Publishing Cross Sections at MINERvA

Noah Harvey Vaughan *(they/them)*
PhD Candidate, Oregon State University

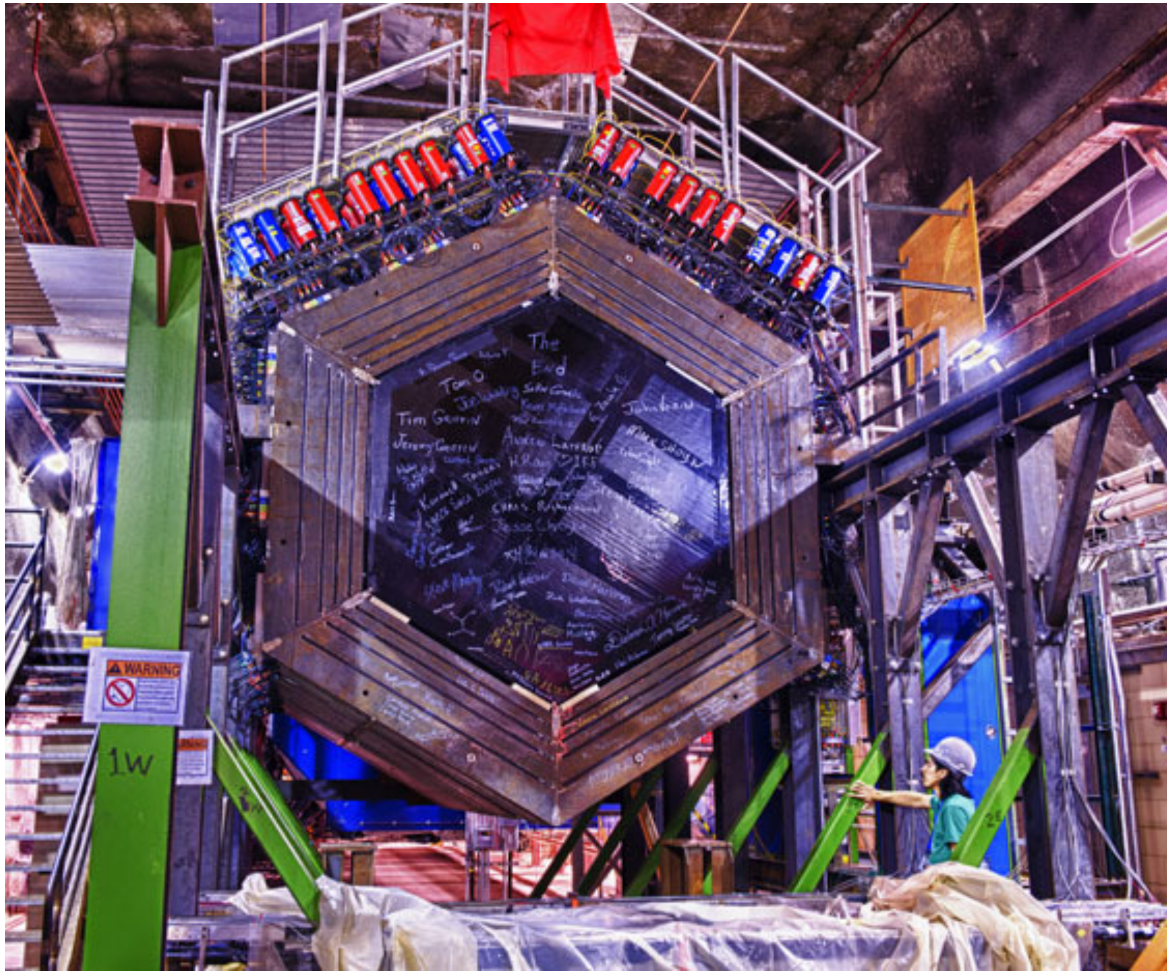
NuXTract 2023

Wednesday October 6th, 2023

Oregon State University

Outline

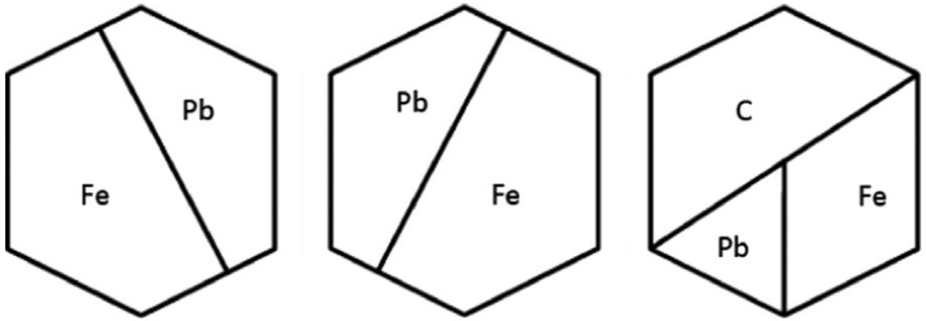
- Overview of MINERvA
- MC tuning
- Background constraint
- Unfolding testing and process



MINERvA Experiment

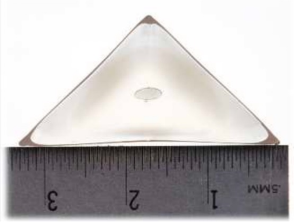
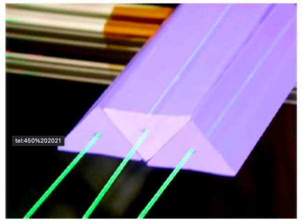
MINERvA Experiment

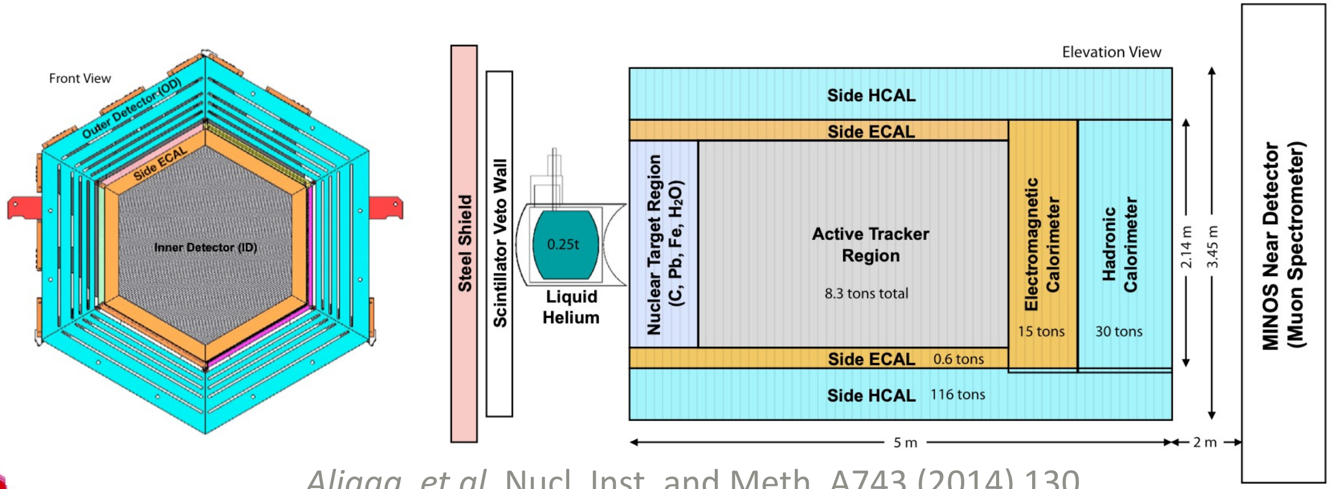
Geometry of nuclear target planes



MINERvA = **M**ain **IN**jector
Expe**R**iment on **v** (nu) **A** (atom)

- Dedicated x-section experiment at FNAL
- Data runs from 2009 – 2019:
  - NuMI LE POT: $4.0 \times 10^{20}\ \nu$, $1.7 \times 10^{20}\ \bar{\nu}$
  - NuMI ME POT: $\sim 3 \times \text{LE}\ \nu$, $\sim 7 \times \text{LE}\ \bar{\nu}$

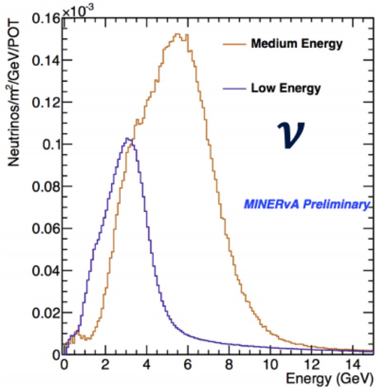


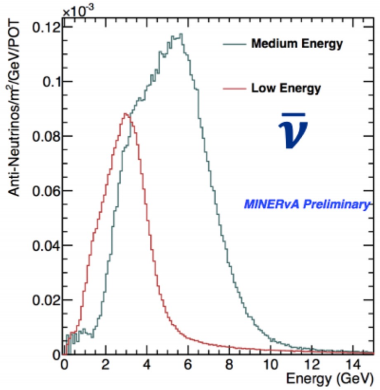
Aliaga, et al. Nucl. Inst. and Meth. A743 (2014) 130

Cross section publication history

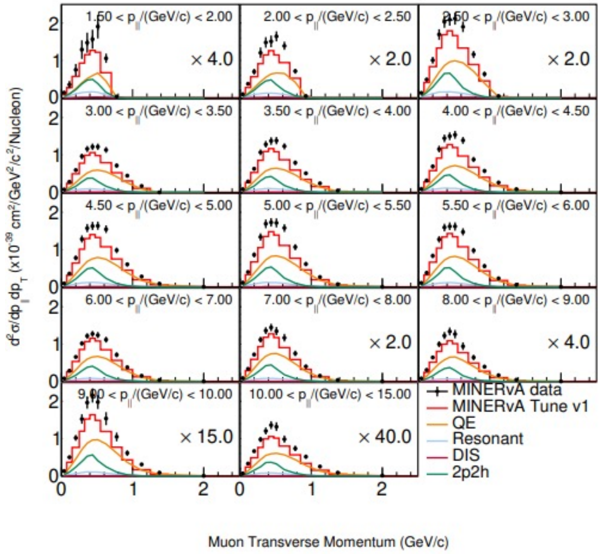
MINERvA Experiment

- LE: 32 published; ME: 12 published, more in progress
- Processes studied:
  - QElike, π production, 2p2h
  - DIS
  - Coherent
- Comparison of targets

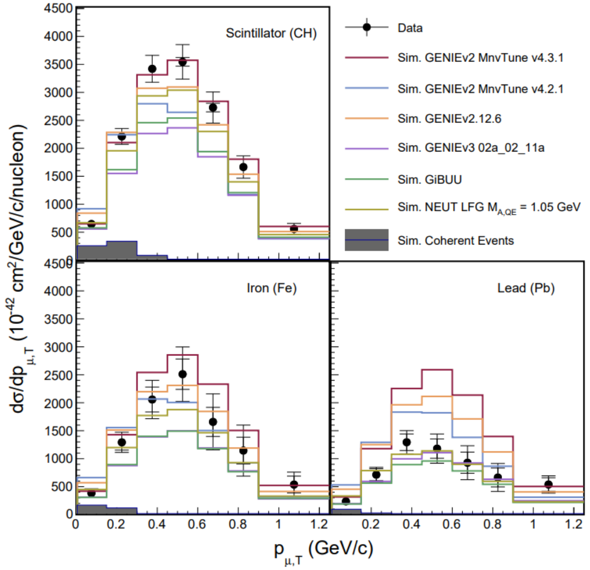




$\overline{\nu_\mu}$ CC0π on CH



ν_μ CC1π^+ on Targets



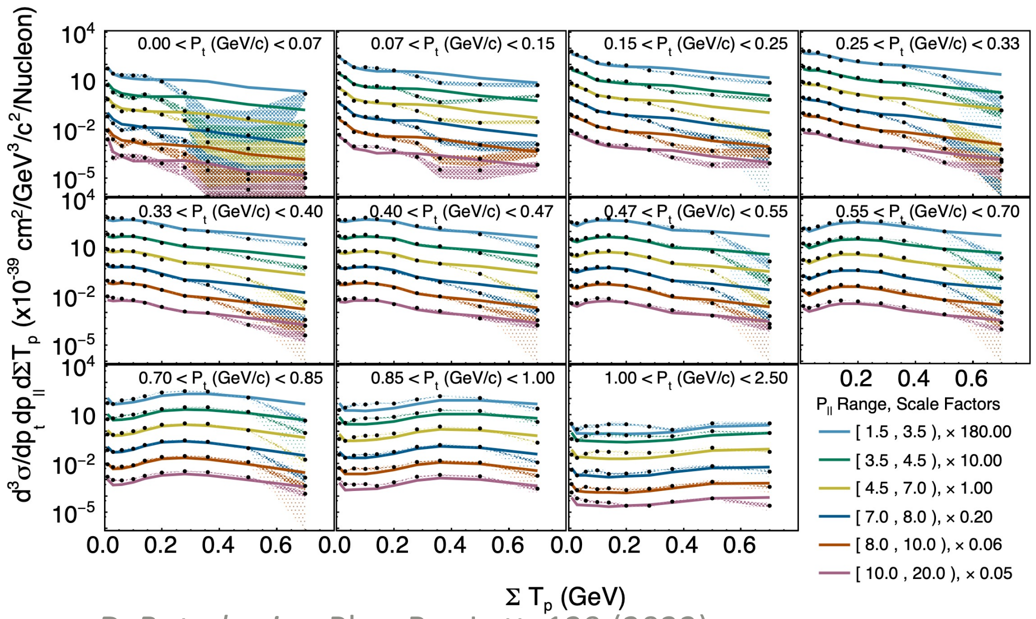
A. Bashyal, Phys.Rev.D 108, 032018 (2023)

A. Bercelle, PRL 131, 011801 (2023)

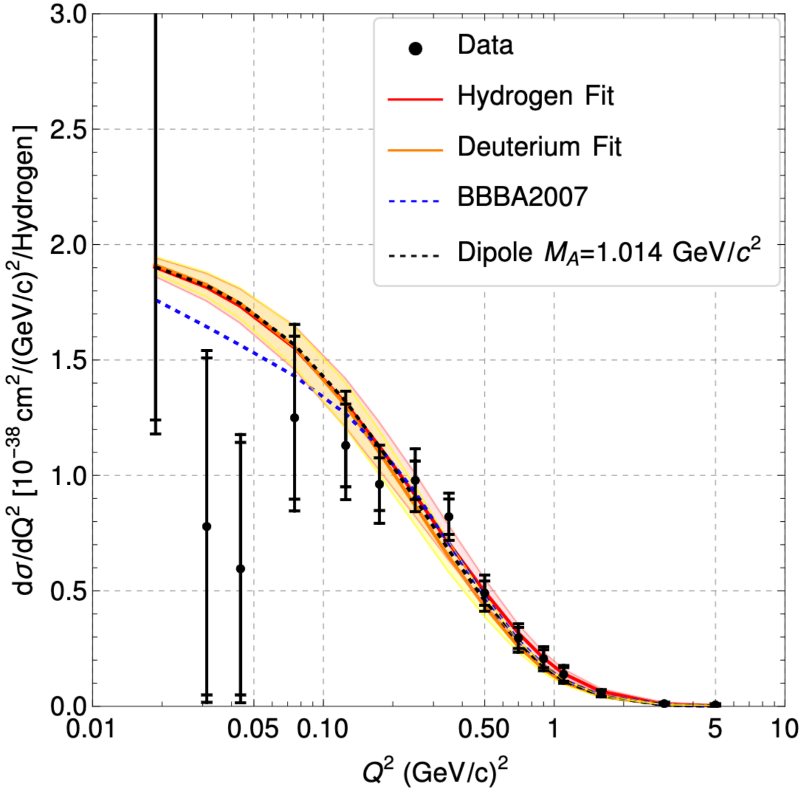
Cross section publication history

More difficult stuff too…

- 1.3 M events in first <u>3D</u> ν_μ *CCQElike*
- Highly exclusive (~5k events) Hydrogen x-sec



D. Ruterbories, Phys.Rev.Lett. 129 (2022)



T. Cai, Nature **614**, 48-53 (2023)

Extracting a Cross-section

Steps analyzers take for most cross-sections at MINERvA

Extraction process

Somewhat simplistically, analyzers start by:

- Choosing a process(es) they want to study (inclusively or exclusively)
- Creating a truth level signal definition of this process
- Choosing reconstruction cuts to select events in the data, reconstructed MC
- Cuts and sig def chosen to maintain compatibility with preexisting measurements

Extraction process

Somewhat simplistically, analyzers start by:

- Choosing a process(es) they want to study (inclusively or exclusively)
- Creating a truth level signal definition of this process
- Choosing reconstruction cuts to select events in the data, reconstructed MC
- Cuts and sig def chosen to maintain compatibility with preexisting measurements

For variable x in truth bins i with reconstruction bins α:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}x}\right)_i = \frac{\sum_\alpha U_i^\alpha \left(N_{data,\alpha} - N_{data,\alpha}^{bkg}\right)}{\epsilon_i (\Phi T)(\Delta x_i)}$$

Higher Dimensions

2D: For variables x, y, in truth bins i, j with reconstruction bins α, β:

$$\left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}x\,\mathrm{d}y}\right)_{ij} = \frac{\sum_{\alpha\beta} U_{ij}^{\alpha\beta}\left(N_{data,\,\alpha\beta} - N_{data,\,\alpha\beta}^{bkg}\right)}{\epsilon_{ij}(\Phi T)\left(\Delta x_i \Delta y_j\right)}$$

3D: For variables x, y, z in truth bins i, j, k with reconstruction bins α, β, γ:

$$\left(\frac{\mathrm{d}^3\sigma}{\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z}\right)_{ijk} = \frac{\sum_{\alpha\beta\gamma} U_{ijk}^{\alpha\beta\gamma}\left(N_{data,\,\alpha\beta\gamma} - N_{data,\,\alpha\beta\gamma}^{bkg}\right)}{\epsilon_{ijk}(\Phi T)\left(\Delta x_i \Delta y_j \Delta z_k\right)}$$

Ingredients

For variable x in truth bins i with reconstruction bins α:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}x}\right)_i = \frac{\sum_\alpha U_i^\alpha \left(N_{data,\,\alpha} - N_{data,\alpha}^{bkg}\right)}{\epsilon_i (\Phi T)(\Delta x_i)}$$

Background Subtraction

Ingredients

For variable x in truth bins i with reconstruction bins α:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}x}\right)_i = \frac{\sum_\alpha U_i^\alpha \left(N_{data,\alpha} - N_{data,\alpha}^{bkg}\right)}{\epsilon_i (\Phi T)(\Delta x_i)}$$

Background Subtraction

Unfolding

Ingredients

For variable x in truth bins i with reconstruction bins α:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}x}\right)_i = \frac{\sum_\alpha U_i^\alpha \left(N_{data,\,\alpha} - N_{data,\alpha}^{bkg}\right)}{\epsilon_i (\Phi T)(\Delta x_i)}$$

Background Subtraction

Unfolding

Efficiency & Acceptance Correction

Ingredients

For variable x in truth bins i with reconstruction bins α:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}x}\right)_i = \frac{\sum_\alpha U_i^\alpha \left(N_{data,\,\alpha} - N_{data,\alpha}^{bkg}\right)}{\epsilon_i (\Phi T)(\Delta x_i)}$$

Background Subtraction

Unfolding

Efficiency & Acceptance Correction

Flux & Target Normalization

Ingredients

For variable x in truth bins i with reconstruction bins α:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}x}\right)_i = \frac{\sum_\alpha U_i^\alpha \left(N_{data,\,\alpha} - N_{data,\alpha}^{bkg}\right)}{\epsilon_i (\Phi T)(\Delta x_i)}$$

Background Subtraction

Unfolding

Efficiency & Acceptance Correction

Flux & Target Normalization

Bin Width Normalization

Ingredients

For variable x in truth bins i with reconstruction bins α:

$$\left(\frac{d\sigma}{dx}\right)_i = \frac{\sum_\alpha U_i^\alpha \left(N_{data,\alpha} - N_{data,\alpha}^{bkg}\right)}{\epsilon_i (\Phi T)(\Delta x_i)}$$

Background Subtraction

Unfolding

Efficiency & Acceptance Correction

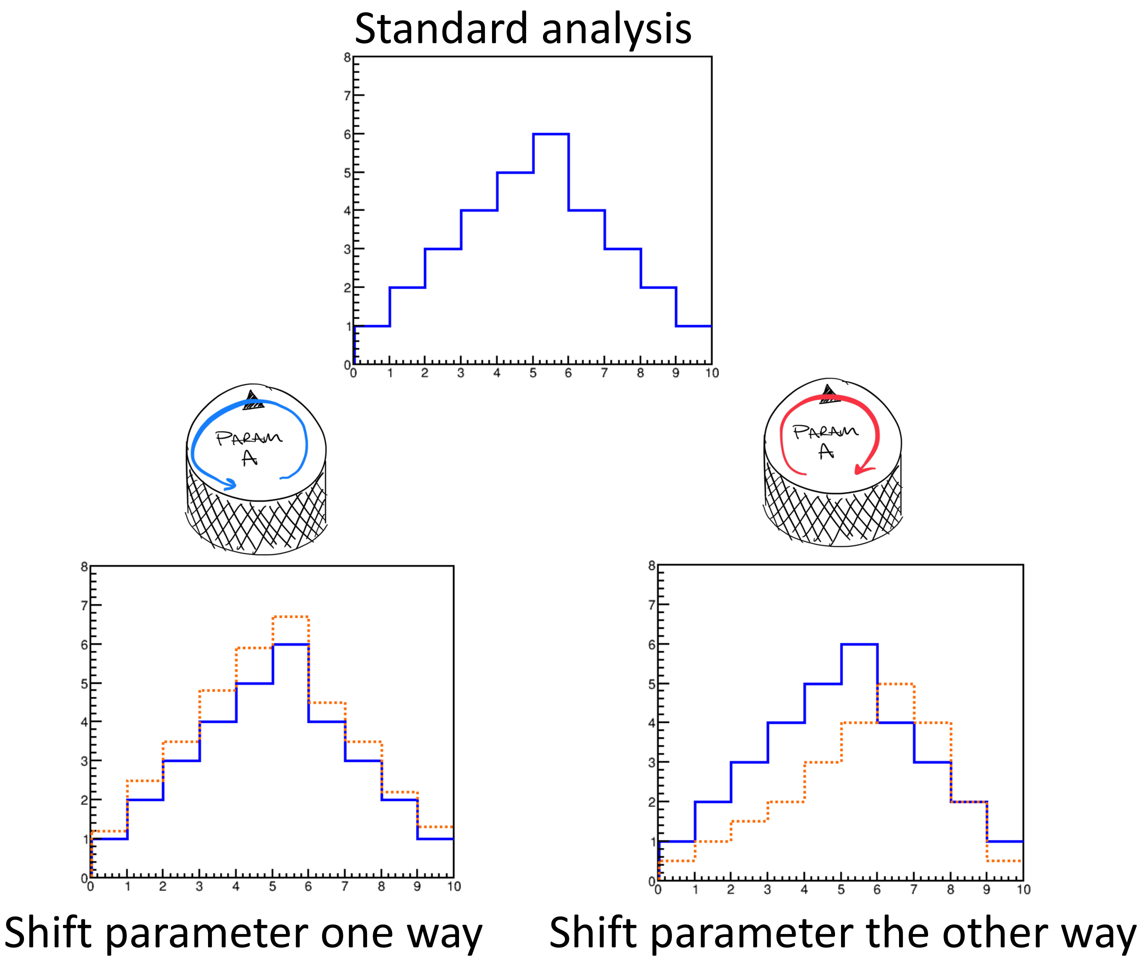
Flux & Target Normalization

Bin Width Normalization

Note on MINERvA Systematics

MINERvA handles systematics through `MnvHND`'s, inherited `THND`

- Central value universe: `CVUniverse`, stored as `ROOT` histogram
- *Systematic universes* (error bands): shifts and migrations on `CVUniverse`
- Typically, a $\pm 1\sigma$ shift on some parameter
- More detail on data preservation talk



C. Patrick, Fermilab JETP, 2016

MINERvA Models and Tunes

Brief overview of empirical and physics tunes on MINERvA simulation

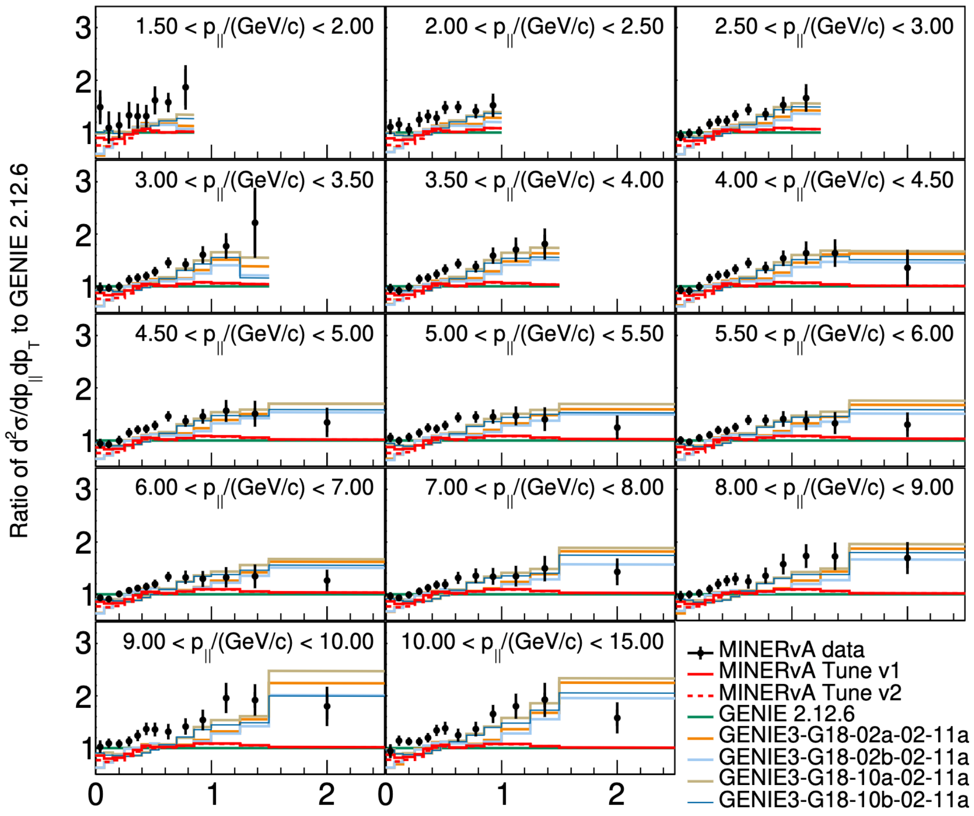
Making and using MnvTunes

MINERvA creates tunes on simulation from measurements

- Base simulation GENIE 2.12.6
- Theory/physics based if not included in generator
- Based off MINERvA measurements and other experiments

Using MnvTunes

- Applied as weights, event-by-event
- Analyzers choose appropriate tune for unfolding, analysis



A. Bashyal, Phys.Rev.D 108 (2023)

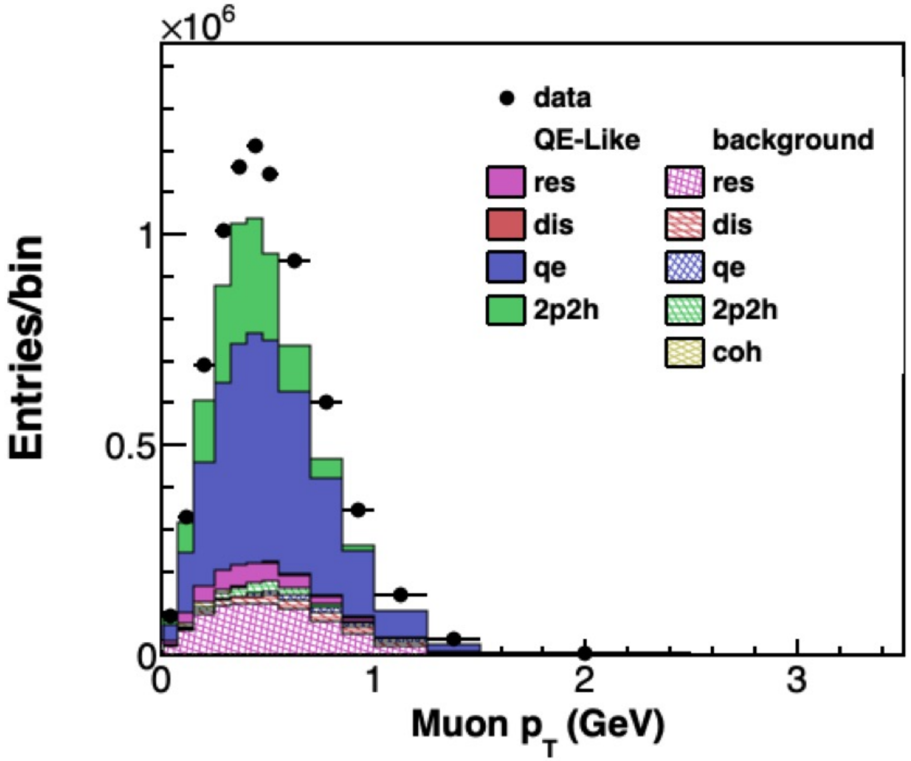
Constraining the Background

How analyzers constrain the background using the simulated sample

General idea

Selected data contaminated with background events

- Can use simulation (with *truth* info) to inform removal
- Directly subtracting MC background introduces model dependence
- Use data driven methods to constrain the background instead

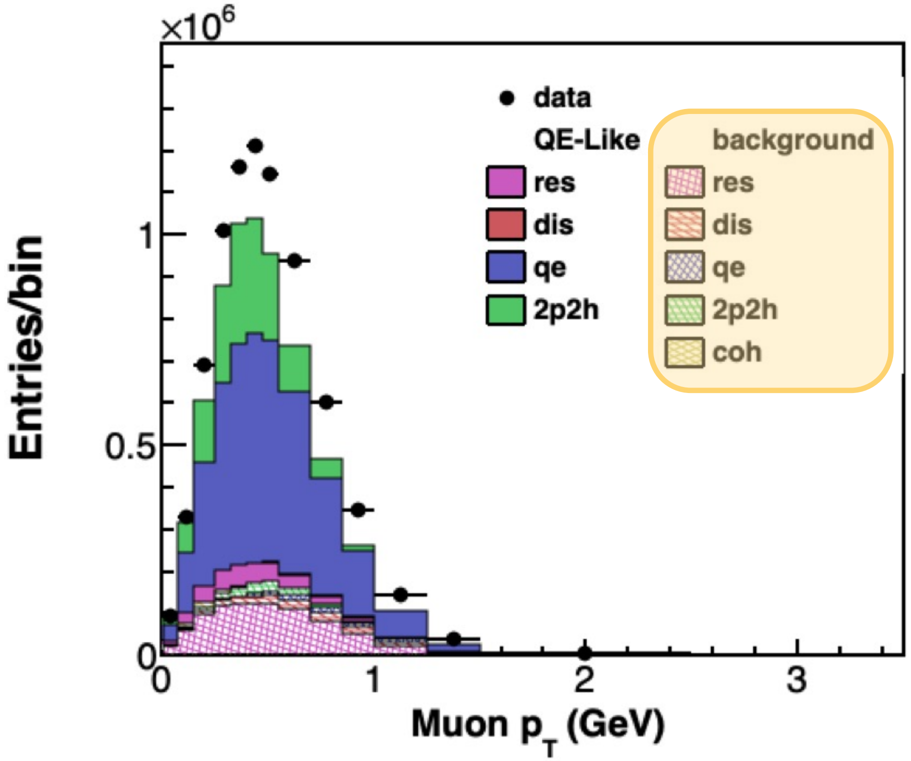


A. Bashyal, Fermilab JETP, 10 March 2023

General idea

Selected data contaminated with background events

- Can use simulation (with *truth* info) to inform removal
- Directly subtracting MC background introduces model dependence
- Use data driven methods to constrain the background instead



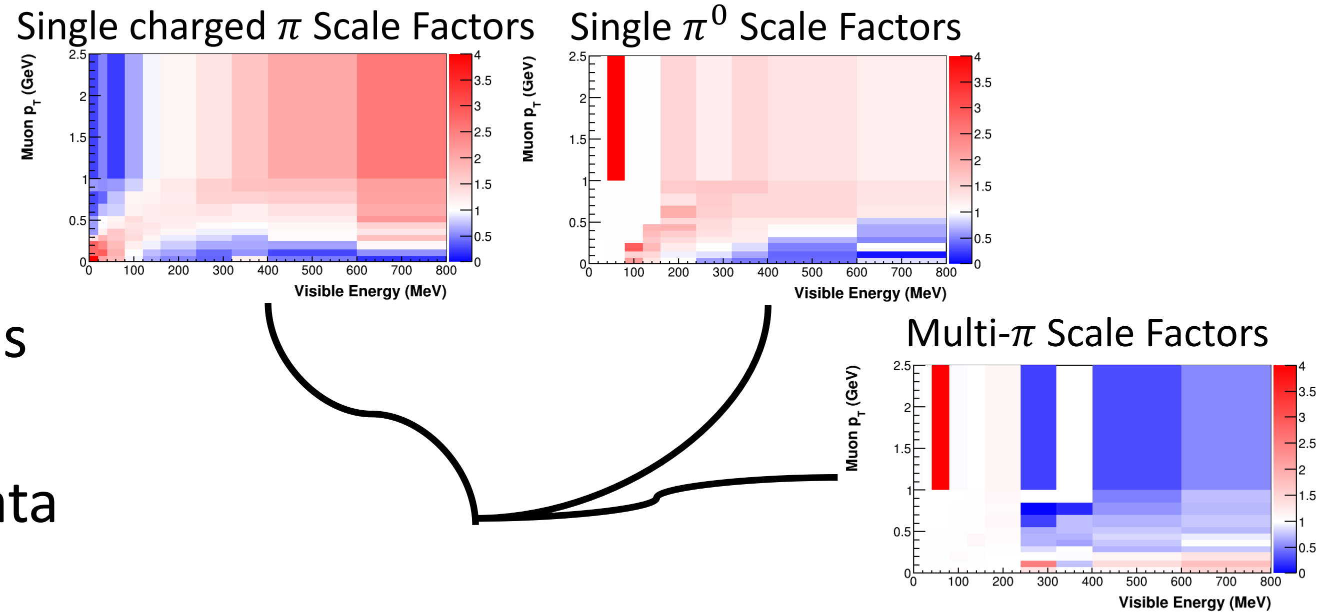
A. Bashyal, Fermilab JETP, 10 March 2023

Sidebands

General idea:

1. Choose a signal sample, sidebands that select for main backgrounds
2. Fit background distributions to data
3. Use fits to tune predicted backgrounds in signal sample

Most analyzers use this method, but fitting & tuning methods vary



D. Ruterbories, Phys.Rev.Lett. 129 (2022)

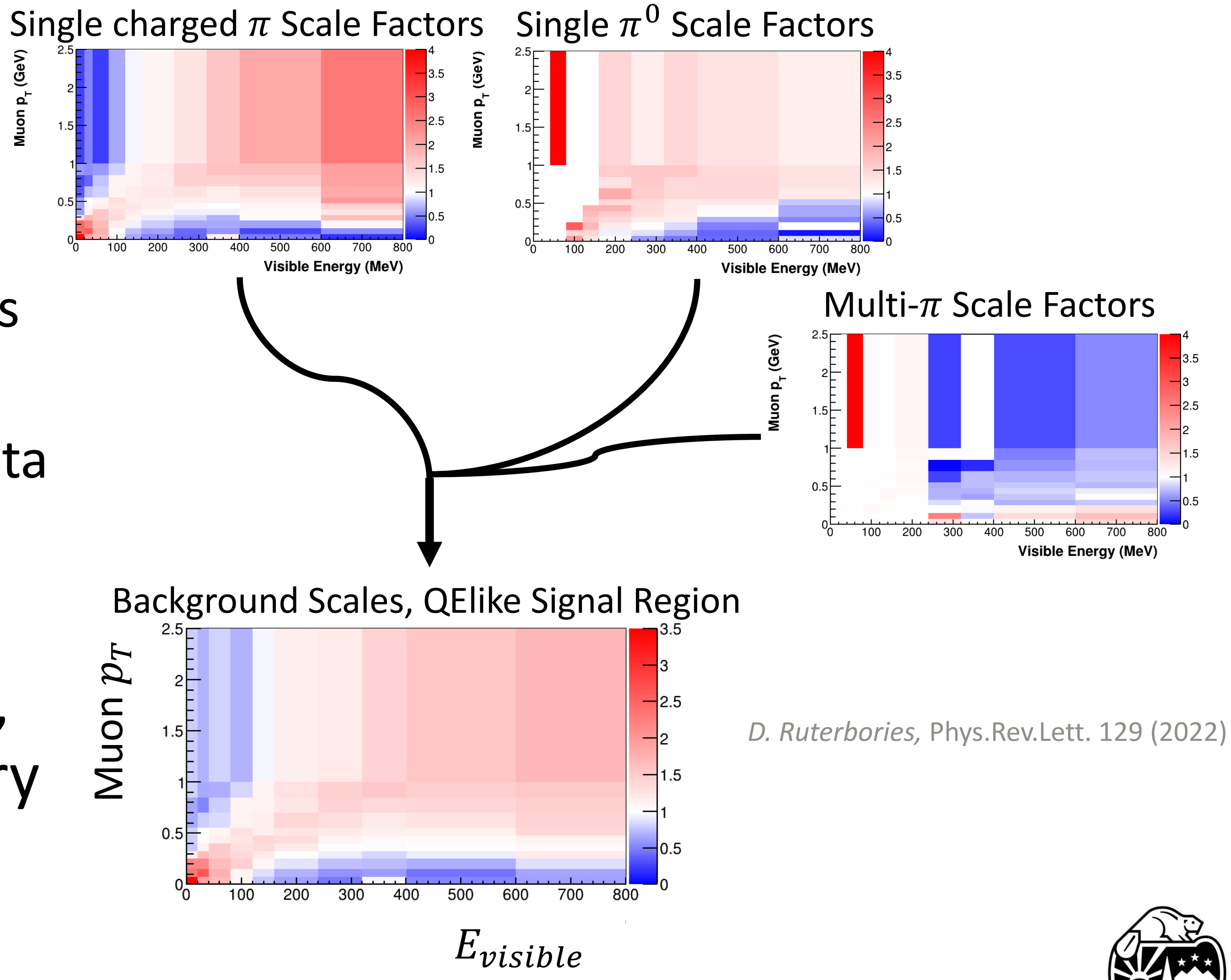
Sidebands

Background Constraint

General idea:

1. Choose a signal sample, sidebands that select for main backgrounds
2. Fit background distributions to data
3. Use fits to tune predicted backgrounds in signal sample

Most analyzers use this method, but fitting & tuning methods vary



D. Ruterbories, Phys.Rev.Lett. 129 (2022)

Unfolding

How analyzers do unfolding in MINERvA using the D'Agostini method

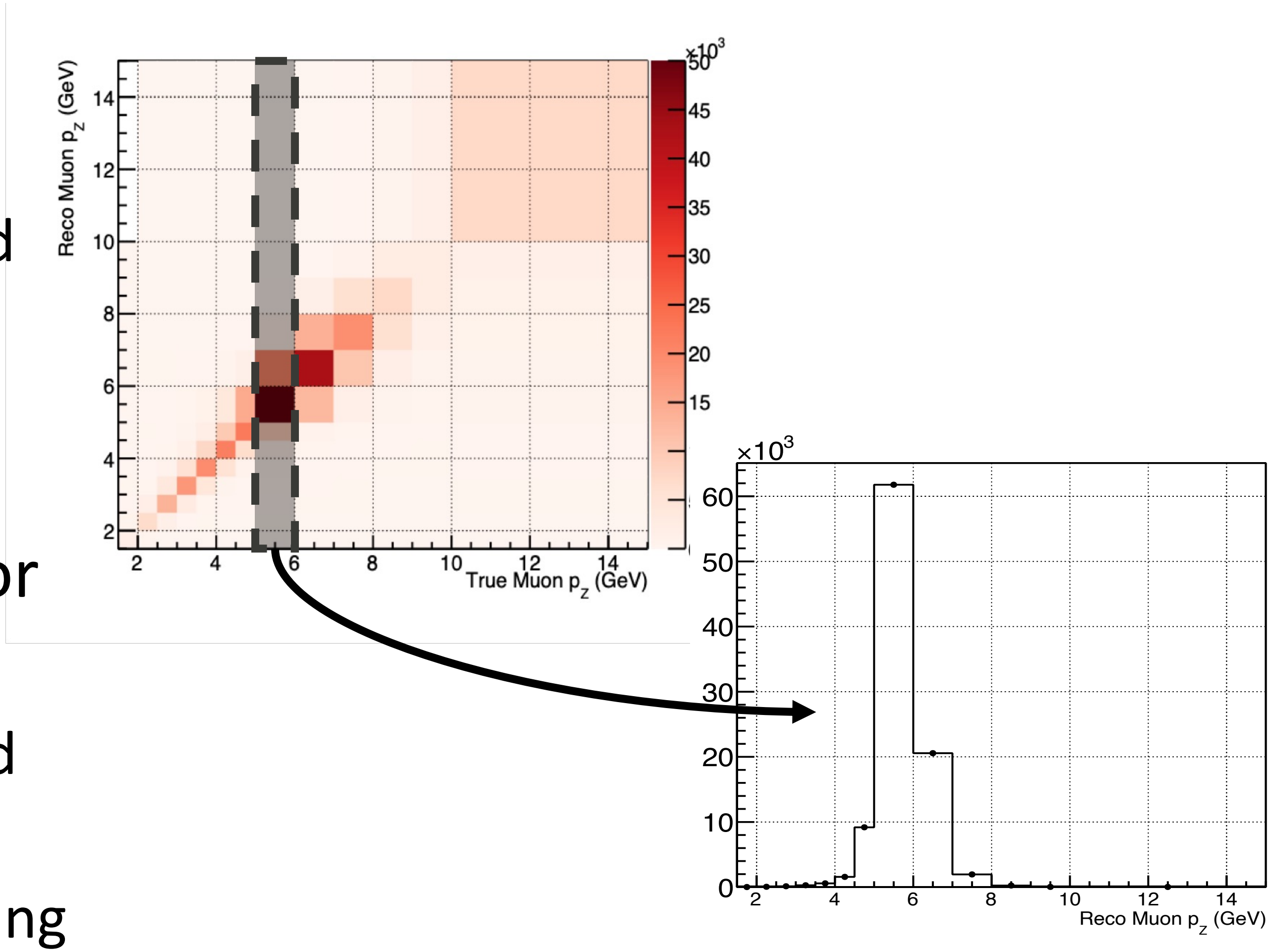
Smearing in Reconstruction

Data is in reconstructed bins

- "Smearing" of events ⇒ measured different than *true* value
- Detector resolution and reco algorithms sources of smearing

Unsmear reconstructed data for detector-independent result

- Use simulation, has both *truth* and *reco* information
- Migration matrix describes smearing from *truth* → *reconstruction*



A. Bashyal, Fermilab JETP, 10 March 2023

D'Agostini Unfolding

A method of iterative "Bayesian" unfolding, outlined here and here

- Distributions unfolded binwise
- Regularized by number of iterations
- Multidimensional unfolding
- Implemented in `RooUnfold`

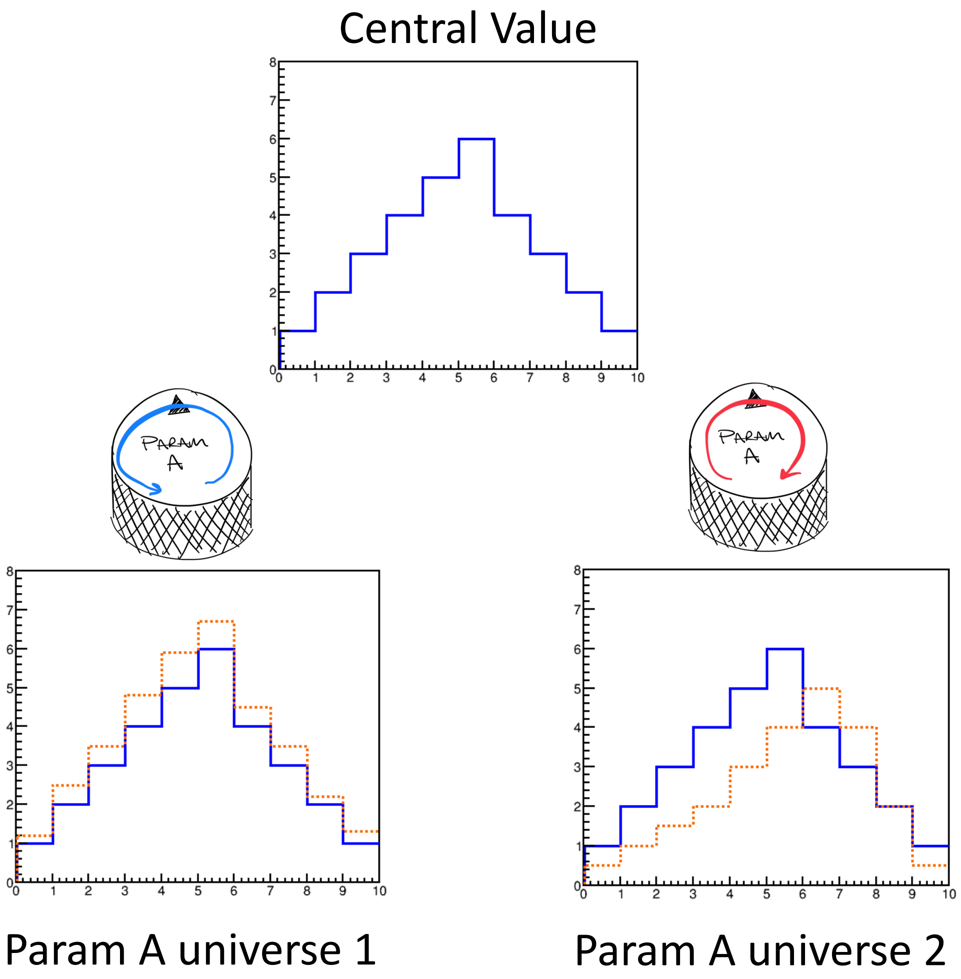
$$\lambda_j^{(k+1)} = \frac{\lambda_j^{(k)}}{\sum_{i=1}^{N} U_{ij}} \sum_{i=1}^{N} \frac{U_{ij} y_i}{\sum_{l=1}^{p} U_{il} \lambda_l^{(k)}}$$

- k is number of iterations
- $\lambda^{(0)}$ is initial truth distribution
- U is response
- y is smeared data

D'Agostini Unfolding

Implementation in MINERvA:

- `RooUnfold` based, extended to `UnfoldUtils` to handle *systematic universes*
- Square migration matrices typical, but non-square possible and explored
- Unfolds each *systematic universe* individually
- Unfolding uncertainties calculated from *central value universe*

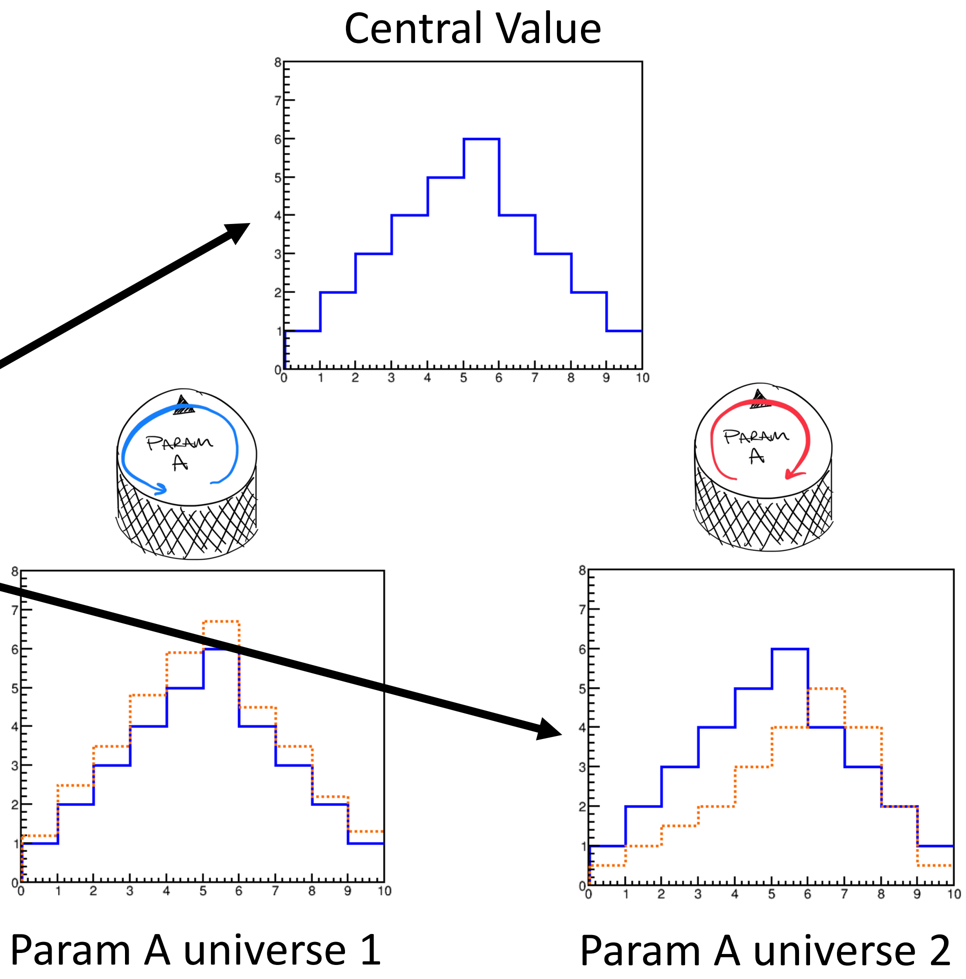


D'Agostini Unfolding

Implementation in MINERvA:

- `RooUnfold` based, extended to `UnfoldUtils` to handle *systematic universes*
- Squa … but non…
- Unfolds each *systematic universe* individually
- Unfolding uncertainties calculated from *central value universe*

Unfold each of these



Unfolding Studies

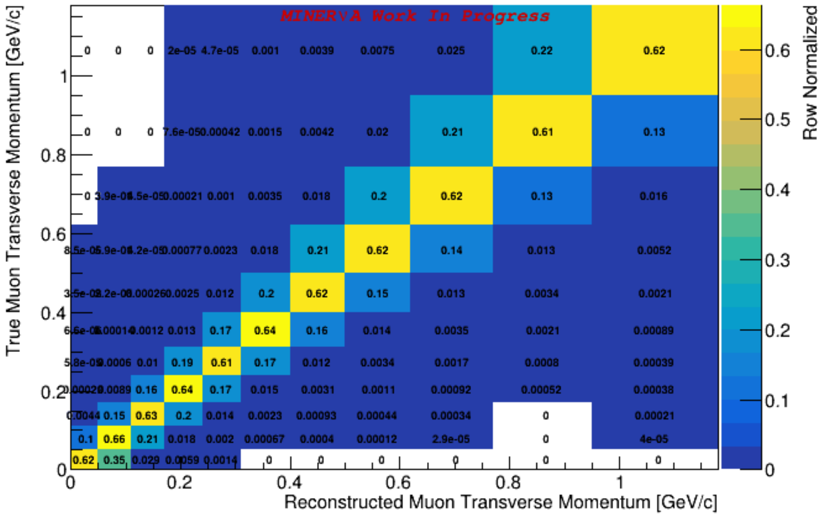
Each analyzer does unfolding studies

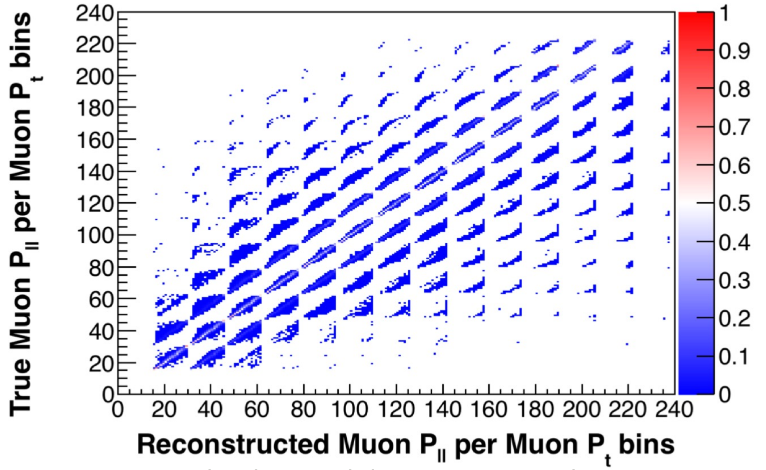
- Check behavior of migration matrix
- Ensure full consideration of unfolding uncertainties, closure tests
- Optimize number of iterations

What can be changed to fix issues?

- Adjust binning to reduce correlation, off-diagonal migration matrix
- Introduce new systematics to parameterize some cause of smearing
- Try new physics models to better match data

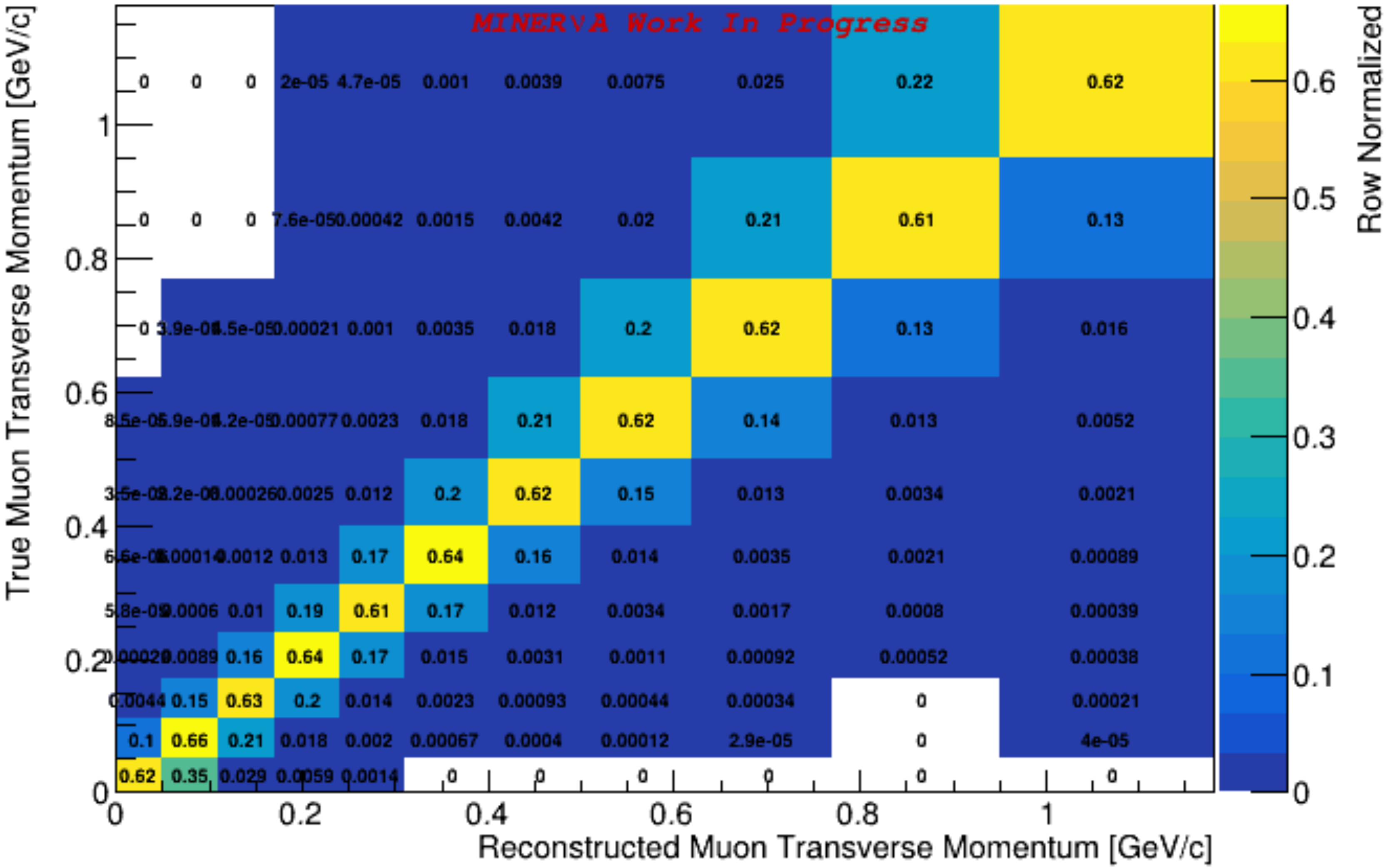
A. Olivier, Neutrino 2022





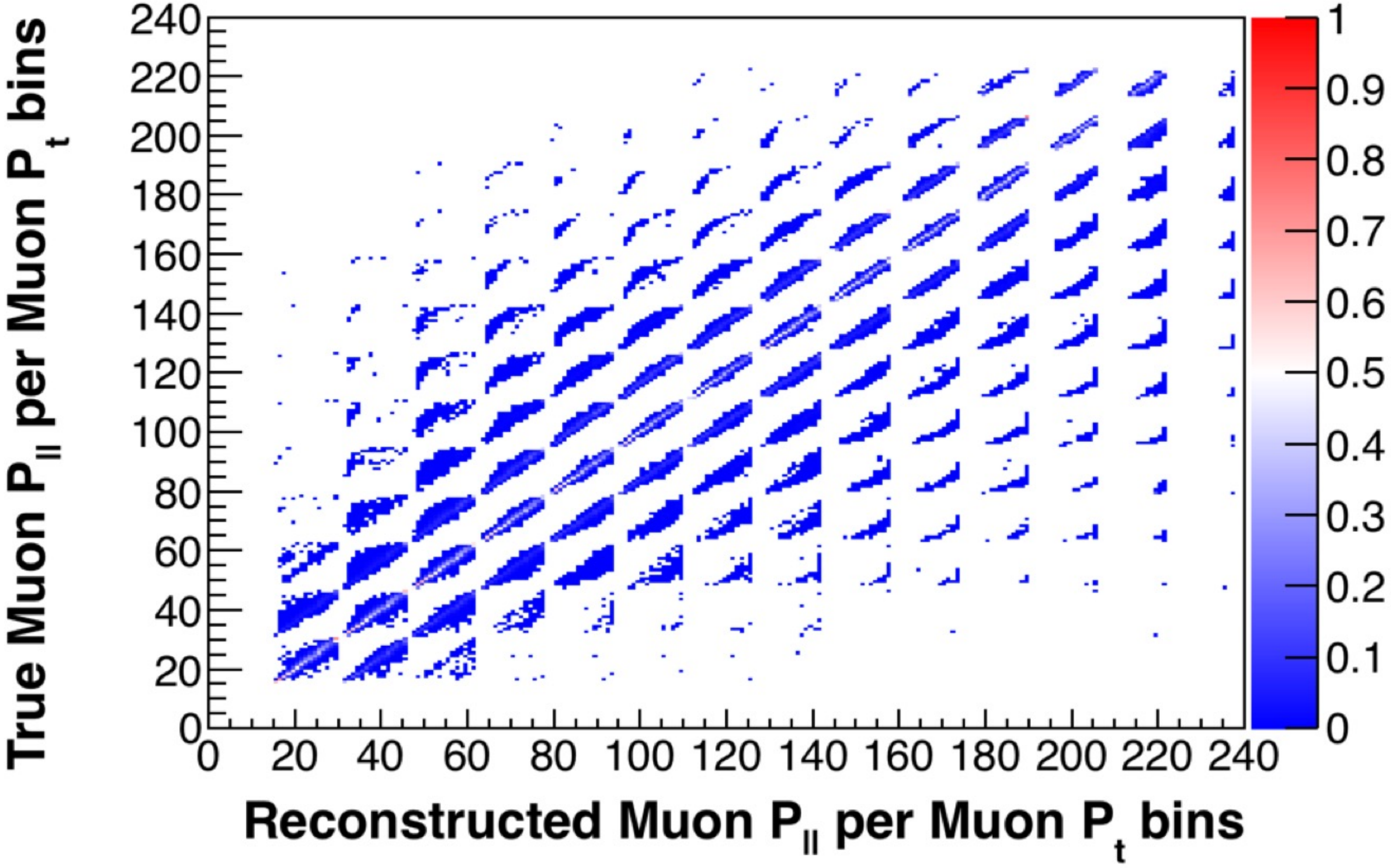
A. Bashyal, Fermilab JETP, 10 March 2023

Migration Matrix 1D



A. Olivier, Neutrino 2022

Migration Matrix 2D



A. Bashyal, Fermilab JETP, 10 March 2023

`TransWarp` for unfolding studies

Analyzers use "`TransWarp`" macro to test unfolding:

- 5 histograms as inputs:
  - *Migration matrix*
  - *Reco & True MC*
  - *Reco & True fake data* as specified number of *statistical universes* of random Poisson throws on MC

Using TransWarp

Conditions when filling:

- All hists data exposure equivalent, typically POT scaled
- Sometimes single model for all hists
- Sometimes different models for *fake data* and *migration matrix*

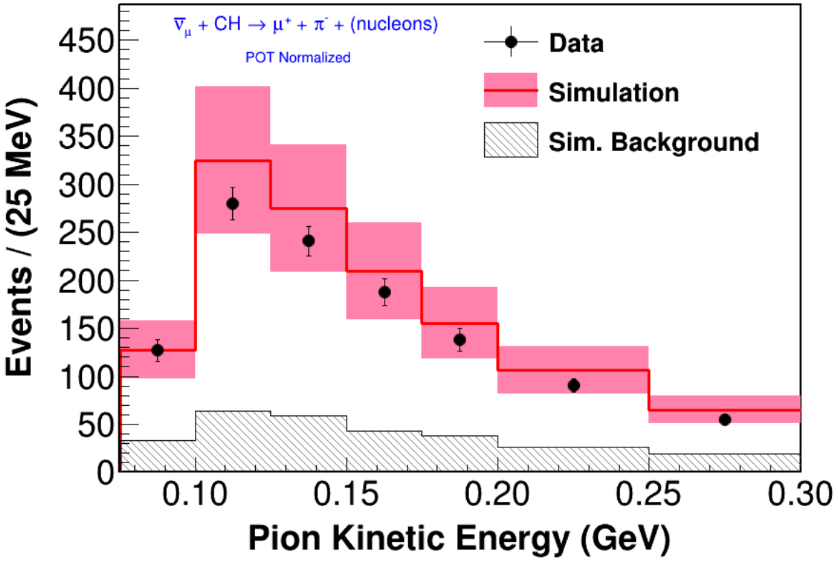
Often comparing an unfolded distribution to a true distribution

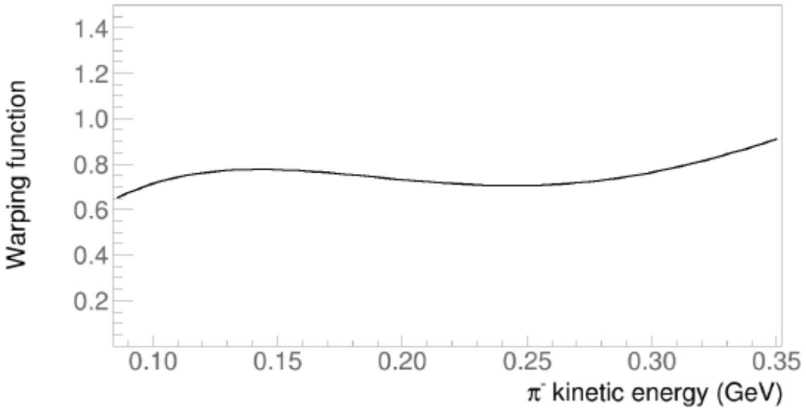
$$\chi^2 = \sum_{\alpha\beta} \left(\hat{\lambda}^{(k)} - \lambda\right)_\beta V_{\beta\alpha}^{-1} \left(\hat{\lambda}^{(k)} - \lambda\right)_\alpha$$

- λ is true kinematic
- $\hat{\lambda}^{(k)}$ is unfolded kinematic after k iterations
- V is unfolding covariance matrix

“Warping” to Determine N_{iter}

1. Options to fill *fake data*:
   - Different physics model than *migration matrix* (e.g., *MnvTune v1* vs *v2*)
   - Same model, weighted by some function from a data/MC ratio fit (“warp”)
   - Same model, up- or down-scaling (“CV scaling”)

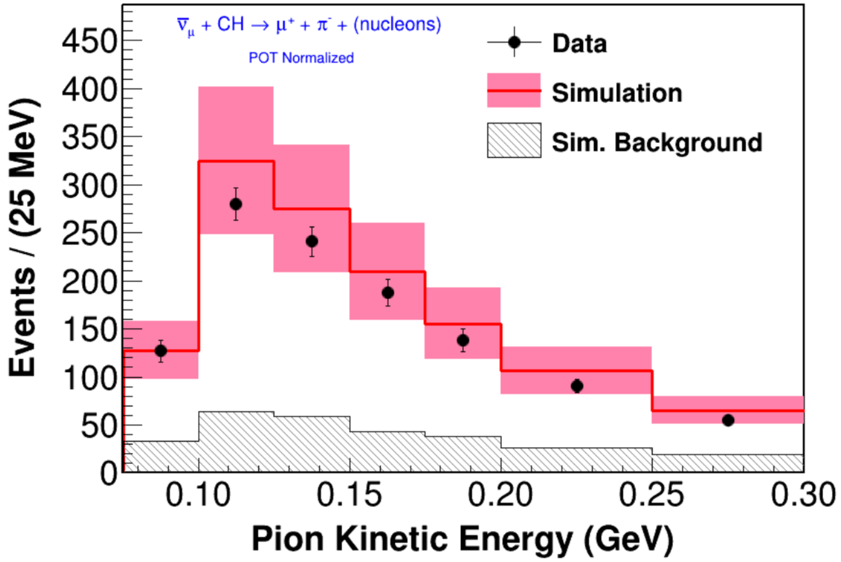


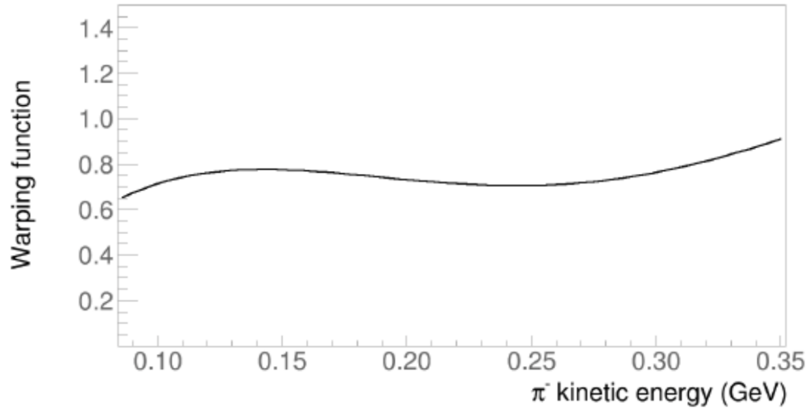


P. Rodrigues, PhySTAT, 24 Jan 2019

“Warping” to Determine N_{iter}

1. Options to fill *fake data*:
   - Different physics model than *migration matrix* (e.g., *MnvTune v1* vs *v2*)
   - Same model, weighted by some function from a data/MC ratio fit (“warp”)
   - Same model, up- or down-scaling (“CV scaling”)



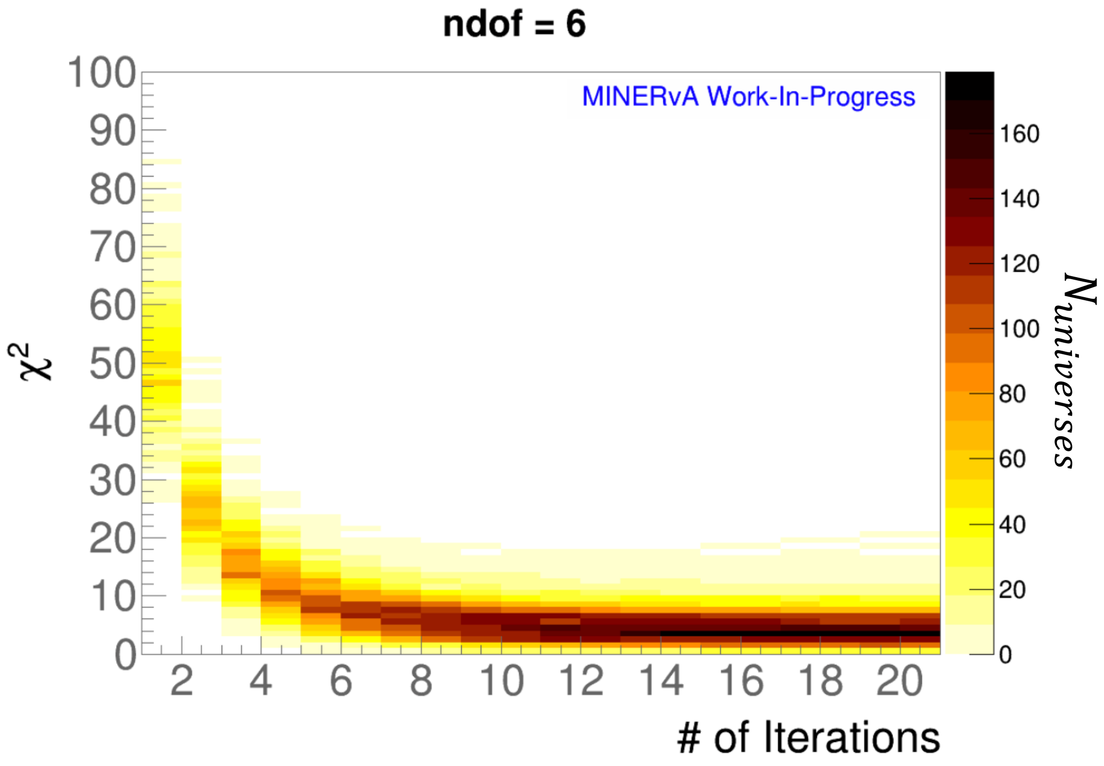


P. Rodrigues, PhySTAT, 24 Jan 2019

“Warping” to Determine N_{iter}

Continued…

2. Check χ^2 of *unfolded fake data* compared to *true fake data* vs. N_{iter}
   - All *statistical universes*, median, average
   - See minimum N_{iter} where χ^2 is minimized and stable



P. Rodrigues, PhySTAT 2019

Closing

MINERvA Cross sections

MINERvA uses its dedicated design to extract many cross sections

- MINERvA produces results that help improve generators for the next generation of neutrino physics
  - *MnvTunes* inform how generators can improve to better match observation
- Analyzers expected to use methods that are well tested
  - Signal definition set as standards
  - Flexibility in constraining the background
  - Robust testing and implementation of D'Agostini unfolding

Acknowledgements

MINERvA

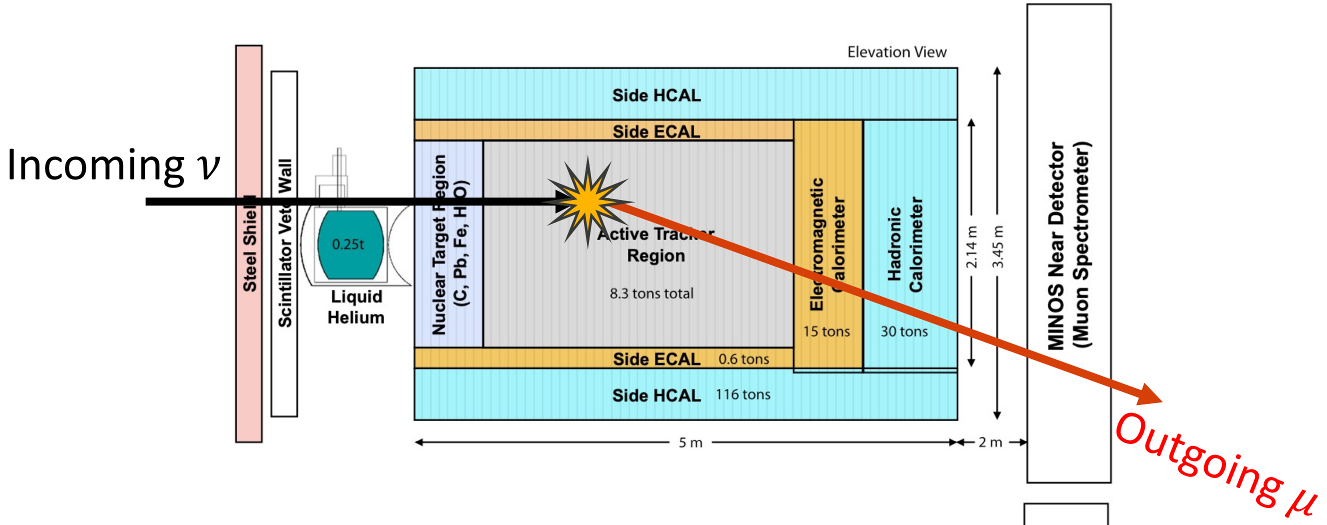
U.S. DEPARTMENT OF ENERGY

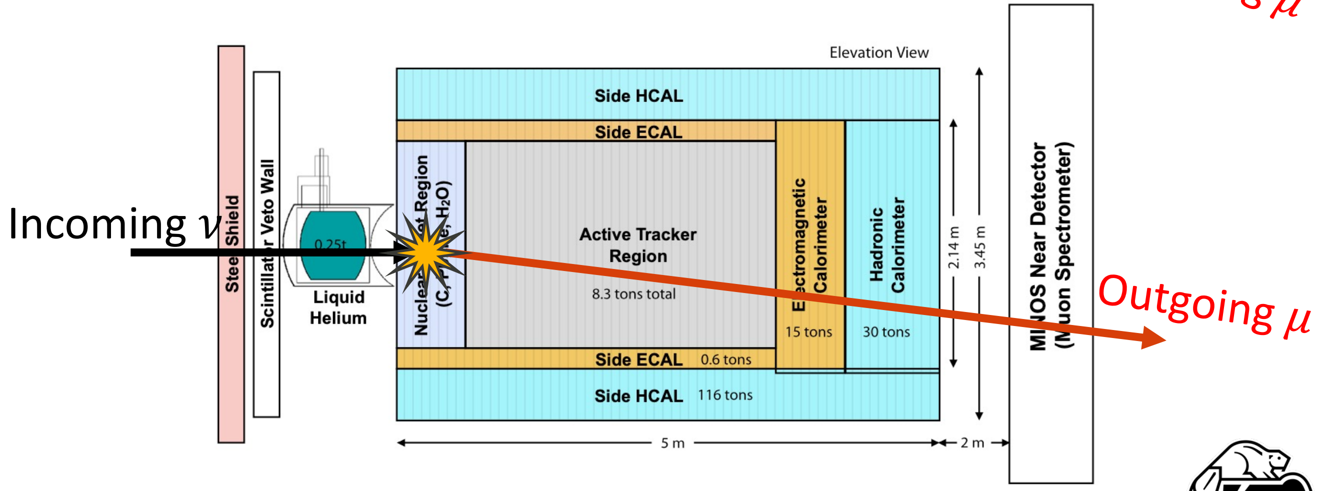
Fermilab

NSF

Common reconstruction cuts

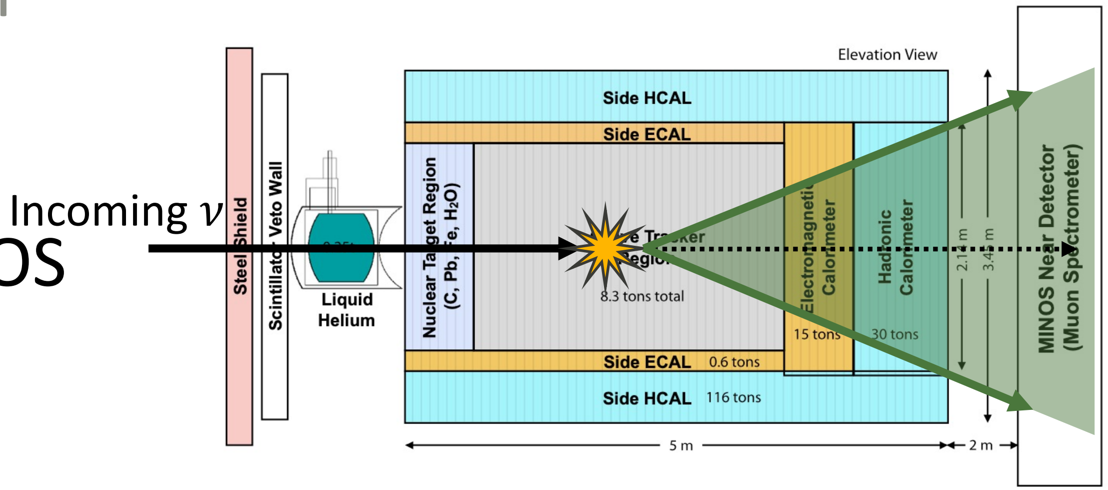
- Event vertex location, whether tracker or targets analysis





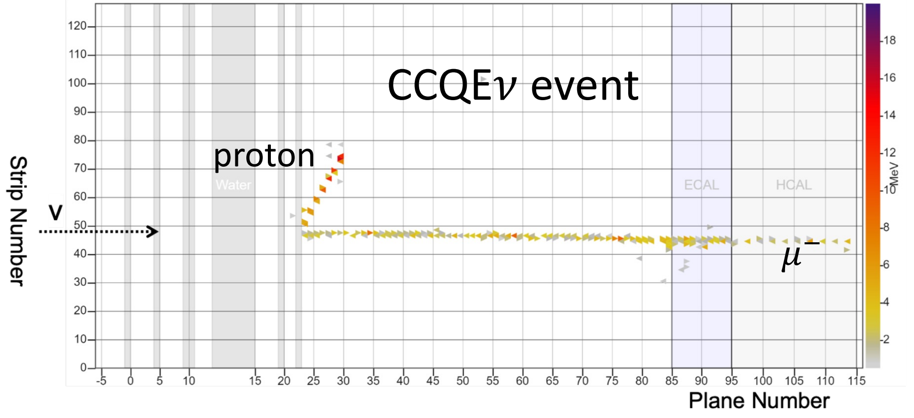
Common reconstruction cuts

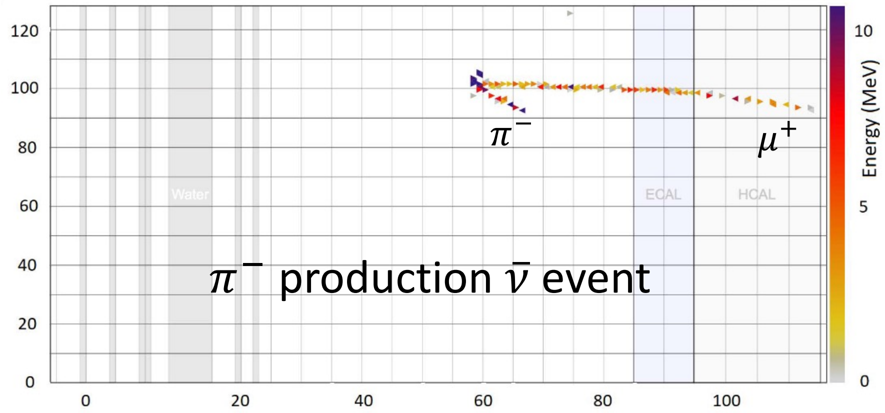
- Event vertex location, whether tracker or targets analysis
- Cuts on outgoing μ angle to ensure reconstruction in MINOS (momentum & helicity)



Common reconstruction cuts

- Event vertex location, whether tracker or targets analysis
- Cuts on outgoing μ angle to ensure reconstruction in MINOS (momentum & helicity)
- Final state particles/event topologies





Summary of MnvTune vX.Y.Z

X	Description
1	Original tune: Valencia RPA on QE (RFG), reduce non-RES π, low-recoil fit from LE to Valencia 2p2h
2	Same as 1, w/ Stowell low-Q^2_{QE} π suppression
3	Same as 1, w/ SuSA 2p2h, enhanced Bodek-Ritchie tail in Valencia QE, 25 MeV removal on E_{avail} in π events w/ FS protons
4	Same as 1, w/ π bubble chamber fit, CCNormRes scaled 1.15, MaRES set to 0.9, full MaRES and CCNormRes correlations

Y	Description
1	Normalization on coherent π production based on E_π
2	Normalization on coherent π production based on E_π and θ_π
3	Additional Low-Q^2_{QE} π suppression, normalization on coherent π production based on E_π
4	Replace dipole form of axial formfactor of QE with z-expansion from Meyer et al.
5	Replace QE RFG nuclear model with NuWro SF

Z	Description
1	π and proton elastic FSI bug fix

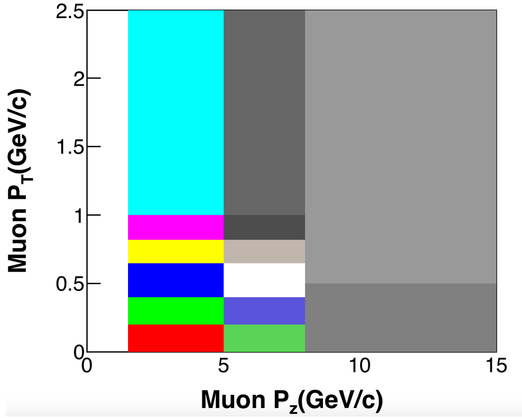
Tunes with a "0" or missing number indicate no tune from that list: e.g., *MnvTune v1*, *MnvTune v1.0.1*

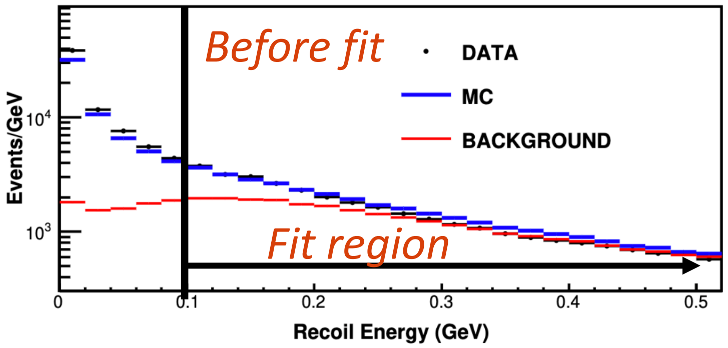
Other methods

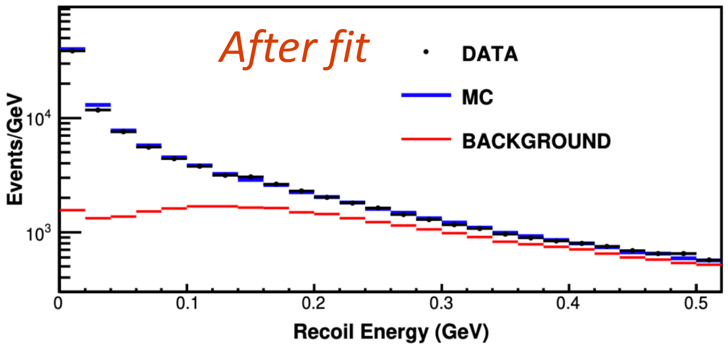
Backup: Background Constraint

Published 2D CCQE-like $\overline{\nu_\mu}$:

- Fit modified signal sample
- Treat background holistically ("*QElikenot*")
- LE: Fit full $recoil$ spectrum
- ME: Fit only in background rich, high $recoil$ (>100 MeV) region







A. Bashyal, Fermilab JETP, 10 March 2023

Closure tests

1. Fill *fake data* & *migration matrix*, using same model
2. Compare *unfolded fake data* to *true fake data*,
   - Ratio vs. N_{iter} for un-thrown *statistical universe*, should be 1
   - χ^2 vs. N_{iter} of all *statistical universes* & their average, should be stable

Additional Stat Uncertainty

Stat uncertainty for $N_{iter} > 1$ missing in some implementations

- Greatly underpredicts uncertainty without this
- Enhancement applied as scale, studies must be done to find scale factor
- Scale factor is typically similar between similar analyses

$$\frac{\partial \hat{n}(\mathrm{C}_i)}{\partial n(\mathrm{E}_j)} = M_{ij} + \sum_{k=1}^{n_\mathrm{E}} M_{ik} n(\mathrm{E}_k) \left(\frac{1}{n_0(\mathrm{C}_i)} \frac{\partial n_0(\mathrm{C}_i)}{\partial n(\mathrm{E}_j)} - \sum_{l=1}^{n_\mathrm{C}} \frac{\epsilon_l}{n_0(\mathrm{C}_l)} \frac{\partial n_0(\mathrm{C}_l)}{\partial n(\mathrm{E}_j)} M_{lk} \right)$$

T. Adye, arXiv:1105.1160 [physics.data-an]

Additional Stat Uncertainty

Stat uncertainty for $N_{iter} > 1$ missing in some implementations

- Greatly underpredicts uncertainty without this
- Enhancement applied as scale, studies must be done to find scale factor
- Scale factor is typically similar between similar analyses

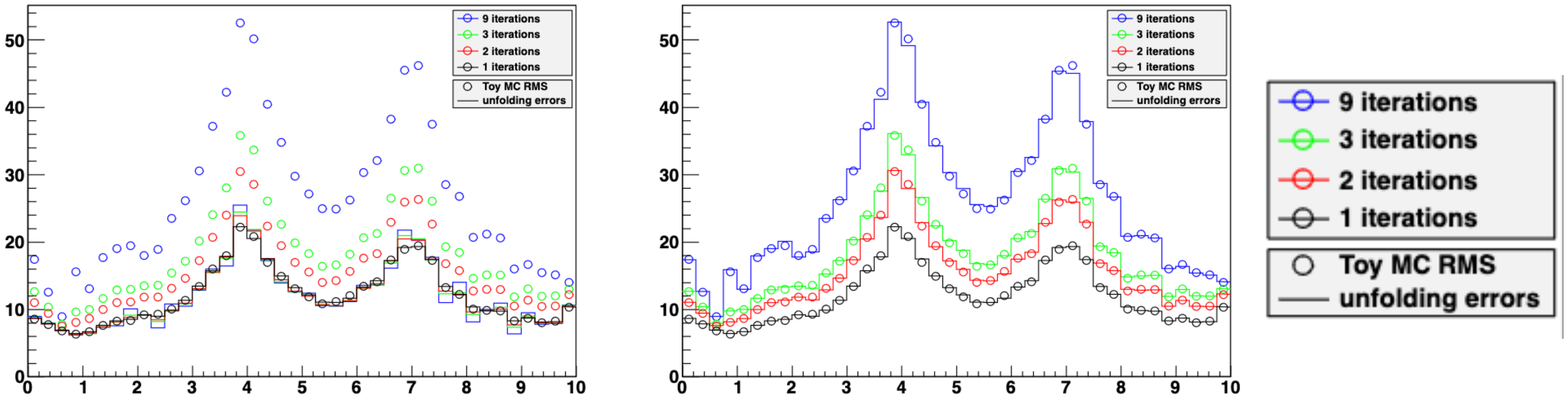


Fig. 5: Bayesian unfolding errors (lines) compared to toy MC RMS (points) for 1, 2, 3, and 9 iterations on the Fig. 2 test. The left-hand plot shows the errors using D'Agostini's original method, ignoring any dependence on previous iterations (only the M_{ij} term in Eq. (3)). The right-hand plot shows the full error propagation.

T. Adye, arXiv:1105.1160 [physics.data-an]

Finding Stat Uncertainty Scale

Analyzers can use the `TransWarp` macro here too:

1. Fill *fake data* and *migration matrix* using a single model
2. Check average χ^2 vs. N_{iter} for different *MC sample sizes*
   - Typically, change scale in integer steps 1 - 10, inc. the normal data exposure POT scale
   - Check how far χ^2 lines for each MC scale gets to ${}^{\chi^2}/_{ndf} = 1$
3. Determine and save *uncertainty scale* to bring average $\chi^2 = ndf$
4. Scale unfolding covariance & error using the *uncertainty scale*, and run closure test
   - Ratio of *unfolded fake data* to *true fake data* should be unchanged
   - χ^2 vs. N_{iter} of all *statistical universes* will shift down slightly, remain stable

Other things to look for

Other things to check to find issues (e.g., overly aggressive fake data)

- Migration matrices for *fake data*, the CV, ratios between migration matrices
- Warping weight vs. *true* & *reco* distributions
- *True fake data*, unfolded fake data *statistical universes*, average *unfolded fake data*
- Ratio of average *unfolded fake data* to *true fake data*

Upcoming results…

Quasielastic

- ν 3D transverse kinematic imbalance
- $\bar{\nu}$ 3D simultaneous leptonic & hadronic measurement
- $\nu/\bar{\nu}$ ratios
- Neutron tagging

Low hadronic recoil

- Interactions with 2+ neutrons
- ν_e & $\overline{\nu_e}$
- Charged π's

Inelastic

- Several DIS results
- SIS results
- Interactions on Helium