Using Neutrino Data Releases
NUISANCE Experiences

Patrick Stowell, Luke Pickering,
Clarence Wret, Callum Wilkinson

08/09/17



Introduction

- Core NUISANCE developers have been using MINERvA data since you started releasing it.
- My masters project used your very first CCQE data!
- Between us we have ~10-13 PhD years experience using MINERvA cross-section data.
- Have almost all of your datasets inside NUISANCE

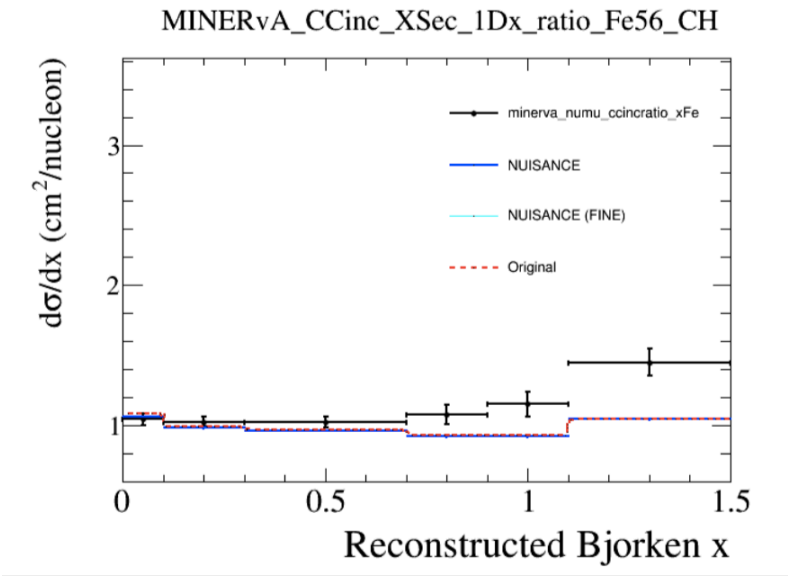

THE
GOOD

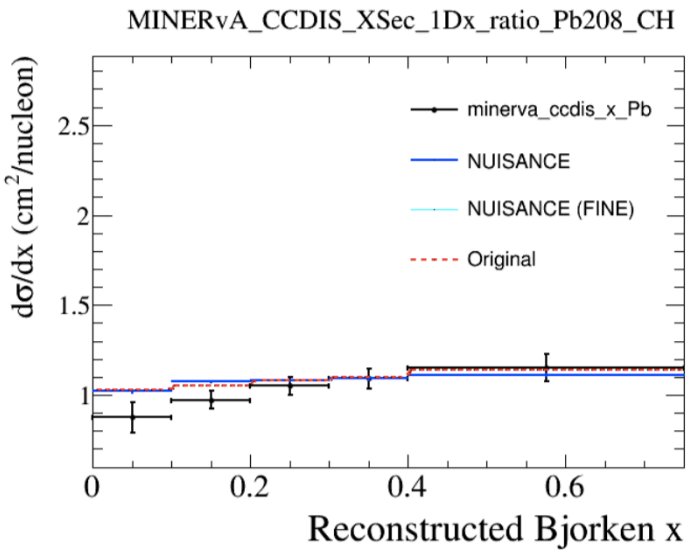
1. Covariances

- MINERvA always releases full covariance information for each distribution which is awesome.

- Some datasets have nu/nubar cross-correlations too. Are there limitations to why this can't be done for all datasets?

2. Reco Data

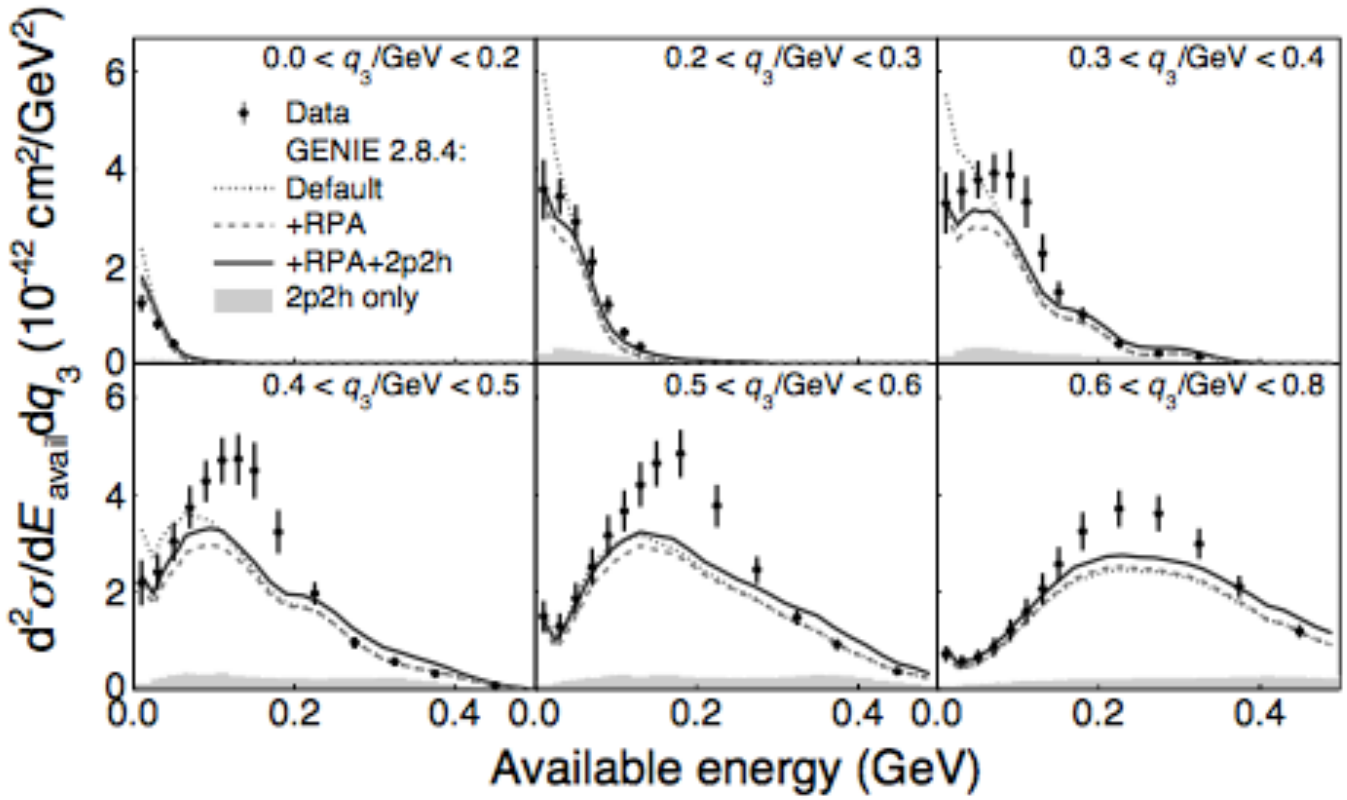




- We try very hard to remove model dependence in our fits.
- Smearing our MC and comparing to the reco data is one way to reduce model dependence.

- Already implemented some of your reco data into NUISANCE.

3. Alternative Kinematics



- I personally think that distributions other than just pure muon kinematics (e.g. pmu-costheta) are going to be the key to breaking the degeneracy between different interaction models.

4. 2D Data Releases

	0.2	0.3	0.4	0.5	0.6	0.8
0.02	0	3	10	22	36	51
0.04	1	4	11	23	37	52
0.06	2	5	12	24	38	53
0.08		6	13	25	39	54
0.10		7	14	26	40	55
0.12		8	15	27	41	56
0.14		9	16	28	42	57
0.16			17	29	43	58
0.20			18	30	44	59
0.25			19	31	45	60
0.30			20	32	46	61
0.35			21	33	47	62
0.40				34	48	63
0.50				35	49	64
0.60					50	65
0.80						66

- Phil's analysis had a covariance matrix and mapping table which tells you how to correctly project the 2D distribution into a 1D histogram that matches the covariance.

- This is perfect, release any new distributions following that papers example.


THE
BAD

Experimental Issues

- Compiled a list of the issues we have with peoples data releases.
- These points don't just apply to you, pretty much every experiment we have dealt with has had these same issues.

5. Signal Definitions

- This is the thing we spend most of our time on.
- **Sometimes it can be very unclear what are the most appropriate signal cuts that should be applied to MC are.**
- Generally these are given in two sections in most papers:
  - Actual analysis cuts
  - MC Modelling
- Sometimes have to come up with our own combined interpretation as its not clear in either section.

5. Signal Definitions

- **Example:** Our first CC1pip comparisons used W_{exp} instead of W_{true} in the signal definition because we didn't realise the data also used corrected W_{exp}->W_{true}

- Think people need to move to providing explicit table of cuts that should be applied to simulation to avoid any confusion.

#	**Cut**
1	1 initial state muon neutrino
2	1 final state muon
3	1 pi+ or 1pi- (total pions == 1)
4	Theta mu < 20 degrees (see Equation 1)
5	Wtrue < (see Equation 2)

6. Self Contained Data Release

- At the moment, if I want to compare to the CC1pi0 data on the data release page I have this info.

Overview

These measurements all use the low energy NUMI beams (average neutrino energy of 3.5-4.0 GeV), running in both neutrino and antineutrino configurations.

Publication 1 above has both pion and muon kinematics with derived quantities.

Publications 2 and 3 above emphasized pion kinematics.

Publication 1 extends publications 2 and 3, also supercedes data from publication 3.

Data

- Data files (including the original charged and neutral pion production results updated with the published flux) can be found at the following location: latex file of cross sections as function of both muon and pion kinematics
- The new data superceding data from publication 3 are provided at this this link (pdf file).
  A new publication with these new data is in preparation, please reference Publication 3 until the new reference is available.

- Publications 2 and 3 are the older pre-flux distributions, but publication 2 doesn't contain the flux used.
- Publication 1 is with the new flux but the paper doesn't contain the flux distribution used (the latex file does).

6. Self Contained Data Release

- A data release (especially the one on your nice release site) can be larger than what is in the paper.

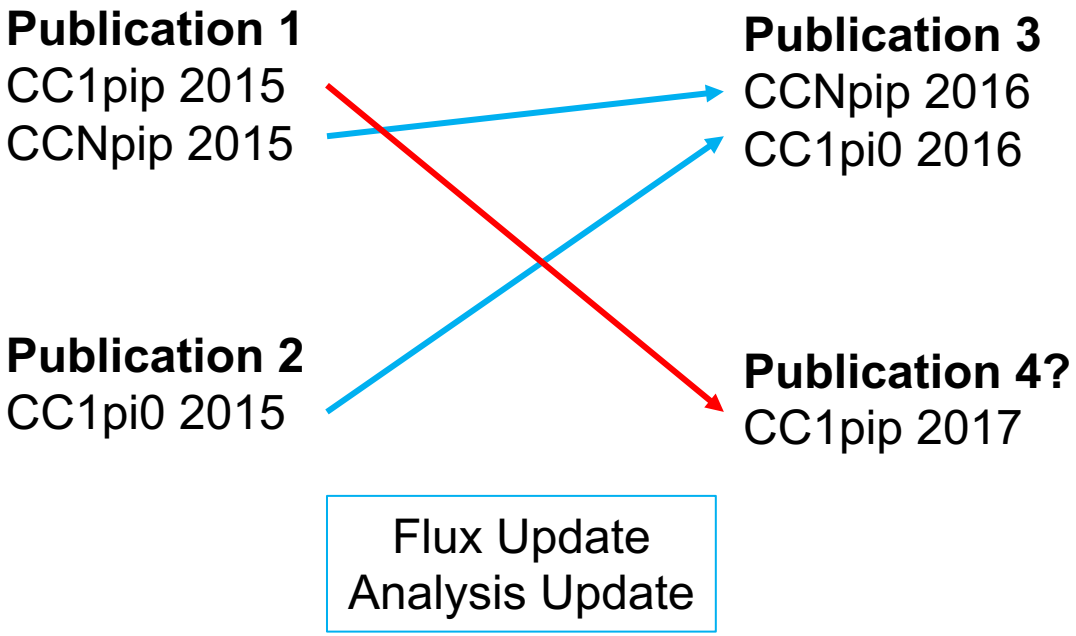
- When creating the data release treat it as if this is the only data release you have ever released.

- **Data release + paper should try to contain everything an analyzer needs to use the data.**
  - Exact target definitions
  - **Exact signal definitions** ← This needs to be crystal clear
  - Binning Methods
  - Flux Prediction
  - Data points
  - MC points

7. New Data

- I think its great that you update data as you improve analyses.

- We have found it can be unclear what data should be used if you are not constantly keeping track of MINERvA's data releases, so could be tricky for anyone new.



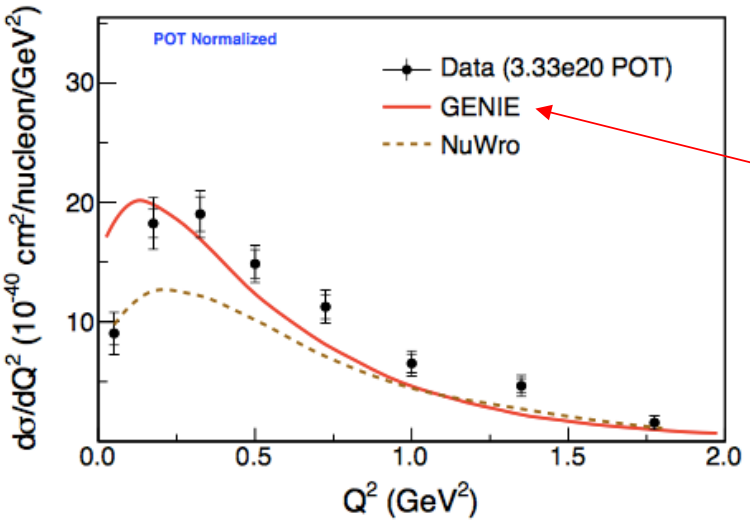
CC1pip has a 2017 data release but no accompanying paper.

But apparently signal definition changed and yet its not mentioned anywhere.

8. Unfolded Covariance

- If you are going to provide unfolded distributions and covariance, but would like to quote chi2 values from reco data...

- **Please quote the chi2 values you get from unfolded distributions**



Personal preference is to include it in the legend here, but that is just my opinion.

- Helps us validate our calculations, and is an extra cross-check to help make sure the covariances you release are actually valid.

Every experiment has given us an un-invertable covariance at least once…

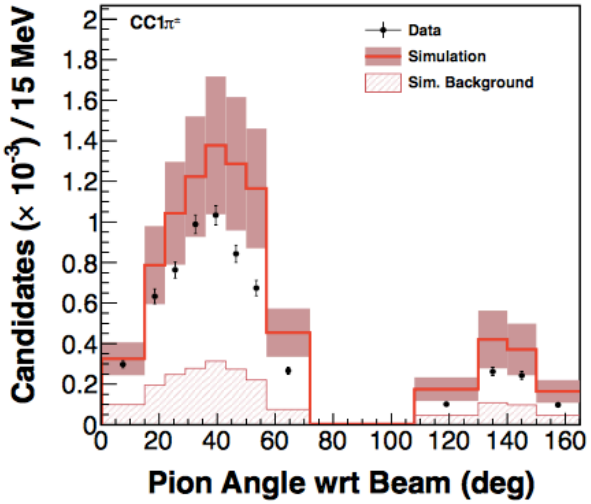
9. MC Points

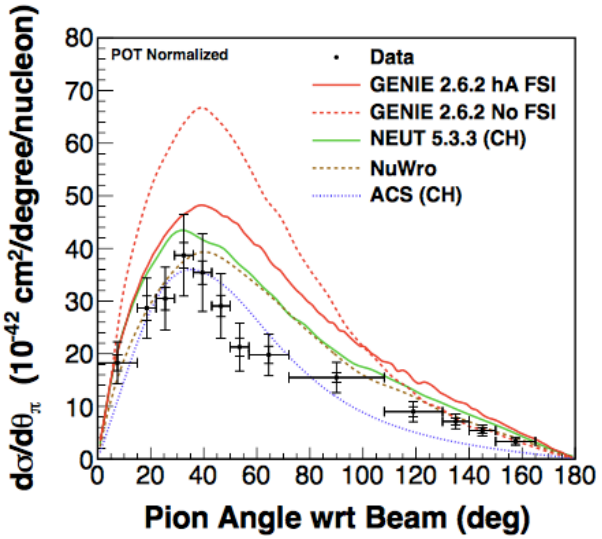
- Checking predictions match what you got by eye is awkward.
- I ended up having to manually digitise a huge stack of MINERvA MC distributions to validate our code.
- **Providing a table of MC points is also important in the data release. You've worked hard to make those comparisons, let people use them!**
- Also means theorists comparing to your paper can include a GENIE/NuWro MC curve if they want...

10. Restricted Phase Space

- Some data distributions are unfolded into a full phase space measurement which covers regions you didn't actually measure.

- **Restricted phase space distributions are way more useful to generator builders than a prediction that has been sprinkled with MC to get full phase space coverage.**

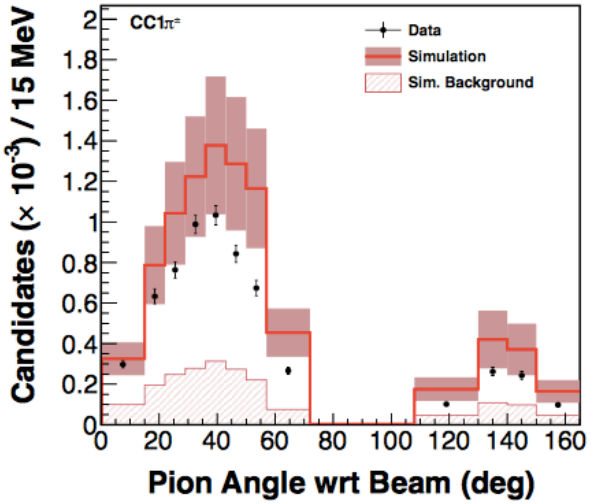


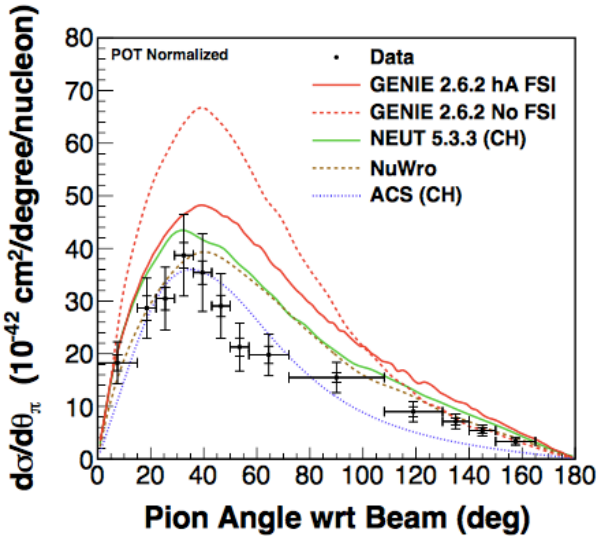


10. Restricted Phase Space

- If you are going to unfold to full phase space to appease some theorists then **always** provide both restricted and full phase space. (prioritise restricted)

- 2015 CCNpip had angular cuts, but the 2016 numubar CC1pi0 and CCNpip data releases are only full phase space!






THE
UGLY

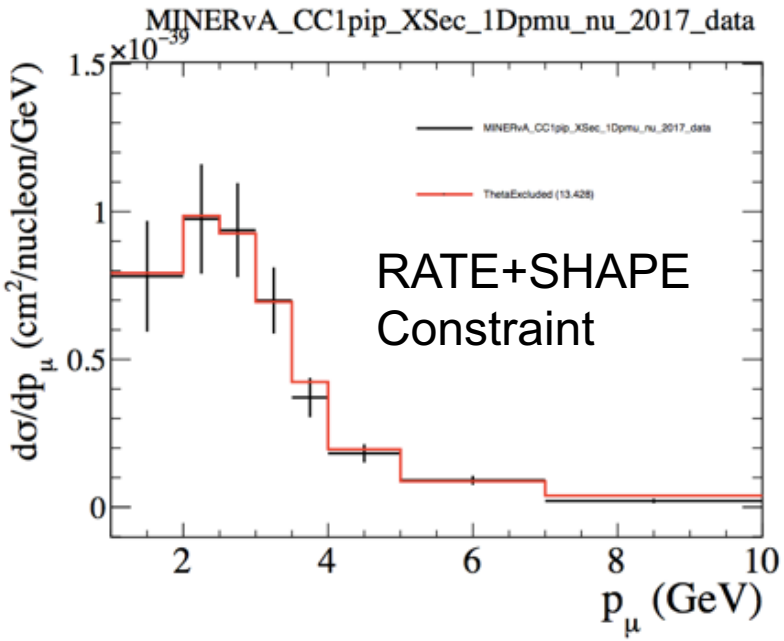
Longer Term Projects

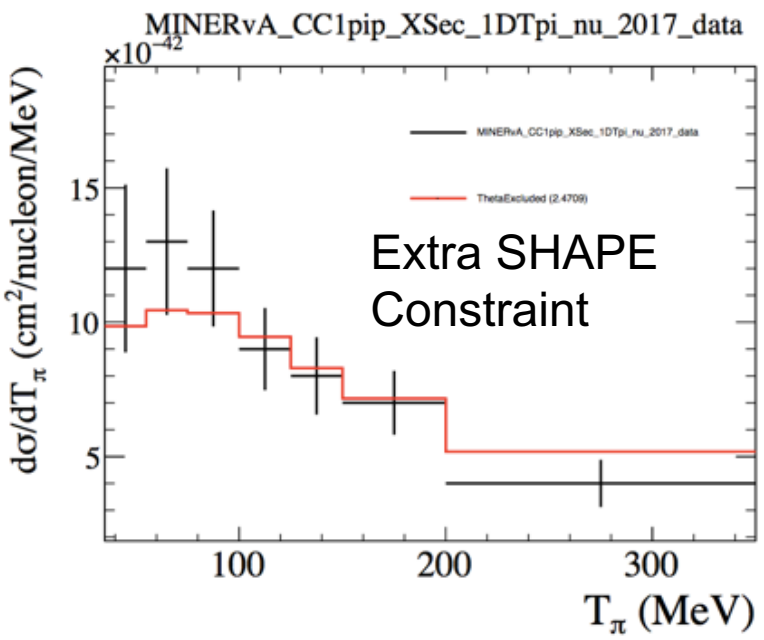
- There are few things we think will be important in the future when trying to address model dependence in data.
- None of these things are quick fixes, as each would probably be breaking new ground, but are where we think the field could go to address some of the issues seen.
- Really keen for a lot of these things, so if you have early outputs you would like to test consider using NUISANCE!

11. Sample Correlations

- Have neutrino/antineutrino covariance matrices for some samples which is awesome.

- Biggest issue we are facing is combining distributions from the same dataset. At the moment we fit RATE for one distribution and SHAPE for the others.





11. Sample Correlations

- No-one has every really given us a covariance that covers multiple distributions from the same measurement.

- Is it possible?
- You've successfully made nu/nubar cross-correlations.

- Moving to a format where we have access to this information has the potential to help us set stronger yet reliable model constraints.

12. More Reco Data

- Would like to see more Reco distributions, not just ones for distributions where model-dependence is a **major** problem in the unfolding.

- MiniBooNE NCEL paper is very nice example of reco data which also tries to cater to pure theorists.

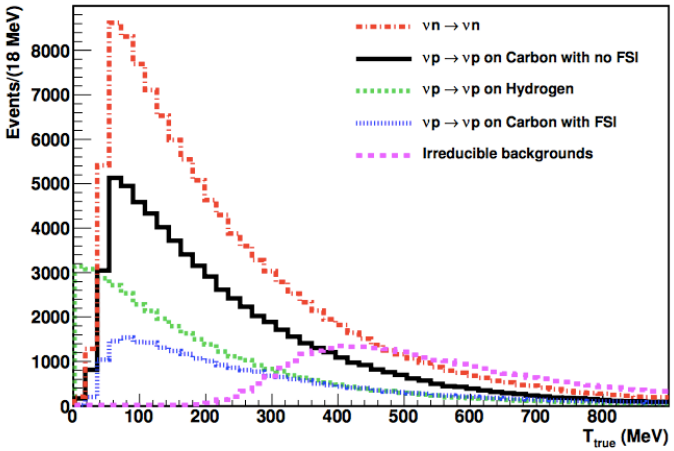


Figure B.1: True energy histograms for possible signal events before any cuts are applied, used in the NCE cross-section measurement.

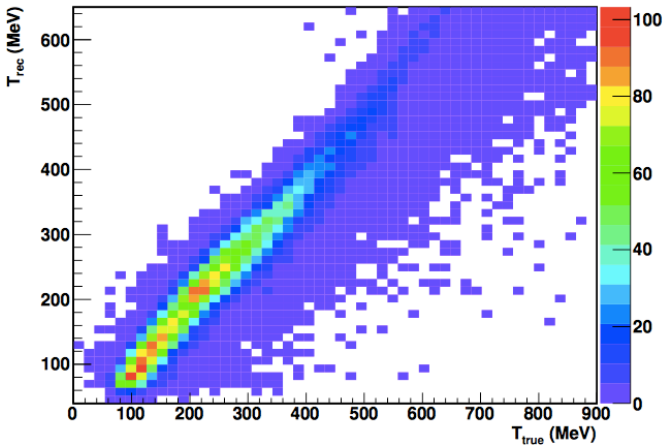


Figure B.3: Migration matrix for the NCE events on hydrogen in the NCE event sample.

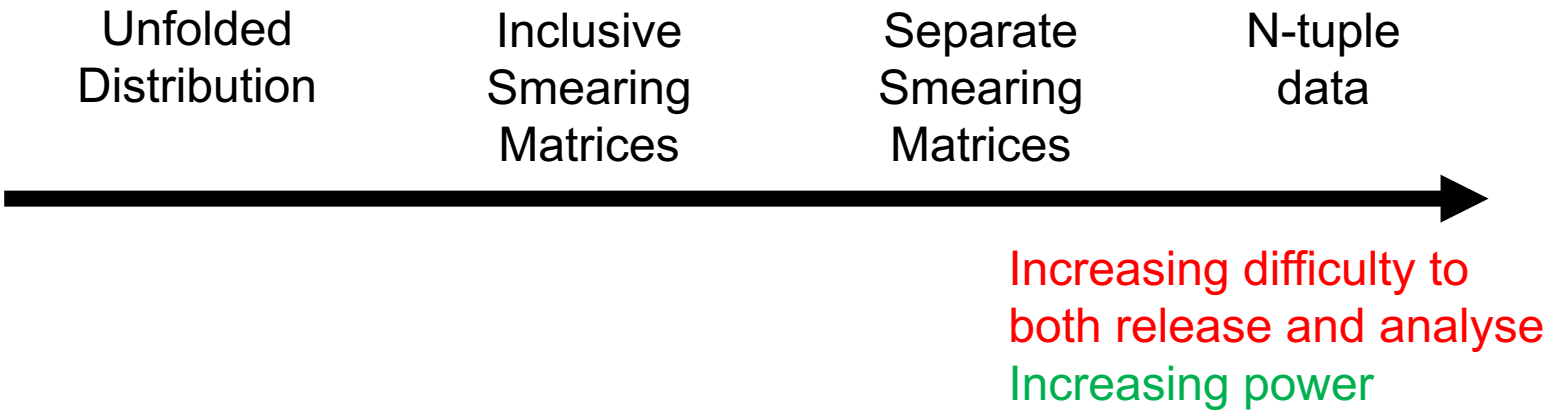
http://www-boone.fnal.gov/publications/Papers/denis_thesis.pdf

12. More Reco Data

- Lots of possible ways to release reco-data.

- Experiments finding a proper way to release N-tuple data that we can apply parametrised detector smearing to could be very powerful.

- Luke Pickering, is looking at something similar in NUISANCE in relation to CLAS data so you should speak to him if interested.

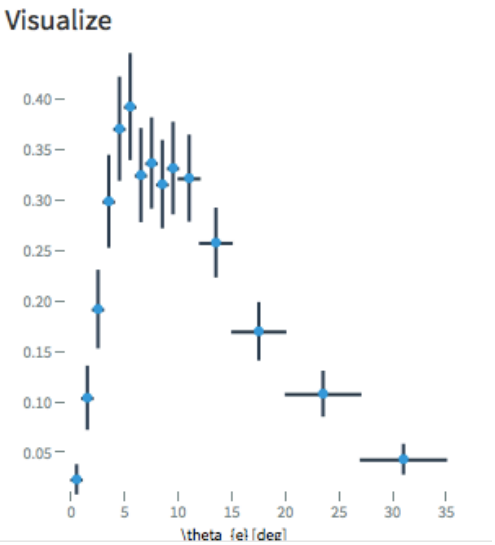


13. HepData

- HepData project at Durham wants to also start including neutrino cross-section data.

- They already have some MINERvA datasets, so I'm guessing have you already spoken to them, but thought I'd mention it just in case.

θ_e [deg]	$d\sigma/d\theta_e$ [10^{-39} cm^2/deg/nucleon]
0 - 1	0.023 ±0.012 stat ±0.009 sys
1 - 2	0.104 ±0.022 stat ±0.023 sys
2 - 3	0.192 ±0.025 stat ±0.03 sys
3 - 4	0.299 ±0.03 stat ±0.035 sys
4 - 5	0.371 ±0.034 stat ±0.039 sys
5 - 6	0.393 ±0.036 stat ±0.039 sys
6 - 7	0.325 ±0.032 stat ±0.034 sys
7 - 8	0.337 ±0.033 stat ±0.031 sys



Summary

1. Covariance
2. Reco Data for model dependent variables
3. Alternative Kinematics
4. 2D Data Releases

5. Clarify signal definitions
6. Completely self contained data releases
7. Make clear what data should be used for each channel
8. Provide unfolded chi2 results
9. Include MC predictions in tables
10. Always try to include restricted phase space

11. Provide distribution cross-correlations
12. Produce more reco data
13. More HepData entries