Kaon SIDIS modulations with RGA

Extraction of the cos ϕ and cos 2ϕ moments with multidimensional unfolding

Aron Kripko

University of Connecticut
Justus-Liebig University Giessen

Financial support

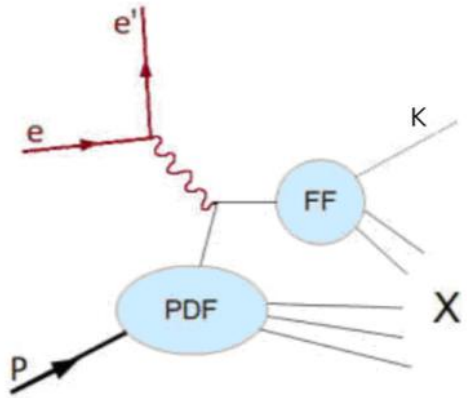
U.S. Department of Energy, Office of Science
Grant No. **DE-FG02-04ER41309**

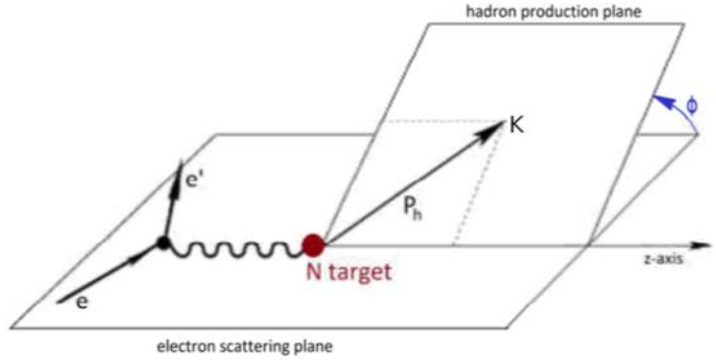
Deutsche Forschungsgemeinschaft (DFG)
Project No. **508107918**

CLAS Collaboration Meeting, Nov. 2026

The measurement: two harmonics with different physics content

- Kaon Sidis: $e^- p^+ \rightarrow e^- K^{\pm} X$
- Moments, averaging beam polarisation, un-polarised target:
- $d\sigma = A_0(1 + A_{UU}^{\cos\phi} \cos\phi + A_{UU}^{\cos 2\phi} \cos 2\phi)$





The measurement: two harmonics with different physics content

- Structure functions - moments - convolution of FFs and TMDs:
- $F_{UU} = \frac{Q^2(1-\epsilon)A_0}{\pi\alpha^2(1+\frac{\gamma^2}{2x_B})}$
- $\frac{F_{UU}^{\cos\phi}}{F_{UU}} = \frac{A_{UU}^{\cos\phi}}{\sqrt{2\epsilon(1+\epsilon)}} \quad - \quad F_{UU}^{\cos\phi} = \frac{2M}{Q}\zeta\left(-\frac{\hat{h}k_T}{M_h}xhH_1^{\perp} - \frac{\hat{h}p_T}{M}f_1D_1 + \ldots\right)$
- $\frac{F_{UU}^{\cos 2\phi}}{F_{UU}} = \frac{A_{UU}^{\cos 2\phi}}{\epsilon} \quad - \quad F_{UU}^{\cos 2\phi} = \zeta\left(-\frac{2(\hat{h}k_T)(\hat{h}p_T)-k_Tp_T}{MM_h}h_1^{\perp}H_1^{\perp} + \ldots\right)$
- Access to Boer-Mulders - $h_1^{\perp}$ - distribution of transversely polarized quarks in an unpolarized hadron
- Cahn effect - kinematic effect due to the intrinsic momentum of quarks in the nucleon

B

The first harmonic is sensitive to intrinsic transverse momentum and subleading-twist contributions. It is expected to carry strong kinematic dependence.

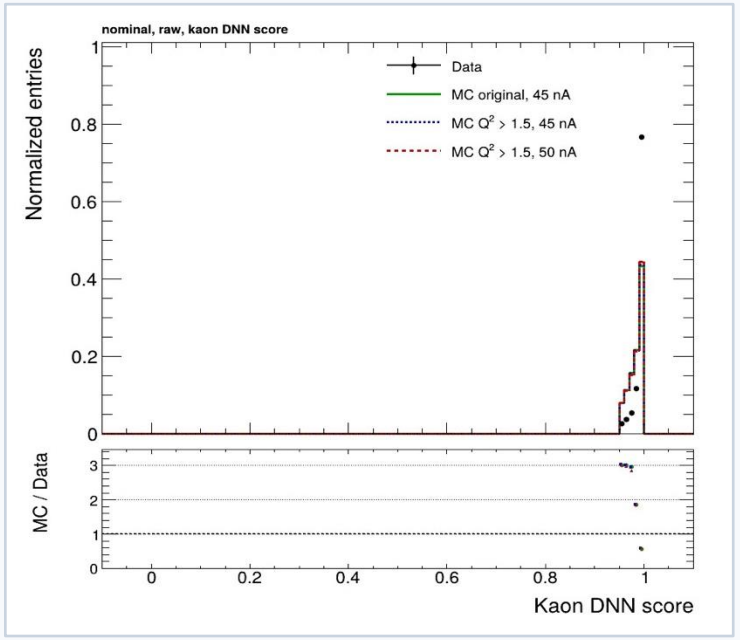
C

The second harmonic contains Boer–Mulders × Collins sensitivity together with kinematic contributions. It is generally smaller and harder to resolve.

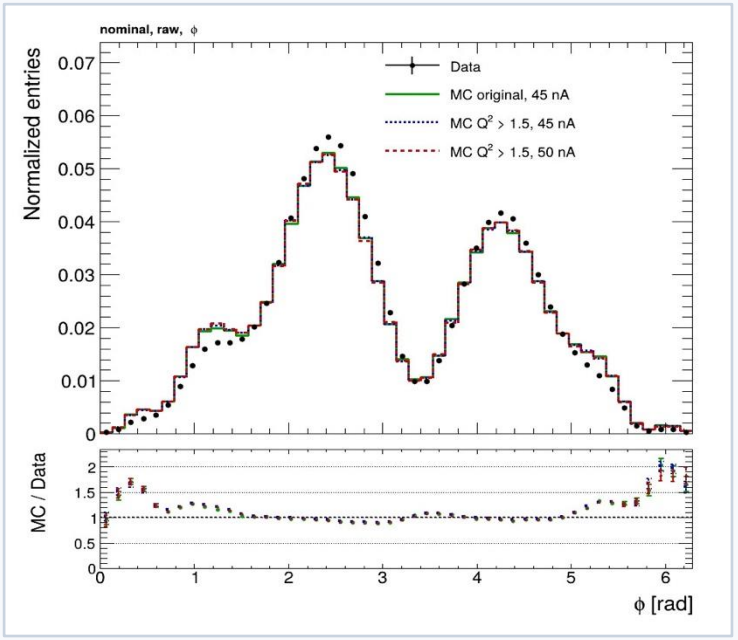
$$N(\phi) = A_0 [\, 1 + B \cos\phi + C \cos(2\phi) \,]$$

Why this is difficult

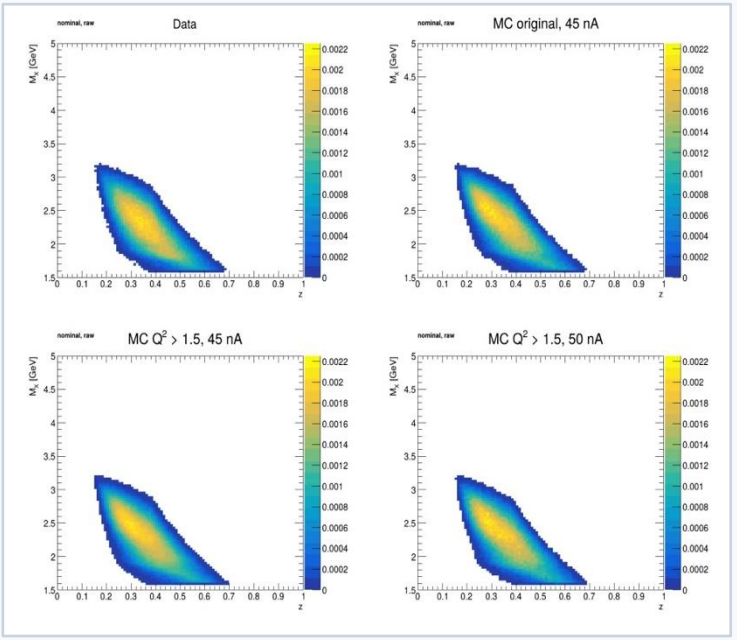
Three effects can create or distort the same azimuthal structure we want to measure.



Kaon DNN score: data and simulation.



Raw ϕ acceptance is strongly nonuniform.



Kinematic correlations drive migration and exclusivity.

1 · Particle ID

Pions dominate the hadron yield and leak into the kaon sample.

2 · Detector response

Acceptance and ϕ migration must be unfolded.

3 · Correlated kinematics

Projection can hide or create apparent trends.

Dataset and accepted phase space

Use IGUANA implementation when available

RG-A data

10.604 GeV electron beam
Liquid-hydrogen target
Fall 2018 inbending Pass 2
Runs 5032–5419
Forward detector
QA: OkForAsymmetry

3 slightly different MC sample

- EventBuilderFilter PIDs
- QADB applied; required flags include TotalOutlier, TerminalOutlier, MarginalOutlier and SectorLoss.
- Misc rejection is conditional on the run allow-list logic. Use only FD.
- Momentum correction or smearing

DIS / SIDIS cuts

$Q^2 > 2\ \mathrm{GeV}^2$
$W > 2\ \mathrm{GeV}$
$y < 0.75$
$xF > 0$
$MX > 1.6\ \mathrm{GeV}$
$0.3 < z < 0.7$

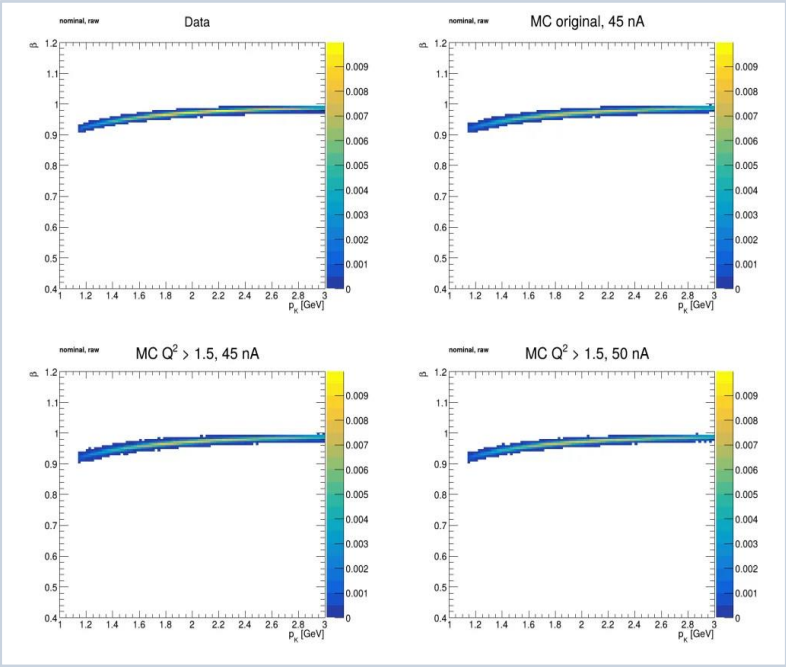
- Select highest-momentum accepted electron.
- HTCC Nphe > 2.
- PCAL energy > 0.07 GeV.
- Sector-dependent total sampling fraction within ±3.5σ.
- Additional sector/momentum-dependent PCAL-versus-ECIN partial-SF rejection.
- Vertex: −8 < vz < 2 cm.
- Standard pass 2 and more stricter versions from Valerii

Kaon selection

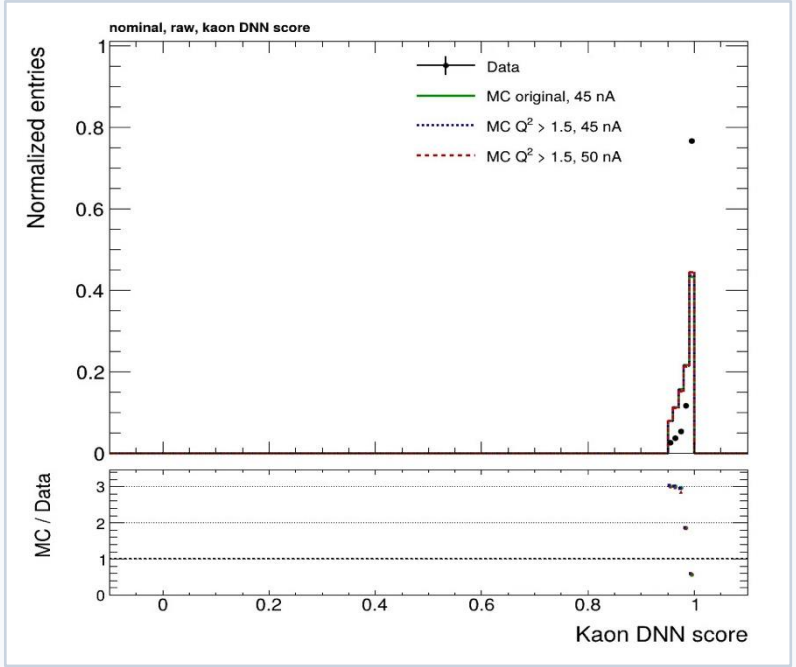
$1.25 < pK < 3\ \mathrm{GeV}$
$5° < \theta K < 35°$
Event-builder PID refinements
$|\chi^2\mathrm{PID}| < 3$
DNN score > 0.95
24 ϕ bins per cell

- Vertex: −20 < vz(e) − vz(h) < 20 cm.
- Standard pass2 fiducial cuts.
- Standard pass 2 momentum correction

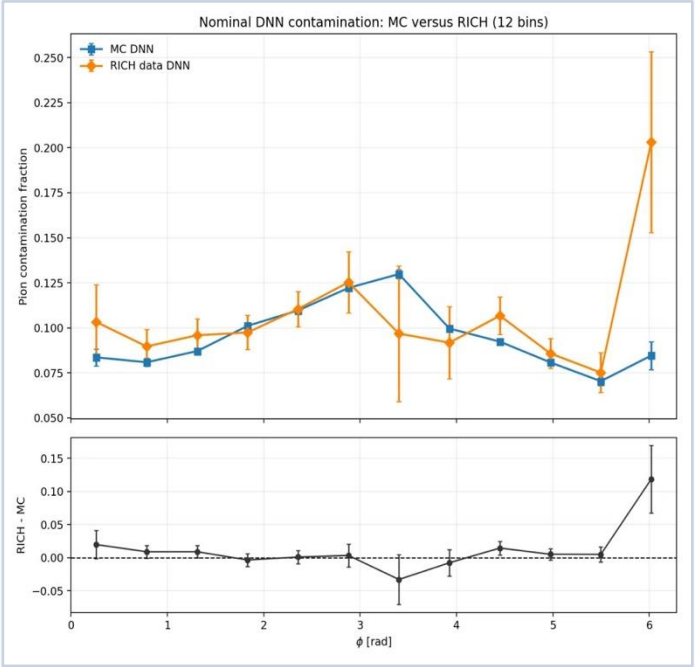
Kaon PID is controlled



β versus pK: the expected kaon band is reproduced.



DNN response: the selected tail is sensitive to data–MC differences.



RICH gives an independent PID label.

Nominal cut

DNN score > 0.95

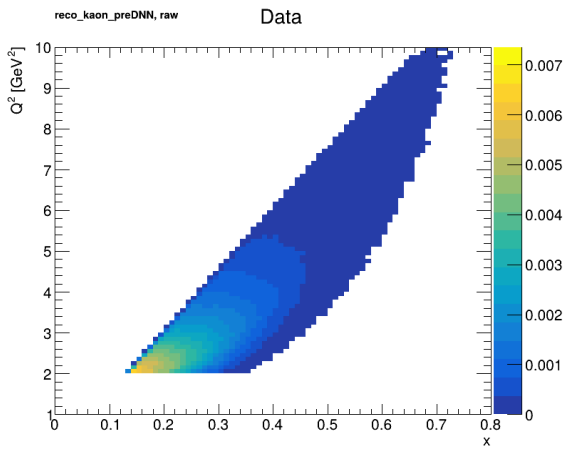
Residual background

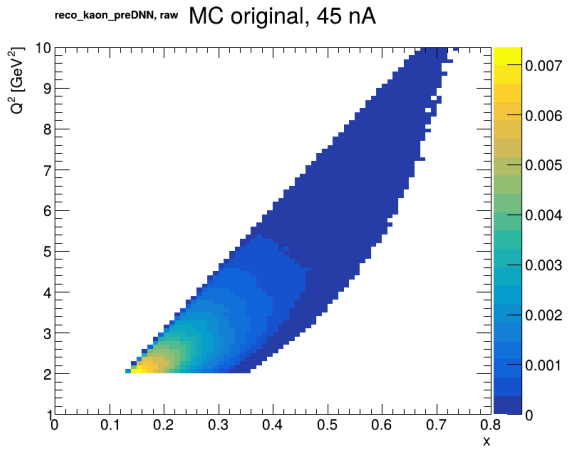
Global MC pion fraction ≈ 10.1%

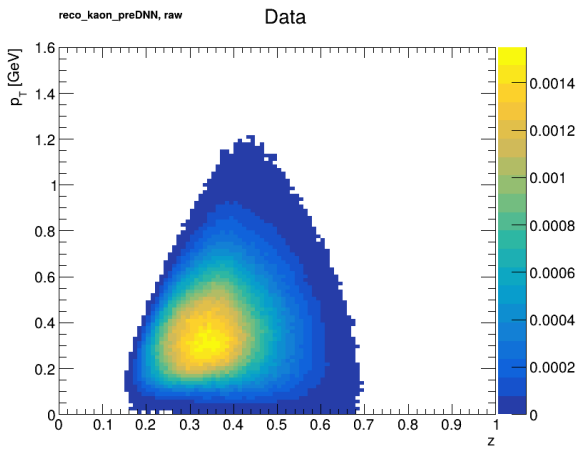
Independent check

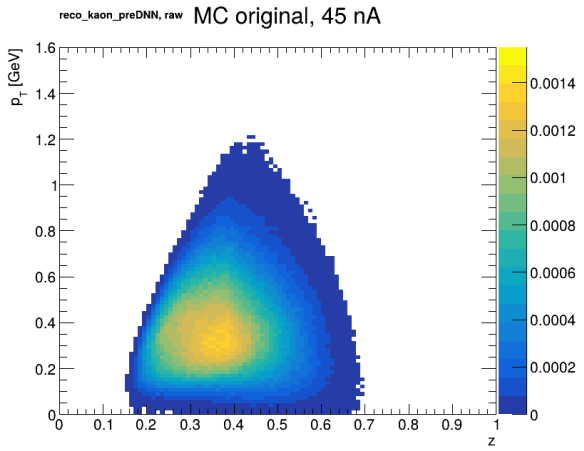
RICH-tagged data agree within available precision.

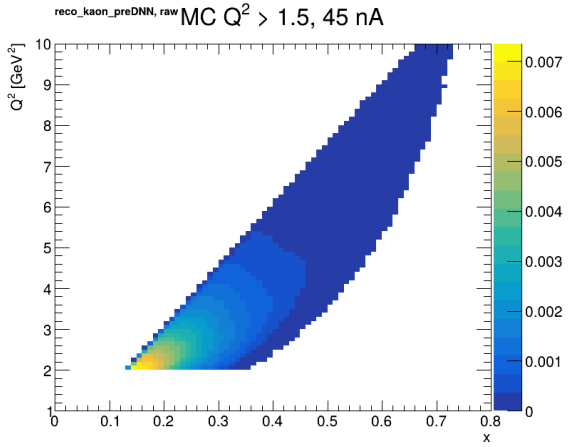
Kinematics

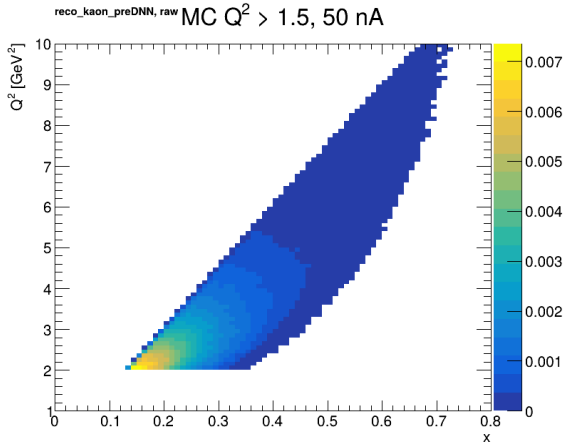

reco_kaon_preDNN, raw
Data
Q^2 [GeV^2]
x

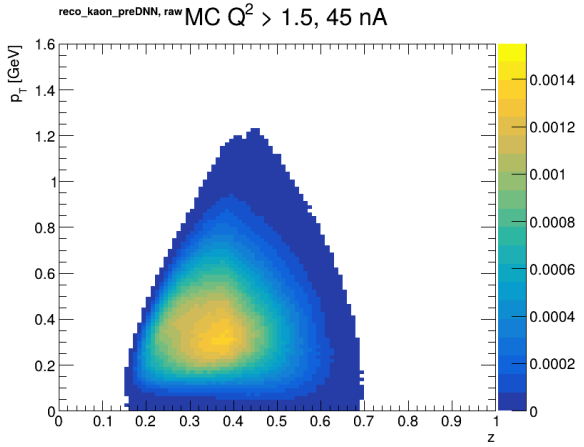

reco_kaon_preDNN, raw
MC original, 45 nA
Q^2 [GeV^2]
x

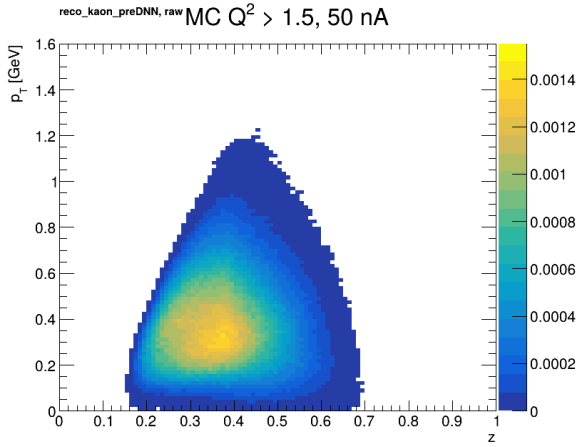

reco_kaon_preDNN, raw
Data
p_T [GeV]
z


reco_kaon_preDNN, raw
MC original, 45 nA
p_T [GeV]
z

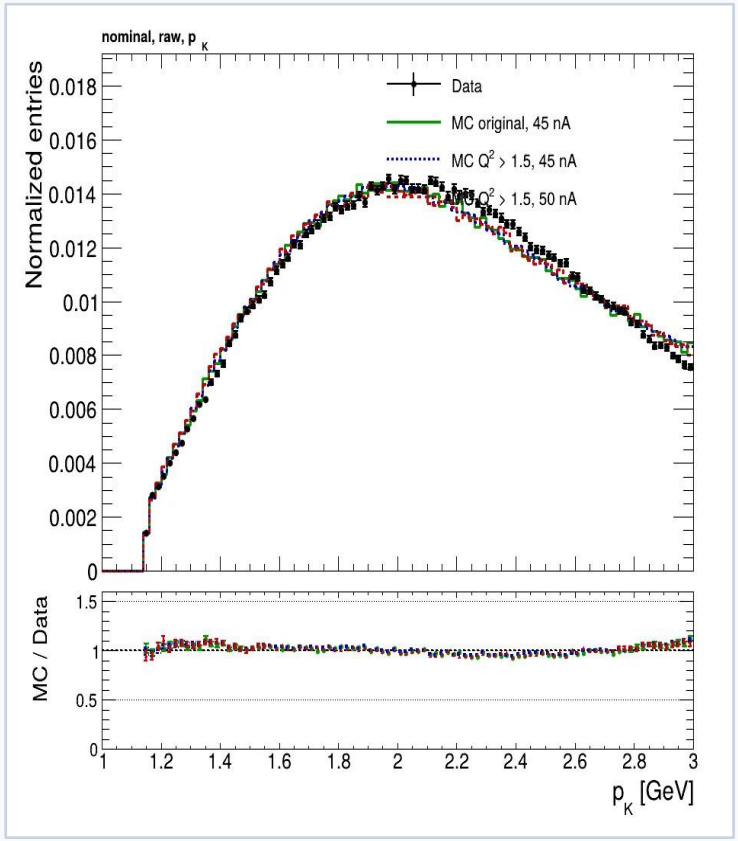

reco_kaon_preDNN, raw
MC Q^2 > 1.5, 45 nA
Q^2 [GeV^2]
x


reco_kaon_preDNN, raw
MC Q^2 > 1.5, 50 nA
Q^2 [GeV^2]
x

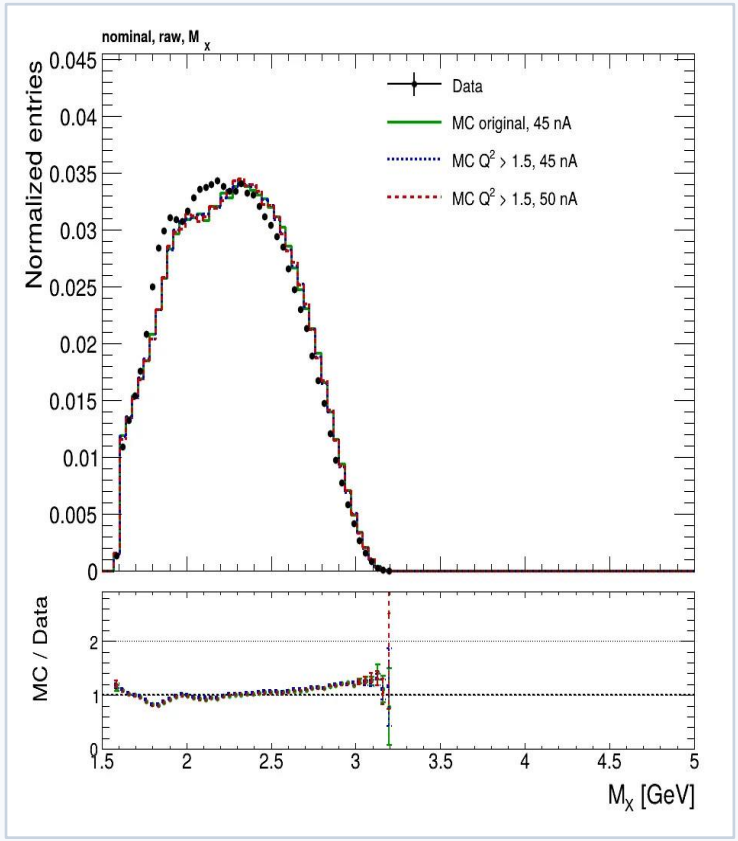

reco_kaon_preDNN, raw
MC Q^2 > 1.5, 45 nA
p_T [GeV]
z


reco_kaon_preDNN, raw
MC Q^2 > 1.5, 50 nA
p_T [GeV]
z

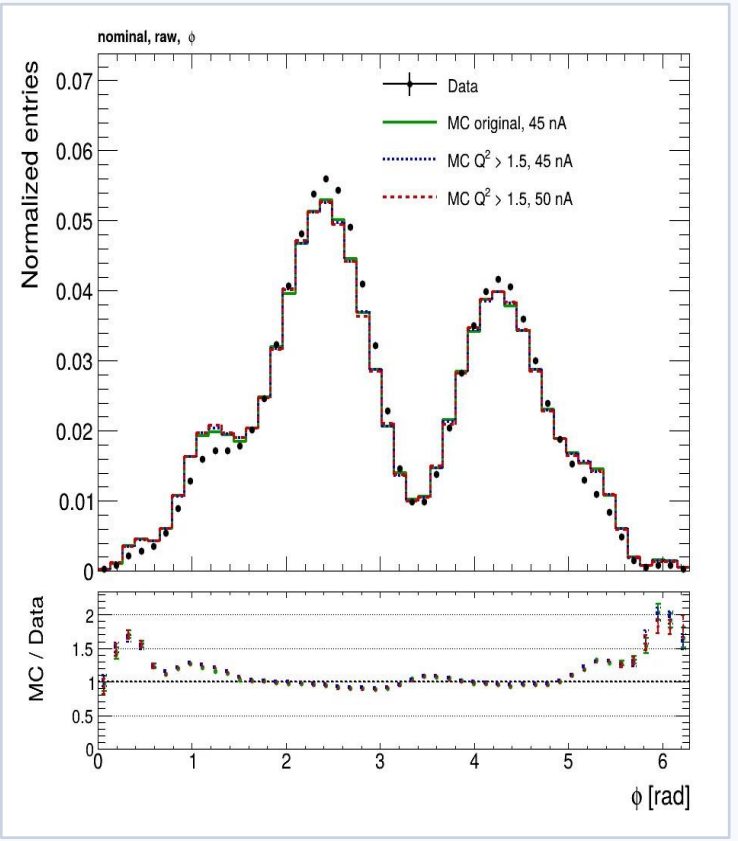
Data–MC agreement



Kaon momentum



Missing mass



Azimuthal angle

DNN trained to reduce pion contamination

Input variables

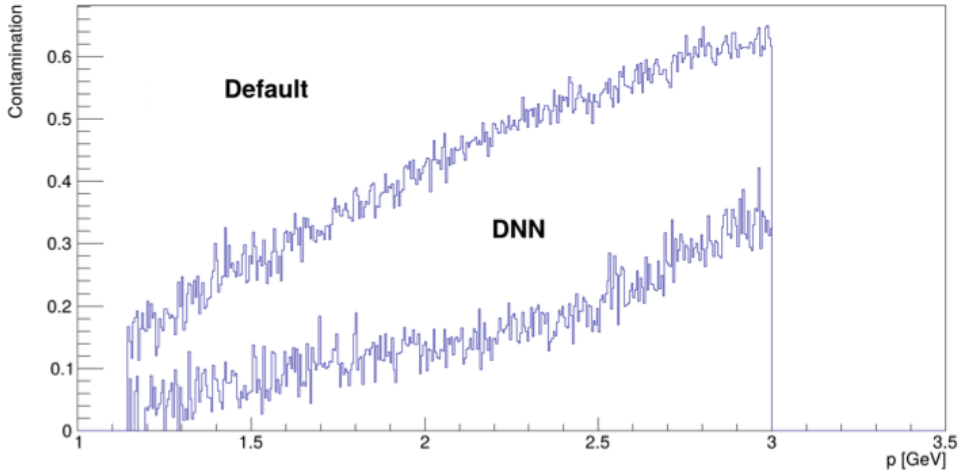
- Momentum and β
- θ and ϕ
- Cal PCAL energy
- Cal ECin energy
- Cal ECout energy
- Cal energy total
- Cal PCAL path/Cal PCAL time
- Cal ECin path/Cal ECin time
- Cal ECout path/Cal ECout time
- χ^2_{PID}
- FTOF energy
- FTOF path/part FTOF time
- FTOF energy layer1
- FTOF path layer1/FTOF time layer1

Available in IGUANA

- ROOT TMVA
- Deep neural network
  - 3 hidden layers with 128 neurons per layer
  - Fully connected
  - tanh activation function, linear for the last neuron
- Training
  - Crossentropy loss function
  - Xavier weight initialisation
  - Optimizer: ADAM
  - Batch size: 30
  - Learning rate: 10^{-5}
  - The input variables are normalised, invalid events skipped
  - The same amount of π-s and K-s are used during the training

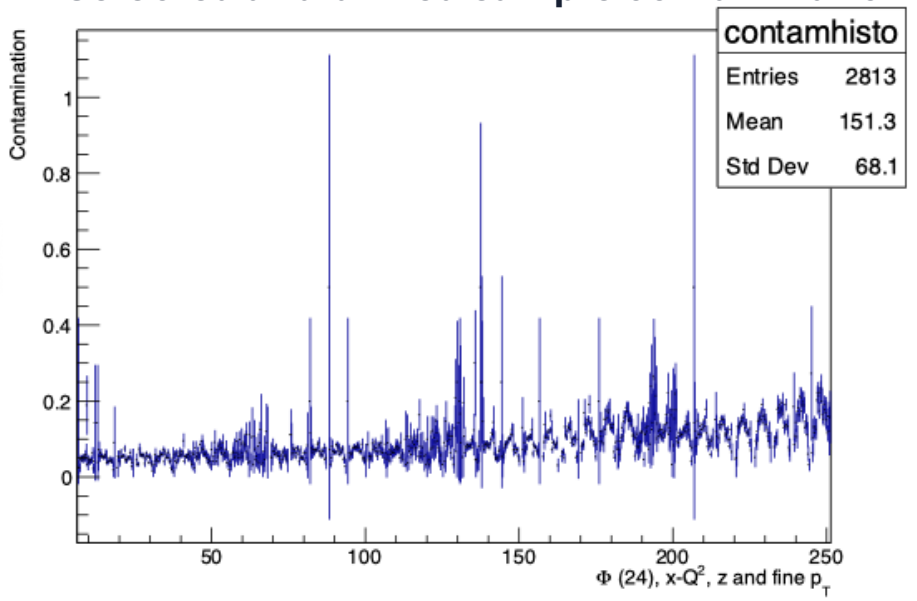
MC – Data smearing and shifting

Variable	Shift	MC sigma	Data sigma
FTOF path layer1/FTOF time layer1	1.6	4.30694e-02	1.37476e-01
FTOF path/FTOF time	1.6	0.0348996	0.134454
Cal PCAL path/Cal PCAL time	1.61	0.0273053	0.142171
FTOF path layer3/FTOF time layer3	1.6	5.18899e-02	1.29125e-01
Cal ECin path/Cal ECin time	1.7	5.42561e-02	1.66168e-01
Cal ECout path/Cal ECout time	1.8	5.25559e-02	1.58020e-01
CC HTCC path/CC HTCC time	0.6	3.56741e-02	9.77466e-02

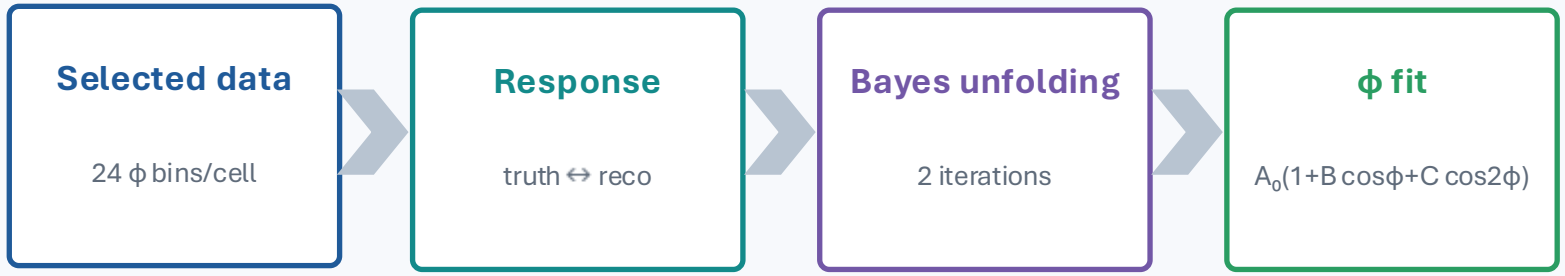


All Kaon contamination

Selected and binned sample contamination



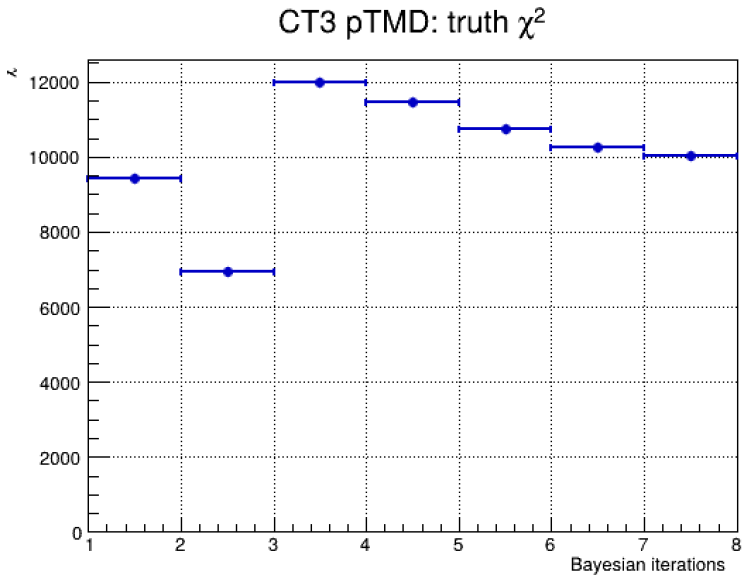
From selected events to the moments – Bayes unfolding with RooUnfold

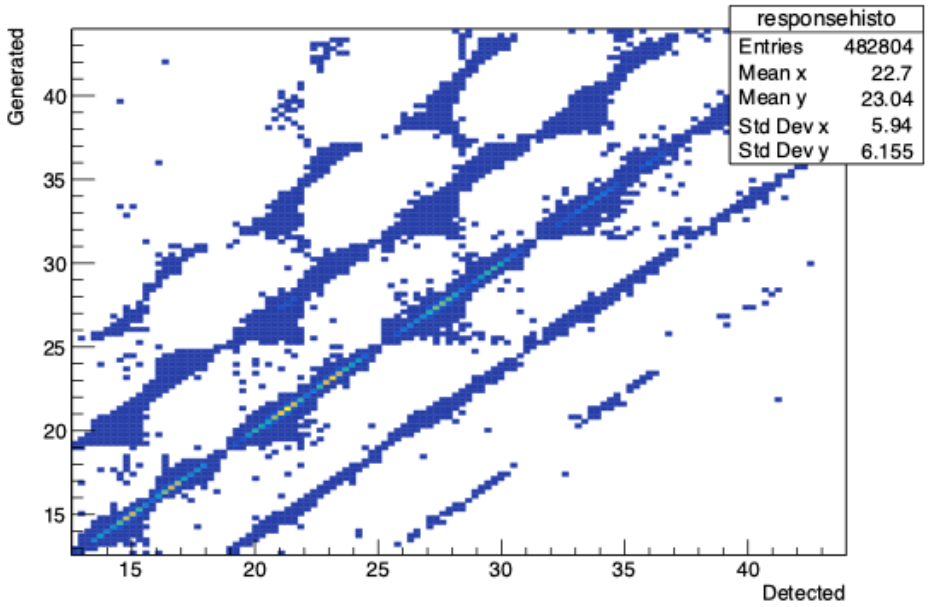


Corrections

Pion misidentification, acceptance, and multidimensional migration.

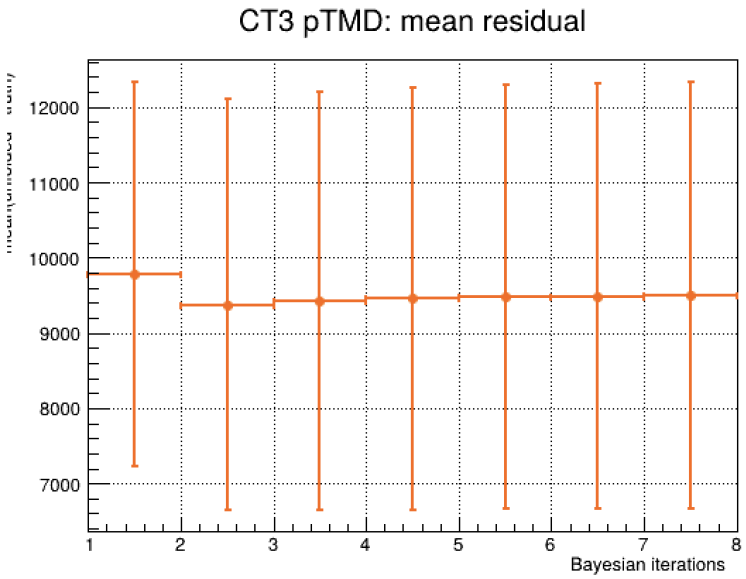
Binning schemes: 2 multidimensional and 4 one dimensional





Fitting:

The unfolded distributions are fit to extract the moments

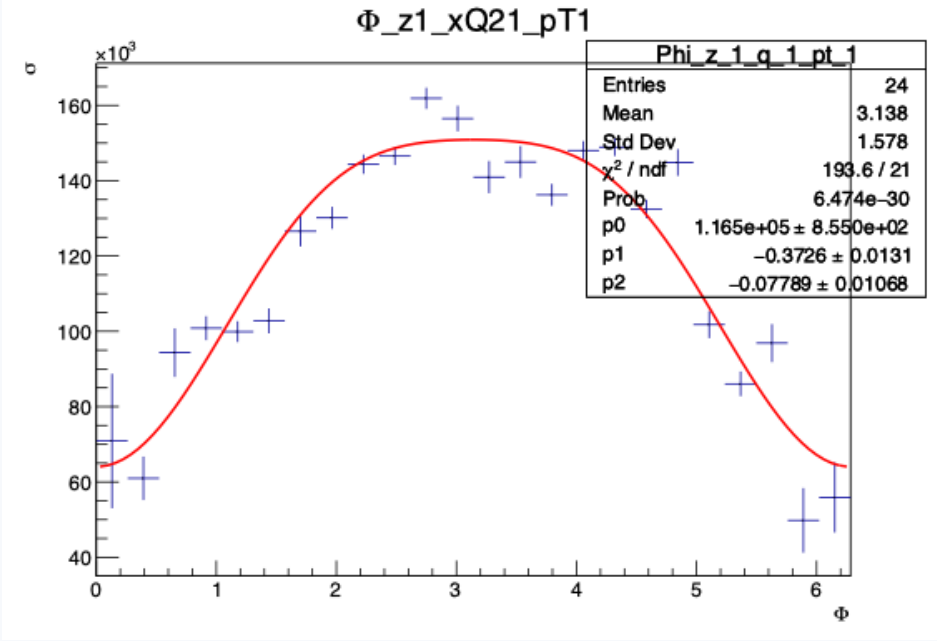


Bayesian unfolding:

2 iterations for all binning schemes

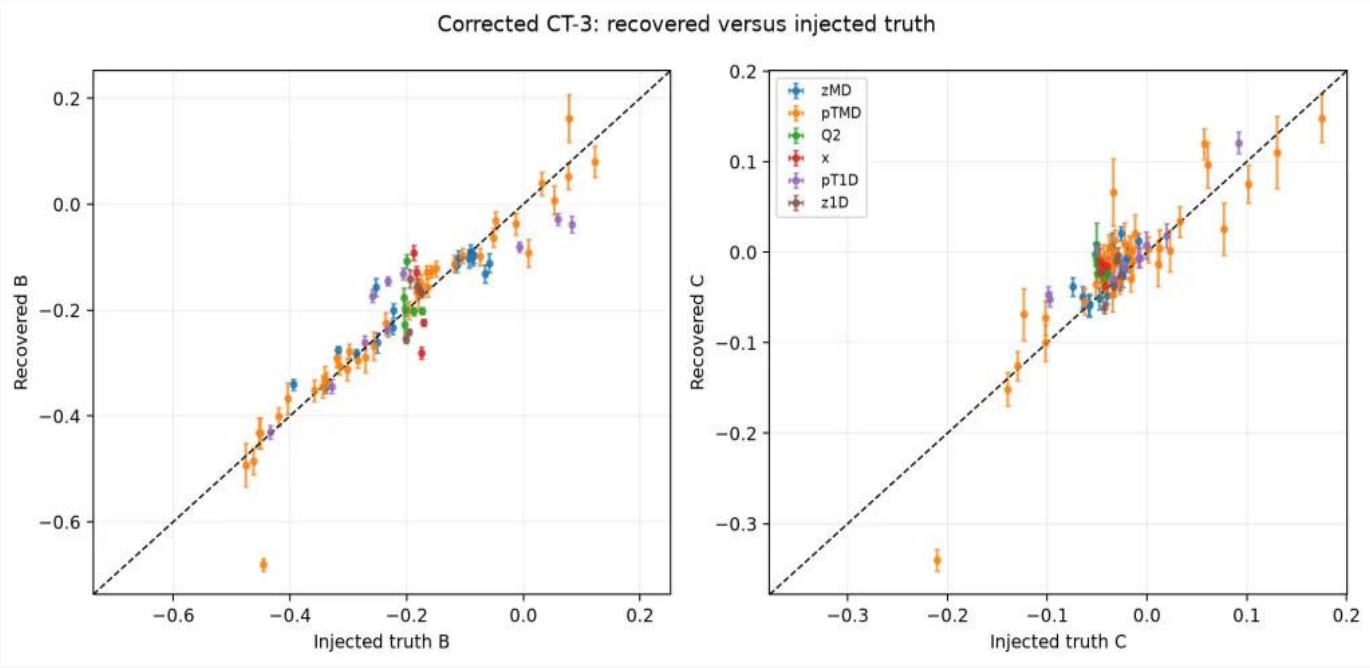
The multiple dimensions are flattened into a single bin number

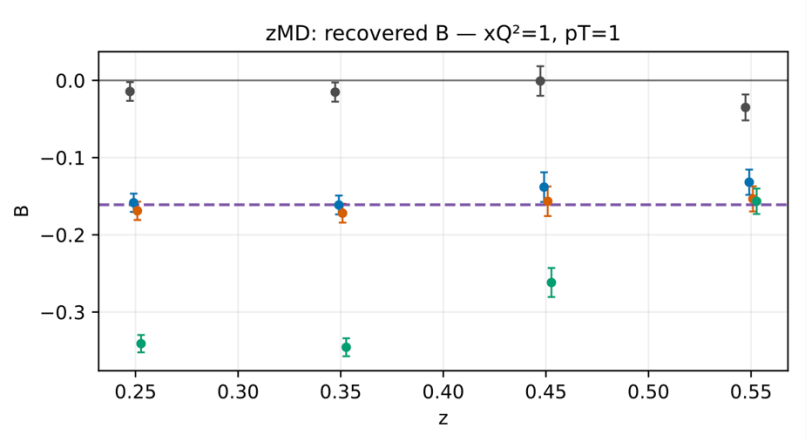
It handles detector acceptance and pion contamination

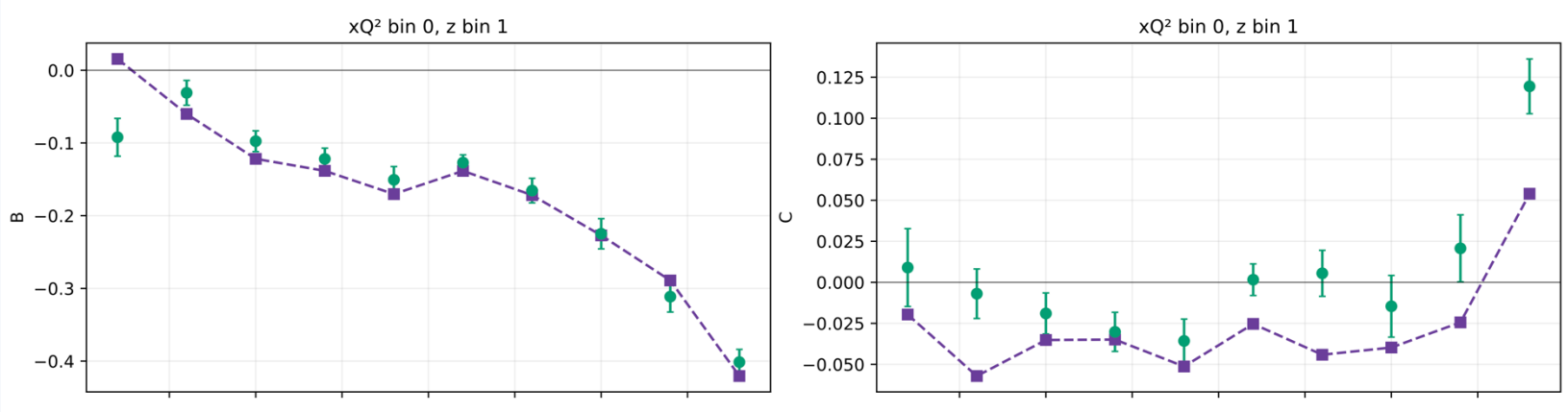


Closure test

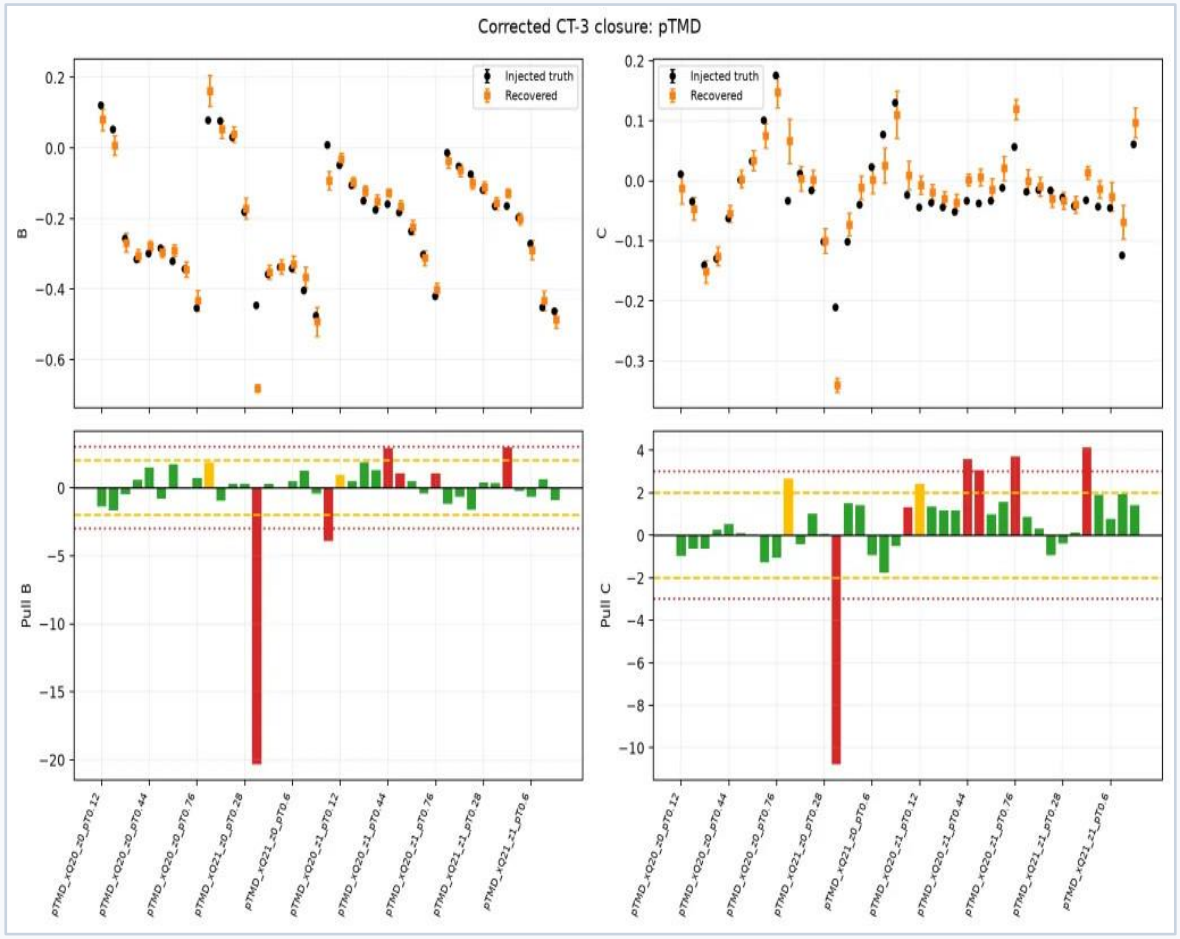
Multiple versions – different injection levels
Repeat analysis but on MC not data
Nominal method – inject measured data for both kaons and pions
The pT multi-D data is used with spline fits as interpolation
Difference is treated as uncertainty

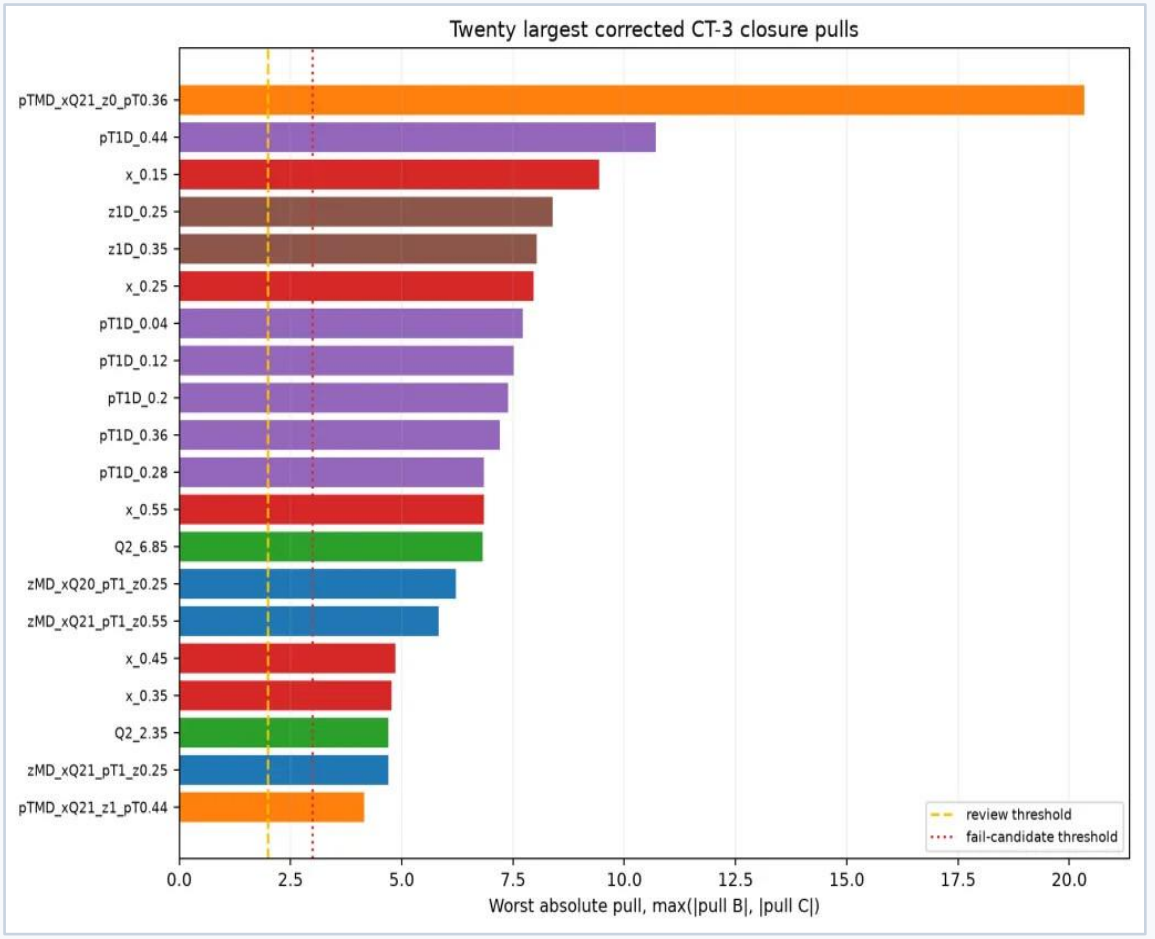






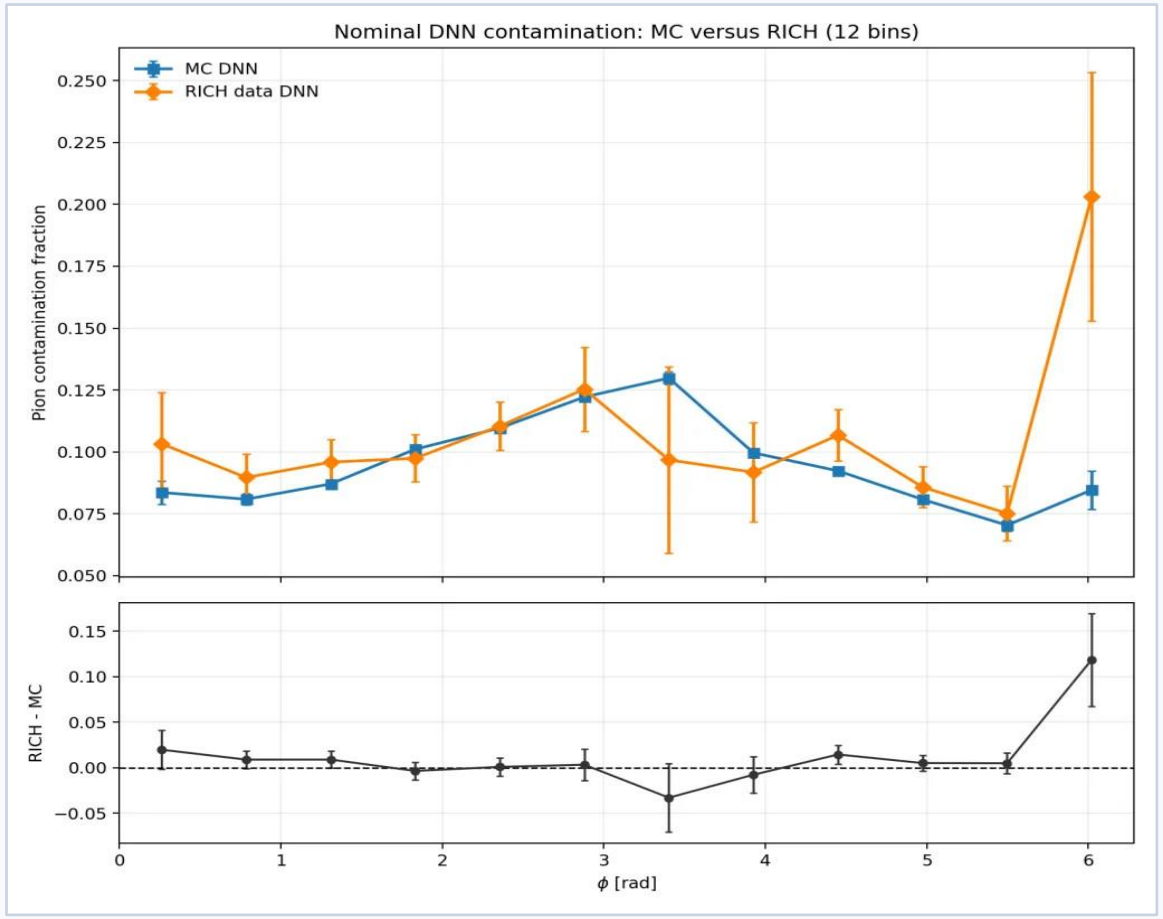
Where CT succeeds — and where it fails



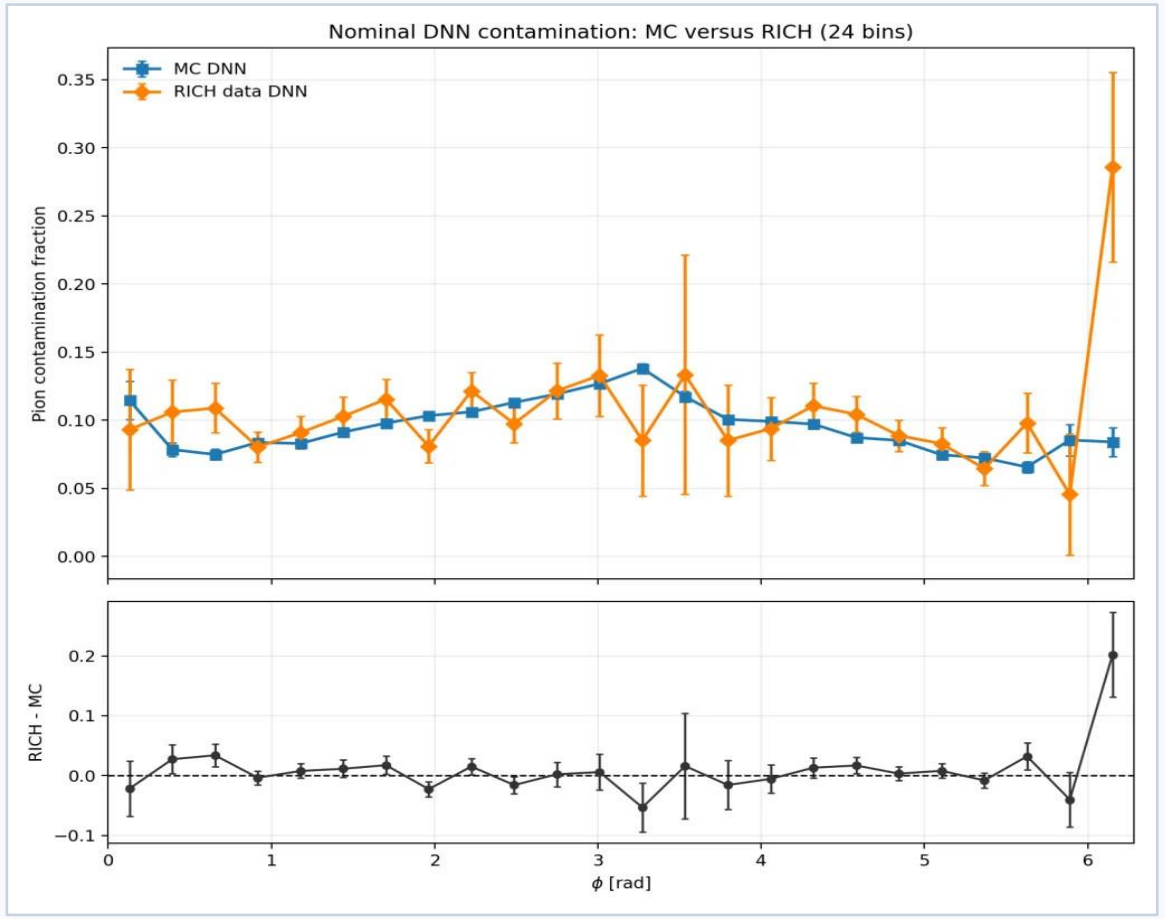


Failures – Low statistics or bin edges – invalid otherwise as well

RICH data to check the MC pion contamination



12-bin primary comparison.



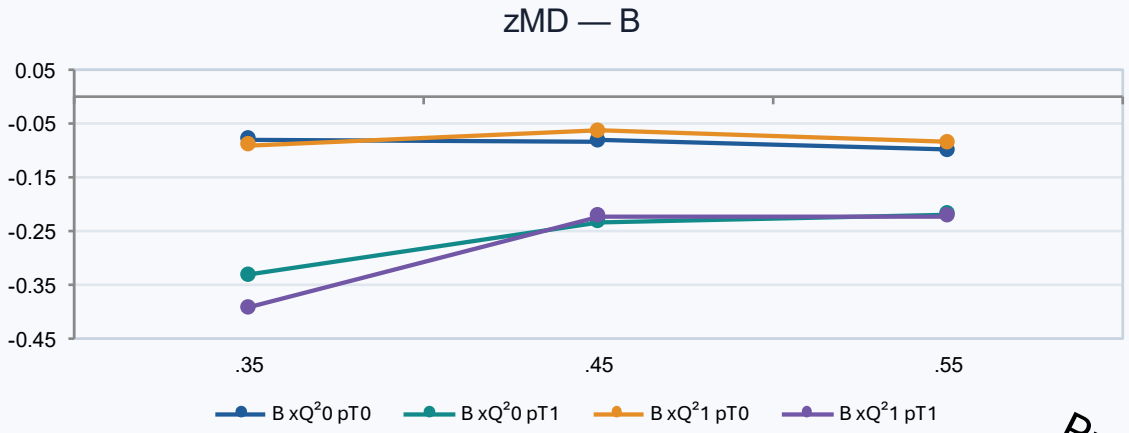
24-bin stability comparison.

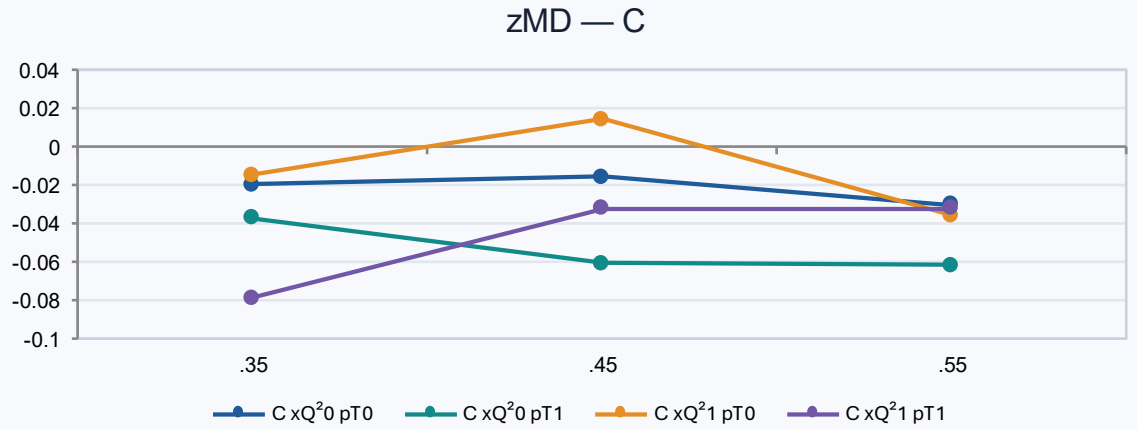
Shape p-value

0.281 (12 bins) – we cannot prove that they disagree

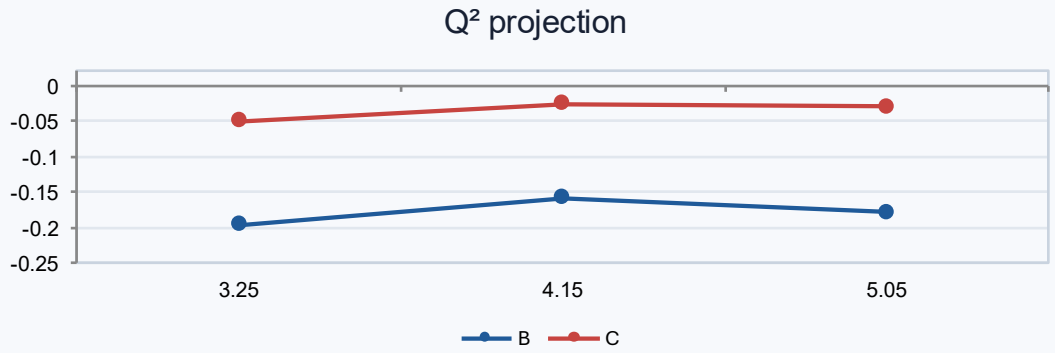
Closure tests - the extracted moments respond linearly to the contamination scale
The RICH – MC difference is propagated as a systematic uncertainty

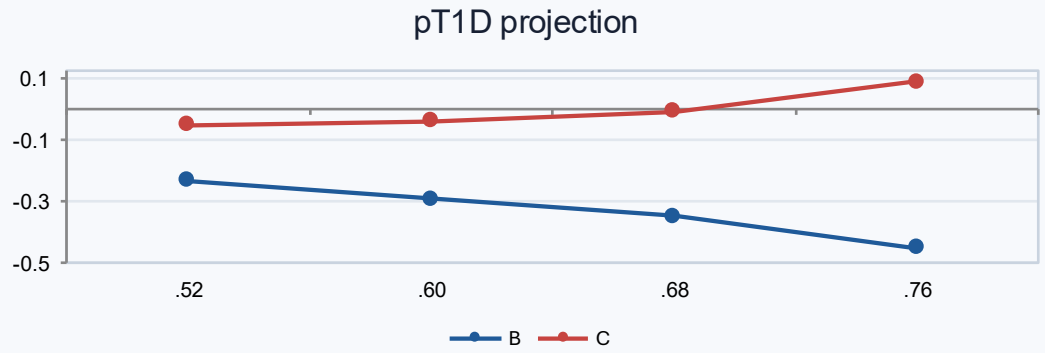
z dependence and one-dimensional projections



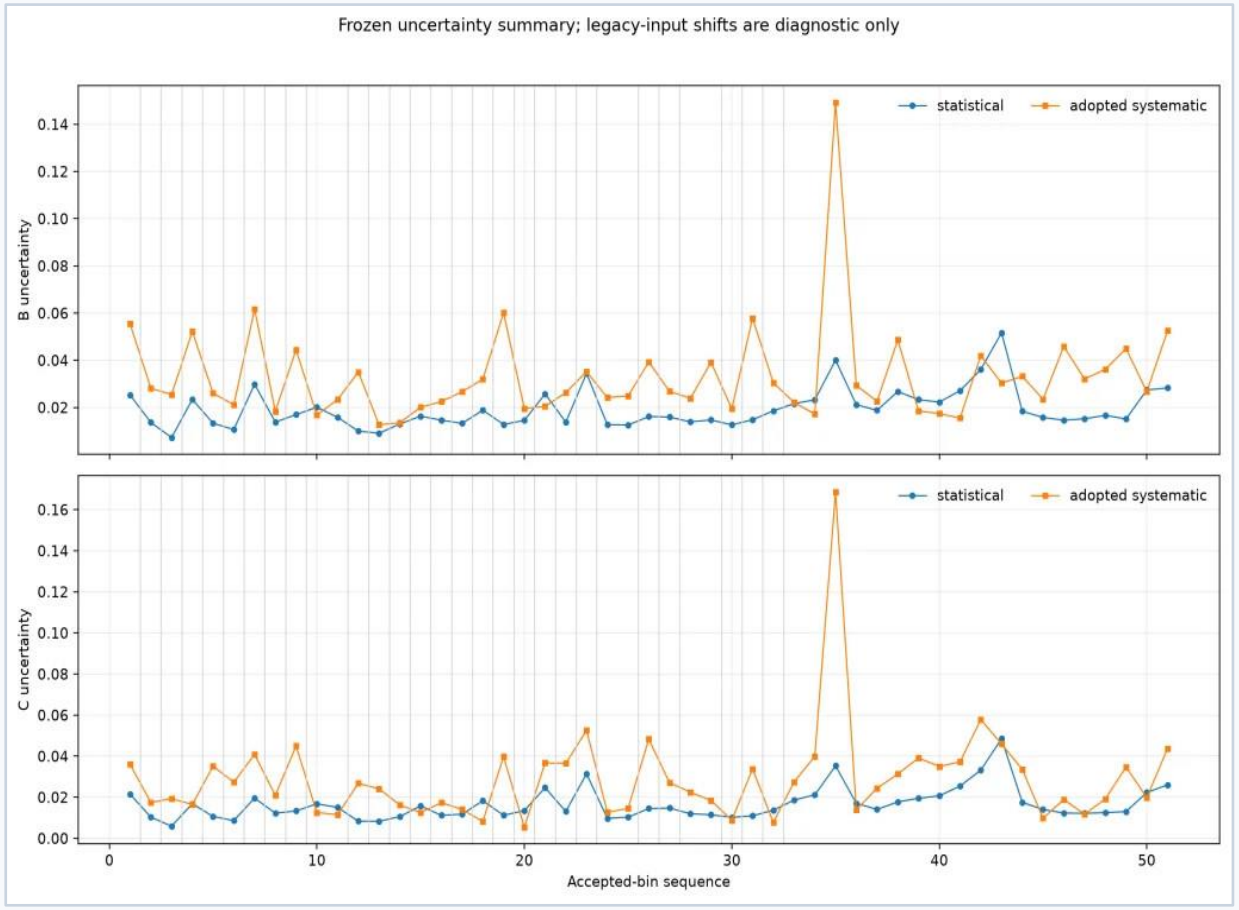


Preliminary

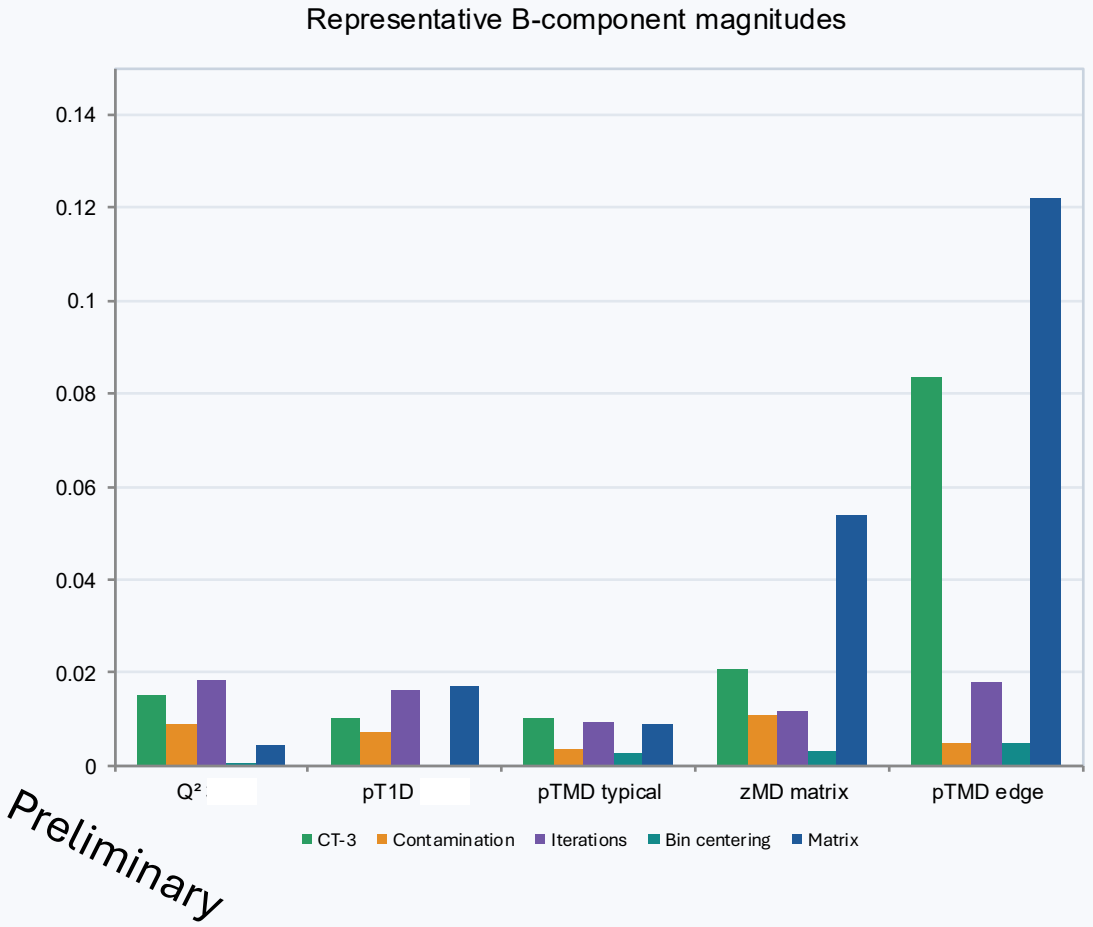




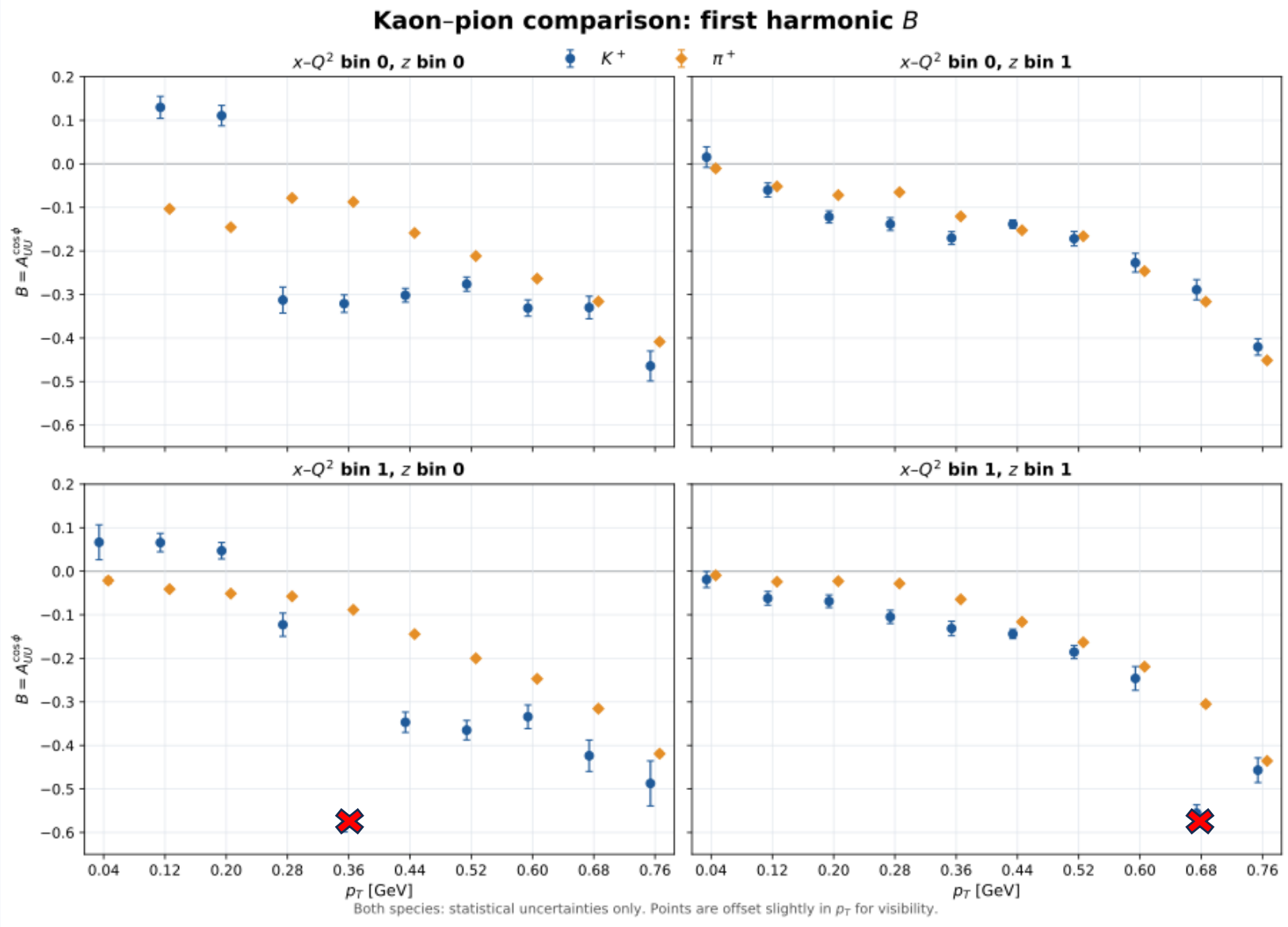
The uncertainty budget is bin dependent – main systematics

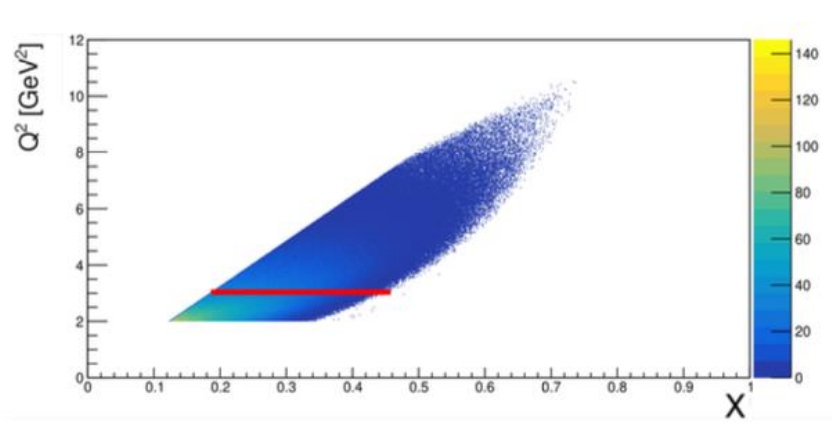


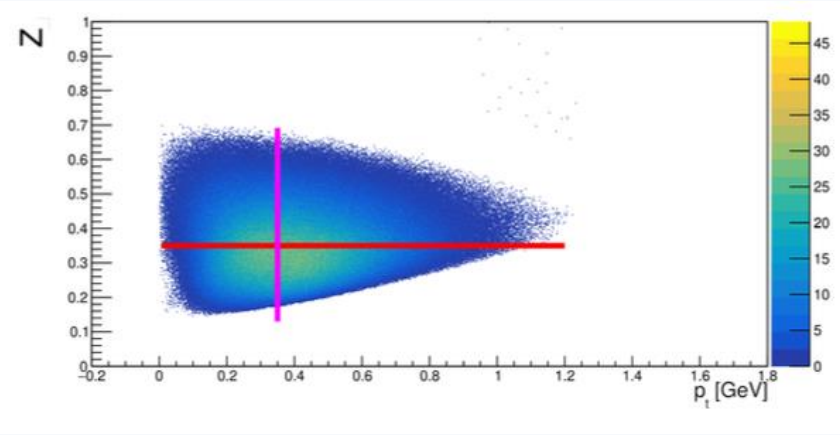
Statistical and adopted systematic uncertainties across the accepted sequence.



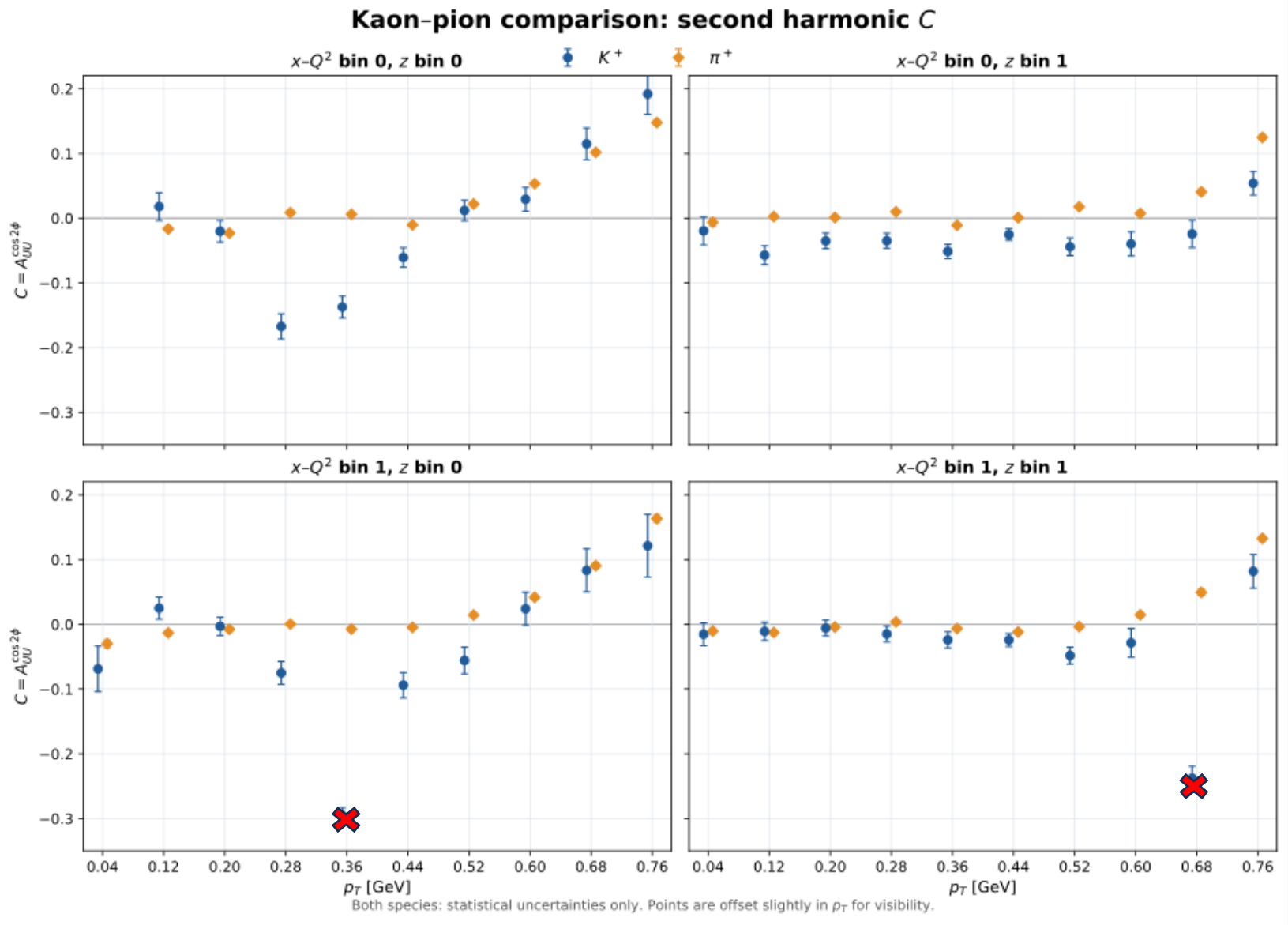
There is no single dominant source: CT-3, iteration and matrix terms trade places from bin to bin.

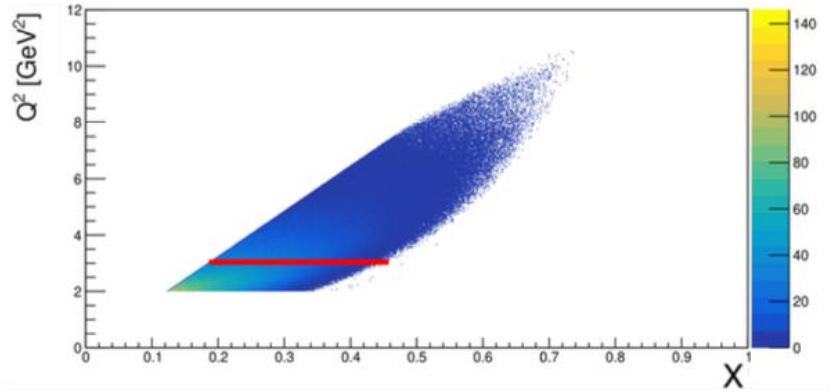


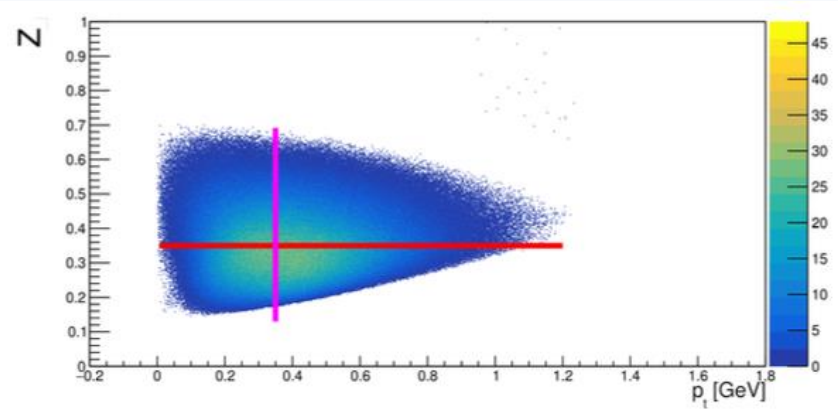




Preliminary







Preliminary

Conclusions and outlook

- Complete data processing rewritten to use iguana
- New ML DNN trained on pass 2
- Full MC sample processed
- Dedicated systematics, closure tests and validation
- Pion contamination effect studied
- Bayesian unfolding stable and reproducible
- Enough statistics for a fully multidimensional binning
- Systematic studies almost finished

- Radiative corrections
- Analysis note
- Cross section – A_0