Preview of new techniques for future MicroBooNE cross-section data releases

Steven Gardiner

NuXTract Workshop

4 October 2023

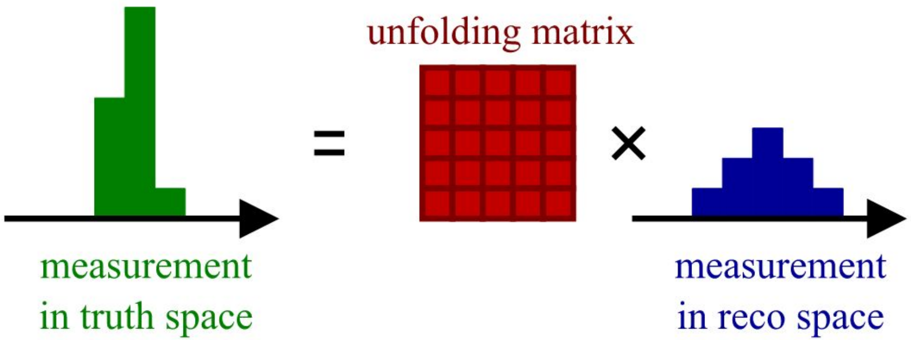
Introduction

- Present methods in other MicroBooNE talks, this is about the (near) future
  - Exploration ongoing in maturing cross-section analyses
- Content is abstract and mathematical, but with immediate applications
  - Easiest to apply with "traditional" unfolding (e.g., D'Agostini)
  - Techniques are generic, not specific to observables or interaction channels
- Three innovations for neutrino cross-section data releases
  - Analytic propagation of uncertainties, Generalized "A_C" matrix, Blockwise unfolding

Typical method (1)

$$\left\langle \frac{d\sigma}{dx} \right\rangle_\mu = \frac{\sum_j U_{\mu j}(D_j - B_j)}{\Phi\, T\, \Delta x_\mu}$$

- Flux-averaged differential cross section
  - true bins μ, reco bins j



- Unfolding matrix U accounts for inefficiency and smearing

- Unfolded space ≈ true space
  - Sometimes with unfolding-related systematics

$$U_{\mu j} = \frac{P(\mu|j)}{\epsilon_\mu}$$

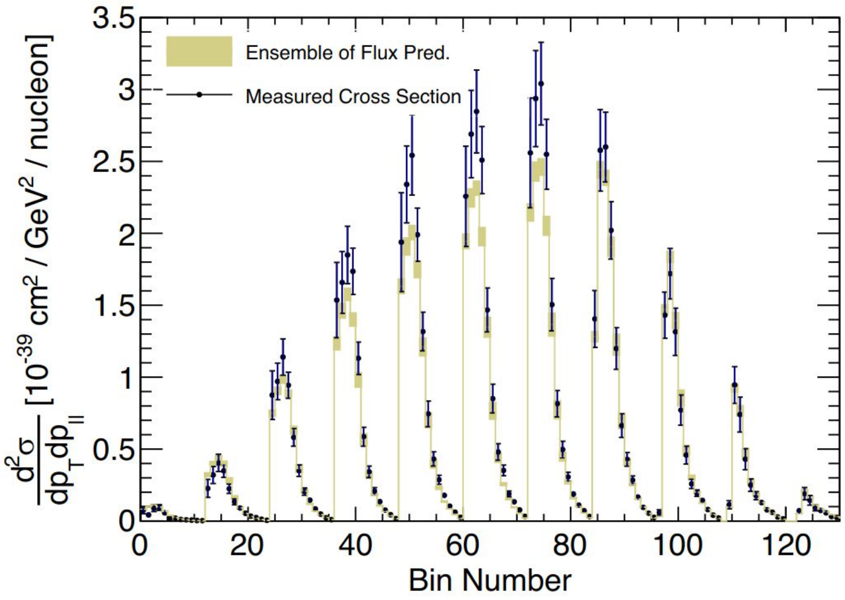
Typical method (2)

$$\left\langle \frac{d\sigma}{dx} \right\rangle_\mu = \frac{\sum_j U_{\mu j}(D_j - B_j)}{\Phi\, T\, \Delta x_\mu}$$

- U built using D'Agostini procedure

- U, B, Φ, and T varied, unfolding repeated in each systematic universe
  - Spread of results used for covariances
  - Systematics and unfolding intertwined
- Koch + Dolan "first approach"
  - Measurement in *real* flux
  - Flux shape uncertainties should be applied to models in comparisons

Phys. Rev. D 102, 113012 (2020)



Recent MicroBooNE analyses make different choices (1)

- U built using Wiener-SVD recipe
  - Relatively minor adjustment
- Uncertainties applied to expected reco events $n_j \approx D_j$
  - Extraction proceeds with *fixed* parameter central values
  - Covariances propagated analytically through unfolding
- MC expectation must be a reasonably good estimate for the data
  - Model validation procedures, etc.

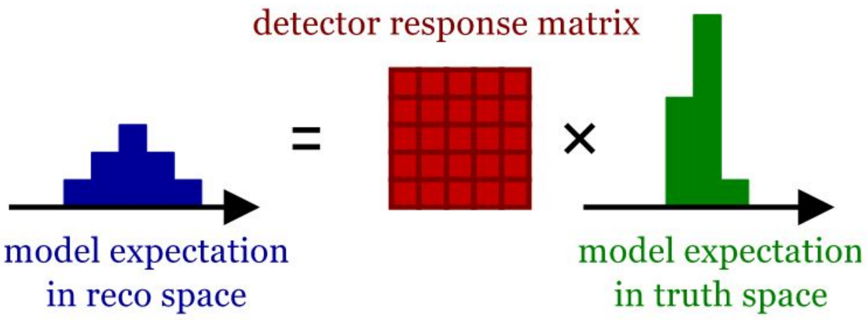
$$\left\langle \frac{d\sigma}{dx} \right\rangle_\mu = \frac{\sum_j U_{\mu j}(D_j - B_j)}{\Phi\, T\, \Delta x_\mu}$$

$$\mathrm{Cov}(D_j, D_k) \approx \mathrm{Cov}(n_j, n_k)$$

$$= \frac{1}{N_{\mathrm{univ}}} \sum_{u=1}^{N_{\mathrm{univ}}} (n_j^{\mathrm{CV}} - n_j^u)(n_k^{\mathrm{CV}} - n_k^u)$$

Recent MicroBooNE analyses make different choices (2)

- Detector response matrix Δ
  - Inverse of unfolding matrix (roughly speaking)
- Signal model uncertainties
  - Flux variations applied directly
  - Cross-section variations through Δ only
- Koch + Dolan "second approach"
  - Measurement in *reference* flux, xsec models directly comparable
  - Assumes CV MC energy dependence is sufficient to assess flux systematics



$$n_j = \sum_\mu \Delta_{j\mu}\, \phi_\mu^{\mathrm{CV}} + B_j$$

Dj now an estimator for the number of reco events MicroBooNE ***would have measured*** in the reference flux (despite same central value)

Some benefits of this treatment

- Full uncertainty assigned to quantity actually measured (selected events)
  - Extraction procedure becomes a linear transformation of the data

- Simpler implementation: no systematic throws needed during unfolding

- Data release can allow unfolding specifics to be revisited
  - Report reco-space data event counts, covariances, Δ matrix, and central-value scaling factors (Φ, T) for unit conversions
  - Remaining procedure can be executed by people external to the experiment

Uncertainty propagation

- Error propagation matrix $\mathfrak{E}$
  - Connects reco-space event uncertainties to final result
  - Identical to unfolding matrix U when there is no direct dependence on the data (Wiener-SVD)
  - Known solution for D'Agostini
- Overall scaling factors are just constants in MicroBooNE approach

$$\mathfrak{E}_{\mu j} \equiv \frac{\partial \hat{\phi}_\mu}{\partial d_j}$$

$\hat{\phi}_\mu$ **Unfolded signal event count in true bin μ**

$$d_j = D_j - B_j$$

Background-subtracted measured event count in reco bin j

$$\mathfrak{E}^{i+1}_{\mu j} = \frac{\partial \hat{\phi}^{i+1}_\mu}{\partial d_j} = U^i_{\mu j} + \frac{\hat{\phi}^{i+1}_\mu}{\hat{\phi}^i_\mu}\,\mathfrak{E}^i_{\mu j} - \sum_{\nu,k} \epsilon_\nu \, \frac{d_k}{\hat{\phi}^i_\nu}\, U^i_{\mu k}\, U^i_{\nu k}\, \mathfrak{E}^i_{\nu j}$$

i-th D'Agostini unfolding iteration

$$\mathrm{Cov}\left(\left\langle \frac{d\sigma}{dx} \right\rangle_\mu, \left\langle \frac{d\sigma}{dy} \right\rangle_\nu\right) = \frac{1}{\Phi^2\, T^2\, \Delta x_\mu\, \Delta y_\nu} \sum_{j,k} \mathfrak{E}_{\mu j}\, \mathrm{Cov}(D_j, D_k)\, \mathfrak{E}^T_{k\nu}$$

Regularization and the "A_C" matrix

- Naive unfolding: directly invert the response matrix
  - Strong anticorrelations, large uncertainties

$$U^{\mathrm{naive}} = (\Delta^T \Delta)^{-1} \Delta^T$$

- Regularization: add prior information as a fix
  - Add (hopefully) small bias for smaller variance, unfolding methods differ on recipe
- Additional smearing matrix A_C

$$U = A_C \cdot U^{\mathrm{naive}} = A_C \cdot (\Delta^T \Delta)^{-1} \Delta^T$$

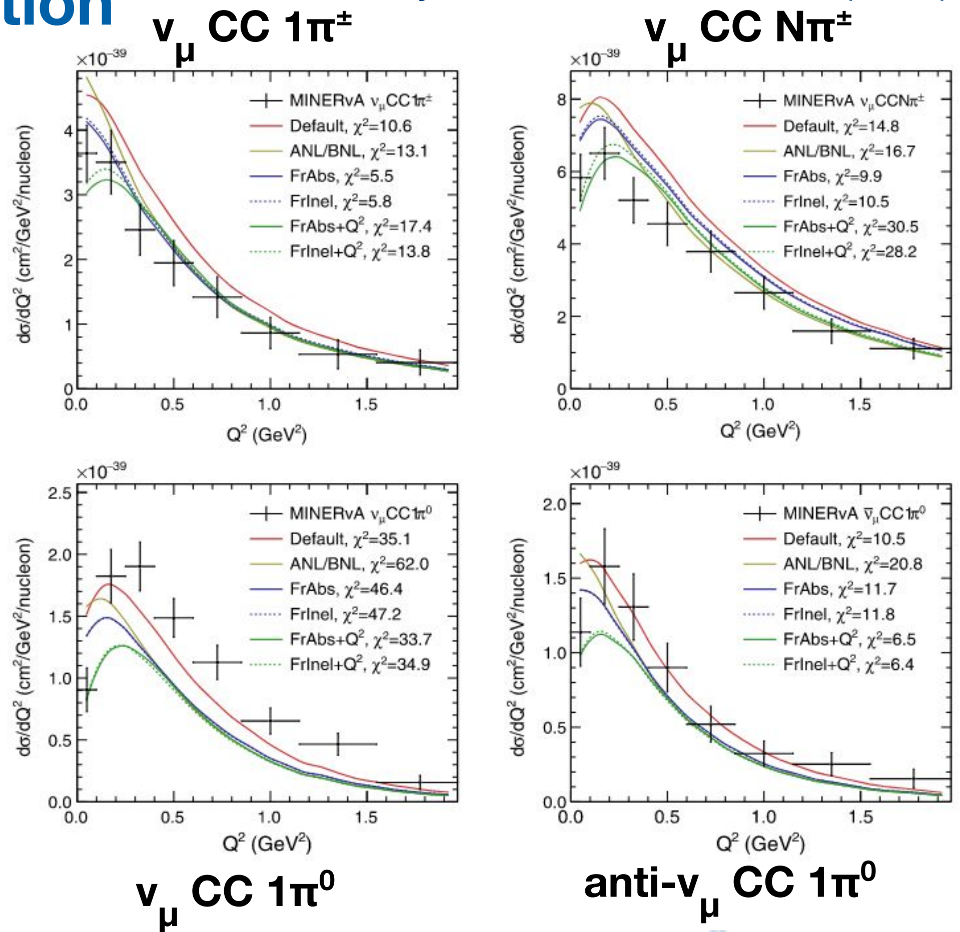
  - Introduced in Wiener-SVD method:
    W. Tang *et al.*, J. Instrum., 12, P10002 (2017)
  - Apply consistent regularization to theory models → avoids need for extra unfolding-related uncertainty
  - **Calculable for an arbitrary unfolding matrix U**

$$A_C = U \cdot \Delta$$

"Blockwise unfolding": motivation

Phys. Rev D. 100, 072005 (2019)

- Experiments report multiple kinematic distributions
  - Same analysis or complementary ones
- **Correlated uncertainties** between distributions are not typically reported
  - Important! (e.g., flux)
- Limitations discussed in MINERvA paper tuning GENIE to π production data



"Blockwise unfolding": motivation

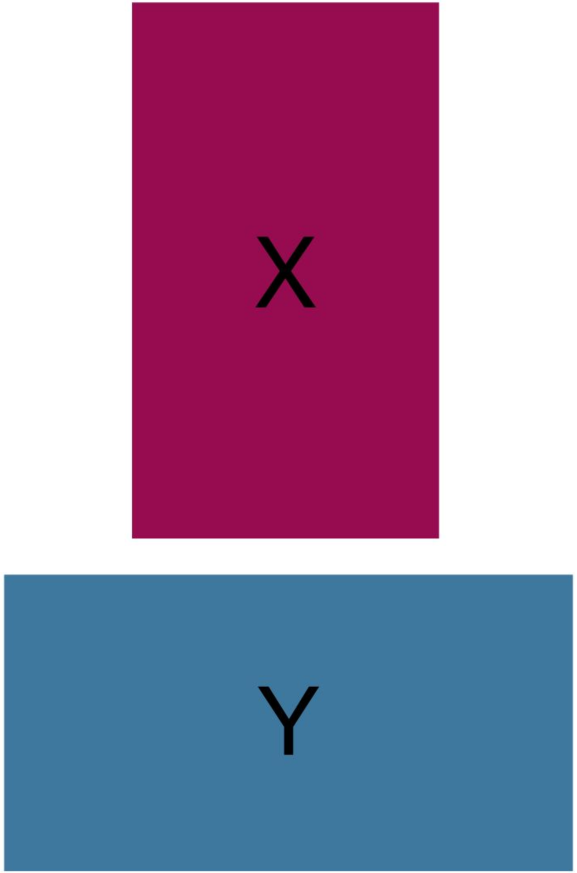
Phys. Rev D. 100, 072005 (2019)

The published cross sections are one dimensional with correlations provided between the bins within each distribution. No correlations are provided between measurements of different final states, or between different one-dimensional projections of the same measurement. These correlations are expected to be large, coming predominantly from flux and detector uncertainties. Additionally, the $\nu_\mu \mathrm{CC}1\pi^\pm$ event sample is a subset (~64%) of the $\nu_\mu \mathrm{CC}N\pi^\pm$ event sample, and including both channels introduces a statistical correlation. Not assessing correlations between the distributions, while a common practice in this field, is a limitation when tuning models to multiple datasets. It introduces a bias in the χ^2 statistic that is difficult to quantify, and requires imposing *ad hoc* uncertainties [4] as the test statistic is not expected to follow a χ^2 distribution for the given degrees of freedom.

- Not trivial to add this information after the fact
- Correlations calculable with suitable planning ahead
- Two issues
  - Event overlaps (*statistical* covariances)
  - Unfolding treatment

Statistical covariances (1)

- Events belong to multiple bins ⇒ correlated stat uncertainties
- Easily calculable if the problem is framed properly
- Arbitrary bins X and Y
- Event count n_X in bin X follows a Poisson distribution
  - Estimator for the mean: n_X
  - Estimator for the variance: n_X
- Bin Y is similar. How to get the covariance?

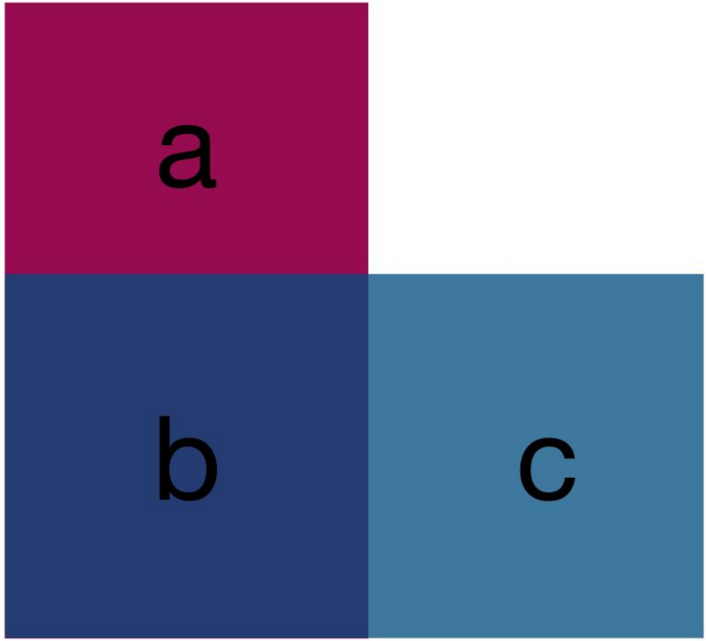


Statistical covariances (2)

- The trick: one may always rebin 2 → 3
- Bins a, b, and c are **non-overlapping**
  - Independent Poisson distributions

$$\begin{aligned}\text{cov}(X, Y) &= \text{cov}(a + b, b + c) \\ &= \text{cov}(a, b) + \text{cov}(a, c) + \text{cov}(b, b) + \text{cov}(b, c) \\ &= 0 + 0 + \text{var}(b) + 0 \\ &\approx n_b\end{aligned}$$

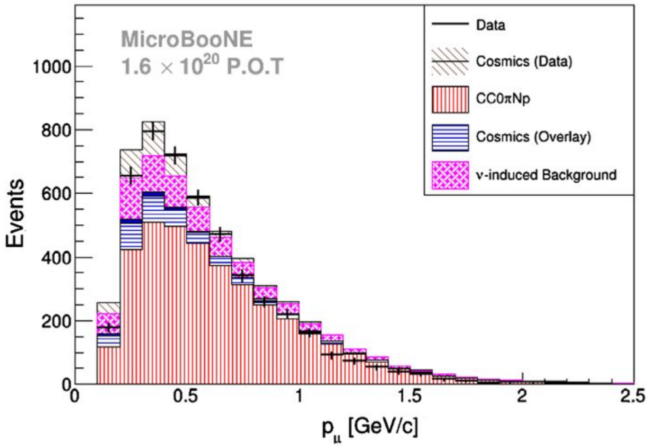
- Estimator for statistical covariance is just the **number of events that bins X and Y have in common**



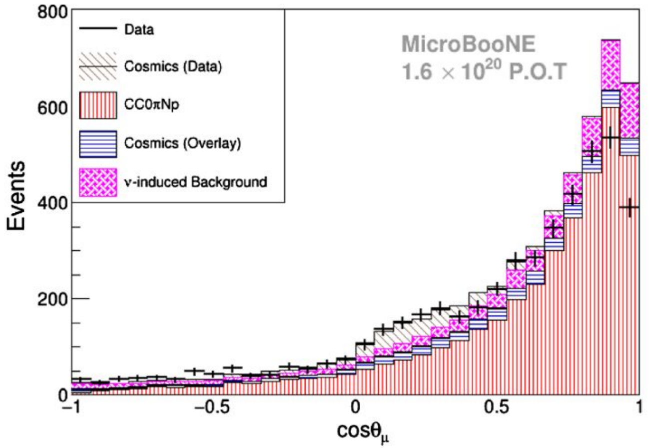
Note that this behaves as expected for X = Y as well as disjoint bins

Organizing blocks of bins

- Group bins belonging to the same kinematic distribution in a "block"
  - An event should belong to a maximum of one reco bin and one true bin in each block → avoids double-counting
- Observables can be abstracted away by working in "bin number space"
  - Trivially generalizes to 2D, 3D, etc.
- Example:
  - Bins 0-19 represent p_μ → block #0
  - Bins 20-49 represent $\cos\theta_\mu$ → block #1



Phys. Rev. D 102, 112013 (2020)



Blockwise unfolding procedure

- Build an unfolding matrix U_b for the b-th block according to one's preferred approach
- Overall unfolding matrix U is block-diagonal
  - Results for individual blocks are the same as for a stand-alone measurement
- Error propagation matrix $\mathfrak{E}$ and additional smearing matrix A_C
  - Block-diagonal, calculable from the U_b
- **Covariances between all bins** in all measured distributions can thus be

$$U = \bigoplus_{b=0} U_b = U_0 \oplus U_1 \oplus \cdots = \begin{pmatrix} U_0 & 0 & 0 & \ldots \\ 0 & U_1 & 0 & \ldots \\ 0 & 0 & \ddots & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$$\mathfrak{E} = \bigoplus_{b=0} \mathfrak{E}_b = \mathfrak{E}_0 \oplus \mathfrak{E}_1 \oplus \ldots$$

Fermilab

Outlook for blockwise unfolding

- Theorists and generator developers can fit to all measured distributions simultaneously
  - Increases discrimination power of the data: can the model describe the correlations as well as each individual block?
  - No need for ad hoc estimates of flux uncertainty, etc. All systematics come from the experiment itself

- Potential for *inter-analysis* covariances
  - Bookkeeping for event overlaps (statistical uncertainties)
  - Consistent systematic universes
- MicroBooNE analyses in progress aim to report model goodness-of-fit χ^2 over hundreds of bins in this way

Summary

- Recent MicroBooNE papers extract cross sections with a distinct strategy
  - Subtleties in systematics treatment important
- Enables innovations in data releases
- Analytic error propagation
  - Potential to revisit unfolding choices after the fact
- Generalized calculation of A_C matrix
  - Quantify regularization in D'Agostini using same approach as Wiener-SVD
- Blockwise unfolding
  - Maximize usefulness of data by reporting full set of covariances

Backup