What we measure when we measure σ

Cross-section extraction best practices at ND280

Stephen Dolan

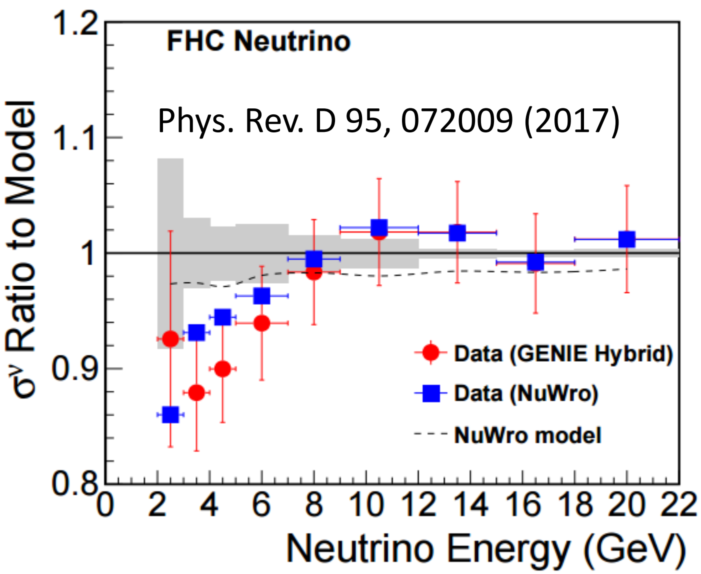
For the T2K collaboration

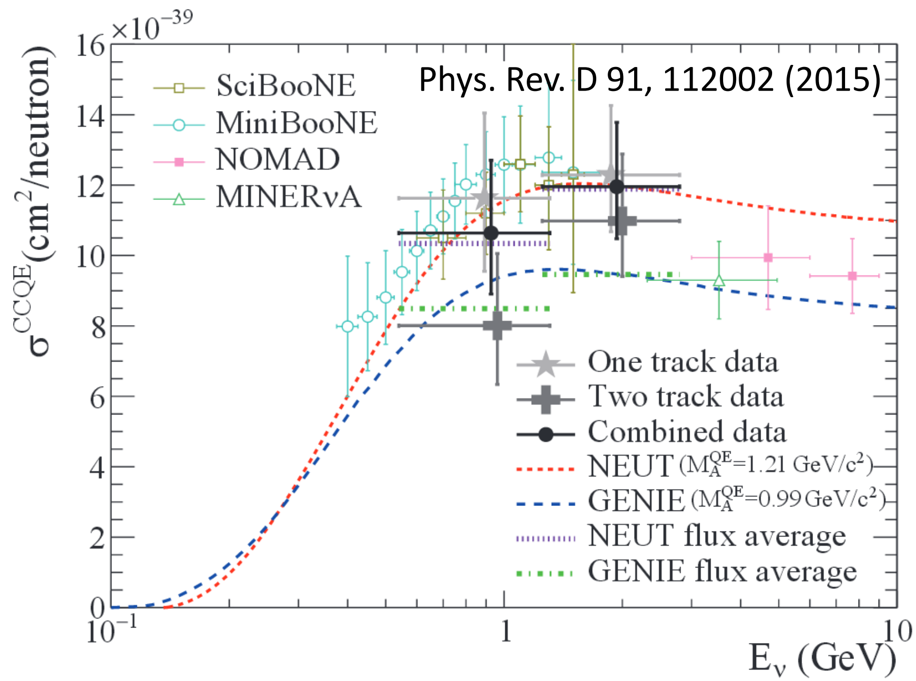
s.dolan@physics.ox.ac.uk

UNIVERSITY OF OXFORD

Model dependence is important!

- In this talk "model dependence" = dependence on the signal we are trying to measure
- Can obfuscate the interesting physics in our results
- Tension between results from different experiments in global fits (e.g. Phys. Rev. D 93, 072010 (2016))
  – Model dependence could be partially responsible
- Difficult to avoid entirely





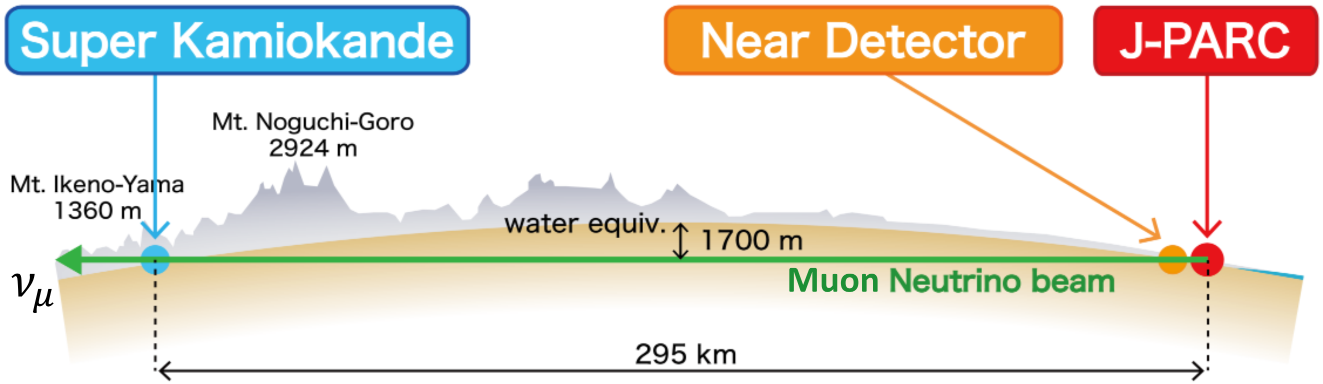
Hypocrisy warning

- Will present ND280 *best* practices for cross-section extraction
- This does not mean that all ongoing or previous analyses adhere to these
- Aim of this is to:
  - Understand whether our best practices are sensible
  - Provide useful methodologies beyond ND280
  - Converge on a global set of best practices

Overview

- T2K and ND280
- Choosing a signal definition
- Choosing a selection
- Choosing a binning
- Efficiency corrections
- Unfolding, uncertainties and model dependence
  - Regularisation
  - D'Agostini (1995)
  - Likelihood fitting
  - How not to unfold
- Fake data and bias studies
- Conclusions

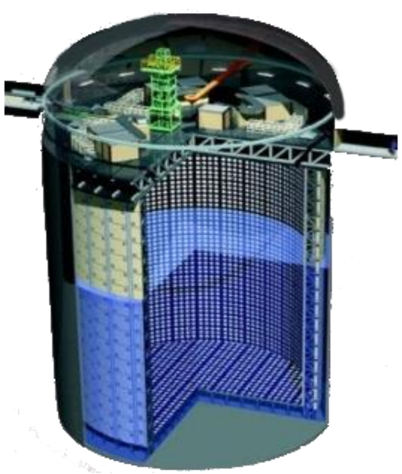
The T2K Experiment



Use off-axis beam to give a narrow neutrino energy spread

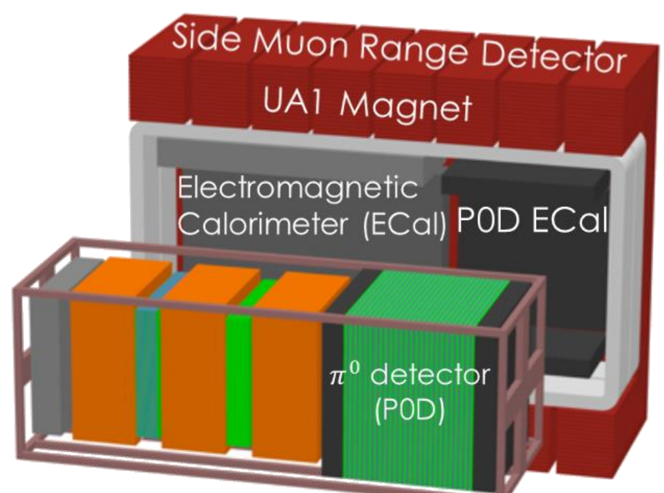
Far Detector (Off-Axis)

Super-Kamiokande

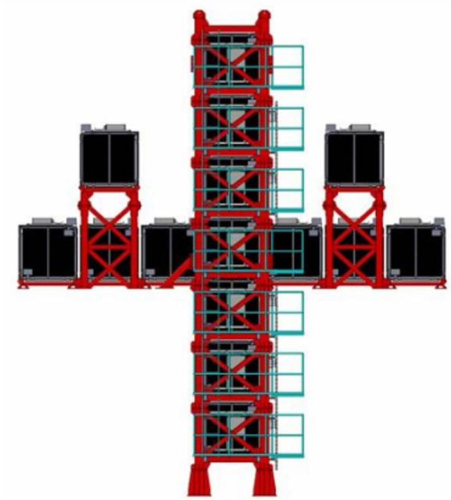


Near Detectors

Off-Axis: ND280



On-Axis: INGRID



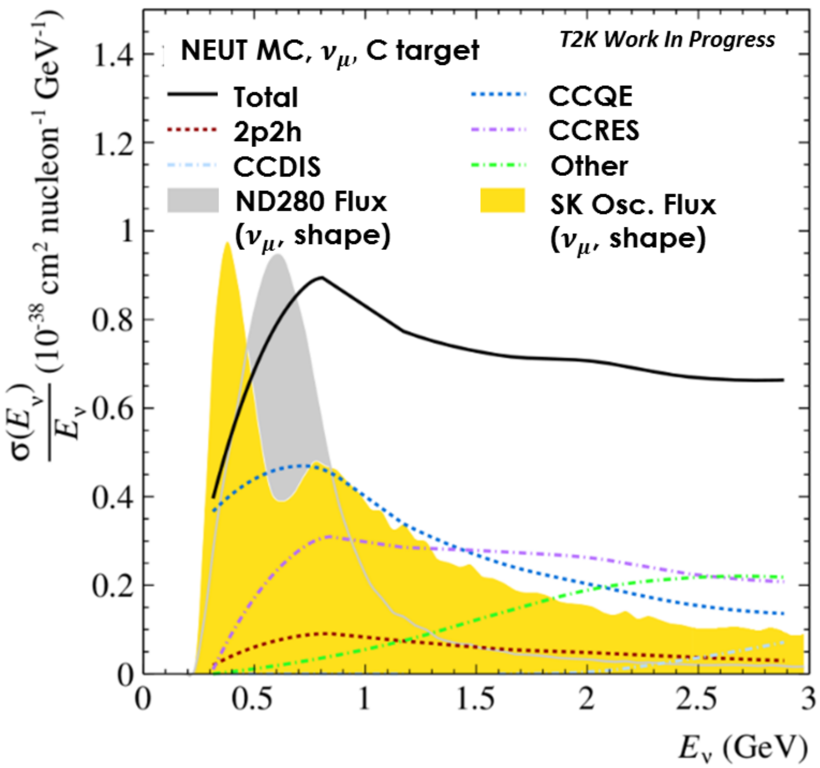
ν-Interactions and Osc. Analysis

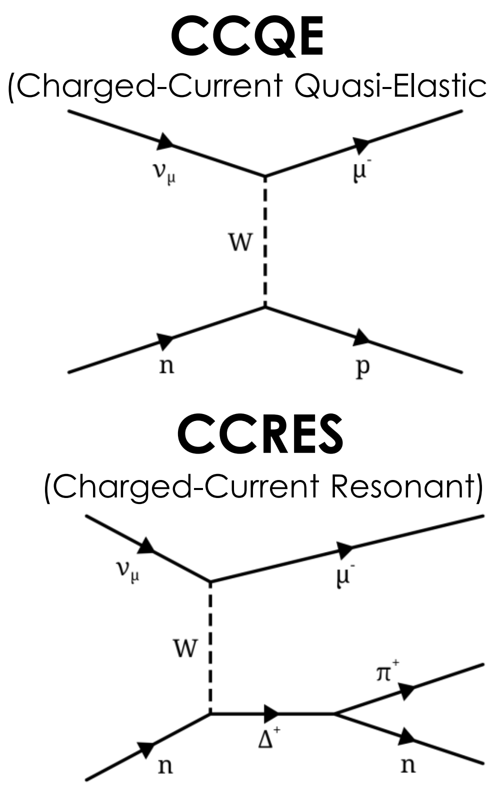
Fractional error on the number of expected events at SK with and without ND280

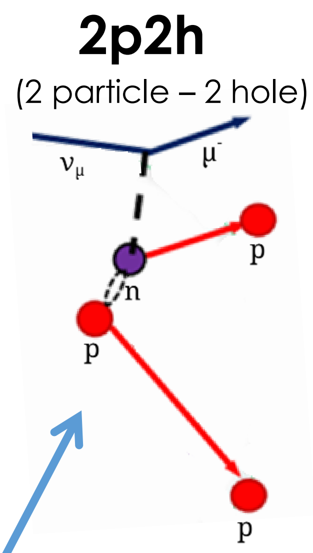
	ν_μ sample $1R_\mu$ FHC	ν_e sample $1R_e$ FHC	$\bar{\nu}_\mu$ sample $1R_\mu$ RHC	$\bar{\nu}_e$ sample $1R_e$ RHC
ν flux w/o ND280	7,6%	8,9%	7,1%	8,0%
ν flux with ND280	3,6%	3,6%	3,8%	3,8%
ν cross section w/o ND280	7,7%	7,2%	9,3%	10,1%
ν cross section with ND280	4,1%	5,1%	4,2%	5,5%
ν flux+cross section	2,9%	4,2%	3,4%	4,6%
Final or secondary hadron int.	1,5%	2,5%	2,1%	2,5%
Super-K detector	3,9%	2,4%	3,3%	3,1%
Total w/o ND280	12,0%	11,9%	12,5%	13,7%
Total with ND280	**5,0%**	**5,4%**	**5,2%**	**6,2%**

- Largest systematic uncertainty comes from neutrino interaction uncertainties

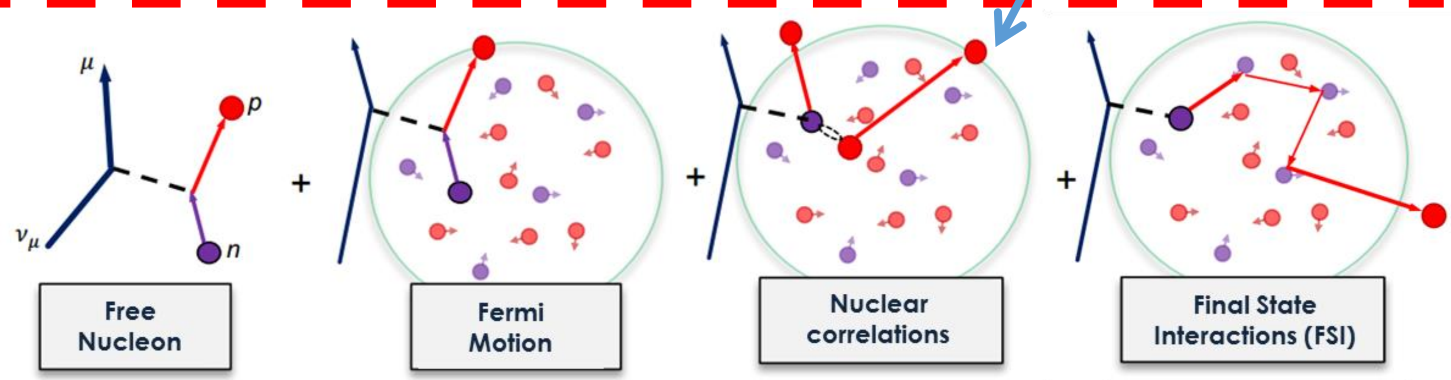
Neutrino interactions at T2K





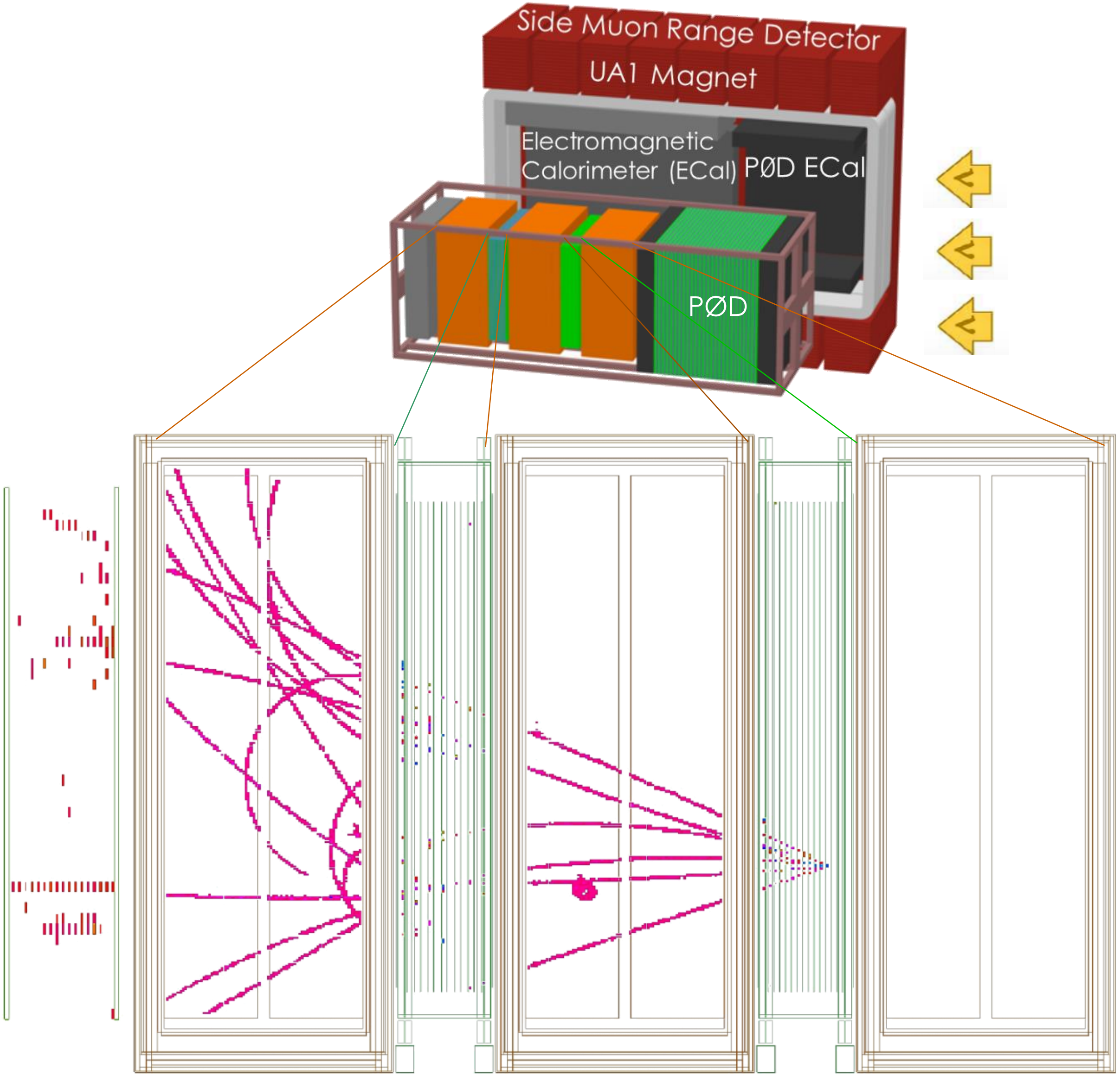


Nuclear Effects

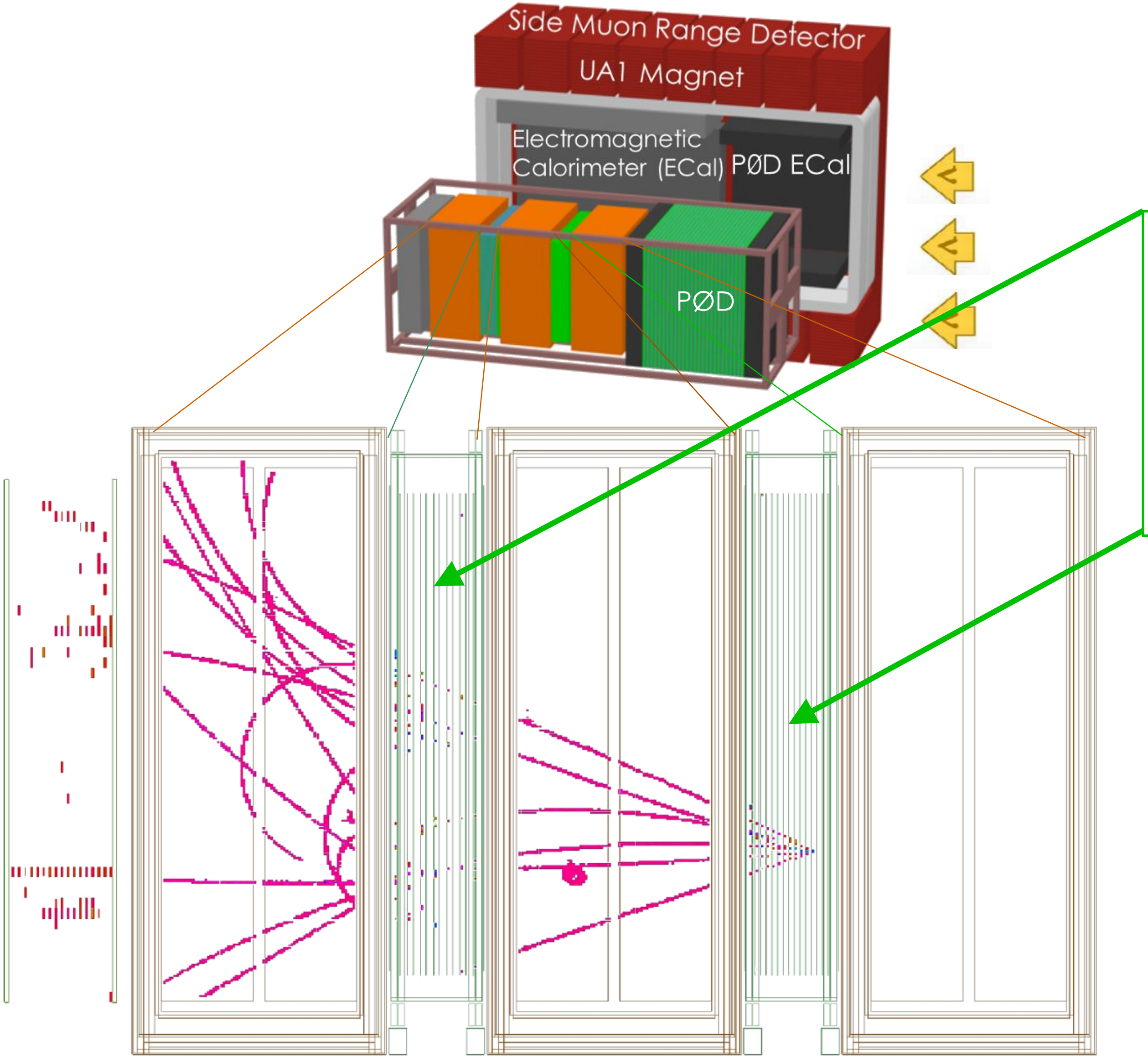


Diagrams by Patrick Stowell

ND280 (off axis near detector)



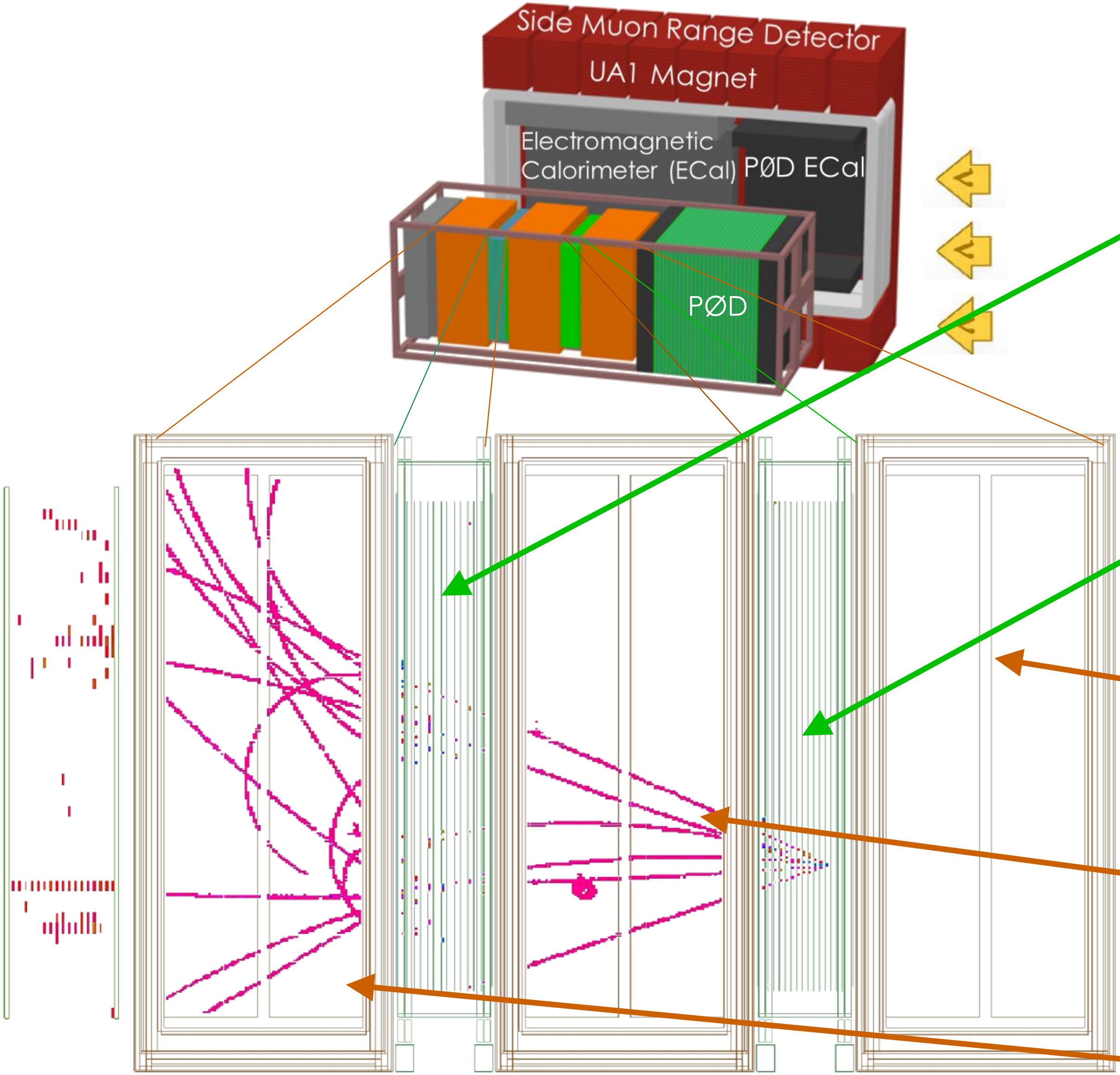
ND280 (off axis near detector)



Fine-Grained Detectors (FGD 1 & 2):

- CH scintillator tracker
- Target for ν
- FGD2 contains water

ND280 (off axis near detector)



Fine-Grained Detectors (FGD 1 & 2):

- CH scintillator tracker
- Target for ν
- FGD2 contains water

Time Projection Chambers (TPC):

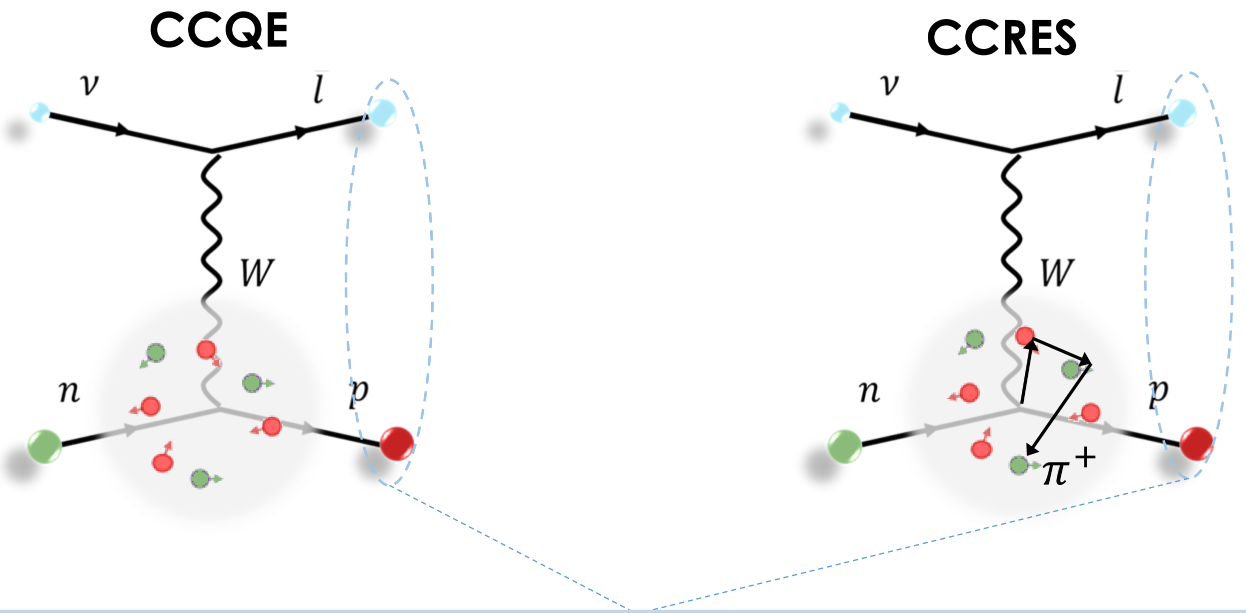
- Excellent tracking
- High-res charged-particle momenta
- Accurate particle ID

Overview

- T2K and ND280
- Choosing a signal definition
- Choosing a selection
- Choosing a binning
- Efficiency corrections
- Unfolding, uncertainties and model dependence
  - Regularisation
  - D'Agostini (1995)
  - Likelihood fitting
  - How not to unfold
- Fake data and bias studies
- Conclusions

What can't we measure

- Naively it would be great to measure $\sigma_{CCQE}(E_\nu), \sigma_{2p2h}(E_\nu), \sigma_{Oth}(E_\nu)$
- Why not?

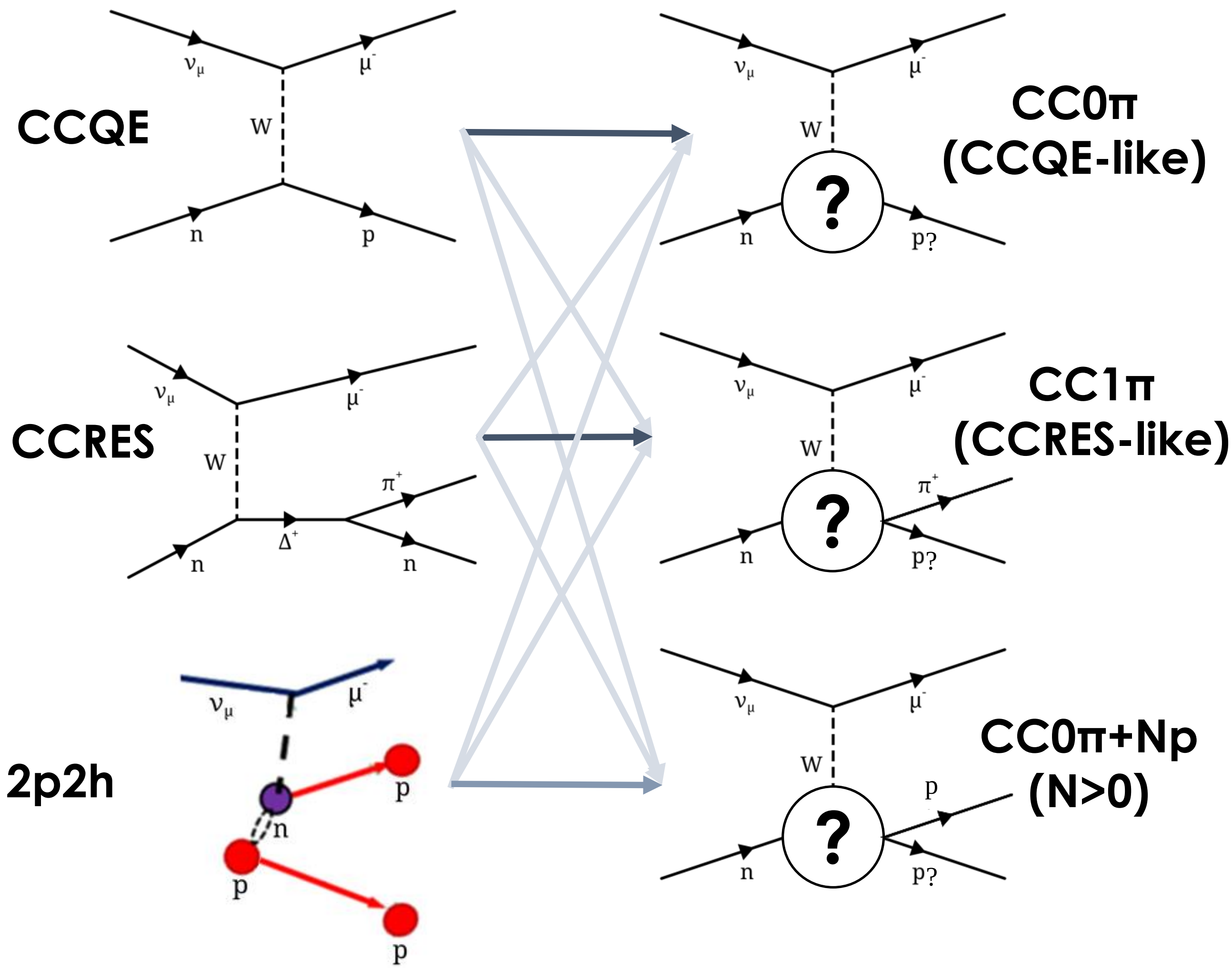


Final state interactions (FSI) can cause different interaction modes to have the same final state

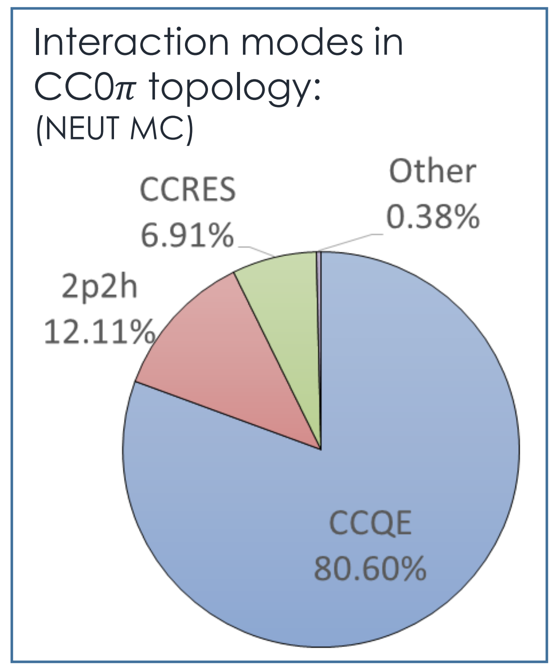
- Can't separate interaction modes on an event by event basis
- Entirely reliant on the input simulation to tell us contamination

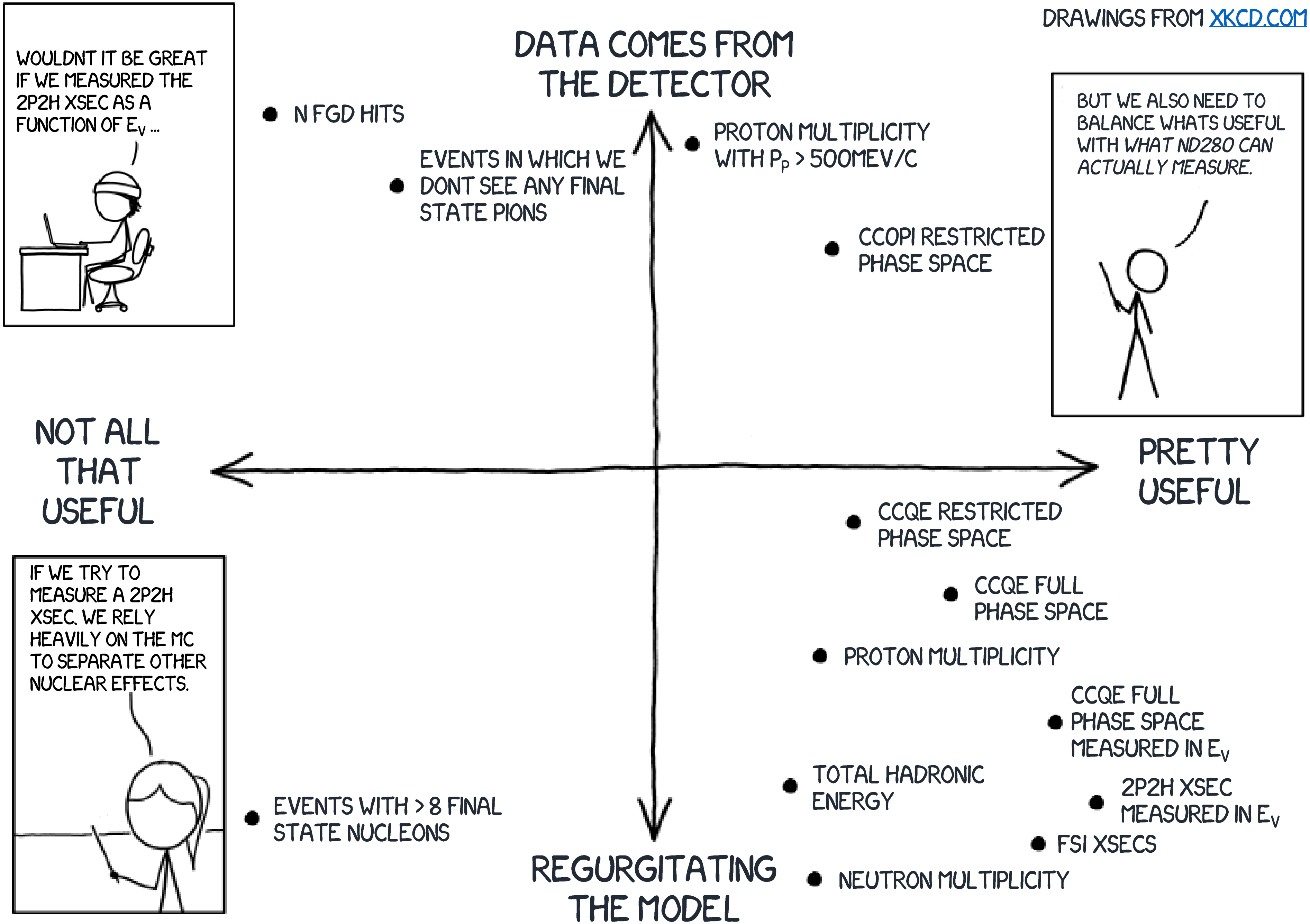
What can we measure

Interaction Modes | **Interaction Topologies**



- Nuclear effects obfuscate interaction mode
- To minimise model dependence we measure interaction topologies




DRAWINGS FROM XKCD.COM
WOULDNT IT BE GREAT IF WE MEASURED THE 2P2H XSEC AS A FUNCTION OF E_ν ...
DATA COMES FROM THE DETECTOR
BUT WE ALSO NEED TO BALANCE WHATS USEFUL WITH WHAT ND280 CAN ACTUALLY MEASURE.
N FGD HITS
EVENTS IN WHICH WE DONT SEE ANY FINAL STATE PIONS
PROTON MULTIPLICITY WITH P_p > 500MEV/C
CCOPI RESTRICTED PHASE SPACE
NOT ALL THAT USEFUL
PRETTY USEFUL
CCQE RESTRICTED PHASE SPACE
CCQE FULL PHASE SPACE
PROTON MULTIPLICITY
CCQE FULL PHASE SPACE MEASURED IN E_ν
TOTAL HADRONIC ENERGY
2P2H XSEC MEASURED IN E_ν
FSI XSECS
NEUTRON MULTIPLICITY
IF WE TRY TO MEASURE A 2P2H XSEC, WE RELY HEAVILY ON THE MC TO SEPARATE OTHER NUCLEAR EFFECTS.
EVENTS WITH > 8 FINAL STATE NUCLEONS
REGURGITATING THE MODEL

Choosing a signal definition

- Need to provide a useful cross section that can easily be compared to model predictions (without needing a detector simulation).
- Should only attempt to measure what ND280 can actually reconstruct.
- Avoid extracting interaction mode xsecs from measurements of an interaction topology.
  - State signal definition clearly, e.g.:
    - $\nu_\mu CC0\pi \; on \; H_2O$
    - $\nu_\mu CCN\pi \; on \; Pb, N \geq 1$
    - $\nu_\mu CC0\pi + Np \; on \; CH, N \geq 1$

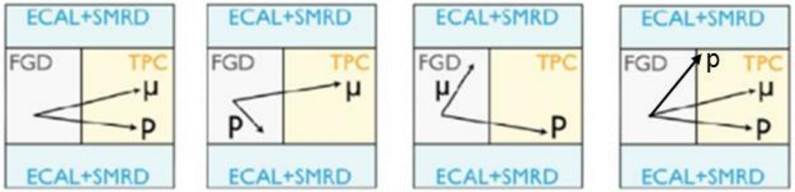
TL;DR: ENSURE A BALANCE BETWEEN USEFULNESS AND MODEL INDEPENDENCE

Overview

- T2K and ND280
- Choosing a signal definition
- Choosing a selection
- Choosing a binning
- Efficiency corrections
- Unfolding, uncertainties and model dependence
  - Regularisation
  - D'Agostini (1995)
  - Likelihood fitting
  - How not to unfold
- Fake data and bias studies
- Conclusions

Choosing a selection

- Selections for topology based signal definitions are conceptually simple
- Avoid cutting on “interaction-level” variables
- Wish list:
  - High proportion of signal events in selection → High Purity
  - High proportion of total number of signal events → High Efficiency
  - Control regions to constrain backgrounds
    - Details of how we use control regions in the backups
- Typically have multi-sample selections to maximise kinematic acceptance
  - Example from $CC0\pi + Np$ selection:
    (more details in backups)

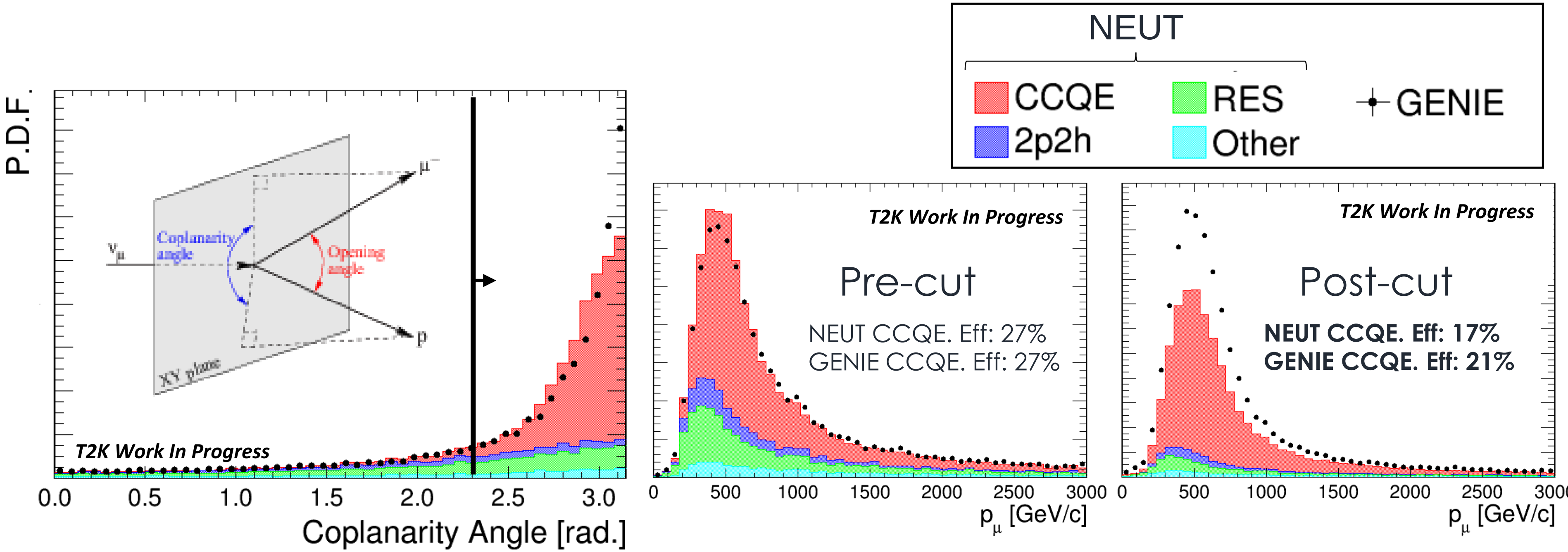


- The basic selection should look something like this (see backups for details):

Event Quality → Vertex in subdetector → PIDs → Vetos

Cutting on "interaction-level" variables

- Cutting on variables like **vertex activity** or **coplanarity angle** can give interesting interaction mode enhancements.
- But the change in efficiency and purity from the cut may strongly depend on the interaction model.
- **Potential for strong model dependence**



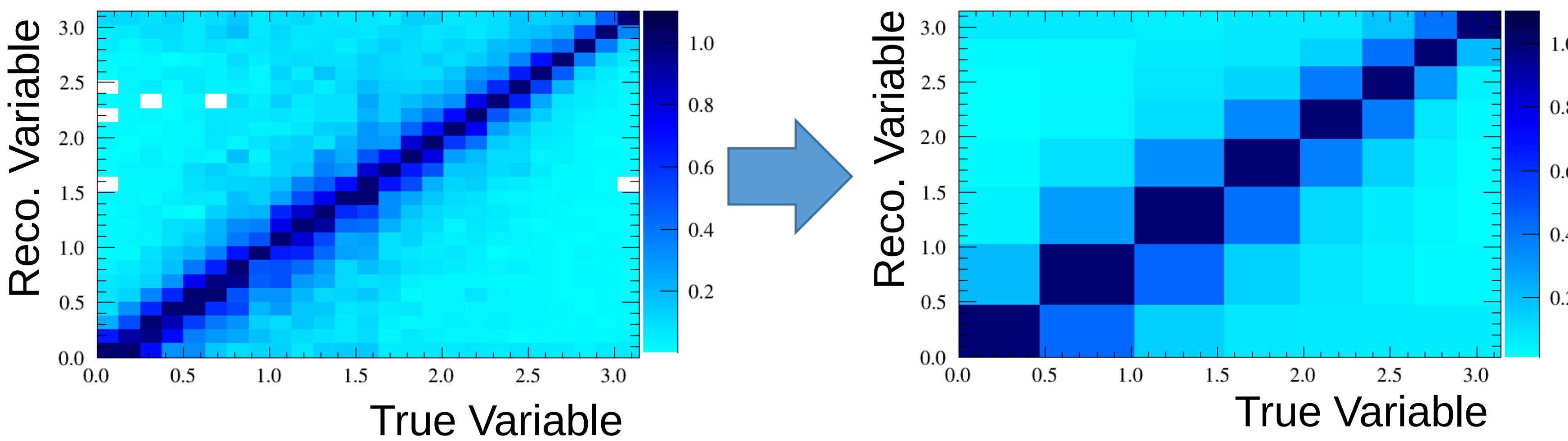
Overview

- T2K and ND280
- Choosing a signal definition
- Choosing a selection
- Choosing a binning
- Efficiency corrections
- Unfolding, uncertainties and model dependence
    - Regularisation
    - D'Agostini (1995)
    - Likelihood fitting
    - How not to unfold
- Fake data and bias studies
- Conclusions

Choice of binning

Binning choice is a careful balance between a number of factors:

- **Bin width should not be finer than the detector resolution**
  - Overly fine bins increases ill-posedness of the unfolding problem
  - Requires stronger regularisation during unfolding
  - More potential for model dependence

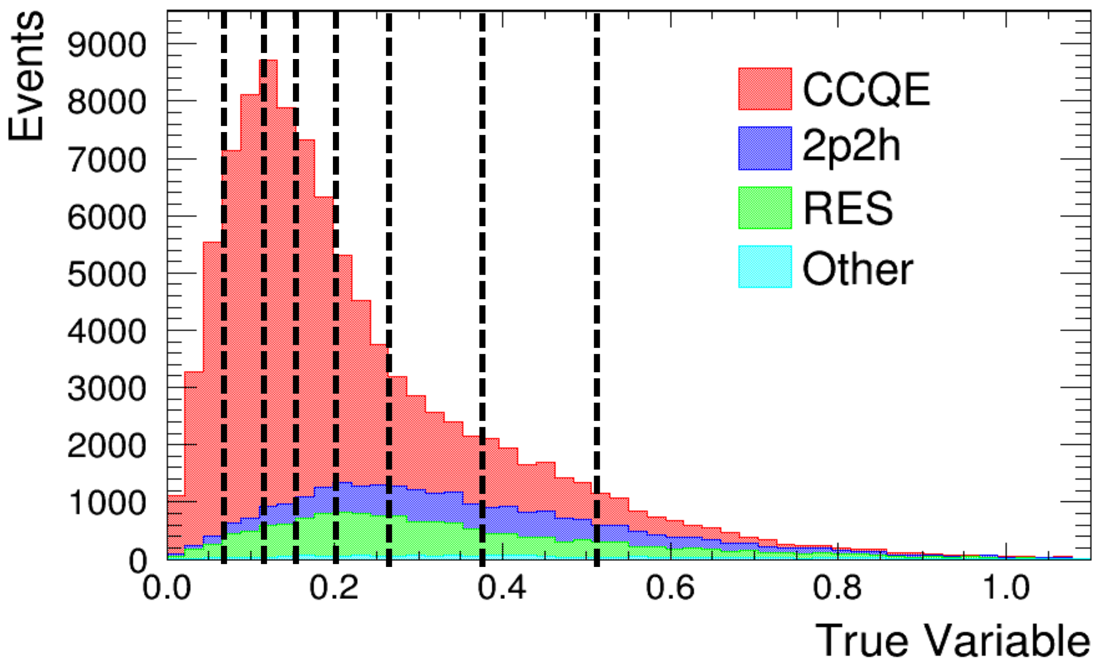


- In general: $$Bin\ Width < RMS\left(\frac{N_{reco} - N_{true}}{N_{true}}\right)$$

Choice of binning

Binning choice is a careful balance between a number of factors:

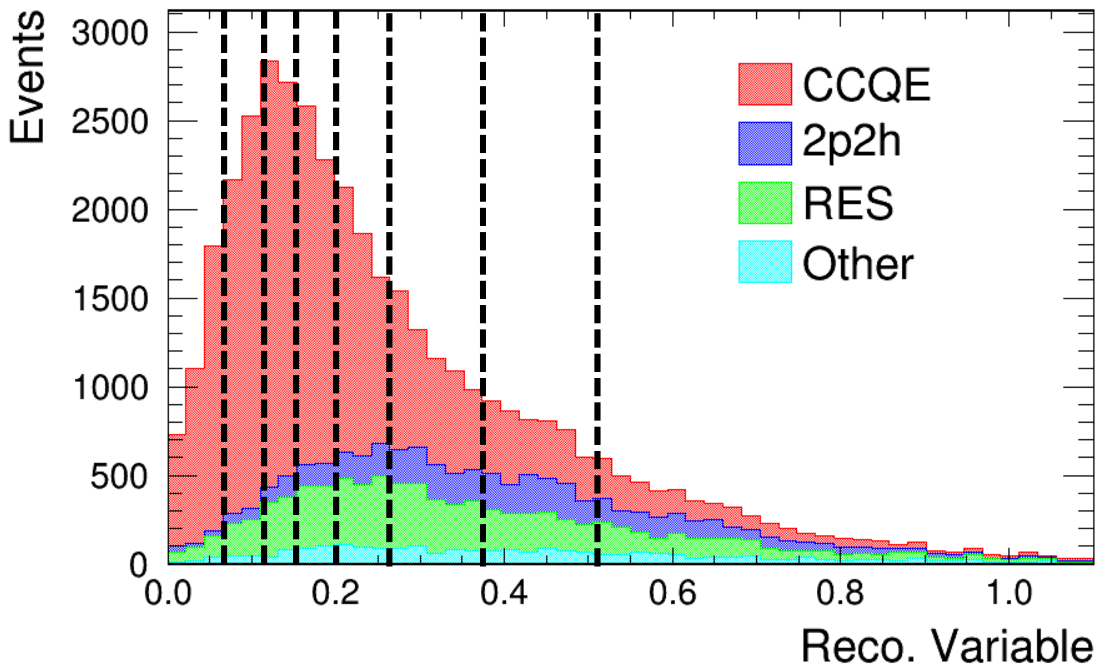
- **Expected signal variation within a bin should be small and smooth**
  - Coarse bins lose sensitivity to interesting cross-section variation
  - Trust input simulation to describe distribution of events within a bin
  - Can lead to model bias in efficiency correction (more on this later)



Choice of binning

Binning choice is a careful balance between a number of factors:

- **Stat. error should not be much greater than syst. error in each bin**
  - Can always combine bins later (with some caveats)
  - Helpful if stat. uncertainty ~ Gaussian



Overview

- T2K and ND280
- Choosing a signal definition
- Choosing a selection
- Choosing a binning
- Efficiency corrections
- Unfolding, uncertainties and model dependence
  - Regularisation
  - D'Agostini (1995)
  - Likelihood fitting
  - How not to unfold
- Fake data and bias studies
- Conclusions

Efficiency correcting

- Whether we unfold or not, we need to correct for detector acceptance

- Normally we go from a number of selected signal events to a cross section as:

$$\left(\frac{d\sigma_{flux\ integrated}}{dx}\right)_i = \frac{N_i^{signal}}{\epsilon_{i,sim}^{signal}\Phi N_{targets}^{FV}} \times \frac{1}{\Delta x_i}$$

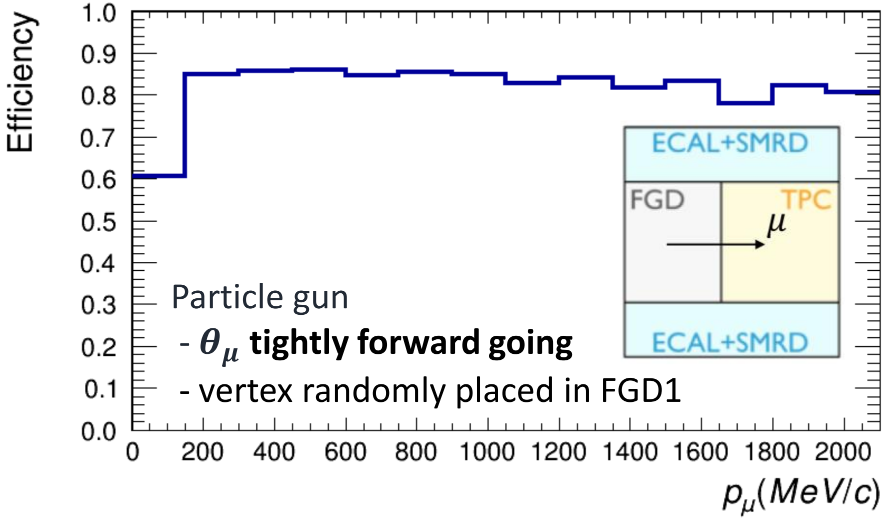
(i is the bin index)

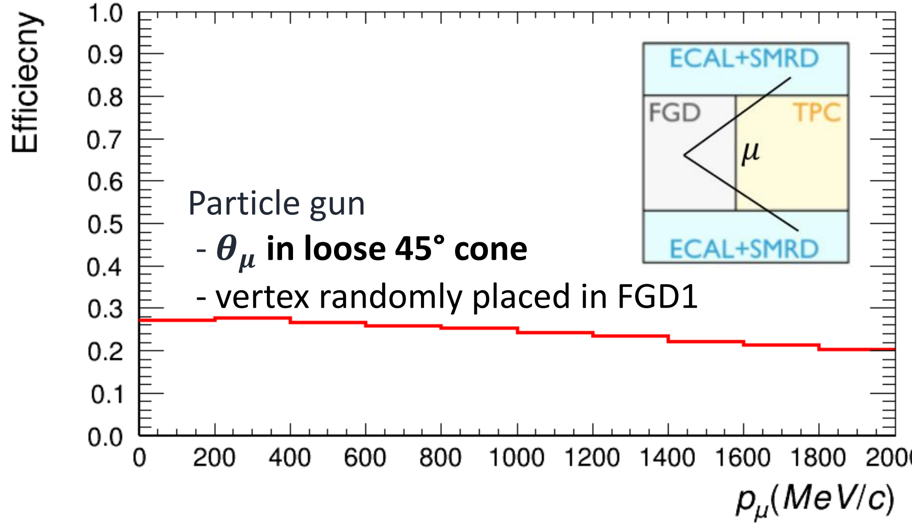
- Where we usually efficiency correct in each analysis bin

- Doing this without caution has substantial scope for model dependence

Efficiency and model dependence

- **Toy example** – want to measure p_μ for single muons using TPC.





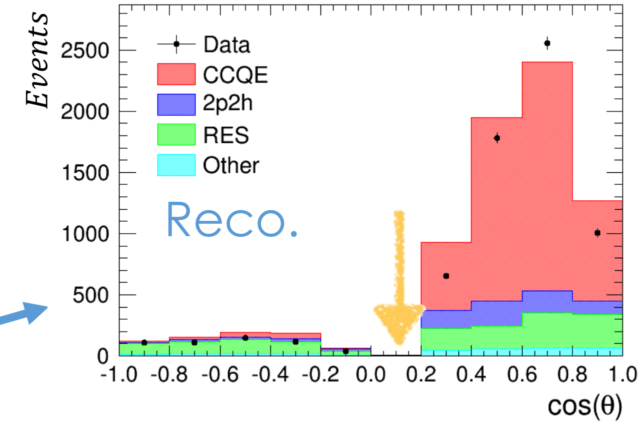
- Efficiency of seeing events with a single muon with a particular momentum starting somewhere in FGD1 depends on distribution of muon angles
  - **This depends on the neutrino scattering model**
- In general: Model dependence enters when integrating over a "model-dependent variable" with **a non-flat efficiency**
- Significant issue for single-bin cross-section measurements

Efficiency and model bias

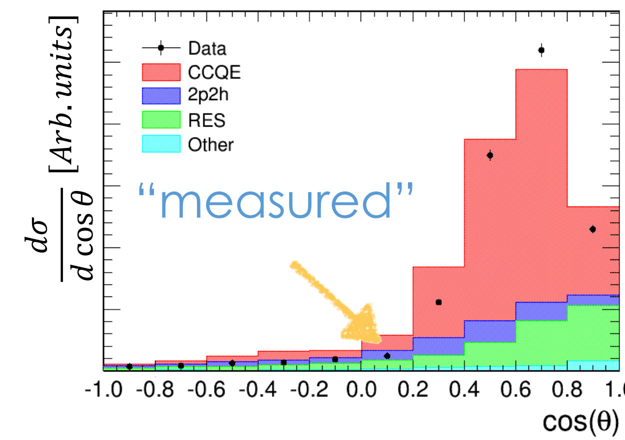
Realistic Example (but not a real result!):

Measure some particle kinematics

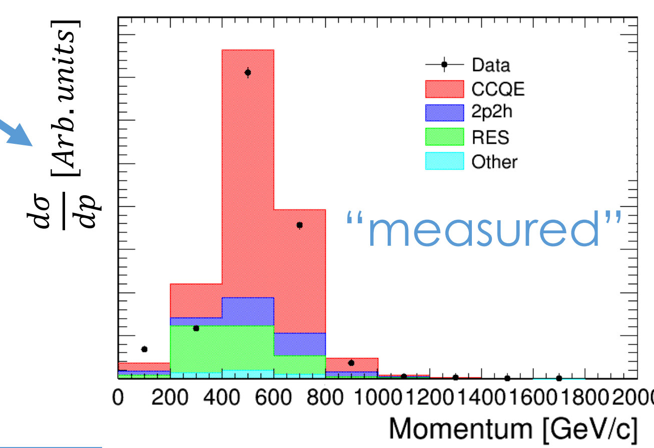
- $\sim 0\%$ efficiency at $\theta \sim 90°$



- Trying to measure the $\theta \sim 90°$
- Certainly not much data from the detector
- Can only reproduce the MC!



- Trying to measure momentum (p) over all θ
- **Spreads model dependence over all** p



What's the solution?

Ideal case:

- Extract cross section in all variables that:
  - **Characterise the detectors acceptance**
  - Whose distributions would **vary with a change of neutrino interaction model**
- After this can marginalise variables that are not of interest (See backups for more details)
- Can get very complicated with multiple particles
  - E.g. for measuring $\mu + p$ need $p_\mu, \cos(\theta_\mu), p_p, \cos(\theta_p)$ and $\cos(\theta_{\mu p})$
- Sometimes impractical

What's the solution?

Alternative:

- Measure **fiducial** cross-sections (restricted phase-space) rather than full phase space.
  - Change to the signal definition.

- Should restrict phase space to regions of well-understood, relatively flat efficiency in the underlying kinematics

- E.g. $CC0\pi + Np$ analysis measuring transverse imbalance between μ and p restricts phase space

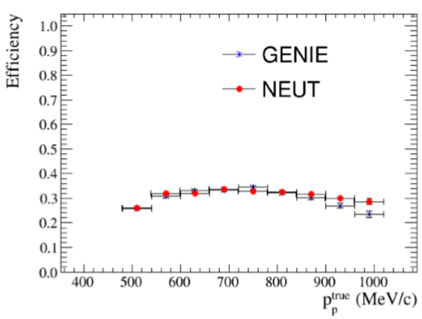
$p_\mu > 250\ MeV/c$

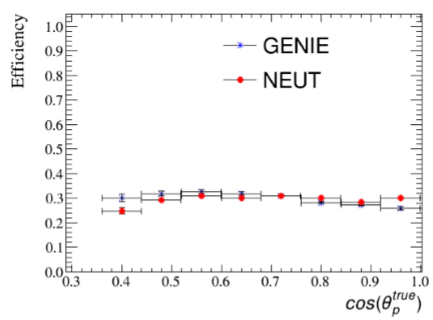
$\cos(\theta_\mu) > -0.6$

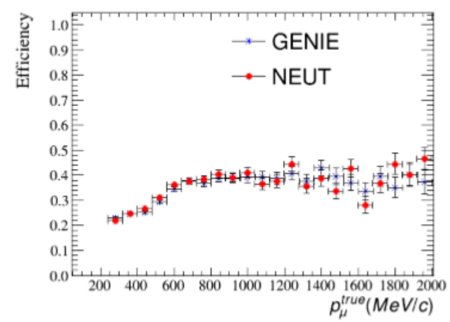
$450\ MeV/c < p_\mu < 1\ GeV/c$

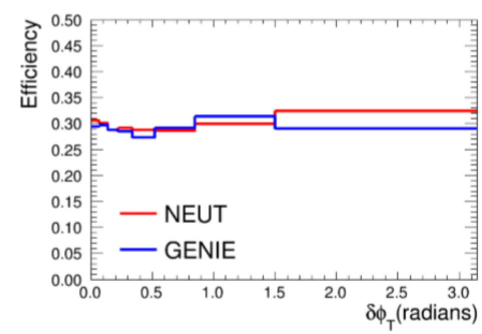
$\cos(\theta_p) > 0.4$

(Based on particle gun studies)





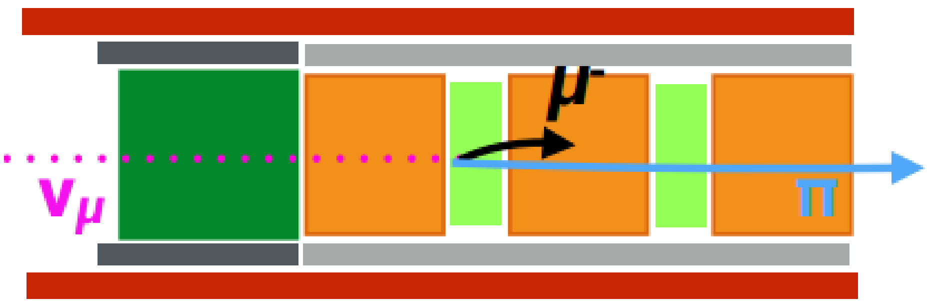


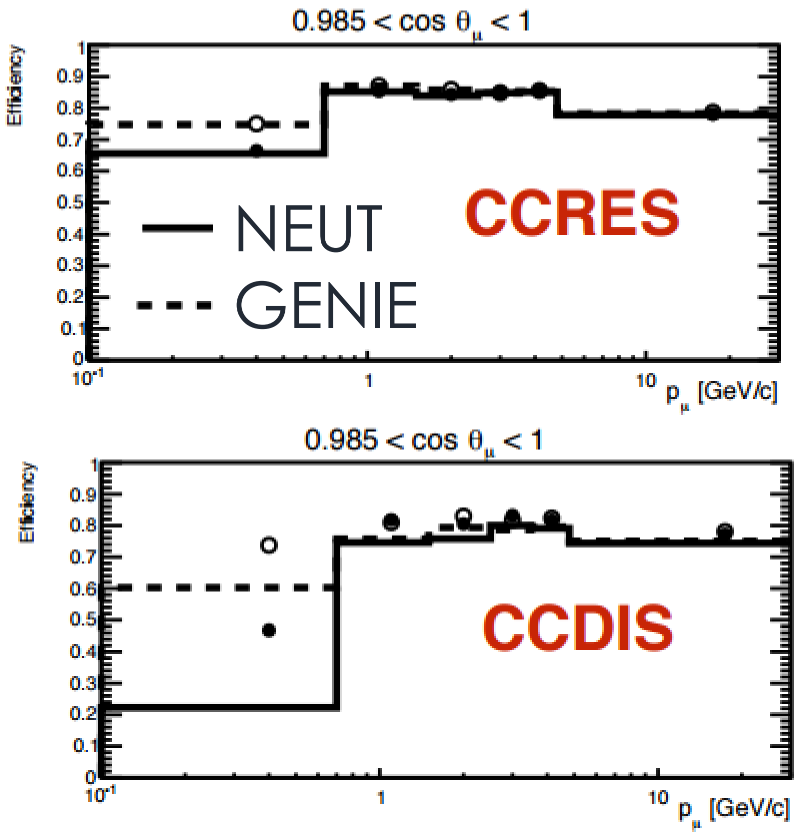


Further complications

- Even if only measuring the muon, inclusive channels can have their efficiency depend on the underlying kinematics of other particles.

- Example from T2K CC-Inclusive analysis:





- Efficiency depends on outgoing pion kinematics (different in NEUT and GENIE)

Overview

- T2K and ND280
- Choosing a signal definition
- Choosing a selection
- Choosing a binning
- Efficiency corrections
- Unfolding, uncertainties and model dependence
  - Regularisation
  - D'Agostini (1995)
  - Likelihood fitting
  - How not to unfold
- Fake data and bias studies
- Conclusions

Overview

- T2K and ND280
- Choosing a signal definition
- Choosing a selection
- Choosing a binning
- Efficiency corrections
- Unfolding, uncertainties and model dependence
  - Regularisation
  - D'Agostini (1995)
  - Likelihood fitting
  - How not to unfold

DISCLAIMER: I am not a statistician

- Conclusions

Unfolding

**G. Cowan, Statistical Data Analysis, oxford science publications (1998), c. p 163*
N.B this is no longer generally true if we impose a non-negativity constraint on the unfolded result

- Measure ***selected*** number of signal events in bins of a ***reconstructed*** quantity

Efficiency correct

Unfolding

- Want the ***total*** number of signal events in bins of a ***true*** quantity

---

Number of events in reco bin j

$$R_j = \sum_{True\ Bins, i} S_{ji} T_i$$

Number of events in true bin i

Smearing matrix

Number of events in true bin i

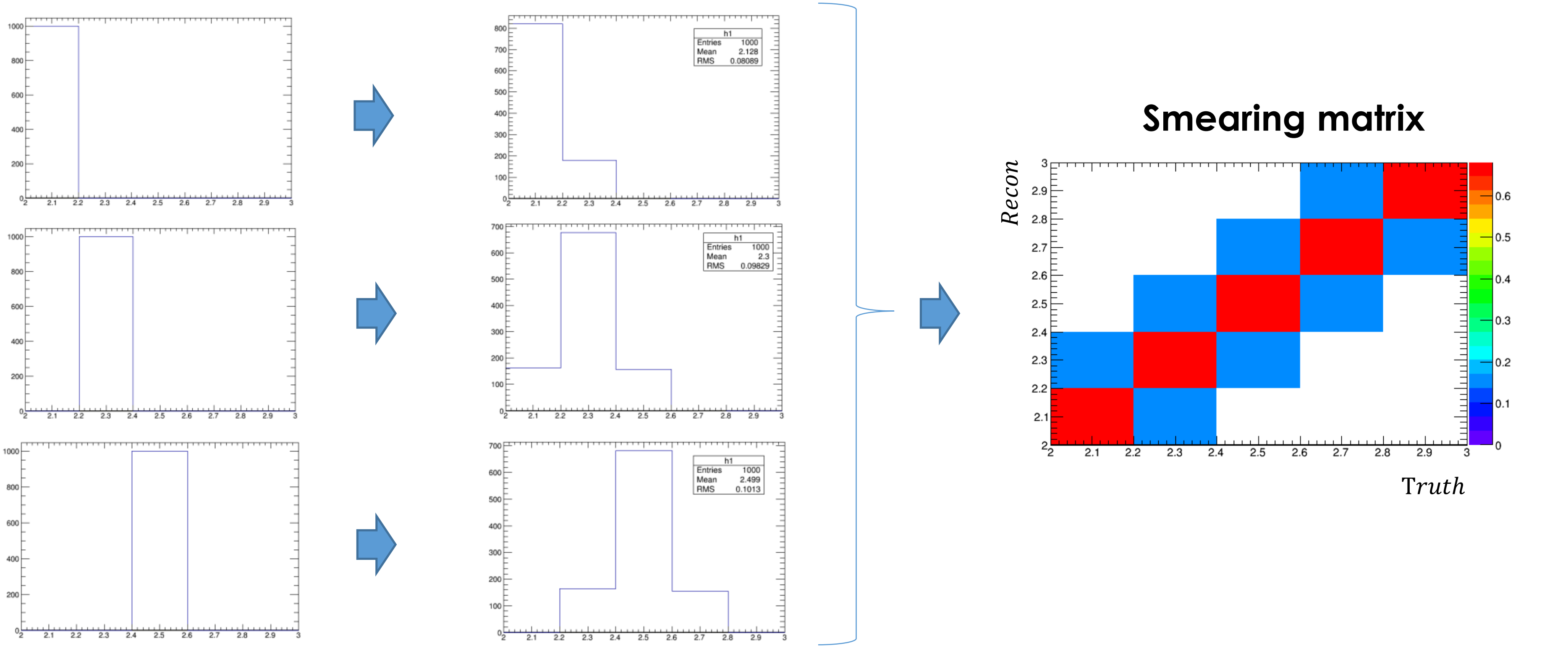
$$T_i = \sum_{Reco\ Bins, j} U_{ij} R_j$$

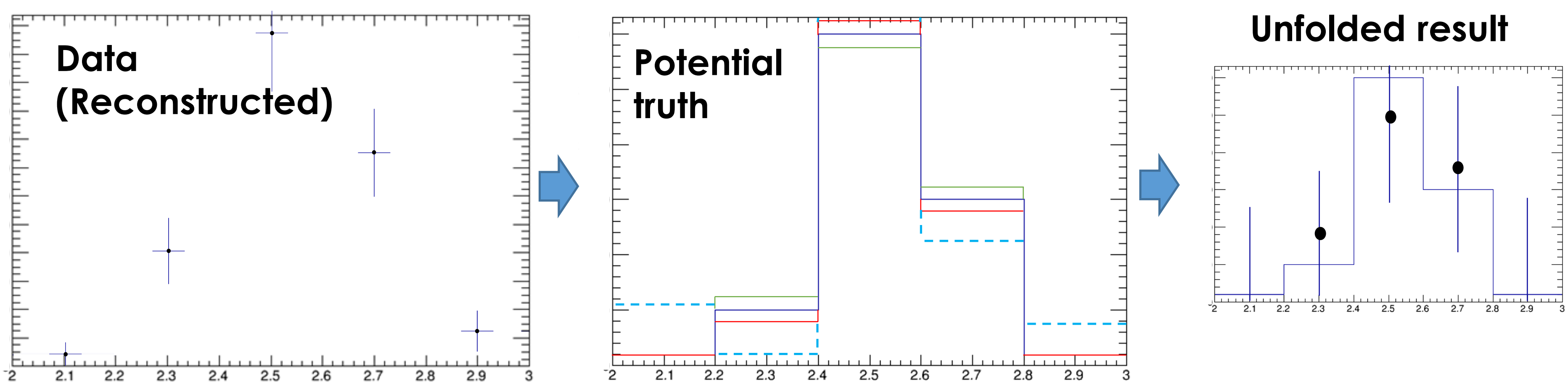
Number of events in reco bin j

Unsmearing matrix

- Smearing matrices can have little model dependence
  - In principle can build without any neutrino scattering model at all
- Unfolding is finding U_{ij} from S_{ji}. Simplest method: use S_{ji}^{-1}
  - **Any other method of getting U_{ij} gives larger errors or is biased***
  - But lots of ways to arrange reco bins to give same true bin contents → degeneracies in solution → strong anti-correlations -"ill-posed problem"

True Distributions **Reconstructed Distribution**





Resolving the ill-posed problem

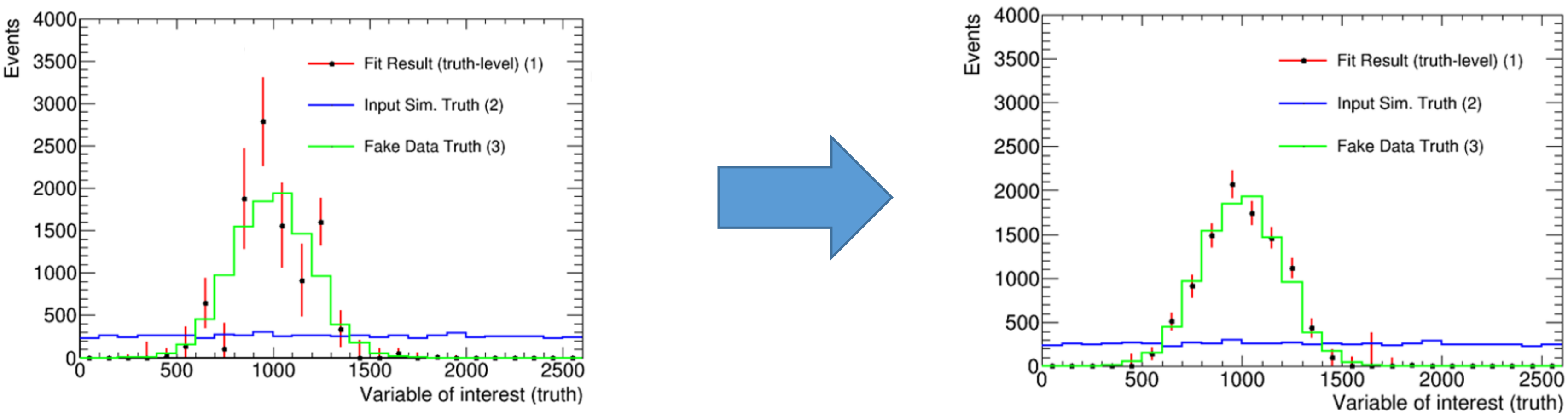
- Unfolding methods mostly differ in the way they resolve these degeneracies (i.e. their **regularisation** implementation)
- Ideally, regularisation should be selecting the "smoothest" of many (almost) degenerate solutions

Most common methods on T2K:

- *Previously*: D'Agostini (1995) "Baysian" Iterative Unfolding
- *Now*: Likelihood template fitting (with optional Tikhonov regularisation)
- *Future*: D'Agostini (2010) Iterative Unfolding / MCMC ???

Resolving the ill-posed problem

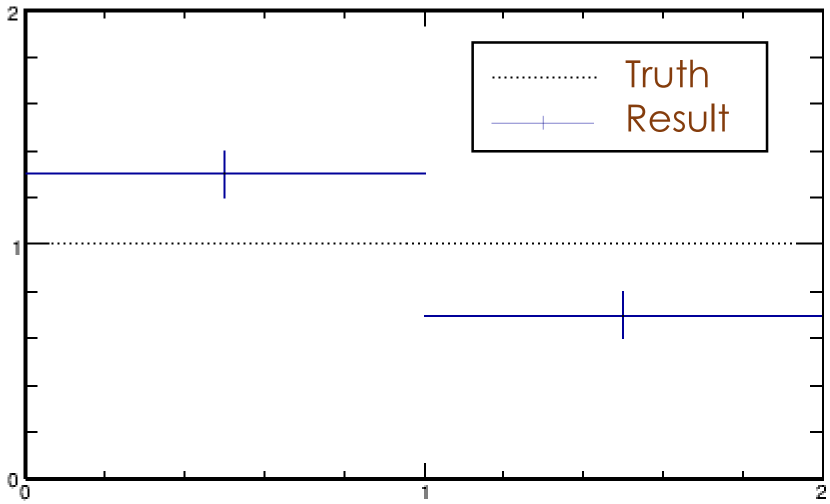
- Unfolding methods mostly differ in the way they resolve these degeneracies (i.e. their **regularisation** implementation)
- Ideally, regularisation should be selecting the "smoothest" of many (almost) degenerate solutions

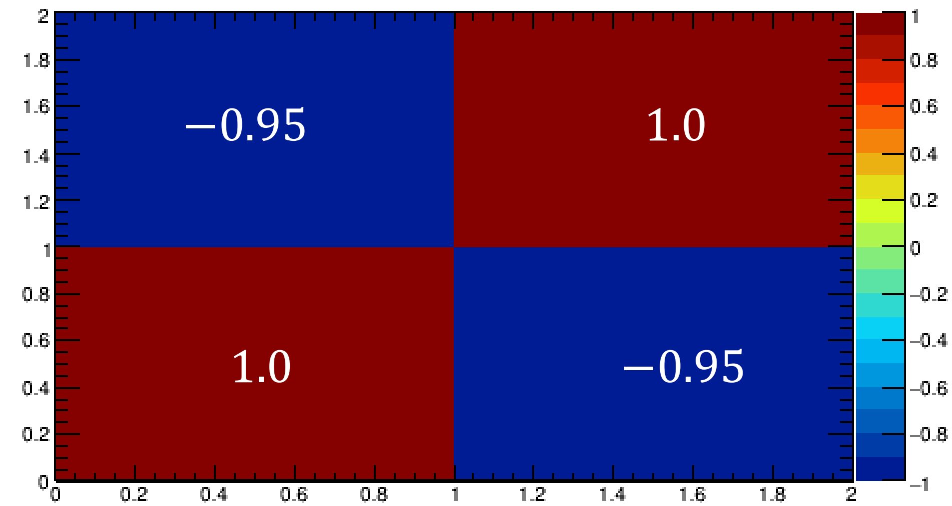


- **Regularisation always adds some bias**
- The unregularised result is the most "correct" representation of the true unfolded result

But the result looks awful!?

- Consider a two bin result:





$$pull_i = \frac{N_{fit} - N_{true}}{Error}$$

$$pull_0 = 3$$
$$pull_1 = 3$$

Fairly awful pull

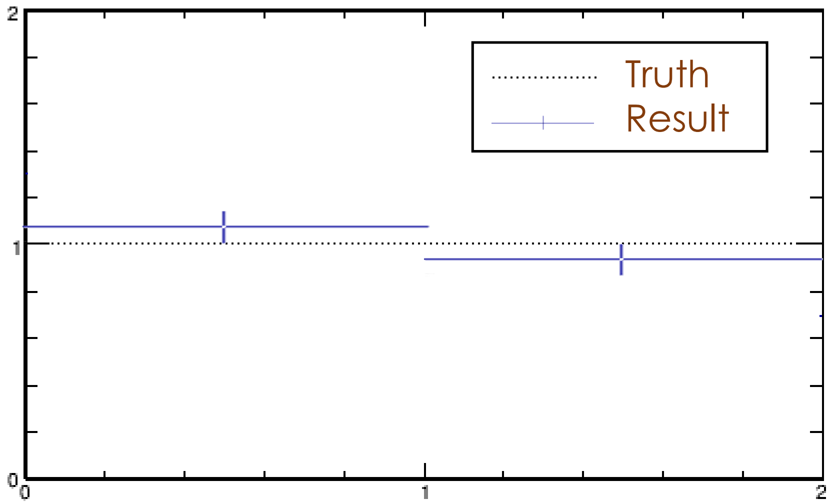
$$\chi^2 = (\overline{N_{fit}} - \overline{N_{true}})(V_{cov})^{-1}(\overline{N_{fit}} - \overline{N_{true}})$$

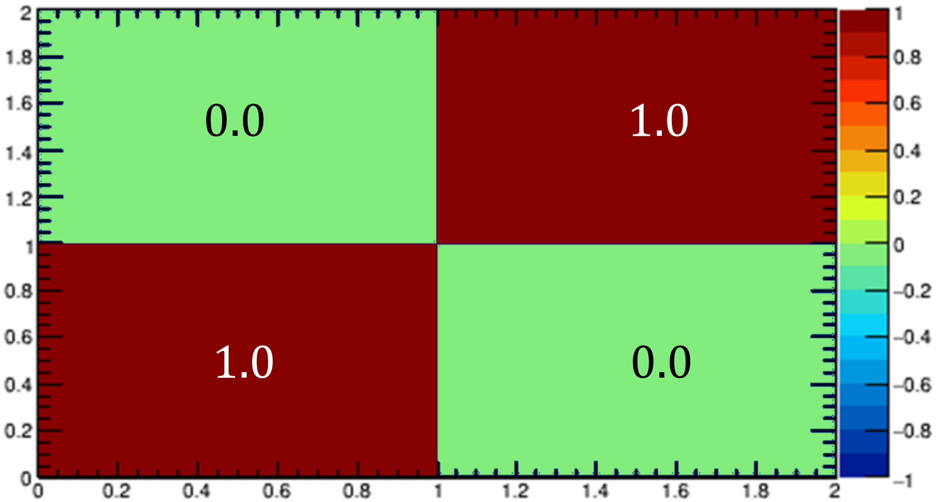
$\chi^2 = 1.69$ **Good χ^2**

- Need to see the correlation matrix to tell whether the result is good or not.

But the result looks awful!?

- Consider a two bin result:





$$pull_i = \frac{N_{fit} - N_{true}}{Error}$$

$$pull_0 = 1$$
$$pull_1 = 1$$

Better pull

$$\chi^2 = (\overline{N_{fit}} - \overline{N_{true}})(V_{cov})^{-1}(\overline{N_{fit}} - \overline{N_{true}})$$

$$\chi^2 = 2.0$$

Worse χ^2

- Pulls/bin-to-bin bias doesn't tell the whole story

D'Agostini (1995) iterative unfolding

- Using Bayes' theorem*:

Smearing Matrix

MC Prior

Unsmearing Matrix

$$U_{ij} = \frac{P_{rel}(r_j|t_i)P_0(t_i)}{\sum_{i=1}^{i=N_t} P(r_j|t_i)P_0(t_i)}$$

$$P(r_j|t_i) = N_{ij}^{MC}/T_i^{MC}$$

$$P_{rel}(r_j|t_i) = \frac{P(r_j|t_i)}{\sum_{j=1}^{j=N_r} N_{ij}^{MC}/T_i^{MC}}$$

$$P_0(t_i) = \frac{T_i^{Prior}}{\sum_{i=1}^{i=N_t} T_i^{Prior}}$$

N_{ij} - number of events in true bin i and reco bin j
T_i - number of events in true bin i
r_j/t_i - reco/true bin j/i

- If prior formed from MC (as it typically is), model dependence is explicit

- To mitigate:
  1. Found U_{ij} → calculate T_i^{Unfold} (Unfolded result)
  2. Use T_i^{Unfold} as T_i^{Prior} and recalculate U_{ij}
  3. Return to step 1

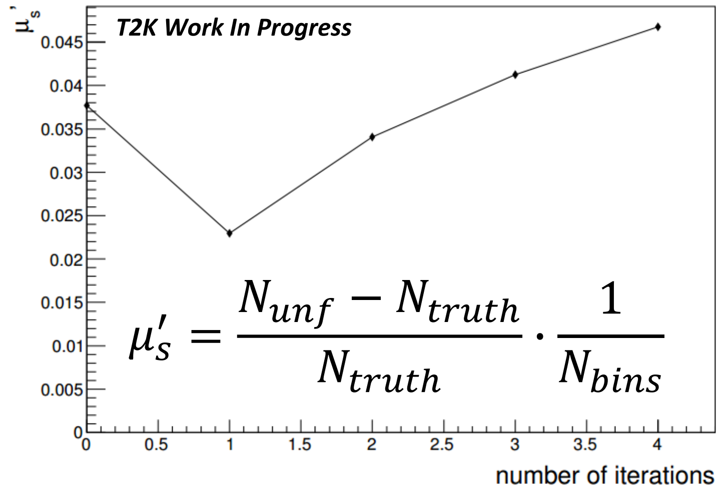
- Each step reduces reliance on the MC but decreases reg. strength
- Stat. errors increase with each step
- Many steps ≡ S_{ji}^{-1} **

- Can attempt to balance smoothness and bias by truncating iterations
  - This is a fairly ill defined procedure that was only optimised on the MC

**Although this method uses Bayes' theorem, it is not a Bayesian technique (in fact it's equivalent to the widely-used "Expectation-maximisation algorithm") [M.Kuusela]*
***This is only generally true if we do not enforce a non–negaitivity constraint*

D'Agostini (1995) iteration optimisation

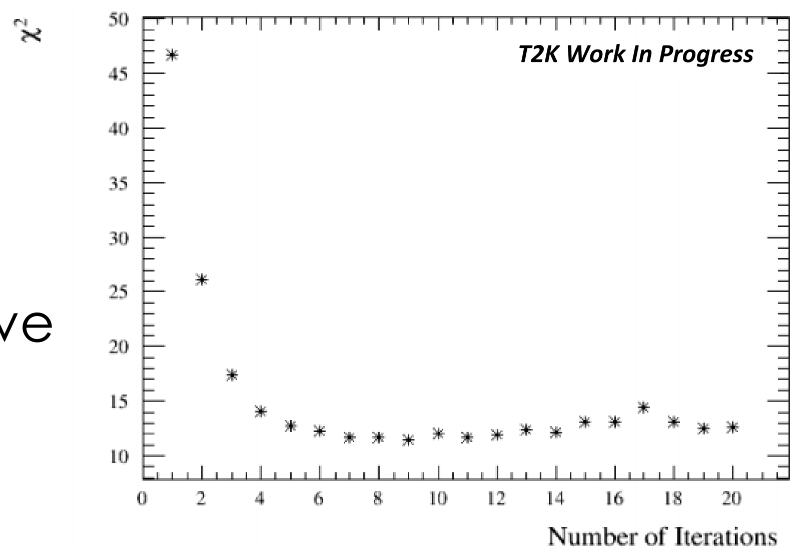
- Previously looked for convergence in bin-by-bin bias in fake data studies:
- Choose point of minimal bias
- Not so good - disregards correlations



- Alternative: calculate χ^2 for each iteration

$$\chi^2 = (\vec{N}_{unf} - \vec{N}_{true})(\bar{\bar{V}}_{cov}^{unf})^{-1}(\vec{N}_{unf} - \vec{N}_{true})$$

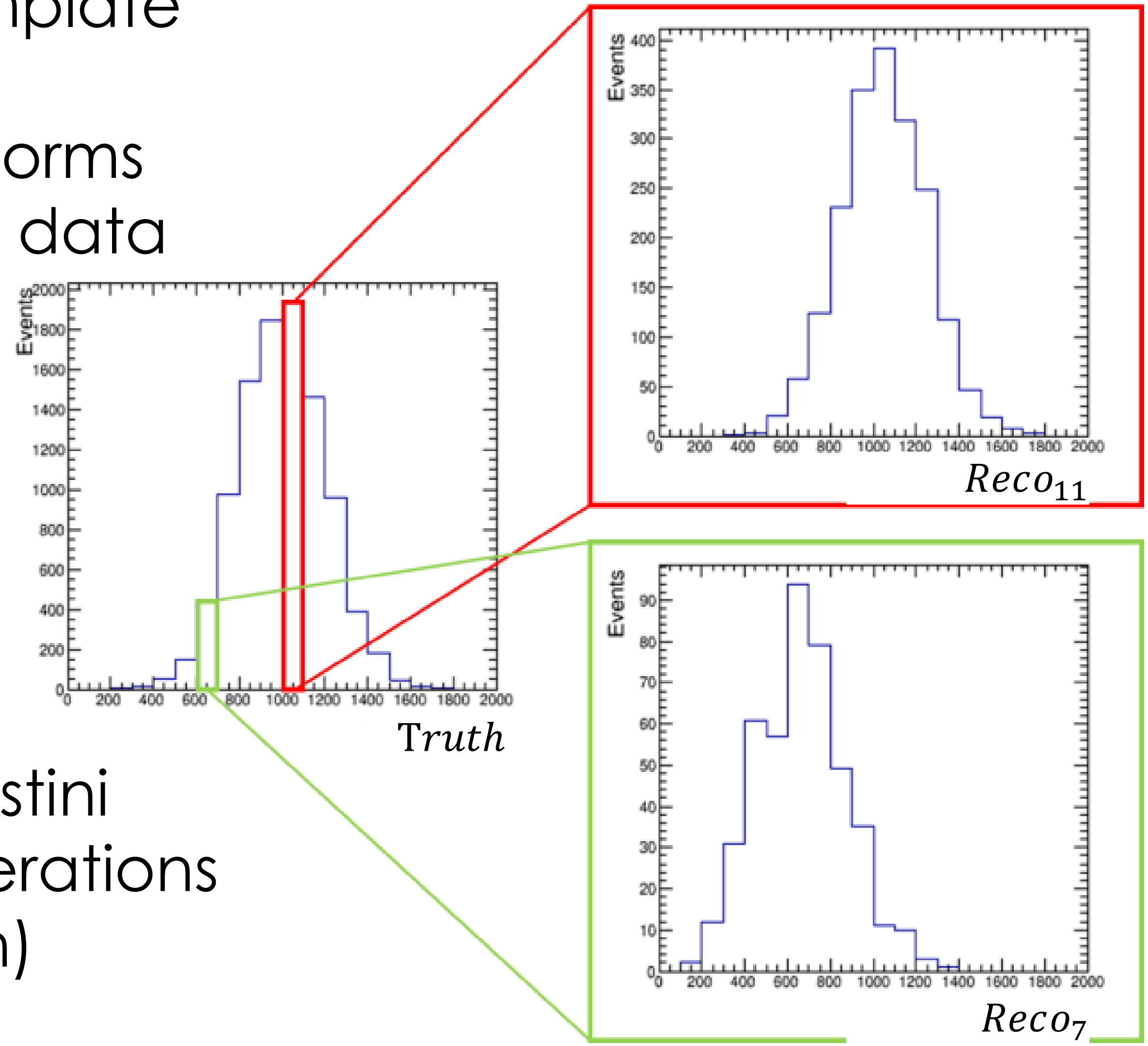
- Choose highest curvature on the curve
- Better – includes correlations



Still, this optimisation has to be **tuned based on fake data studies** and regularisation **biases the result**.

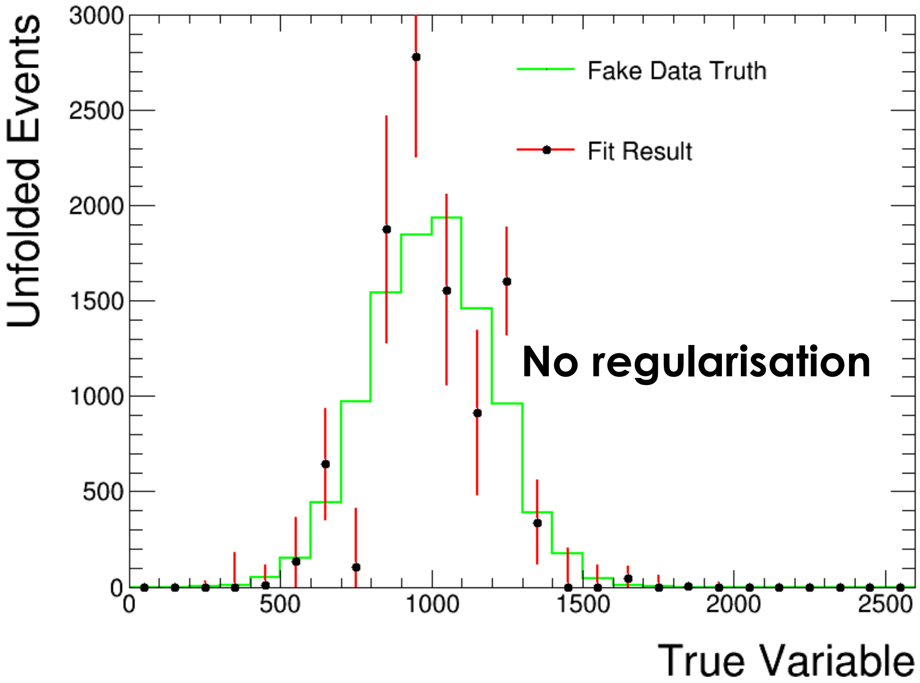
Binned likelihood fitting

- True bin → Reco. template
- Vary MC template norms (c_i) and compare to data
- Maximise Poisson likelihood + syst. penalty term
  (using max. gradient decent)
- Equivalent to D'Agostini (1995) with infinite iterations (for no regularisation)



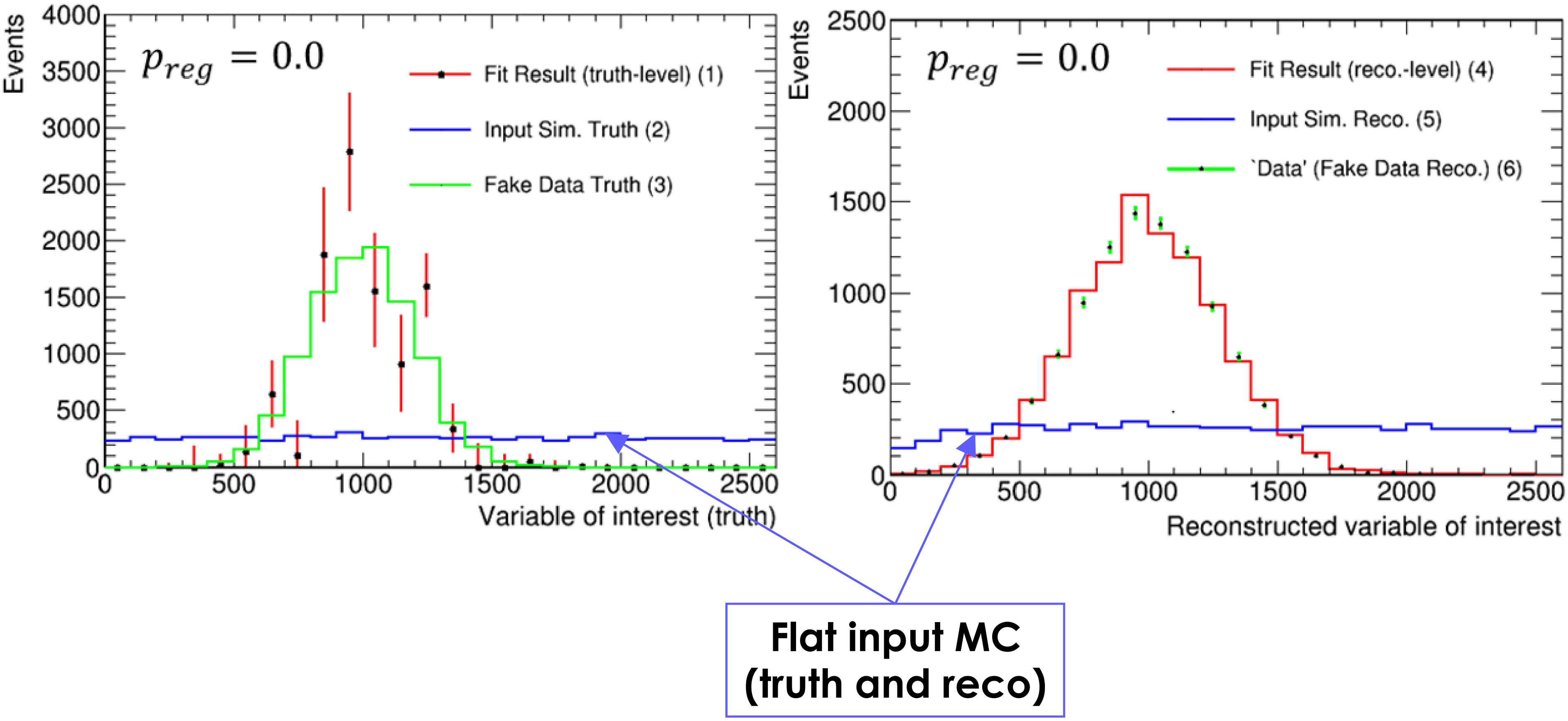
The ill-posed problem in fit results

- If there is significant smearing between bins → ill-posed problem
- Seen as a "zig-zagging" result with strong anti-correlations between bins
- Can apply regularisation to penalise such results.
- Many ways to regularise, best method depends on the analysis.
- One option:

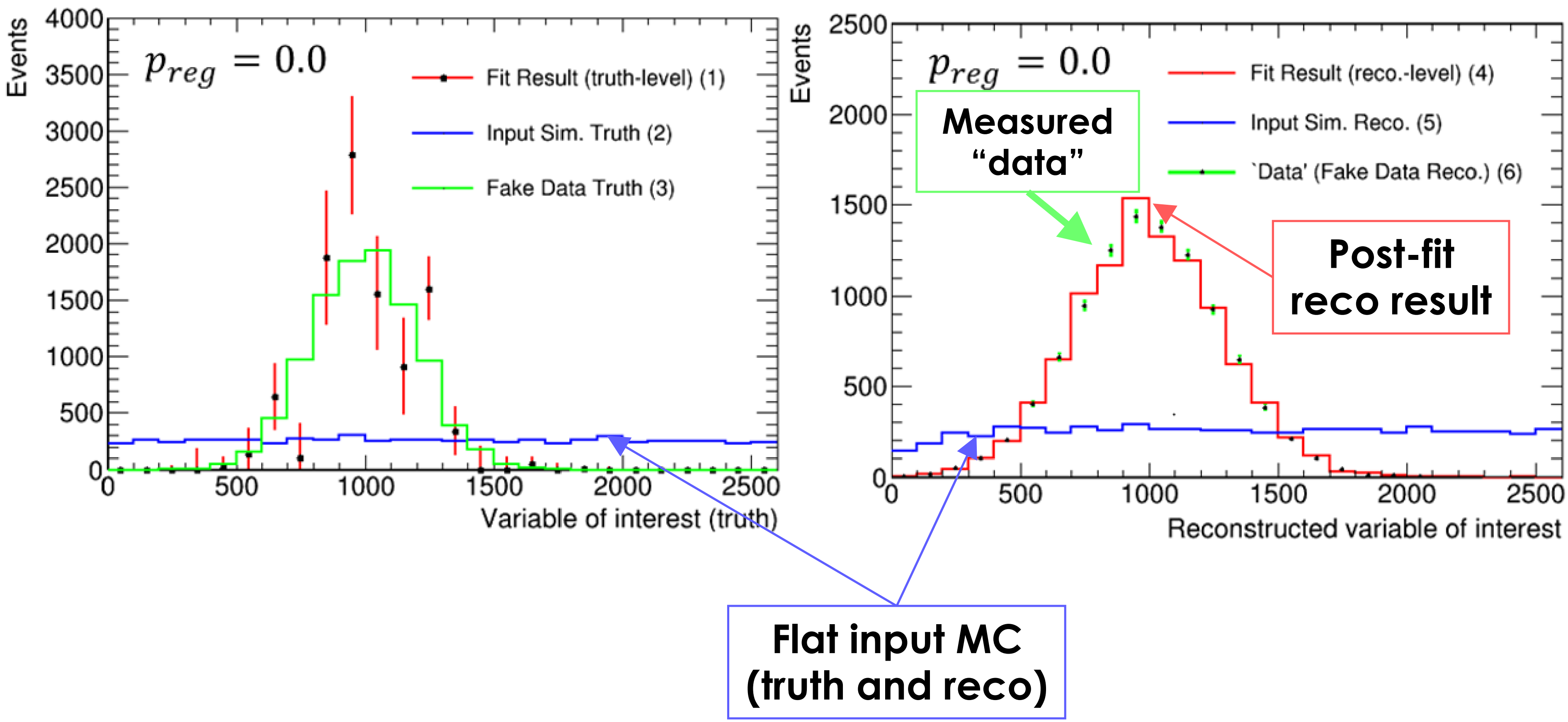


$$\chi^2_{reg} = p_{reg} \sum_{i}^{truebins-1} (c_i - c_{i+1})^2 = p_{reg}(\vec{c} - \vec{c}_{prior})(V^{reg}_{cov})^{-1}(\vec{c} - \vec{c}_{prior})$$

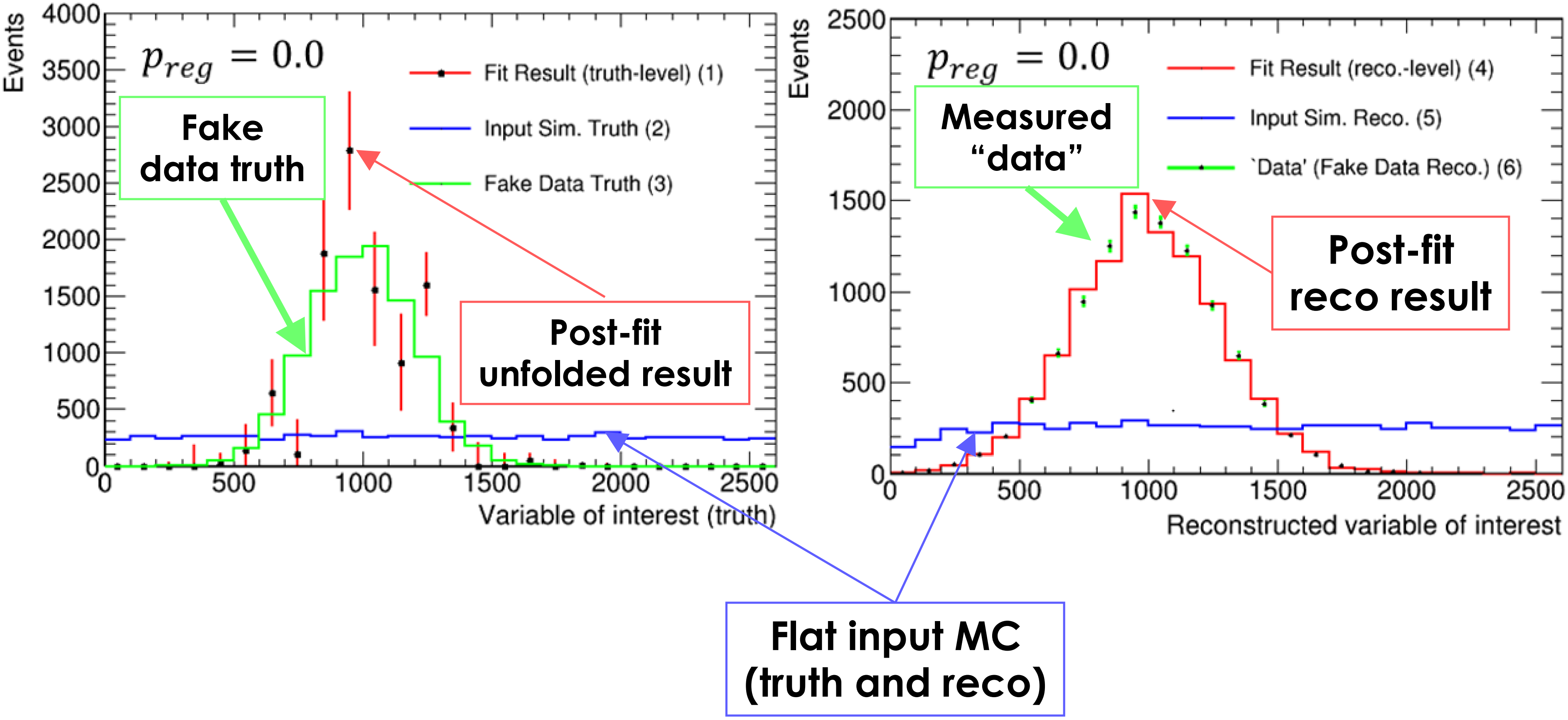
The role of regularisation



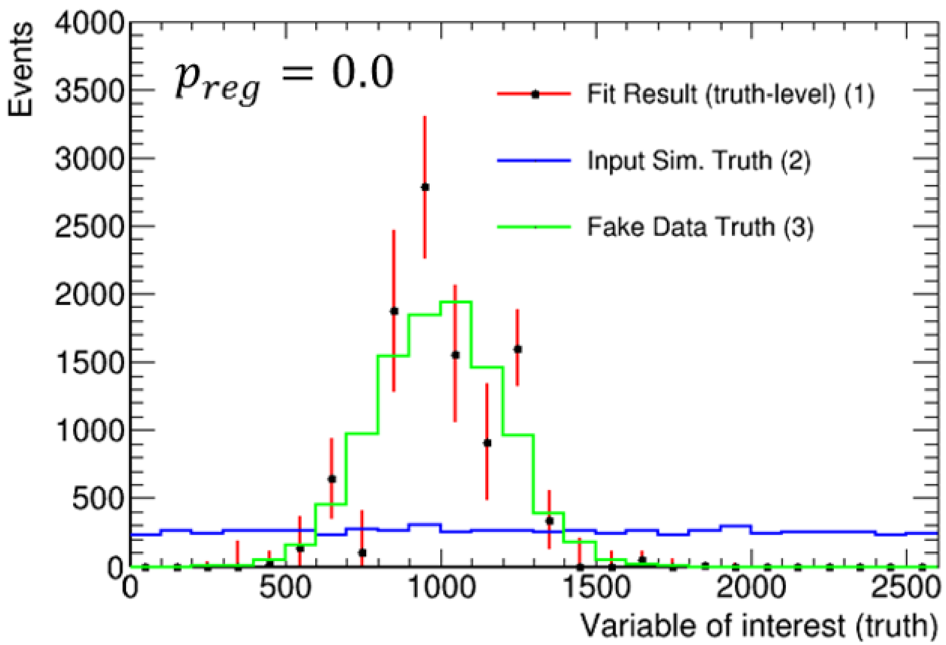
The role of regularisation

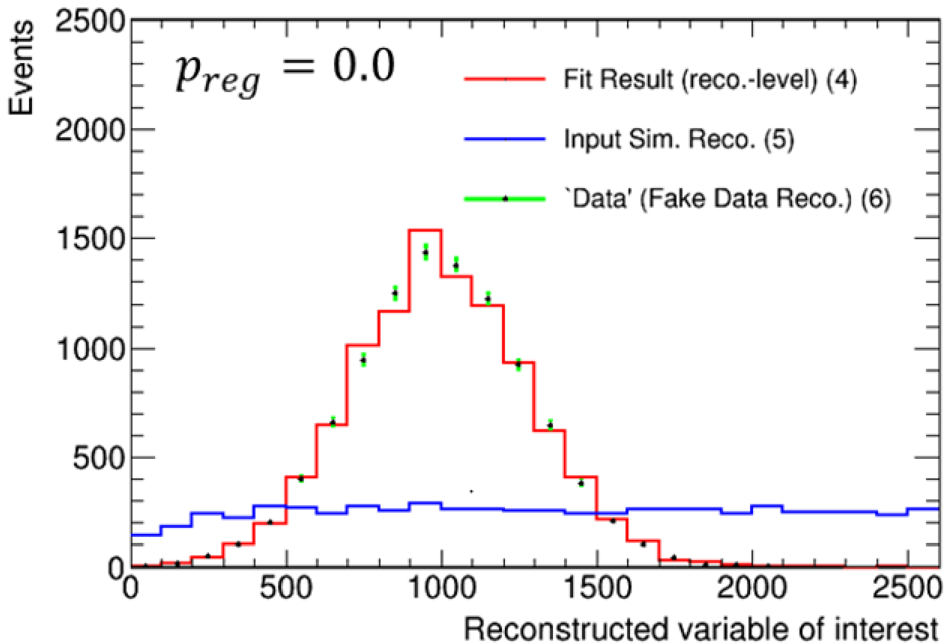


The role of regularisation

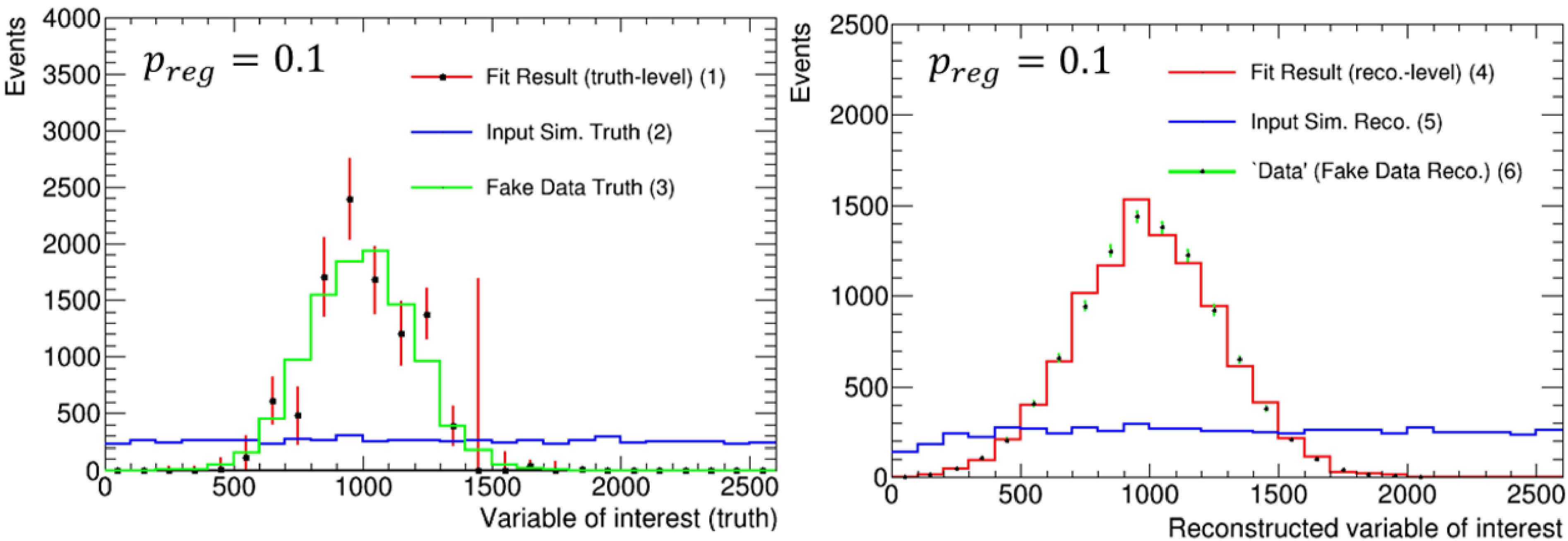


The role of regularisation

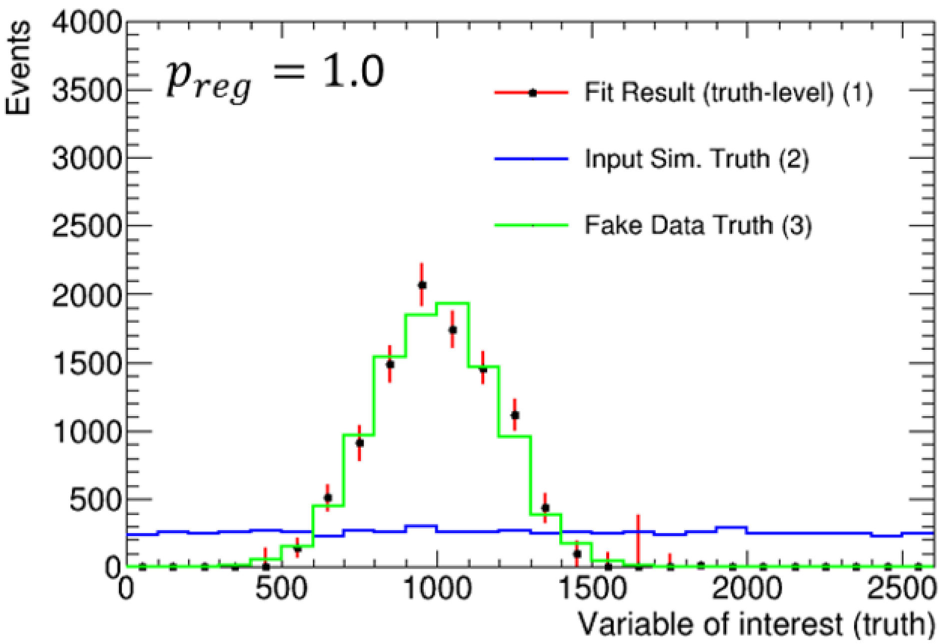


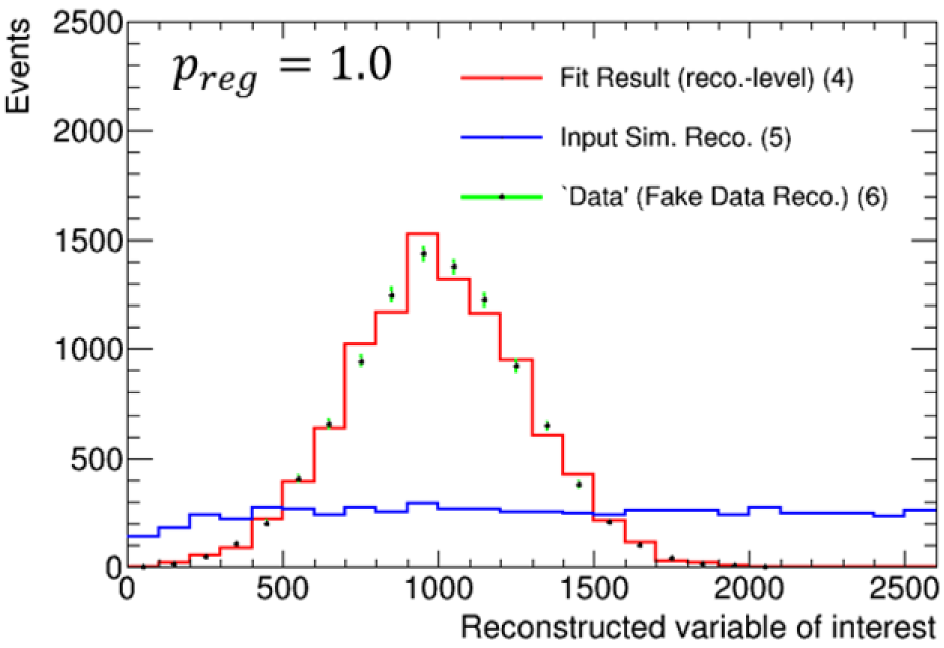


The role of regularisation

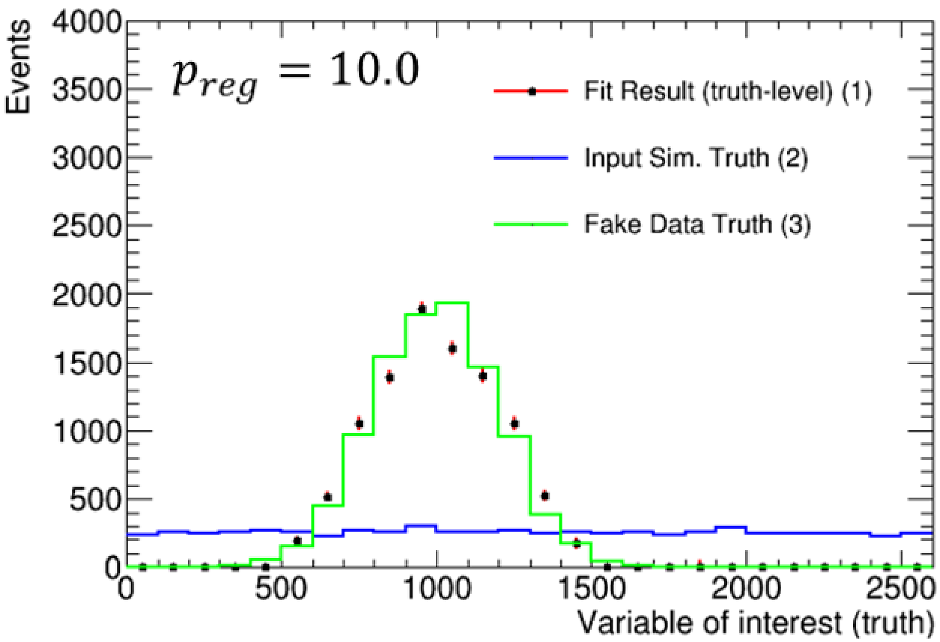


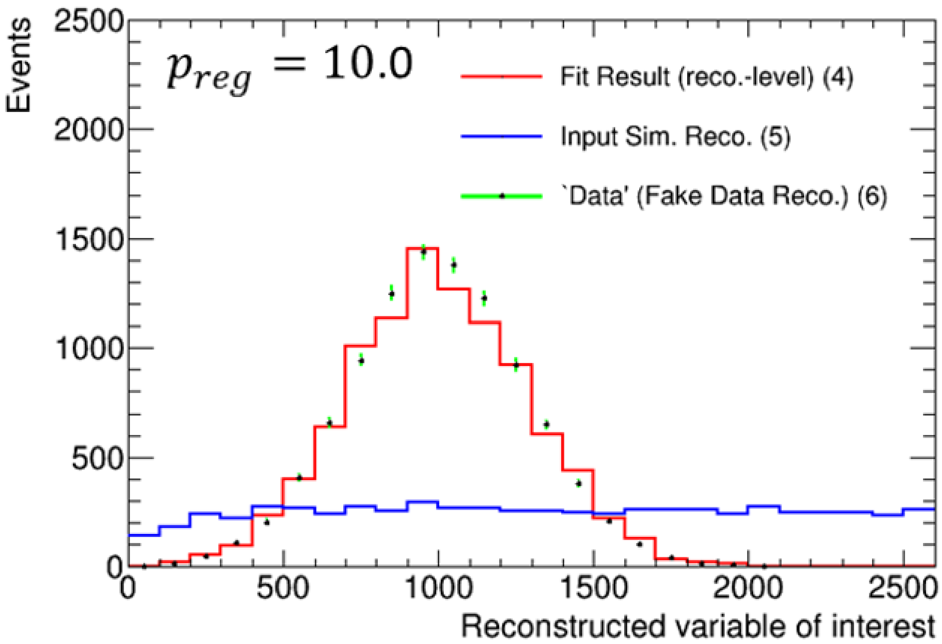
The role of regularisation



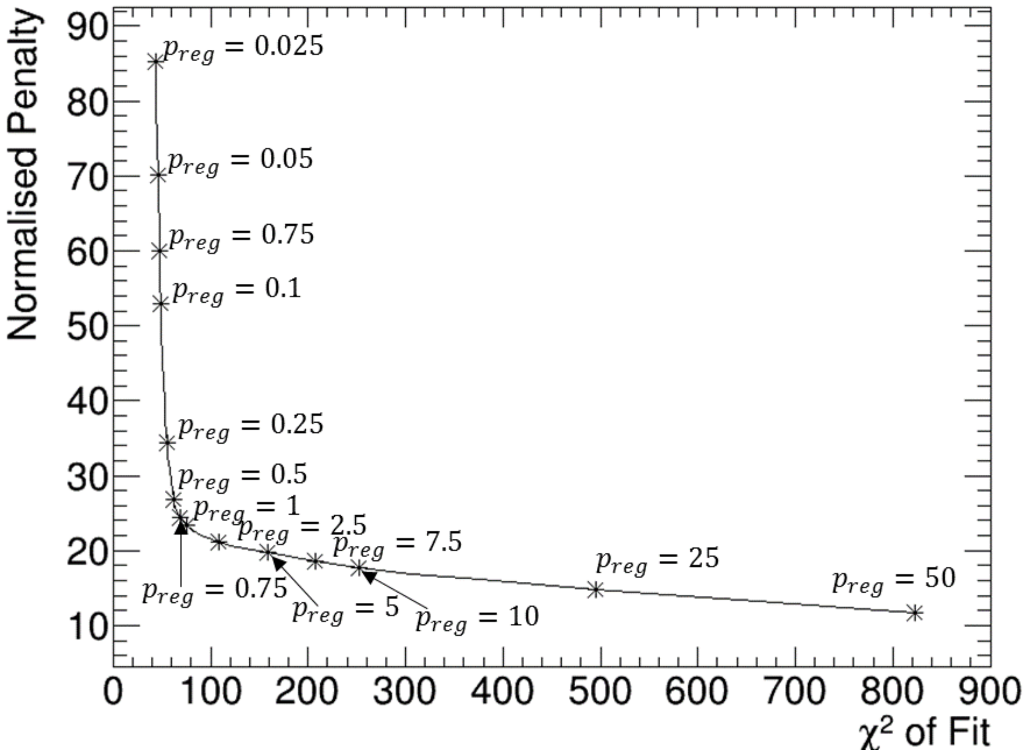


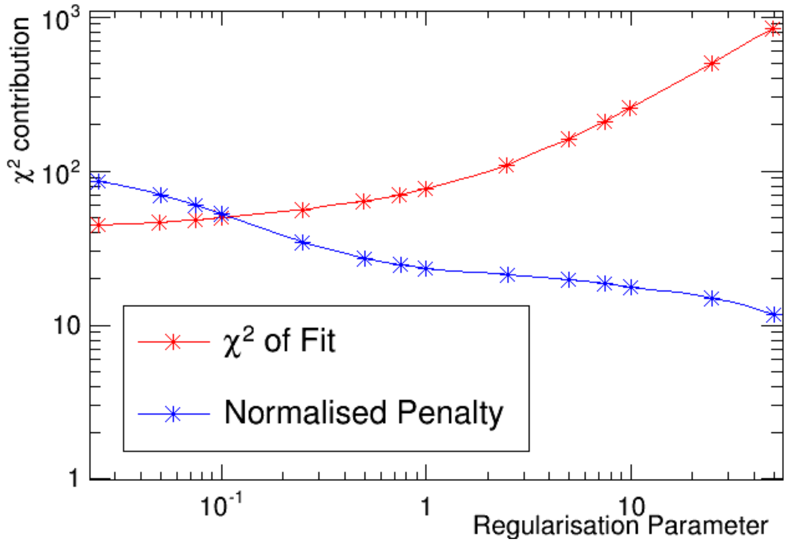
The role of regularisation





The role of regularisation





$$Normalised\ Penalty = \chi^2_{reg}/p_{reg}$$

- Best p_{reg} is the kink of the curve (in this case ~1)
- Balances regulation (in this case smoothness) with bias
- **L-curve can be formed on real data** – data driven regularisation

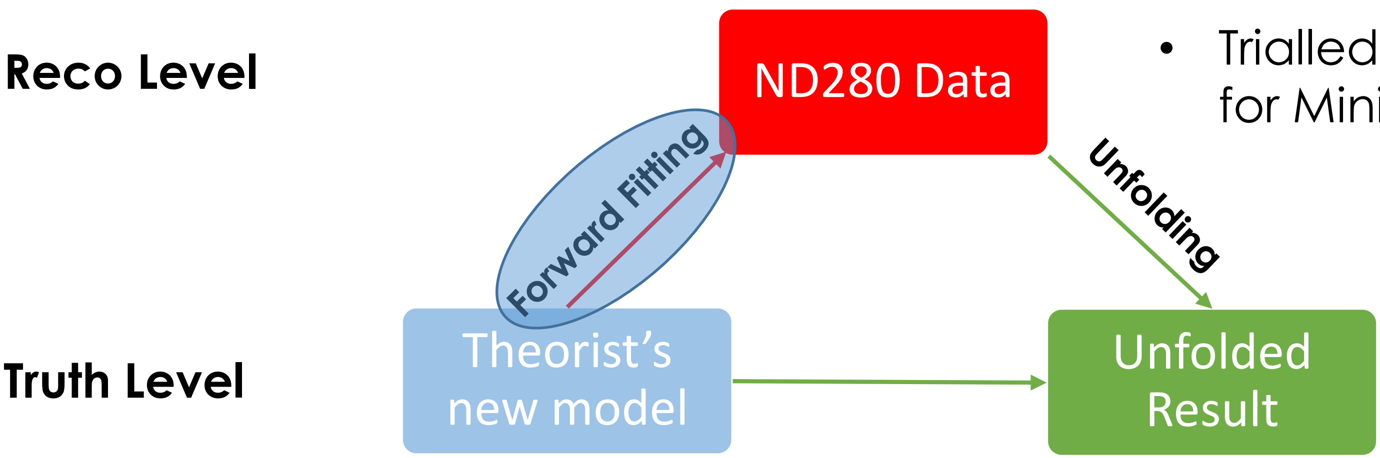
http://epubs.siam.org/doi/abs/10.1137/1034115
http://epubs.siam.org/doi/abs/10.1137/0914086 http://arxiv.org/pdf/1205.6201v4.pdf - use in TUnfold

How not to unfold

- **Unfolding is often either ill-posed or biases the result**
  "Only useful for ~~comparing data~~ or obtaining a plot for posterity" – Louis Lyons

- The best way to gain useful physics from data is to compare the model and measurement **at the recon level.**

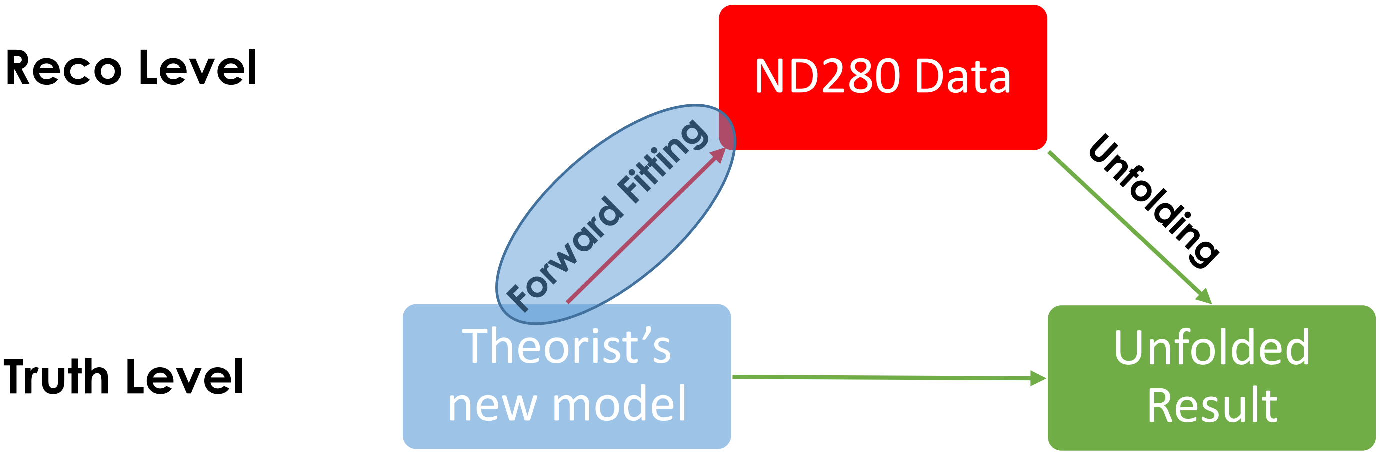
- Need to provide tools for model builders to smear their models
  - Provide the smearing matrix to facilitate forward fitting!



- Trialled by D. Perevalov for MiniBooNE NCE (2009)

How not to unfold

- Feasibility of providing useful tools is an open question:
- Do we need a smearing matrix for every permutation of final state topology?
- How do we deal with backgrounds that theorists can't predict (e.g. OOFV)?



Unfolding, uncertainties and model dependence

- Very easy to introduce subtle model dependence to cross-section measurements

- Naïve implementations of D'Agostini-like unfolding can cause this

- Correct propagation of uncertainties is not trivial

- Wish list for unfolding:
  - **Provision of the unregularised result**
  - Transparent optimisation of the regularisation
  - Ideally a data-driven regularisation
  - Well-motivated and clear propagation of uncertainties
  - Fake data/bias studies (next section)

Overview

- T2K and ND280
- Choosing a signal definition
- Choosing a selection
- Choosing a binning
- Efficiency corrections
- Unfolding, uncertainties and model dependence
  - Regularisation
  - D'Agostini (1995)
  - Likelihood fitting
  - How not to unfold
- Fake data and bias studies
- Conclusions

Fake data and bias studies

- Input simulation is typically NEUT with either SF or RFG+RPA nuclear model
- Need to ensure that the cross-section extraction method used is not biased toward this
- Test cross-section extraction on a variety of fake data to check that we can recover the truth – fake data studies
- Test the impact of regularisation and of control regions to evaluate possible bias – bias studies

Validating the result – fake data

- Want to validate the cross-section extraction is working properly and that there is sufficient freedom to fit a comprehensive range of plausible data.

- Test cross-section extraction on a variety of fake data to check:

  - Asimov (input=fake data)
  - Stat. and syst. fluctuations of input

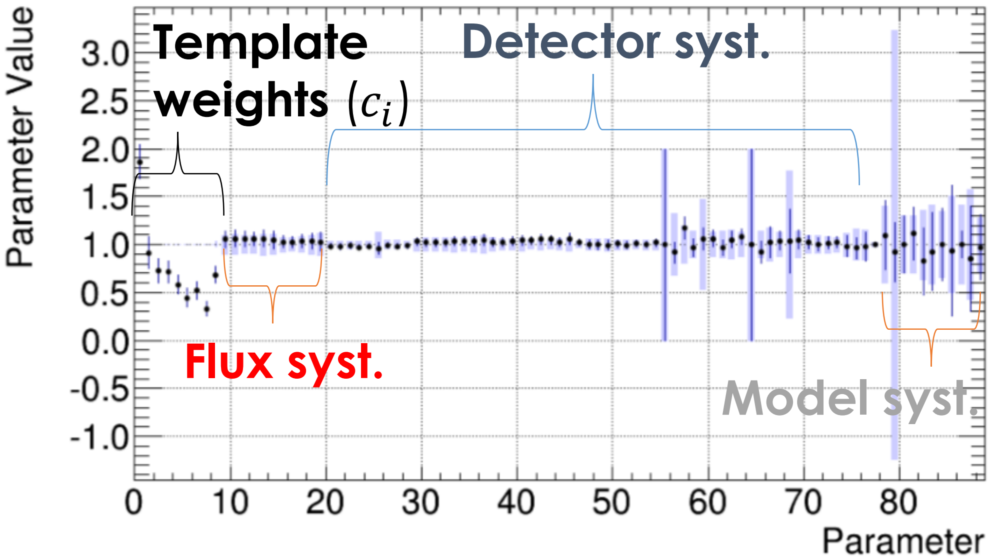
These test whether the fitter is actually working

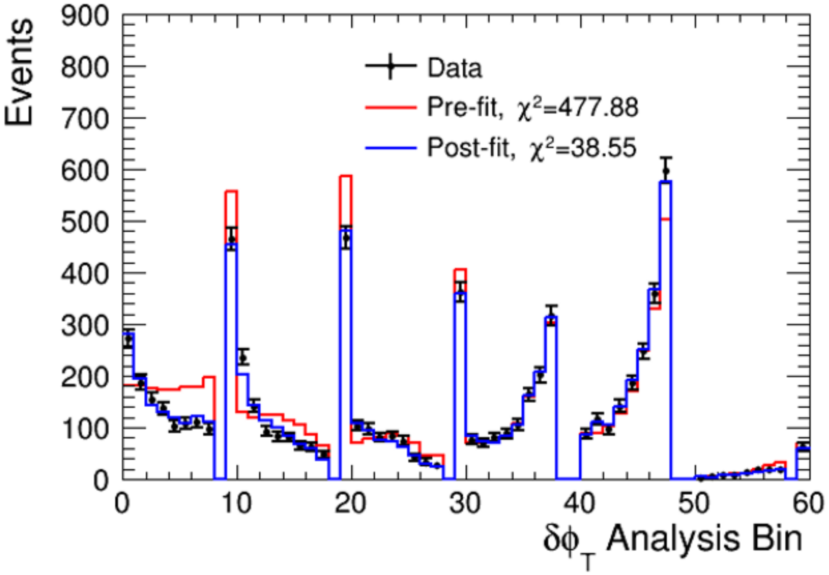
  - NEUT with different parameters
  - GENIE (2.8.0 – no 2p2h and BR-RFG)
  - NuWro (11q – LFG)
  - Custom reweightings

These test model dependence: tests if background model systematics sufficient and extent of bias to signal model.

Example: GENIE fake data

Fit result

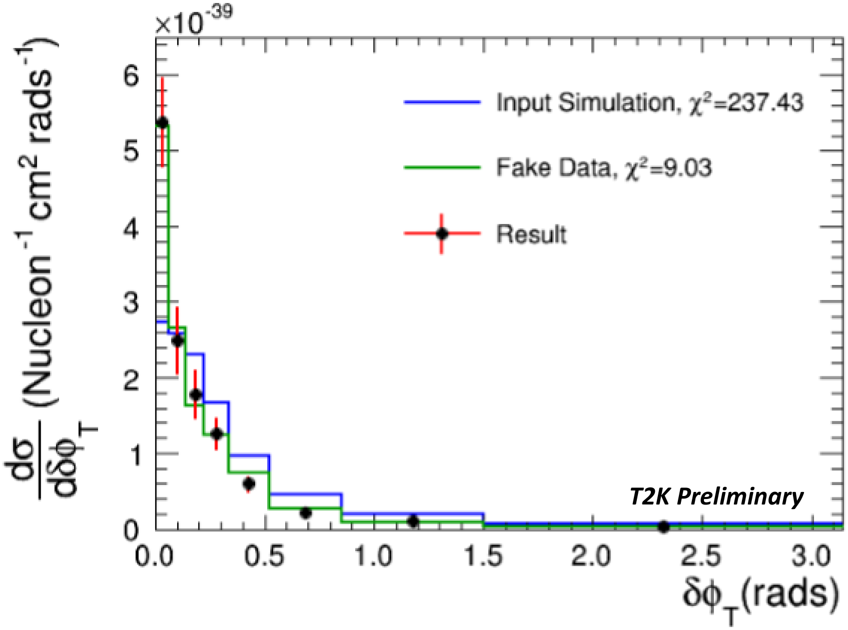


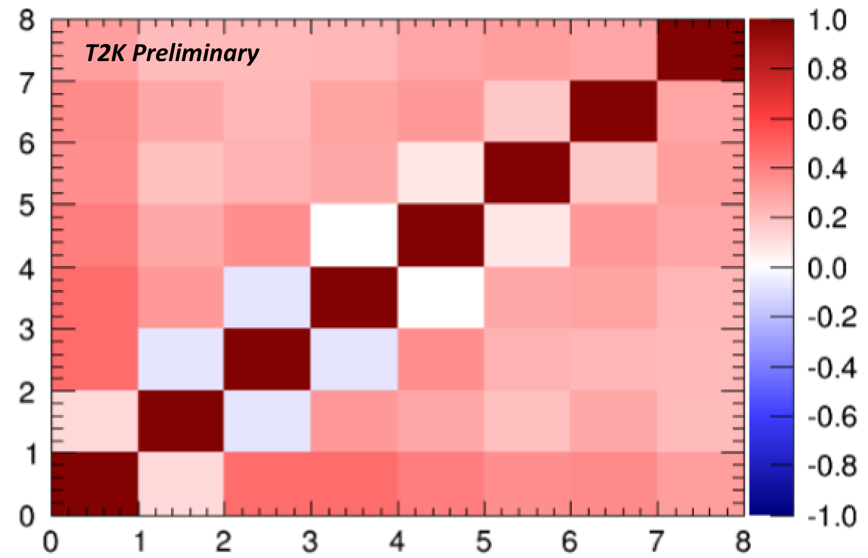


- Parameters other than template weights should be ~1 since GENIE simulation used similar flux, background model and detector simulation
- Covariances should be understood
- Post-fit result should well characterise the fake data (can test this for real data too)

Example: GENIE fake data

Cross-section result



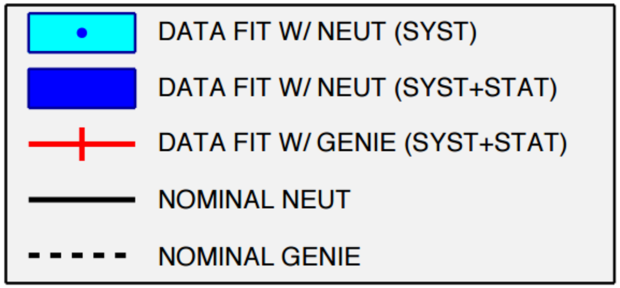


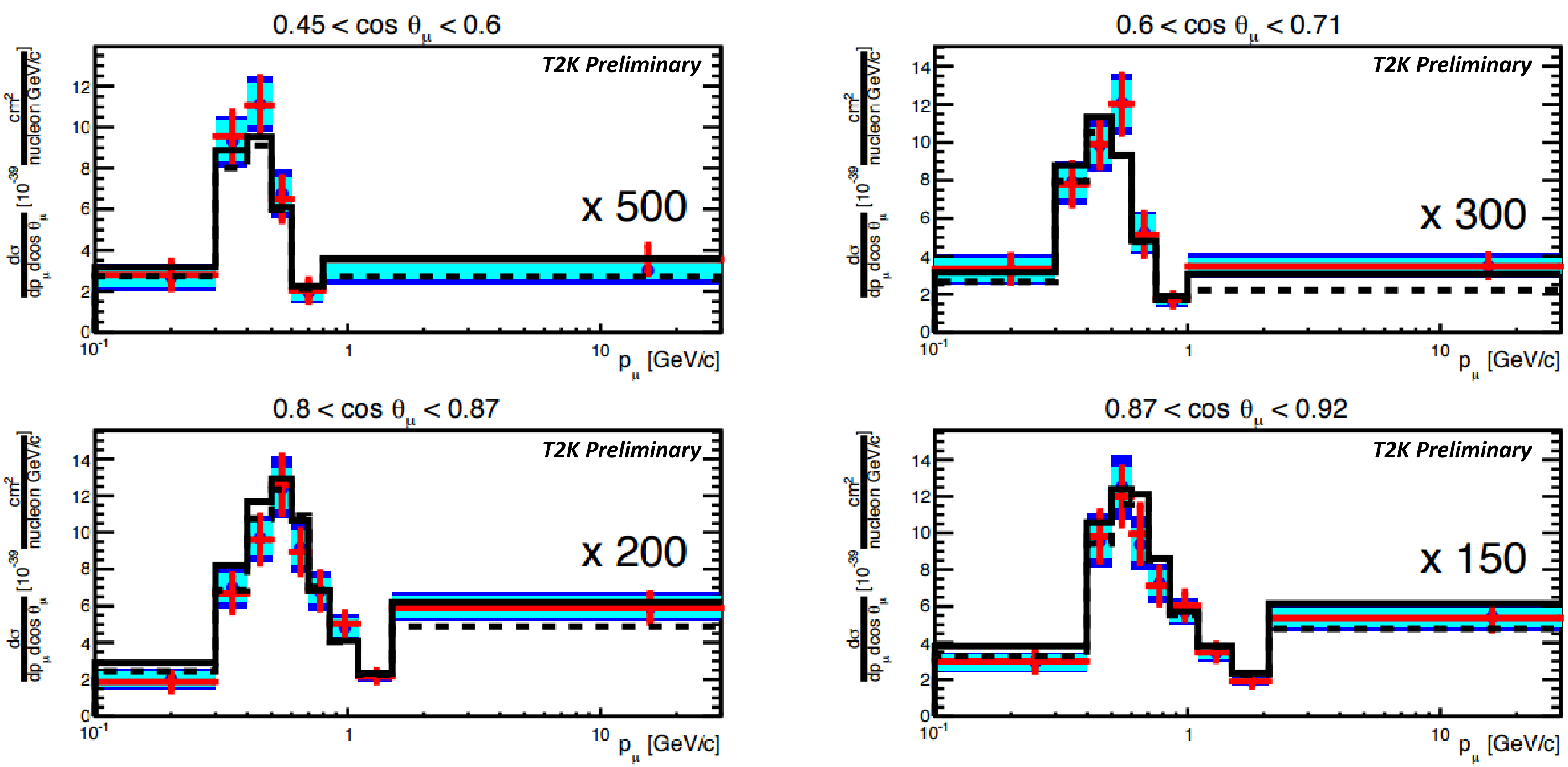
- Check that post-fit cross-section result is in good agreement with the fake data truth
- Check covariance look sensible

Example: Real data with different inputs

Cross-section result

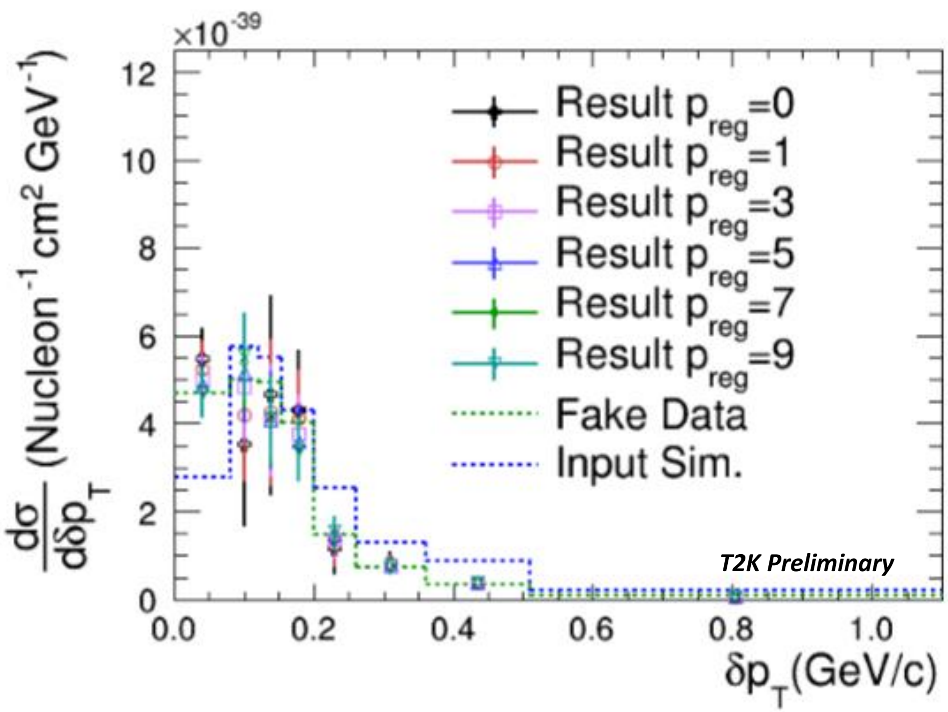
- Result is independent of whether GENIE or NEUT is used as an input

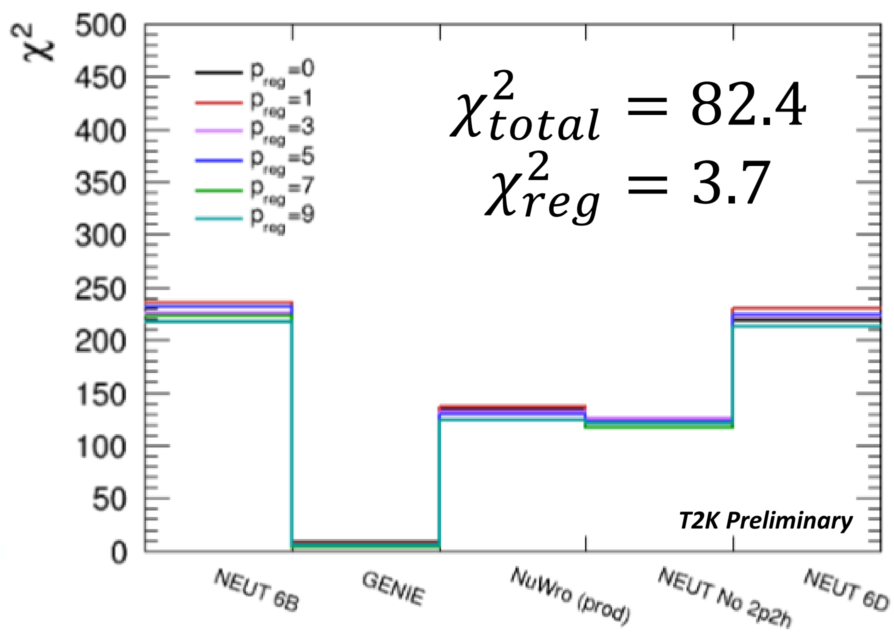




Regularisation bias studies

- Also important to explicitly check bias due to regularisation:
  - Compare reasonable alterations in regularisation strength
  - Check sensitivity isn't much altered by regularisation
  - If fitting: Assess regularisation contribution to the likelihood
  - Check compatibility of result with unregularised result





Overview

- T2K and ND280
- Choosing a signal definition
- Choosing a selection
- Choosing a binning
- Efficiency corrections
- Unfolding, uncertainties and model dependence
  - D'Agostini (1995 and 2010)
  - Likelihood fitting
  - Interpretation of cross-section errors
  - How not to unfold
- Fake data and bias studies
- Conclusions

Conclusions

- Discussed best practices for cross-section extraction at ND280
  - Signal definition
  - Selection
  - Binning
  - Efficiency correcting
  - Unfolding
  - Fake data/bias studies
- **Key points:**
  - Important to have a clear signal definition
  - Important that signal definition is accessible to the detector
  - **Cutting on "interaction level" variables in a selection is dangerous**
  - As are purity corrections
  - **Extracting a useful differential cross-section in variables other than those that characterise a detectors acceptance is hard!**
  - **Plenty of room for model dependence / dangerous handling of cross-section uncertainties in unfolding** (particularly 1995 D'Agostini)
  - Regularisation is helpful but is ultimately a bias
  - Maybe it's worth considering how to avoid unfolding completely
  - Important to demonstrate minimal model dependence with fake data/bias studies

Backups

See Laura Monfregola's thesis for an excellent breakdown of cuts for a CC0π selection

All sub-detector data quality flags are good. Beam trigger is good.

Construct an appropriate "pull" – the likelihood of the track being a particular particle based on its dE/dx – and cut on it.

Event Quality → Track with vertex in subdetector FV → PIDs → Vetos

Typically uses TPC to identify HMN/P track. Start of track (vertex) is then required to be within subdetector fiducial volume.

Reject events with unwanted tracks.

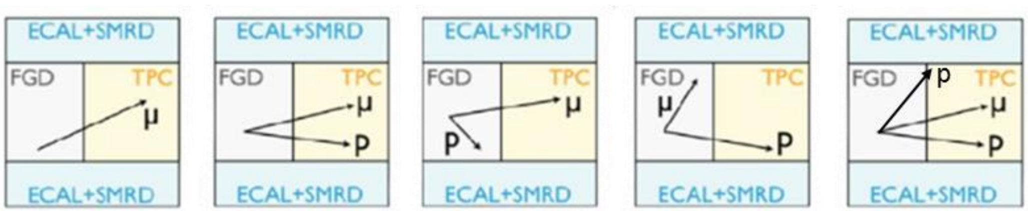
Ensure nothing is entering the target subdetector from the upstream detectors/magnet.

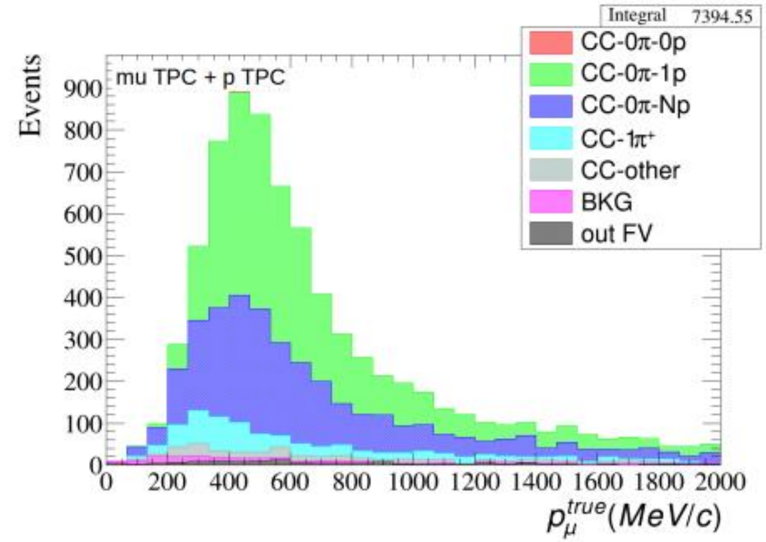
Use Michel electron tags or ECal EM shower vetos to reject/select $\pi^{+/-}$ or π^0.

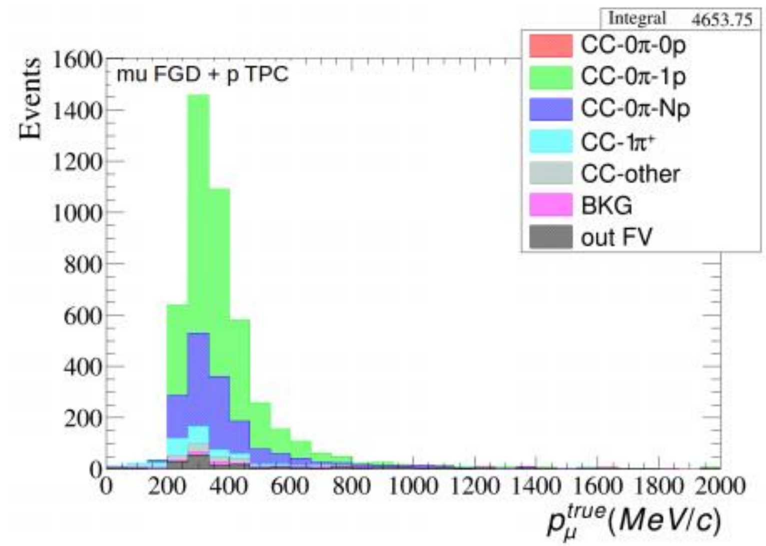
HMN/P – highest momentum negative/positive

Multi-sample selections

- Different subdetectors have very different reconstruction capabilities
- Difficult to untangle detector response if we consider them all together
- Split selection depending on which subdetectors used for recon
- Use of many samples gives a wide kinematic acceptance
- Example from $CC0\pi + Np$ selection:





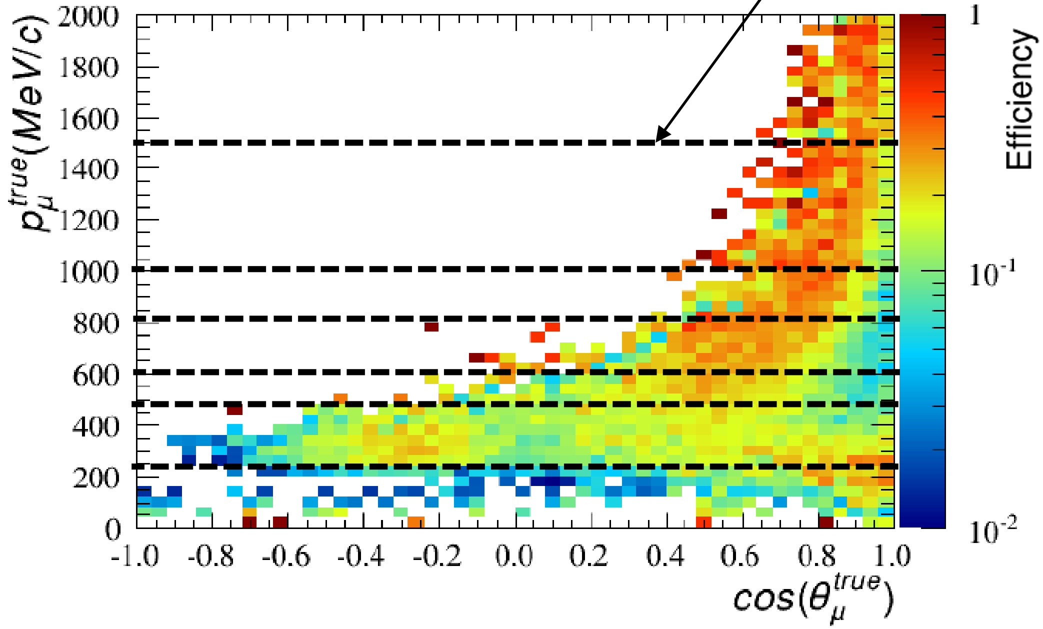


Back to the toy example

Want to measure single-differential p_μ

1D desired p_μ binning

- Consider 2D efficiency

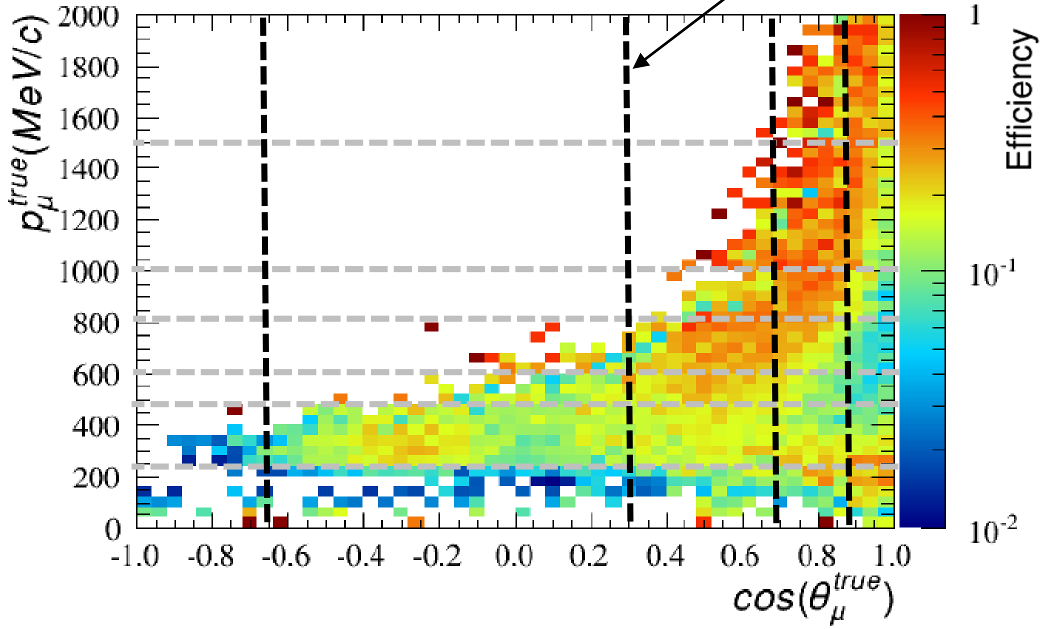


Back to the toy example

Want to measure single-differential p_{μ}

Coarse $\cos(\theta_{\mu})$ binning

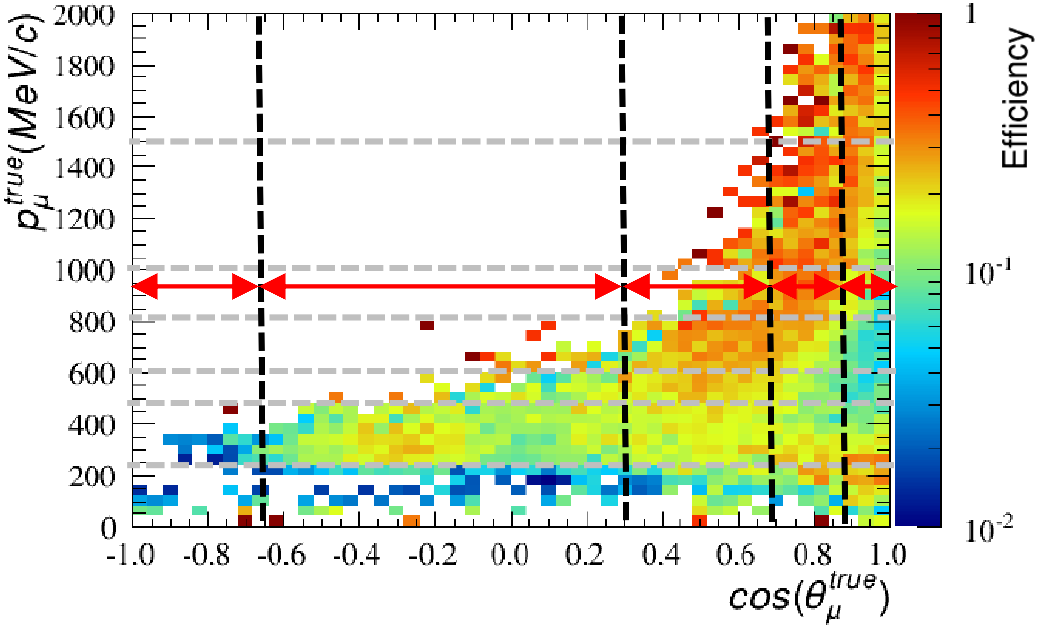
- Consider 2D efficiency
- Bin just fine enough in $\cos(\theta_{\mu})$ such that ϵ~flat



Back to the toy example

Want to measure single-differential p_μ

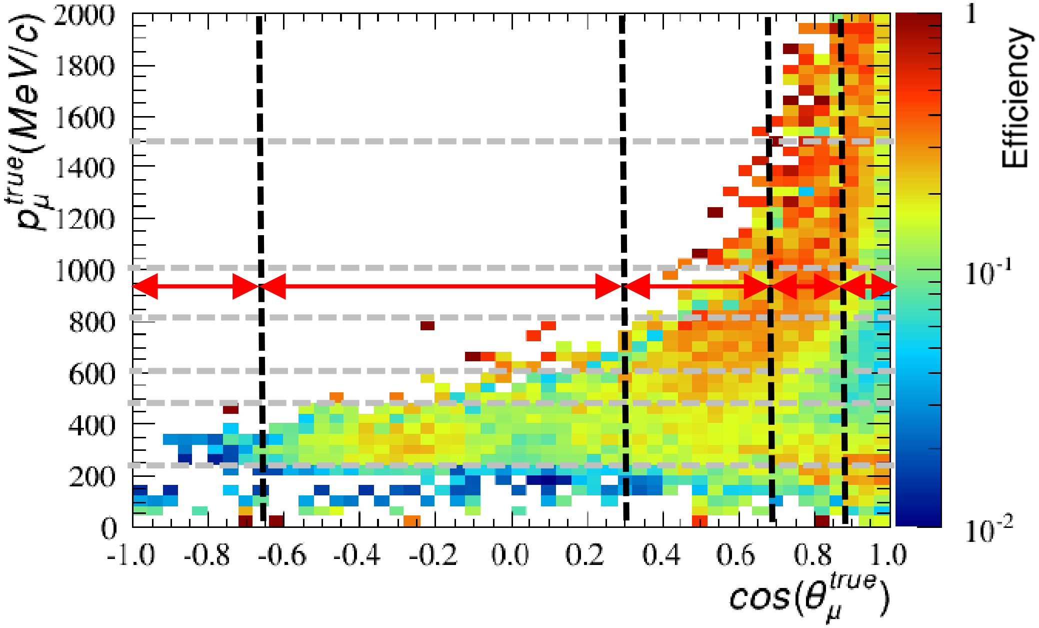
- Consider 2D efficiency
- Bin just fine enough in $\cos(\theta_\mu)$ such that ϵ~flat
- Extract 2D cross-section
- Marginalise over $\cos(\theta_\mu)$ bins



Back to the toy example

Want to measure single-differential p_μ

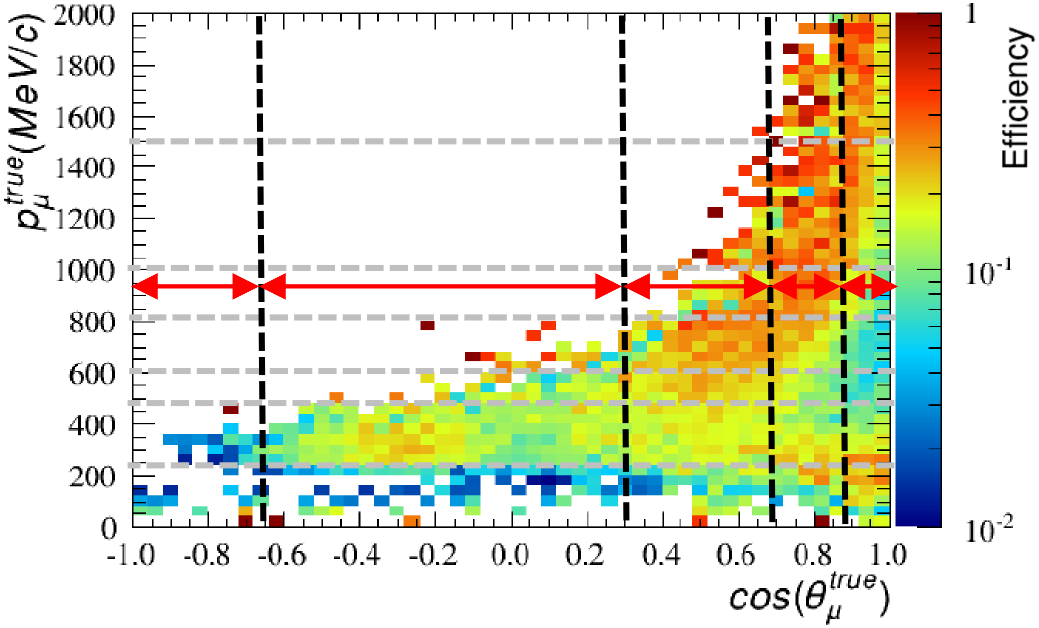
- Consider 2D efficiency
- Bin just fine enough in $\cos(\theta_\mu)$ such that ϵ~flat
- Extract 2D cross-section
- Marginalise over $\cos(\theta_\mu)$ bins
- Report 1D p_μ cross-section



Back to the toy example

Want to measure single-differential p_μ

- Consider 2D efficiency
- Bin just fine enough in $\cos(\theta_\mu)$ such that ϵ~flat
- Extract 2D cross-section
- Marginalise over $\cos(\theta_\mu)$ bins
- Report 1D p_μ cross-section



Measuring a multi-differential cross section in fine bins is clearly often impractical, but even very coarse binning in the variables to be marginalised can be sufficient to mitigate the worst of the model dependence.

E.g. In the T2K CC0$\pi + Np$ analysis in δp_T, just marginalise over 2 bins in the 4D underlying kinematics ($p_\mu, \cos(\theta_\mu), p_p, \cos(\theta_p)$) to achieve a fairly flat efficiency.

Efficiency correcting examples

Double-differential $CC0\pi$ in $p_\mu, \cos(\theta_\mu)$ (using ND280 TPC)

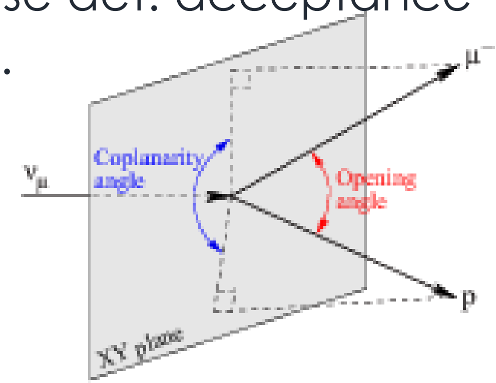
- $p_\mu, \cos(\theta_\mu)$ well characterise detector acceptance
- No need for extra binning

Single-differential $CC0\pi$ in $Q^2_{QE,\mu}$ (using ND280 TPC)

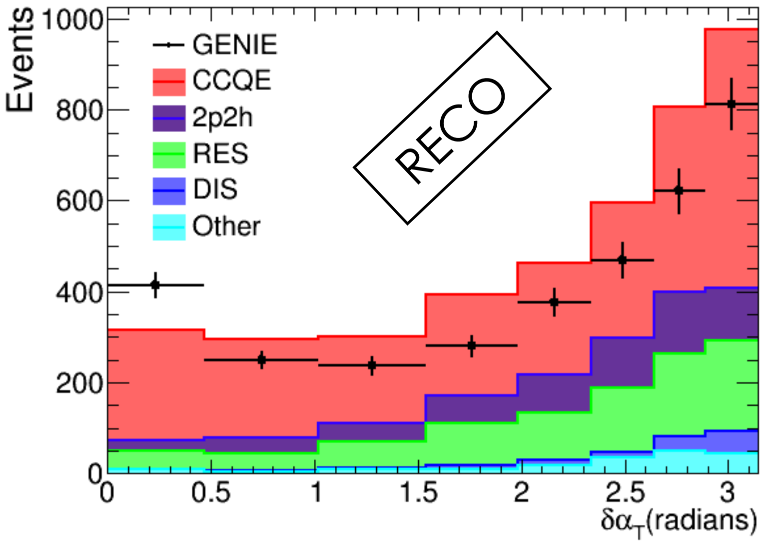
- $p_\mu, \cos(\theta_\mu)$ well characterise detector acceptance
- Measure triple-differential cross section in $Q^2_{QE,\mu}, p_\mu, \cos(\theta_\mu)$
- Marginalise over $p_\mu, \cos(\theta_\mu)$ to report $Q^2_{QE,\mu}$

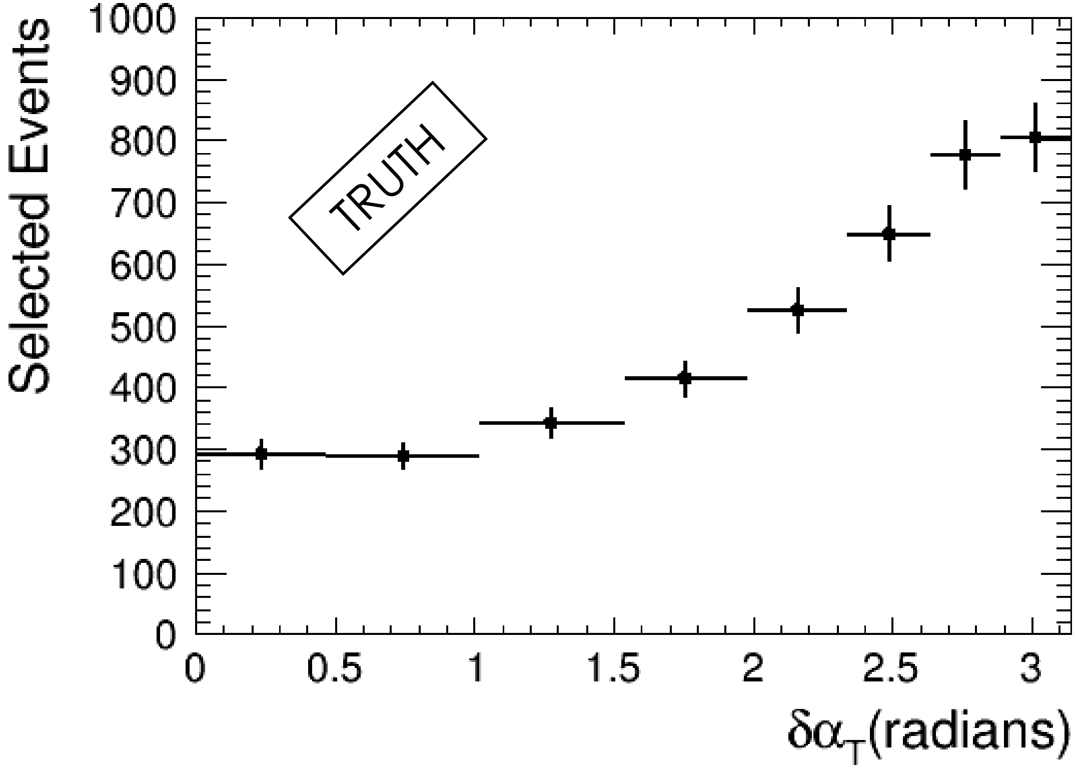
Single-differential $CC0\pi + Np$ in coplanarity angle, ϕ (using ND280 TPC)

- $p_\mu, \cos(\theta_\mu), p_p, \cos(\theta_p)$ (and $\cos(\theta_{\mu p})$) well characterise det. acceptance
- Measure quin(hex)tuple-differential cross section …
- Marginalise over all but ϕ

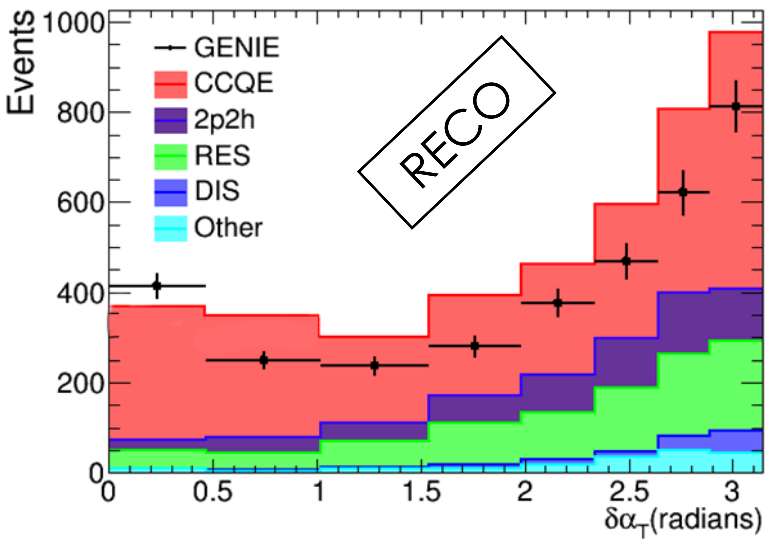


How does it work?

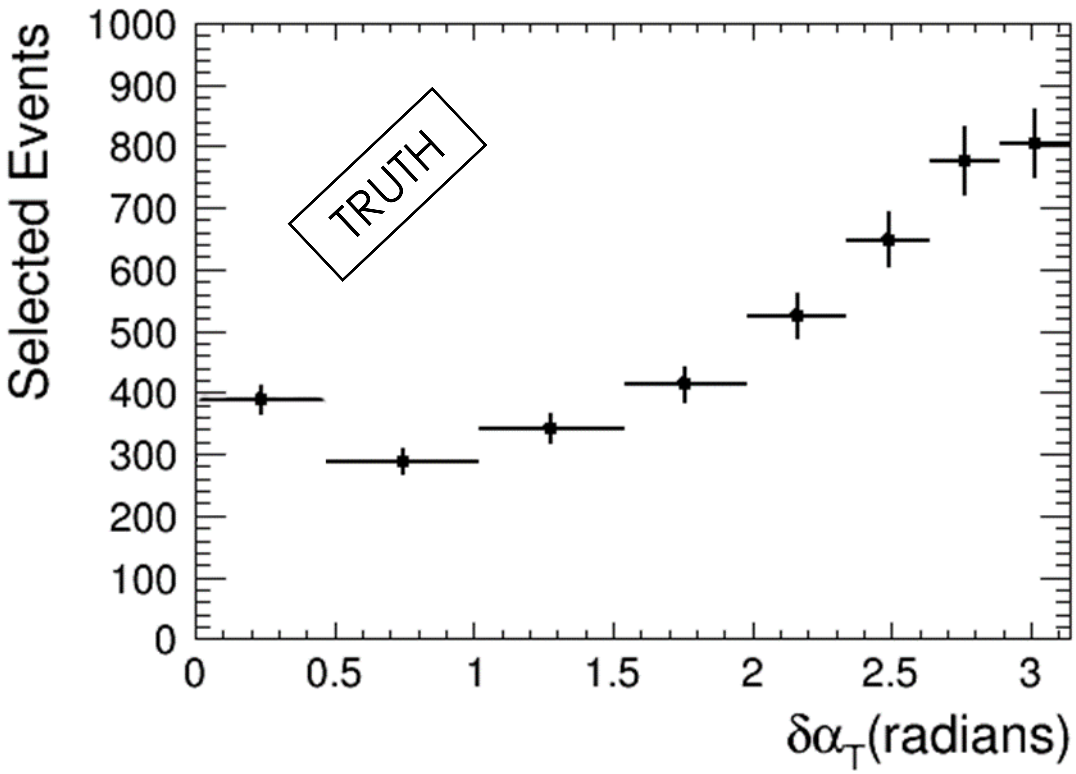




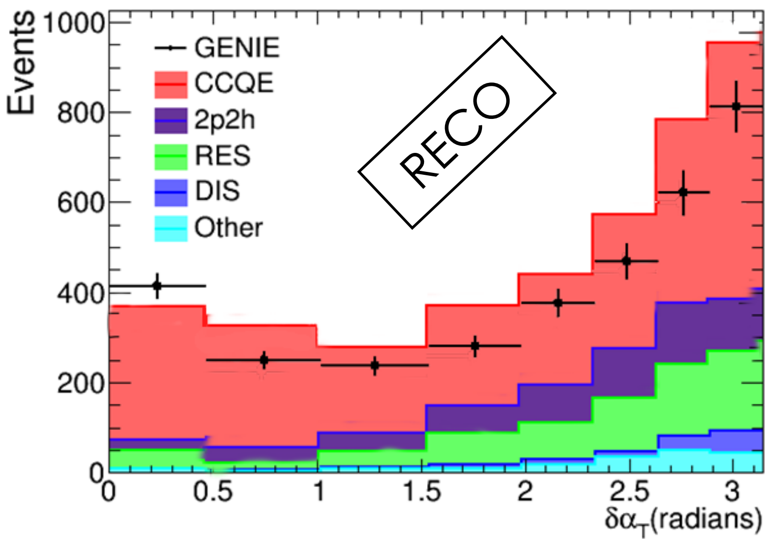
Unsmearing



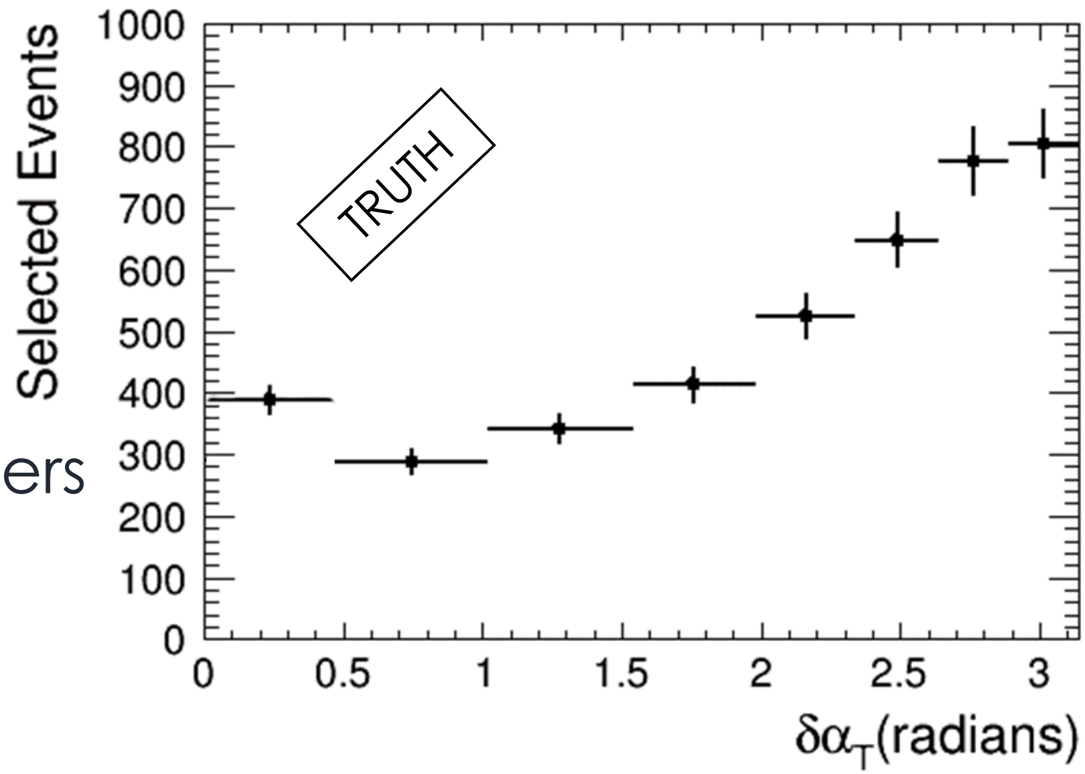
- Scale template weights



Unsmearing

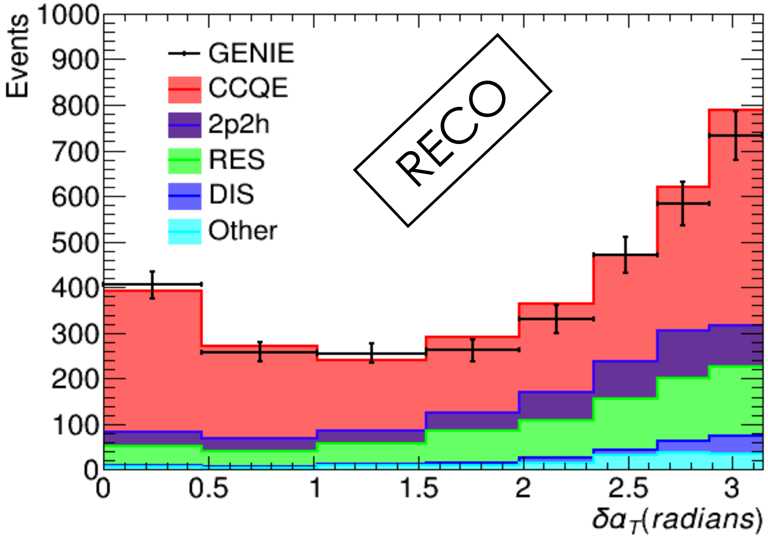


- Scale background systematics
- These should ideally be constrainable by control regions

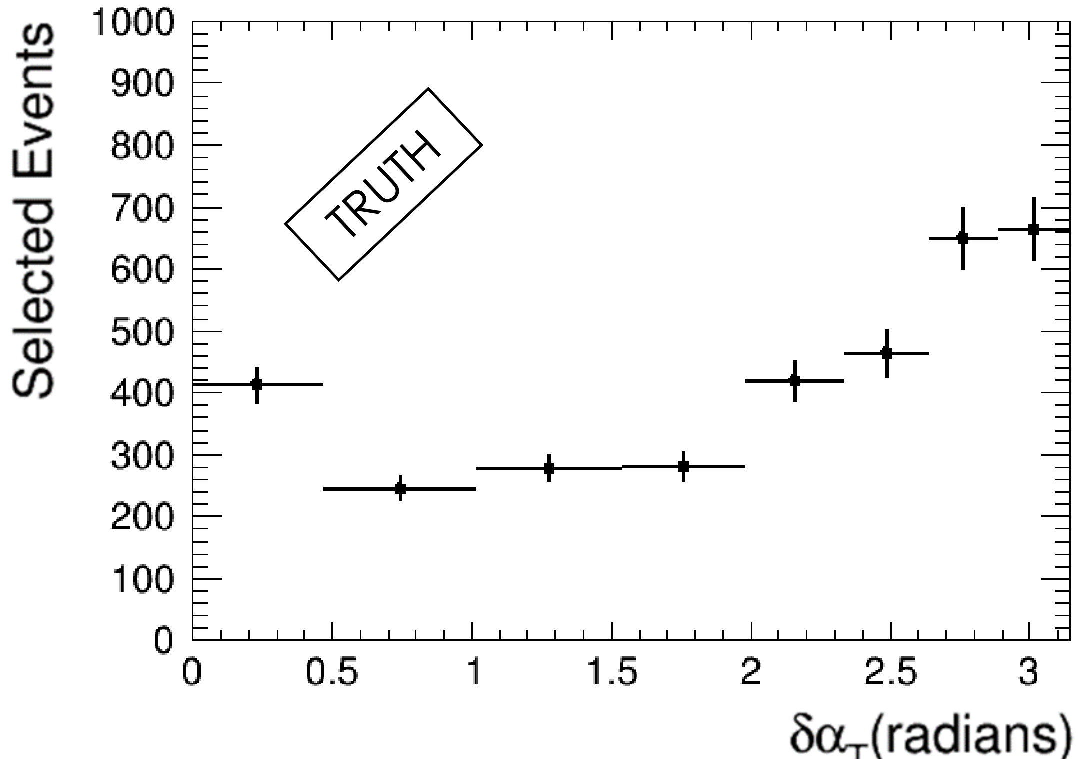


- Overall can alter:
  - Template weights
  - BG Model parameters
  - Flux
  - Detector response

Unsmearing



- Keep iterating to minimize the $-2(\log(L)) \approx \chi^2$



- Maximise likelihood / minimise $-2\ln(L) \approx \chi^2$

Fitting components

- The best fit parameters are those that minimise the following likelihood (here $\chi^2 \approx -2\ln(L)$):

$$\chi^2 = \chi^2_{stat(fit\,goodness)} + \chi^2_{syst(penalty)} + \chi^2_{reg}.$$

$$\chi^2_{stat} = \sum_{j}^{recobins} 2(N_j^{MC} - N_j^{obs} + N_j^{obs} ln \frac{N_j^{obs}}{N_j^{MC}})$$

$$\chi^2_{syst} = (\vec{a}^{syst} - \vec{a}^{syst}_{prior})(V^{syst}_{cov})^{-1}(\vec{a}^{syst} - \vec{a}^{syst}_{prior})$$

$$\chi^2_{reg} = p_{reg} \sum_i (c_i - c_{i-1})^2$$

Fitting components

- The best fit parameters are those that minimise the following likelihood (here $\chi^2 \approx -2\ln(L)$):

$$\chi^2 = \chi^2_{stat(fit\,goodness)} + \chi^2_{syst(penalty)} + \chi^2_{reg}.$$

$$\chi^2_{stat} = \sum_{j}^{recobins} 2(N_j^{MC} - N_j^{obs} + N_j^{obs} ln \frac{N_j^{obs}}{N_j^{MC}})$$

Poisson likelihood:
Characterises how well the reco MC matches the data.

Fitting components

- The best fit parameters are those that minimise the following likelihood (here $\chi^2 \approx -2\ln(L)$):

Penalty term:
Penalises fit for moving systematic parameters far from their nominal

$$\chi^2_{syst} = (\vec{a}^{syst} - \vec{a}^{syst}_{prior})(V^{syst}_{cov})^{-1}(\vec{a}^{syst} - \vec{a}^{syst}_{prior})$$

$$\chi^2_{reg} = p_{reg}\sum_i (c_i - c_{i-1})^2$$

Fitting components

- The best fit parameters are those that minimise the following likelihood (here $\chi^2 \approx -2\ln(L)$):

$$\chi^2 = \chi^2_{stat(fit\,goodness)} + \chi^2_{syst(penalty)} + \chi^2_{reg}.$$

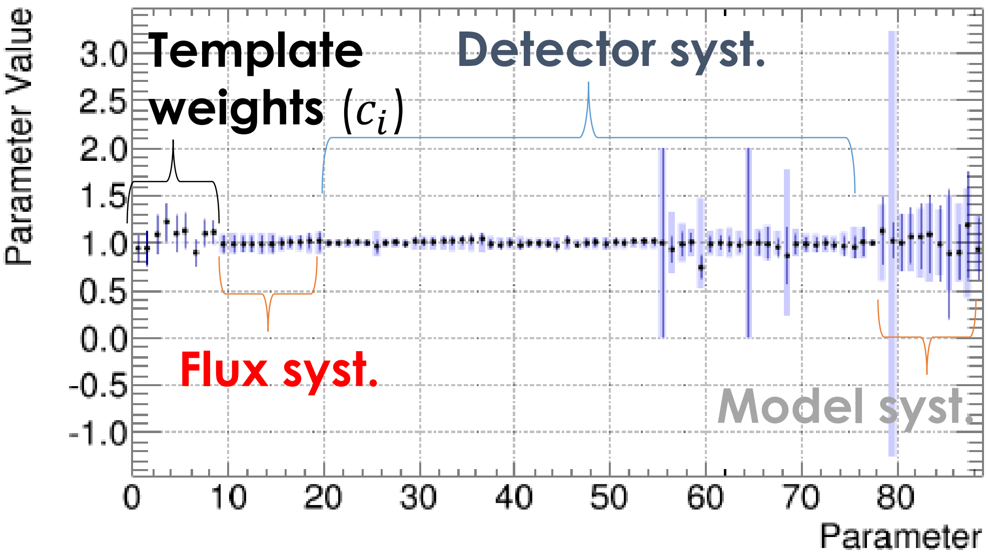
Regularisation term:
Penalises "spiky" truth spectra

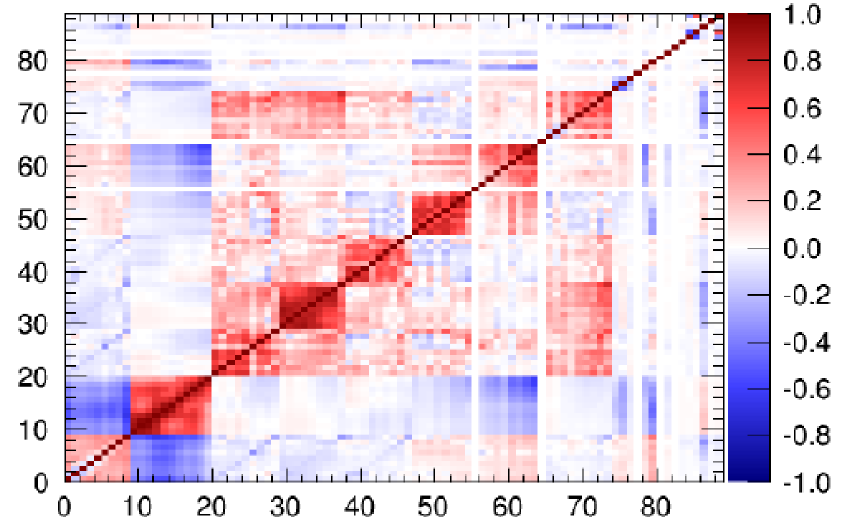
$$\chi^2_{reg} = p_{reg}\sum_i (c_i - c_{i-1})^2$$

Fitting output

** Can't just take c_i since some model syst. parameters also effect signal*

- The fit returns a set of post-fit parameters and a covariance matrix
- Covariance matrix built considering curvature of likelihood surface close to best fit point and assuming Gaussian likelihood.





- Best fit unsmeared selected signal events built from re-weighting input simulation with post-fit parameters*

Post-unfolding uncertainties

- Now we have a cross section, we need uncertainties (normally within a covariance matrix)
- There's several different ways of getting to these, will discuss the two most commonly used at ND280

Option 1 – "Fluctuated input"

- Systematically fluctuate input MC and statistically fluctuate data
- Unfold data with fluctuated MC as the input
- Repeat many times – spread of results gives uncertainties

- Frequentist-like uncertainty
- Prone to under-coverage when using regularisation (I think?)
- When used with the likelihood fitting: Assumes the pre-fit uncertainties are valid for systematic fluctuations of input MC
  - E.g. for a systematic toy in which the flux is moved 3σ from it's nominal, the prior uncertainty is still assumed to be the nominal uncertainty?

Option 2 – "Post-fit propagation"

- Fit data with your favourite input MC
- Throw from post-fit cov. matrix to make toy fit result
- Reweight input MC with toy post-fit parameters to get **post-fit result** (subtleties of what exactly this should be in the backups)
- Repeat for many toys – spread of results gives xsec uncertainties
  - Bayesian-like uncertainty (with a very specifically defined credible interval)
  - Assumes post-fit likelihood is well characterised by a multi-variate Gaussian ("Gaussian errors approximation")
    - Resolvable by using an MCMC rather than a likelihood fit

Option 2 – What is the "post-fit result"?

- Could calculate a **differential xsec** for each toy set of best-fit parameters.
- Inherently includes constraints from fit on the flux/model/efficiency uncertainty
  - Small errors (flux normalisation ~ 5%)
- But potential for unrealistic over constraint of parameters
  - E.g. incomplete model parameterisation could lead to sidebands giving strong flux constraints
- Instead use fit result to find the unsmeared distribution of signal events
- Include the pre-fit uncertainties on the subsequent efficiency correction and flux + $N_{targets}$ normalisation.
  - Larger errors – more conservative

Background Removal

$$purity, p = \frac{N_{signal\ selected}}{N_{total\ selected}}$$

$$N_{sig} = pN_{sel}$$

Purity Correction

- Lower stat. errors
- Requires both the signal and background models to calculate p

$$N_{sig} = N_{sel} - N_{BG}$$

Background Subtraction

- Includes stat. error from selection and BG.
- Requires only the background model

- Can use sidebands (control regions) to further constrain a particular background. The most simple implementation:

New prediction for number of BG events in the selection

$$N_{BG,sig}^{rescale} = \frac{N_{SB}^{data}}{N_{SB}^{MC}} N_{BG,sig}^{MC}$$

MC prediction for number of BG events in the selection

Purity Correction vs. BG subtraction

Analytical example of bias from purity correction / background subtraction

- MC predicts 60 signal events and 40 background events
- Assume flux is actually 20% higher than simulated
- Consider following cases:

A. Signal is wrong by +/- 50%

B. BG is wrong by +/- 50%

C. Signal and background are both wrong by +/- 50%

Assumes a perfectly pure sideband

Case study	A (signal wrong)		B (background wrong)		C (norm. wrong)	
Deviation in data	$\mathcal{S}+50\%$	$\mathcal{S}-50\%$	$\mathcal{B}+50\%$	$\mathcal{B}-50\%$	$N+50\%$	$N-50\%$
Method #1	Trust initially predicted MC purity (60%):					
Obtained signal events	93.6	50.4	86.4	57.6	108	36
Deviation from truth	−13%	+40%	+20%	−20%	±0%	±0%
Method #2	Trust initially predicted MC background (40 events):					
Obtained signal events	116	44	104	56	140	20
Deviation from truth	+7%	+22%	+44%	−22%	+30%	−44%
Method #3	Rescale MC purity using sideband constraint:					
Corrected purity	55.6%	55.6%	45.5%	71.4%	45.5%	71.4%
Obtained signal events	86.7	46.7	65.5	68.6	81.8	42.9
Deviation from truth	−20%	+30%	−9%	−5%	−24%	−19%
Method #4	Rescale MC background using sideband constraint:					
Obtained signal events	108	36	72	72	108	36
Deviation from truth	±0%	±0%	±0%	±0%	±0%	±0%

- To avoid bias – BG reduction with a SB seems like the best option
- Should consider case by case

Sidebands

The requirements for a sideband are:

- **Must** be mutually exclusive from signal selection
- **Should** contain a high purity of the BG to be constrained
- May be used to constrain more than one BG

- The BG constrained by a SB should be representative of that BG in the signal region
  - A SB should contain BG events in a similar range of kinematic phase space as is found in the signal samples
  - E.g.: should not use a standard CC0π selection to constrain the CCQE background to a DIS event selection (completely different q_0, q_3 phase space)

Sidebaand Implementation

- **Directly rescale the background** $N_{BG,sig}^{rescale} = \frac{N_{SB}^{data}}{N_{SB}^{MC}} N_{BG,sig}^{MC}$
  - Assumes SB perfectly characterises the BG in the signal region
  - No direct use of BG model, but require BG model to verify the above

- **Simultaneous fitting**
  - Fit the BG simultaneously to the signal in a template fit to constrain model parameters
  - Relies on the BG model to extrapolate between the signal and SB regions

- **Simultaneous unfolding**
  - Unfold the signal and the BG using the signal + SB regions
  - Subtract BG