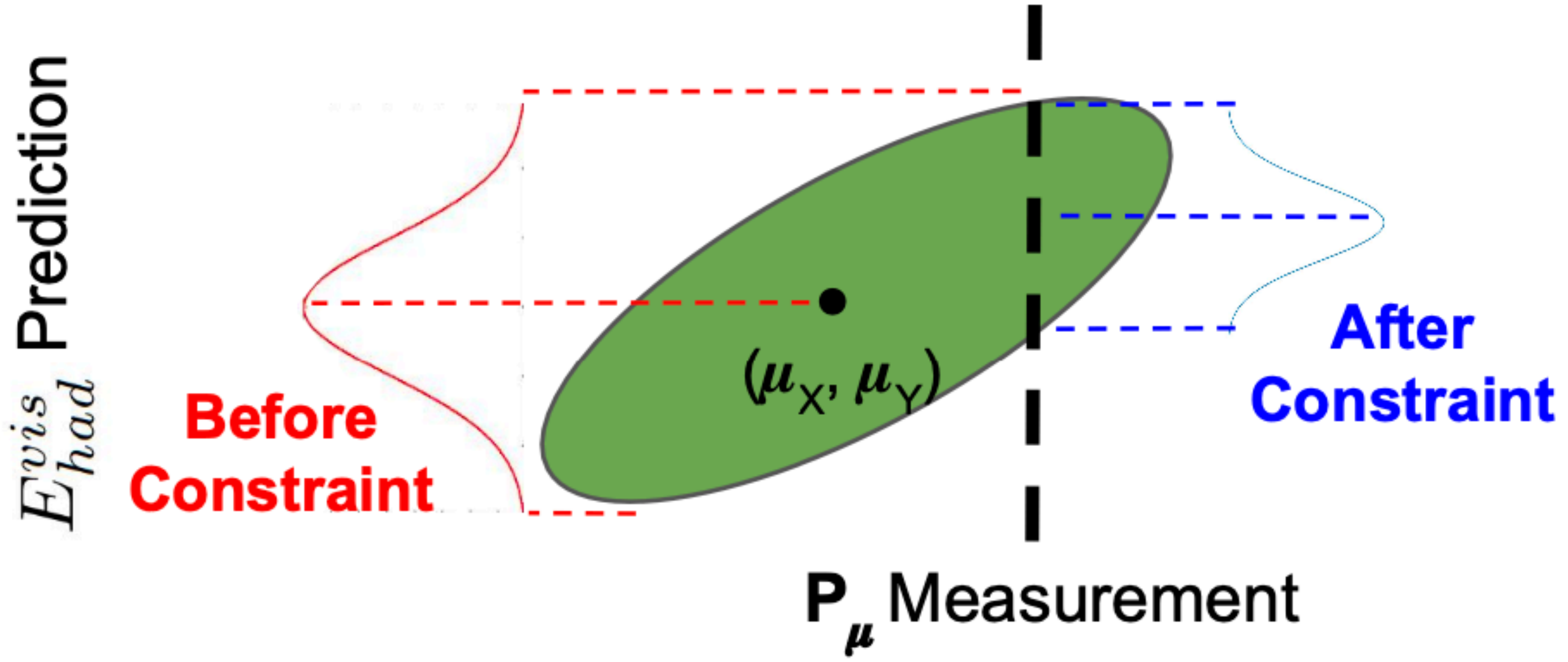
Conditional Constraint χ^2/ndf Tests

$$\Sigma = \begin{pmatrix} \Sigma^{XX} & \Sigma^{XY} \\ \Sigma^{YX} & \Sigma^{YY} \end{pmatrix}, \quad n: \text{measurement}, \quad \mu: \text{prediction}$$

$$\mu^{X,\text{const.}} = \mu^X + \Sigma^{XY} \cdot (\Sigma^{YY})^{-1} \cdot (n^Y - \mu^Y)$$

$$\Sigma^{XX,\text{const.}} = \Sigma^{XX} - \Sigma^{XY} \cdot (\Sigma^{YY})^{-1} \cdot \Sigma^{YX}$$

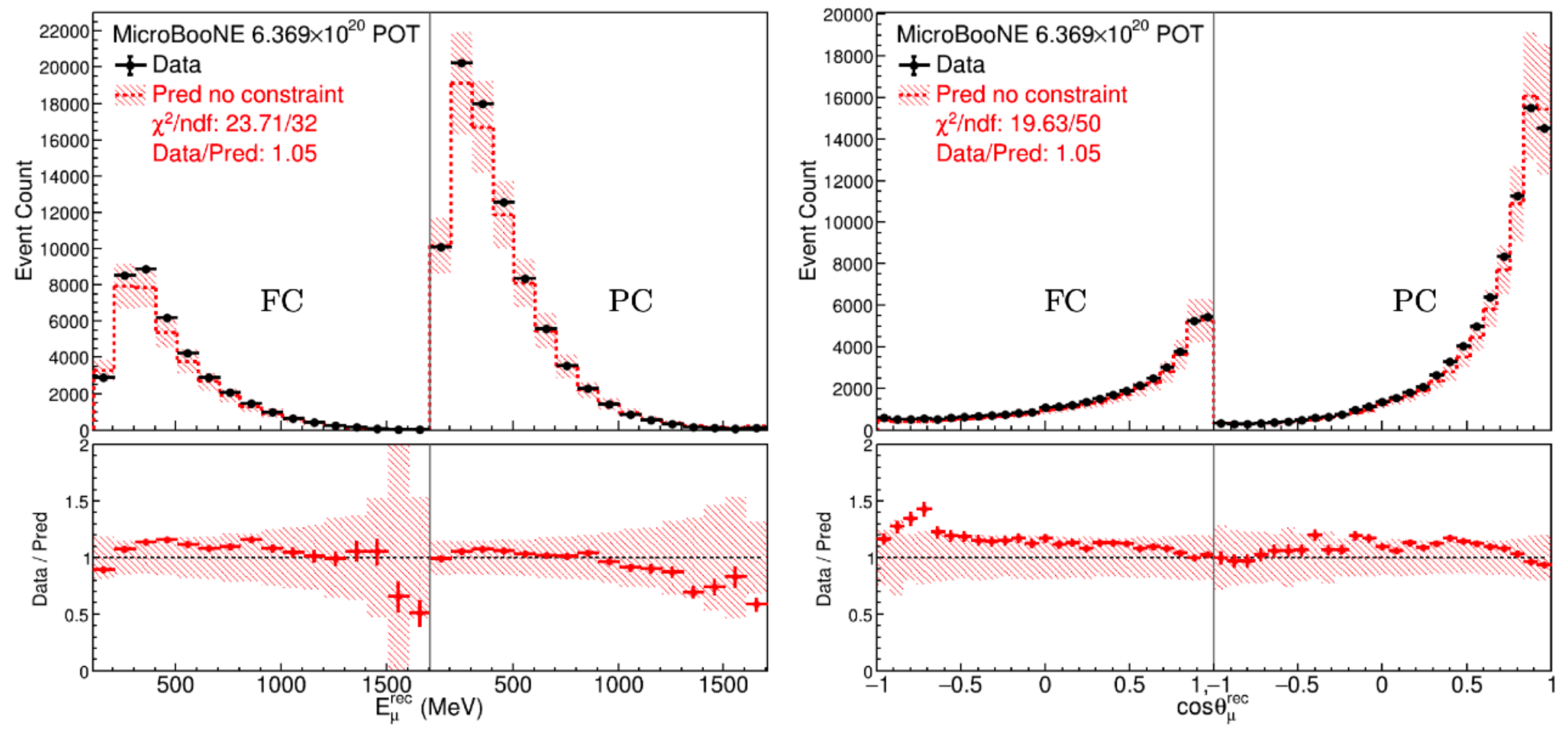
- Our cross section model doesn't just predict individual channels, it also predicts correlations between channels
- We can test this correlation by using a conditional constraint
- This uses Bayes' theorem to update the prediction and uncertainty on one channel after constraining with another channel



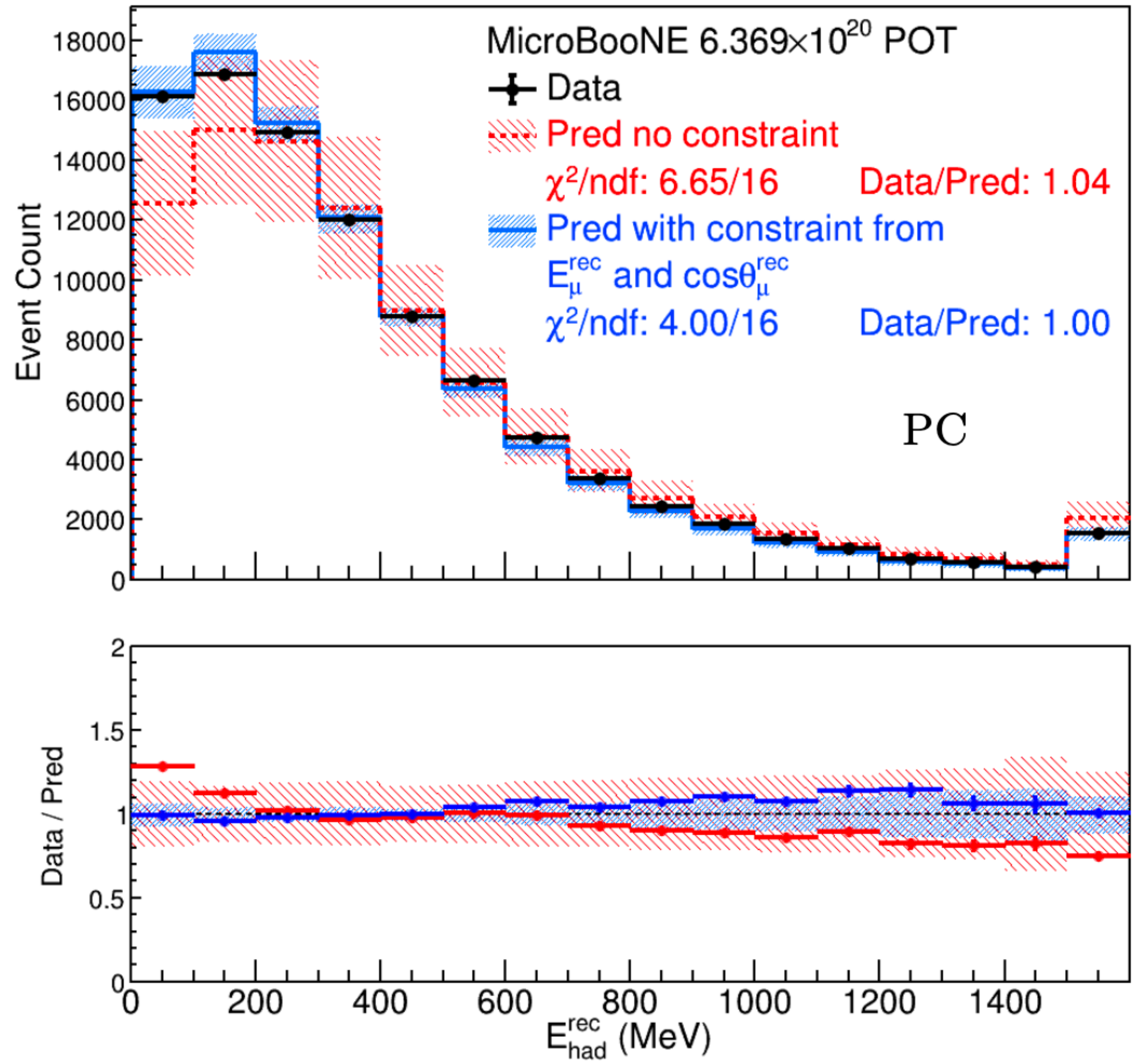
Conditional Constraint χ^2/ndf Tests

- With MicroBooNE ν_μCC data, after using muon kinematics to constrain the reconstructed hadronic energy, we see a substantial update, with good agreement and good χ^2/ndf
- This type of test can be much more sensitive to cross section model deficiencies

Fully Contained (FC) and Partially Contained (PC)
Constraining $E_\mu^{\rm rec}$ and $\cos\theta_\mu^{\rm rec}$ observations:



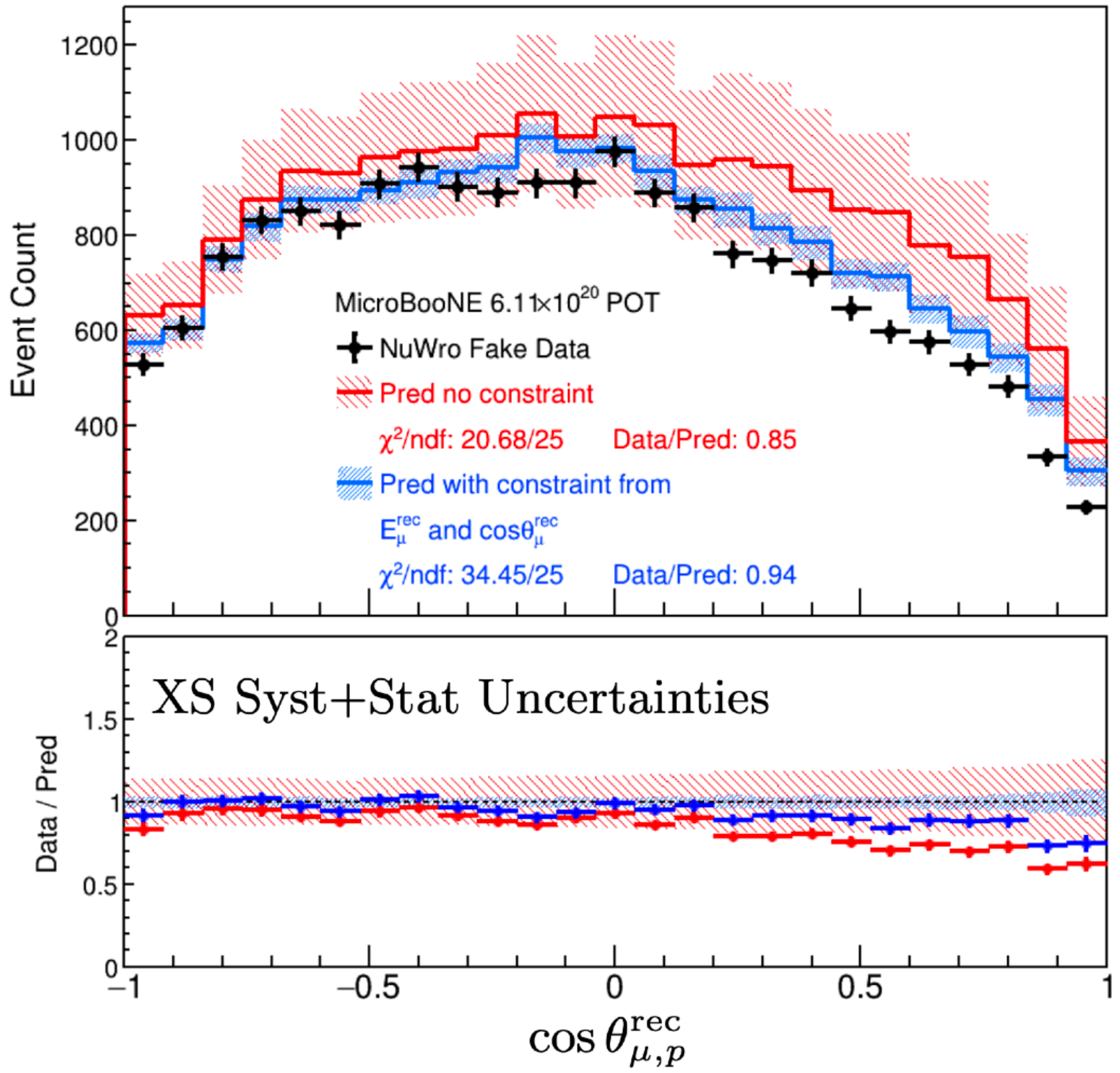
Conditional constraint updating red to blue:

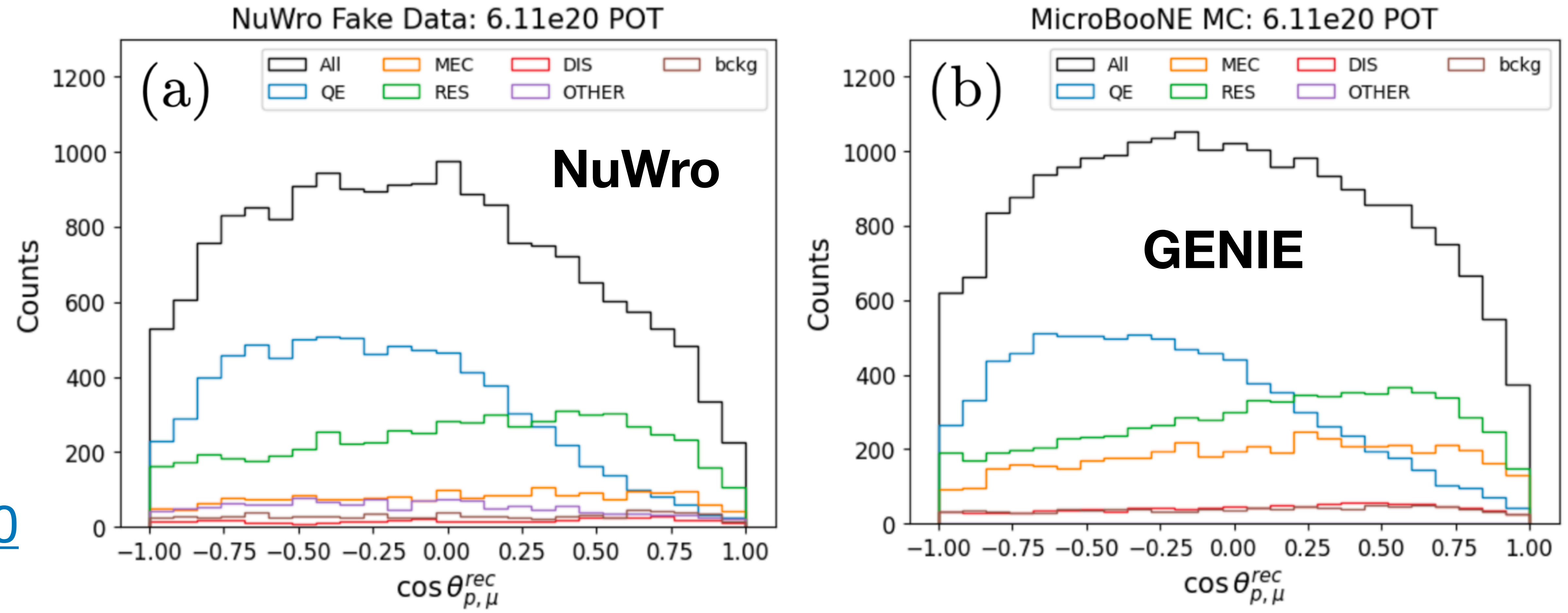


arXiv:2411.03280

The Conditional Constraint Isn't A Black Box!
NuWro Fake Data: MEC Rate and $\theta_{\mu,p}$

- The effect of a conditional constraint can be easily interpreted in some cases
- As an example, with NuWro fake data, we use muon kinematics to constrain the muon-proton opening angle distribution
- NuWro has a much smaller MEC prediction relative to GENIE
- After the update from the constraint, this difference is reflected in the muon-proton opening angle distribution
    - Good agreement for the QE-dominant (large angle) region
    - Bad agreement for the MEC/RES-dominant (small angle) region
- This tells us that the NuWro MEC prediction seems to be outside of GENIE uncertainties (more on NuWro/GENIE comparisons later)

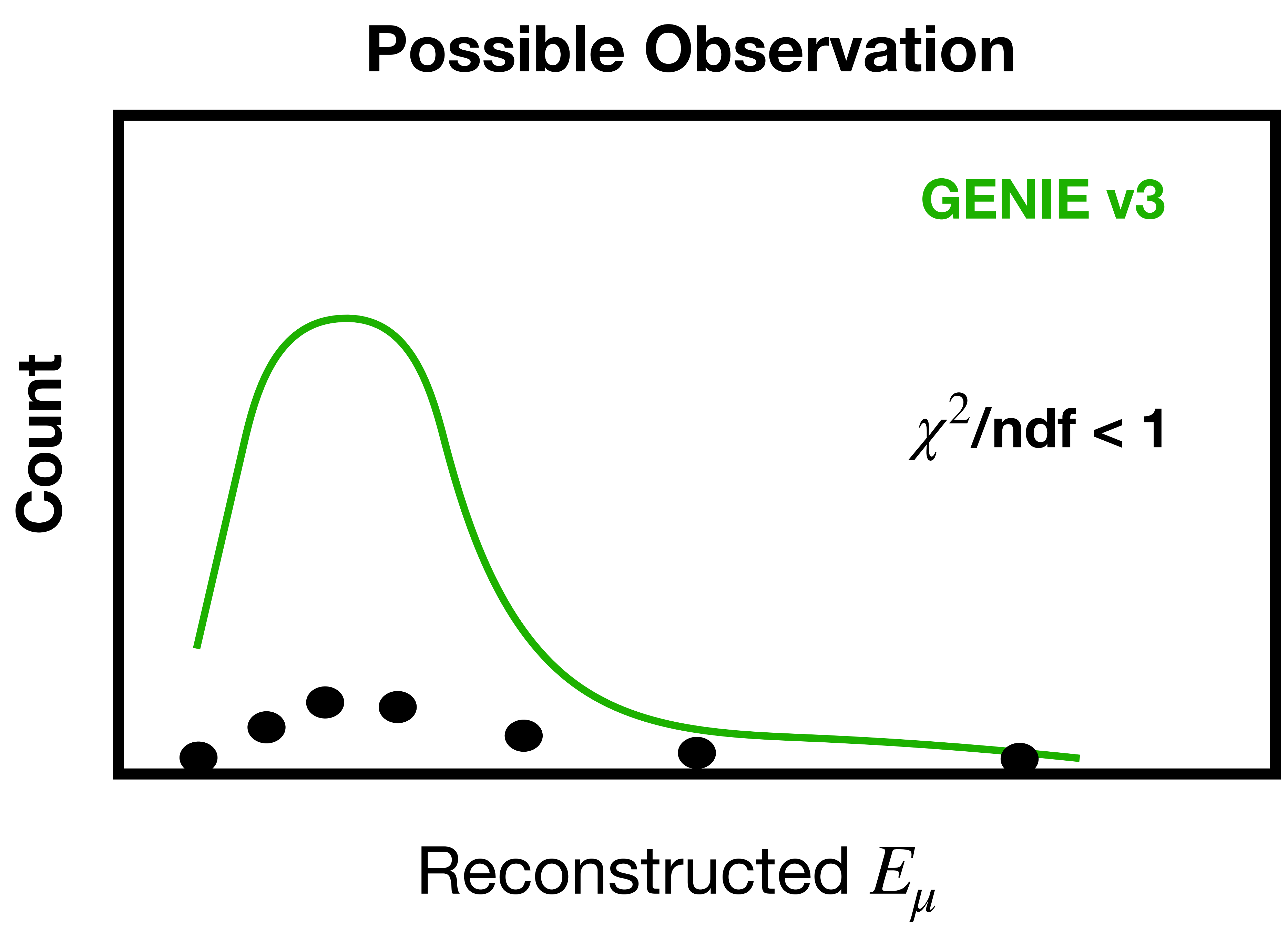




arXiv:2411.03280

χ^2 Test Failures

- Even after a conditional constraint, a χ^2/ndf test isn't necessarily good enough
- It can be tricked if some conservative uncertainties mask other inadequate uncertainties
- Example: You can get a good χ^2/ndf if you have perfect shape agreement, but normalization far outside of uncertainties
    - For this specific case, we could collapse to one bin and then calculate χ^2/ndf for a more sensitive test
    - But other similar issues are also possible, is there a more general solution?



χ^2 Decomposition Tests

- Using an eigenvalue decomposition, we can transform the covariance matrix into a space with no bin-to-bin correlations
- We get the same χ^2 value as normal, but now each bin is uncorrelated, and corresponds to a specific type of prediction change across all reco bins
- We can calculate a p-value for each bin according to a 1-bin χ^2 distribution, and then account for the look elsewhere effect, translating that set of p-values to a global p-value using the largest discrepancy
- This can be performed any time we have a covariance matrix
  - Before or after constraint, in reco space or unfolded truth space

$$\chi^2 = (M-P)^T \cdot V^{-1} \cdot (M-P)$$

$$V = \tilde{Q} \cdot \Lambda \cdot \tilde{Q}^T$$
$$Q = \tilde{Q}^{-1}$$
$$\Delta = (M-P)$$

Λ: diagonal matrix of V eigenvalues
$\tilde{Q}$: corresponding matrix of V eigenvectors as columns

$$\chi^2 = (Q \cdot \Delta)^T \cdot (Q \cdot V \cdot Q^T)^{-1} \cdot (Q \cdot \Delta)$$

$$\Delta' = Q \cdot \Delta$$
$$\epsilon_i = \Delta'_i / \sqrt{\Lambda_{ii}}$$

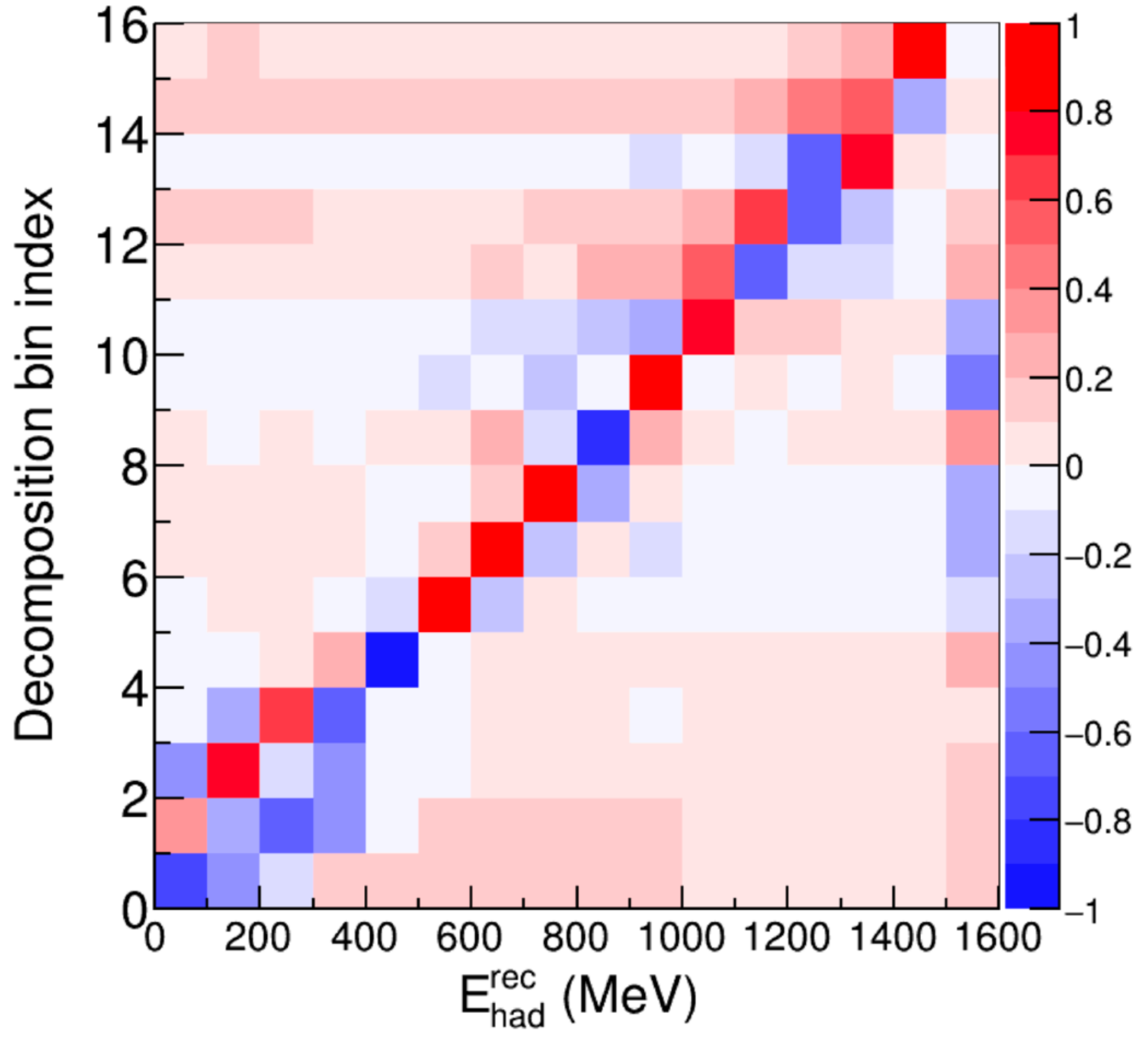
$$\chi^2 = \Delta'^T \cdot \Lambda^{-1} \cdot \Delta'$$
$$\chi^2 = \sum_i \epsilon_i^2$$

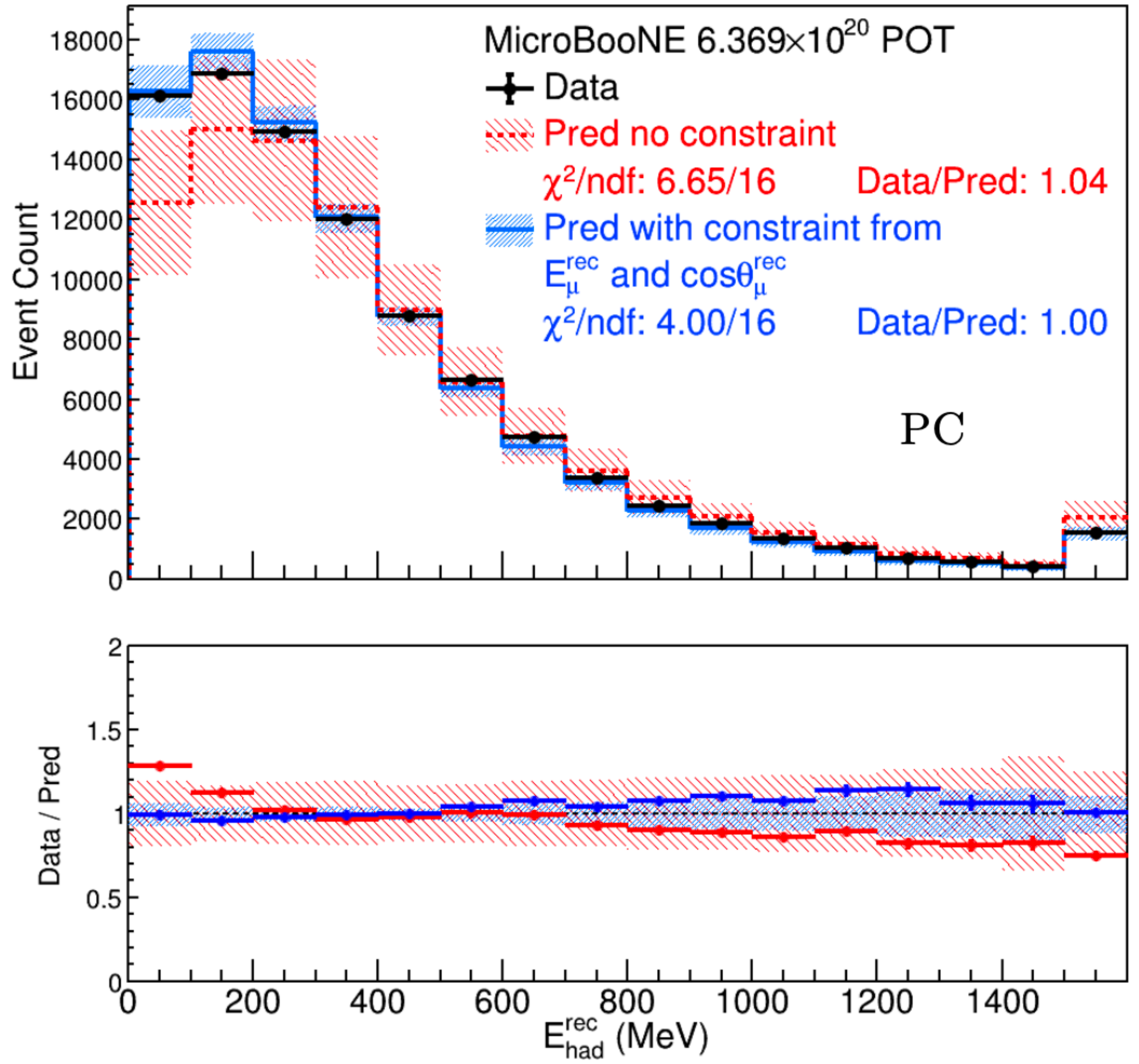
$$p_{\text{global}} = 1 - (1 - p_{\text{local}})^n$$

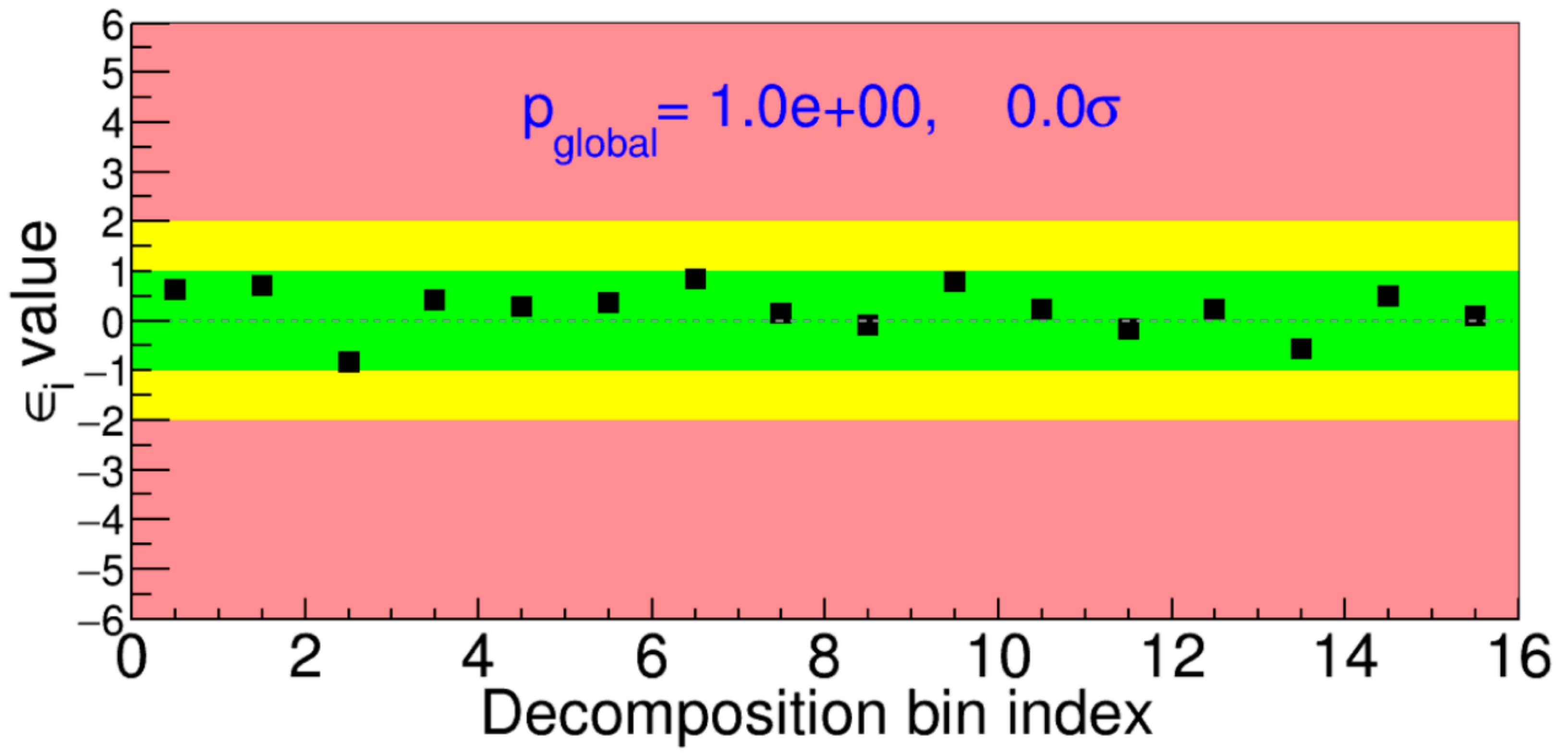
arXiv:2411.03280

χ^2 Decomposition Example

- Looking at the χ^2 decomposition for the previous GoF test, we see good agreement in each bin
- We can look at the matrix of eigenvectors to see how decomposition bins relate to the original bins
- For example, the first decomposition bin corresponds to a decrease in the first three energy bins, and an increase in the remaining bins
    - Resembles a bin migration effect
    - Our data does show some preference for this type of shape change, but the corresponding ϵ_0 value tells us that this movement is within uncertainties







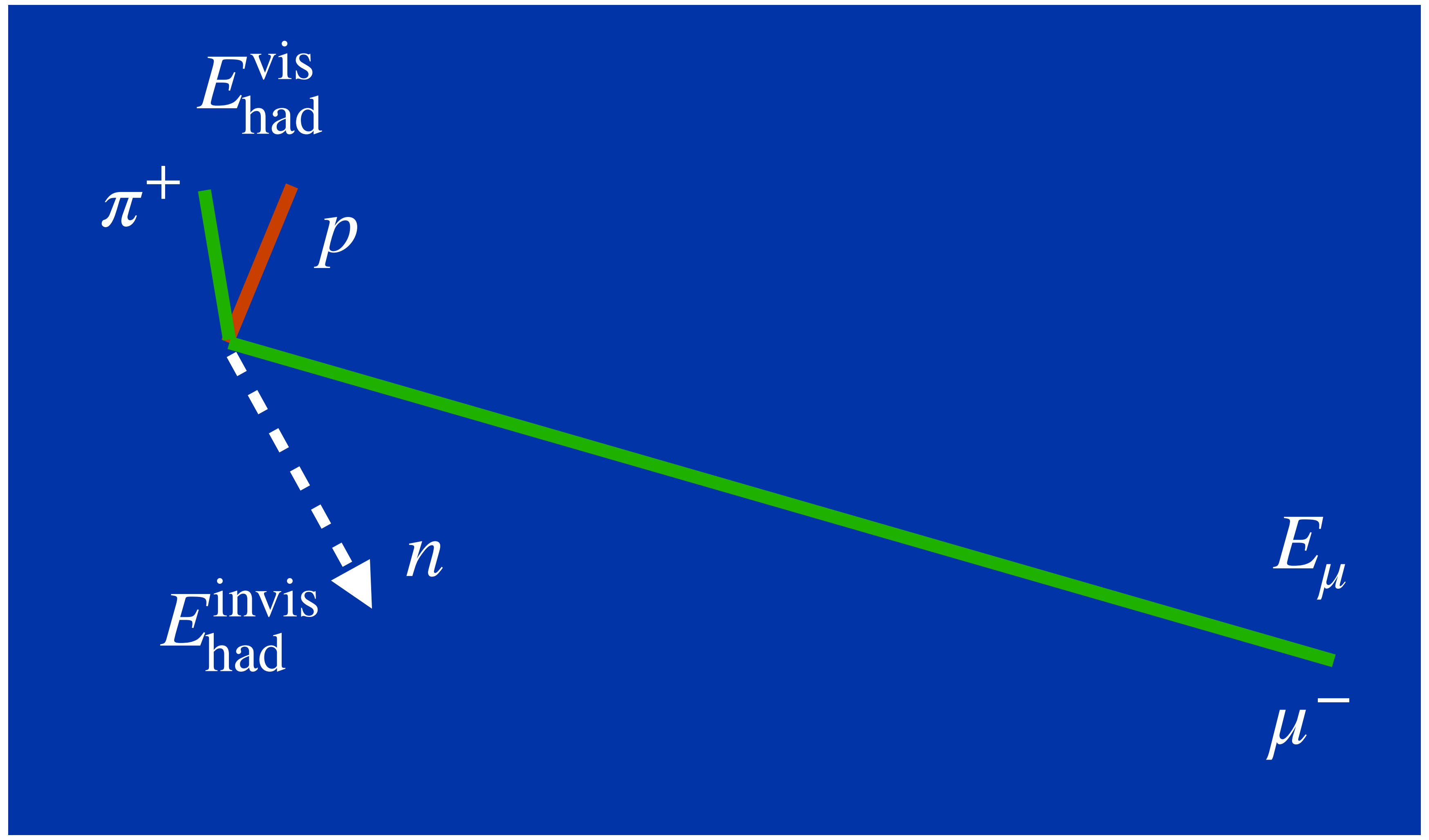
arXiv:2411.03280

Invisible Neutrino Energy Modeling Validation

- In general, the best model validations to perform depends on the type of measurement being made
- One particularly important part of the model to validate is the $E_\nu^{\text{true}} \rightarrow E_\nu^{\text{rec}}$ mapping
  - Important for future oscillation analyses
  - Important for $\sigma(E_\nu)$ cross sections
  - Important for "nominal flux" extractions
- For calorimetric energy reconstructions, E_ν^{true} and E_ν^{rec} differ by $E_{\text{had}}^{\text{invis}}$
  - No direct measurement possible

Consider ν_μ CC interactions:

$$E_\nu^{\text{true}} = E_\mu + E_{\text{had}}^{\text{vis}} + E_{\text{had}}^{\text{invis}}$$



Invisible Neutrino Energy Modeling Validation

- Energy conservation: if modeling of $E_\nu^{\text{true}}, E_\mu, E_{\text{had}}^{\text{vis}}$ is correct, then our modeling of $E_{\text{had}}^{\text{invis}}$ must be correct
  - We can't test this event-by-event, but we can test this for a distribution of many events
- Specifically, we want to validate our modeling of the correlation between hadronic and leptonic energy
- A conditional constraint test shows that $E_{\text{had}}^{\text{vis}}$ data matches the prediction within uncertainties when using E_ν^{true} (from our flux model) and E_μ (from our data measurement)
- So, these three distributions tell a consistent story, and therefore we have increased confidence in our modeling of $E_{\text{had}}^{\text{invis}}$

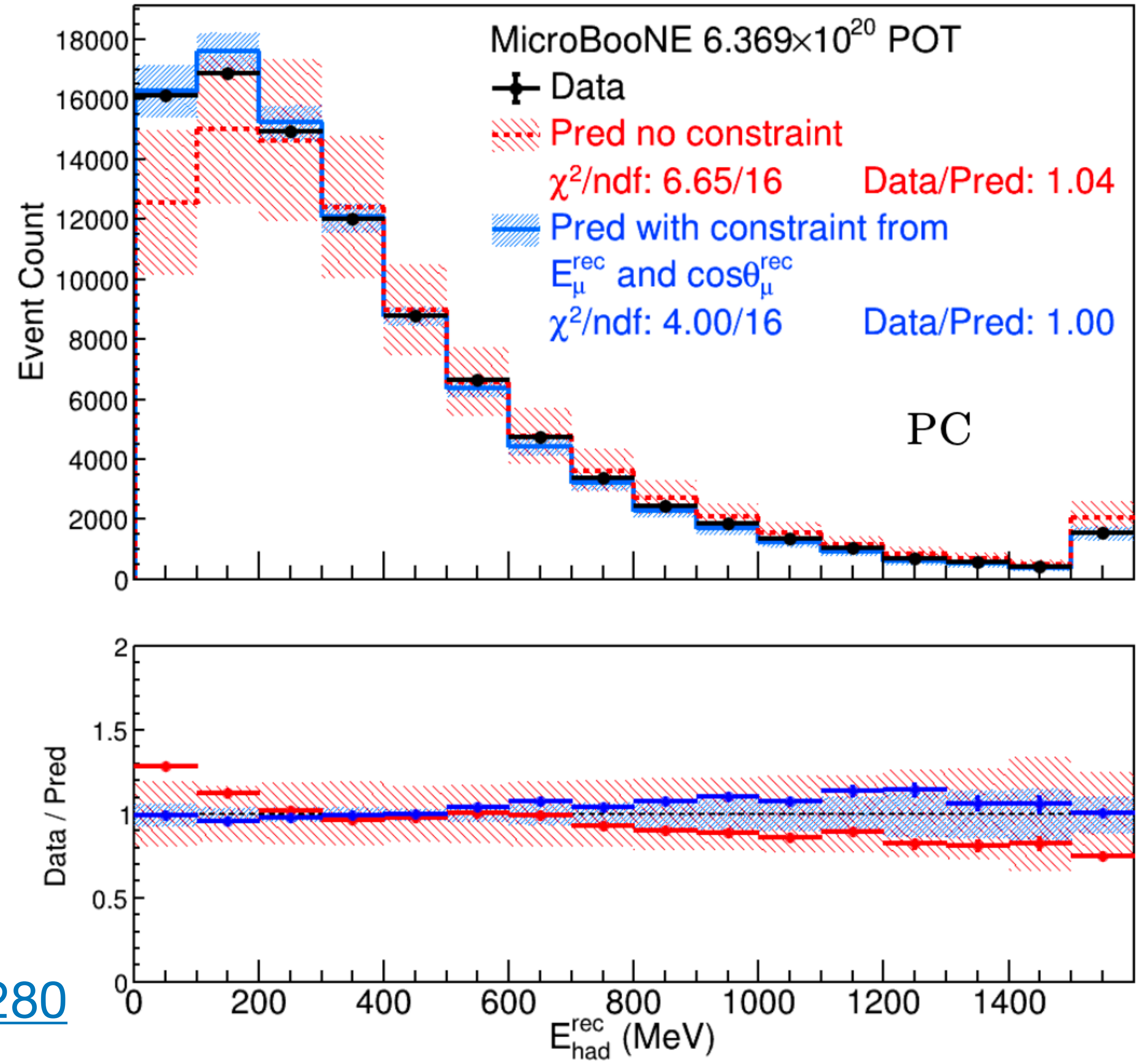
Modeled from beam simulation

Measured from muon range

Measured from hadronic ionization

Initially unknown (Neutrons, low energy photons, etc.)

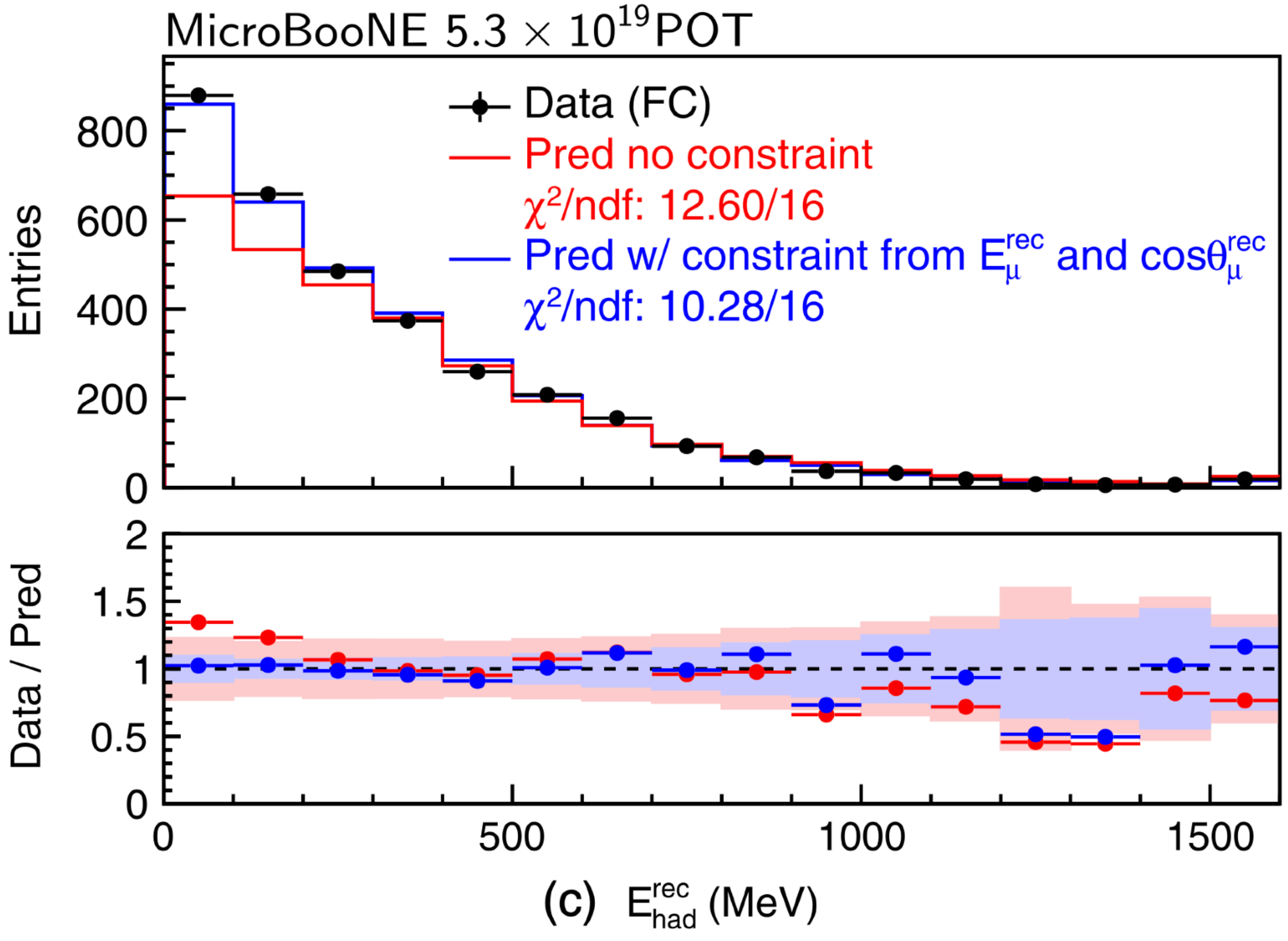
$$E_\nu^{\text{true}} = E_\mu + E_{\text{had}}^{\text{vis}} + E_{\text{had}}^{\text{invis}}$$

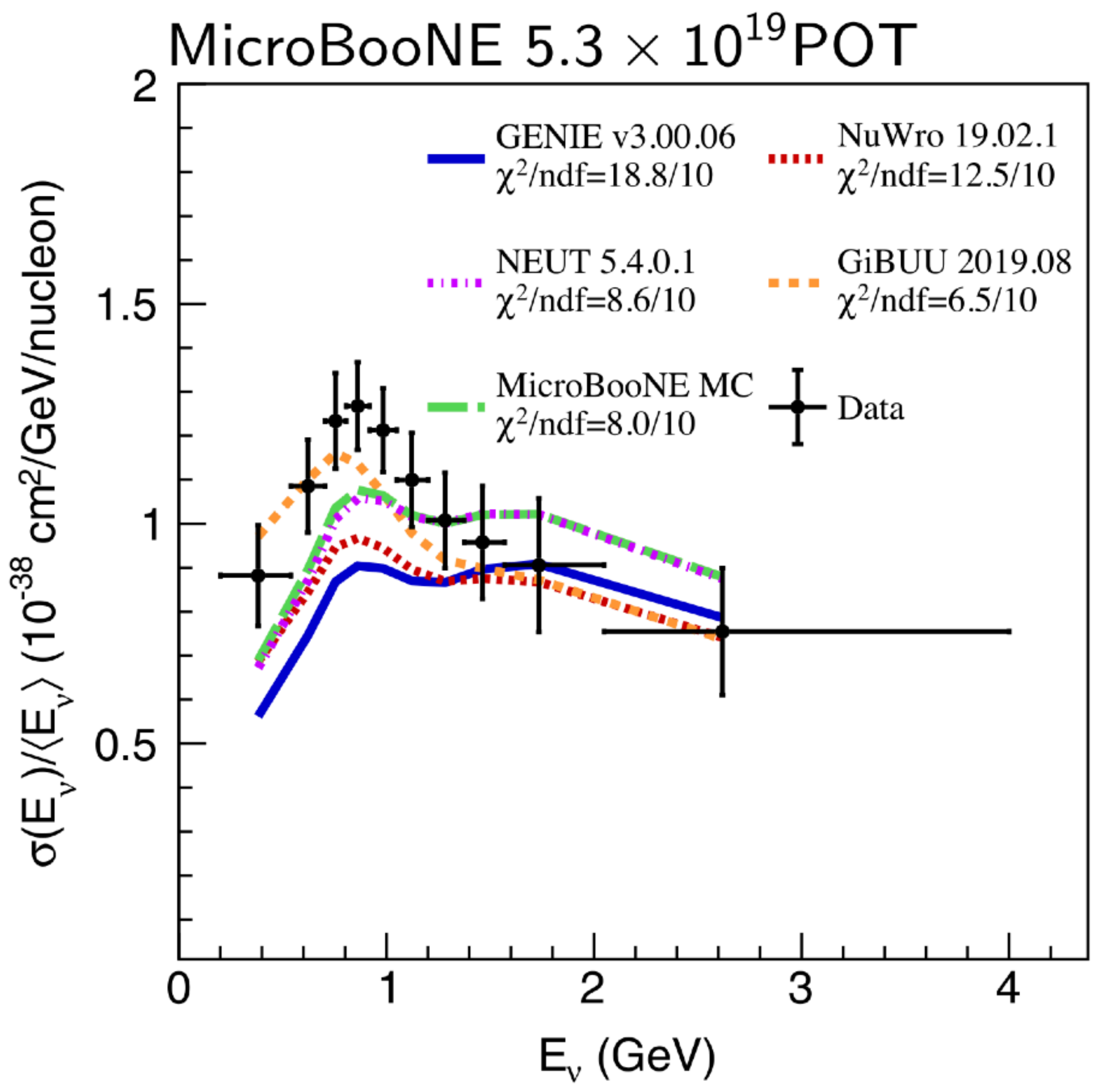


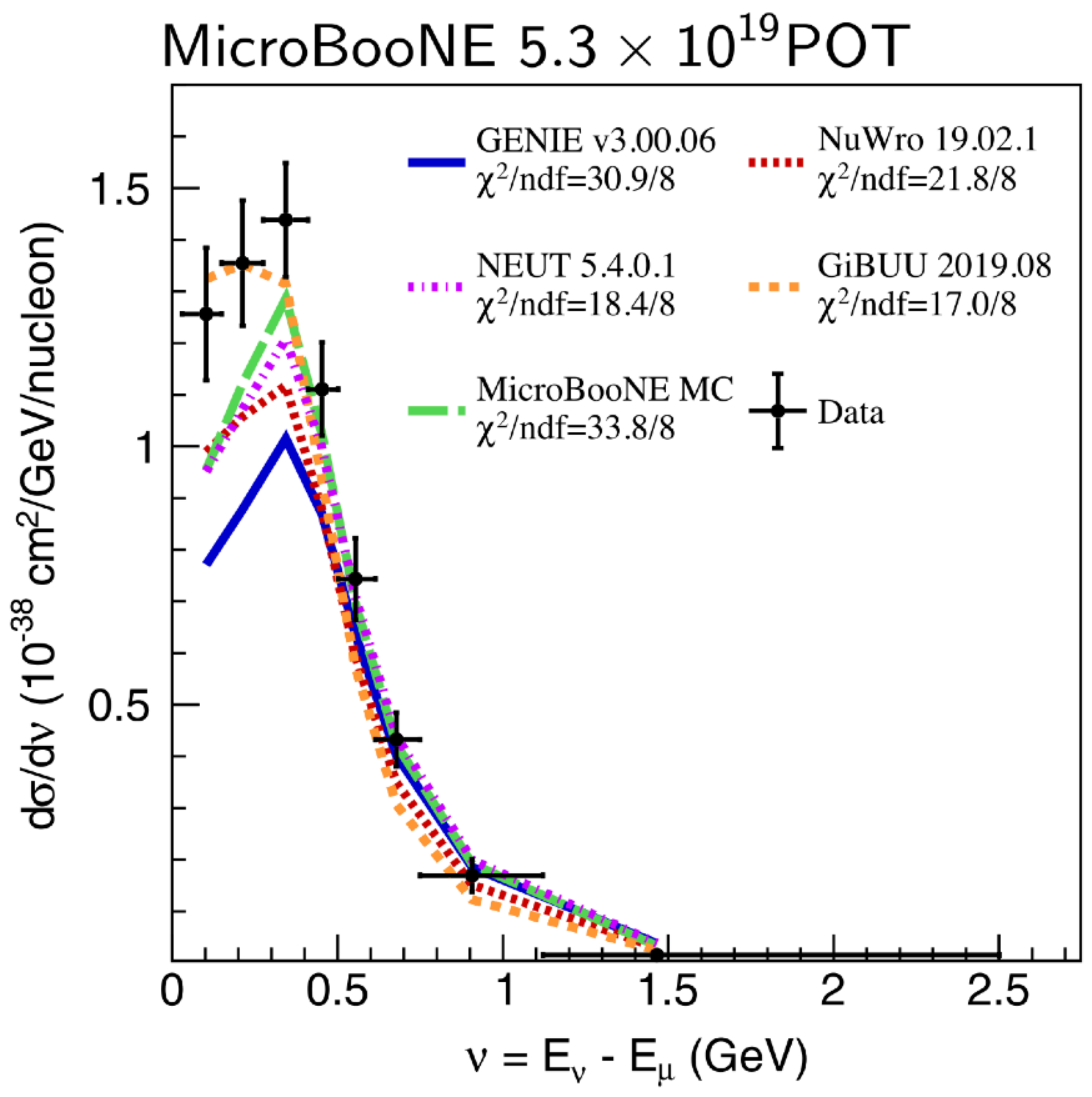
arXiv:2411.03280

Invisible Neutrino Energy Modeling Validation

- We first used this type of constraint in 2022 with lower data statistics
- Validated our modeling before extracting true neutrino energy $\sigma(E_\nu)$ and true energy transfer $d\sigma/d\nu$ cross sections



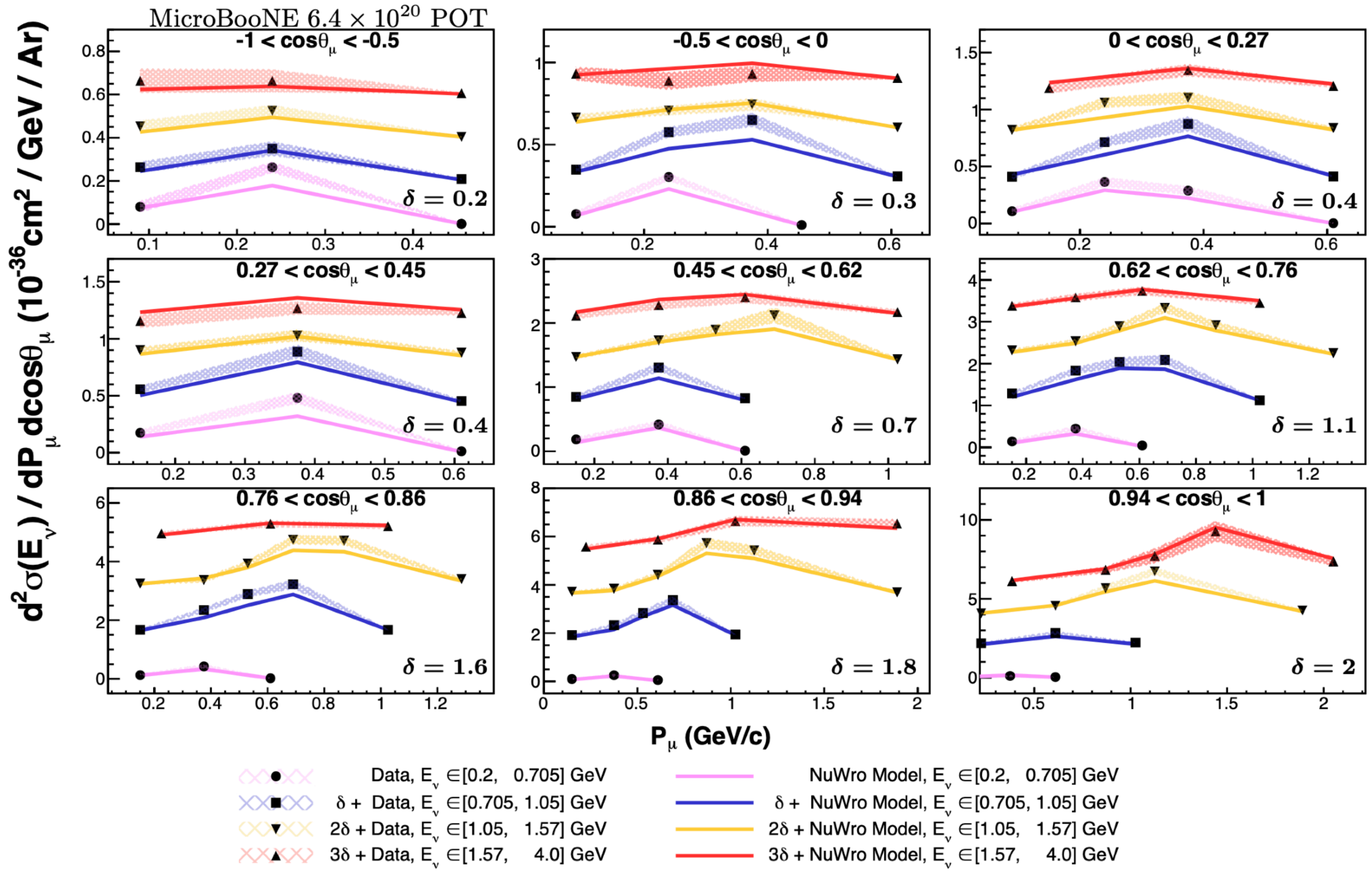




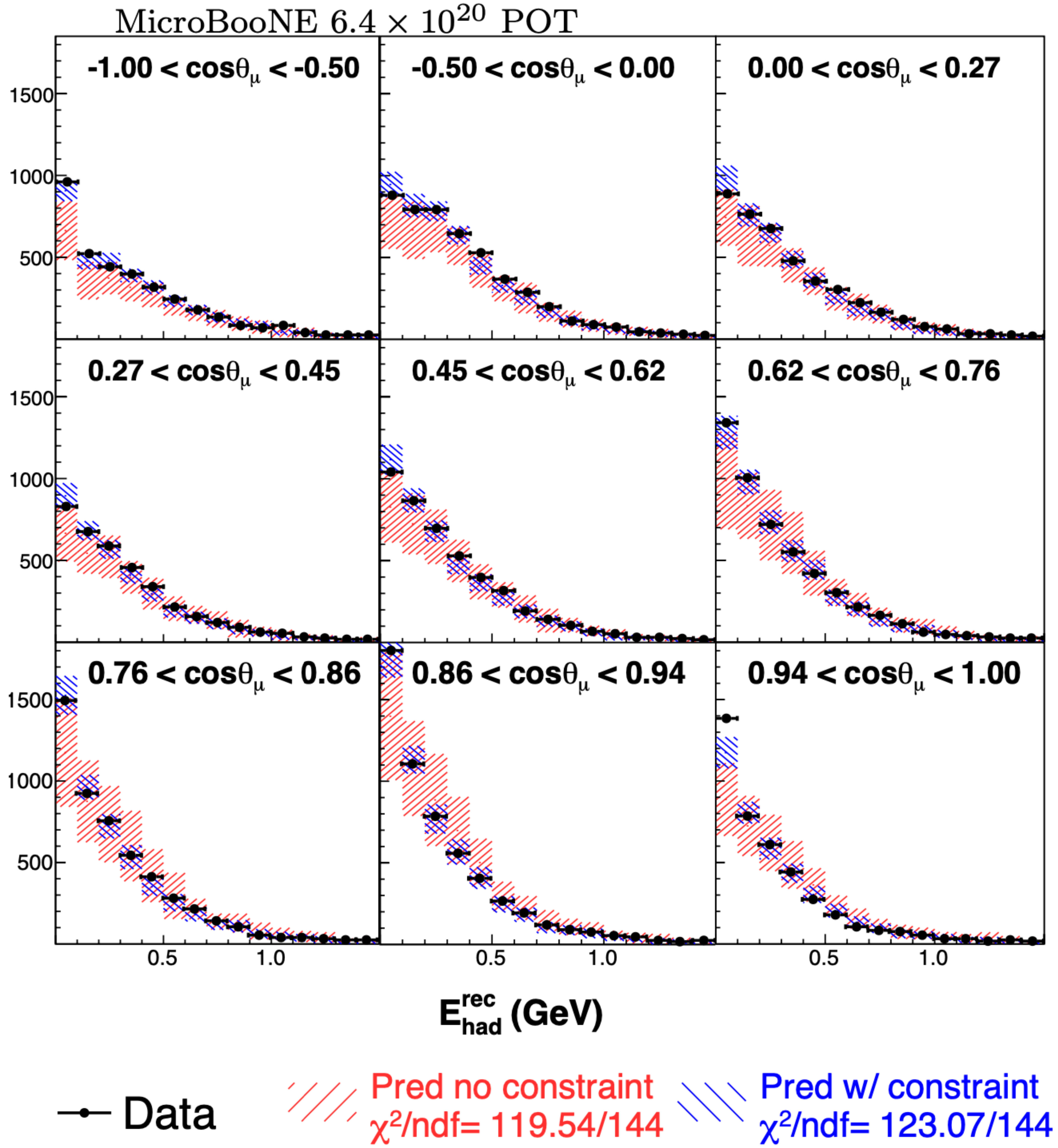
Phys. Rev. Lett. 128, 151801 (2022)

Invisible Neutrino Energy Modeling Validation

- In 2023, we expanded a similar model validation to multiple dimensions before extracting 3D cross sections including an E_ν dimension



$\{E_{\rm rec}^{\rm had}, \cos\,\theta_\mu\}$ constrained by $\{P_\mu, \cos\,\theta_\mu\}$



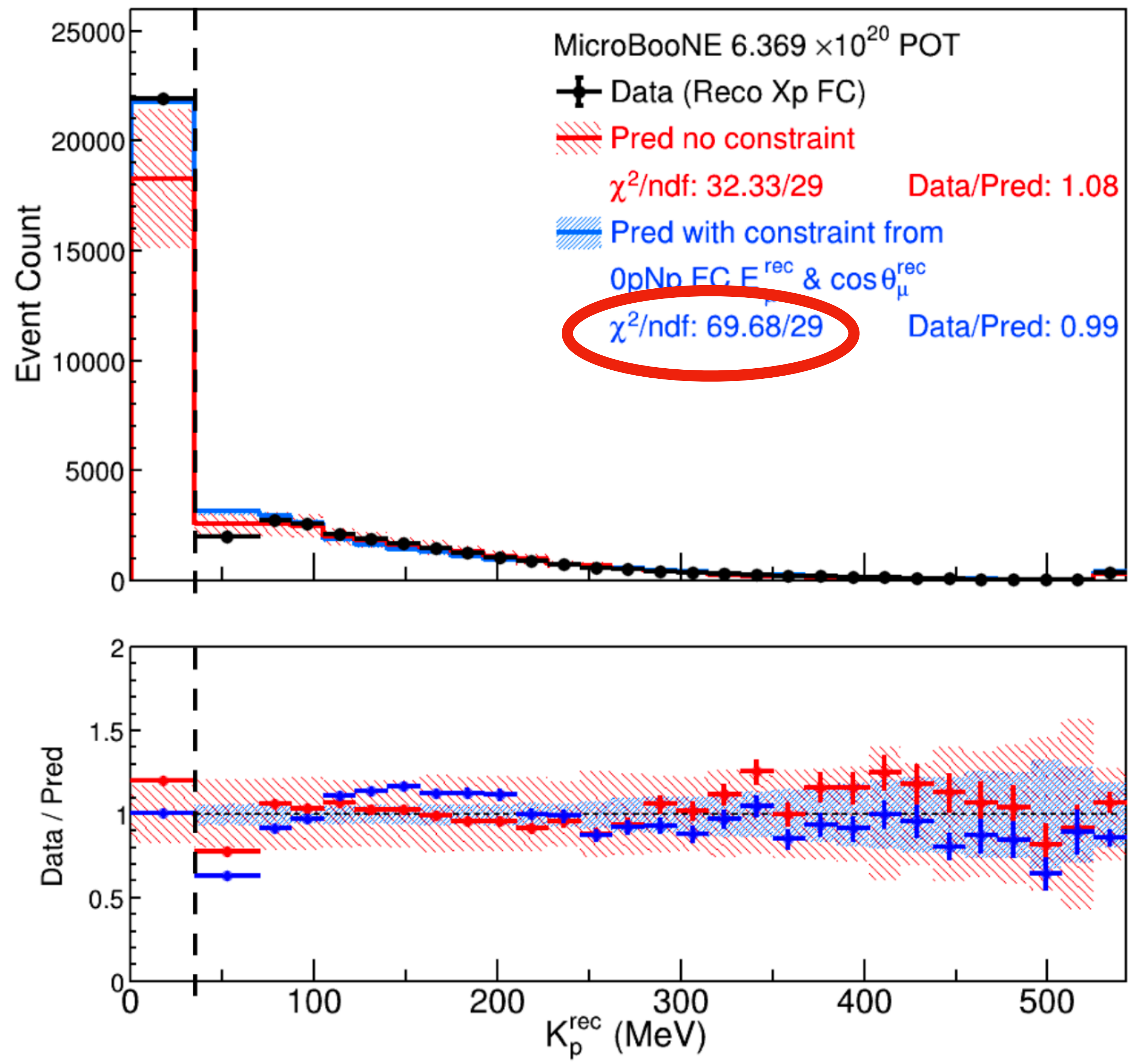
arXiv:2307.06413

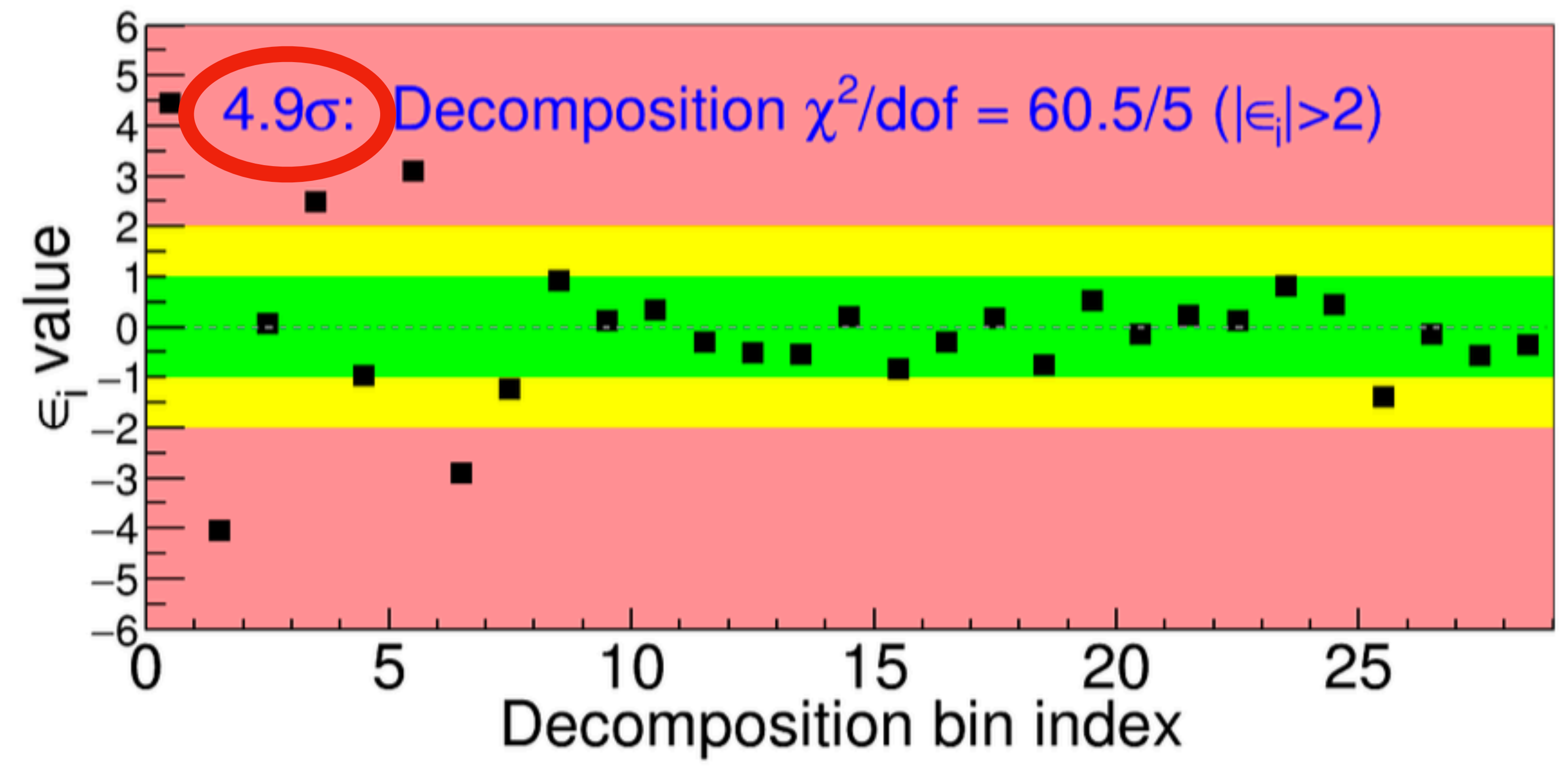
Data-driven Model Validation Has Its Own Similar Limitations

Traditional Fake-Data Closure Testing	Data-driven Model Validation
• 1: What if the spread of models included in fake data tests do not describe the real cross section?	• 1: What if the variety of model validation tests performed does not detect relevant mis-modeling?
• 2: What if the spread of models included in fake data tests is very large, and makes you expand to very large uncertainties, even when your original model is good?	• 2: What if a model validation test fails, but the failure is actually in a phase space irrelevant to the analysis (not significantly affecting the detector response, efficiency, and background prediction), leading to an unnecessary expansion of uncertainties?
• 3: How do you know when to stop? There are many generators, and many configurations and tunes, is testing just one or a few alternate generators enough?	• 3: How do you know when to stop? There are many model validation tests you can think of, how many should you perform?

What About When Model Validation Fails?

- For one analysis, we examined ν_μCC inclusive final states with and without protons
- Our model validation for the reconstructed leading proton energy K_p^{rec} fails!
- A low energy proton connected to a muon is a topology that LArTPCs can study much more precisely than some other detectors
- Low energy proton mis-modeling could potentially cause incorrect neutrino background estimates in searches for coherent interactions or BSM decay-in-flight events

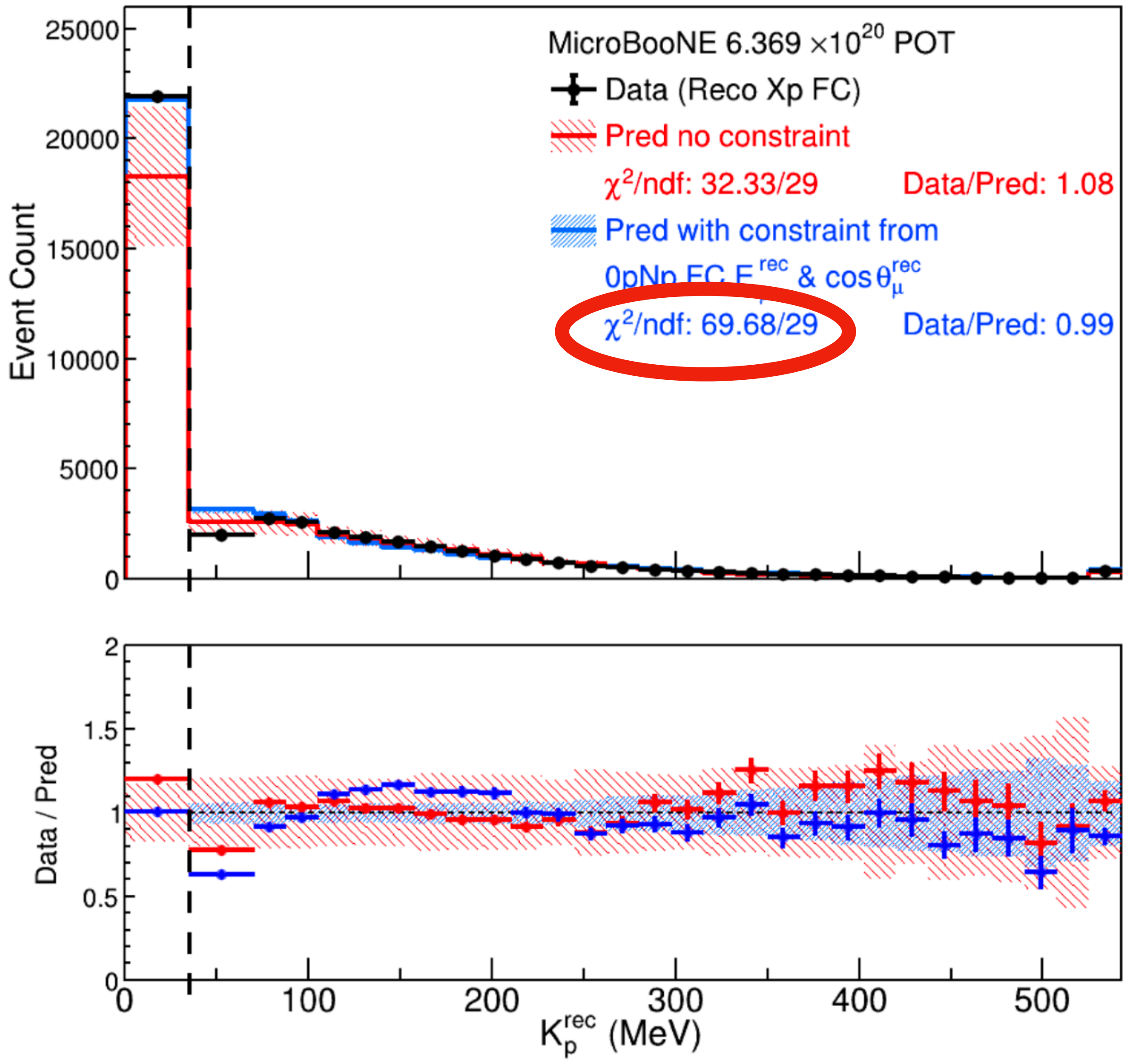


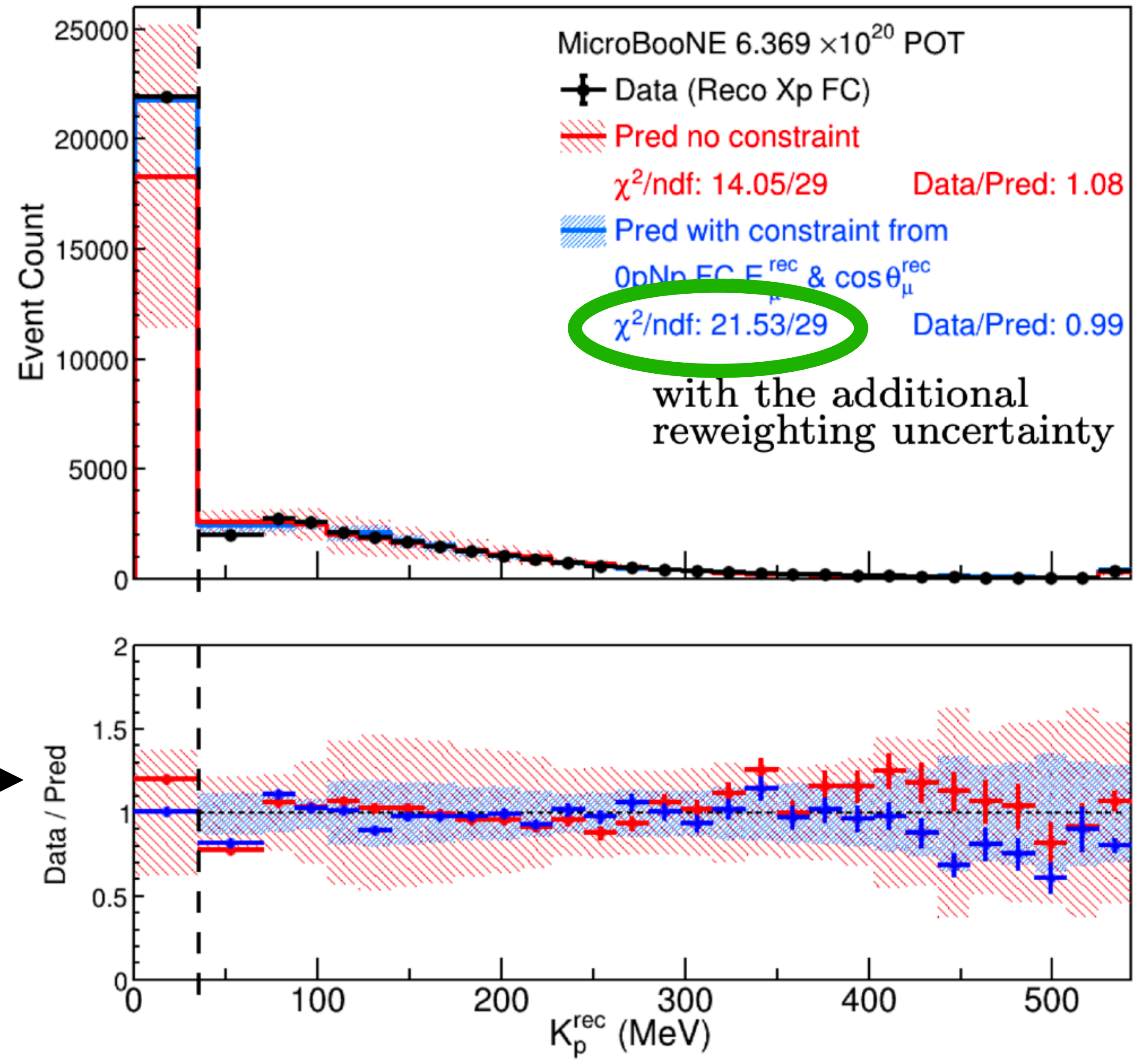


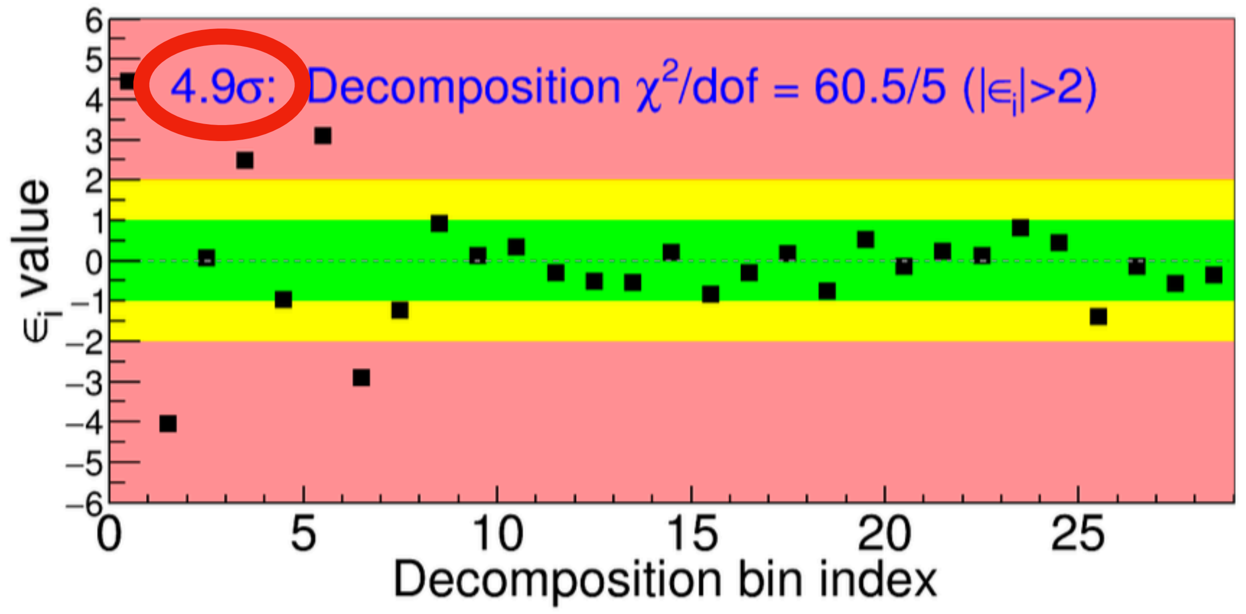
Phys. Rev. D 110, 013006 (2024) Phys. Rev. Lett. 133, 041801 (2024)

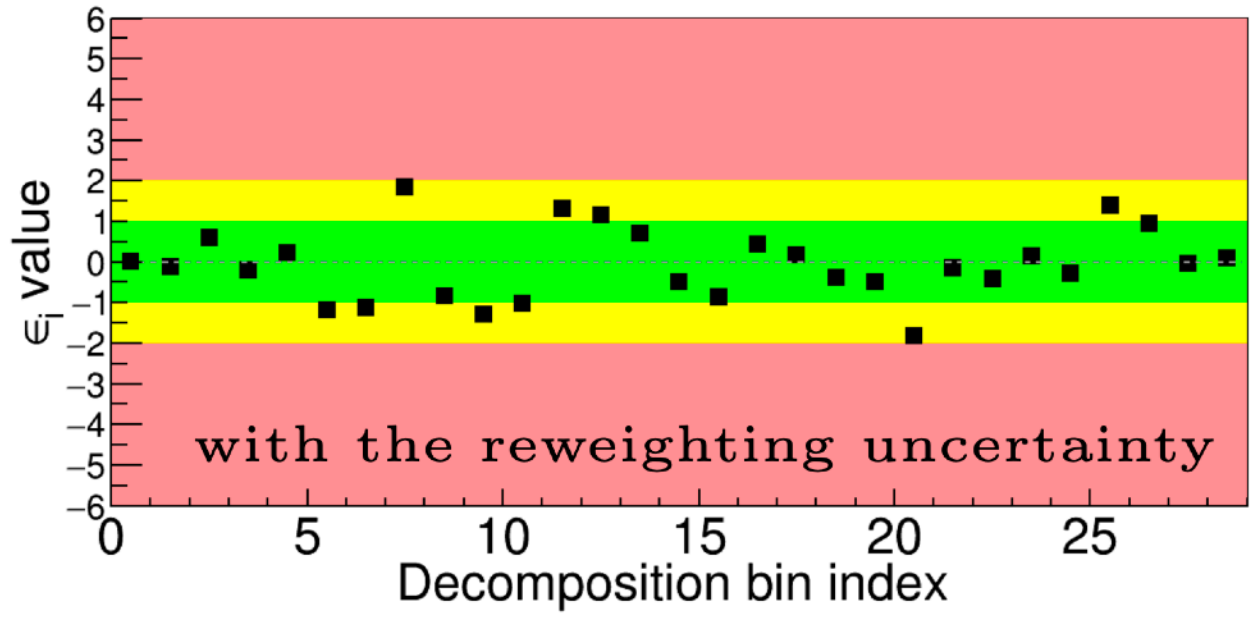
What About When Model Validation Fails?

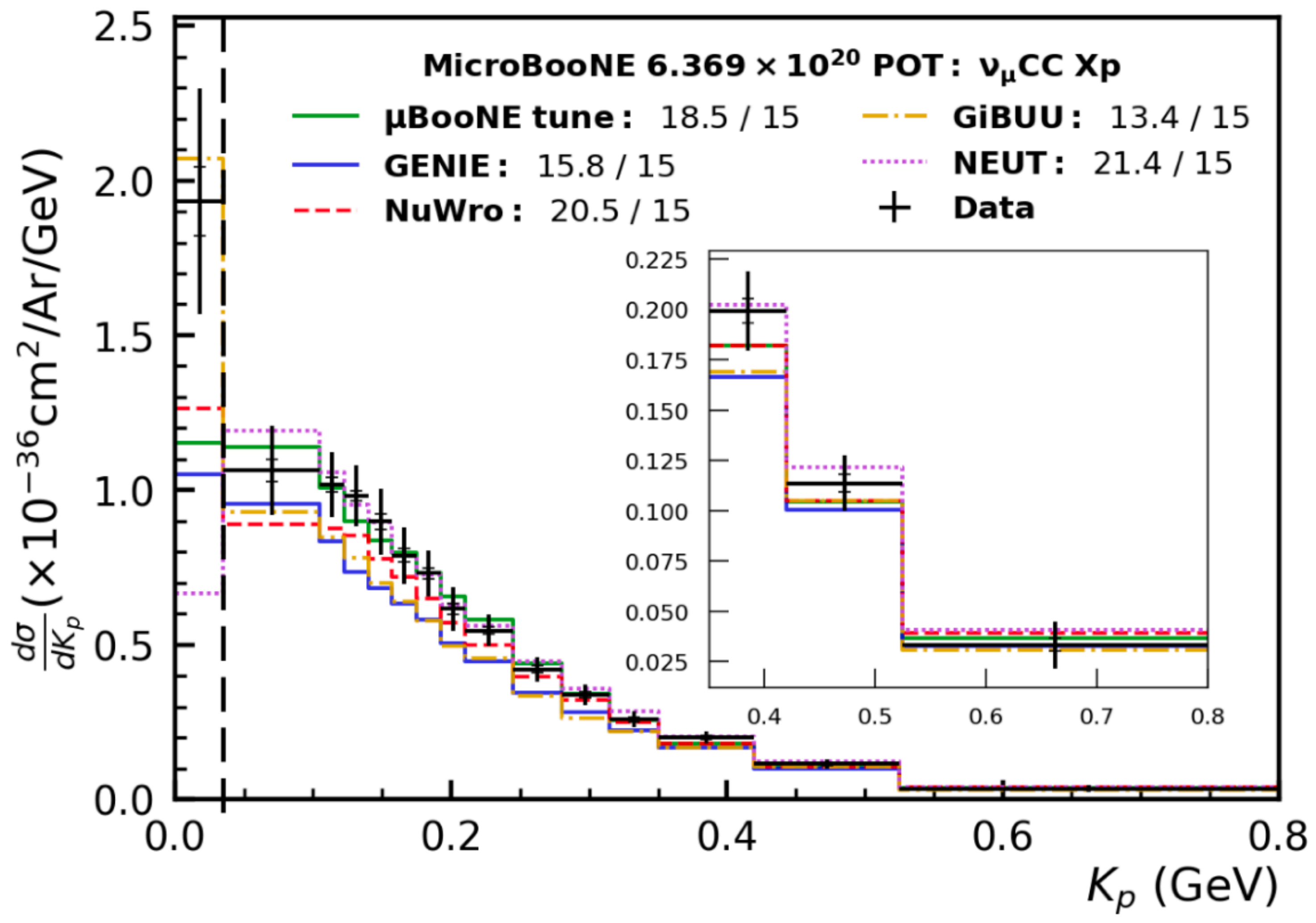
- We use this data-simulation difference to expand our cross section uncertainty
- Unfold this distribution (statistical uncertainty only) to get a reweighting binned in true K_p
- We use this reweighting function to form a new covariance matrix describing this data/MC difference, including correlated and uncorrelated terms
- When we use this to expand our cross section uncertainty, we pass all model validation tests
- Then we can extract cross sections related to protons

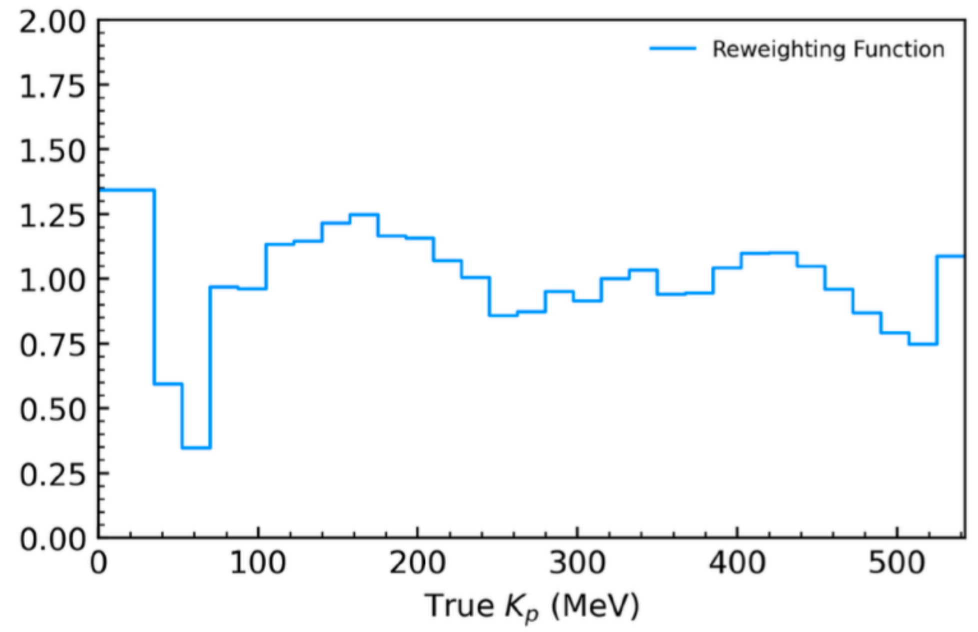












Phys. Rev. Lett. 133, 041801 (2024)

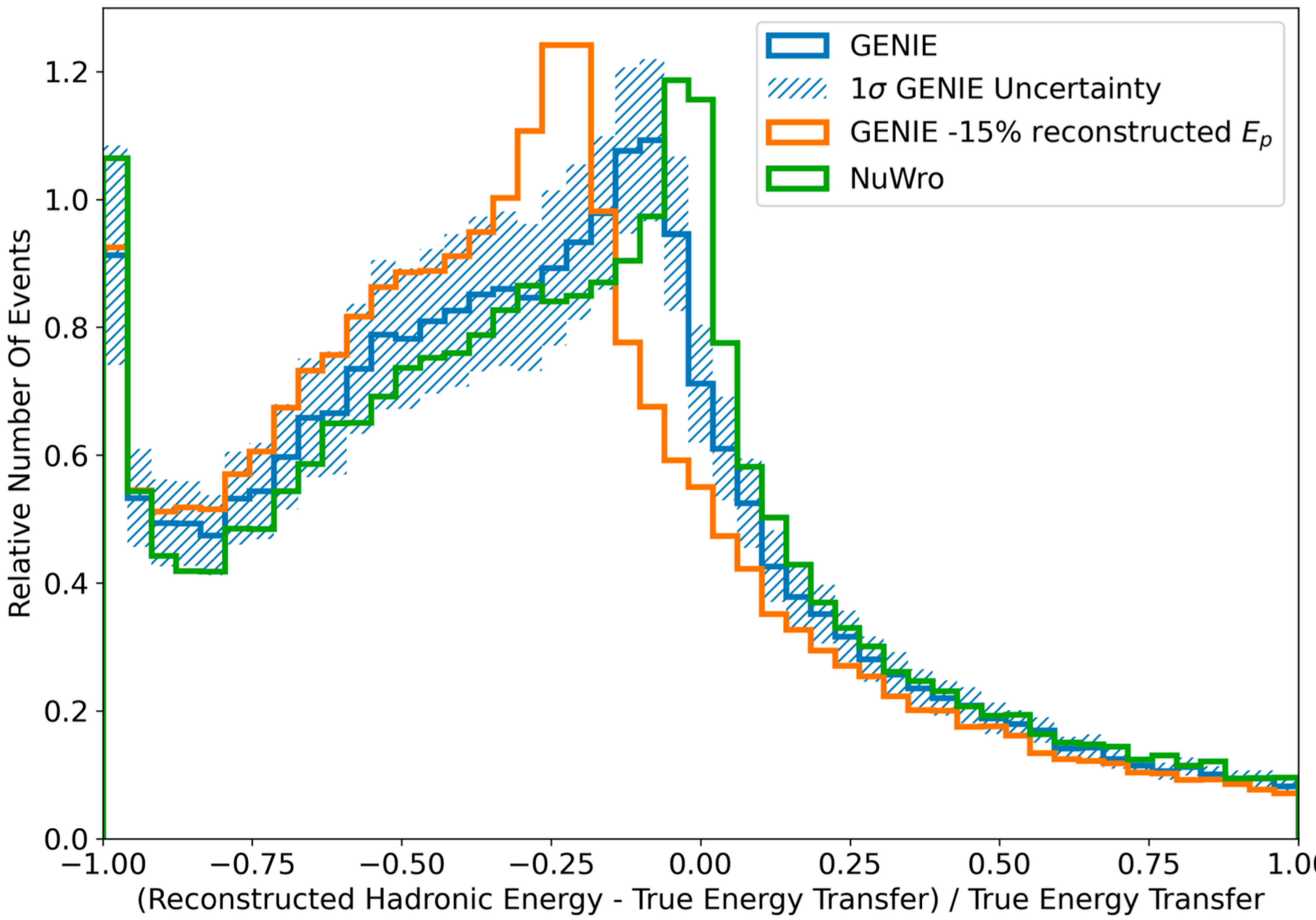
Phys. Rev. D 110, 013006 (2024)

Fake Data Sets For Data-driven Model Validation

- How do we know if a data-driven model validation is sensitive to the types of discrepancies that would bias our cross section results?
  - We demonstrate the sensitivity using fake data sets
- Fake data plays a specific role here:
  - If a fake data set passes model validation, we must ensure that we extract a correct cross section
  - If a fake data set fails model validation, then we conclude that it is too far outside our modeled cross section uncertainty, and that an extracted result would have too much bias
- We don’t use fake data to test the robustness of our cross section model; instead, we use fake data to test the robustness of our model validation procedure

Proton Energy Scaling Fake Data

- One concern about $\sigma(E_\nu)$ modeling: the splitting between $E_{\text{had}}^{\text{vis}}$ and $E_{\text{had}}^{\text{invis}}$ could be modeled wrong
  - For example, maybe we model that 50% of hadronic energy in an event goes to protons and 50% goes to neutrons, but in reality that split is 40%/60%
- We can create a fake data test for this type of scenario by scaling reconstructed proton energies down by 15%
- This shift is significant, outside our modeled GENIE uncertainties
- This type of reconstructed hadronic energy shifting relative to GENIE is also seen in alternate generators, for example NuWro



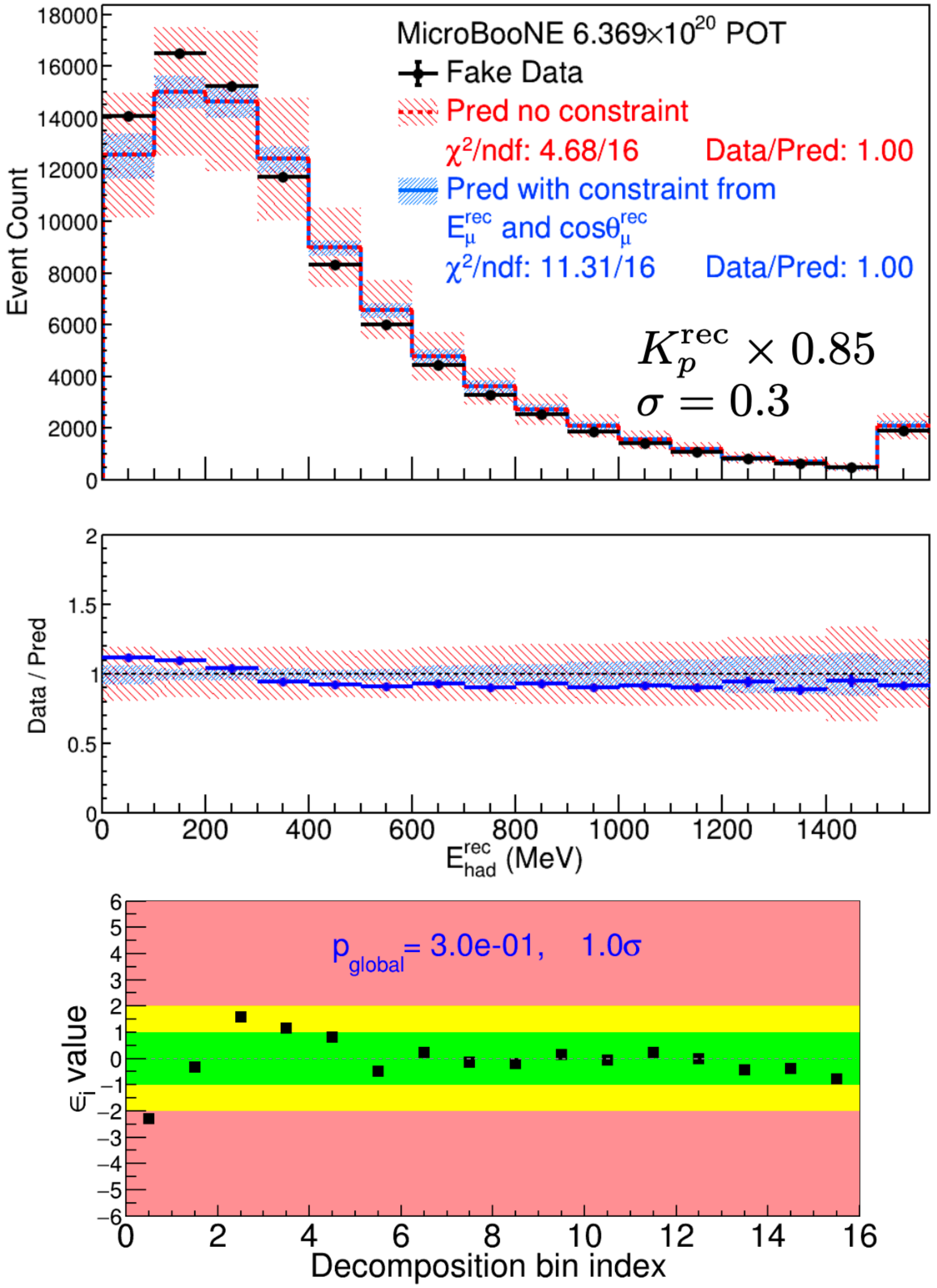
True Energy Transfer $\nu = E_\nu - E_\mu$

arXiv:2411.03280

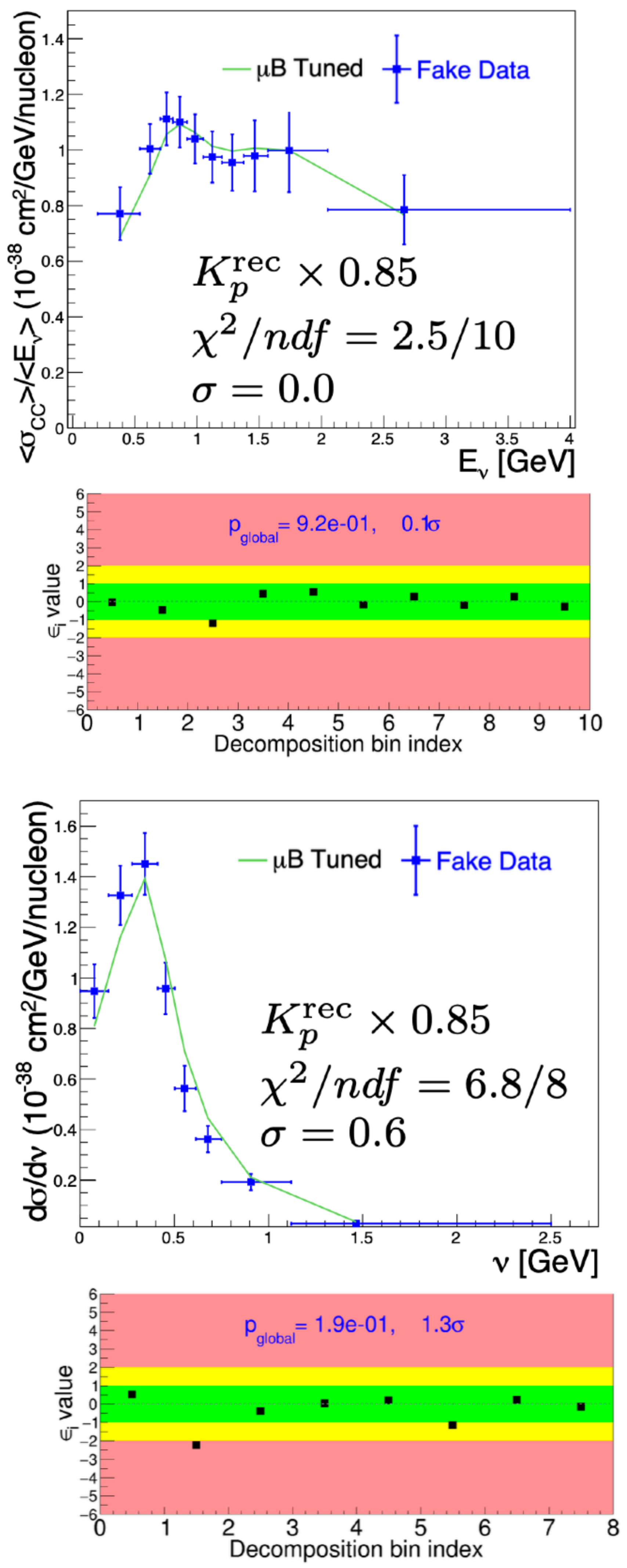
Proton Energy Scaling Fake Data

- With a 15% downscaling of reconstructed proton energy:
  - Good model validation
  - Good cross section extraction
- This tells us that if our modeling of the split between visible and invisible hadronic energy is off by 15%, it would not bias our $\sigma(E_\nu)$ and $d\sigma/d\nu$ cross section results

Good Model Validation



Good XS Extractions

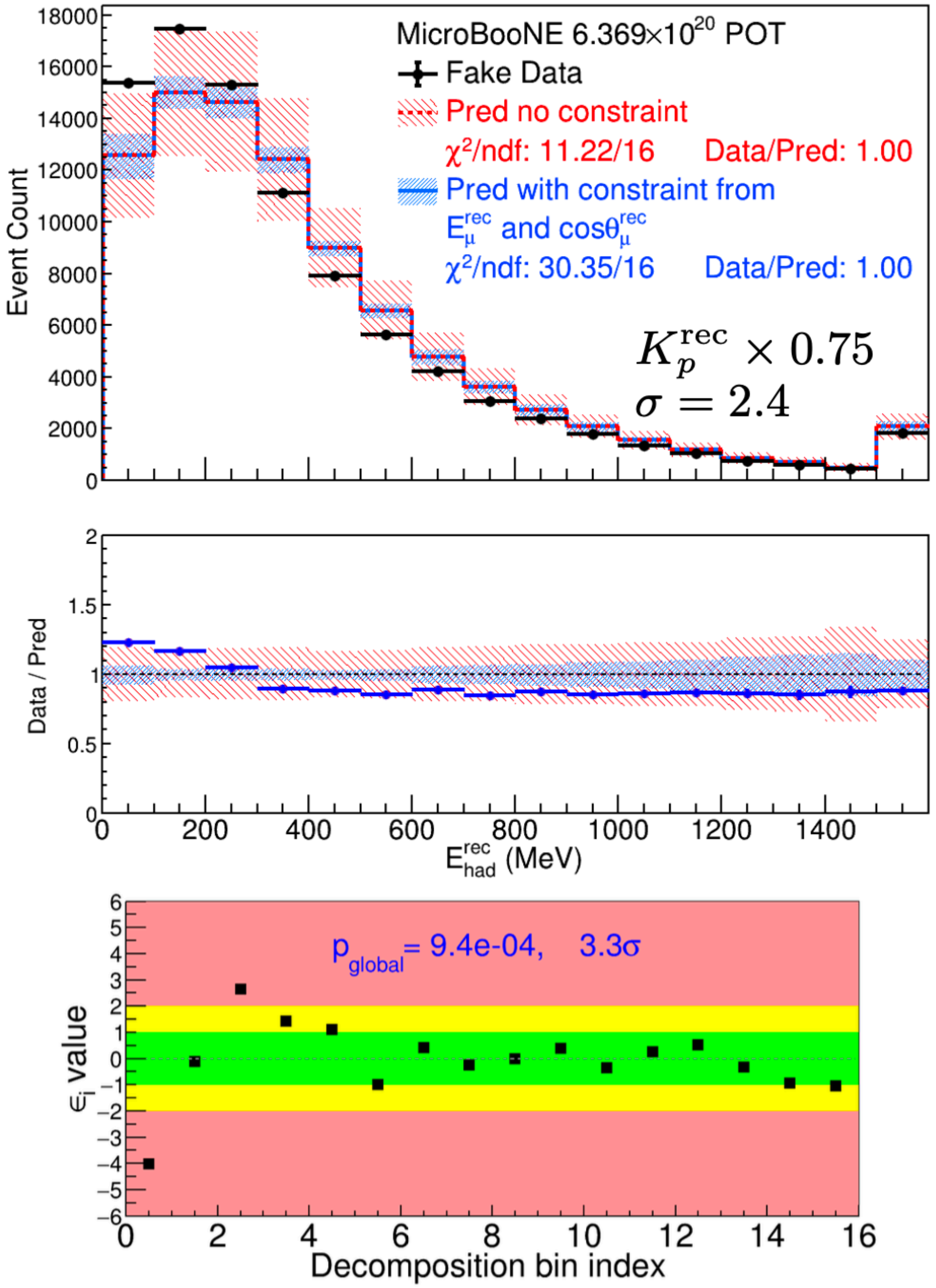


arXiv:2411.03280

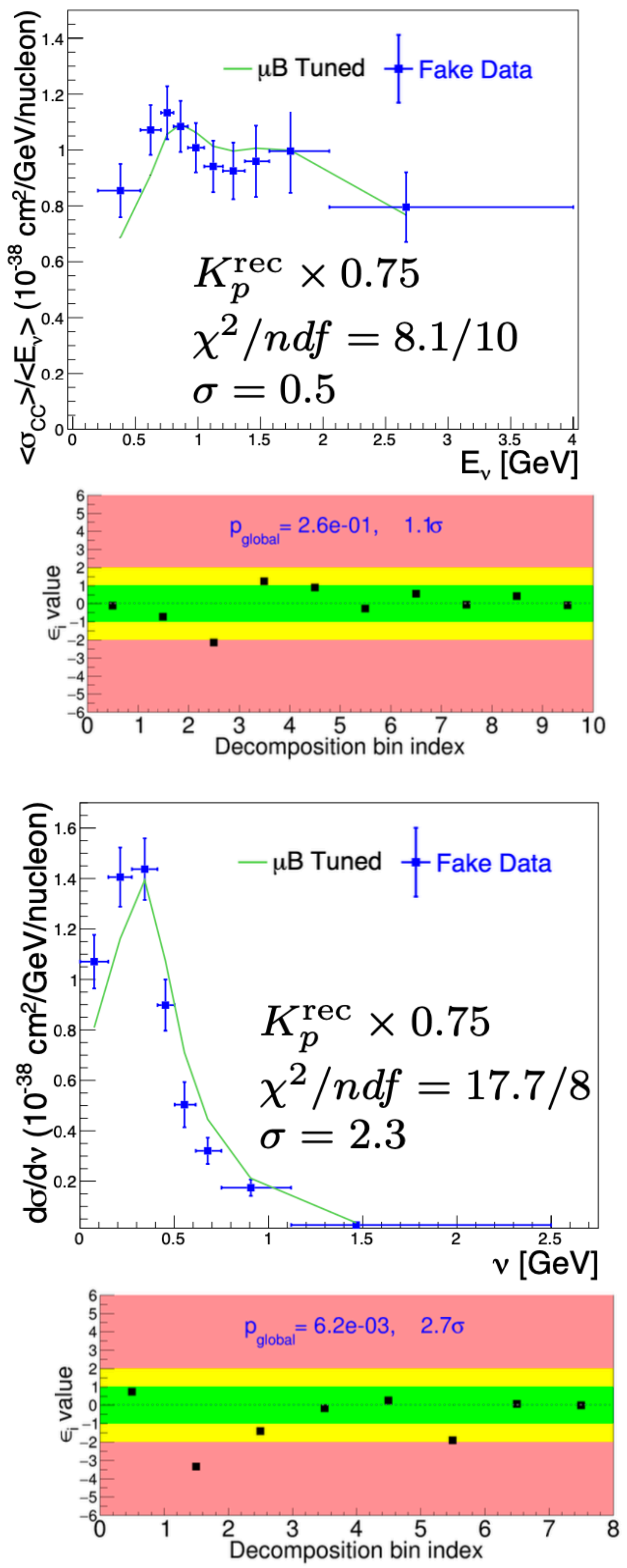
Proton Energy Scaling Fake Data

- With a larger 25% downscaling of reconstructed proton energy:
  - Bad model validation
  - Bad cross section extraction
- This tells us that if our modeling of the split between visible and invisible hadronic energy was this bad, we would notice this ahead of time with model validation tests, and would stop before extracting any biased cross sections

Bad Model Validation



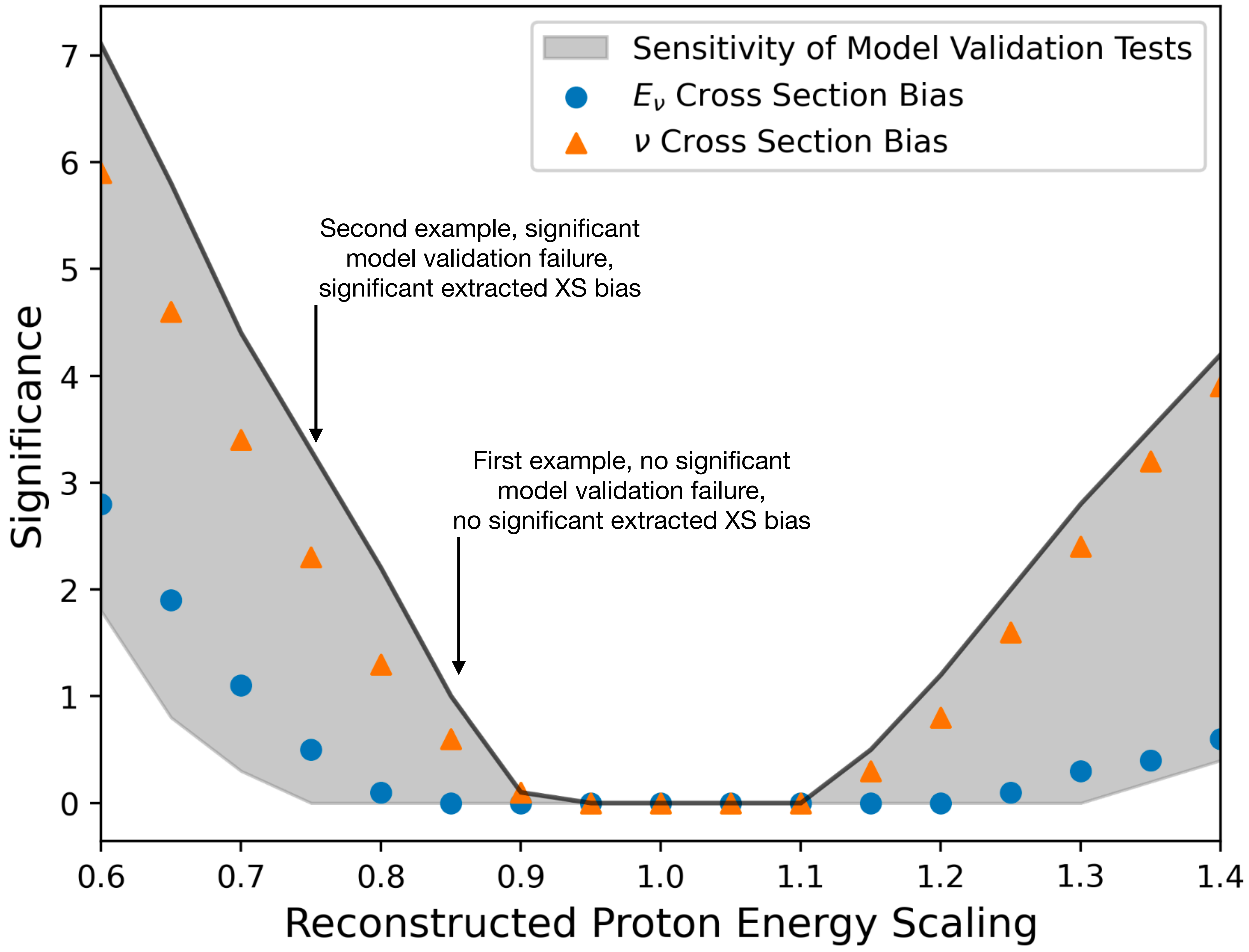
Bad XS Extractions



arXiv:2411.03280

Proton Energy Scaling Fake Data

- Those were examples at proton energy scalings of 0.85 and 0.75, and with a specific model validation test: constraining $E_{\rm had}^{\rm rec}$ using FC&PC $E_\mu^{\rm rec}$ and $\cos\theta_\mu^{\rm rec}$ observations
- We also test with more proton energy scalings, and with more model validation tests
- For every scaling, we see that model validation is more sensitive to mis-modeling than extracted cross section results
- None of these scenarios would lead us to extract a biased cross section
  - In every case, we would either extract a correct cross section, or we would stop when model validation failed



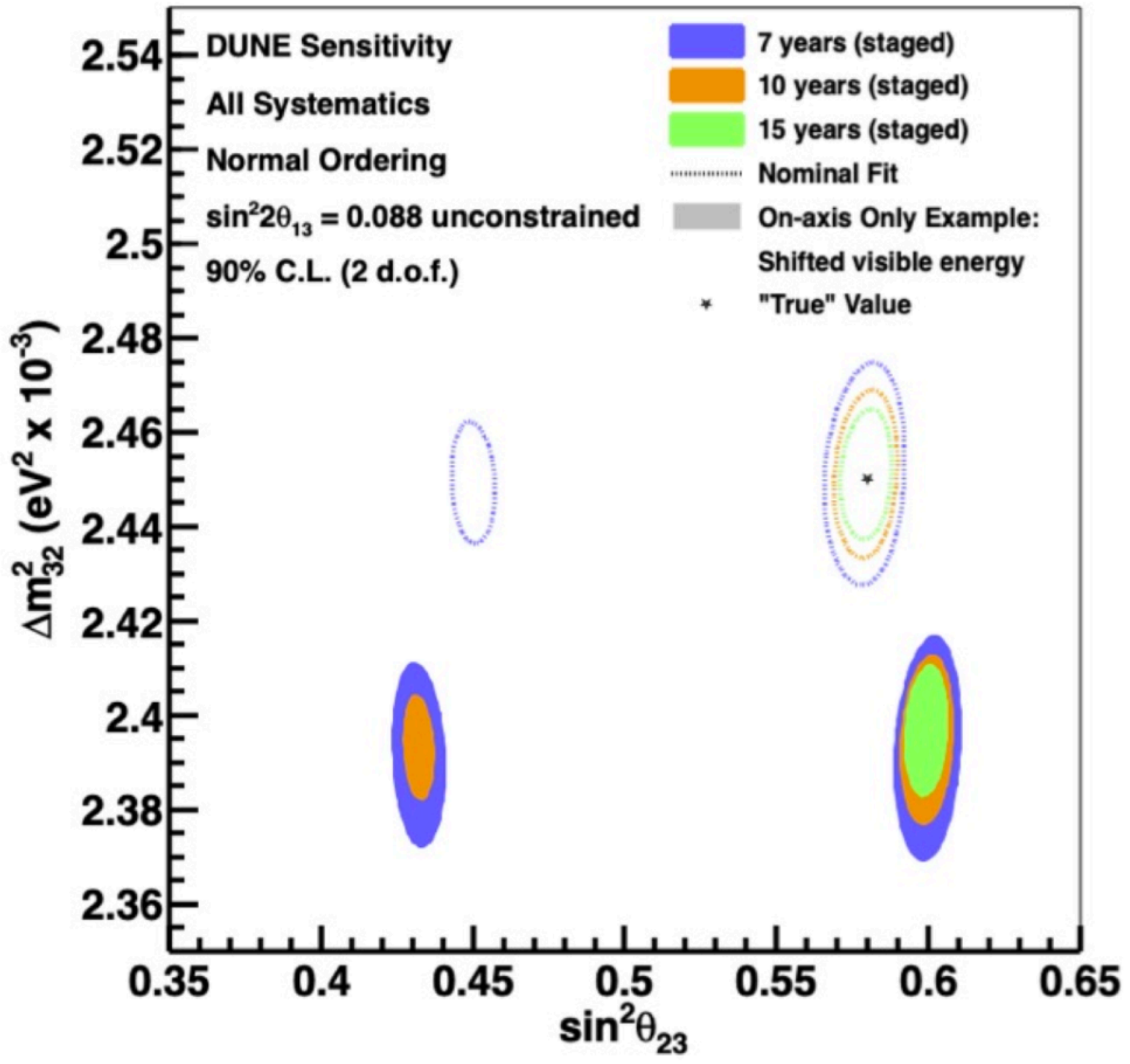
List of 50 different model validation tests performed at each scaling:

- Evaluation of the $E_\mu^{\rm rec}$ FC, PC and FC&PC distributions through overall χ^2 GoF tests and χ^2 decompositions (6 total tests).
- Evaluation of the $E_\mu^{\rm rec}$ PC distribution after constraint from the analogous FC distribution. The overall χ^2 GoF test and χ^2 decomposition are examined (2 total tests).
- Evaluation of the $\cos\theta_\mu^{\rm rec}$ FC, PC and FC&PC distributions through overall χ^2 GoF tests and χ^2 decompositions (6 total tests).
- Evaluation of the $\cos\theta_\mu^{\rm rec}$ PC distribution after constraint from the analogous FC distribution. The overall χ^2 GoF test and χ^2 decomposition are examined (2 total tests).
- Evaluation of the $\cos\theta_\mu^{\rm rec}$ FC, PC and FC&PC distributions after constraint from the FC&PC $E_\mu^{\rm rec}$ distribution. The overall χ^2 GoF test and χ^2 decomposition is examined for each distribution (6 total tests).
- Evaluation of the $E_\nu^{\rm rec}$ FC, PC and FC&PC distributions through overall χ^2 GoF tests and χ^2 decompositions (6 total tests).
- Evaluation of the $E_\nu^{\rm rec}$ PC distribution after constraint from the analogous FC distribution. The overall χ^2 GoF test and χ^2 decomposition are examined (2 total tests).
- Evaluation of the $E_\nu^{\rm rec}$ FC, PC and FC&PC distributions after constraint from the FC&PC $E_\mu^{\rm rec}$ and $\cos\theta_\mu^{rec}$ distributions. The overall χ^2 GoF test and χ^2 decomposition is examined for each distribution (6 total tests).
- Evaluation of the $E_{\rm had}^{\rm rec}$ FC, PC and FC&PC distributions through overall χ^2 GoF tests and χ^2 decompositions (6 total tests).
- Evaluation of the $E_{\rm had}^{\rm rec}$ PC distribution after constraint from the analogous FC distribution. The overall χ^2 GoF test and χ^2 decomposition are examined (2 total tests).
- Evaluation of the $E_{\rm had}^{\rm rec}$ FC, PC and FC&PC distributions after constraint from the FC&PC $E_\mu^{\rm rec}$ and $\cos\theta_\mu^{\rm rec}$ distributions. The overall χ^2 GoF test and χ^2 decomposition is examined for each distribution (6 total tests).

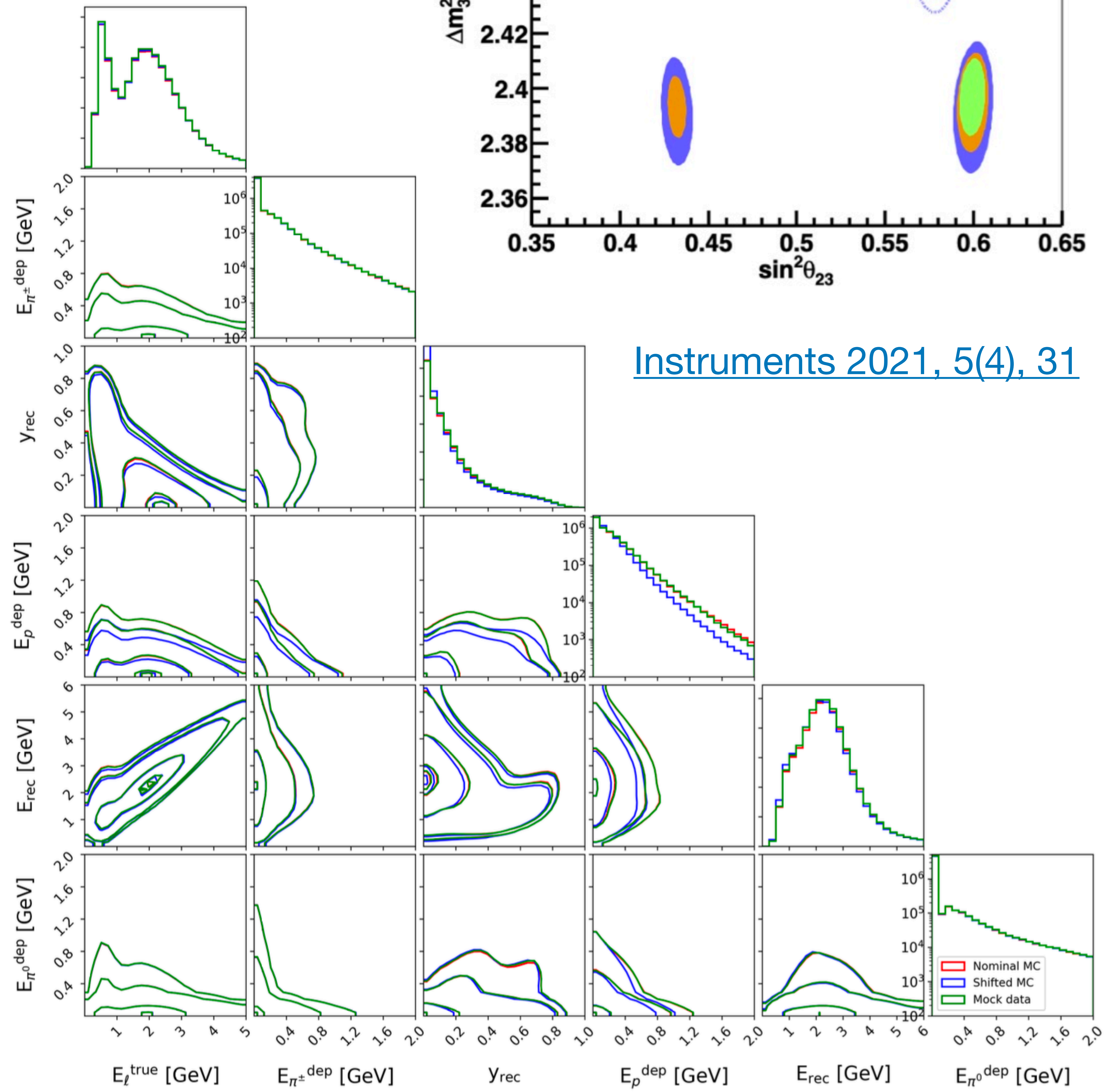
arXiv:2411.03280

DUNE-ND Proton Energy Scaling Fake Data

- What if the modeling is wrong, but the model validation passes anyway?
    - A “nightmare scenario”, unluckily having multiple mis-modelings cancel each other out to give good agreement to data
- This was a possibility considered in the DUNE Near Detector Conceptual Design Report
- What if 20% of the modeled proton energy is actually carried away by neutrons?
    - This would be noticed in the near detector data, but what if there’s additional cross section mis-modeling and tuning that causes the near detector data to look good?
- In this case, we could extract biased oscillation parameters, despite passing all data driven model validation tests in the near detector!

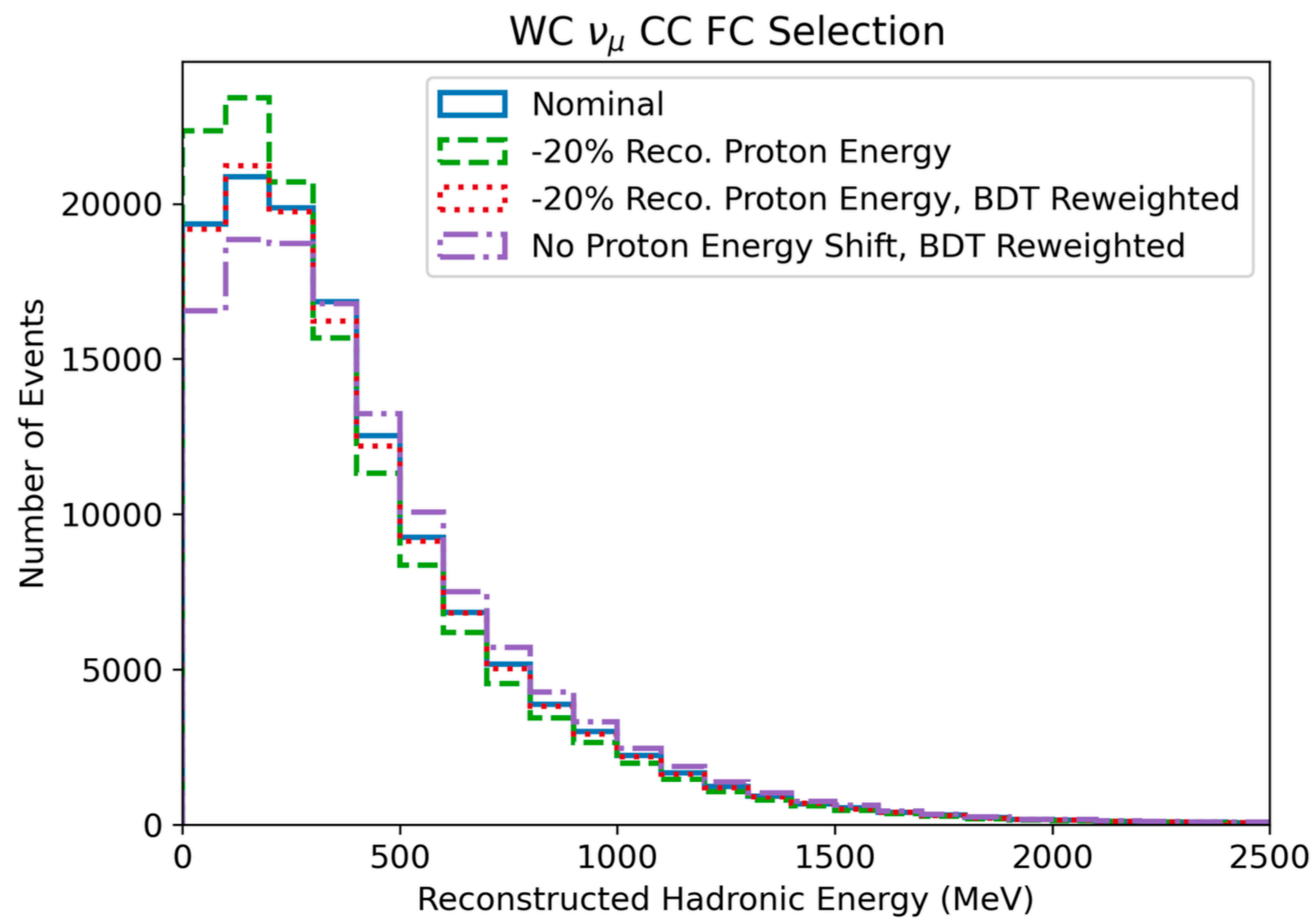


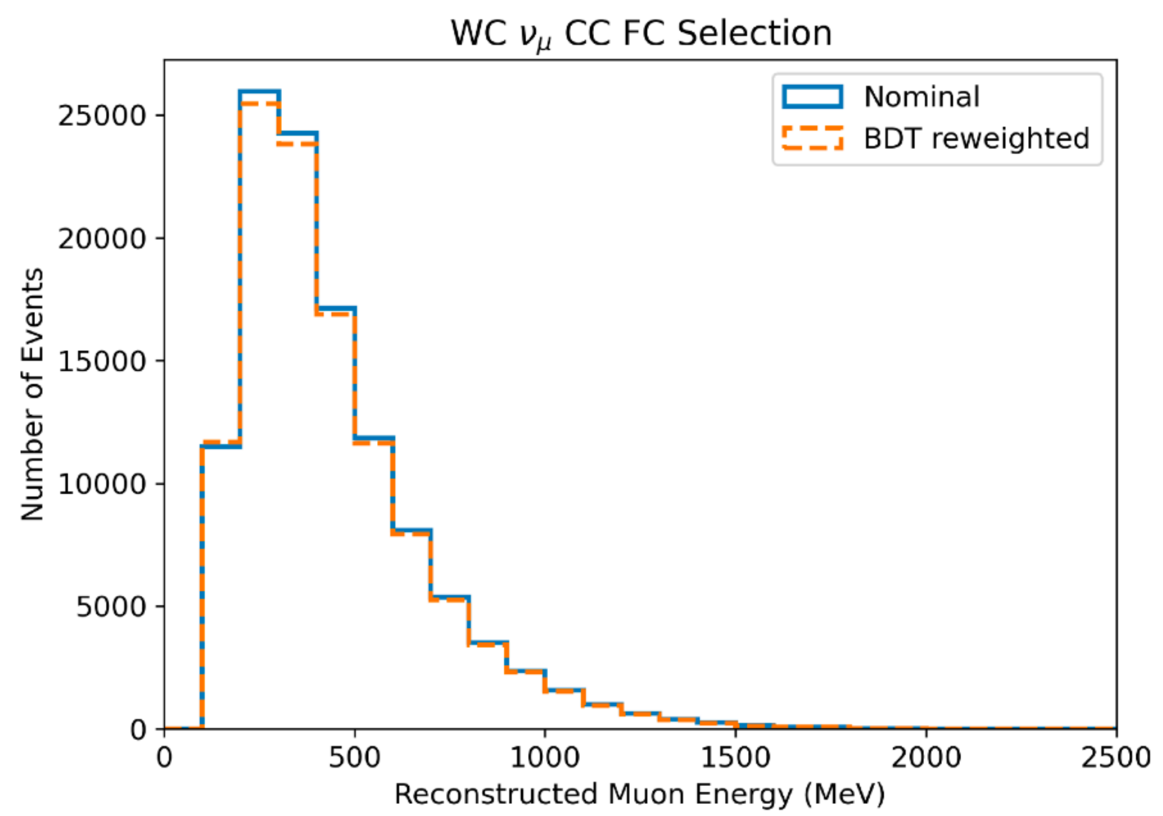
[Instruments 2021, 5(4), 31]

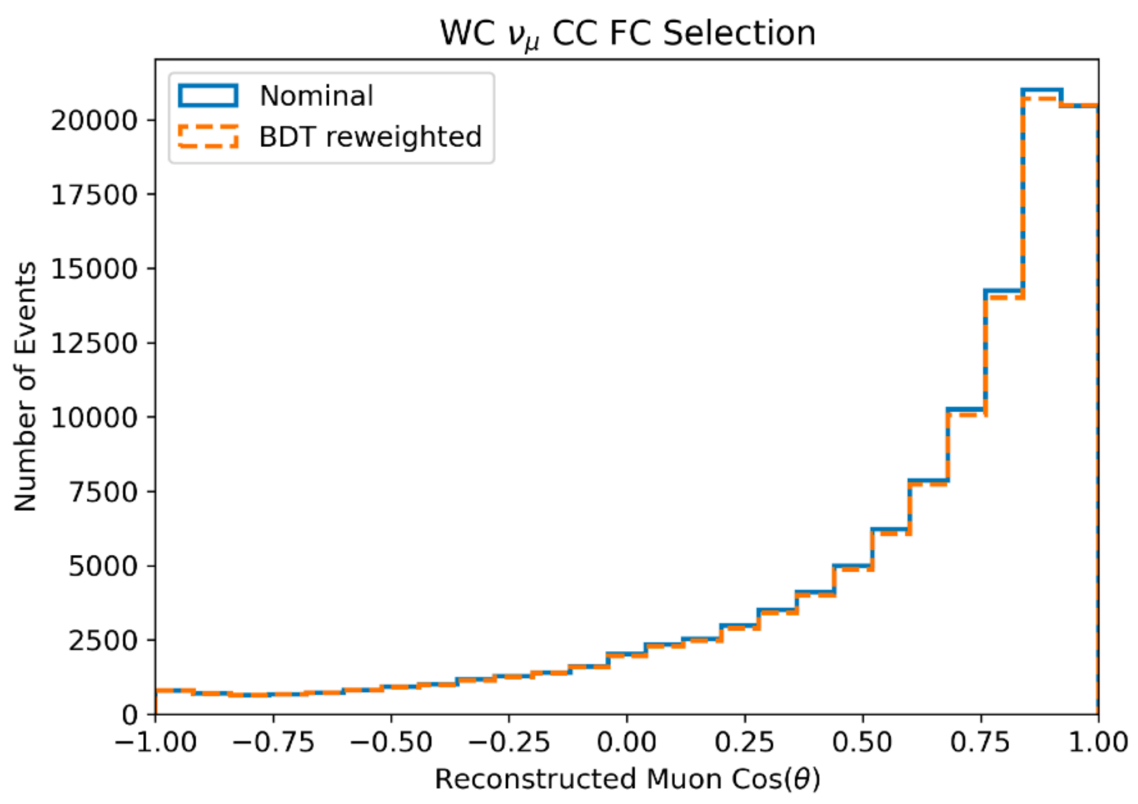


Recreating DUNE-ND Proton Energy Scaling Fake Data

- For illustration, we recreate a similar scenario using MicroBooNE simulation
- We shift the reconstructed proton energy down by 20%
- Then, we have a multivariate BDT reweighting to restore the distributions of E_μ^{true}, $\cos\theta_\mu^{\mathrm{true}}$, and $\nu = E_\nu - E_\mu^{\mathrm{true}}$
  - Simulating a mis-modeled cross section that happens to cancel out the proton energy shift
- This reweighting results in good distributions of E_μ^{rec}, $\cos\theta_\mu^{\mathrm{rec}}$, and $E_{\mathrm{had}}^{\mathrm{rec}}$





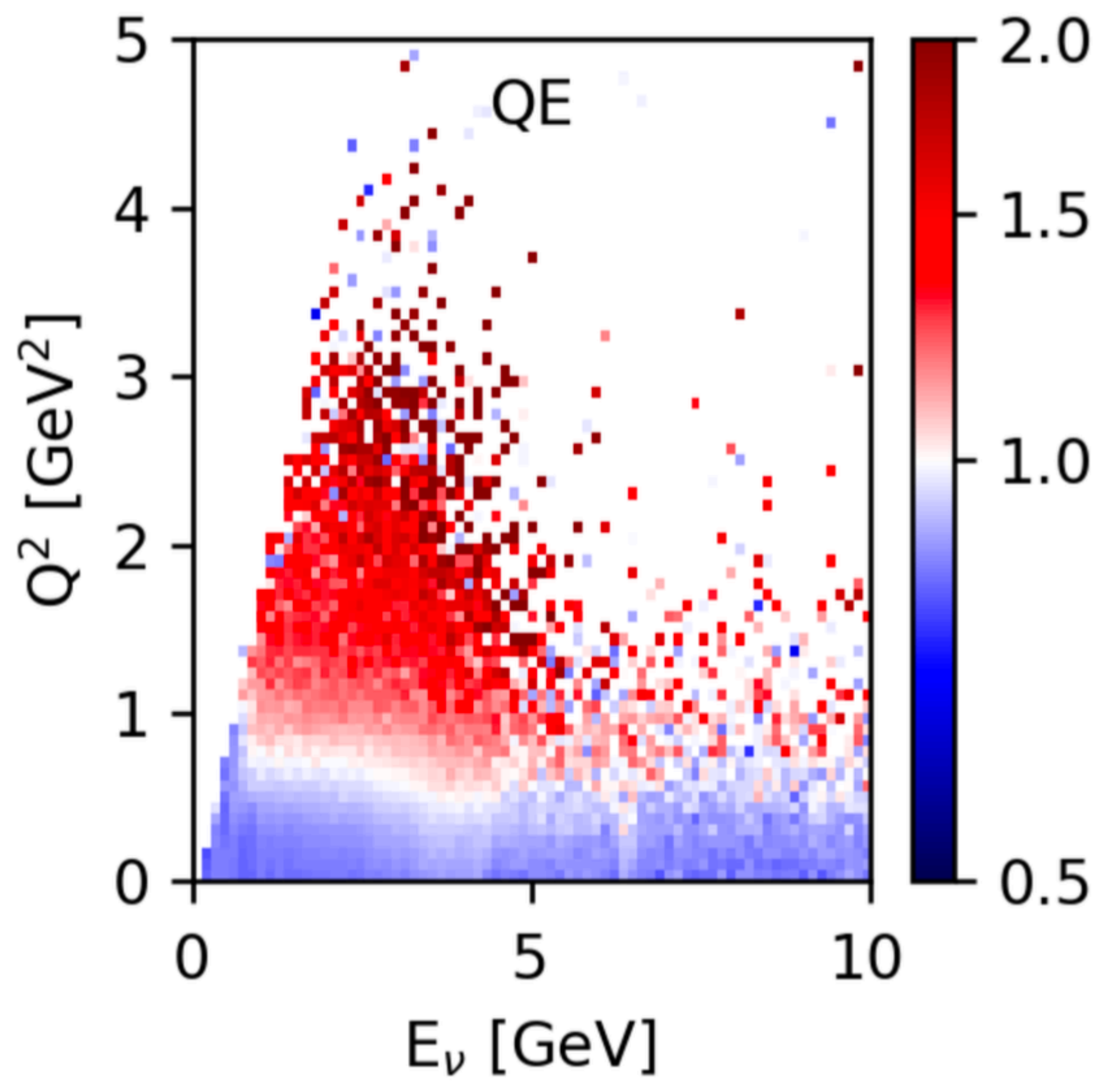


arXiv:2411.03280

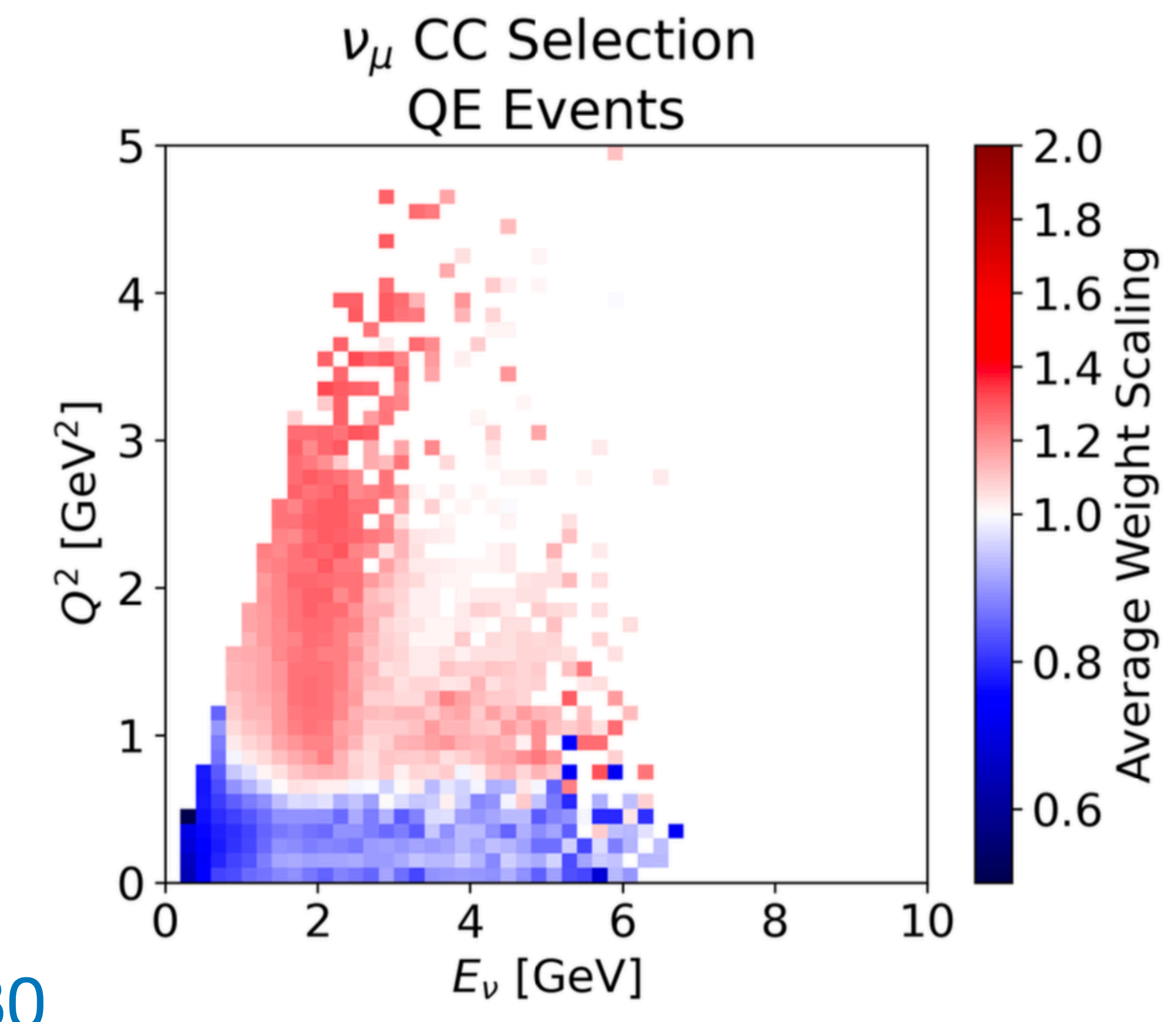
Recreating DUNE-ND Proton Energy Scaling Fake Data

- Looking at QE events:
- This BDT reweighting results in a similar cross section change as DUNE saw
- Significantly scaling up the cross section at large Q^2, and down at low Q^2

From DUNE ND CDR study:



From MicroBooNE study:



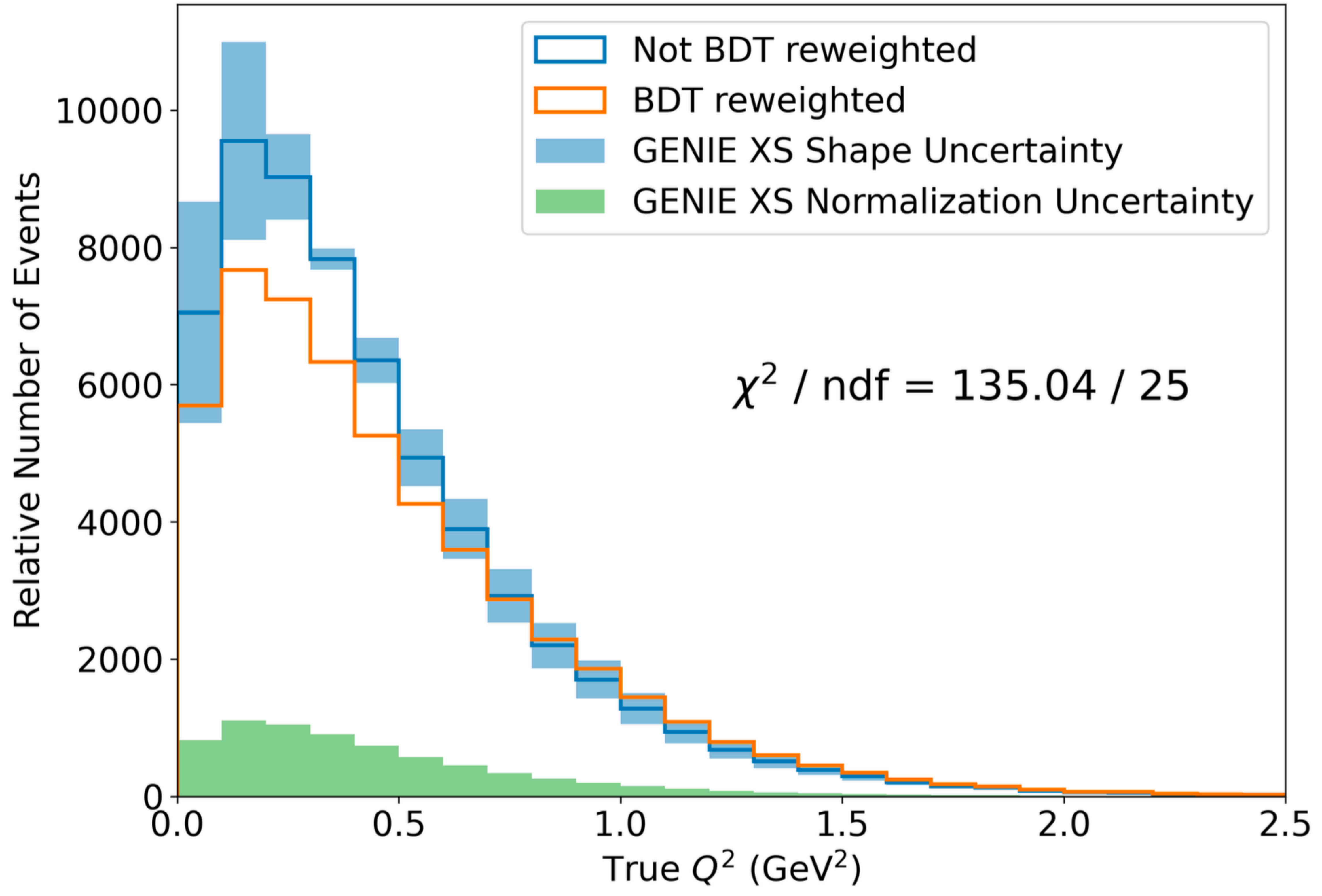
arXiv:2411.03280

Recreating DUNE-ND Proton Energy Scaling Fake Data

- In this fake data test:
    - 0 σ disagreement in our model validation tests
    - 0.7 σ disagreement in the $d\sigma/d\nu$ cross section
- This is a worst-case-scenario, where there is more bias in an extracted cross section than in any model validation test
- How likely are we to end up in this situation, where multiple mis-modelings cancel each other out to give good agreement to data?
- This is hard to quantify, but as an illustration, we can see how this BDT-reweighted model holds up from some other perspectives

Recreating DUNE-ND Proton Energy Scaling Fake Data: QE Events

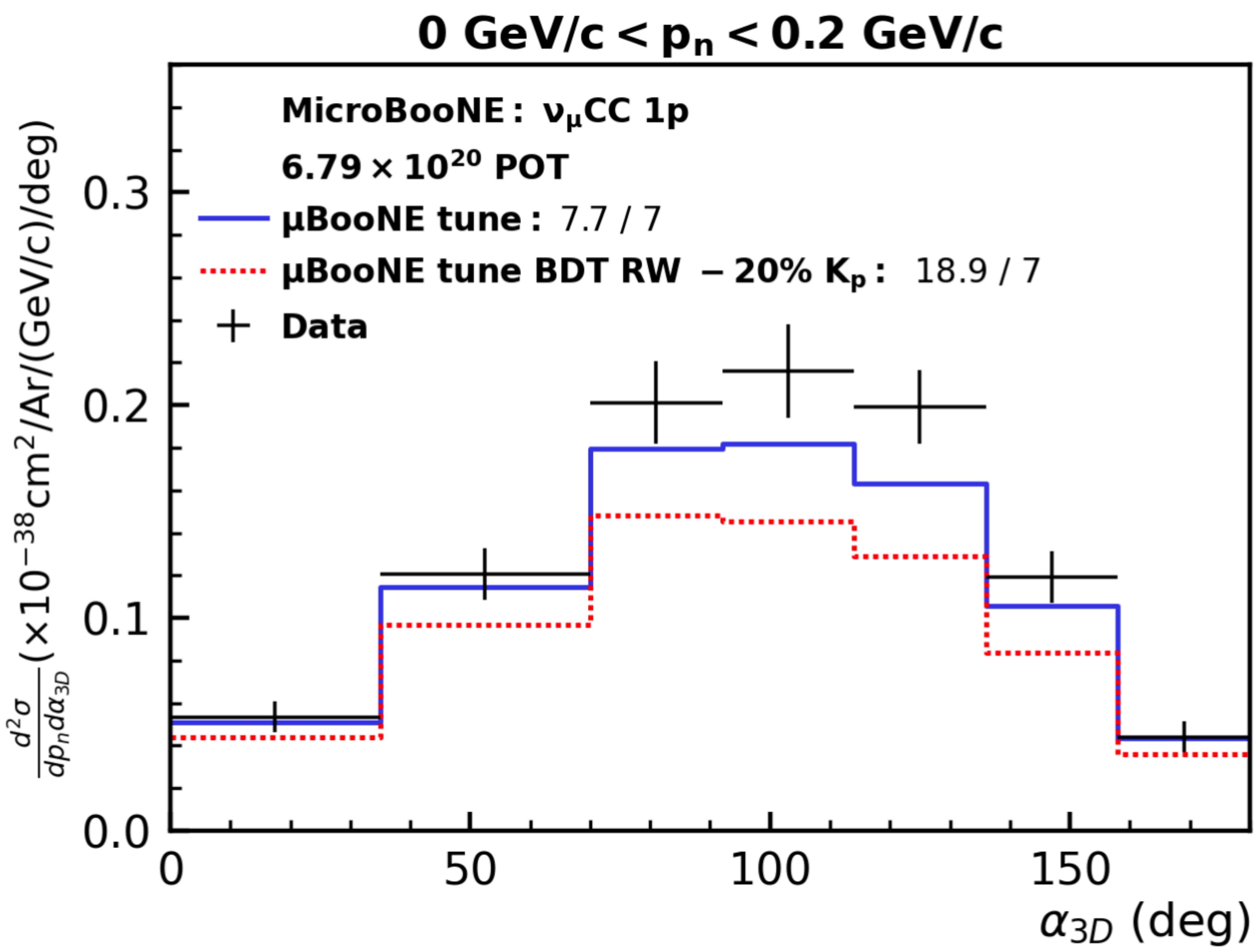
- The cross section for QE events as a function of Q^2 is among the most well understood parts of our cross section models
  - Constrained by theoretical modeling and electron scattering data
- Looking at this distribution, the BDT reweighting results in a cross section very far outside of theoretical uncertainties
- Such a large change that could cancel out the proton energy scaling seems implausible



arXiv:2411.03280

Recreating DUNE-ND Proton Energy Scaling Fake Data: Real Data GKI Measurements

- We can also see how this proton energy shifting and BDT reweighting would affect other types of measured neutrino cross sections
- Here, we examine how this would change our data-prediction consistency for a multidimensional generalized kinematic imbalance (GKI) distribution, [Phys. Rev. D 109, 092007 (2024)]
- In some of the phase space, it moves the prediction notably further from real data, again indicating that this energy shift + XS modification does not seem plausible



arXiv:2411.03280

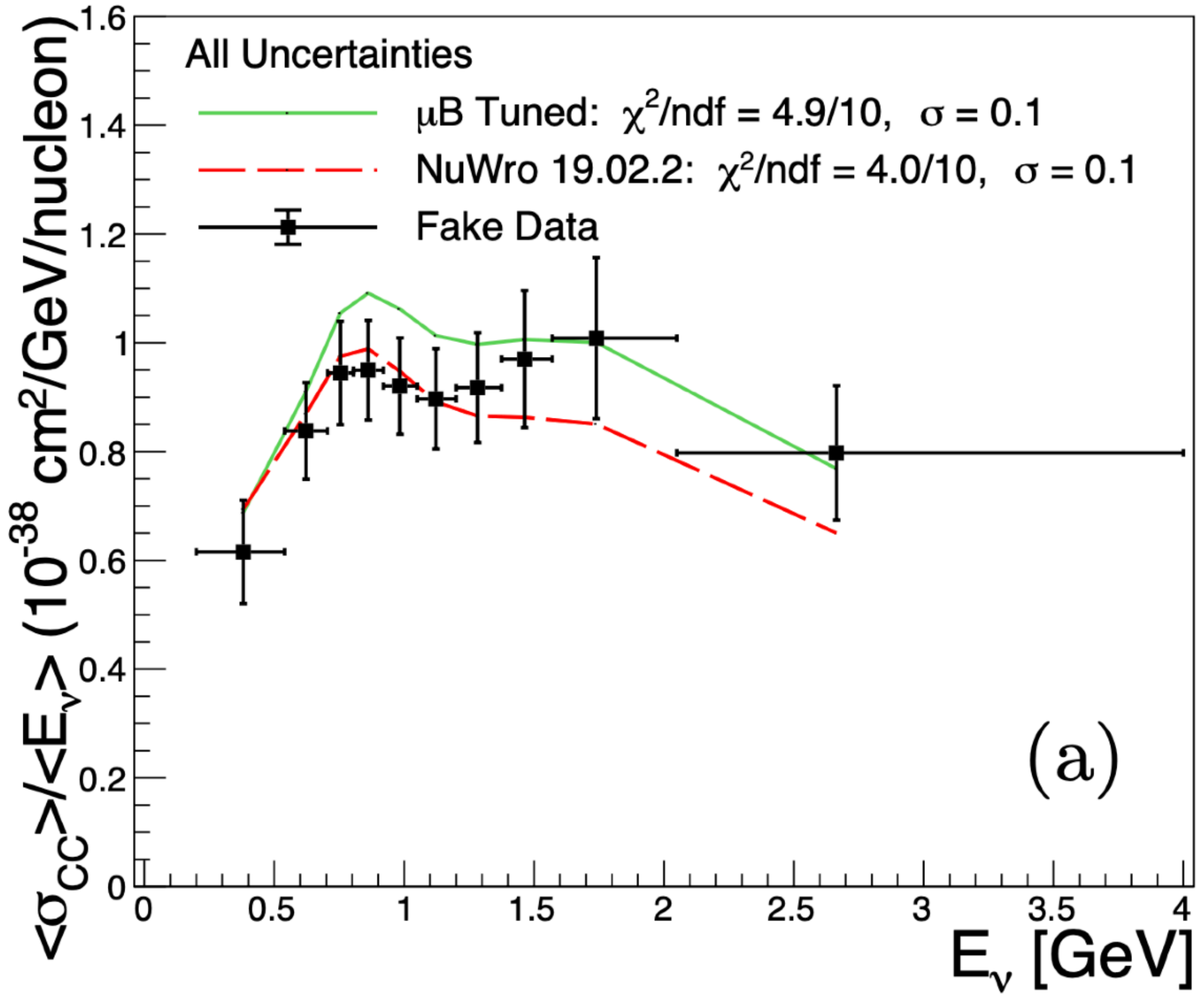
Recreating DUNE-ND Proton Energy Scaling Fake Data

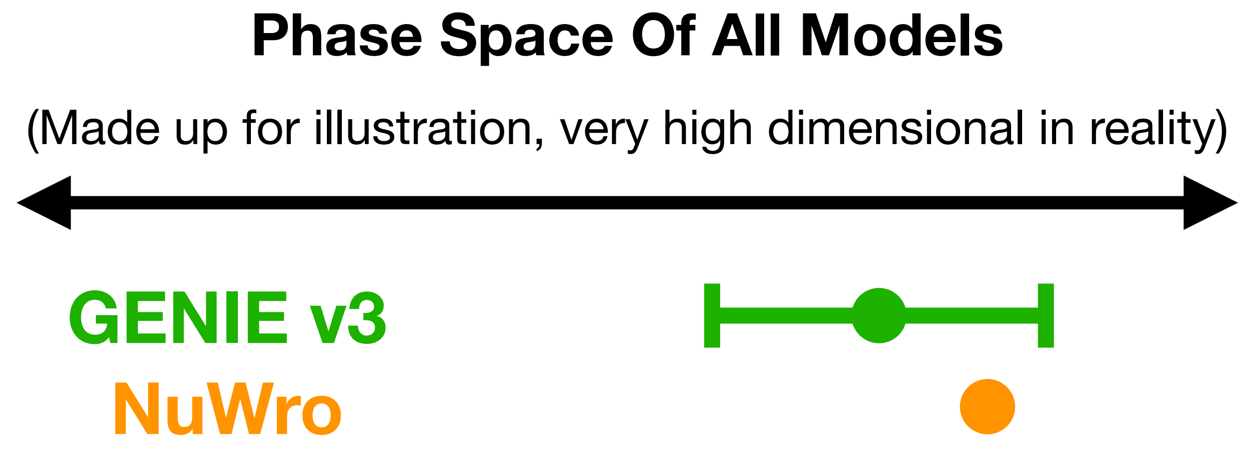
- This is one example of a case where different effects could conspire to make data-driven model validation fail
- In this specific case, there are signs that such a scenario would be inconsistent with other data
- We can't prove this in general for all possible model validation failures, but this is an illustration of the fact that there are a lot of constraints on our models, so it is not easy for them to be badly wrong in such specific ways

NuWro Fake Data: E_ν Cross Section Extraction
All Systematic Uncertainties

All Uncertainties				
Bias in Extraction			Model Validation $\cos\theta_\mu^{\mathrm{rec}} \vert E_\mu^{\mathrm{rec}}$	
E_ν	E_μ	ν	GoF	Decomposition
$\boxed{0.1}$	0.0	0.1	$\boxed{0.0}$	0.4

- Used NuWro 19.02.2 as fake data
- First, we include all systematic uncertainties
  - This is most accurately testing the sensitivity of the data-driven model validation procedure, which will use all uncertainties
  - We conclude that if this was real data, there would be no bias beyond uncertainties in a resulting cross section extraction



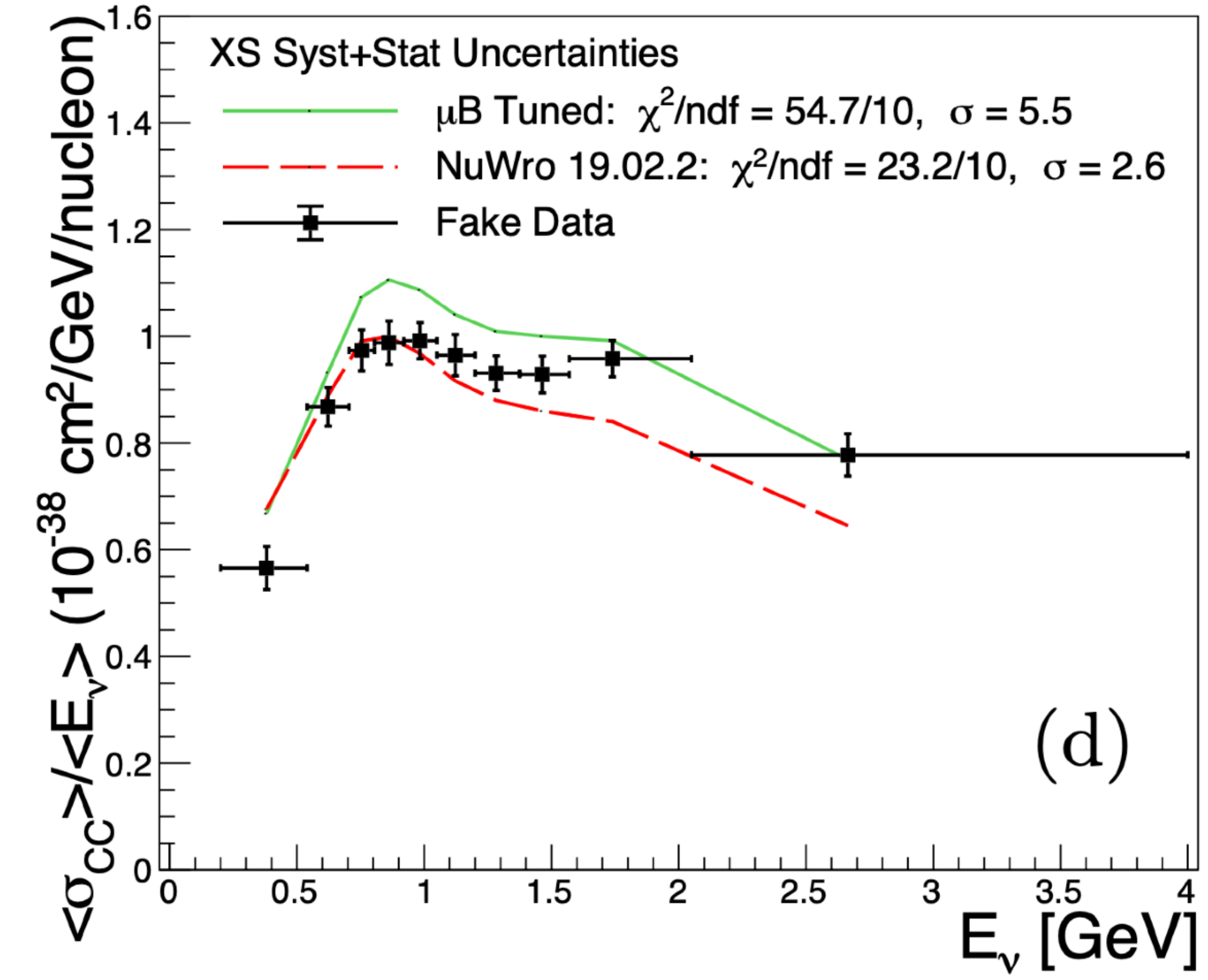


arXiv:2411.03280

NuWro Fake Data: E_ν Cross Section Extraction
Cross Section And Statistical Uncertainties

- Secondly, we included only cross section and statistical uncertainties
  - This is testing specifically the cross section model, since the flux and detector response modeling is identical between the prediction and the fake data
  - We conclude that NuWro is outside of our GENIE v3 MicroBooNE tune cross section model uncertainties
  - If this was real data, and if we had drastically smaller flux and detector uncertainties, we would fail model validation and therefore stop before extracting a biased cross section result

XS Syst+Stat Uncertainties				
Bias in Extraction			Model Validation $\cos\theta_\mu^{rec}\vert E_\mu^{rec}$	
E_ν	E_μ	ν	GoF	Decomposition
$\boxed{2.6}$	0.1	1.5	$\boxed{3.1}$	3.9



Phase Space Of All Models

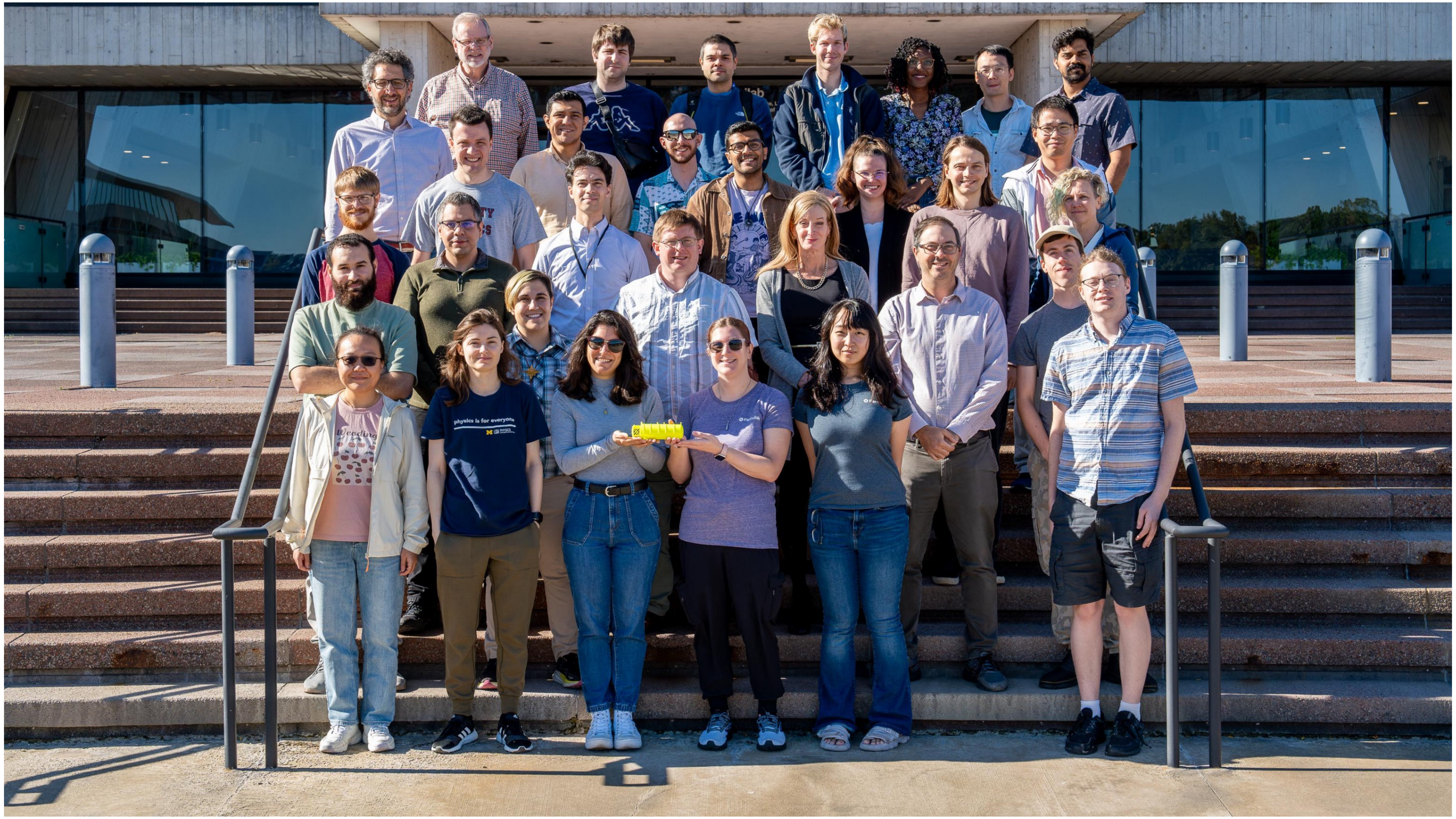
(Made up for illustration, very high dimensional in reality)

arXiv:2411.03280

Summary

- Model dependence is relevant precisely where we don't have quantitative estimates of bias from cross section models, so I think there's never going to be a simple exact answer for this issue
  - When we don't trust our cross section models, real data can be a very valuable resource
- Data-driven model validation is a useful tool for limiting model dependence in cross section results
  - Fake data studies have shown it to be sensitive to relevant mis-modeling in all cases we have examined
  - In a manufactured example, it is possible for a precise cancellation of multiple mis-modelings to cause a failure of the model validation procedure, but such a case seems likely to be excluded by other types of data
- This is not the only valid way to extract cross sections, and it has its own sets benefits and limitations relative to fake data closure testing. In MicroBooNE, we continue to use both techniques in different analyses.
- Our main goal for a discussion is to ensure a mutual understanding of the data-driven model validation technique

Thanks for your attention!



Backup Slides

All MicroBooNE Cross Section Results

- CC inclusive
  - ν_μCC neutron production, BNB, arXiv:2406.10583
  - 3D ν_μCC inclusive 0p/Np, BNB, Phys. Rev. Lett. 133, 041801 (2024), Phys. Rev. D 110, 013006 (2024)
  - 3D ν_μCC inclusive, BNB, arXiv:2307.06413
  - 1D ν_μCC inclusive E_ν, BNB, Phys. Rev. Lett. 128, 151801 (2022)
  - 1D ν_eCC inclusive, NuMI, Phys. Rev. D105, L051102 (2022)
  - One bin ν_eCC inclusive, NuMI, Phys. Rev. D104, 052002 (2021)
  - 2D ν_μCC inclusive, BNB, Phys. Rev. Lett. 123, 131801 (2019)
- Pion production
  - 2D NC π^0, BNB, arXiv:2404.10948
  - 1D CC π^0, BNB, arXiv:2404.09949
  - 1D NC π^0, BNB, Phys. Rev. D 107, 012004 (2023)
  - One bin CC π^0, BNB, Phys. Rev. D 99, 091102(R) (2019)
- Rare channels
  - η production, BNB, Phys. Rev. Lett. 132, 151801 (2024)
  - Λ production, NuMI, Phys. Rev. Lett. 130, 231802 (2023)
  - NC $\Delta \rightarrow N\gamma$ (interpreted as a limit on the cross section), BNB, Phys. Rev. Lett. 128, 111801 (2022)
- CC 0π
  - 2D ν_μ CC Np0π, BNB, arXiv:2403.19574
  - 1D & 2D ν_μ CC 1p0π Generalized Imbalance, BNB, Phys. Rev. D 109, 092007 (2024)
  - 1D & 2D ν_μCC 1p0π Transverse Imbalance, BNB, Phys. Rev. Lett. 131, 101802 (2023), Phys. Rev. D 108, 053002 (2023)
  - 1D ν_eCC Np0π, BNB, Phys. Rev. D 106, L051102 (2022)
  - 1D ν_μ CC 2p0π, BNB, arXiv:2211.03734
  - 1D ν_μ CC Np0π, BNB, Phys. Rev. D102, 112013 (2020)
  - 1D ν_μ CC 1p0π, BNB, Phys. Rev. Lett. 125, 201803 (2020)

How We Unfold To The Nominal Flux

Protons on target — **Number of targets** — **Beam flux** — **Cross section** — **Detector response** — **Selection efficiency** — **Background**

$$M(E_{rec}) = POT \cdot T \cdot \int F\left(E_\nu\right) \cdot \sigma\left(E_\nu\right) \cdot D\left(E_\nu \rightarrow E_{rec}\right) \cdot \varepsilon\left(E_\nu, E_{rec}\right) \cdot dE_\nu + B\left(E_{rec}\right)$$

All of these quantities must consider full flux, cross-section, detector, and statistical uncertainties!

Re-writing this same equation to be useful later (adding more terms that cancel each other out):

$$M(E_{rec}) = \underbrace{\frac{POT \cdot T \cdot \int_j F\left(E_{\nu\, j}\right) \cdot \sigma\left(E_{\nu\, j}\right) \cdot D\left(E_{\nu\, j} \rightarrow E_{rec\, i}\right) \cdot \varepsilon\left(E_{\nu\, j}, E_{rec\, i}\right) \cdot dE_{\nu\, j}}{\cancel{POT \cdot T \cdot \int_j \overline{F}\left(E_{\nu\, j}\right) \cdot \sigma\left(E_{\nu\, j}\right) \cdot dE_{\nu\, j}}}}_{\tilde{\Delta}_{ij}} \cdot \underbrace{\cancel{POT \cdot T} \cdot \cancel{\int_j \overline{F}\left(E_{\nu\, j}\right) \cdot dE_{\nu\, j}}}_{\tilde{F}_j} \cdot \underbrace{\frac{\cancel{\int_j \overline{F}\left(E_{\nu\, j}\right) \cdot \sigma\left(E_{\nu\, j}\right) \cdot dE_{\nu\, j}}}{\cancel{\int_j \overline{F}\left(E_{\nu\, j}\right) \cdot dE_{\nu\, j}}}}_{S_j} + B\left(E_{rec}\right)$$

$$M(E_{rec})_i = \tilde{\Delta}_{ij} \cdot \tilde{F}_j \cdot S_j + B\left(E_{rec}\right)_i$$

How We Unfold To The Nominal Flux

$$M(E_{rec})_i = \tilde{\Delta}_{ij} \cdot \tilde{F}_j \cdot S_j + B\left(E_{rec}\right)_i$$

$$\tilde{\Delta}_{ij} = \frac{POT \cdot T \cdot \int_j F\left(E_{\nu\ j}\right) \cdot \sigma\left(E_{\nu\ j}\right) \cdot D\left(E_{\nu\ j}, E_{rec\ i}\right) \cdot \varepsilon\left(E_{\nu\ j}, E_{rec\ i}\right) \cdot dE_{\nu\ j}}{POT \cdot T \cdot \int_j \overline{F}\left(E_{\nu\ j}\right) \cdot \sigma\left(E_{\nu\ j}\right) \cdot dE_{\nu\ j}}$$

Cross-section uncertainty largely (but not entirely) cancels

$$\tilde{F}_j = POT \cdot T \cdot \int_j \overline{F}\left(E_{\nu\ j}\right) \cdot dE_{\nu\ j}$$

Binned nominal flux

$$S_j = \frac{\int_j \overline{F}\left(E_{\nu\ j}\right) \cdot \sigma\left(E_{\nu\ j}\right) \cdot dE_{\nu\ j}}{\int_j \overline{F}\left(E_{\nu\ j}\right) \cdot dE_{\nu\ j}}$$

Nominal flux-binned cross-section signal

This is what we want to measure!