Early Science Pseudo-Data Production

Analysis Updates and Current Status

Salvatore Franzese

August 25, 2026

In this work

Goal: Generate Early Science pseudo-data for uTMD impact studies.

The Observable: The 4D-differential SIDIS cross section:

$$\frac{d^4\sigma}{dx\, dQ^2\, dz\, dP_{hT}}$$

Output: For each identified hadron ($\pi^{\pm}$, $K^{\pm}$), we produce a table containing the multidimensional bin, the cross section, and its uncertainty (in CSV and YAML formats).

Validation: We also produce analysis plots, including cross sections, multiplicities, and reconstruction efficiencies, using the ECCE proposal (`arXiv:2207.10893`) as the main reference.

Data and Monte Carlo Samples

- Inclusive *ep* events from the **July 2026 ePIC simulation campaign**
- Generated with **PYTHIA 6-eic**, 9 GeV e^- on 275 GeV p (Early Science configuration)
- Each dataset has its production cross section σ_{gen} reported in the **PYTHIA 6 metadata**

Dataset DID
`epic:/RECO/26.07.1/epic_craterlake/SIDIS/pythia6-eic/1.3.0/ep_noradcor/9x275/q2_1to10`
`epic:/RECO/26.07.1/epic_craterlake/SIDIS/pythia6-eic/1.3.0/ep_noradcor/9x275/q2_10to100`
`epic:/RECO/26.07.1/epic_craterlake/SIDIS/pythia6-eic/1.3.0/ep_noradcor/9x275/q2_100to1000`
`epic:/RECO/26.07.1/epic_craterlake/SIDIS/pythia6-eic/1.3.0/ep_noradcor/9x275/q2_1000to100000`

Event and hadron selection

Event selection

- DIS events selected through the scattered-electron kinematics
- Reconstructed electron **truth-matched** to the generated electron
- $|\eta_e| \leq 3.5, \quad 0.01 \leq y \leq 0.95, \quad p_e > 1$ GeV

Hadron selection

- $\pi^\pm$ and $K^\pm$ tracks selected within the DIS sample, **truth-matched**
- $|\eta_h| < 3.5$

PID

- Reconstructed PID required to match generated MC species $\Rightarrow$ no misidentification

For more details, see the previous analysis update.

Cross Section Extraction

Integrated luminosity

For each generated sample, using the PYTHIA generator metadata, we compute:

$$\mathcal{L}_{MC} = \frac{N_{\text{events}}}{\sigma_{\text{gen}}}$$

Differential cross section

$$\frac{d^4\sigma}{dQ^2\,dx_B\,dz\,dP_{hT}} = \frac{N_{\text{rec}}}{\mathcal{L}_{MC} \cdot \varepsilon \cdot \Delta Q^2 \Delta x_B \Delta z \Delta P_{hT}}$$

where $\varepsilon = \frac{N_{\text{rec}}}{N_{\text{gen}}}$ is the reconstruction efficiency in each 4D bin

Kinematic reconstruction

Previous implementation

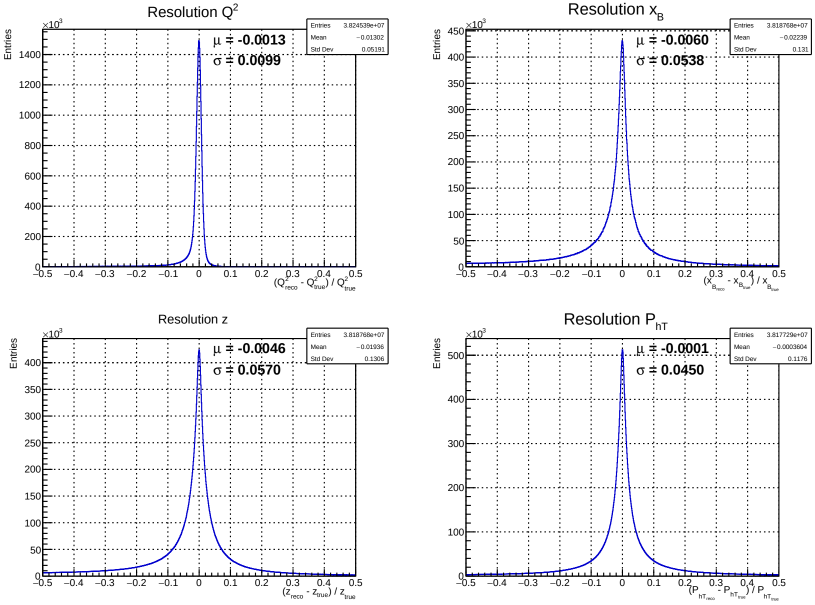
- Kinematic quantities reconstructed using **MC truth information**

Updated implementation

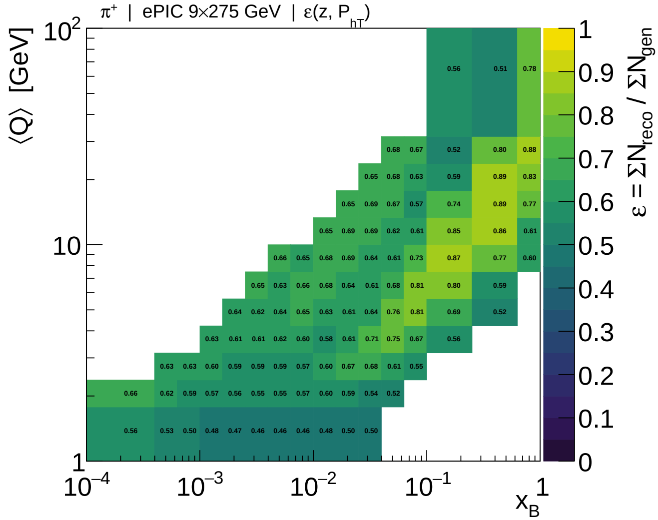
- Kinematic reconstruction now relies **only on reconstructed quantities**
- x, Q^2 and z reconstructed from the scattered electron and hadron four-momenta
- P_{hT} extracted in the reconstructed γ^*–N frame

Validation

- Resolution evaluated by comparing reconstructed and generated quantities
- $\dfrac{\Delta X}{X} = \dfrac{X_{\mathrm{reco}} - X_{\mathrm{MC}}}{X_{\mathrm{MC}}}$


Resolution Q2
Entries 3.824539e+07
Mean −0.01302
Std Dev 0.05191
μ = -0.0013
σ = 0.0099
Entries
×10³
(Q2reco - Q2true) / Q2true
Resolution xB
Entries 3.818768e+07
Mean −0.02239
Std Dev 0.131
μ = -0.0060
σ = 0.0538
(xBreco - xBtrue) / xBtrue
Resolution z
Entries 3.818768e+07
Mean −0.01936
Std Dev 0.1306
μ = -0.0046
σ = 0.0570
(zreco - ztrue) / ztrue
Resolution PhT
Entries 3.817729e+07
Mean −0.0003604
Std Dev 0.1176
μ = -0.0001
σ = 0.0450
(PhTreco - PhTtrue) / PhTtrue

Efficiency in the (x_B, Q) Plane



- Efficiencies evaluated across the full 4D SIDIS phase space, using reconstructed events that were generated in the same bin, divided by the total number of generated events per bin
- For visualization, the map shows the 2D projected efficiency, evaluated by combining events over the remaining variables (z, P_{hT})

Uncertainties on the cross section

Statistical uncertainty

Efficiency (binomial):

$$\delta\varepsilon = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N_{\mathrm{gen}}}}$$

Cross section:

$$\delta\sigma_{\mathrm{stat}} = \sigma\sqrt{\frac{1}{N_{\mathrm{reco}}} + \left(\frac{\delta\varepsilon}{\varepsilon}\right)^2}$$

Systematic uncertainty

We add in quadrature:

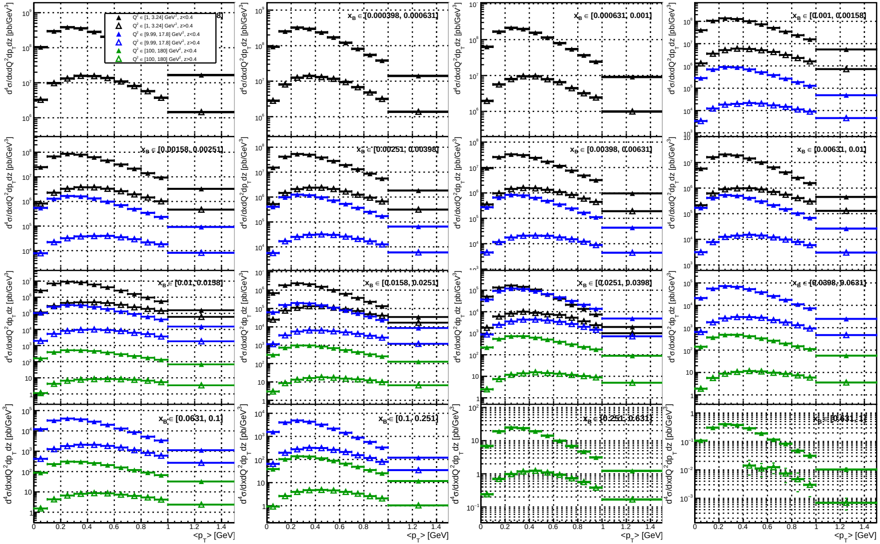
- 3.5% point-to-point uncertainty
- 1.5% luminosity scale uncertainty

DOI: https://doi.org/10.1103/PhysRevD.99.094004

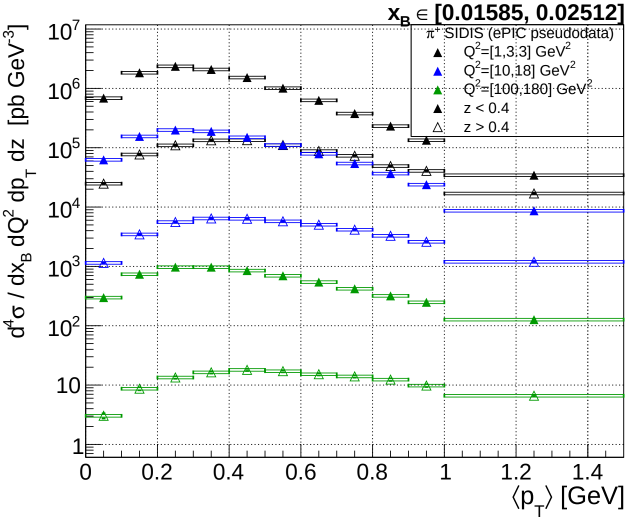
Luminosity scaling:

$$\delta\sigma_{\mathrm{stat}}^{\mathrm{proj}} = \delta\sigma_{\mathrm{stat}}\sqrt{\frac{\mathcal{L}_{\mathrm{MC}}}{\mathcal{L}_{\mathrm{proj}}}}, \quad \mathcal{L}_{\mathrm{proj}} = 2.5\ \mathrm{fb}^{-1}$$

Extracted SIDIS cross sections



Extracted SIDIS cross sections (zoom)



Fixed x bin $(0.016, 0.025)$

Different colors correspond to different Q^2 bins

Results shown for two z regions:

- low z: $0.05 < z < 0.4$
- high z: $0.4 < z < 1$

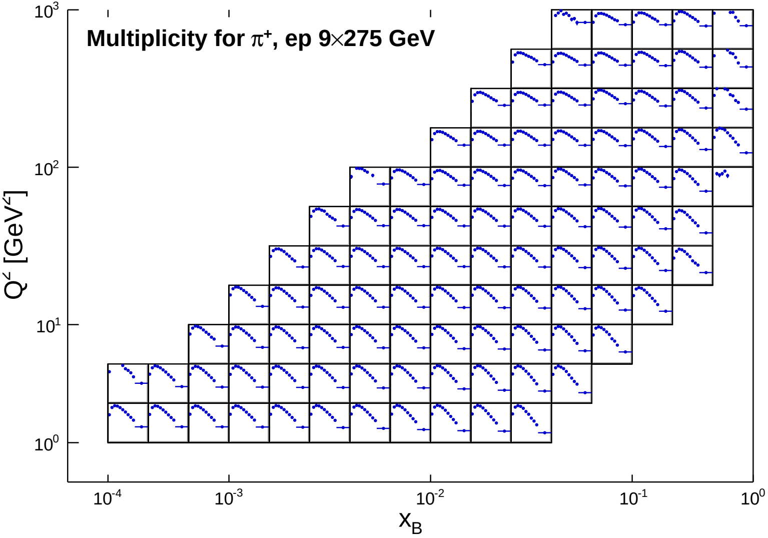
Hadron Multiplicity

The hadron multiplicity is one of the key observables in SIDIS analyses, providing information on the fragmentation process.

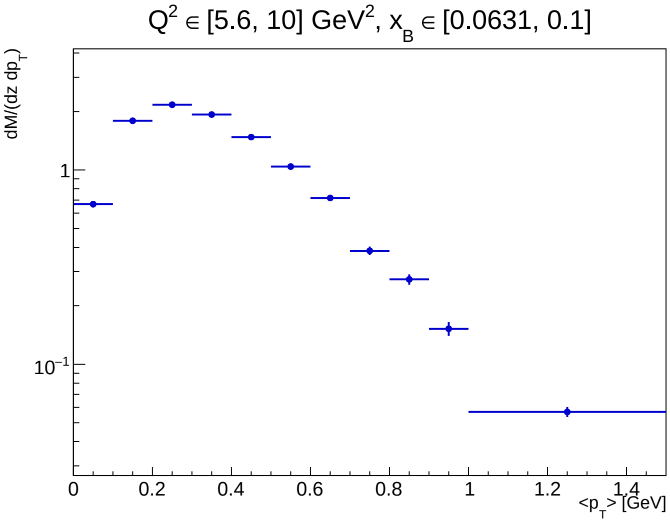
$$M_h(z, P_{hT}) = \frac{N_h(x, Q^2, z, P_{hT})}{N_{\mathrm{DIS}}(x, Q^2)}$$

- Ratio between the number of reconstructed hadrons of type h and the number of DIS events in the same (x, Q^2) bin
- Differential in (z, P_{hT}) bins

π^+ Multiplicity



π^+ Multiplicity



Summary and Outlook

Achievements:

- Developed the SIDIS analysis workflow for identified $\pi^{\pm}$ and $K^{\pm}$ production.
- Produced pseudo-data for 4D differential cross sections in (x, z, Q^2, P_{hT}) bins, including statistical and (simplified) systematic uncertainties.

Outlook:

- Analysis Note is almost ready for PWG review and approval.

Backup

Binning scheme

Q bin edges:

$$1.0,\ 1.78,\ 2.37,\ 3.16,\ 4.22,\ 5.62,$$
$$7.50,\ 10.0,\ 17.79,\ 23.71,\ 31.62,\ 100.0$$

P_{hT} bin edges:

$$0.0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,$$
$$0.6,\ 0.7,\ 0.8,\ 0.9,\ 1.0,\ 1.5$$

x_B bin edges:

$$0.0001,\ 0.0004,\ 0.0006,\ 0.001,\ 0.0016,\ 0.0025,\ 0.0040,\ 0.0063,$$
$$0.010,\ 0.016,\ 0.025,\ 0.040,\ 0.063,\ 0.100,\ 0.251,\ 0.631,\ 1$$

z bin edges:

$$0.05,\ 0.10,\ 0.15,\ 0.20,\ 0.25,\ 0.30,$$
$$0.40,\ 0.50,\ 0.60,\ 0.70,\ 0.80,\ 0.90,\ 1.00$$

RECO Kinematics

- Previous code version (from preTDR studies) relied heavily on MC truth for kinematics reconstruction.
- The code has been updated for upcoming production campaigns to perform a fully RECO kinematics.

Combining z bins

For visualization purposes, neighboring z bins are merged into two regions:

Procedure:

- Events are first selected within the corresponding z range
- Reconstructed and generated yields are summed over the selected z bins
- The cross section is recomputed from the aggregated yields
- Uncertainties are evaluated directly from the aggregated event counts

Event and hadron selection

Beam particles: identified via generator status and PDG codes for the incoming electron and proton.

Electron selection:

- Reconstructed track **truth-matched** to generated electron
- $|\eta_e| \leq 3.5$
- $0.01 \leq y \leq 0.95$
- $p_e > 1$ GeV

Hadron selection ($\pi^{\pm}$, $K^{\pm}$):

- Selection performed within events that passed the electron selection (DIS events)
- Reconstructed track **truth-matched** to generated hadron
- $|\eta_h| < 3.5$

PID Selection

Method: particle identification in current ePIC simulations relies on **parameterized lookup tables**, rather than full likelihood-based reconstruction algorithms.

Requirement: the reconstructed PID assignment is required to match the generated MC particle species (**truth-matched PID**).

Goal: then, misidentification effects are not considered in this analysis, with the scope to obtain a high-purity hadron sample for the cross-section extraction.

SIDIS Kinematic Reconstruction

Electron Method: event kinematics reconstructed from the scattered electron four-momentum, via the virtual photon $q^\mu = k^\mu - k'^\mu$.

For the transverse kinematics, events are boosted to the **γ^*-N frame**, where the virtual photon defines the z-axis. Here we extract the **hadron transverse momentum** P_{hT}, which gives access to unpolarized TMD effects ($\vec{P_{hT}} \approx z\vec{k_T} + \vec{P_T}$).

$$\text{Select } e^-, h^\pm \Rightarrow \text{ Compute the L.I. quantities } (x, Q^2, z) \Rightarrow \text{ Boost to } \gamma^*\text{-N} \Rightarrow \text{ Extract } P_{hT}$$