Supported by U.S. DOE: Award DE-SC0015903

Generator Benchmarks: What a NUISANCE

Luke Pickering, pronouns: he/him/his

NuFACT 2019/08/29

MICHIGAN STATE UNIVERSITY

Team NUISANCE

Comparison tools used in this talk developed as part of NUISANCE with numerous external contributions:

special thanks to A. Mastbaum and S. Dolan!

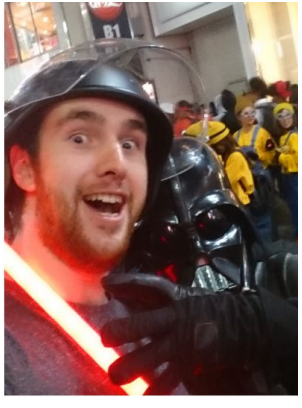
P. Stowell

MICHIGAN STATE UNIVERSITY

C. Wret

The University Of Sheffield.

UNIVERSITY of ROCHESTER

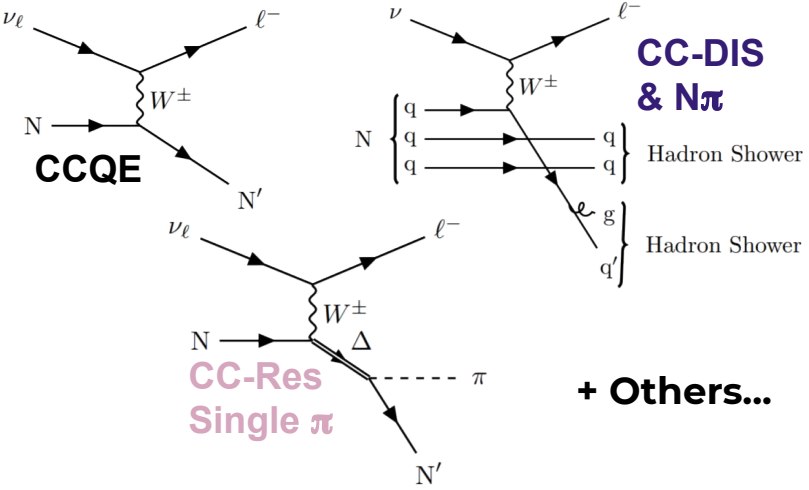


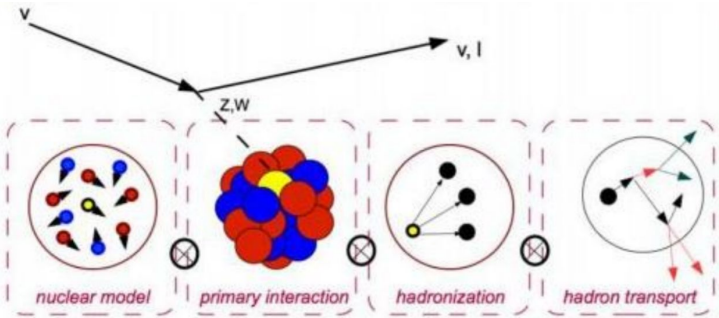
C. Wilkinson

UNIVERSITÄT BERN

What is a Neutrino Event Generator

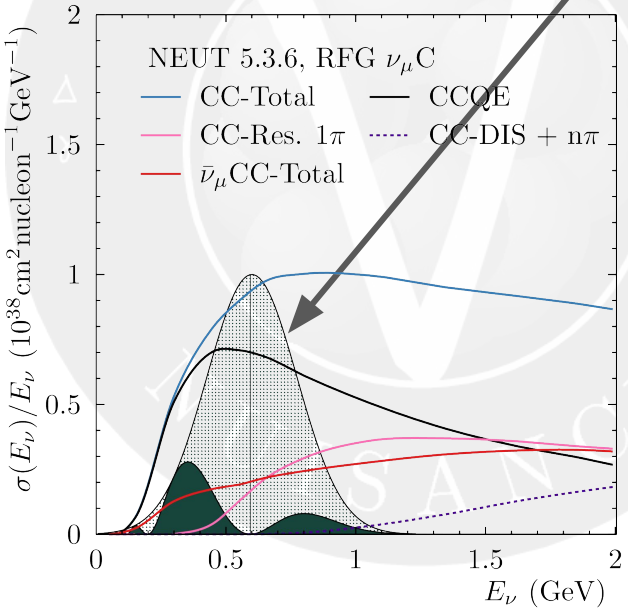
- Selects neutrino ‘events’ from interaction models:
  - **Over a range of neutrino energy and species,**
  - **For a number of ‘primary’ channels**:
    - Neutrino--nucleus (COHPi, CvNS)
    - Neutrino--multi-nucleon (2p2h)
    - Neutrino--nucleon (QE, RESPi)
    - Neutrino--parton (DIS)
  - **In a nuclear environment**:
    - Fermi motion distribution
    - Removal energy
    - Collective effects (RPA)
    - Final state re-interactions of primary particles
- Often factorises the simulation of nuclear model, primary interaction, and FSIs.





Why do we use them?

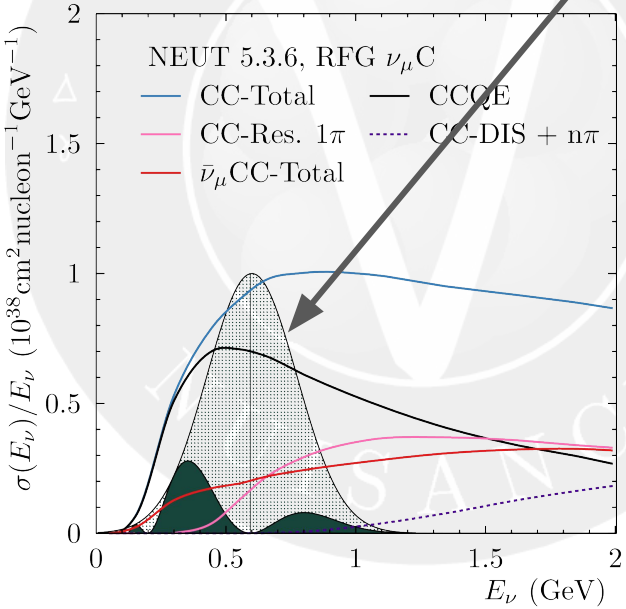
- Want to learn about **neutrinos**.

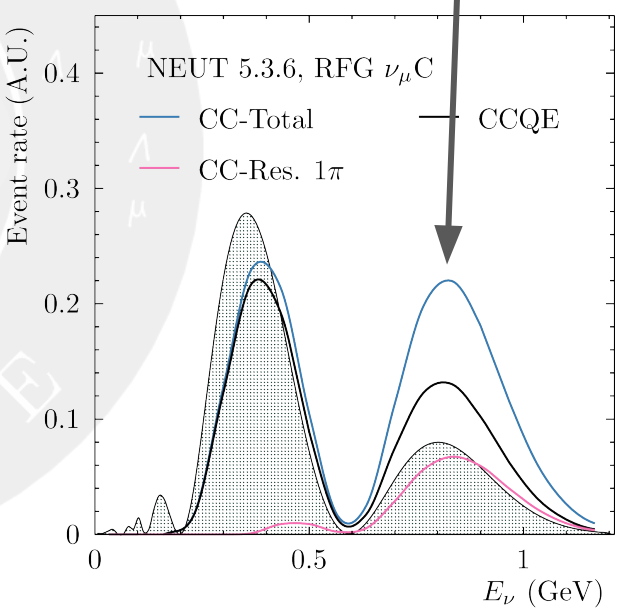


Flux x Cross section

Why do we use them?

- Want to learn about **neutrinos**, but see **interactions**



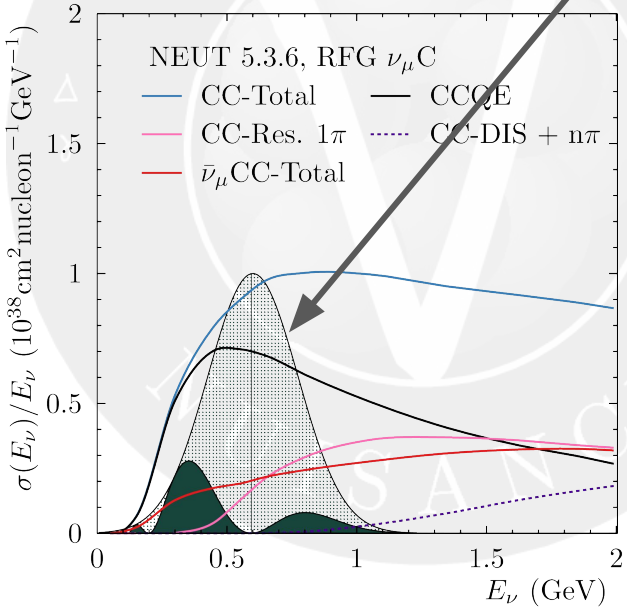


Flux x Cross section **= Event rate**

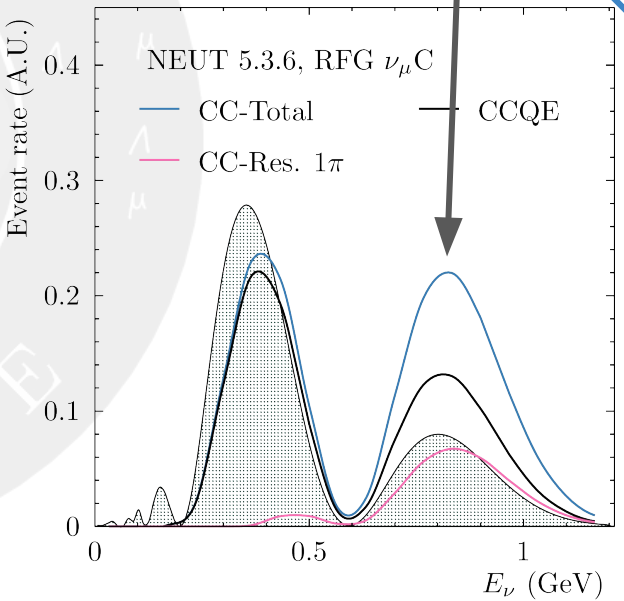
NEUT: Acta Phys.Polon. B40 (2009) 2477-2489

Why do we use them?

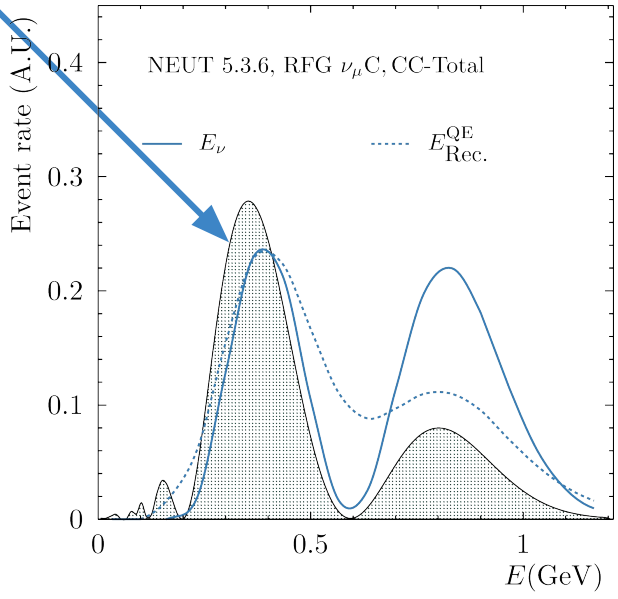
- Want to learn about **neutrinos**, but see **interaction final states**.



Flux x Cross section



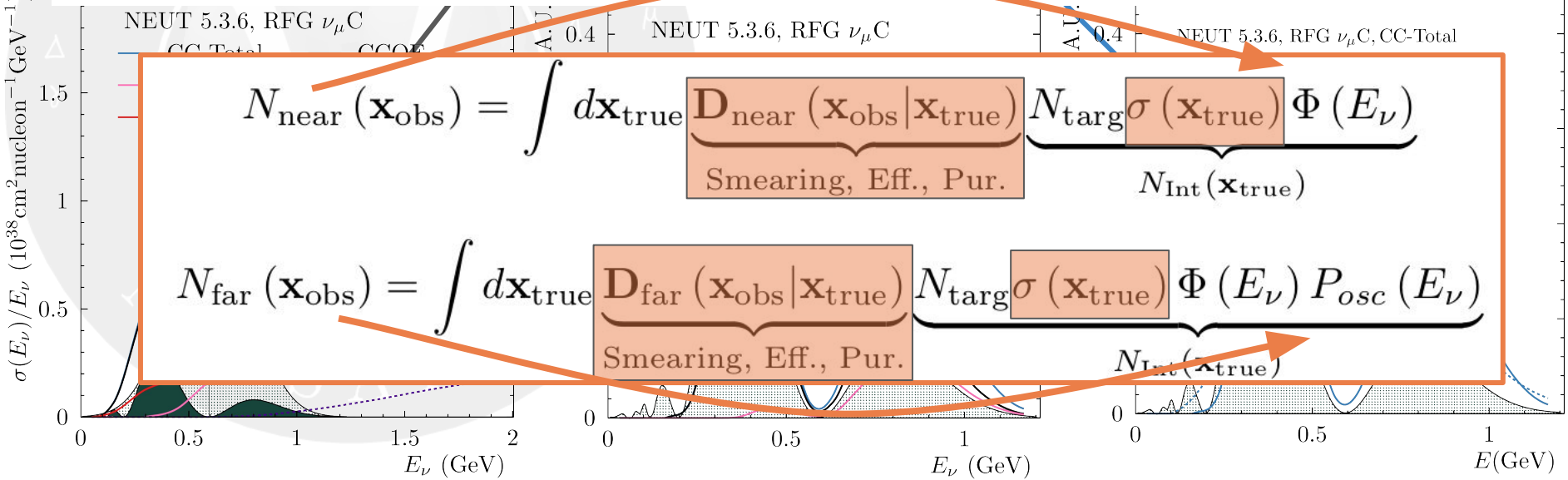
= Event rate



Observed properties

Why do we use them?

Need to work back from observables to learn about neutrinos: Done via generators



Flux x Cross section **= Event rate** **Observed properties**

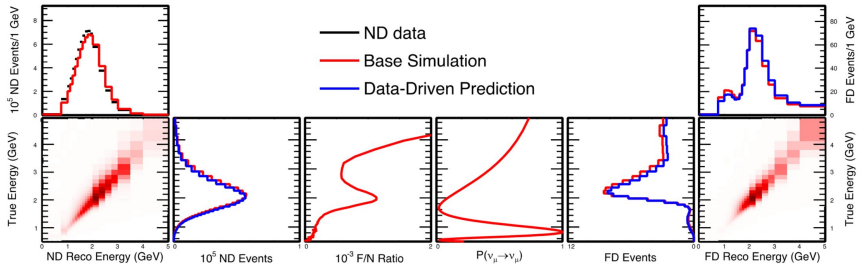
NEUT: Acta Phys.Polon. B40 (2009) 2477-2489

What are they trusted to predict?

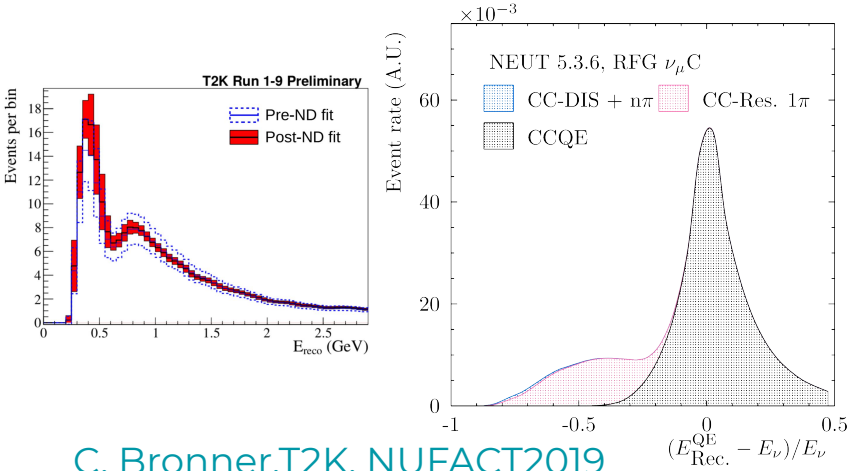
- Predict backgrounds for exotic processes:
  - Nucleon decay, dark matter, ...
- For ν-A cross-sections:
  - Simplifies efficiency determination
  - Predict purity of signal selections
  - Propagate errors correctly to published data

What are they trusted to predict?

- Predict backgrounds for exotic processes:
  - Nucleon decay, dark matter, …
- For ν-A cross-sections:
  - Simplifies efficiency determination
  - Predict purity of signal selections
  - Propagate errors correctly to published data
- For oscillation physics:
  - Predict observable distributions as a function o Energy to correctly infer oscillation parameter values.



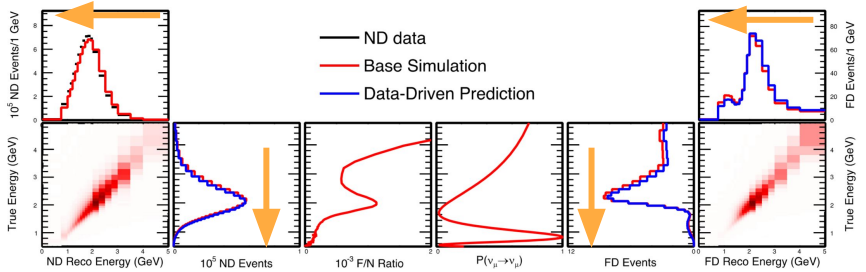
E. Smith, NOvA, NUFACT2019



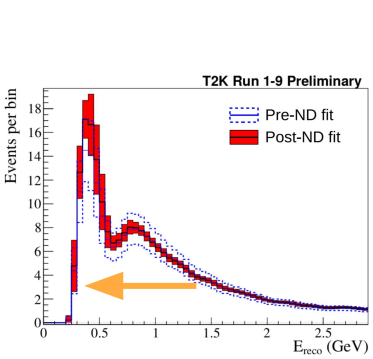
C. Bronner,T2K, NUFACT2019

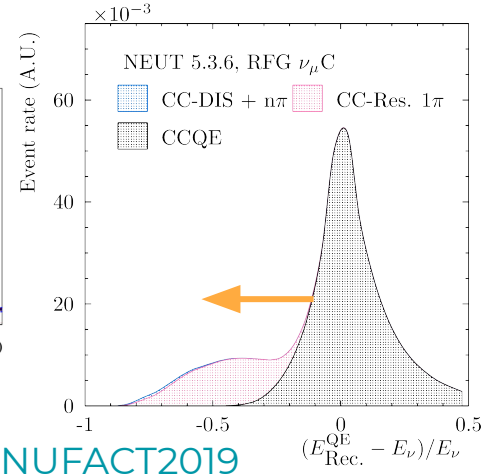
What are they trusted to predict?

- Predict backgrounds for exotic processes:
  - Nucleon decay, dark matter, …
- For ν-A cross-sections:
  - Simplifies efficiency determination
  - Predict purity of signal selections
  - Propagate errors correctly to published data
- For oscillation physics:
  - Predict observable distributions as a function o Energy to correctly infer oscillation parameter values.



E. Smith, NOvA, NUFACT2019

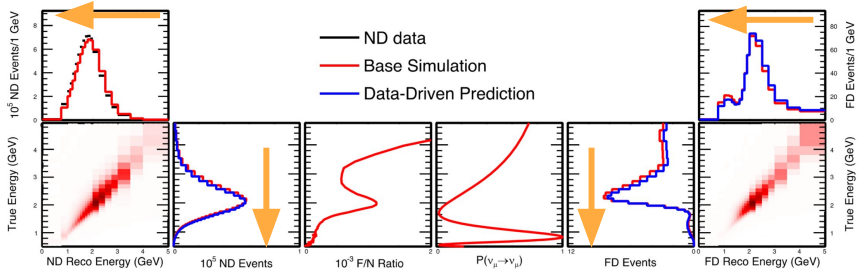




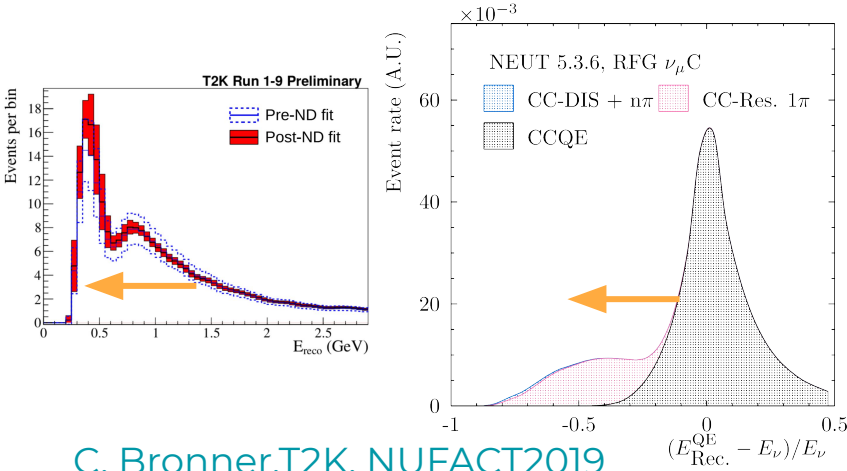
C. Bronner,T2K, NUFACT2019

What are they trusted to predict?

- Predict backgrounds for exotic processes:
  - Nucleon decay, dark matter, …
- For ν-A cross-sections:
  - Simplifies efficiency determination
  - Predict purity of signal selections
  - Propagate errors correctly to published data
- For oscillation physics:
  - Predict observable distributions as a function o Energy to correctly infer oscillation parameter values.
- Progress made in two prong approach:
  - Bottom up: Theory development
  - Top down: Comparison and benchmarking against published data



E. Smith, NOvA, NUFACT2019



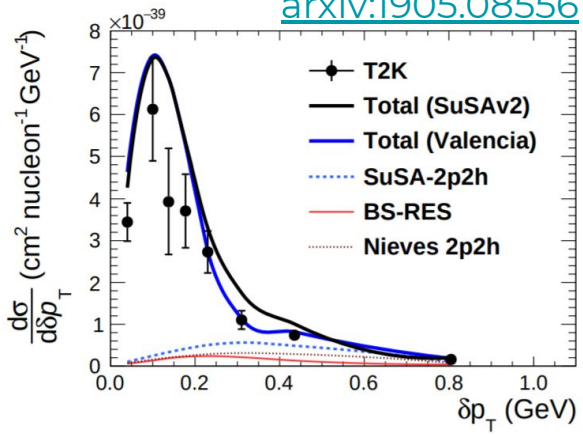
C. Bronner,T2K, NUFACT2019

How do we try and make them right: Theory

- Improve nuclear response models in generators:
  - *e.g.* SuSAv2 1p1h+2ph2 PRD 94, 093004 (2016)
- Improve primary interaction models in generators:
  - *e.g.* MK single pion production PRD 97, 013002 (2018)

χ^2 SuSA: 21/8
χ^2 Val.: 27/8

arxiv:1905.08556

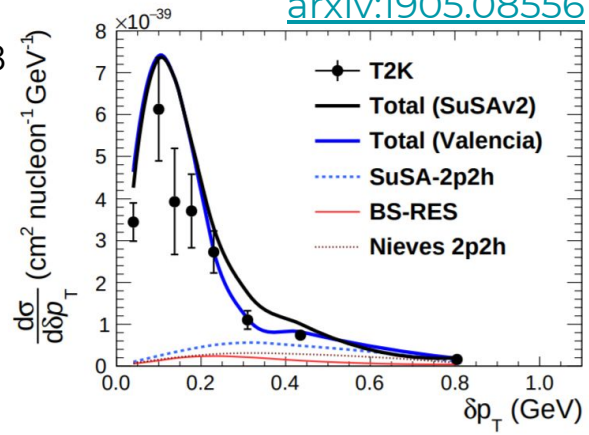


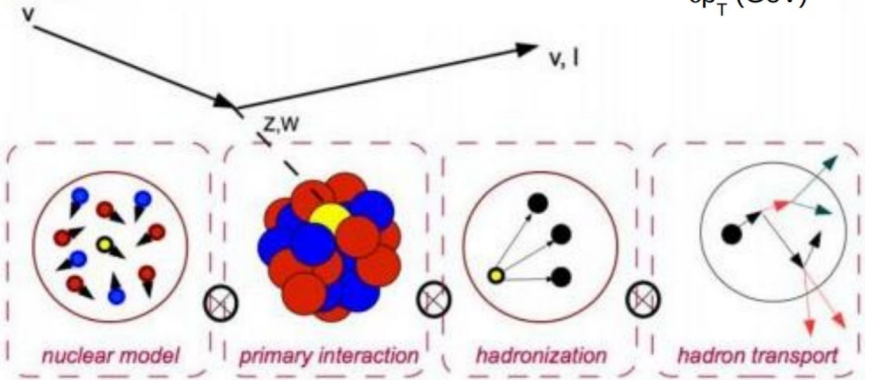
How do we try and make them right: Theory

- Improve nuclear response models in generators:
  - *e.g.* SuSAv2 1p1h+2ph2 PRD 94, 093004 (2016)
- Improve primary interaction models in generators:
  - *e.g.* MK single pion production PRD 97, 013002 (2018)
- Improve simplifications in the MC:
  - Recent interest on un-doing the primary interaction factorisation to better-capture initial and final state physics and lepton-hadron correlations.

χ^2 SuSA: 21/8
χ^2 Val.: 27/8

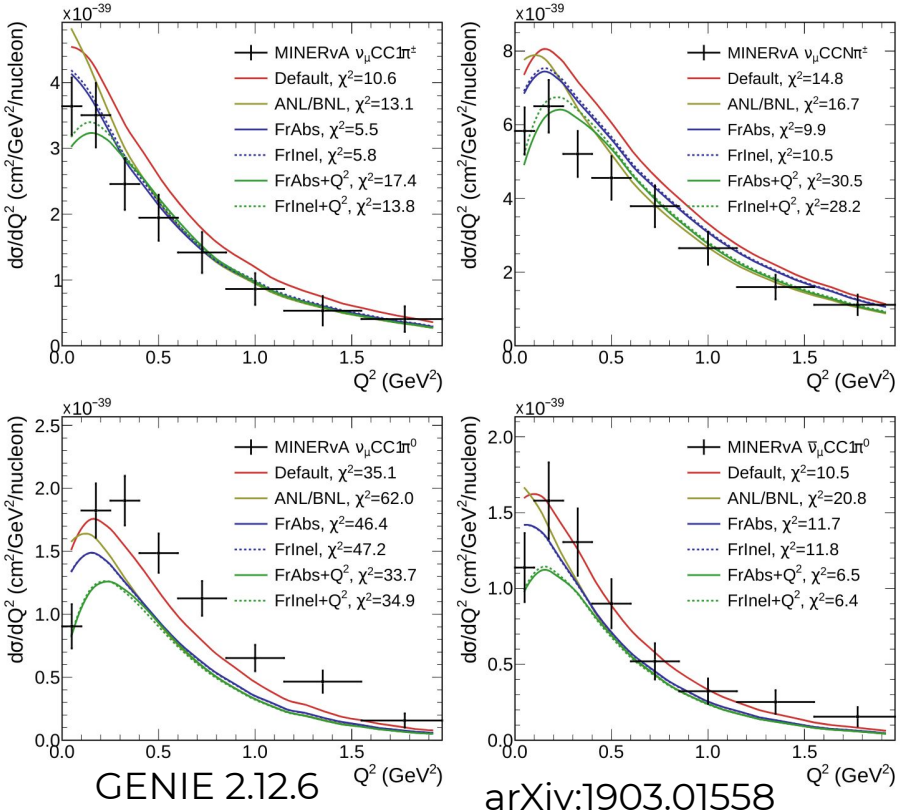
arxiv:1905.08556





How do we try and make them right: Tune

- In an ideal world, the model would describe nature up to some unknown parameters.

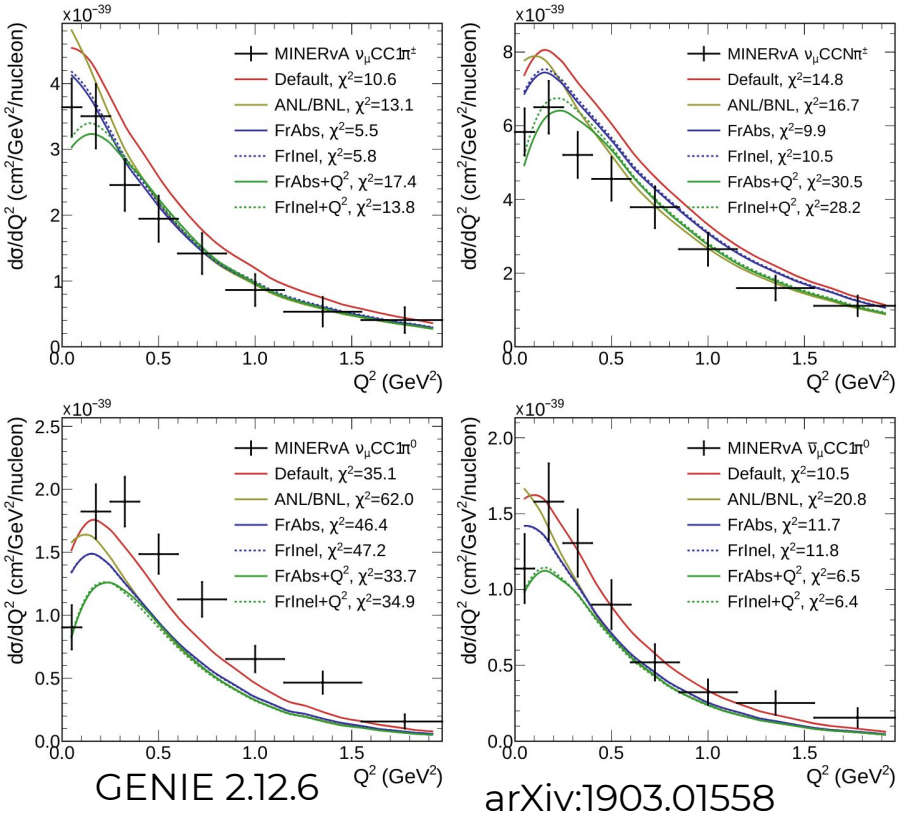


GENIE 2.12.6

arXiv:1903.01558

How do we try and make them right: Tune

- In an ideal world, the model would describe nature up to some unknown parameters:
  - We don't live in that world.
  - Confronting the models with a variety of data will improve predictions and highlight areas for theory development.

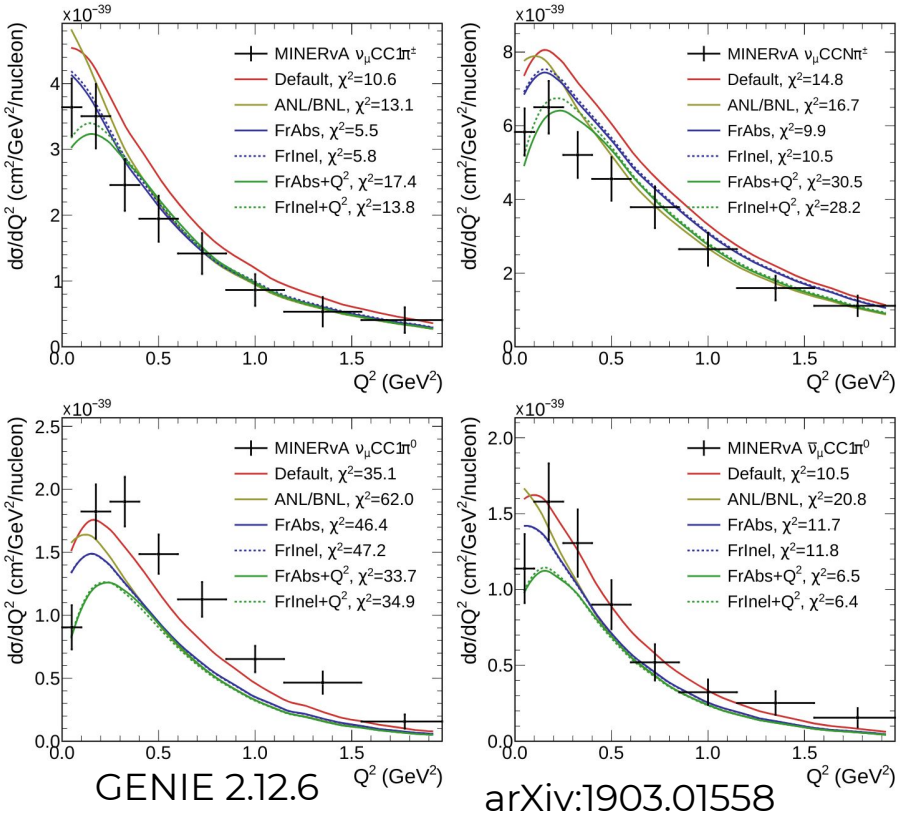


GENIE 2.12.6

arXiv:1903.01558

How do we try and make them right: Tune

- In an ideal world, the model would describe nature up to some unknown parameters:
  - We don't live in that world.
  - Confronting the models with a variety of data will improve predictions and highlight areas for theory development.
- **Dangers of tuning:**
  - Propagate CV+uncerts from well-described projection to poorly described projection without extra uncertainties.
  - *e.g.* Tune in inclusive lepton variables and predict hadronic shower variables.



GENIE 2.12.6

arXiv:1903.01558

Meet the Generators

	Version/ Tune Used	Nuclear-model + QE-like	Single Pion Production	Higher W	Fragmentation	FSI
NEUT	5.4.0	Valencia: - 1p1h+RPA - 2p2h	Rein-Sehgal + lepton mass effects	Bodek-Yang low Q^2	Pythia 5	Tuned Salcedo-Oset cascade
GENIE	v3.0.4 G1810a_0211 + bug-fixed splines	Valencia: - 1p1h+RPA - 2p2h	Rein-Sehgal 16 resonances non-interfering (BC Tuned)	Bodek-Yang low Q^2	AGKY+Pythia 6	Tuned effective single interaction (hA)
NuWRO	v19.02	- Benhar SF w/ opt. pot. - Valencia: RPA & 2p2h	Delta + Pythia Low W	Bodek-Yang low Q^2	Pythia 6	Tuned Salcedo-Oset cascade

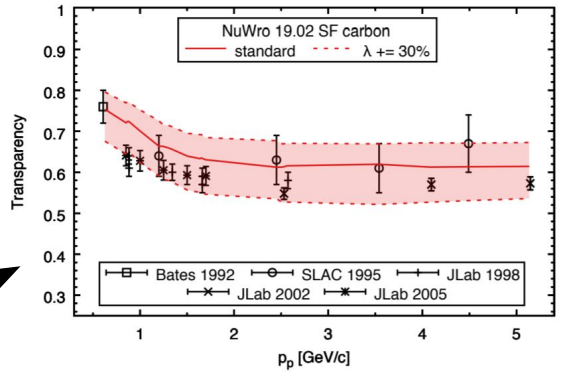
Notable Recent Developments

- NEUT:
  - Nieves 1p1h, LFG nuclear model
  - Improved multi-pion production from BC tune
  - MK pion production, Bug fixes in R-S pion production

Notable Recent Developments

Phys. Rev. C 100, 015505 (2019)

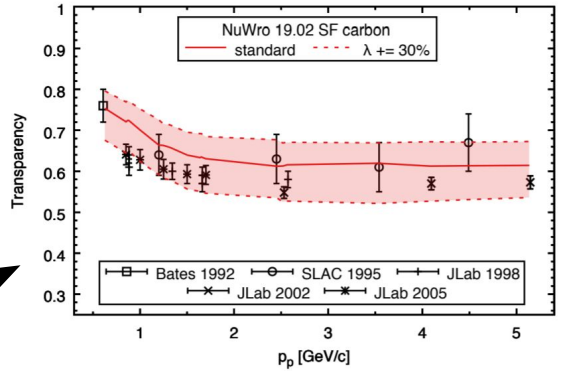
- NEUT:
  - Nieves 1p1h, LFG nuclear model
  - Improved multi-pion production from BC tune
  - MK pion production, Bug fixes in R-S pion production
- NuWro:
  - Updates to spectral function
  - Update of FSI cascade by comparison to nuclear transparency data.
  - Integration of electron scattering simulation.

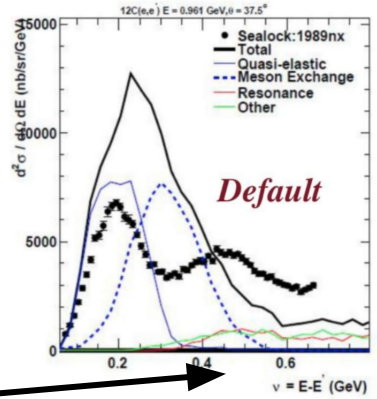


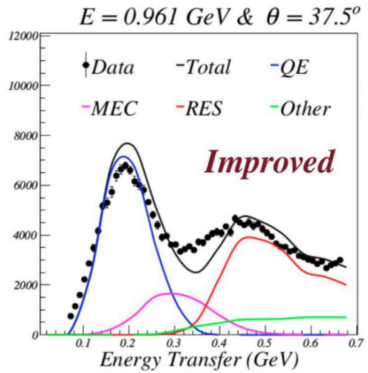
Notable Recent Developments

Phys. Rev. C 100, 015505 (2019)

- NEUT:
  - Nieves 1p1h, LFG nuclear model
  - Improved multi-pion production from BC tune
  - MK pion production, Bug fixes in R-S pion production
- NuWro:
  - Updates to spectral function
  - Update of FSI cascade by comparison to nuclear transparency data.
  - Integration of electron scattering simulation.
- GENIE:
  - Version 3 released!
  - Extensive ν-N tuning to bubble chamber data
  - **Many improvements to electron scattering simulation (c.f. Or Hen e4nu Plenary)**
  - Some significant bug fixes



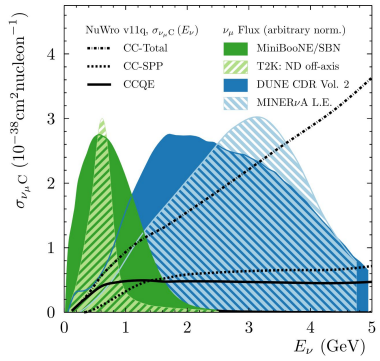




*Genie R-2_12_10

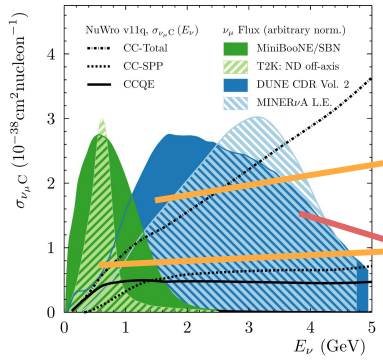
What we want out of comparisons to data

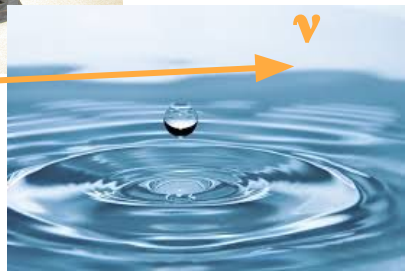
- Range of:
  - Neutrino energies



What we want out of comparisons to data

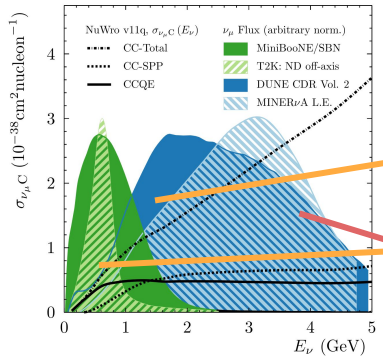
- Range of:
  - Neutrino energies
  - Targets

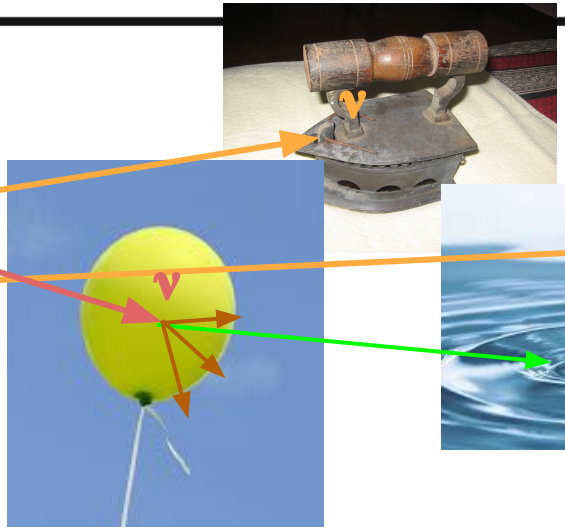




What we want out of comparisons to data

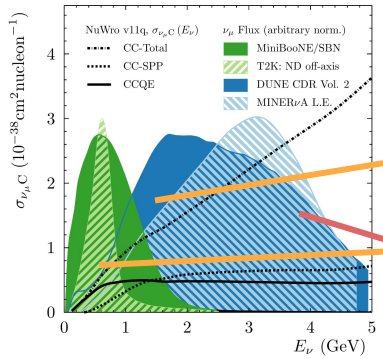
- Range of:
  - Neutrino energies
  - Targets
  - Final state topologies
  - Observable projections

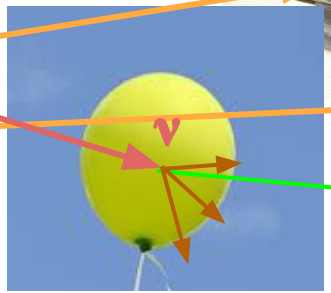


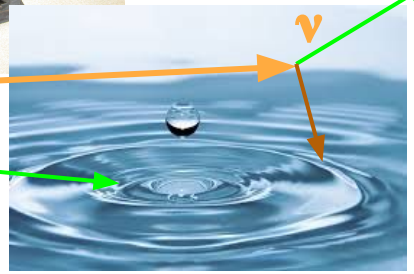


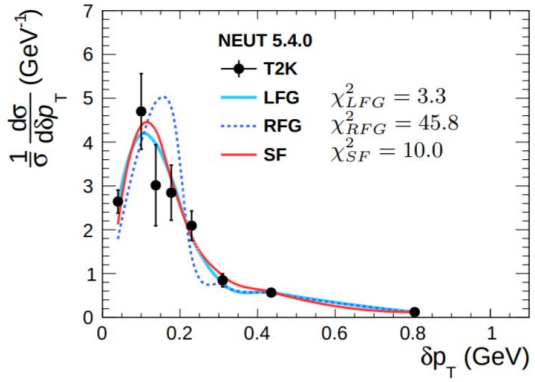
What we want out of comparisons to data

- Range of:
  - Neutrino energies
  - Targets
  - Final state topologies
  - Observable projections
- Sensitivity to:
  - Model choice
  - Free parameter central values
  - Free parameter uncertainties





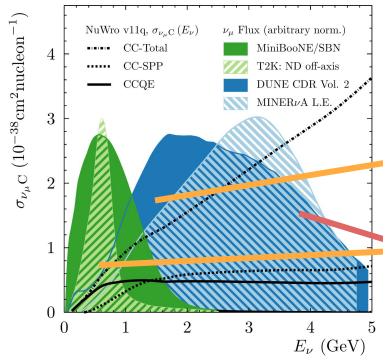


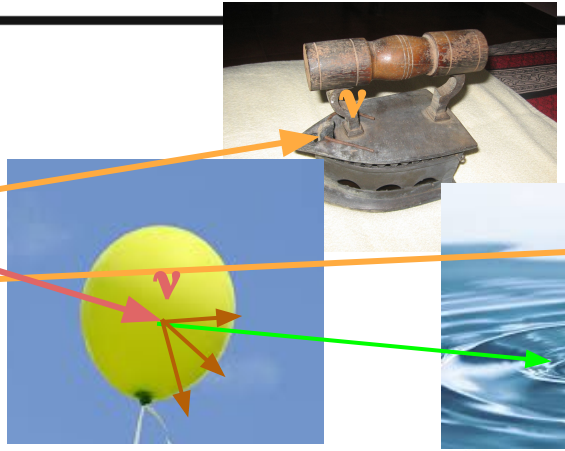


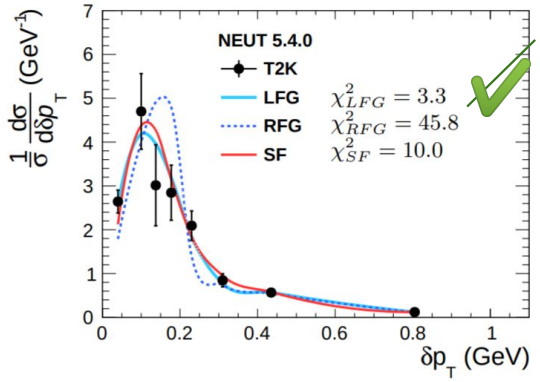
T2K data: PRD98, 032003 (2018)
Plot: arXiv:1810.06043

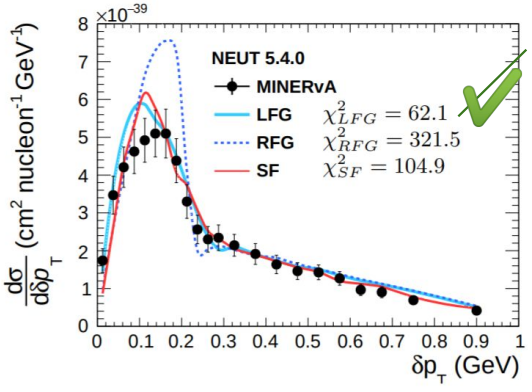
What we want out of comparisons to data

- Range of:
  - Neutrino energies
  - Targets
  - Final state topologies
  - Observable projections
- Sensitivity to:
  - Model choice
  - Free parameter central values
  - Free parameter uncertainties
- Ability to make quantitative statements about GOF
- Give nature fewer places to hide!





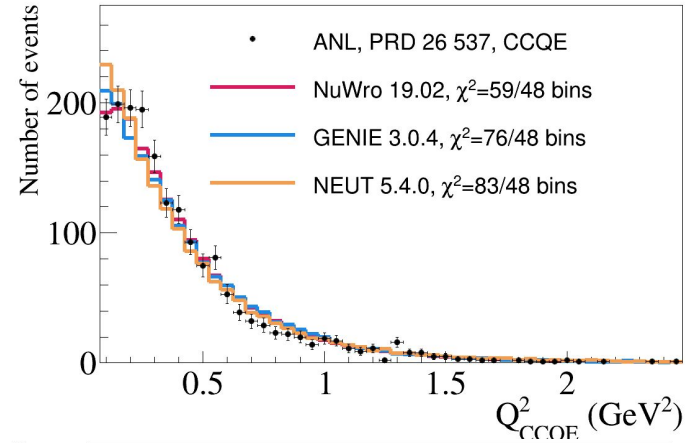


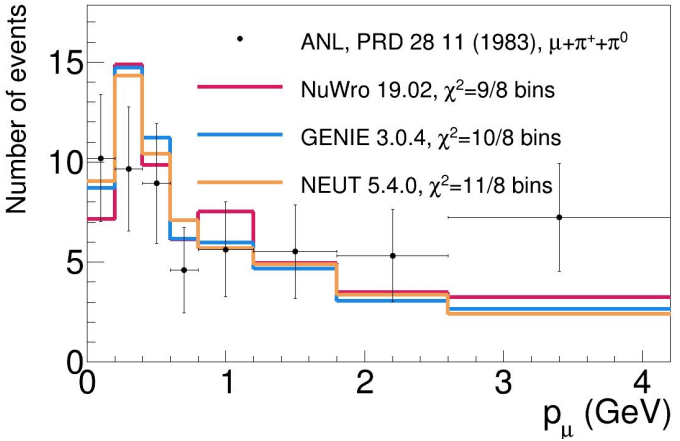


T2K data: PRD98, 032003 (2018) **MINERvA data**: PRL 121 (2018) no.2, 022504
Plots: arXiv:1810.06043

Comparisons to Bubble Chamber data

- (quasi-)free of any nuclear effects.
  - Granular reconstruction and unambiguous final state topologies.
  - Allows tuning of 'primary' neutrino nucleon/part interaction.

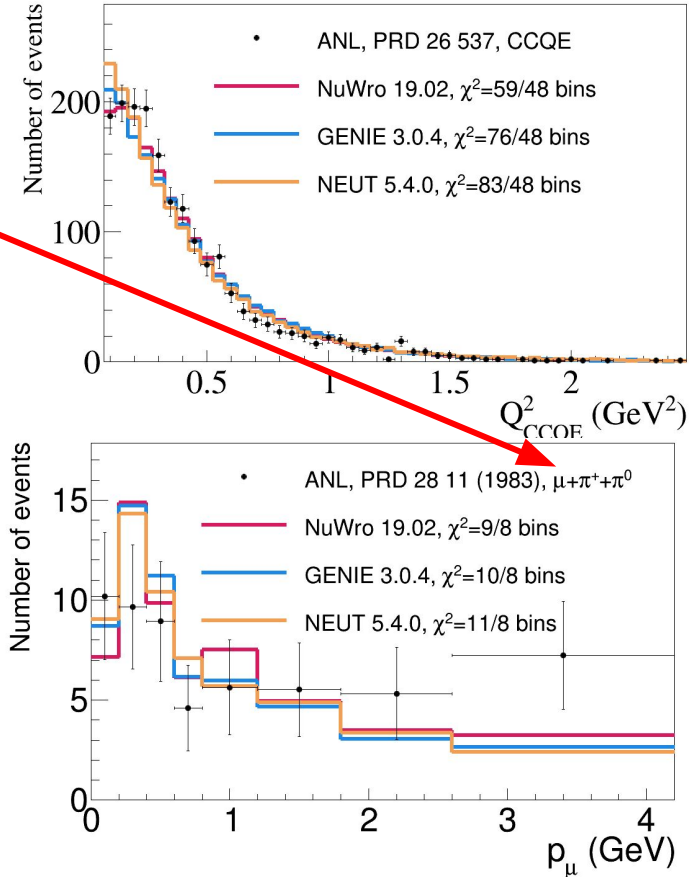




MICHIGAN STATE UNIVERSITY

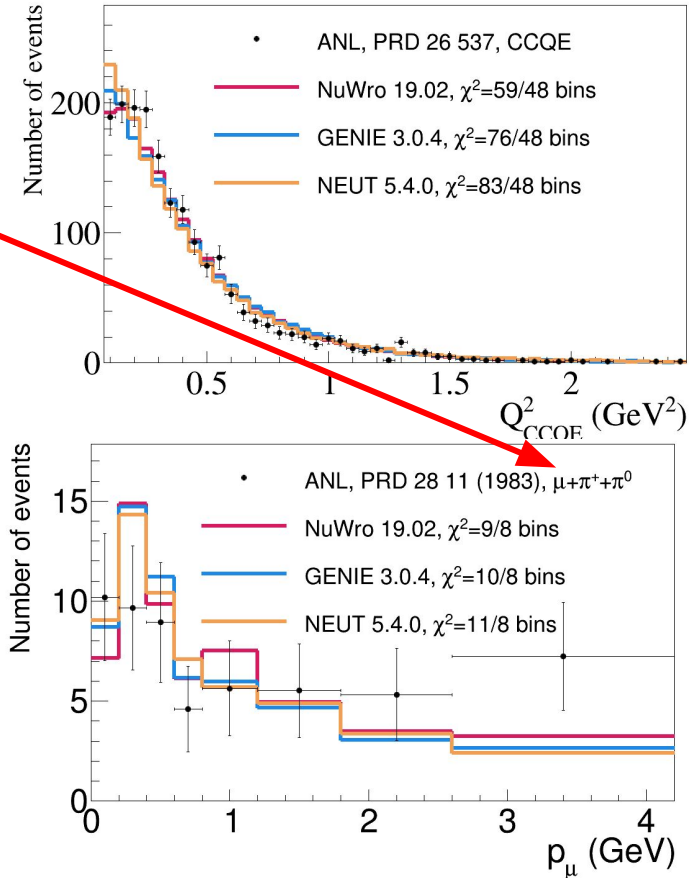
Comparisons to Bubble Chamber data

- (quasi-)free of any nuclear effects.
  - Granular reconstruction and unambiguous final state topologies.
  - Allows tuning of ‘primary’ neutrino nucleon/part interaction.



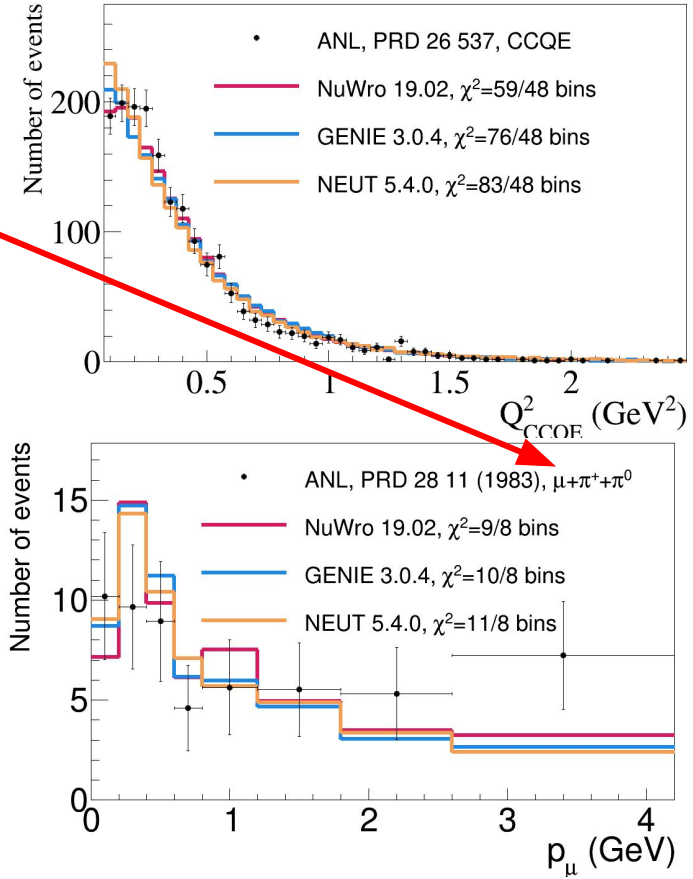
Comparisons to Bubble Chamber data

- (quasi-)free of any nuclear effects.
  - Granular reconstruction and unambiguous final state topologies.
  - Allows tuning of 'primary' neutrino nucleon/part interaction.
- Data is old with large statistical errors and often unknown systematic errors (largely flux).



Comparisons to Bubble Chamber data

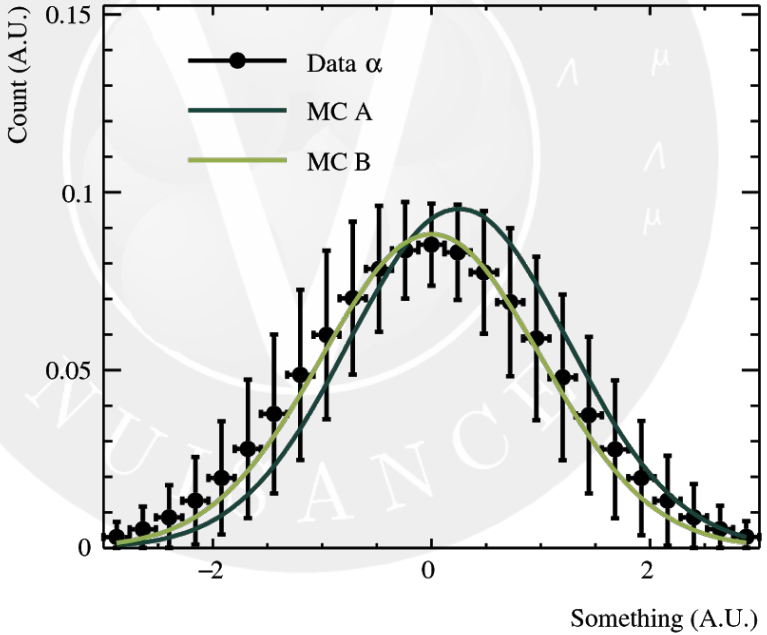
- (quasi-)free of any nuclear effects.
  - Granular reconstruction and unambiguous final state topologies.
  - Allows tuning of 'primary' neutrino nucleon/part interaction.
- Data is old with large statistical errors and often unknown systematic errors (largely flux).
- GENIE v3 provides tuned models through extensive comparisons to a wide range of BC data.

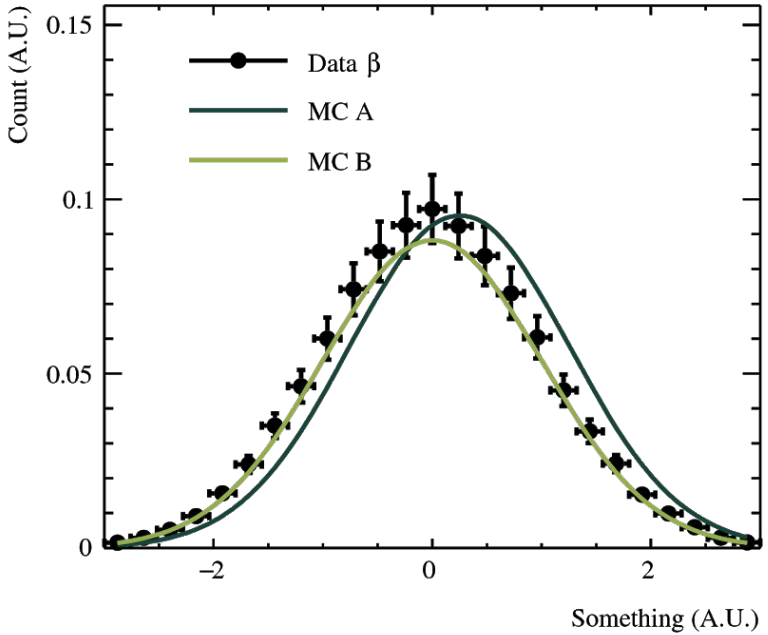


Let's Play... χ-by-eye!

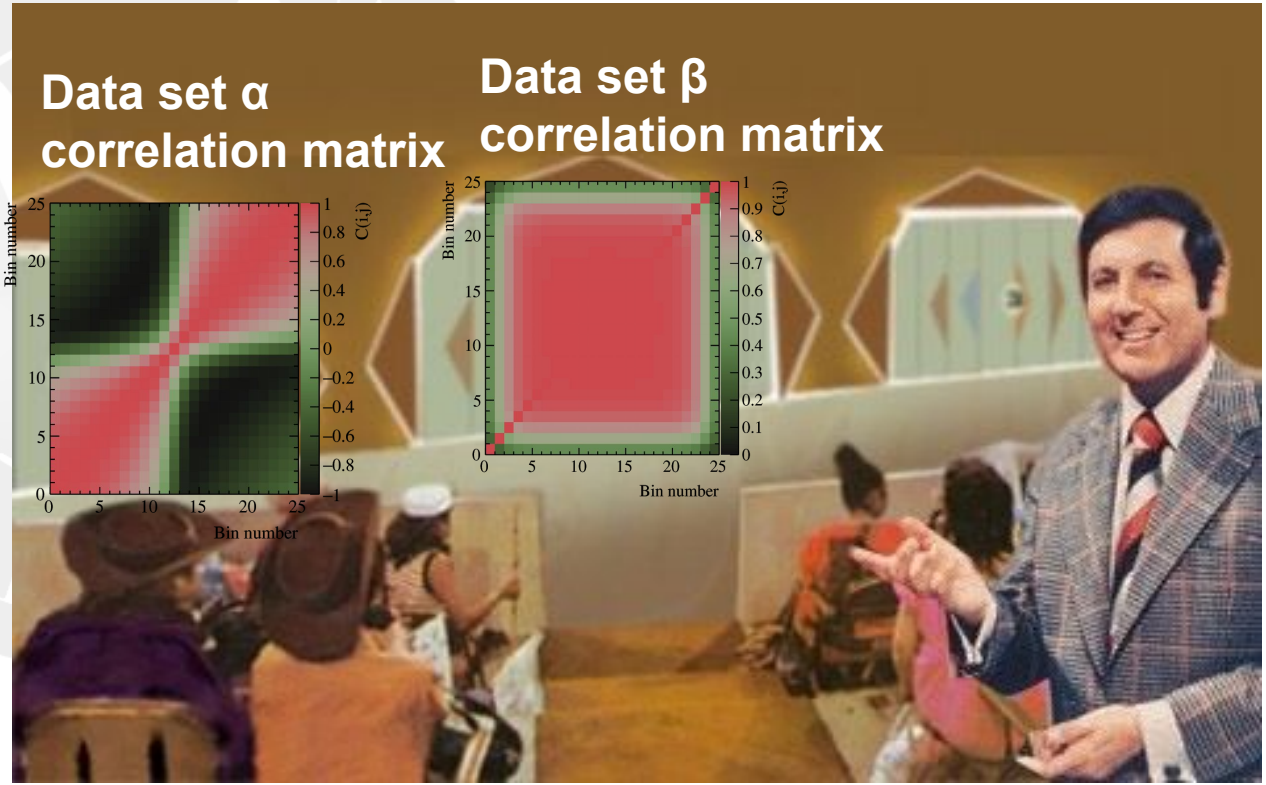
Let's Play... χ-by-eye!

- For each 'data set', guess which MC prediction fits the data better.

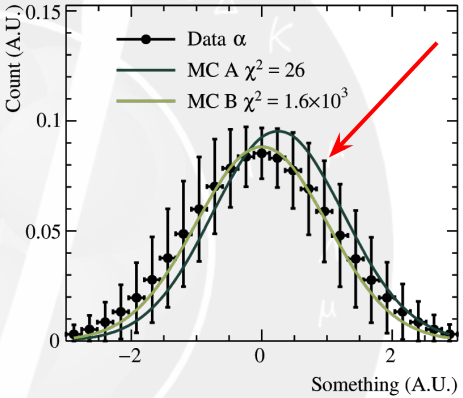


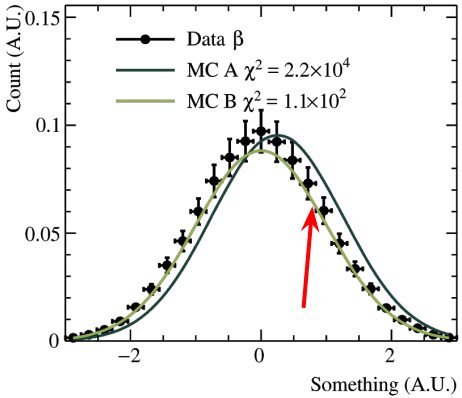


How About Now?

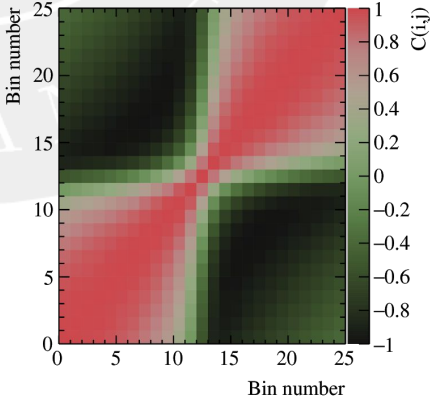


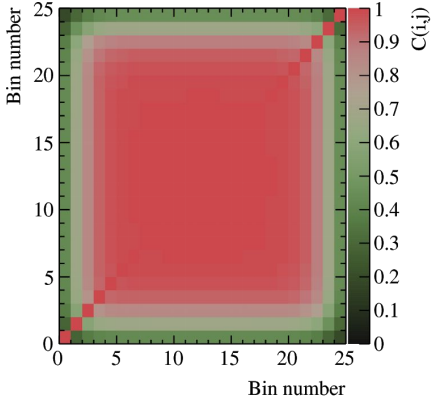
What you expected?





Systematic parameter allows shift in Something. *e.g.* separation energy



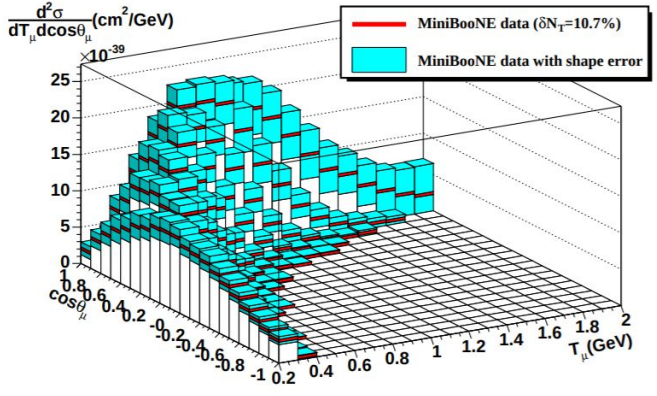


Systematic parameter allows normalization change. *e.g.* flux uncertainty.

Nuclear data: MiniBooNE CCQE

- Data sets without published correlated errors are difficult to use in a global fit.
- MiniBooNE CCQE(like):
  - Many bins, no published error matrix.

PRD 81 092005



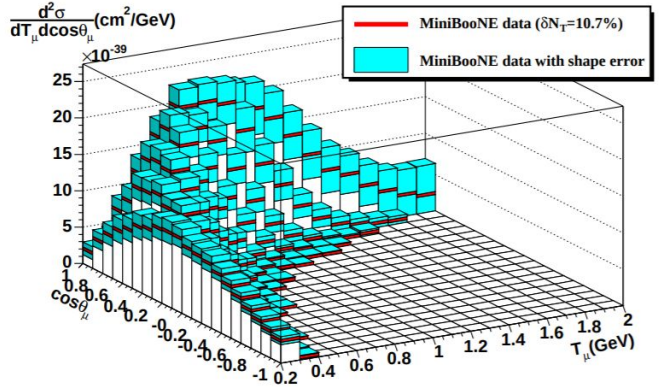
PRD 93 072010

	$\chi^2_{\text{min}}/N_{\text{DOF}}$
All	117.9/228
MINERνA	30.3/13
MiniBooNE	65.7/212
ν	69.1/142
$\bar{\nu}$	46.1/83
MνA vs MB	117.9/228
ν vs $\bar{\nu}$	117.9/228

Nuclear data: MiniBooNE CCQE

- Data sets without published correlated errors are difficult to use in a global fit.
- MiniBooNE CCQE(like):
  - Many bins, no published error matrix.
  - What should the contribution to the global GOF be
    - **Fully uncorrelated**: $\sim \sum_{i \in \text{bins}} (\text{Data} - \text{MC})_i^2$
    - **Fully correlated**: $\sim \sum_{i \in \text{bins}} (\text{Data} - \text{MC})_i^2 / \text{NBins}$
  - In reality, probably somewhere in between.
  - If used naively, will incorrectly dominate a tune **and more data won't help**...
- But, we want to use the information that this data holds, so cannot just ignore it...

PRD 81 092005

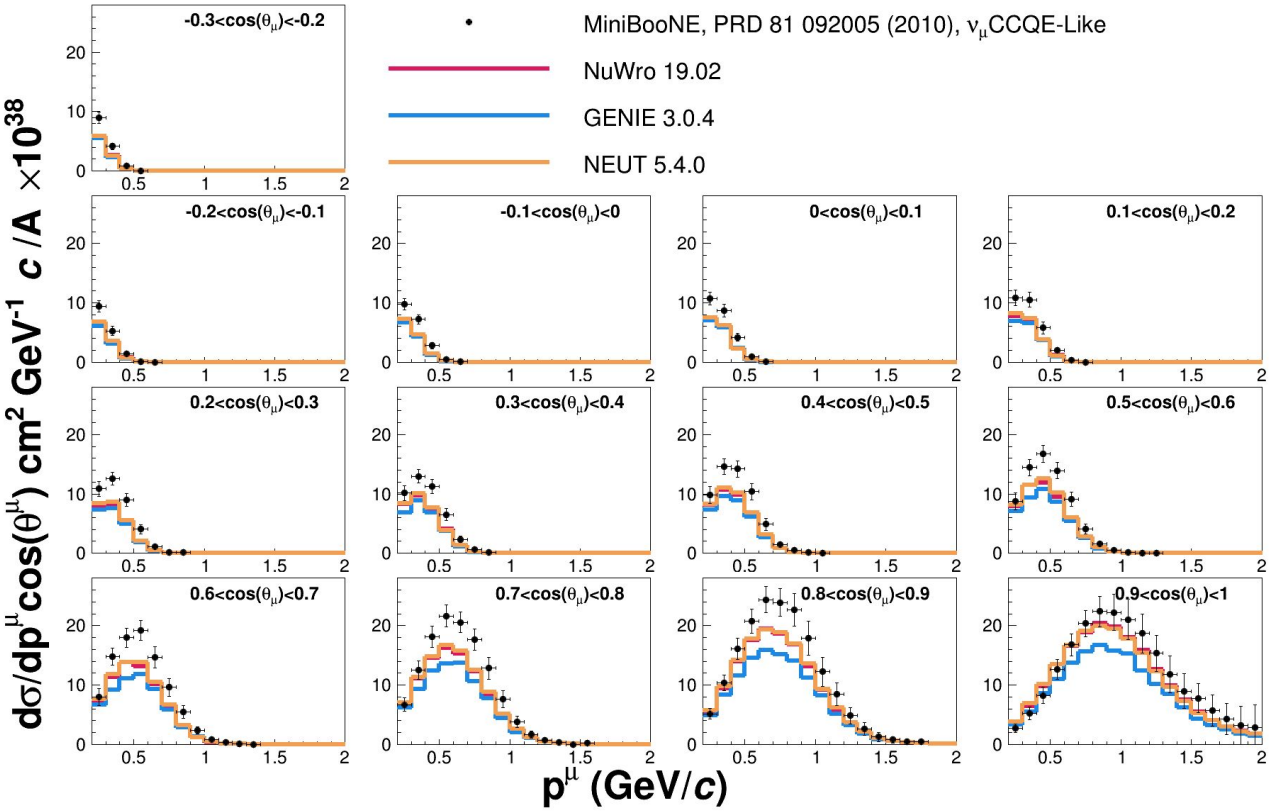


PRD 93 072010

	$\chi^2_{\text{min}}/N_{\text{DOF}}$
All	117.9/228
MINERνA	30.3/13
MiniBooNE	65.7/212
ν	69.1/142
$\bar{\nu}$	46.1/83
MνA vs MB	117.9/228
ν vs $\bar{\nu}$	117.9/228

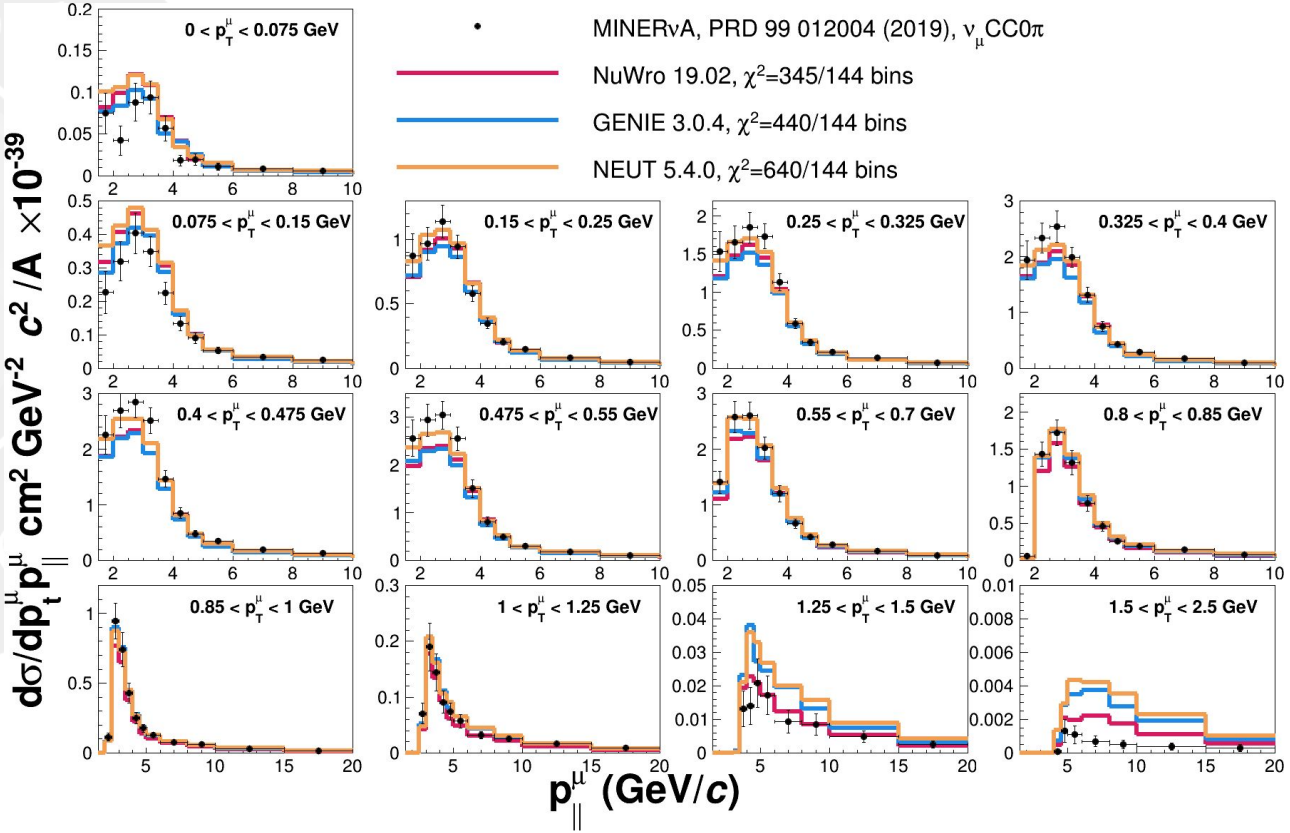
MiniBooNE CCQE-Like

- Not possible to calculate useful GOF, so I'm not going to attempt to…
- The data here is the 'less corrected' CCQE-like data:
  - No pionless delta decay subtraction (subset of MEC diagrams).



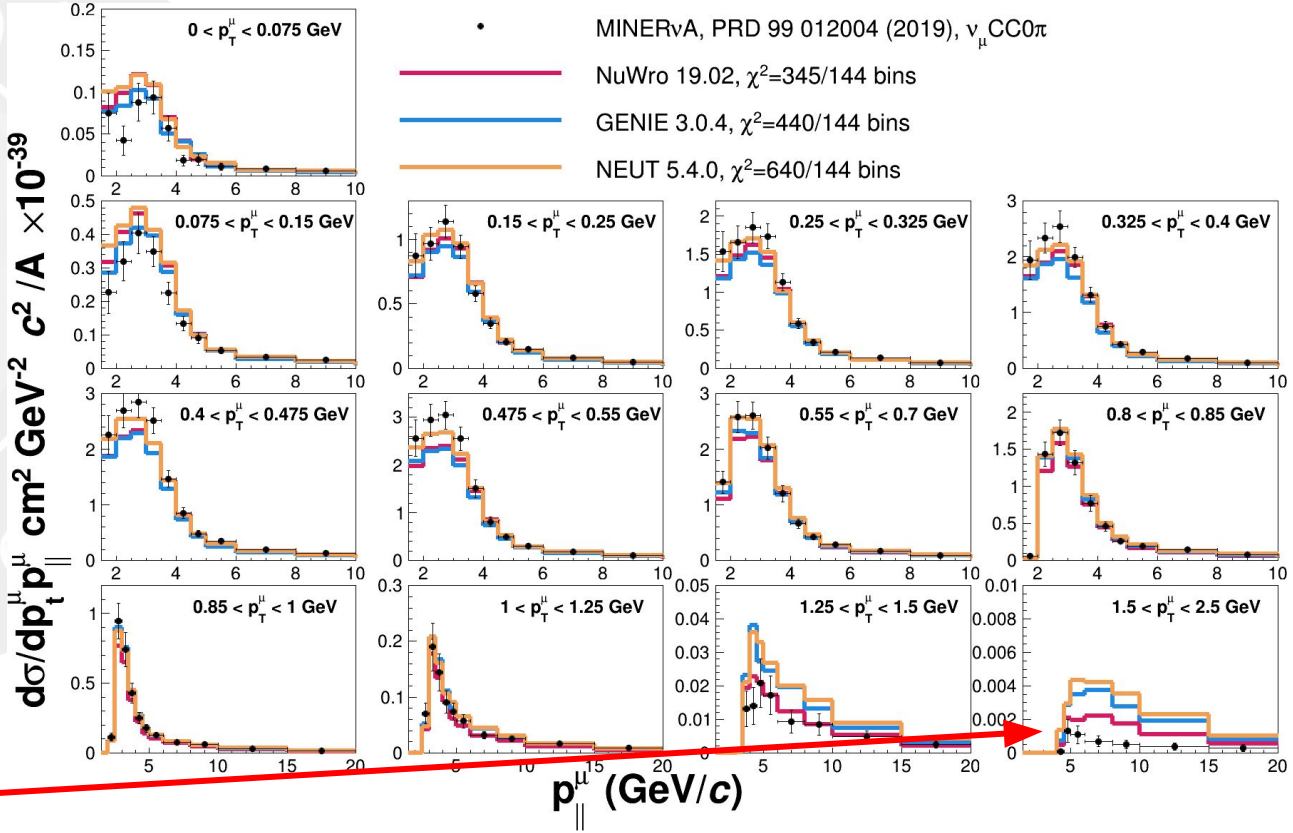
MINERvA 0pi neutrino-mode

- Transverse and longitudinal lepton momenta
  - Kinematics observed by detector: minimal correction required
  - Sensitive to energy and momentum transfer in a known flux



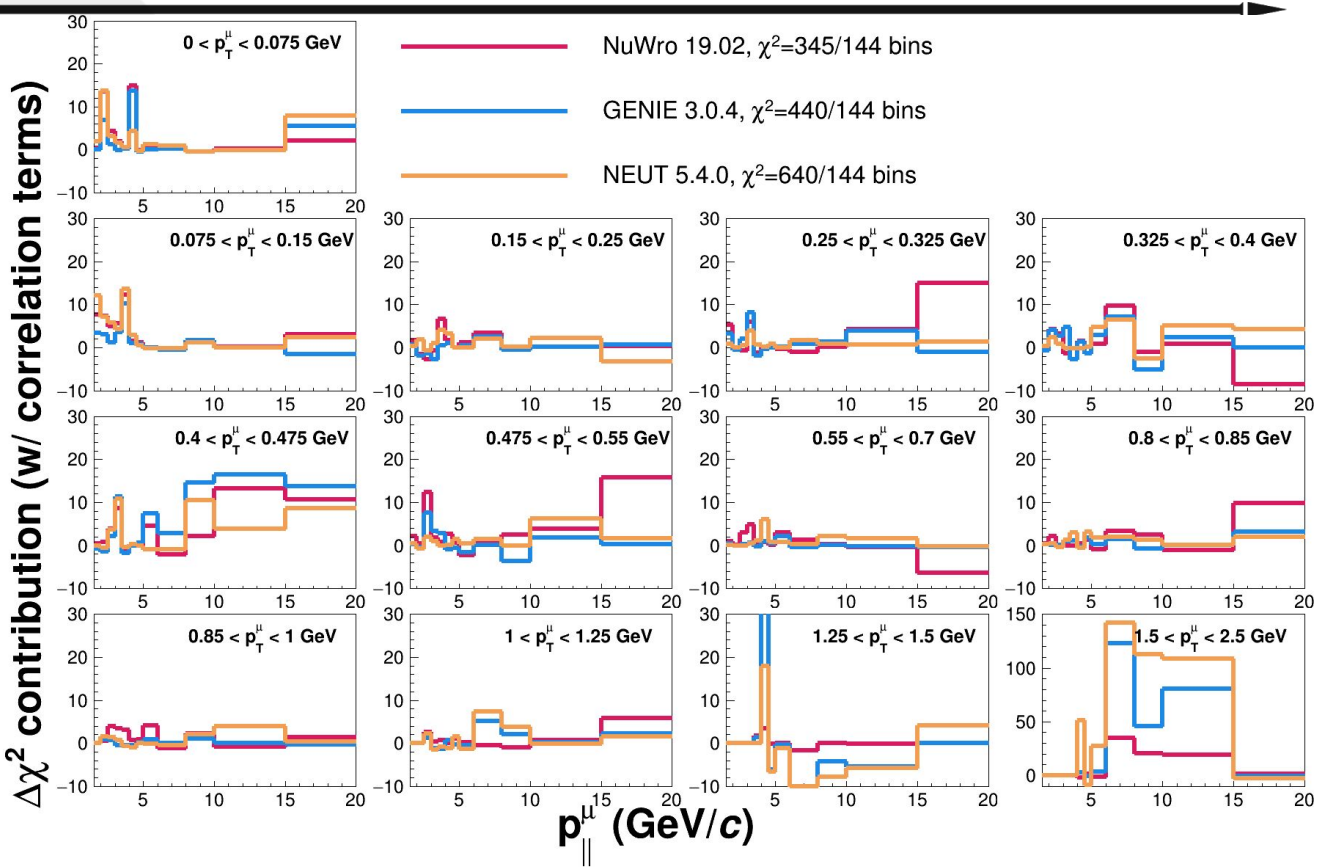
MINERvA 0pi neutrino-mode

- Transverse and longitudinal lepton momenta
  - Kinematics observed by detector: minimal correction required
  - Sensitive to energy and momentum transfer in a known flux
- Predicted ~well for bulk of distribution:
  - Higher angle poorly predicted



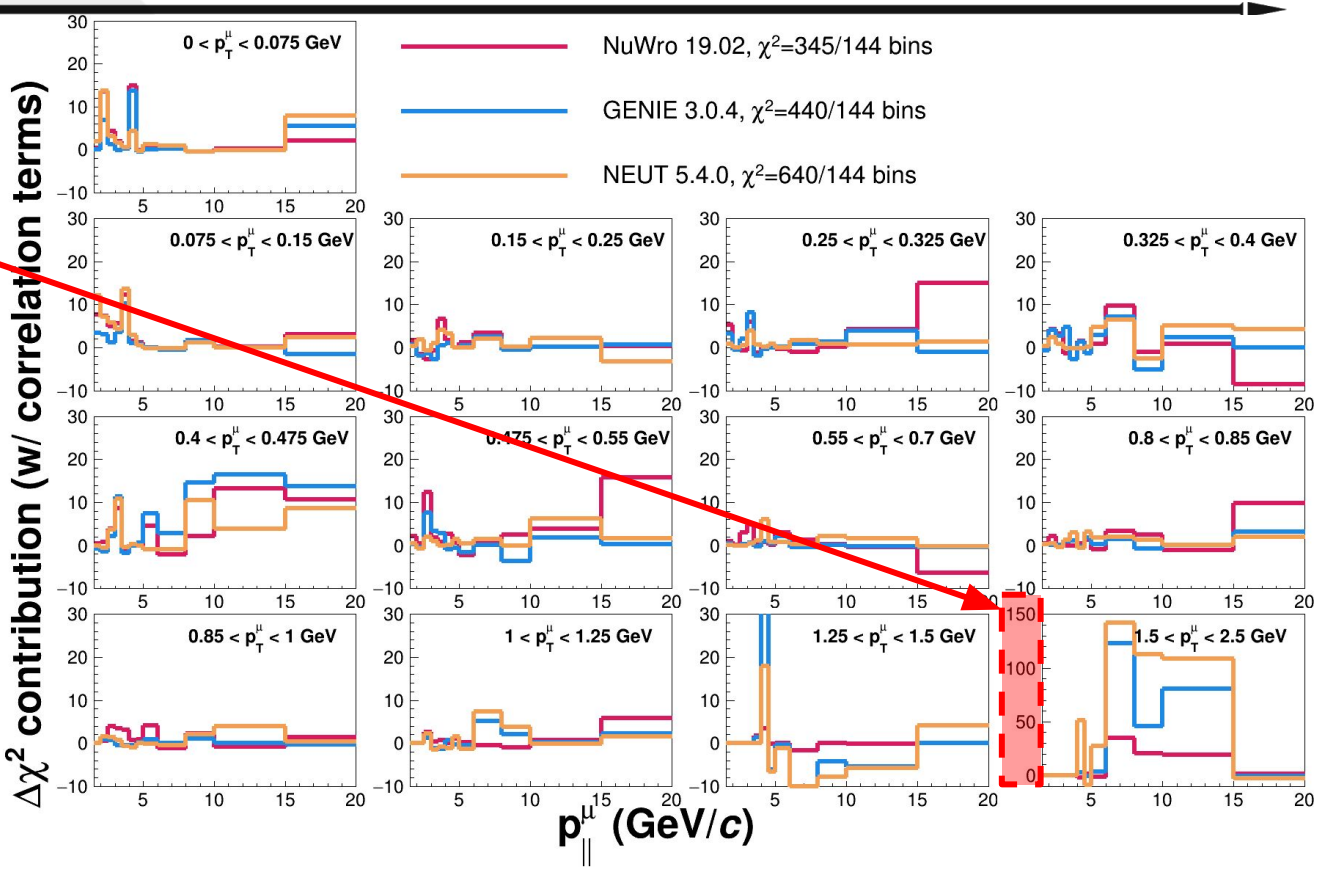
MINERvA 0pi neutrino-mode

- Majority of difference comes from high angle bins.



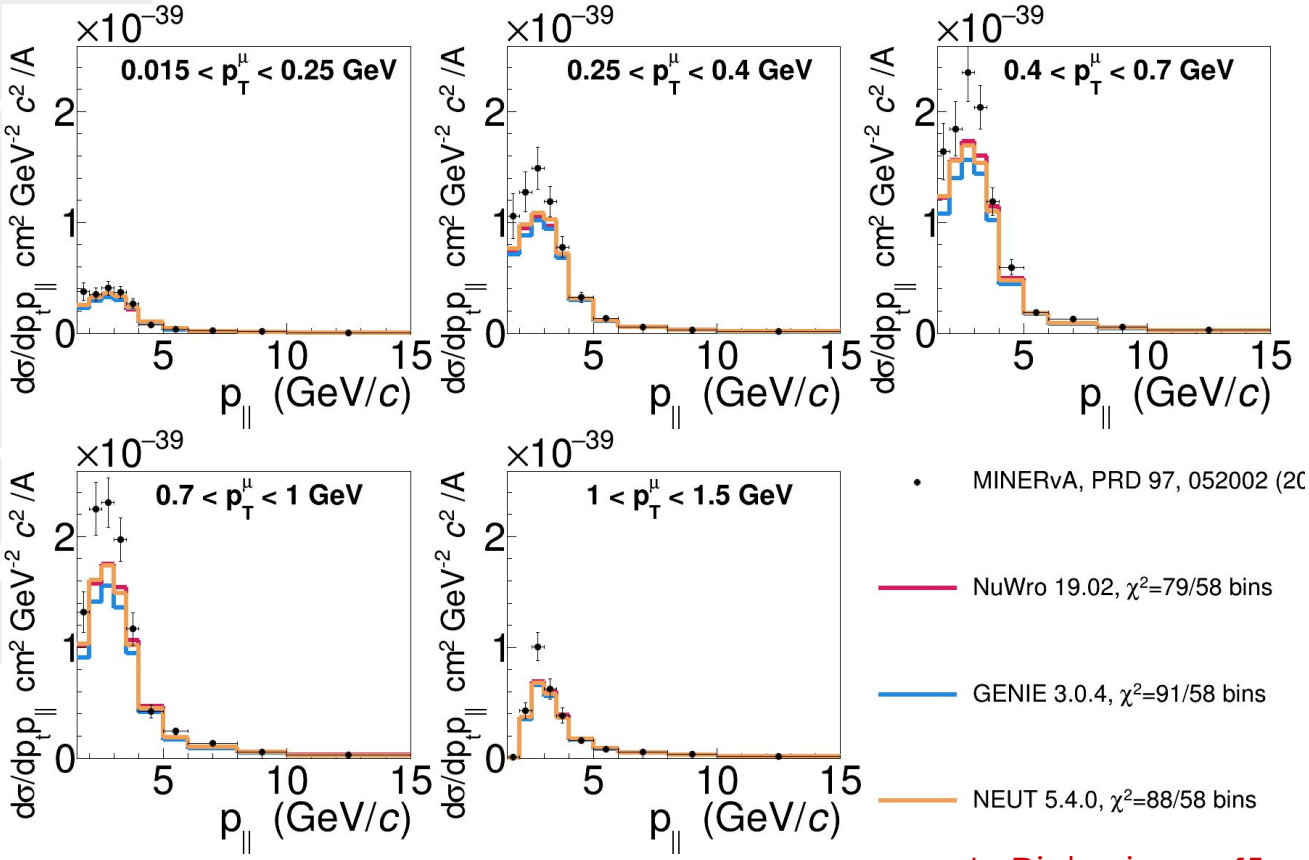
MINERvA 0pi neutrino-mode

- Majority of difference comes from high angle bins.
- Could mask out bad bins, but when to stop p-hacking...



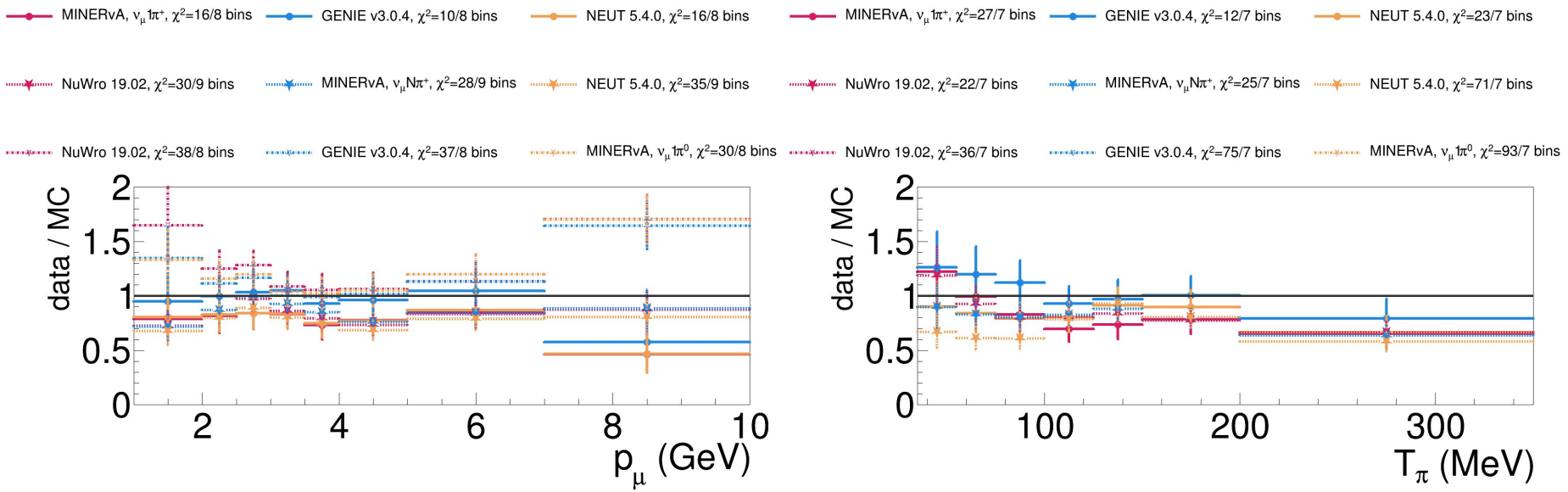
MINERvA 0pi anti-neutrino-mode

- χ-by-eye GOF seems worse (to me) than calculated GOF.
- Possibly because of PPP:
  - Smaller MC normalization can give 'artificially' low χ^2 if uncertainty is not fully characterized.
- Need to be wary of PPP when fitting.



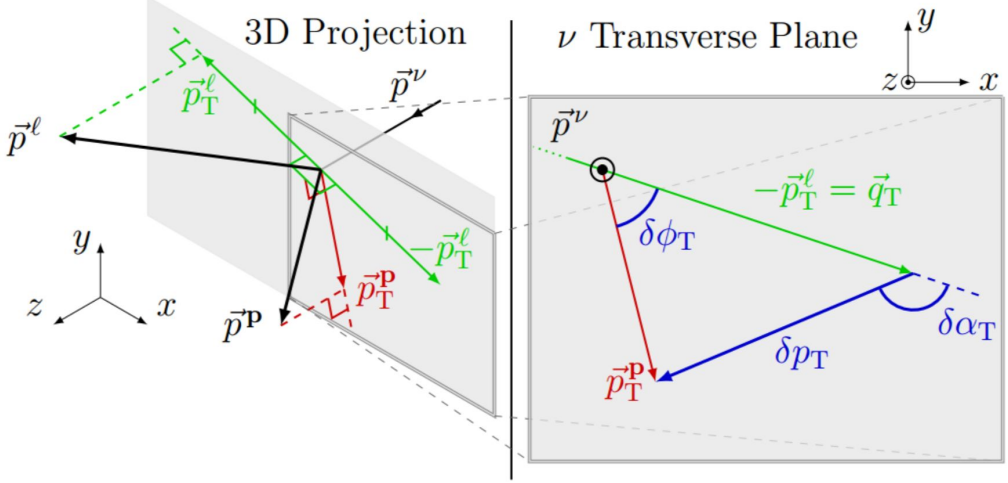
MINERvA 1pi neutrino-mode

- MINERvA have released a number of pion datasets, each with multiple projections
  - Lots of information, much more than shown here.
  - Fairly poorly predicted all around.
- arXiv:1903.01558: discusses some of the difficulties seen fitting these data.



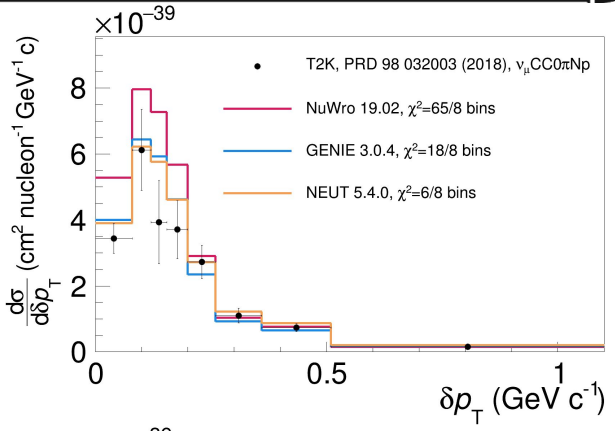
Single Transverse Variables

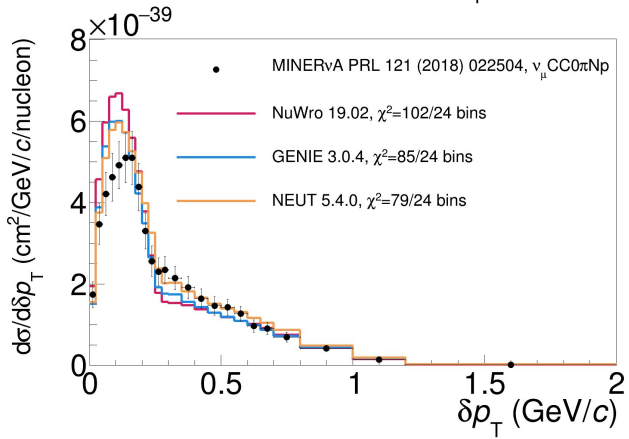
- Recent interest in lepton-hadron correlations:
  - Can be more sensitive to certain effects than lepton-/hadron-only
  - Efficiency/smearing corrections need to be treated with more care.
- Direction/magnitude of momentum imbalance is sensitive to initial and final state effects PRD 98 032003 (2018).



Transverse missing momentum

- Signal phase space cuts chosen for detector capabilities:
  - Results in less model-dependent efficiency correction.
  - T2K:
    - 500 MeV < p_p
    - 250 MeV < pμ, 1 < cos(θ) < -0.6
  - MINERvA:
    - 450 < p_p < 1200 MeV, 0 < θ < 70°
    - 1.5 < pμ < 10 GeV, 0 < θ < 20°

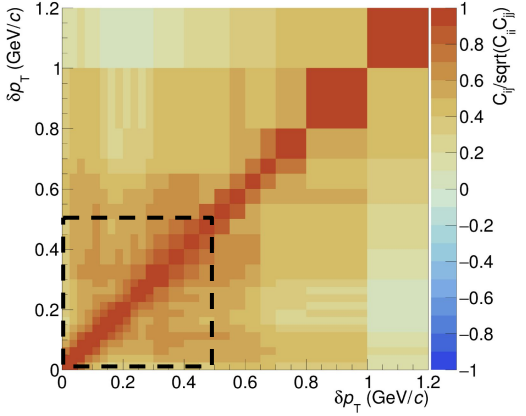


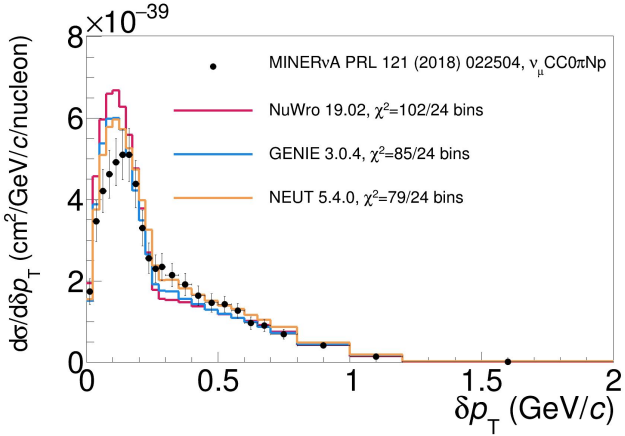


Transverse missing momentum

- MINERvA error matrix provides a tight shape constraint around the peak which drives the high GOF.

MINERvA: PRL 121 (2018) 2, 022504

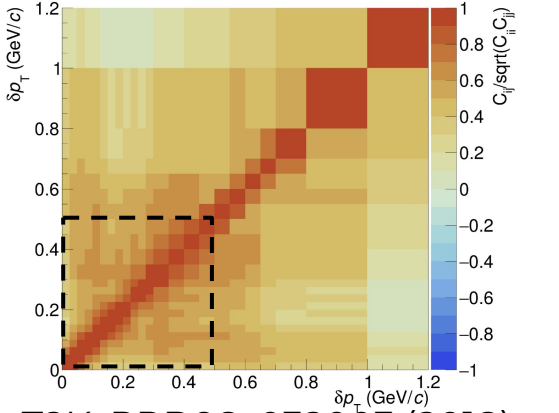




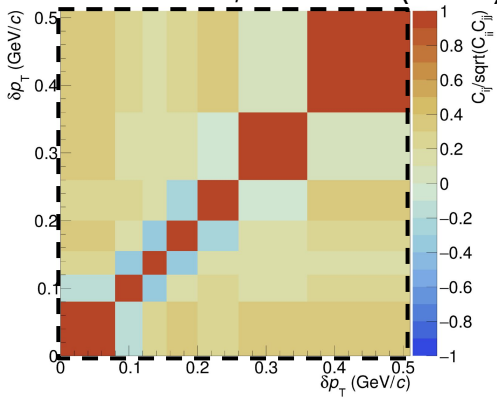
Transverse missing momentum

- MINERvA error matrix provides a tight shape constraint around the peak which drives the high GOF.
- Equivalent matrix for the T2K result exhibits anti-correlations between neighbouring bins:
  - More expected for uncertainties that cause bin migrations.

MINERvA: PRL 121 (2018) 2, 022504

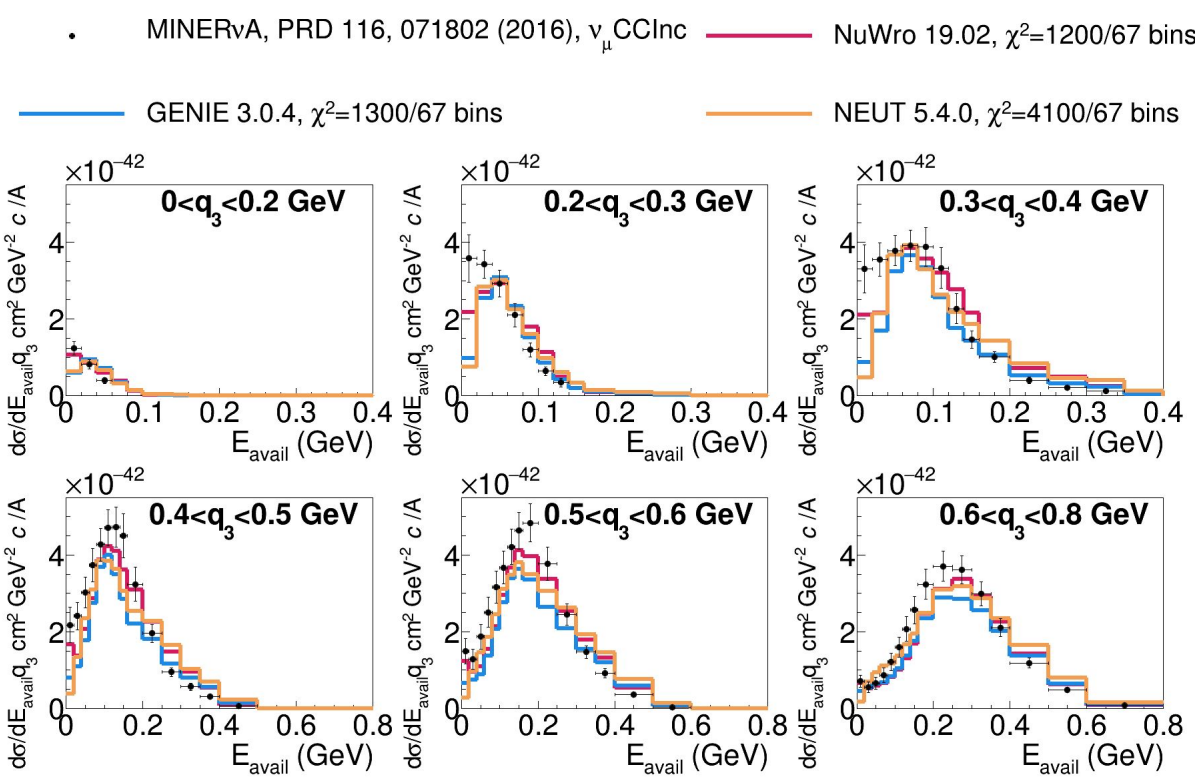


T2K: PRD98, 032003 (2018)



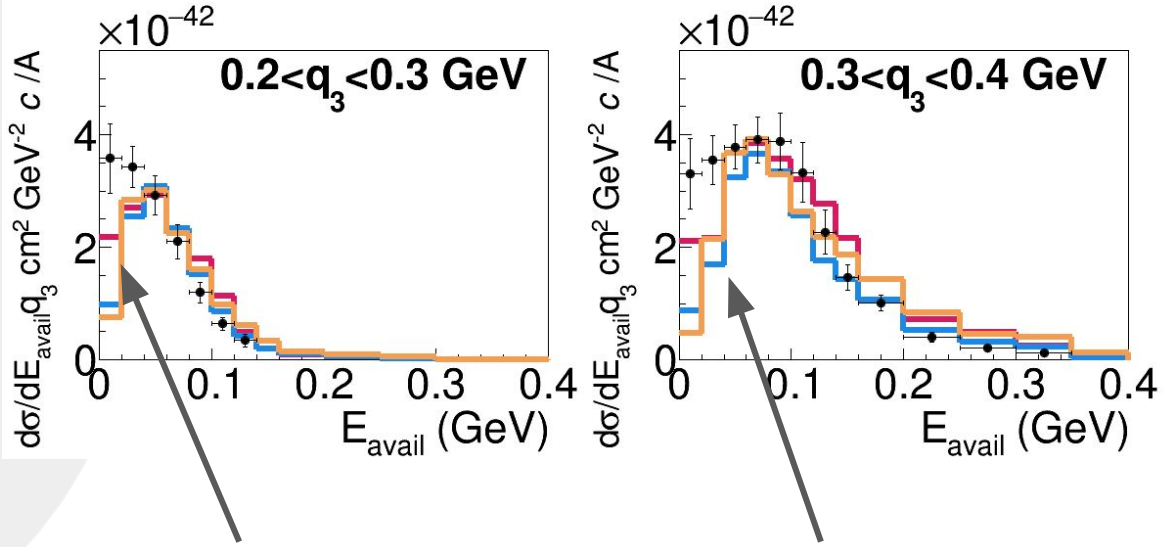
MINERvA CCInclusive: Low recoil

- Interesting/enlightening projections:
  - Inclusive models described by q0/q3
  - But requires model-dependent reconstruction of EAvail and true momentum transfer.
- GOF is awful for all available models:
  - Inconclusive when comparing one bad fit to another bad fit.



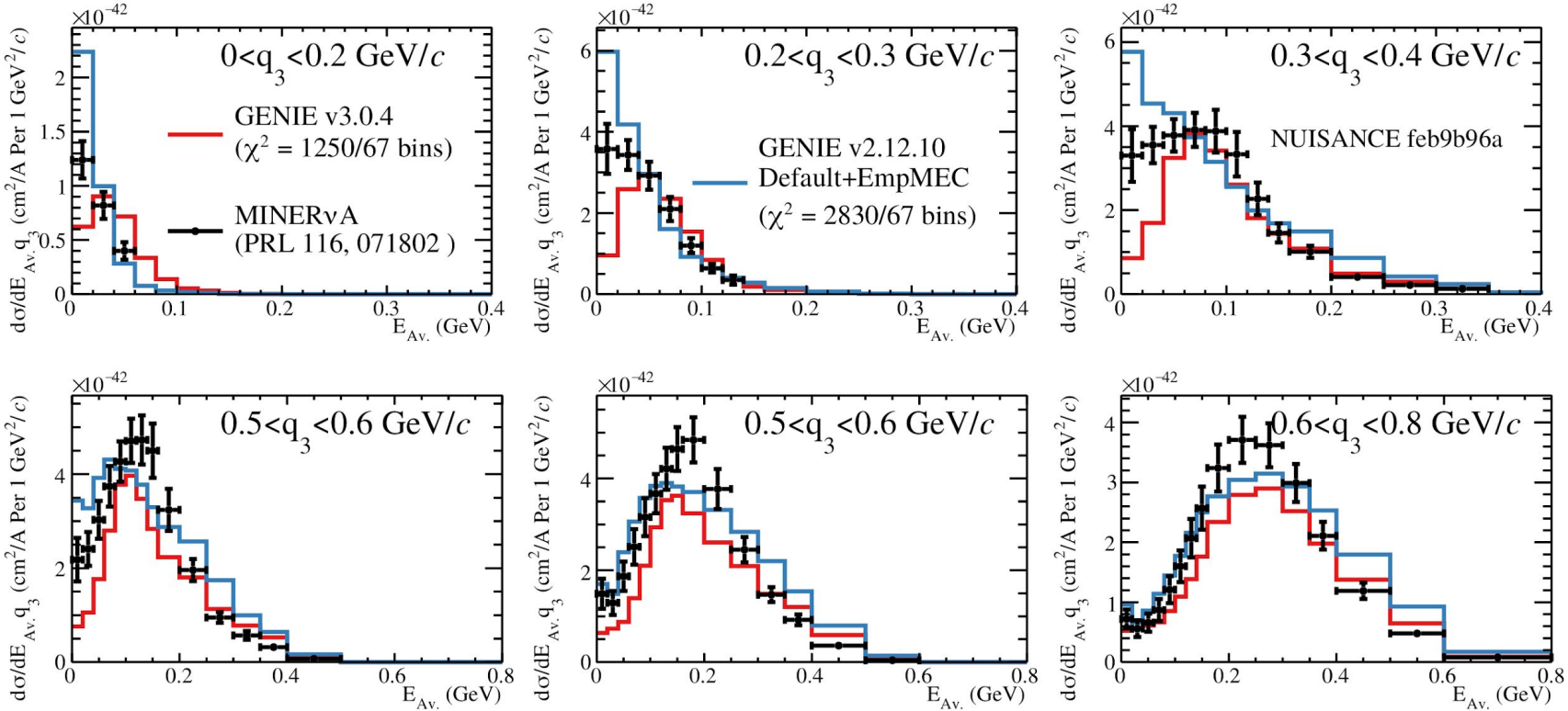
MINERvA CCInclusive: Low recoil

- Interesting/enlightening projections:
  - Inclusive models described by q0/q3
  - But requires model-dependent reconstruction of EAvail and true momentum transfer.
- GOF is awful for all available models:
  - Inconclusive when comparing one bad fit to another bad fit.



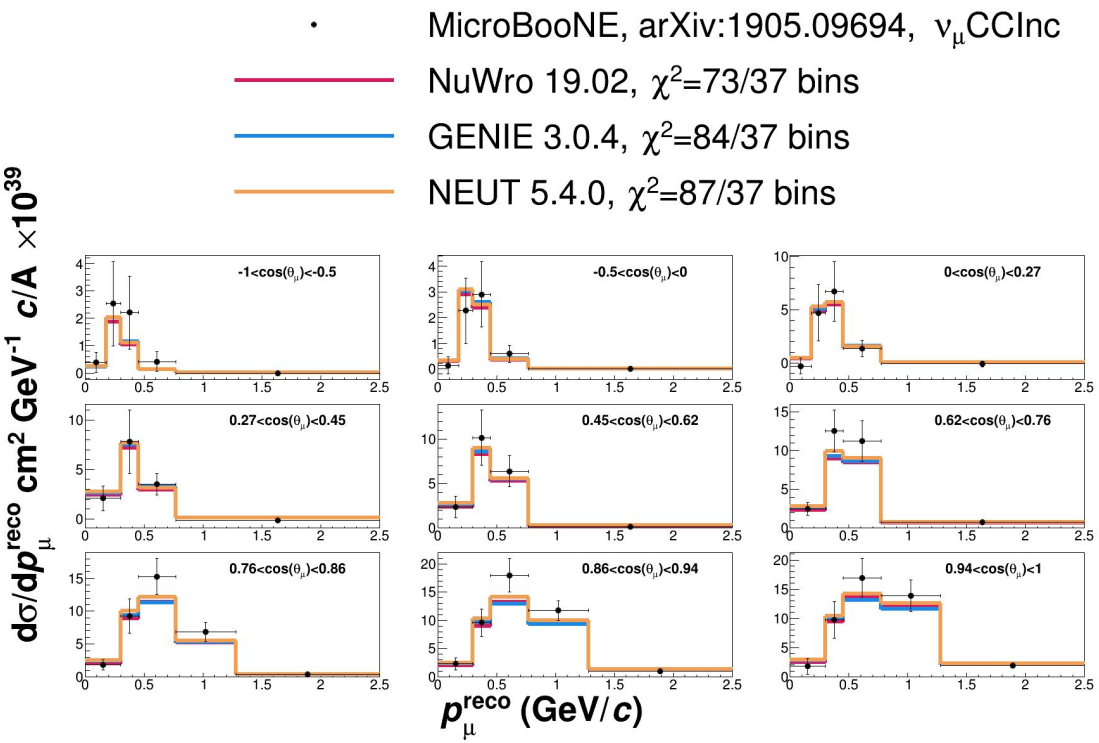
Low energy transfer region especially poorly predicted.

MINERvA CCInclusive: Low recoil



Comparisons to Nuclear data: MicroBooNE

- Liquid Argon is a key detector technology for the next generation of experiments
  - Need to understand neutrino interactions on Ar40 target.
- Data release:
  - Reconstructed distributions
  - True→reco folding matrix
- Potentially useful technique to reduce model bias in published data.

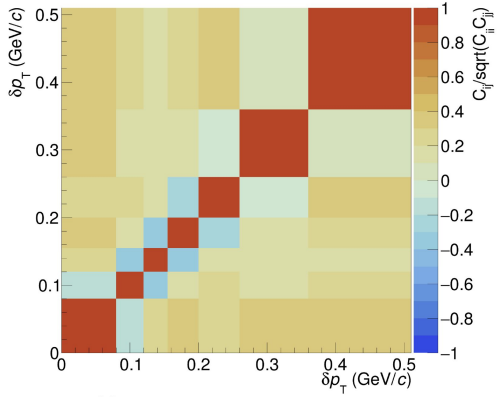


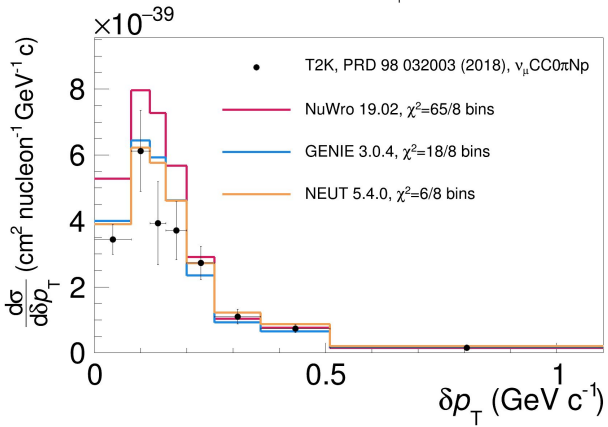
What is needed from Data Measurements

- Minimize model bias while maximising efficacy of data:
  - Lots of recent and rediscovered work on robust statistical techniques to avoid bias in unfolding.
  - Thoughtfully chosen observable event projections:

What is needed from Data Measurements

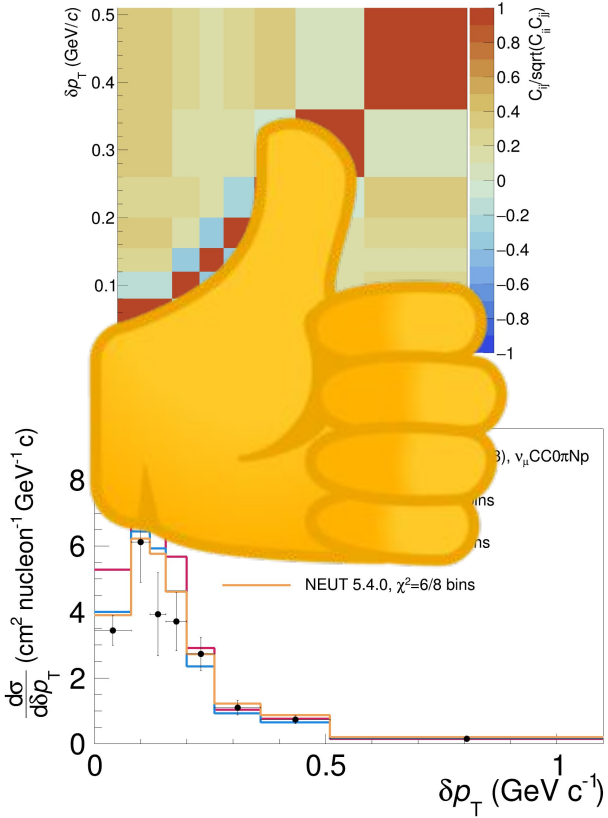
- Minimize model bias while maximising efficacy of data:
  - Lots of recent and rediscovered work on robust statistical techniques to avoid bias in unfolding.
  - Thoughtfully chosen observable event projections:
    - What can a detector measure with good, well-understood efficiency?
    - What projections require minimum interaction model-dependent corrections?
  - Sensible phase space restrictions.





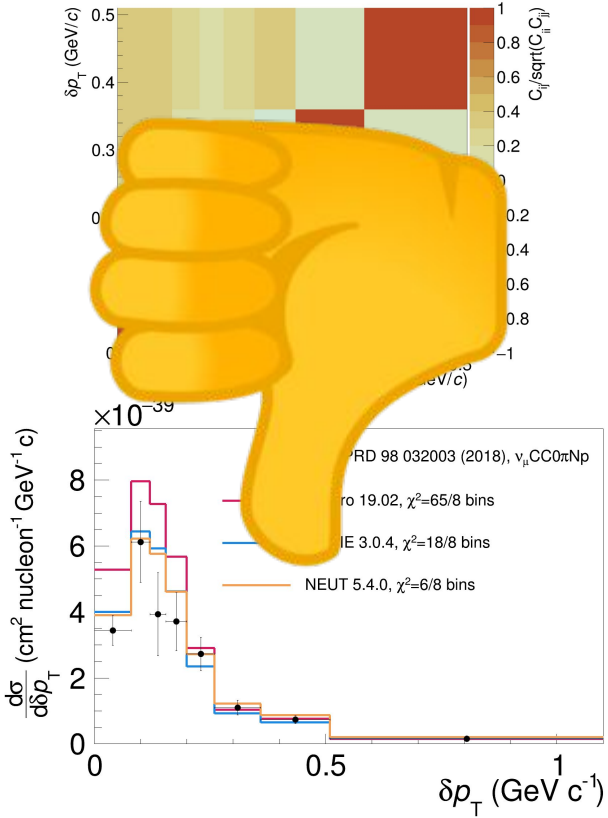
What is needed from Data Measurements

- Minimize model bias while maximising efficacy of data:
  - Lots of recent and rediscovered work on robust statistical techniques to avoid bias in unfolding.
  - Thoughtfully chosen observable event projections:
    - What can a detector measure with good, well-understood efficiency?
    - What projections require minimum interaction model-dependent corrections?
  - Sensible phase space restrictions.



What is needed from Data Measurements

- Minimize model bias while maximising efficacy of data:
  - Lots of recent and rediscovered work on robust statistical techniques to avoid bias in unfolding.
  - Thoughtfully chosen observable event projections:
    - What can a detector measure with good, well-understood efficiency?
    - What projections require minimum interaction model-dependent corrections?
  - Sensible phase space restrictions.
  - **Publish correlated errors wherever possible**:
    - **Between projections**
    - **Between datasets.**



Future: 1

- Last few years seen increase in sophistication of 0pi analyses
  - Lepton/hadron correlations
  - Less Model-dependent selections and projections
  - Would be very useful to see similar renaissance in pion production datasets.
- Future MicroBooNE (and SBND) data sets will be critical for model builders to benchmark and develop before DUNE and Fermilab Short Baseline program.

Future: 2

- These last two years have seen an uptick in model development:
  - GENIE tuning, v3, NEUT and NuWro model developments, ECT* Trento workshops
  - Lots of progress due to closer interaction with theory community, need to continue!
- But given how much LBL programs will rely on the predictions and uncertainties, the community is quite under person-powered…
  - Plenty of room for important work and novel intellectual contribution
- Can learn a lot of the necessary nuclear physics from electron scattering: GENIE + NuWro have e-A modes, ongoing work by **e4nu**.
- See what GiBUU has to say for itself…

Summary

- The loftiest goals of neutrino oscillation physics depend on the accuracy of event generator predictions and associated uncertainties.
- Recent $\nu_\mu \rightarrow 0\pi$ data releases have been more statistically robust, but GOF between available models is generally poor
  - Room for improvement in generator predictions, xsec analyses and data releases and global fitting methodology.
  - Correct, correlated errors are a comparators best friend!
- More recent work on removing assumptions in generator factorization and implementing state-of-the-art predictions is promising!

Thanks for listening

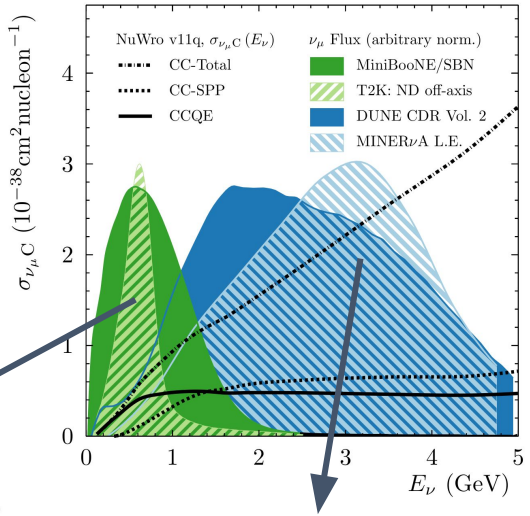
L. Pickering

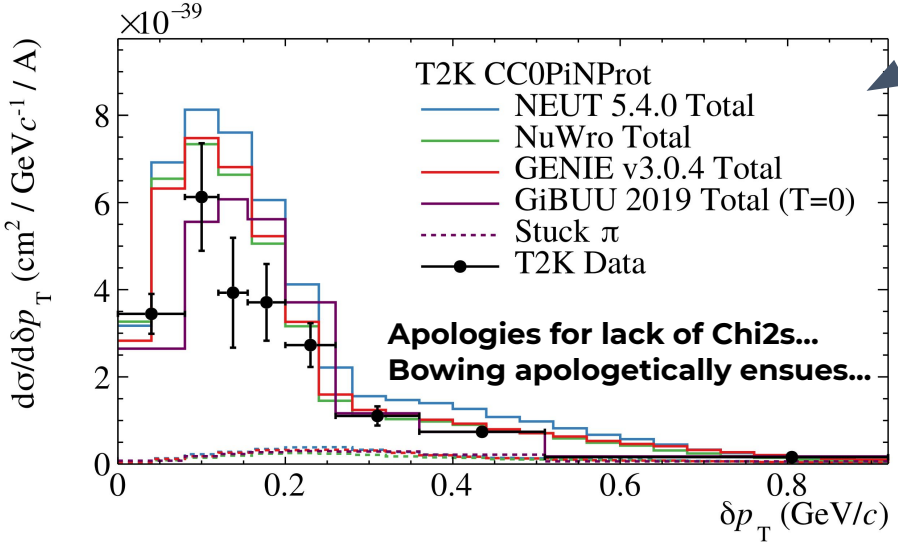
MICHIGAN STATE UNIVERSITY

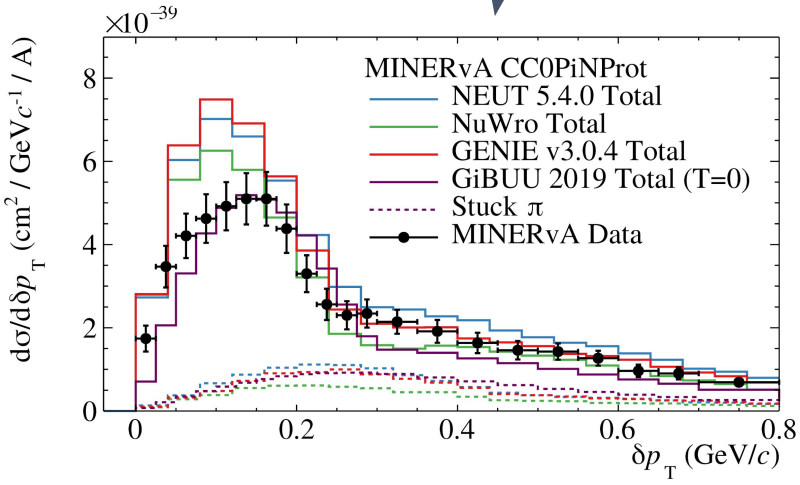
https://doi.org/10.1016/j.physrep.2018.08.003

Data Comparison: δp_T

- T2K: 1802.05078
- MINERvA: 1805.05486
- (GENIE norm may not be quite right to a few %, its fine for here, but probably not best to show these plots as is elsewhere)



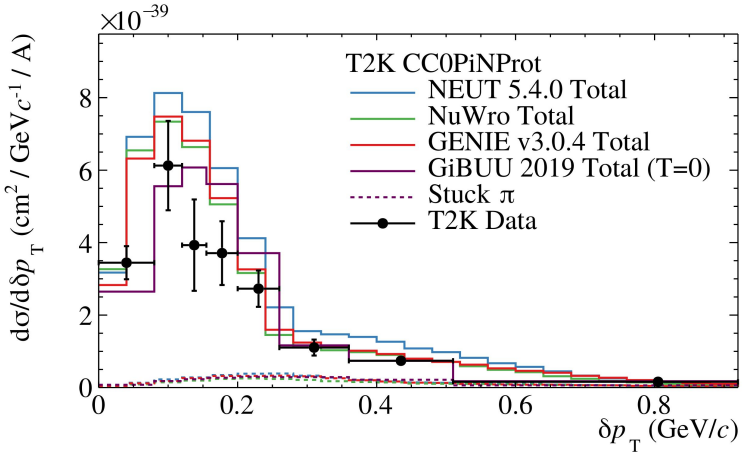




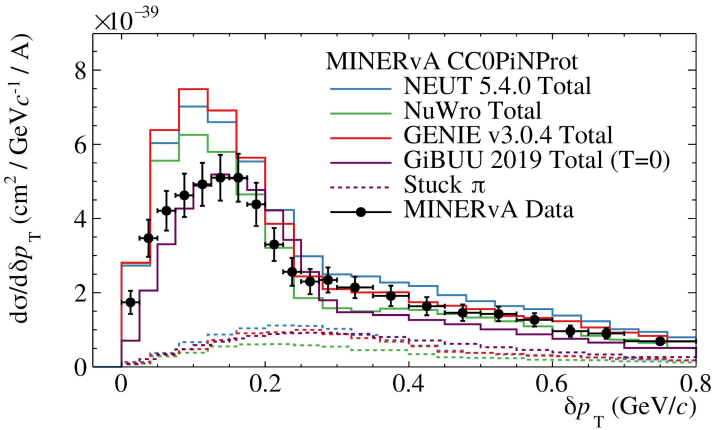
Signal definitions

- T2K: 1802.05078
- MINERvA: 1805.05486
- (GENIE norm may not be quite right to a few %, its fine for here, but probably not best to show these plots as is elsewhere)

500 MeV < pp
250 MeV < pmu, 1 < cos(theta_mu) < -0.6

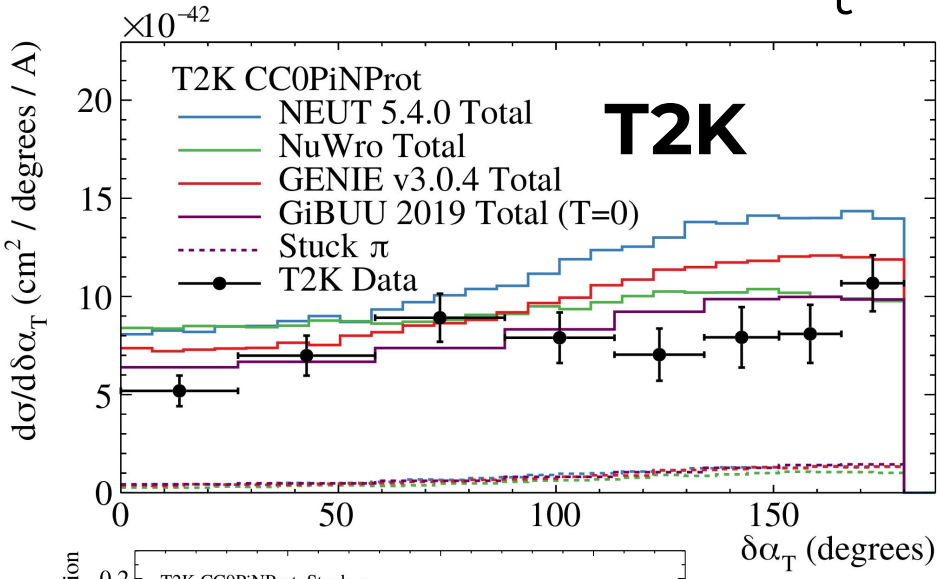


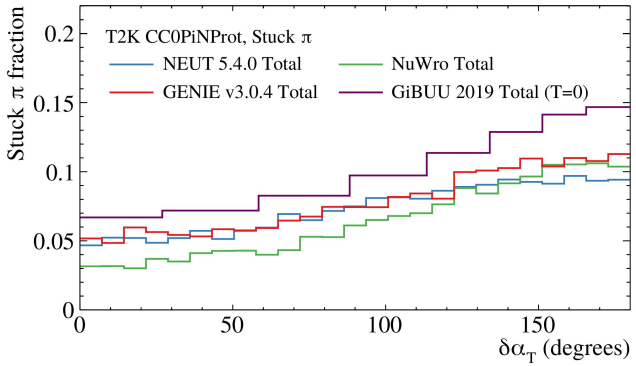
450 < pp < 1200 MeV, 0 < theta_p < 70°
1.5 < pmu < 10 GeV, 0 < theta_mu < 20°

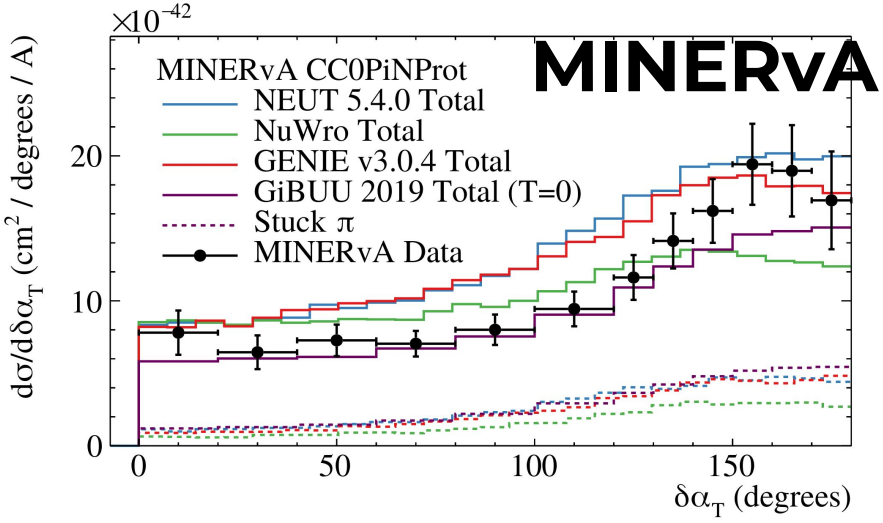


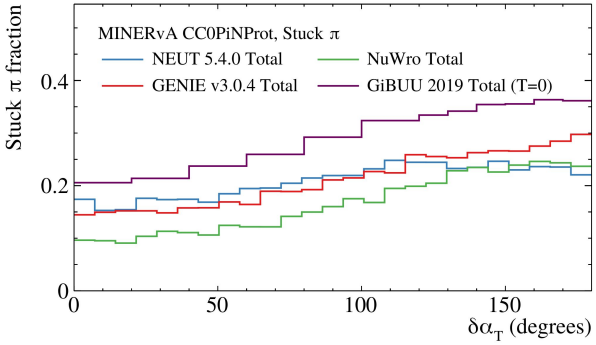
Stuck pion rate: $\delta\alpha_t$

QEL-pure at low δαt
FSI and stuck pion rich at higher δαt





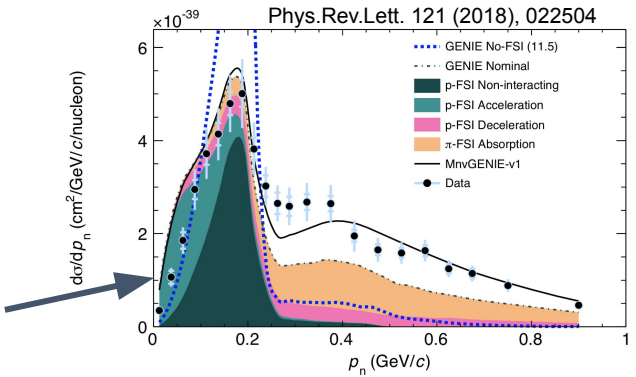


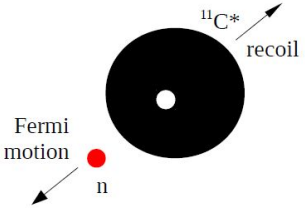


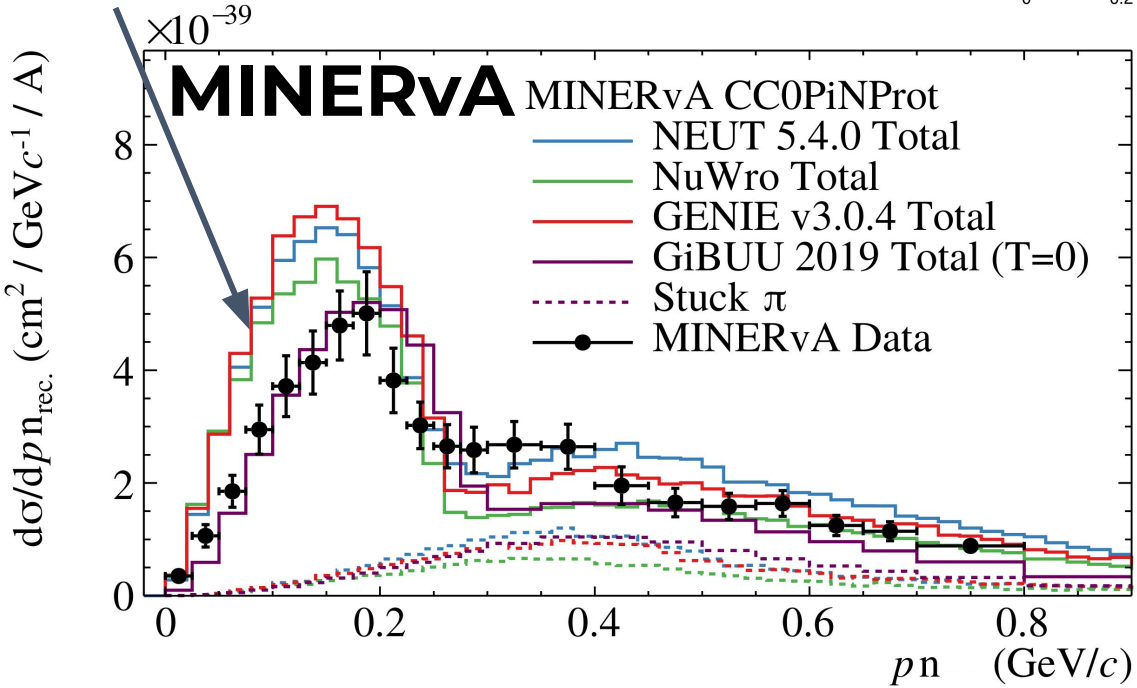
pn

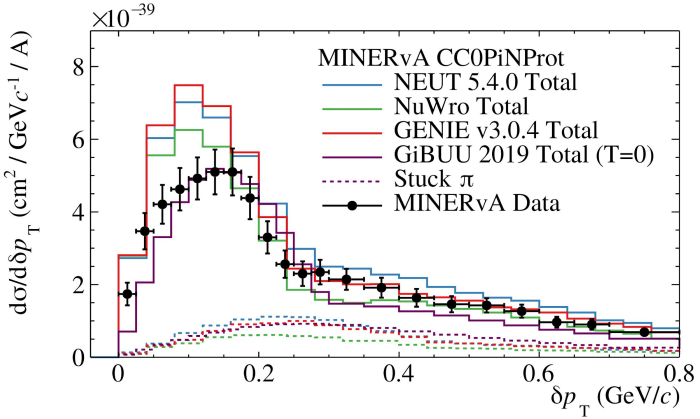
Phys.Rev. C95 (2017) 065501,
see definition in BACKUP

- S. Dolan: Relative to dpt, stuck pions more away from QEL peak **(all non-QE, see later, backup)**
- GENIE V304 below no longer has elastic hA, less lumpy



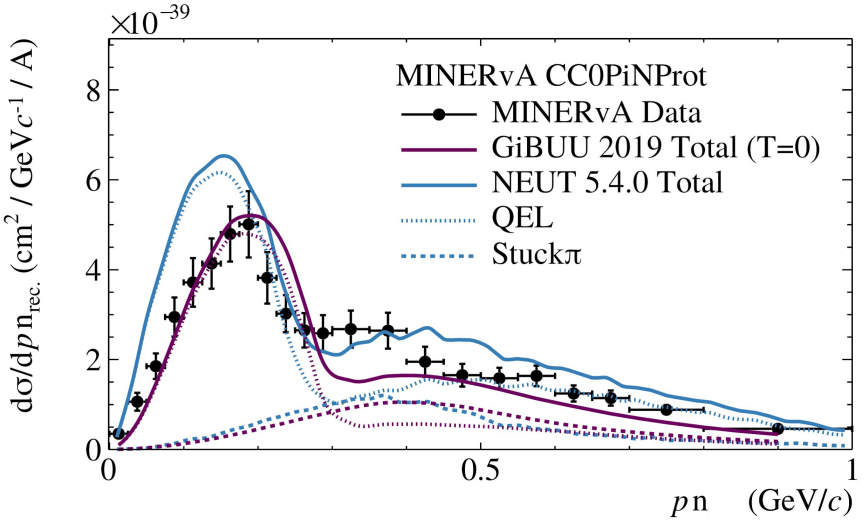


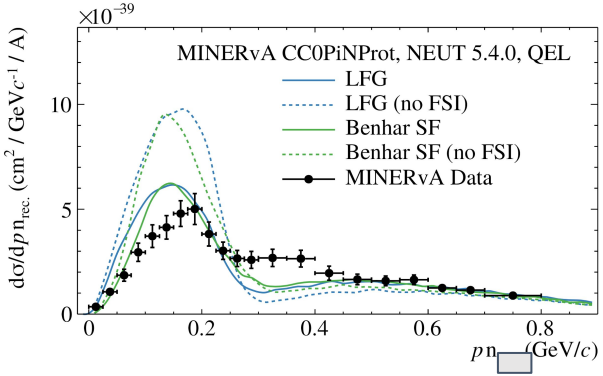


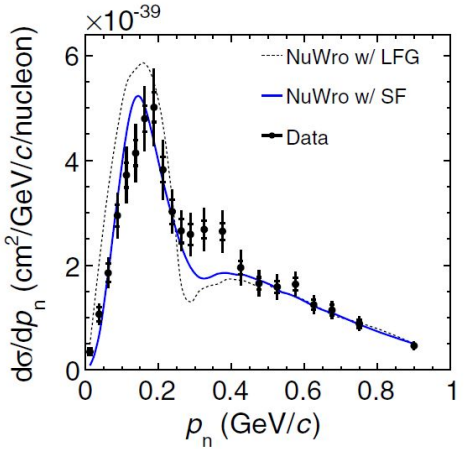


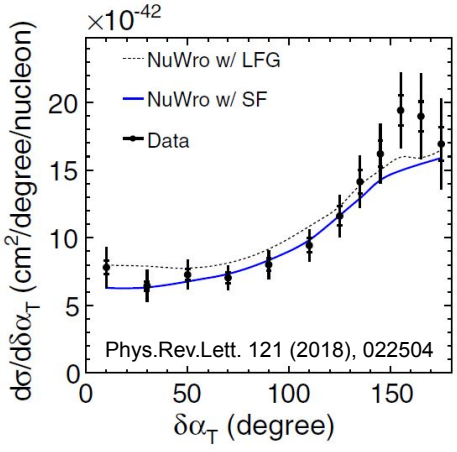
More pn

- Also wanted to look at stuck pi vs. 2p2h
  - GiBUU predicts no second peak for QEL, but NEUT does.
- And FSI/Nuclear momentum/binding model changes:
  - LFG/SF in NEUT qualitatively similar, **contrary to NuWro**
  - FSI mostly interacts with signal selections
- May be interesting to look at energy evolution as well (see last BACKUP)





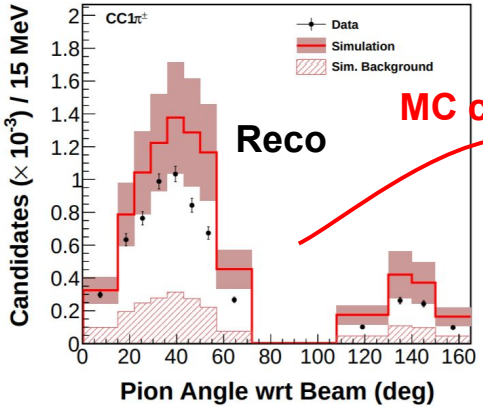




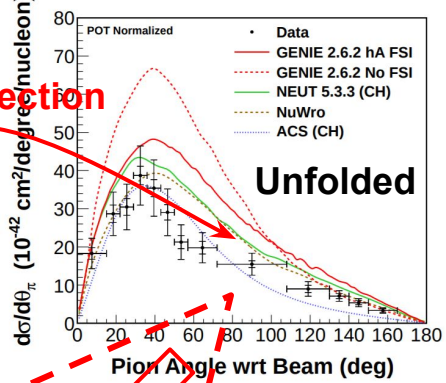
MINERvA 1pi neutrino-mode

- For the charged pion analyses:
  - ~100% efficiency correction at high angle.
  - Where is this ‘MC fill-in’ in other distributions?
- Upcoming re-analysis still no phase space cuts.
- No covariance between distributions ($p\mu$, $\theta\mu$, $T\pi$, $\theta\pi$, Q^2) or samples ($\pi+$, $\pi0$, ν, $\bar{\nu}$):
  - Difficult to consistently use together in a meta-analysis.

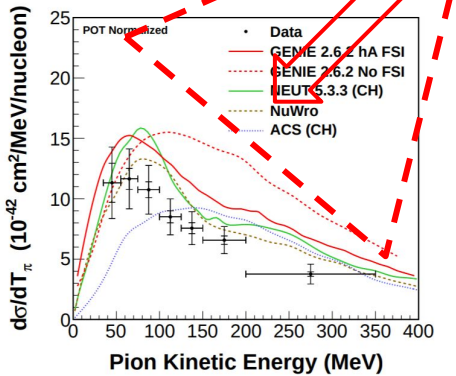
PRD 92 092008



MC correction

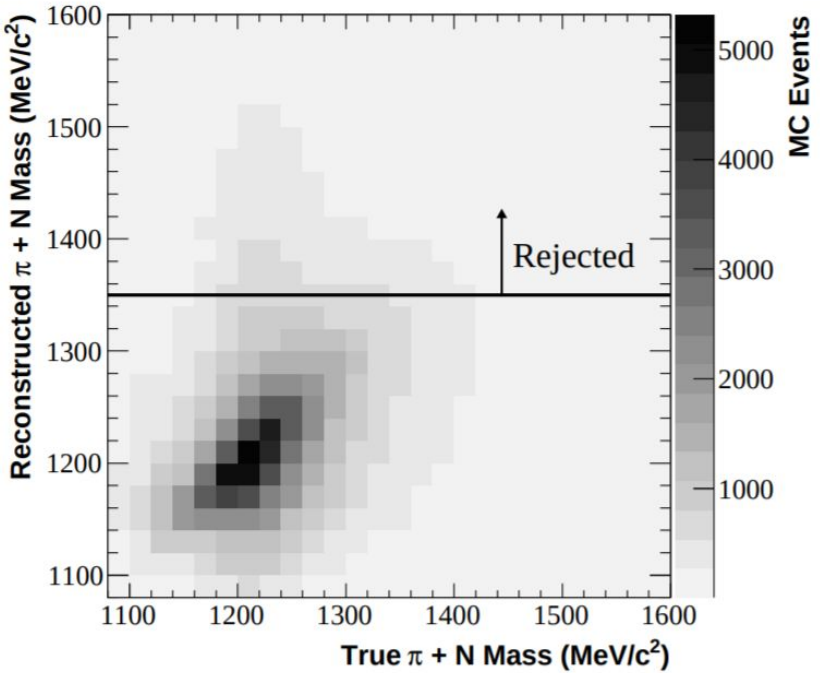


Where is MC in different projection…?



MiniBooNE 1Pi+

- Rejection only in selection, not signal definition:
  - Will be efficiency corrected back with NUANCE-calculated efficiency.
  - Better to include analysis cuts in both signal and selection where possible, then handle new out-of-phase space backgrounds, but smaller, less model dependent efficiency corrections.



MINERvA: Initial state neutron momentum

- Momentum imbalance in all three dimensions is sensitive to initial state fermi nucleon momentum distribution.
  - GOF is poor for all models.