ZEUS–xx–xxx
14th December 2020

Azimuthal correlations in photoproduction at HERA and indications of multiparton interactions

Dhevan Gangadharan[1],
Silvia Masciocchi[1,2],Ilya Selyuzhenkov[2],
[1] Universität Heidelberg; [2] GSI

Abstract

Multiparton interactions and collective behaviour are studied in high-multiplicity ep photoproduction at a centre-of-mass energy $\sqrt{s} = 318\,\text{GeV}$ with the ZEUS detector at HERA. In contrast to point-like deep inelastic scattering, photoproduction refers to the exchange of quasi-real photons with an extended partonic substructure. Multiple distinct $2 \to 2$ partonic scatterings can therefore occur as in high-energy hadronic collisions. Measurements are made of two- and four-particle azimuthal correlations, multiplicity, transverse momentum, and pseudorapidity distributions for event multiplicities $N_{\text{ch}} \geq 20$. The observed correlations in photoproduction do not indicate significant collective behaviour like those observed in high multiplicity hadronic collisions at RHIC and the LHC. Comparisons to PYTHIA predictions are made and provide a strong indication of between two and four multiparton interactions on average.

Contents

1 Introduction

To be written for the paper and copied here...

2 ZEUS detector

The primary components of the ZEUS detector used in this analysis are the Central Tracking Detector (CTD), the Micro Vertex Detector (MVD), Barrel Calorimeter (BCAL), and Rear Calorimeter (RCAL). The CTD and MVD are enclosed within a 1.43 T thin superconducting solenoidal magnetic field.

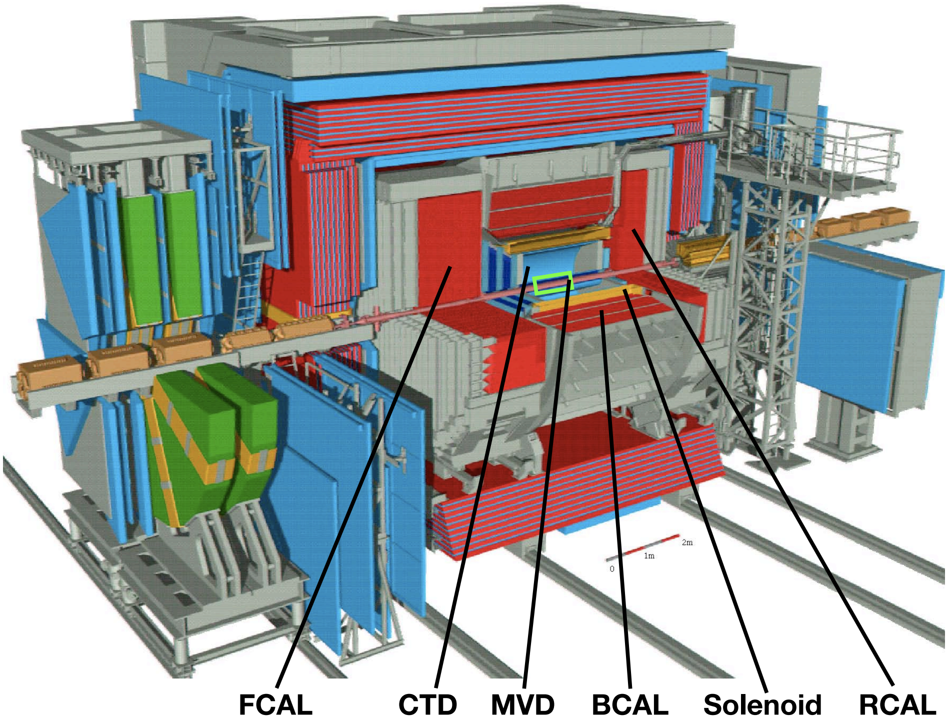


Figure 1: ZEUS detector pic 1

The CTD is the primary tracking detector providing tracking for polar angles between 15° ($\eta = 2.03$) and 164° ($\eta = -1.96$). The inner active radius is 18.2 cm and the outer active radius is 79.4. The physical radial extent for the inner and outer ends is 16.2 and 85 cm, respectively. Within the drift volume of the CTD there are 9 superlayers which alternate between axial and stereo types (5 axial, 4 stereo).

The MVD is composed of two components: the Barrel MVD (BMVD) and the Forward MVD (FMVD). The BMVD is composed of three silicon strip layers. Each layer is further composed of two joined r-ϕ and r-z planes to reconstruct the 2D coordinates of the traversing particle. Tracking with the BMVD with all three layers covers the approximate polar

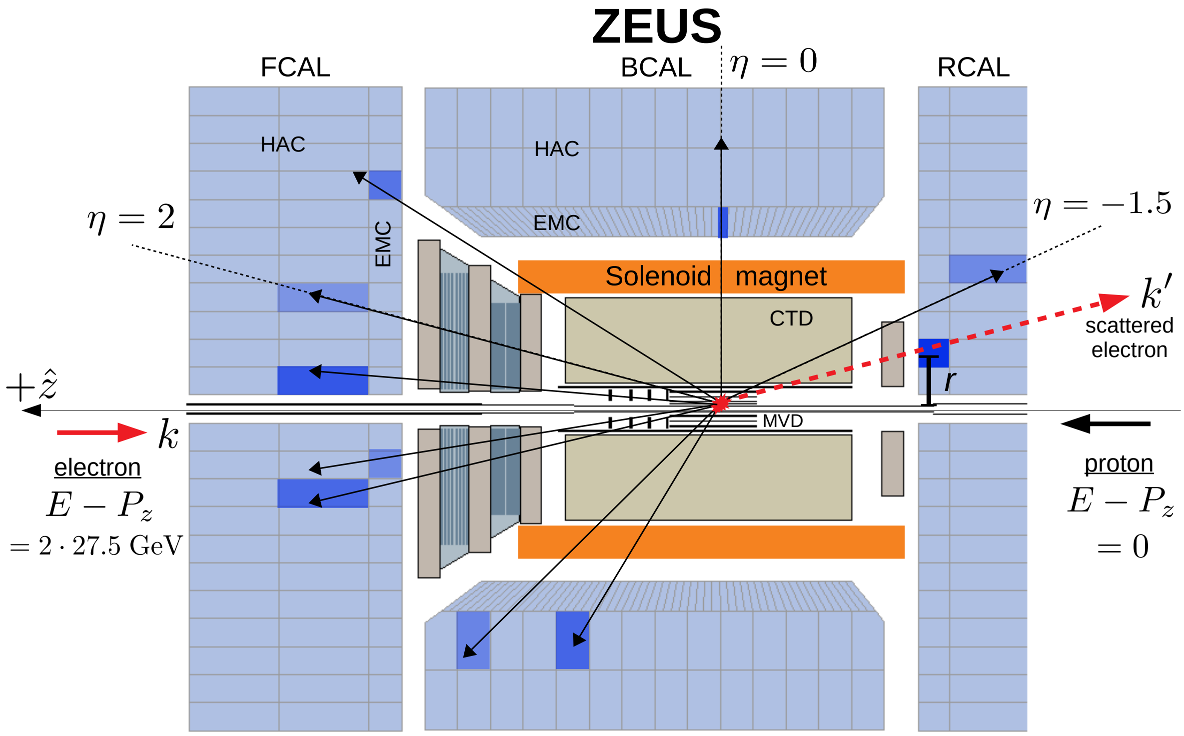


Figure 2: ZEUS detector pic 2

interval between 21.6° ($\eta = 1.65$) and 158° ($\eta = 1.66$). The FMVD is composed of four wheels sequentially placed longitudinally. Similar to the BMVD, each wheel is composed of two joined planes with silicon strips aligned in different directions such that together they determine the 2D coordinate of the traversing particle. The FMVD can extend the tracking down to about 7° ($\eta = 2.8$).

The BCAL and RCAL are used to measure the energy of the scattered lepton. There is also a FCAL but this is not used for our analysis. All three calorimeters are depleted Uranium calorimeters. In Fig 2, BAC stands for backing calorimeter and is not used in this analysis.

3 Analysis method

The main analysis is performed at high multiplicity, which we define as $N_{\mathbf{ch}} \geq 20$, due to an unavoidably large trigger bias to inclusive PhP at low multiplicity. All figures correspond to high multiplicity unless otherwise indicated.

3.1 $c_n\{2\}$ correlation functions

In heavy-ion collisions, particle correlations in the transverse plane are often used to probe the dynamics of the produced medium. In general, the transverse energy-density distribution within the collision zone is non-uniform and its shape fluctuates event-by-event. A development of a final state anisotropy along any of the common planes of symmetry (Ψ_m) in the initial state signifies the emergence of a collectivity. The azimuthal part of the momentum spectrum with respect to the common symmetry planes is usually expanded in a Fourier series [1–5]:

$$E\frac{d^3N}{d^3p} = \frac{1}{2\pi}\frac{d^2N}{p_{\mathrm{T}}\ dp_{\mathrm{T}}\ dy}\left[1 + 2\sum_{n=1}^{\infty} v_n\{m\}\cos\left[n(\varphi - \Psi_m)\right]\right]. \tag{1}$$

The Fourier coefficients, $v_n\{m\}$, quantify the degree of asymmetric particle production relative to Ψ_m. In the idealized case of a non-central collision between two spherical nuclei, the plane formed by the impact parameter and beam direction—the reaction plane—is a clear example of a symmetry plane. The resulting overlap zone is close to ellipsoidal, which leads to the second harmonic coefficient v_2 to be the dominant and the most studied in heavy-ion collisions. Higher harmonics are also important as they are sensitive to, or dominated by (in case of odd harmonics), fluctuations of the energy density in the overlap region of the nuclei.

According to Eq. (1), the coefficients $v_n\{m\}$ are related to each symmetry plane m by

$$v_n\{m\} = \langle\cos\left[n(\varphi - \Psi_m)\right]\rangle, \tag{2}$$

where the average goes over all particles in all events. For the case of $n = m$, one can extract $v_n\{n\}$ through the measurement of a two-particle correlation:

$$c_n\{2\} = \langle\cos\left[n(\varphi_1 - \varphi_2)\right]\rangle, \tag{3}$$

with azimuthal angles φ_1 and φ_2. This equation coincides with that of the two-particle cumulant Fourier analysis [6]. In the case of correlated particle production only via Eq. (1), $c_n\{2\}$ is related to $v_n\{n\}$ as

$$v_n^2\{n\} \approx c_n\{2\}. \tag{4}$$

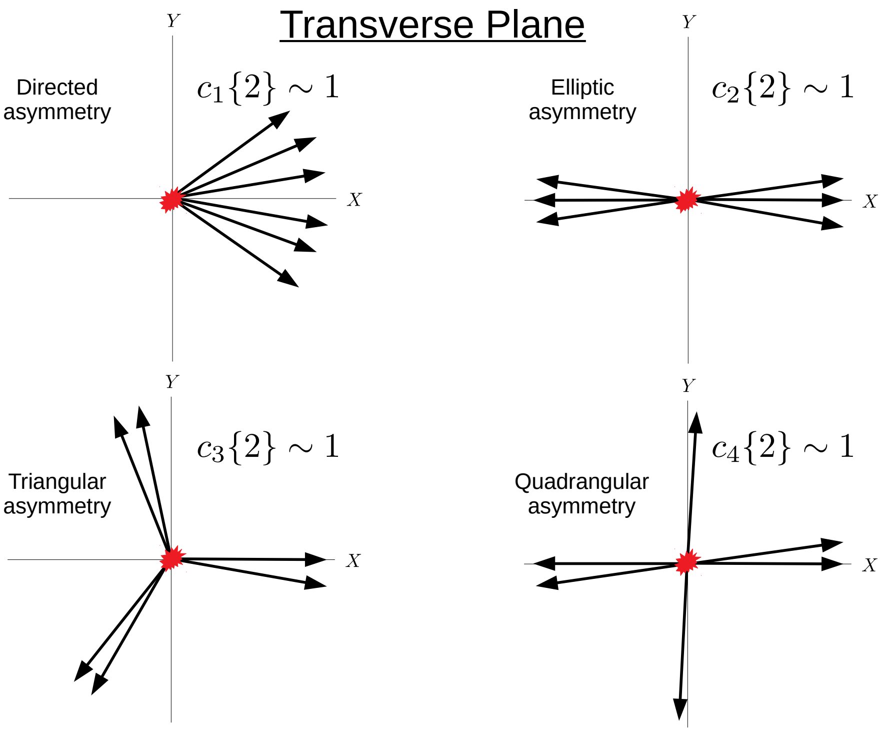


Figure 3: Sample event topologies in the transverse plane to illustrate positive $c_n\{2\}$ for the first four harmonics. The collision vertex is shown in red and the particle trajectories are shown with arrows. In each of the four panels, the value $c_n\{2\}$ for the dominant harmonic is shown while the others are in general non-zero. For instance, in the directed asymmetry panel, all $c_n\{2\}$ are greater than zero. For the elliptic asymmetry panel, $c_1\{2\} < 0$.

The form of Eq. (3) is convenient in that the symmetry-planes need not be explicitly determined. Interpretation of the $c_n\{2\}$ correlations in terms of $v_n\{n\}$ coefficients via Eq. (4) is complicated by contributions from other correlations unrelated to the symmetry plane. In two-particle correlation analyses they can be suppressed by requiring a sufficiently large pair separation in rapidity.

The two-particle correlation functions reported in this paper, is given by

$$c_n\{2\} = \sum_{e}^{N_{\mathrm{ev}}} \left[\sum_{i,j>i}^{N_{\mathrm{rec}}} w_{ij}^{(2)} \cos\left[n(\varphi_i - \varphi_j)\right] \right]_e \Bigg/ \sum_{e}^{N_{\mathrm{ev}}} \left[\sum_{i,j>i}^{N_{\mathrm{rec}}} w_{ij}^{(2)} \right]_e , \tag{5}$$

where the first sum over e is performed for all events, N_{ev}, and the sums over i and j run over the reconstructed tracks passing the track selection criteria in the event, N_{rec}. The correction factor for non-uniform acceptance is given by $w_{ij}^{(2)}$ and is described in Sec. 12.1.

Each harmonic of $c_n\{2\}$ characterises a different type of event topology as illustrated in Fig. 3.

We also measure four-particle cumulant correlations, $c_n\{4\}$, which are defined as [6]:

$$c_n\{4\} = \langle\cos\left(n(\varphi_1 + \varphi_2 - \varphi_3 - \varphi_4)\right)\rangle - 2\langle\cos\left(n(\varphi_1 - \varphi_3)\right)\rangle\langle\cos\left(n(\varphi_2 - \varphi_4)\right)\rangle. \quad (6)$$

The angled brackets denote an average first over all particles in an event followed by an average over all events. Four-particle cumulants are defined such that two-particle "direct" correlations, i.e. correlations arising from two-body resonance decays are explicitly subtracted off, leaving a contribution from genuine four-particle correlations. Due to this feature, they are a more robust probe of the collective behaviour. They are measured as a function of the p_{T} of the so-called particle-of-interest (POI), which is just the p_{T} associated with particle 1 in Eq.6.

It is known [7] from previous studies that four-particle cumulant measurements can be biased when integrating over a wide multiplicity bin, as we do in this analysis. To minimize this bias, we apply event weights to the averages in Eq.6, which are given by the number of pair and quadruplet combinations: $N(N-1)$ and $N(N-1)(N-2)(N-3)$.

In this analysis we present measurements of the two- and four-particle correlation function for the first two harmonics for $N_{\mathrm{ch}} \geq 20$ projected against Q^2, pseudo-rapidity difference $\Delta\eta = |\eta_1 - \eta_2|$, mean transverse momentum $\langle p_{\mathrm{T}}\rangle = (p_{\mathrm{T},1} + p_{\mathrm{T},2})/2$, and p_{T} POI for four-particle correlations.

3.2 $\Delta\eta - \Delta\varphi$ correlation function and event mixing

Two-particle correlations are also shown as a 3-dimensional histogram projected against $\Delta\eta$ and $\Delta\varphi$. The correlation function in this case is:

$$C(\Delta\eta, \Delta\varphi) = \frac{S(\Delta\eta, \Delta\varphi)}{B(\Delta\eta, \Delta\varphi)}, \quad (7)$$

where the pair distributions are formed by taking the first particle from a given event and other from either the same- or mixed-event with similar values of N_{ch} and Z-vertex position. The pair distributions are given by $S(\Delta\eta, \Delta\varphi) = N^{\mathrm{same}}_{\mathrm{pairs}}(\Delta\eta, \Delta\varphi)$ and $B(\Delta\eta, \Delta\varphi) = N^{\mathrm{mixed}}_{\mathrm{pairs}}(\Delta\eta, \Delta\varphi)$, respectively. The S distribution is corrected with w_{ij} while B is corrected with $w_i w_j$. Both distributions are self-normalized before division. Events are mixed if they are similar enough in multiplicity and Z-vertex position (V_z). Events for mixing are binned 10 wide in N_{rec} (0 to 10, 10 to 20,...) and 15 cm wide in V_z (-30 to -15, -15 to 0,...).

4 Data & MC samples

The measurement is based on data collected with the ZEUS detector at the HERA collider during the period 2003–2007 (HERA II) where an electron beam of energy 27.5 GeV collided with a proton beam of 920 GeV. The integrated luminosity was 430 pb^{1} for e-p centre-of-mass energies of 318 GeV. Photoproduction events are the main subject of this analysis although DIS is also studied. The real data samples and corresponding statistics for the PhP event sample are shown in Tab 1

Dataset	All events	T	T+V+O PhP (DIS)	T+V+O+$N_{\mathrm{ch}} > 20$ PhP (DIS)
03p	3.7 M	0.99 M	0.27 M (0.2)	0.031 M (0.001)
04p	47.5	12.6	3.7 (4.7)	0.455 (0.019)
05e	130.0	43.9	14.8 (16.4)	1.972 (0.082)
06e	44.2	13.4	4.5 (7.0)	0.726 (0.034)
06p	86.6	26.3	9.3 (11.8)	1.402 (0.053)
07p	41.2	11.1	3.7 (5.4)	0.524 (0.022)
Total	353.2 M	108.3 M	36.3 (45.5)	5.110 M (0.211)

Table 1: *Real data samples and event tallies for the PhP (DIS) analysis. The analyzed real data samples and number of events. T = Trigger selections, V = Vertex cuts, O = Offline cuts. v08b orange nTuples used.*

The Pythia PhP light-flavor jet dataset (MC $Q^2 < 2$) was the MC dataset used for efficiency and trigger bias corrections. Both the direct and resolved components were summed together. The MC data samples and corresponding statistics are shown in Tab 2.

Dataset and code names	All events	T	T+V+O	T+V+O+$N_{\mathrm{ch}} > 20$
light-flavor jet 0304p cny324, cnx324, cnw324, cn3z24	128.2 M	14.2 M	8.9 M	1.0 M
light-flavor jet 05e dsmr25	121.4	14.0	8.9	0.9
light-flavor jet 06e etrr26	149.5	17.3	11.1	1.3
light-flavor jet 0607p fiw627	195.5	22.7	14.4	1.5
Total	594.6 M	68.2 M	43.3 M	4.7 M

Table 2: *The analyzed Pythia light-flavor jet PhP MC samples and number of events. MC $Q^2 < 2$. Both direct and resolved components were summed together. T = Trigger selections, V = Vertex cuts, O = Offline cuts. v08b orange nTuples used.*

Ariadne and Lepto DIS generators were used to correct the DIS part of this analysis. The statistics from each were similar and given in Tab.3.

Dataset	All events	After DIS trigger	After all cuts
Ariadne 0304p	18.2 M	12.3 M	10.0 M
Ariadne 05e	60.4	40.6	31.4
Ariadne 06e	23.9	15.9	13.0
Ariadne 0607p	62.0	41.4	33.7
Total	164.5 M	110.2 M	88.1 M

Table 3: *The analyzed Ariadne nondiffractive MC data samples and number of events. Technical dataset name Ariadne_Low_Q2_NC_DIS. Lepto nondiffractive MC data has the nearly the same statistics. Technical dataset name for that is Lepto_low_Q2_NC_DIS. v08b orange nTuples used.*

4.1 ROOT version

The analysis is performed using root 6.22/02 on naf-zeus12 (CENT OS 7). The previous analysis of two-particle correlations in DIS was performed using root 5.34 on naf-zeus11 (CENT OS 6). As a test, we compared a few sample histograms (N_{rec}, p_{T}, η, $c_1\{2\}$ versus $\Delta\eta$) obtained using the two mentioned setups and confirm that the histogram contents were identical.

5 Event selection in reconstructed data

5.1 Triggers

We start by stating that there is no ZEUS trigger available to obtain an inclusive PhP data sample. We investigate which subset of the available TLT triggers allow us to obtain a least biased sample and how such a bias can be estimated and corrected. To illustrate the bias, in Fig. 4 we show an MC event from the light-flavor jet PhP sample at high multiplicity (N_{gen} =15, defined in Sec. 8.1) but which does not fire a single TLT. ZEUS triggers were primarily designed to study perturbative processes such as jets and heavy-flavor production. Being such, soft events which have no structures resembling the said processes will not fire any available TLT. This has the unfortunate consequence of rejecting many soft high multiplicity events, which are the events of interest in this analysis. Nevertheless, such soft events sometimes accidentally fire a TLT and it is those which we seek to study.

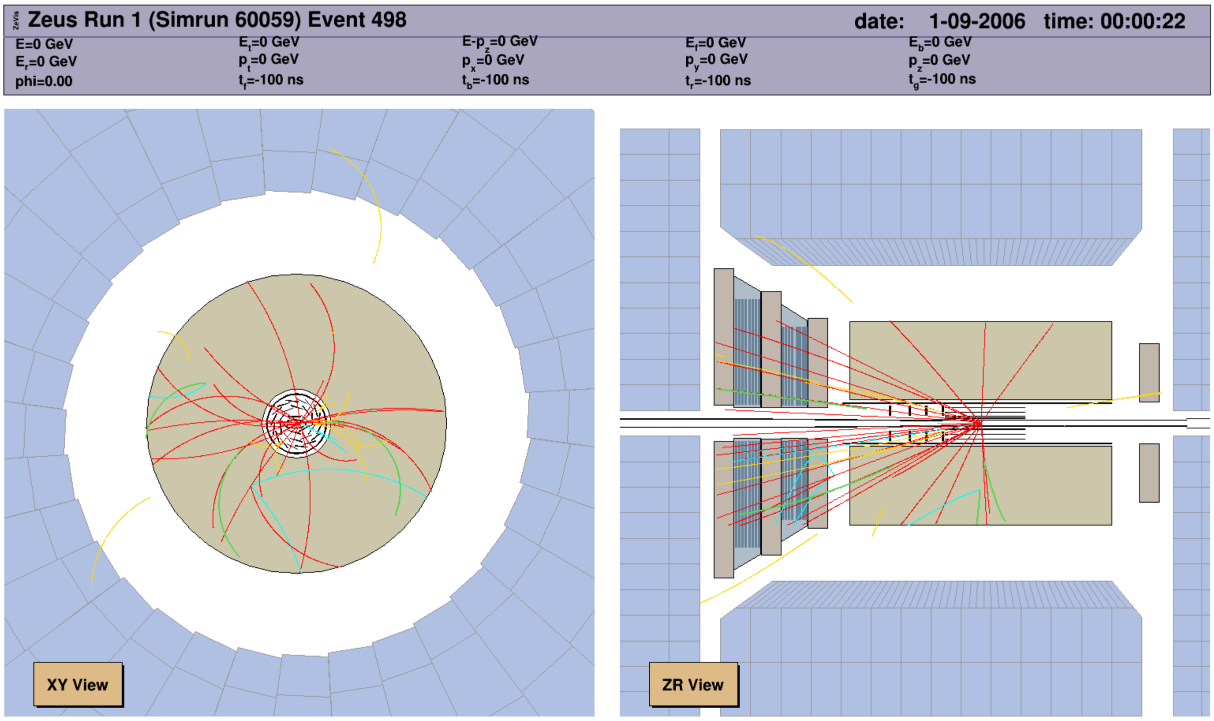
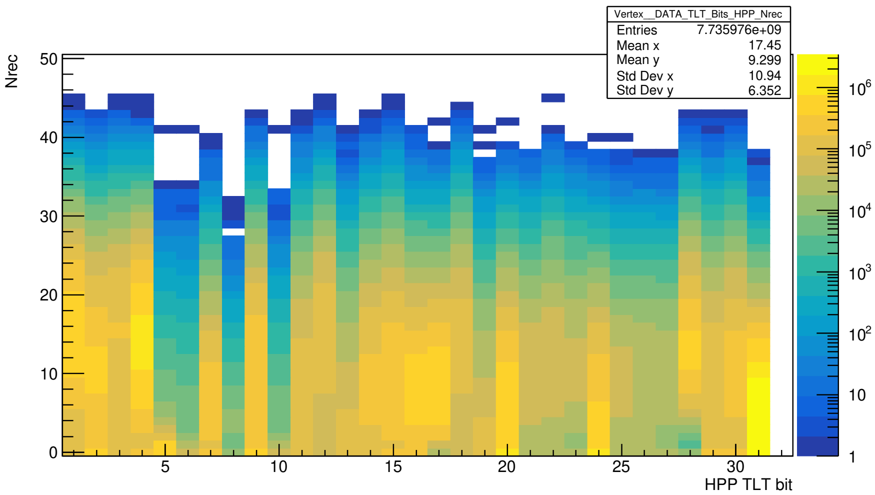
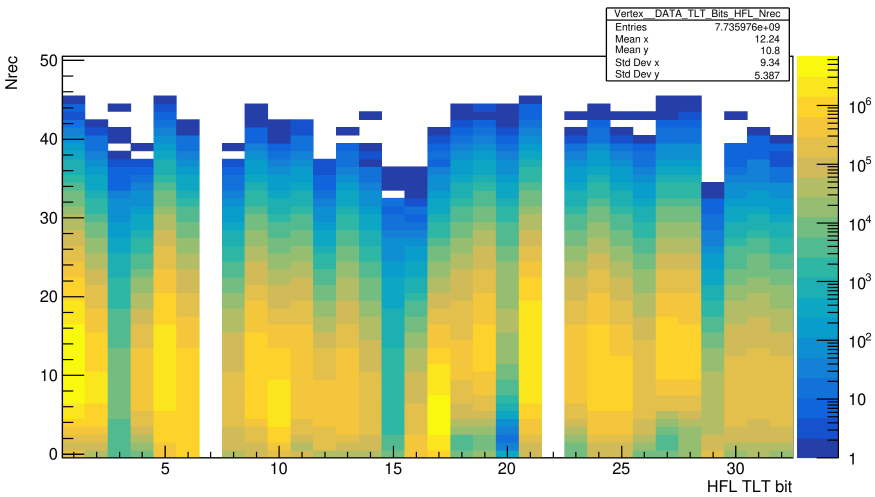


Figure 4: Zevis event display of a MC PhP high multiplicity event which did not fire any TLT. $N_{\rm gen}$ =15. Light-flavor jet 0607p. Red tracks are primary.

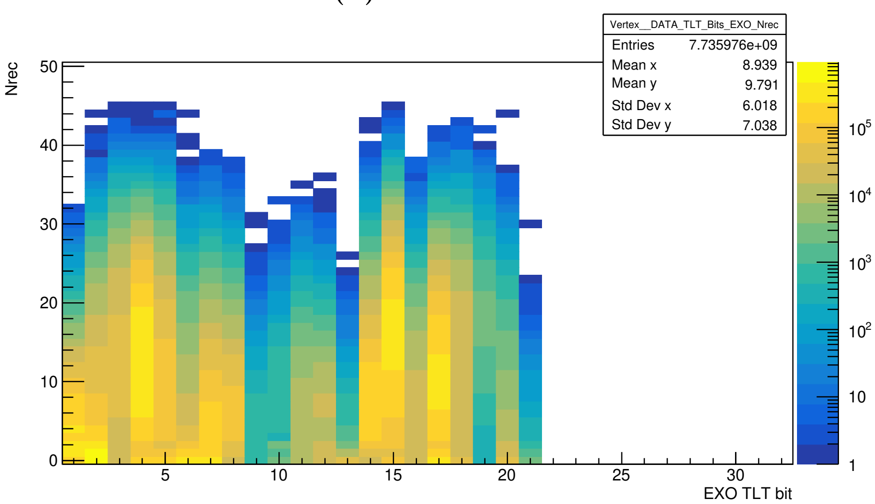
We are mostly interested in high multiplicity events where soft physics dominate. To determine which third-level-triggers (TLTs) to utilize in this analysis we plot the multiplicity distributions for each TLT from the HPP, HFL, and EXO trigger groups. They are shown in Figs. 5a-5c.



(a) HPP TLTs.



(b) HFL TLTs



(c) EXO TLTs.

Figure 5: HPP, HFL, and EXO trigger rates vs N_{rec}. All HERA II ZEUS data.

5.1.1 Trigger bias studies for the top-20

An ordered list of the top-20 high-multiplicity triggers is shown in Fig. 6.

The EXO trigger group only appears starting at 10^{th} place in Fig. 6 and so we do not further pursue it. A brief description of many of the HPP and HFL TLTs is shown in Figs. 7-8.

Using the top-20 TLTs, we assess the trigger bias on a sample of distributions which will be important for our analysis: N_{ch}, $c_1\{2\}$ and $c_2\{2\}$ versus N_{ch}. We compare the distributions from MC events which triggered one of the top-20 TLTs to that from reference events. We use the light-flavor jet MC sample for this. The reference event sample is formed from all MC events which have a reconstructed event satisfying our primary vertex constraint (See Sec. 6) and a sinistra electron candidate with $E_e < 5$GeV. This maximum value for the sinistra candidate energy was used early in the analysis to help isolate PhP events. **Note that the resulting comparison is not an estimate of the full trigger bias since our reference sample implicitly requires any TLT to fire. A reconstructed vertex is only present for MC events which fire at least one TLT. Nevertheless, this type of comparison is valid to judge the relative bias among the top-20 TLTs.** Figures 9-11 show the trigger bias comparisons from the top-20 TLTs in Pythia light-flavor jet MC.

$N_{rec} > 20$

TLT	# events
HFL 1	6613759
HFL 21	5830070
HFL 5	4066308
HPP 4	2567982
HPP 1	2281568
HFL 27	1865013
HFL 19	1827380
HFL 28	1734148
HFL 24	1582638
EXO 15	1579617
EXO 4	1183435
HFL 9	1120284
HFL 2	908181
HPP 3	868207
HPP 28	865885
HFL 6	836168
HPP 2	741468
HFL 18	674742
HPP 12	645157
HFL 25	605287
HFL 11	558036
HFL 23	554917
HPP 9	532315
HPP 15	504376
HPP 18	446312
HFL 10	436954
EXO 17	419884

Top 20

Figure 6: TLT triggers shown in descending order of number of events at high multiplicity. Computed from full HERA II ZEUS dataset.

TLT	Short description	Long description
HPP 1	Very high ET	HPP SLT: 1,2,3,4,5,8 E - pz > 8 GeV ET_cone > 25 GeV pz/E < 0.95 or E - pz > 12 GeV
HPP 2	Inclusive Jet, cone finder	HPP SLT: 1,2,3,4,5,8 >=1 eucell jet with ET > 10 Gev, eta < 2.5
HPP 3	Very high ET	HPP SLT: 1,2,3,4,5,8 ET_cone > 30 GeV
HPP 4	High ET NC DIS	HPP SLT: 1,2,3,4,5,8 E - pz > 30 GeV ET_cone > 20 GeV
HPP 12	Trijet Photoproduction, cone finder	HPP SLT: 1,2,3,4,8 >=3 eucell jets with ET > 5 Gev, eta < 2.5
HPP 28	Double K-short	SLT: HPP1,HFL1 E - pz > 7 GeV At least 4 tracks with pT > 0.25 GeV At least 2 combinations of tracks with invariant mass between 0.447 and 0.547 GeV

Figure 7: HPP TLT triggers with many events at high multiplicity. Details can be found here.

TLT	Short description	Long description
HFL 1	Charmed hadrons in PHP	Or of all HFM triggers with hard cuts: pT thresholds and invariant mass thresholds of decay daughters.
HFL 2	Charmed hadrons in DIS	DIS electron Or of all HFM triggers with loose cuts:
HFL 5	inclusive dijets (similar to old HPP 14)	Two jets ET>4.5, eta<2.5 (EUCELL) Pz/E < 0.95 and E-Pz<100
HFL 6	jets in DIS	Two Jets ET>3.5, eta<2.5 (EUCELL) Pz/E < 1.0 and E-Pz<100
HFL 9	electron in PHP	Number of tracks > 2, Island Energy < 1000 Momentum track > 0 , pt of the track > 1.4 GeV , 0.6 < track theta < 2.55 , DCA < 30. EMC Island energy Fraction eEMCIsland/EIsland < 0.8
HFL 18	D* gold selection	See web pages for longer description.
HFL 19	D0/D0-bar mixing	See web pages for longer description.
HFL 21	MESON + jets	Two Jets ET>3.5, eta<2.5 (EUCELL) Pz/E < 1.0 and E-Pz<100 .or. of any of the 6 D meson low Pt cut channels
HFL 24	jet(s) + electron	See web pages for longer description
HFL 25	jet(s) + muon	See web pages for longer description
HFL 27	MVD inclusive trigger Only active since May 30th 2006 (~40% of HERA II integrated lumi)	All SLT PHP, DIS and MUON slots MVD vertex within -30 cm < z(vtx) < 30 cm at least 4 tracks fitted to the primary vertex Et > 8 GeV (excluding the 1st two inner rings around the beam pipe) At least three tracks with pt > 0.75, 0.6, 0.45 GeV Impact parameter significance cut for the 3rd highest significance track. The impact parameter significance is evaluated with respect to the primary event vertex.
HFL 28	MVD inclusive trigger using beam spot Only active since May 30th 2006	Same cuts as for HFL 27, but the impact parameter significance is evaluated with respect to the beam spot.

Figure 8: HFL TLT triggers with many events at high multiplicity. Details can be found here.

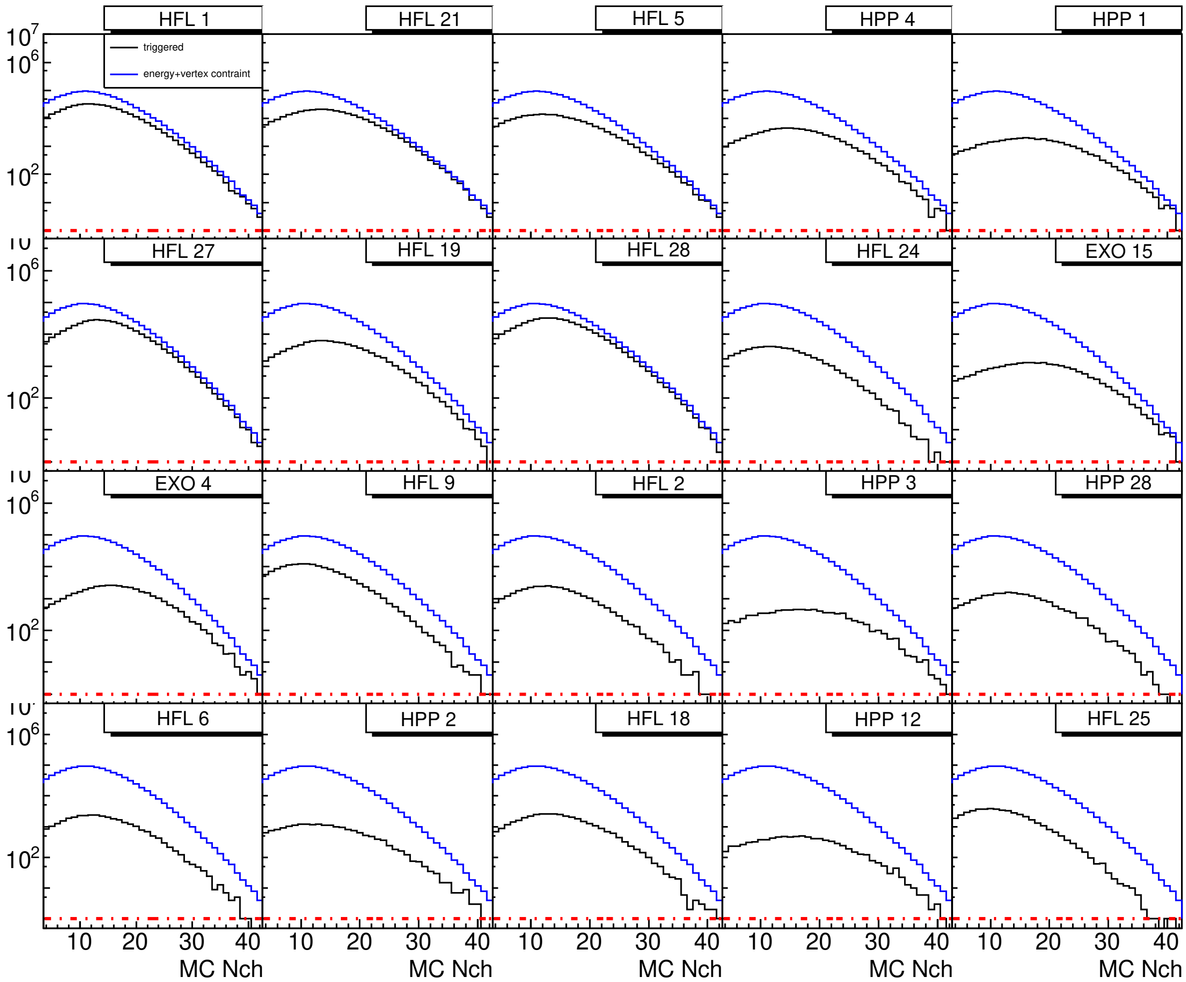


Figure 9: $N_{\mathrm{ch}}^{\mathrm{gen}}$ distribution trigger bias study in Pythia lfjet 0607p. Pure MC generator level correlations. This represents a comparison of trigger biases among the TLTs and is not an absolute estimate of the trigger bias.

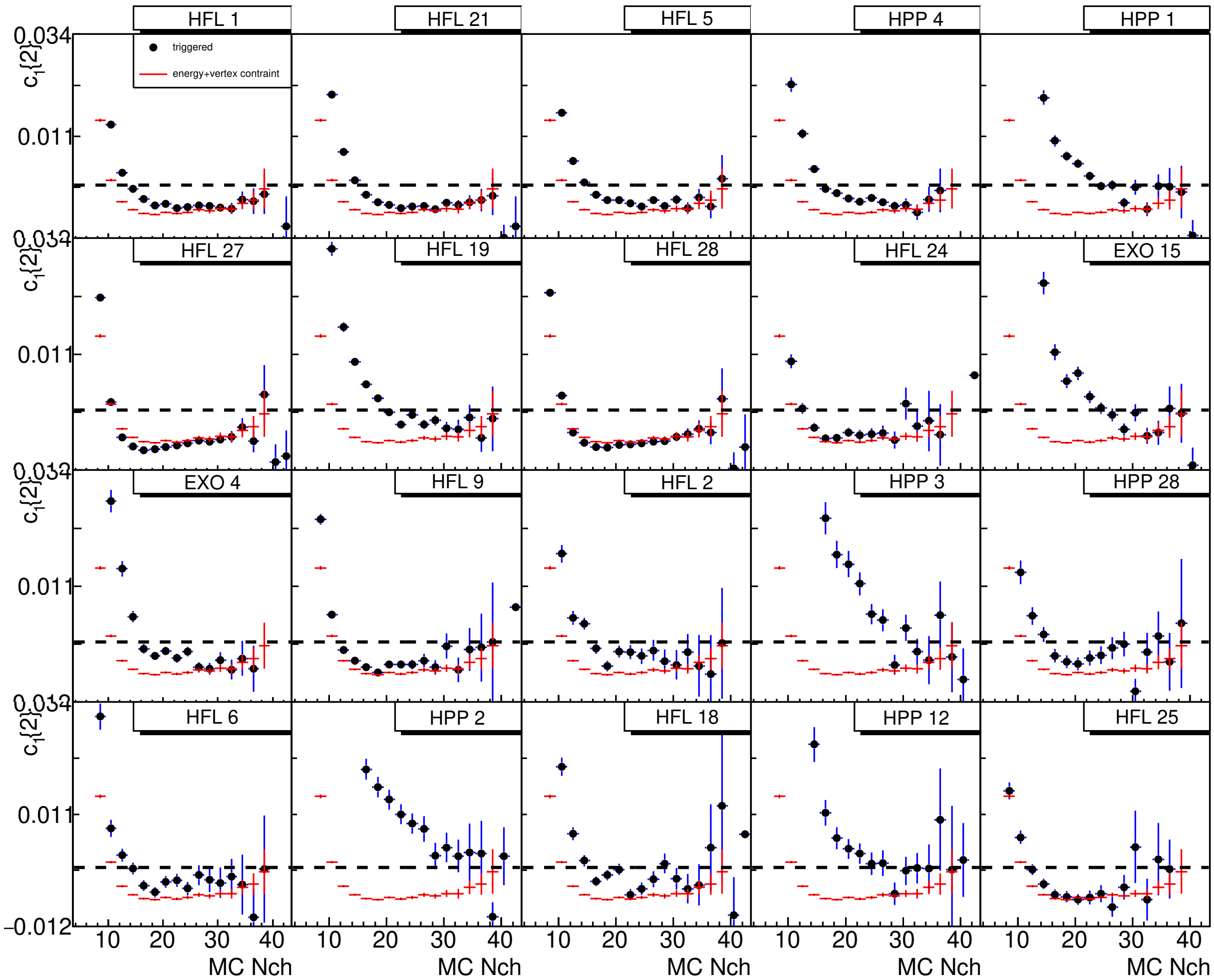


Figure 10: $c_1\{2\}$ trigger bias study in Pythia lfjet 0607p. Pure MC generator level correlations. This represents a comparison of trigger biases among the TLTs and is not an absolute estimate of the trigger bias.

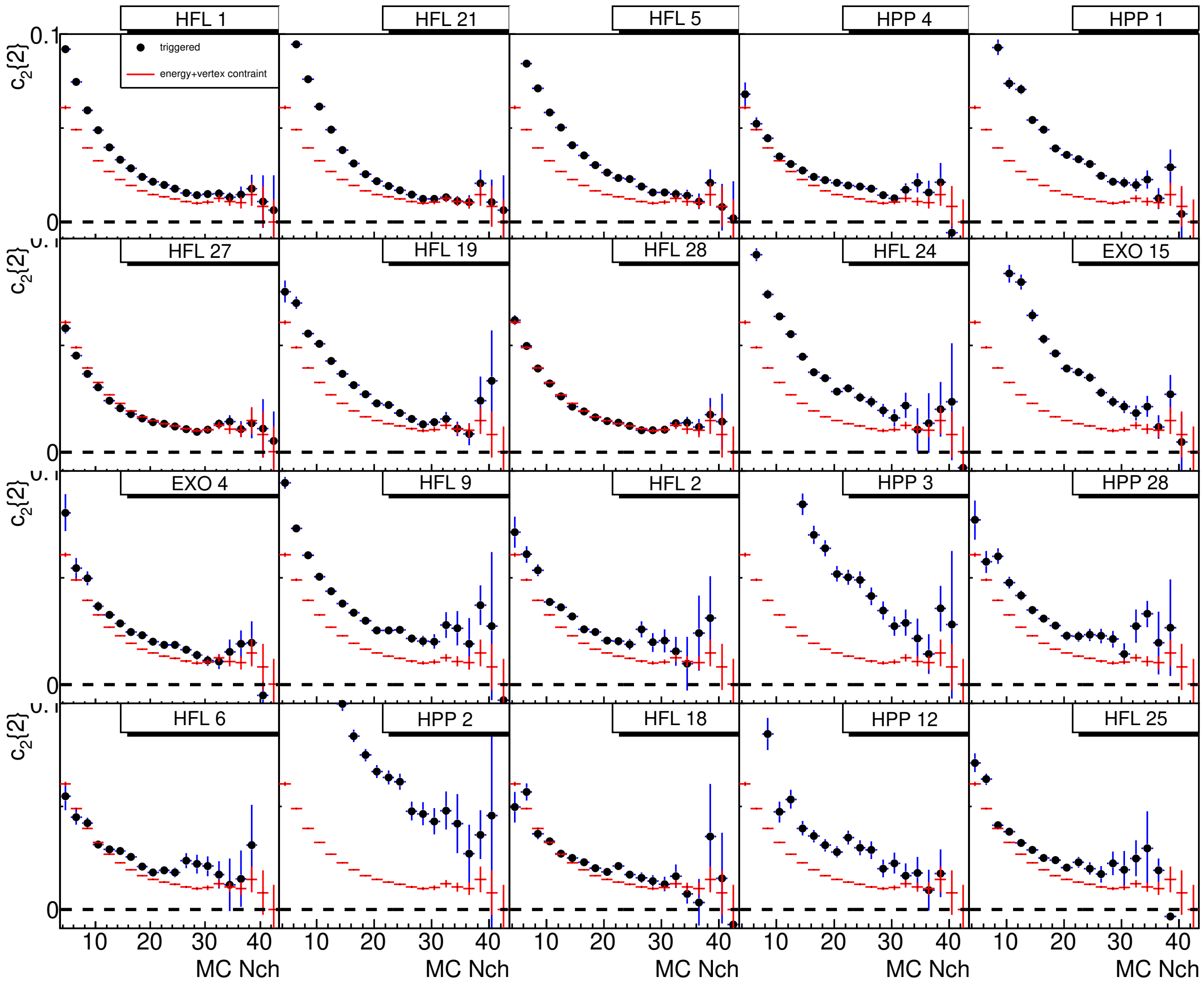


Figure 11: $c_2\{2\}$ trigger bias study in Pythia lfjet 0607p. Pure MC generator level correlations. This represents a comparison of trigger biases among the TLTs and is not an absolute estimate of the trigger bias.

From fig. 11 it is clear that HFL 28 is the least biased TLT from the top-20 in regards to $c_2\{2\}$. It is also one of the least biased in regards to $c_1\{2\}$ and $N_{\rm ch}^{\rm gen}$. For this reason, HFL 28 was chosen early in this analysis as the sole trigger to use for PhP studies. However, now we use an “or" of four of the least biased HFL triggers from the top-20 list: HFL 1, HFL 5, HFL 21, HFL 28. The choice of additional HFL triggers was made based on Fig. 12, where we investigated which TLTs fired when HFL 28 **did not** fire.

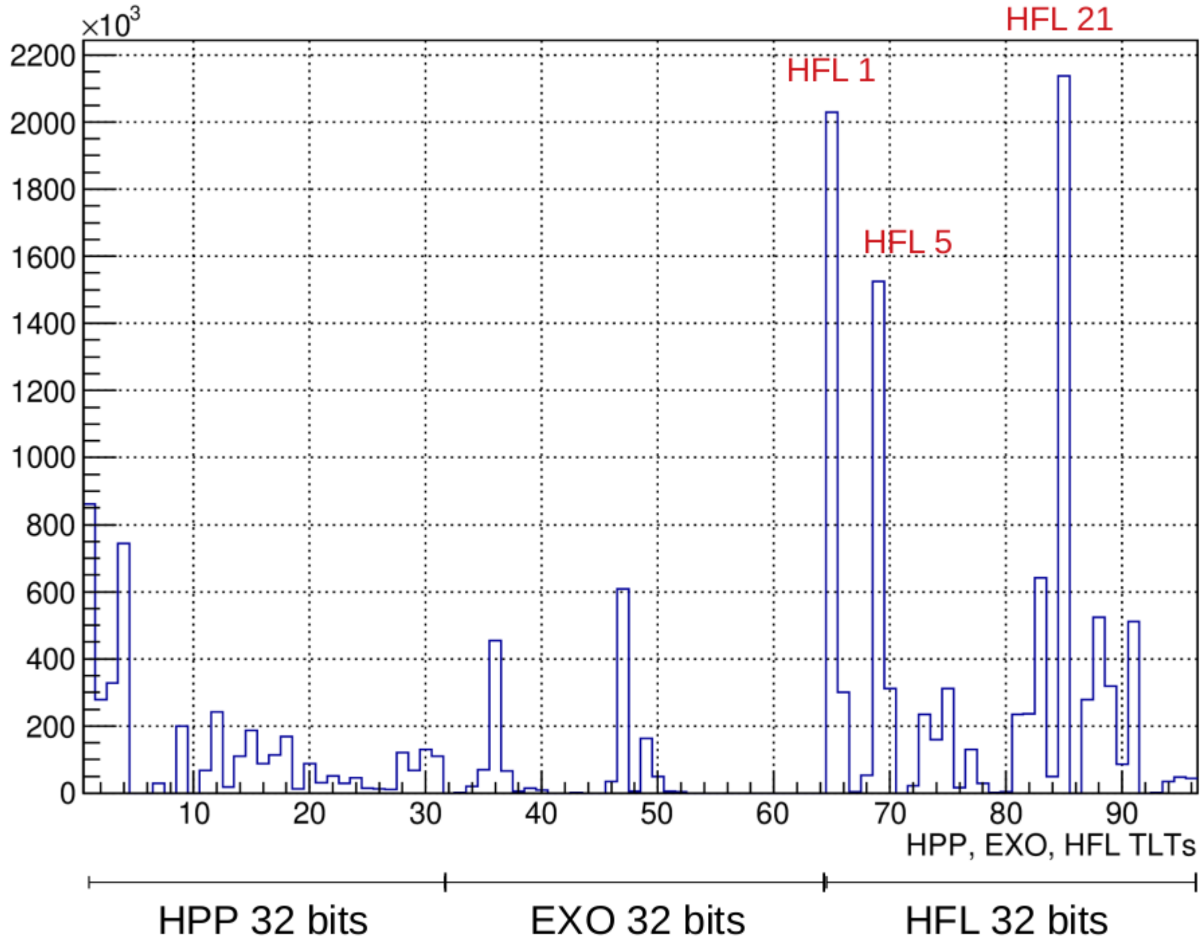


Figure 12: TLTs which fired when HFL 28 did not fire. $N_{\rm rec} > 20$.

5.1.2 HFL trigger bias to a DIS analysis

In this section we study the bias of the HFL 1, 5, 21, and 28 to a known inclusive data sample, DIS, for which ZEUS possesses both an inclusive data and MC sample. For PhP, neither exist. However, as we'll see, much of the bias to our PhP sample induced by our trigger selection can be reduced by focusing only on high multiplicity. Figures 13 and 14 compare the correlations obtained using the standard DIS triggers to an individual HFL trigger. Offline DIS cuts are applied to both. A similar investigation was done by applying DIS trigger and offline cuts throughout and observing how much the “and” of an

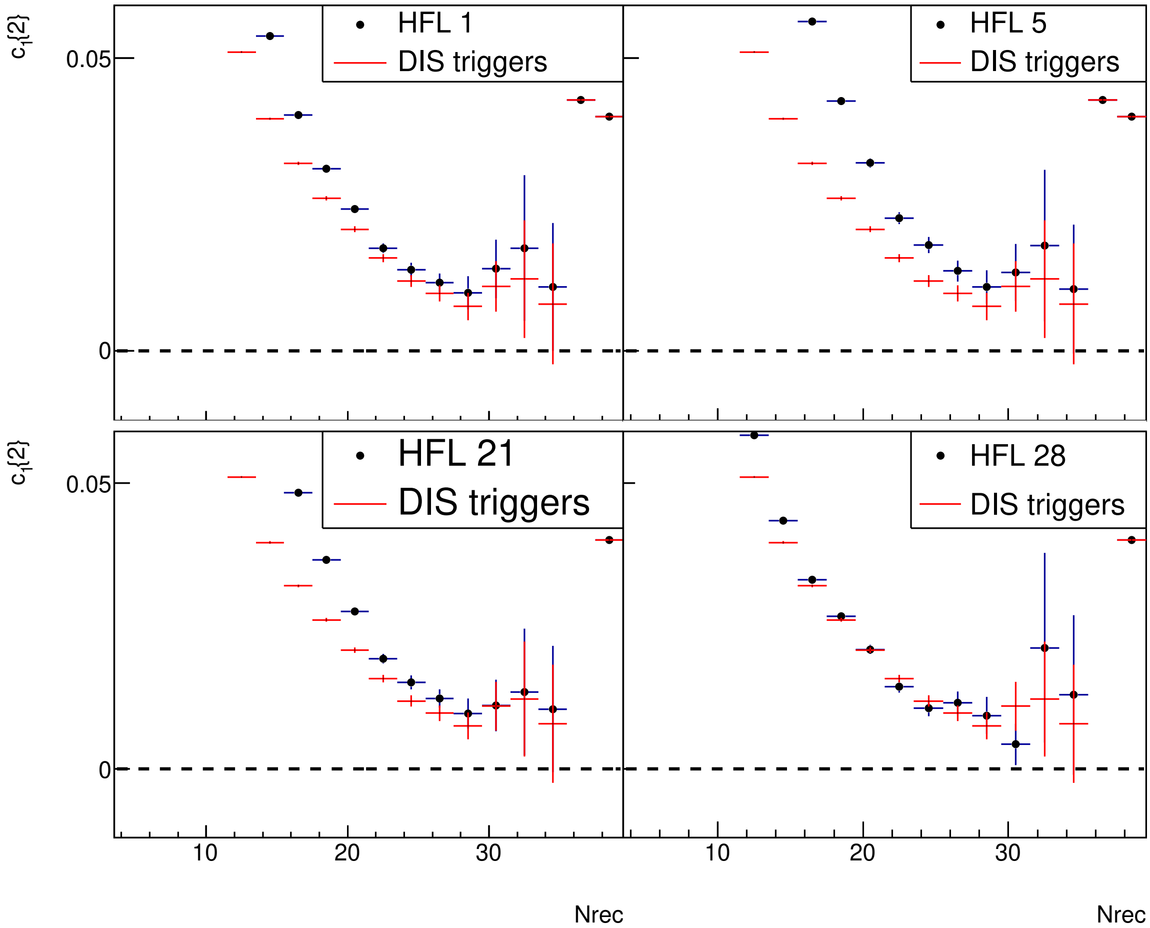


Figure 13: $c_1\{2\}$ obtained with DIS triggers compared to 4 HFL triggers. Offline DIS selection applied to both. ZEUS data.

additional HFL trigger changes the correlations. This is shown in Fig. 15 for ZEUS data and in Fig. 16 for LEPTO data. From these studies we conclude that the bias caused by the HFL triggers to an inclusive DIS sample is sufficiently small above $N_{\text{rec}} \sim 20$.

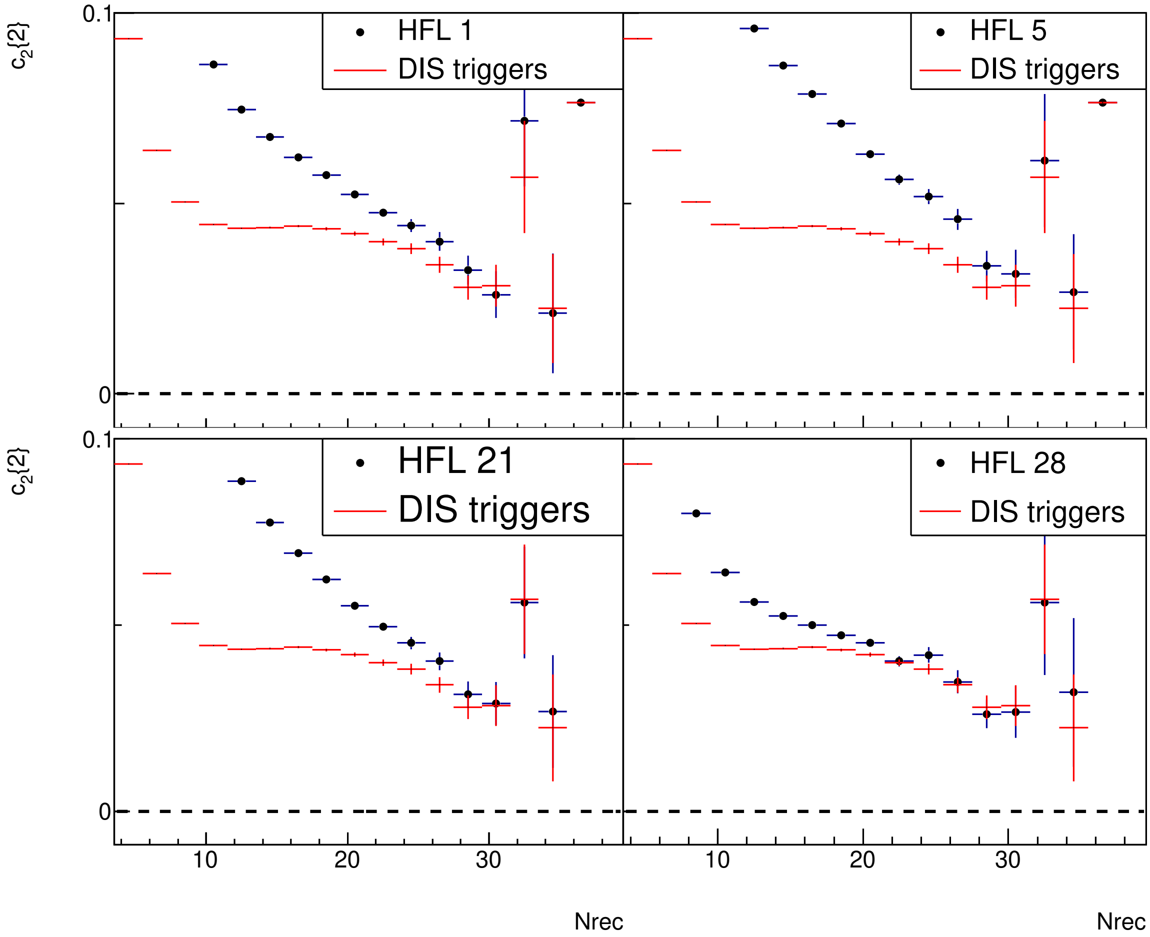
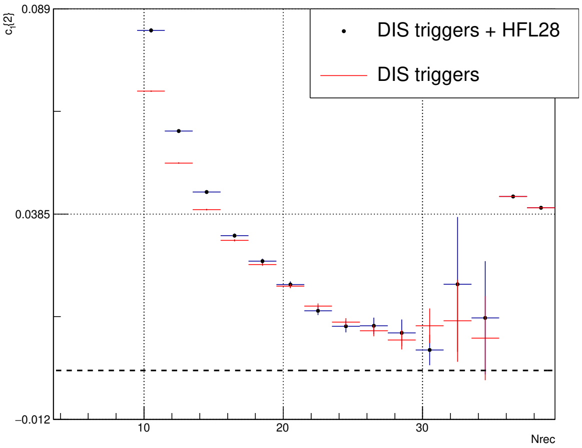
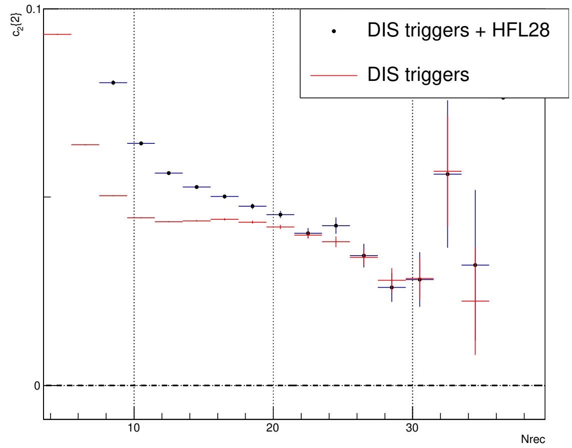


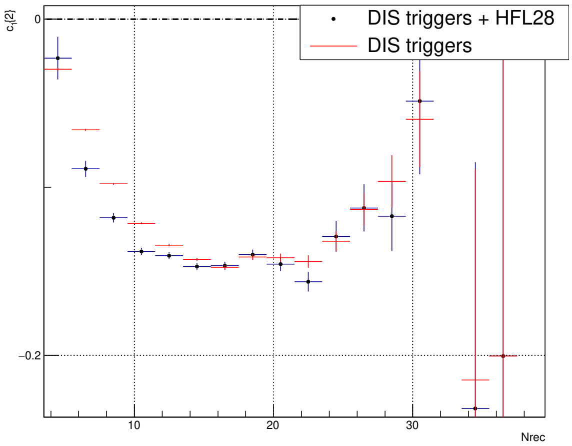
Figure 14: $c_2\{2\}$ obtained with DIS triggers compared to 4 HFL triggers. Offline DIS selection applied to both. ZEUS data.



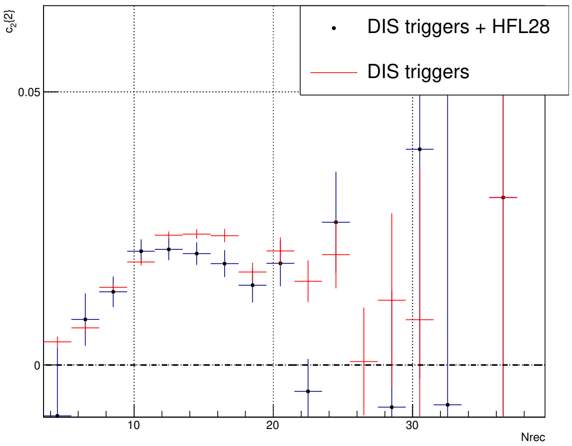
(a) $c_1\{2\}$



(b) $c_2\{2\}$

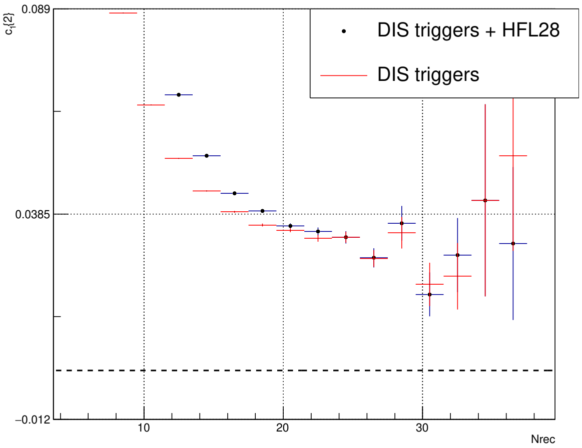


(c) $c_1\{2\}$, $p_{\mathrm{T}} > 0.5, \Delta\eta > 2$.

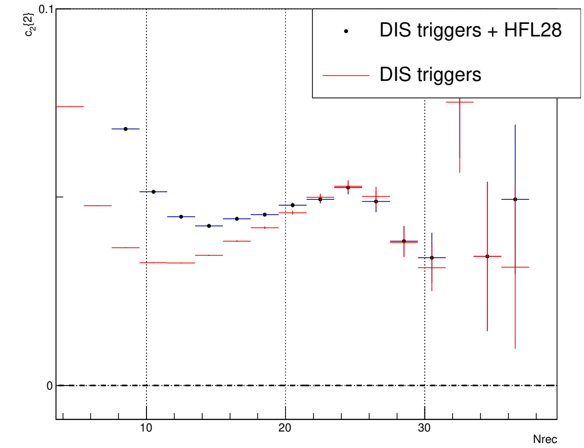


(d) $c_2\{2\}$, $p_{\mathrm{T}} > 0.5, \Delta\eta > 2$.

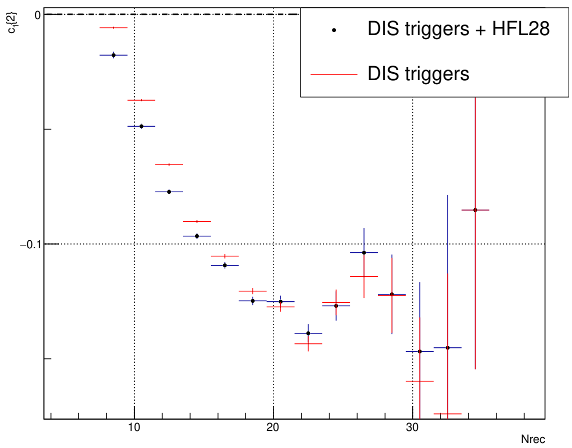
Figure 15: DIS Offline + DIS triggers applied to all. ZEUS data.



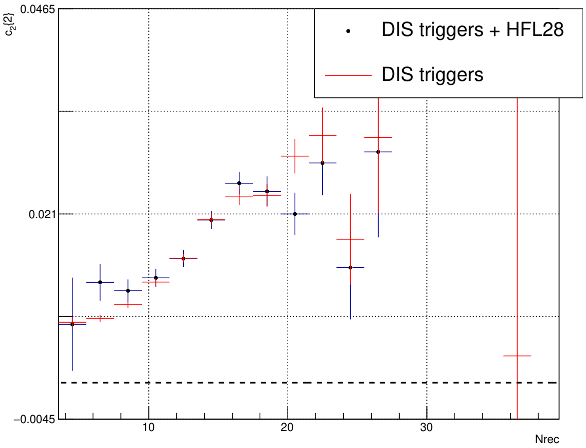
(a) $c_1\{2\}$



(b) $c_2\{2\}$



(c) $c_1\{2\}, p_{\mathrm{T}} > 0.5, \Delta\eta > 2.$



(d) $c_2\{2\}, p_{\mathrm{T}} > 0.5, \Delta\eta > 2.$

Figure 16: DIS Offline + DIS triggers applied to all. LEPTO MC data.

235 5.1.3 HFL trigger rates at high multiplicity

236 The trigger rates of HFL TLTs in ZEUS data, pythia lfjet PhP MC, and Ariadne DIS MC are shown in Figs 17-18.

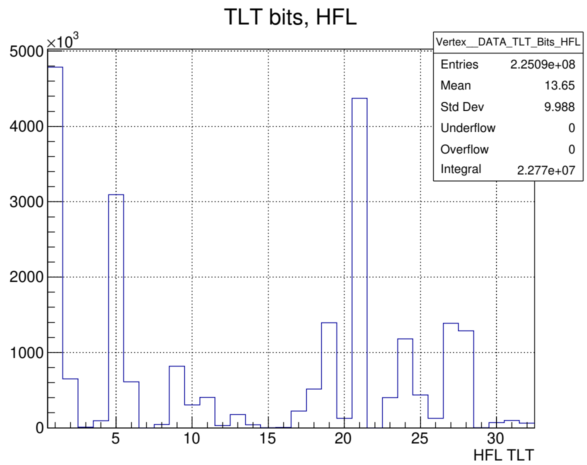
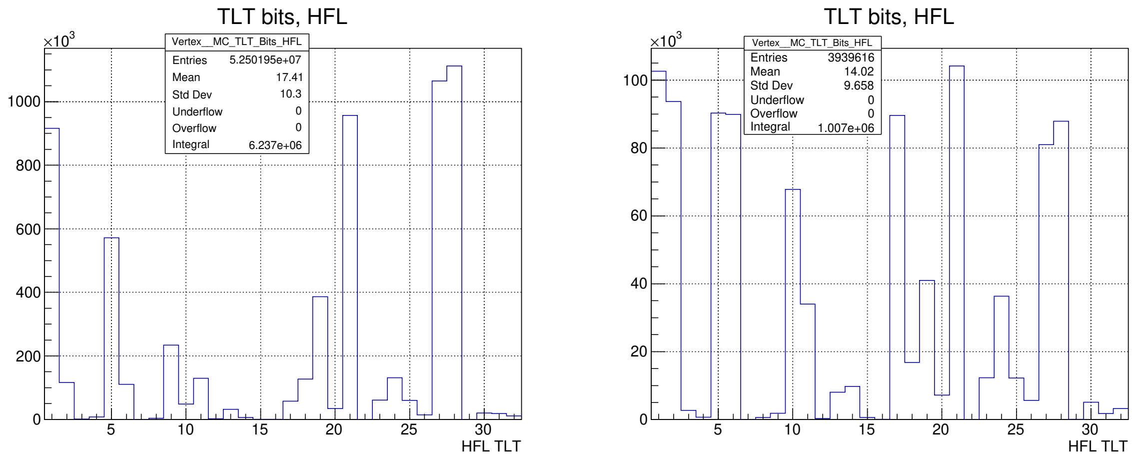


Figure 17: HFL TLT rates for $N_{\mathrm{rec}} > 20$ in ZEUS data, 03p to 07p. Fraction of events which fired HFL 28 $\approx 5\%$

237



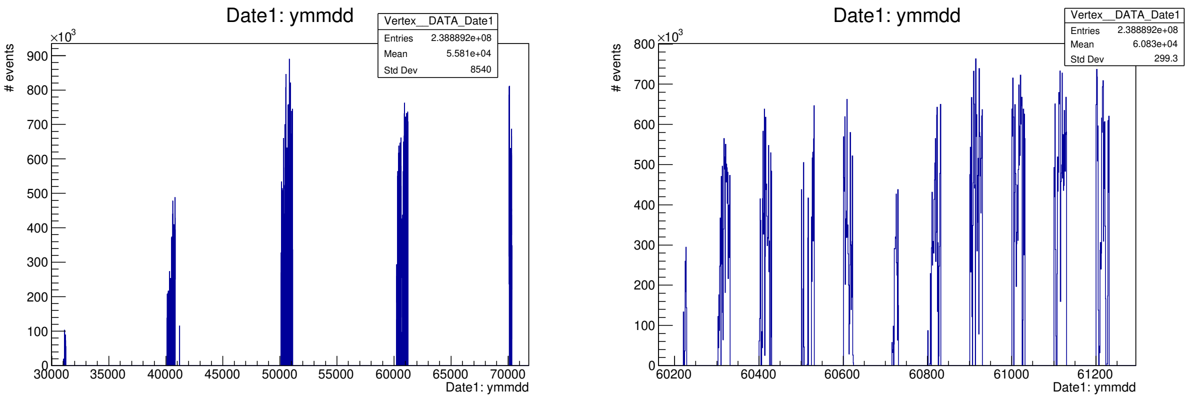
(a) PhP Pythia lfjet 06e. Fraction of events which fired HFL 28 $\approx 18\%$.

(b) DIS Ariadne 0607p. Fraction of events which fired HFL 28 $\approx 9\%$.

Figure 18: HFL TLT rates for $N_{\mathrm{rec}} > 20$ in PhP and DIS MC.

5.1.4 Fraction of events with HFL 28

The least biased trigger at high multiplicity was found to be HFL 28. This trigger only existed starting on May 30th 2006. Figure 19 shows the number of events collected by ZEUS versus date. The fraction of all events containing HFL 28 is 42%. The fraction of



(a) 2003 - 2007 **(b)** 2006

Figure 19: Number of events collected versus date. Date is in the formate of ymmdd: y=1 digit year, mm=2 digit month, dd=2 digit day. HFL 28 present from May 30th 2006 on (60530). In 2006, electron beams were used until the end of June and positron beams afterwards.

events from 2006e containing HFL 28 is 28%.

5.2 Trigger selection for DIS and PhP analyses

DIS trigger name	Description	orange implementation
DIS01	box cut 12x14, E-pz>30, Ee>4, prescaled 10	((ibits(Tltw[3], 16 + 0, 1) > 0) && (ibits(Sltw[5], 7 - 1, 1) > 0))
DIS02	box cut 12x?, E-pz>20, Ee>4, prescaled 20	(ibits(Tltw[3], 16 + 1, 1) > 0)
DIS03	radius cut 25 cm, E-pz>30, Ee>4	((ibits(Tltw[3], 16 + 2, 1) > 0) && (ibits(Sltw[5], 7 - 1, 1) > 0))
DIS04		((ibits(Tltw[3], 16 + 3, 1) > 0) && (ibits(Sltw[5], 7 - 1, 1) > 0))
DIS05		((ibits(Tltw[3], 16 + 4, 1) > 0) && (ibits(Sltw[5], 7 - 1, 1) > 0))
DIS06		(ibits(Tltw[3], 16 + 5, 1) > 0)
DIS11		((ibits(Tltw[3], 16 + 10, 1) > 0) && (ibits(Sltw[5], 7 - 1, 1) > 0))
SPP01		(ibits(Tltw[2], 16 + 0, 1) > 0)
SPP02		(ibits(Tltw[2], 16 + 1, 1) > 0)
SPP03		(ibits(Tltw[2], 16 + 2, 1) > 0)
SPP09		(ibits(Tltw[2], 16 + 8, 1) > 0)
HFL17		(ibits(Tltw[13], 16 + 0, 1) > 0)
HPP31		(ibits(Tltw[11], 16 + 14, 1) > 0)

Table 4: *DIS triggers used in this analysis. An "or" of these triggers is used.*

PhP oriented trigger name	Description	orange implementation
HFL01	Charmed hadrons in PhP, Or of all HFM triggers	ibits(Tltw[9], 16 + 0, 1) > 0
HFL05	inclusive dijets	ibits(Tltw[9], 16 + 4, 1) > 0
HFL21	MESON + jets	ibits(Tltw[13], 16 + 20, 1) > 0
HFL28	MVD inclusive trigger	ibits(Tltw[13], 16 + 27, 1) > 0

Table 5: *PhP oriented triggers used in this analysis. An "or" of these triggers is used.*

6 Vertex selection

Reconstructed events are required to satisfy the following vertex criteria:

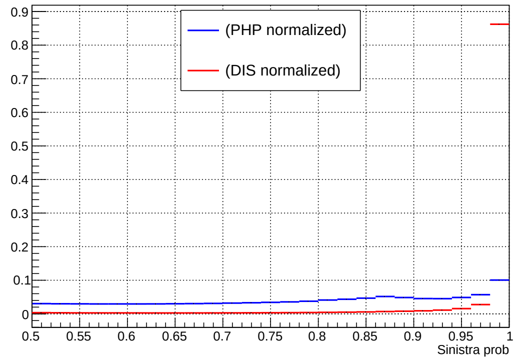
Quantity	orange implementation
$-30 < V_z < 30$ cm	$\mid$ Zvtx $\mid$ < 30
$V_{xy} < 0.5$ cm	((Xvtx - Bspt_x)^2 + (Yvtx - Bspt_y)^2)$^{1/2}$ < 0.5
$N_{\text{vtx tracks}} > 0$	Ntrkvtx > 0
Fraction of tracks associated to primary vertex > 0.1	Ntrkvtx / Trk_ntracks > 0.1
(Event vertex χ^2) / (N_{vtx} tracks) < 50	Chivtx / Ntrkvtx < 50

Table 6: *Primary vertex selection criteria. Same as that used in past DIS analysis of two-particle azimuthal correlations.*

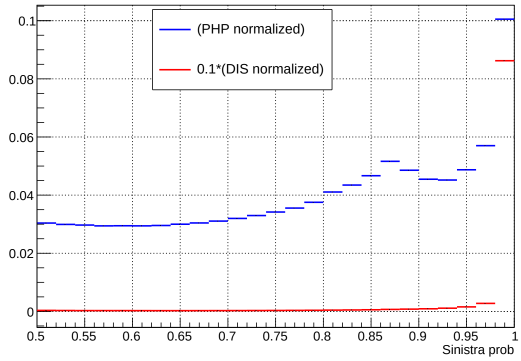
6.1 Offline event selection

For the DIS analysis, the offline event cuts are the same as in the previous analysis on this subject. They are listed in Sec. 6.1.1. For the PhP analysis, the offline event cuts are chosen based on the comparisons of PhP and DIS MC for the distributions in Fig. 20. Generally, we seek to allow as many PhP events as possible while rejecting as many DIS events as possible. It is clear that the sinistra probability distribution shows the clearest distinction since DIS events are highly concentrated near 1.0. To make sure that the PhP event selection is orthogonal to the DIS one, we choose to place the sinistra probability cut at 0.9. The PhP offline cuts are listed in Sec. 6.1.2.

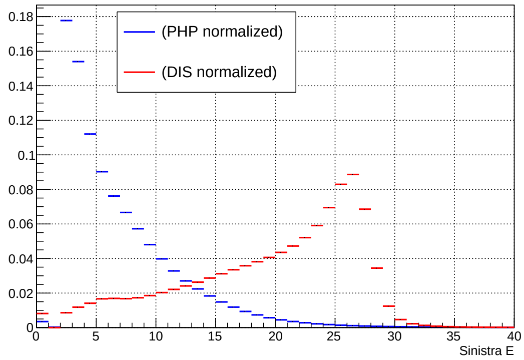
The main analysis is performed at high multiplicity, which we define as $N_{\mathbf{ch}} \geq 20$, due to an unavoidably large trigger bias to inclusive PhP at low multiplicity. All figures correspond to high multiplicity unless otherwise indicated.



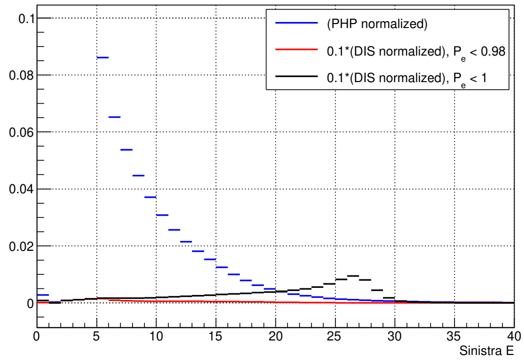
(a) sinistra prob.



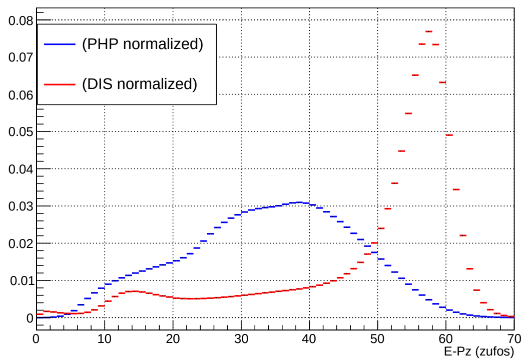
(b) sinistra prob. DIS scaled down by 0.1.



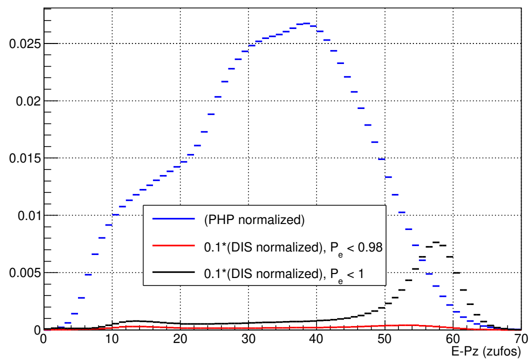
(c) sinistra E.



(d) sinistra E. DIS scaled down by 0.1.



(e) E - Pz.



(f) E - Pz. DIS scaled down by 0.1.

Figure 20: Comparisons of the main offline event quantities for PHP/DIS event selection.

6.1.1 DIS offline selection

Event cut and corresponding orange variable	Purpose
Scattered electron candidates > 0 Sincand	NC DIS must have a scattered electron
Scattered electron energy > 10 GeV Siecorr[0][2]	Ensures good electron identification
scattered electron probability > 0.9 Siprob[0]	Ensures high DIS likelihood from the sinistra neural network algorithm
$\theta_e > 1.0$ Sith[0]	electron id algorithm does not work in forward direction
$Q^2 > 5$ GeV/c Siq2el[0]	Lowest value that can be reasonably reconstructed in HERA II
$E - p_z$: $47 < \sum E_i(1 - \cos\theta_i) < 69$ GeV V_h_e_zu - V_h_pz_zu	E is the total energy summed over every cell in the CAL. p_z represents the summed z-component. If the event is well measured, this should be given by twice the electron beam energy ($2 \times 27.6 = 55.2$ GeV) due to energy and momentum conservation.
scattered electron radial location at entrance into CAL > 15 cm ($\theta_e < 174° = 180 - \tan^{-1}(15/141)$) ((Sipos[0][0]+1)^2 + (Sipos[0][1])^2)$^{1/2}$	Trigger inefficiencies become large at small radii
Remove scattered electrons which entered calorimeter near the chimney (path for cables to leave the inner region of ZEUS) Chimney region given by $-10 < x < 10$ && $y > 110$ && $z < -141$ Sipos[0][0], Sipos[0][1], Sipos[0][2]	scattered electron information can be distorted due to extra material in chimney
HES fiducial/CAL crack cut in RCAL: unwanted region is $5 < x < 11$ && $y > 0$ && $z < -141$ and $-15 < x < -9$ && $y < 0$ && $z < -141$ Sipos[0][0], Sipos[0][1], Sipos[0][2]	Remove regions in Hadron-Electron-Separator which are ineffective or untrustworthy.
0.2 GeV minimum transverse energy in FCAL within 10° cone when there are 0 tracks Cal_et10, Trk_ntracks	Remove off-momentum positron background and diffractive single e events.
For run numbers (Runnr) greater than 44000, Mvdtake > 0	Require an acceptable QA for the MVD.

Table 7: *DIS offline event selection*

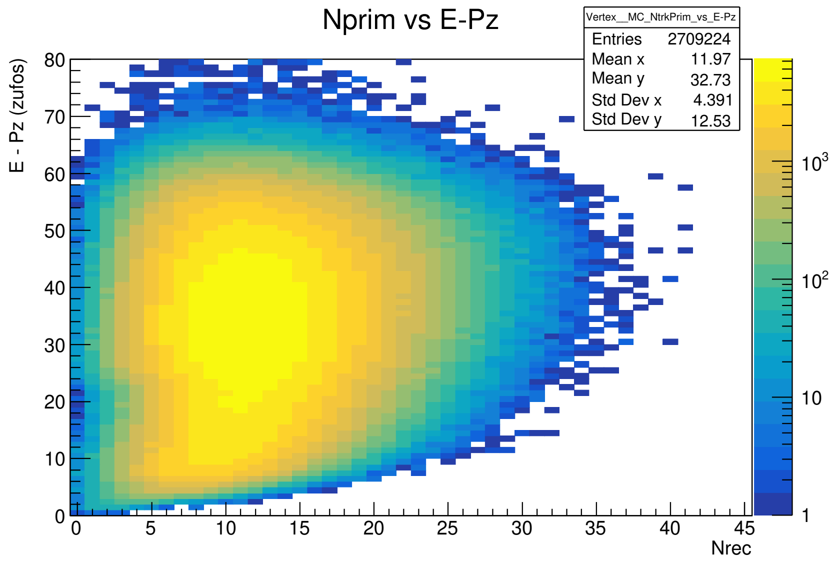


Figure 21: $E - P_z$ versus $N_{\rm rec}$ in pythia light-flavor jet PhP.

6.1.2 PhP offline selection

Event cut and corresponding orange variable	Purpose
scattered electron probability < 0.9 Siprob[0]	Ensures low DIS likelihood from the sinistra neural network algorithm
Scattered electron energy < 15 GeV Siecorr[0][2]	There should be no high energy electron in PhP
$E - p_z$: $\sum E_i(1 - \cos\theta_i) < 55$ GeV V_h_e_zu - V_h_pz_zu	PhP events should be void of a high energy scattered electron. We keep this cut higher than most other PhP analyses due to the interesting high-multiplicity inelastic PhP events at high $E - p_z$ (Fig. 21)
For run numbers (Runnr) greater than 44000, Mvdtake > 0	Require an acceptable QA for the MVD.

Table 8: *PhP offline event selection*

In the PhP part of this analysis, our choice of the $E - P_z$ cut is larger than most other ZEUS analyses. In DIS part of this analysis, we restrict ourselves to $E - P_z < 47$ GeV. Figure 21 shows $E - P_z$ versus $N_{\rm rec}$, which demonstrates that high $E - P_z$ events also contain interesting high multiplicity events. As seen in Fig. 22, high $E - P_z$ also correspond to high inelasticity events.

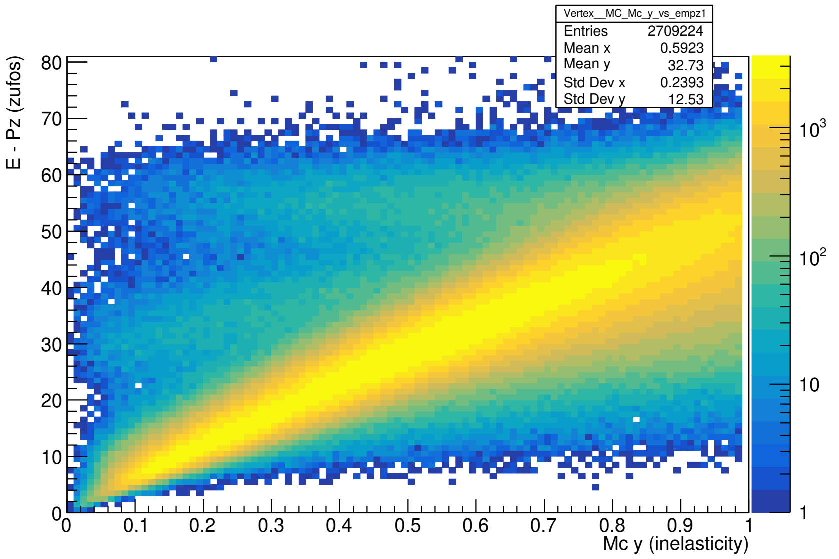


Figure 22: $E - P_z$ versus inelasticity y in pythia light-flavor jet PhP. Highly inelastic events occur when the electron deposits most of its energy into the system.

7 Settings for generating PYTHIA MC

To form a reference to compare our ZEUS data to, we generate MC events using the PYTHIA 8.303 generator. The main "knob" of interest in this analysis is the MPI strength given by pT0Ref. The settings we used are shown in Tab. 9.

Setting type	choice(s)
Beams:frameType	2 (CM)
Beams:eA	920 GeV
Beams:eB	27.52 GeV
Beams:idA	2212
Beams:idB	11
PDF:lepton2gamma	on (enables photon sub-beam from lepton)
Photon:Q2max	1.0 GeV^2
Photon:Wmin	10 GeV
Photon:ProcessType	0 (auto mix of resolved and direct)
PhotonParton:all	on or off (direct component)
SoftQCD:nonDiffractive	on or off (resolved component)
PartonLevel:MPI	on or off (master switch for MPI)
MultipartonInteractions:pT0Ref	2.5, 3.0, 3.5 4.5 GeV(MPI strength)
ColourReconnection:range	0.0, 1.8 (range in p-space of CR)

Table 9: *PYTHIA 8.303 settings used to generate PhP MC data.*

At the generated particle level, similar to the light-flavor jet MC sample, particles are retained if they satisfy the requirements in Tab. 8.1. About 10 M events with $N_{\mathrm{ch}} \geq 20$ are generated for each variation.

Jets were also reconstructed in order to assess the intrinsic jet bias in the light-flavor jet sample. We used the jet finding algorithms provided by pythia. Concretely: SlowJet(power=1, R=1, pTMin=3, etaMax=3, select=2, massSet=2, 0, true). Which corresponds to: kT algorithm, $\Delta R < 1$, $p_{\mathrm{T}_{min}} > 3$, $|\eta| < 3$, final-state massive particles are used, fastJet Core.

7.1 Number of MPI in PYTHIA

The number of distinct $2 \rightarrow 2$ initial parton scatterings (nMPI) in pythia can be counted event-by-event and is inversely related to the pT0Ref parameter. Figure 23 shows the distributions of nMPI for several choices of pT0Ref. **The mean values in Fig. 23, ⟨nMPI⟩, together with comparisons to ZEUS data in Sec. 14, forms our estimate of the average number of MPI in PhP at high multiplicity.**

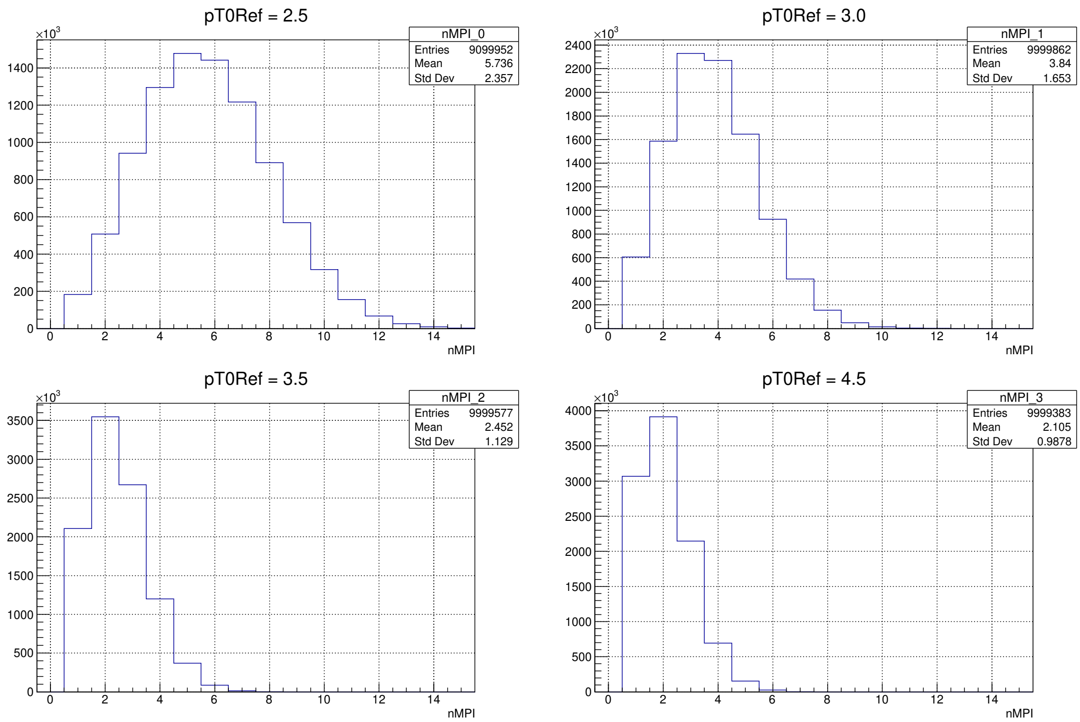


Figure 23: Distribution of the number of MPI (nMPI) in *ep* PhP PYTHIA for different settings of pT0Ref.

8 Track selection

8.1 MC primary particle definition and selection

Generator level primary particles are selected using information about their particle ID, their origin, and their kinematics (p_T and η). Our definition of primary particles will follow the ALICE definition `https://cds.cern.ch/record/2270008`. Primary particles are those with a mean proper lifetime $\tau > 1$ cm/c which was produced directly in the collision or from a decay of a particle with $\tau < 1$ cm/c. A table of particles with $\tau > 1$ cm/c is shown in Fig 24. Of the particles in the table, only charged hadrons are selected

Specie	Width Γ (GeV)	Mean proper lifetime τ (ps)	(cm/c)
p^+	0	∞	∞
γ	0	∞	∞
K^0	0		
e^-	0	∞	∞
n	7.478×10^{-28}	$8.861 \times 10^{+14}$	$2.656 \times 10^{+13}$
μ^-	2.996×10^{-19}	$2.212 \times 10^{+06}$	$6.63 \times 10^{+04}$
K^0_L	1.287×10^{-17}	$5.148 \times 10^{+04}$	1543
π^+	2.528×10^{-17}	$2.621 \times 10^{+04}$	785.7
K^+	5.317×10^{-17}	$1.246 \times 10^{+04}$	373.6
Ξ^0	2.27×10^{-15}	291.9	8.751
Λ	2.501×10^{-15}	264.9	7.943
Ξ^-	4.02×10^{-15}	164.8	4.941
Σ^-	4.45×10^{-15}	148.9	4.464
K^0_S	7.351×10^{-15}	90.14	2.702
Ω^-	8.071×10^{-15}	82.1	2.461
Σ^+	8.209×10^{-15}	80.72	2.42

Figure 24: List of primary particles.

in our analysis. We use the generator information about a particle's mother to reject weak decays of long lived particles. Consequently, an $\Omega^{\pm}$ weak decay will be counted as a primary at the truth level (denominator of efficiencies) but since it decays ~ 2.5 cm from the primary vertex, its daughters will typically not pass our track selection criteria in the reconstructed pool. Finally, we must reject particles which arose from interactions with the ZEUS detector material (UBUF(1) $<$ 0). These are identified with the help of the ZEUS ISTHEP number provided in the MC data. The ISTHEP number is explained here `http://adamo.web.cern.ch/Adamo/zeusddl/FMCZEvt.html` and can be decomposed as follows:

$$\begin{aligned} \text{FMCKIN\_ISTHEP} &= \text{Tn\_ISTHEP} \times 10^4 + \text{Generator\_ISTHEP} \quad (8) \\ \text{Tn\_ISTHEP} &= \text{Int(UBUF(2))} \times 10^3 \quad (9) \\ &+ \text{Int(UBUF(1))} + 500 \quad (10) \end{aligned}$$

The codes to be extracted from the FMCKIN_ISTHEP are:

- Generator_ISTHEP: 1 for final state particle given by generator. 0 otherwise.
- UBUF(1) is the parent's MC track index. It is negative for products of interaction or showers within the detector material.
- UBUF(2) gives information on a particles demise. See the above url for a list of demise possibilities.

The selection of generated primary particles is summarized in Tab 8.1

Quantity	orange implementation
primary particle selection	Fmck_isthep != 2,3,11. $c\tau < 1$ cm (mcprt being one of the charged hadrons in in Fig 24). Not from decay of long-lived particle (Fmck_id, Fmck_daug). UBUF1 >= 0 (Fmck_isthep).
Scattered electron rejection (DIS only)	Fmck_id != Idlepton
$0.1 < p_{\mathrm{T}} < 5$ GeV	(Fmck_px^2 + Fmck_py^2)$^{1/2}$
$-1.5 < \eta < 2$	calculated from Fmck_px, Fmck_py, Fmck_pz

Table 10: *Generator level particle selection*

The total number of generated primary particles is defined as N_{gen}.

8.2 Reconstructed track selection

We select reconstructed tracks as described in Tab 8.2:

Quantity	orange implementation
ZTT tracking type	orange.Trk_type = 3
At least 1 hit point in the MVD barrel + wheels	(Trk_nbr + Trk_nbz + Trk_nwu + Trk_nwv) > 0
DCA_xy < 2.0 cm	Trk_imppar
DCA_z < 2.0 cm	$\vert$Trk_pca[index][2] - Zvtx$\vert$
Reject scattered electron candidate (DIS only)	Trk_id != Sitrknr[0]
Reject tracks which are within a cone radius of 0.4 around the scattered electron candidate: $\Delta R = \sqrt{\Delta\varphi^2 + \Delta\eta^2} > 0.4$ (DIS only)	φ η calculated from Trk_px, Trk_py, Trk_pz. Sinistra φ and η from Sitrkph[0] and Sitrkth[0].
$0.1 < p_\mathrm{T} < 5$ GeV	(Trk_px^2 + Trk_py^2)$^{1/2}$
$-1.5 < \eta < 2.0$	calculated from Trk_px, Trk_py, Trk_pz

Table 11: *Reconstructed track selection*

9 Control figures

To justify the extraction of efficiency corrections from the light-flavor jet PhP MC and their application to ZEUS PhP data, we compare basic reconstructed event and track distributions in data and MC. These are shown in Figs. 25 and 26. Uncorrected correlation function comparisons are shown in Fig. 27.

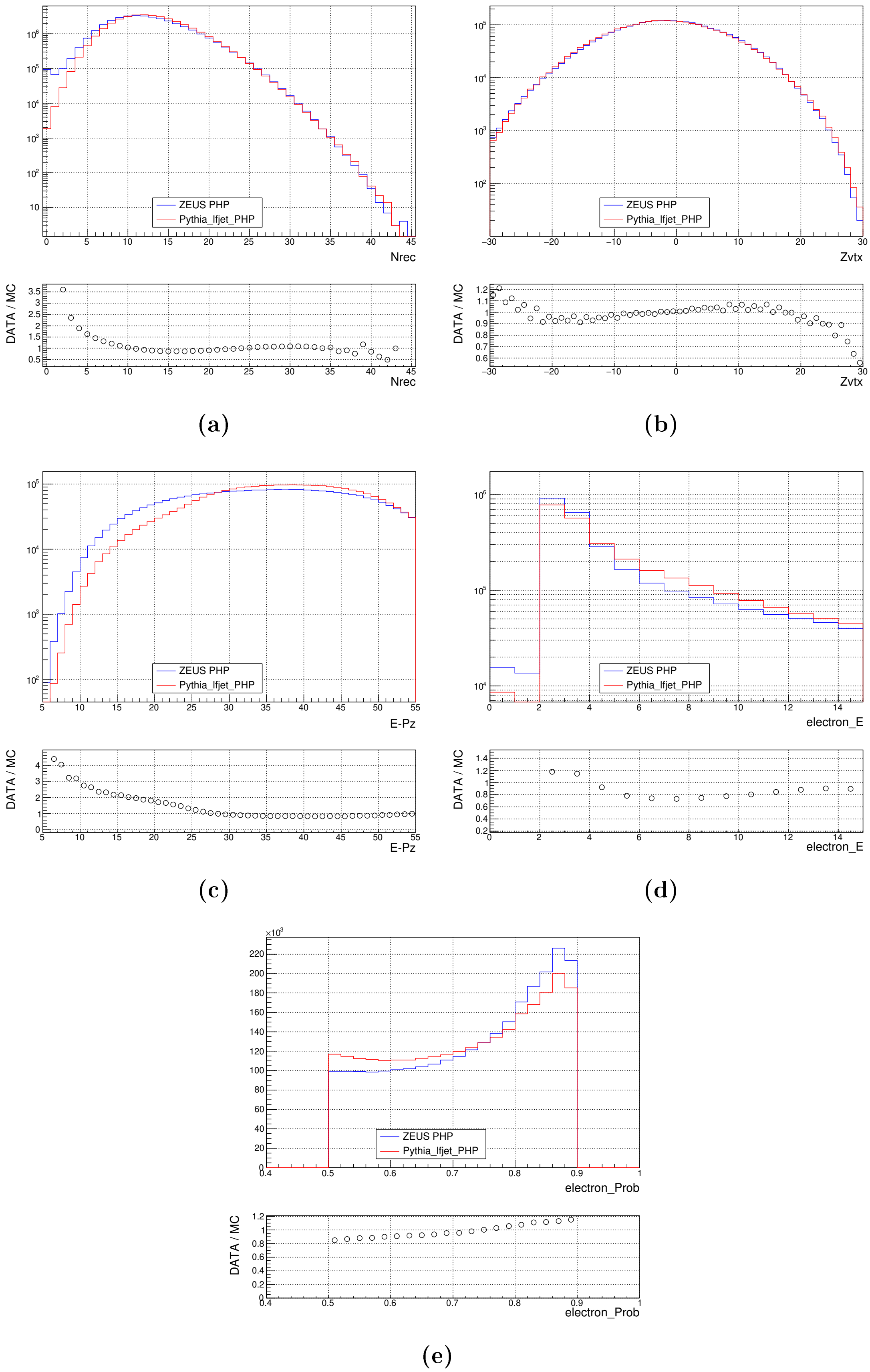


Figure 25: Event control figures. Comparison of raw reconstructed quantities in PhP data to light-flavor jet MC. All default trigger and offline cuts applied. $N_{\rm rec} \geq 20$. MC distributions normalized to that in data.

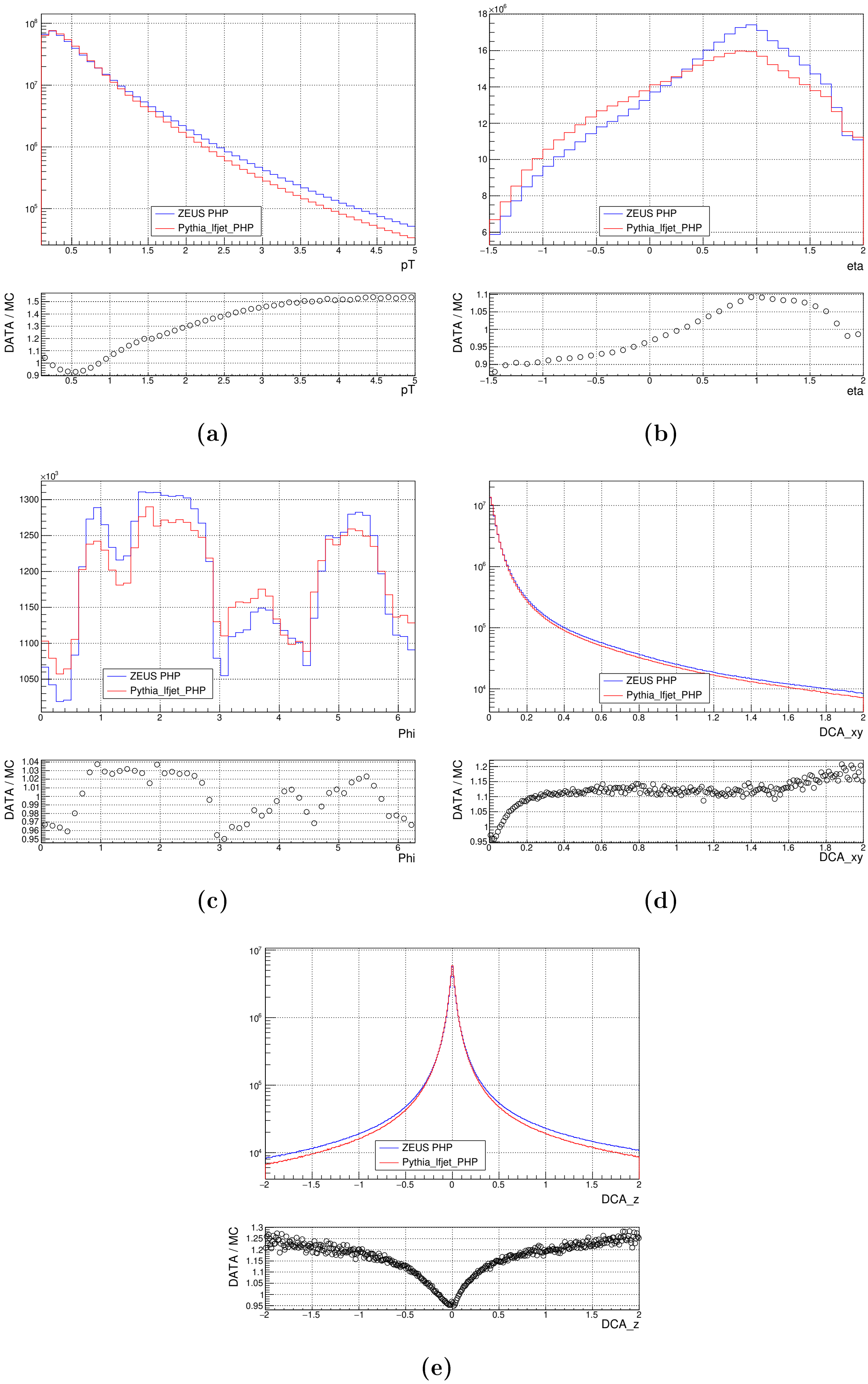


Figure 26: Track control figures. Comparison of raw reconstructed quantities in PhP data to light-flavor jet MC. All default trigger and offline cuts applied. $N_{\text{rec}} \geq 20$. MC distributions normalized to that in data.

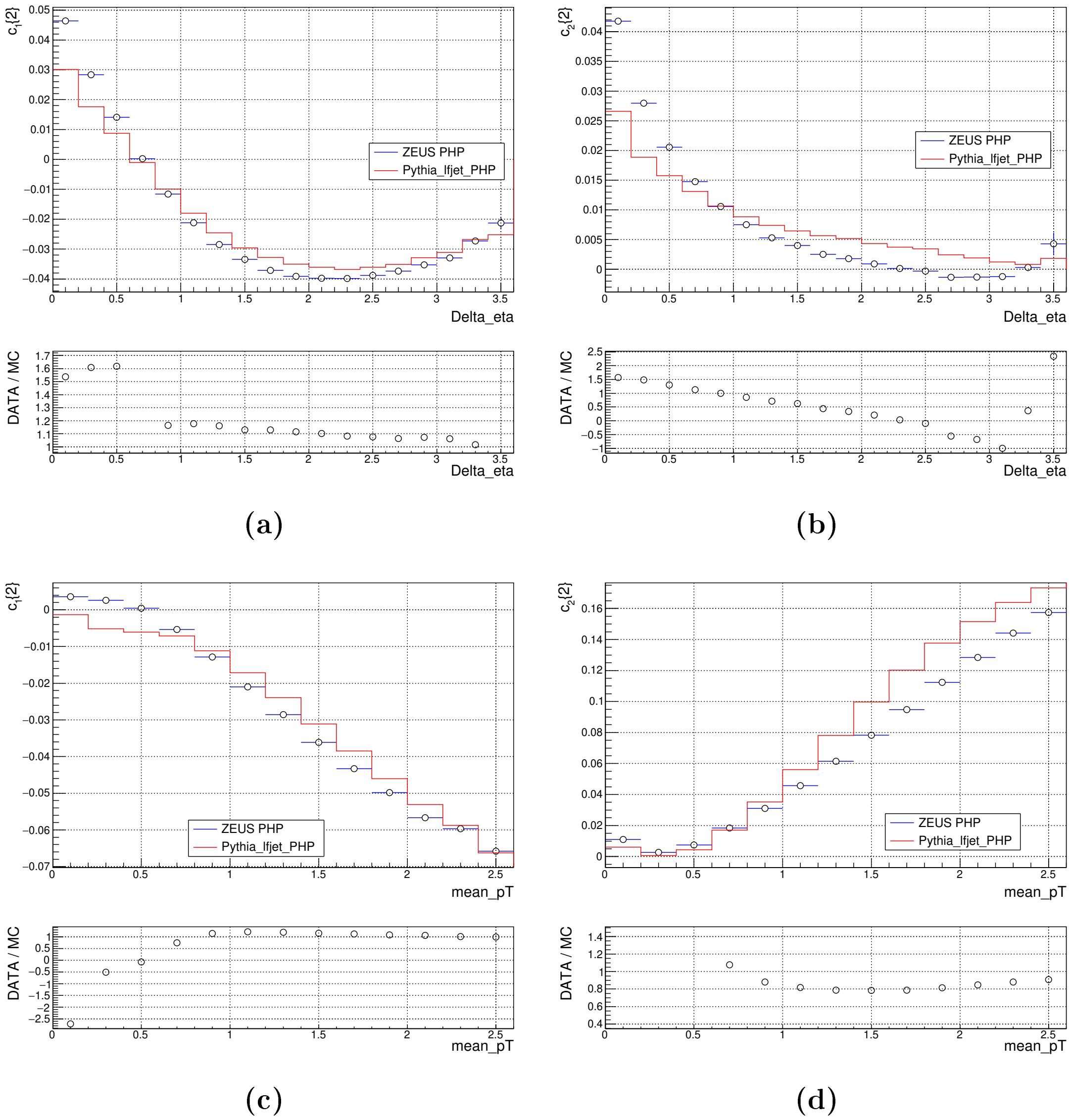


Figure 27: Correlation function control figures. Comparison of raw reconstructed quantities in PhP data to light-flavor jet MC. All default trigger and offline cuts applied. $N_{\rm rec} \geq 20$.

10 DIS contamination studies

To estimate the DIS contamination to our PhP sample, we consider four factors. The first is the ratio of DIS to PhP cross-sections in *ep* collisions, which from other studies is known to be about 10%. However, this is a multiplicity integrated estimate and the value at high multiplicity is likely different. To estimate a correction at high multiplicity we compare the normalized N_{gen} distributions in DIS and PhP MC and take the ratio of all events with $N_{\mathrm{gen}} \geq 20$. From Fig. 28 we calculate the ratio to be 0.09. The second factor is the ratio

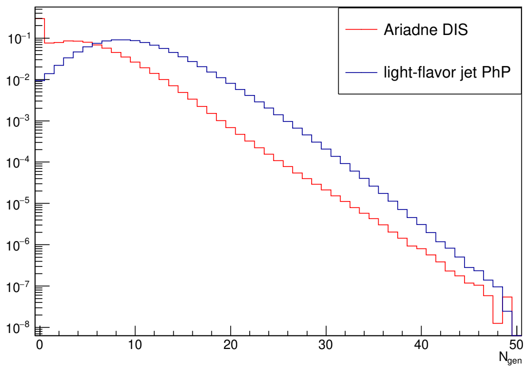


Figure 28: Comparison of Ariadne DIS and light-flavor jet PhP N_{gen} distributions. Each is normalized to unity.

of our cocktail trigger efficiencies in DIS to PhP. This is estimated using DIS and PhP MC by counting the faction of generated events at high multiplicity, $N_{\mathrm{gen}} \geq 20$, surviving the cocktail trigger. It is about 88% in DIS MC and 50% in PhP MC. We cannot use N_{rec} for this since reconstructed data was not stored in the light-flavor jet MC when no triggers were fired. Finally, the third factor using Figs. 29 and 30. PhP offline cuts in DIS MC leave about 6% of the total DIS population as seen in the projections against N_{rec}, p_{T}, and η in Fig. 29. The PhP offline cuts in PhP MC leave about 65% of the PhP population as seen in the same projections in Fig. 30. Multiplying all three factors gives us our estimate of the DIS contamination to the PhP analysis:

$$
\begin{aligned}
&[\text{relative\_cross\_sec}] \times [\text{trigger\_efficiencies}] \times [\text{offline\_efficiencies}] \\
&= [0.10 \times 0.09] \times [\tfrac{0.88}{0.50}] \times [\tfrac{0.06}{0.65}] = 0.0015
\end{aligned}
\tag{11}
$$

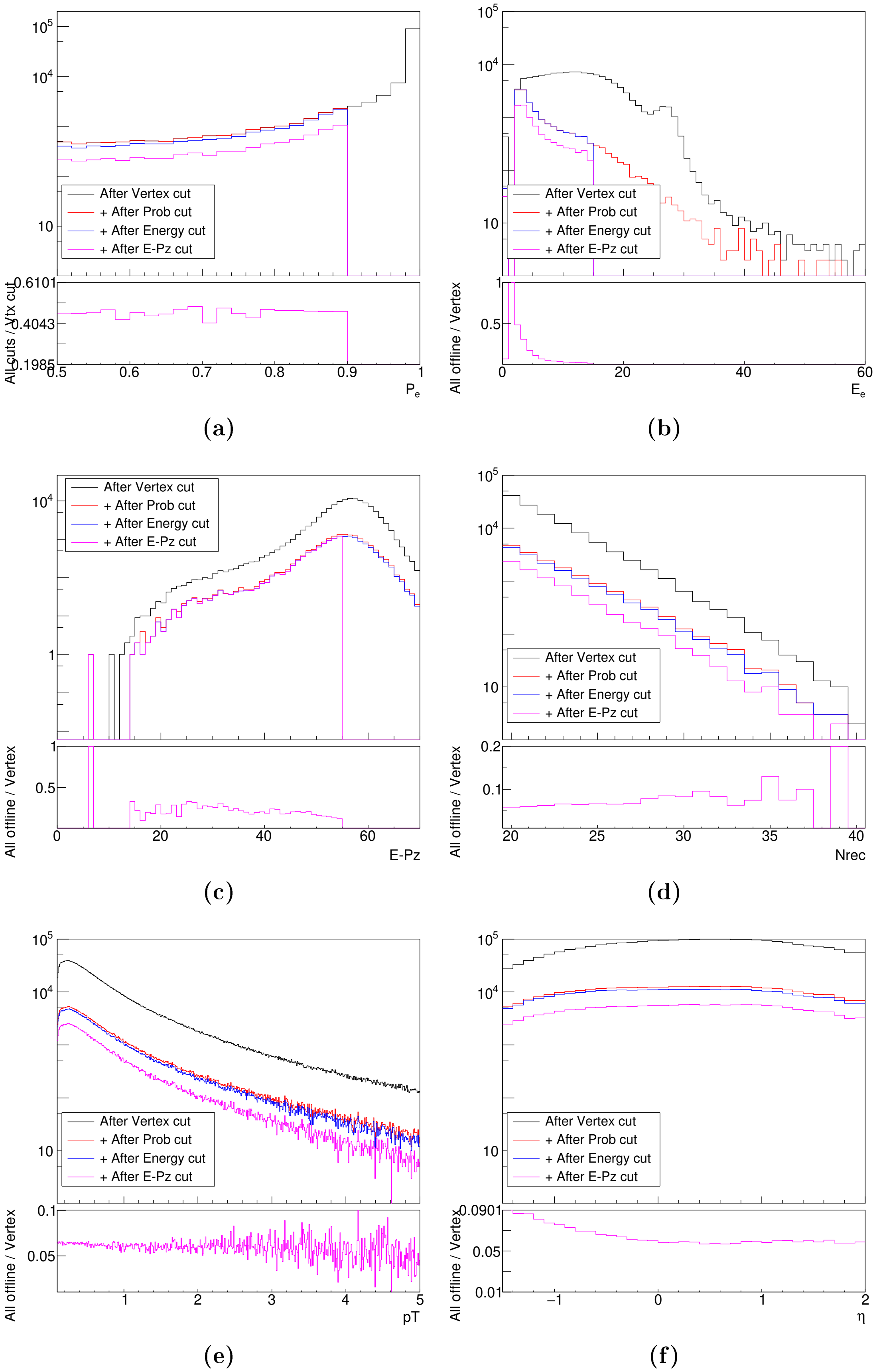


Figure 29: DIS Ariadne distributions after sequential PhP offline cuts. $N_{\text{rec}} > 20$. The lower sub-panels represent the ratio of vertex + all offline cuts to vertex only cuts.

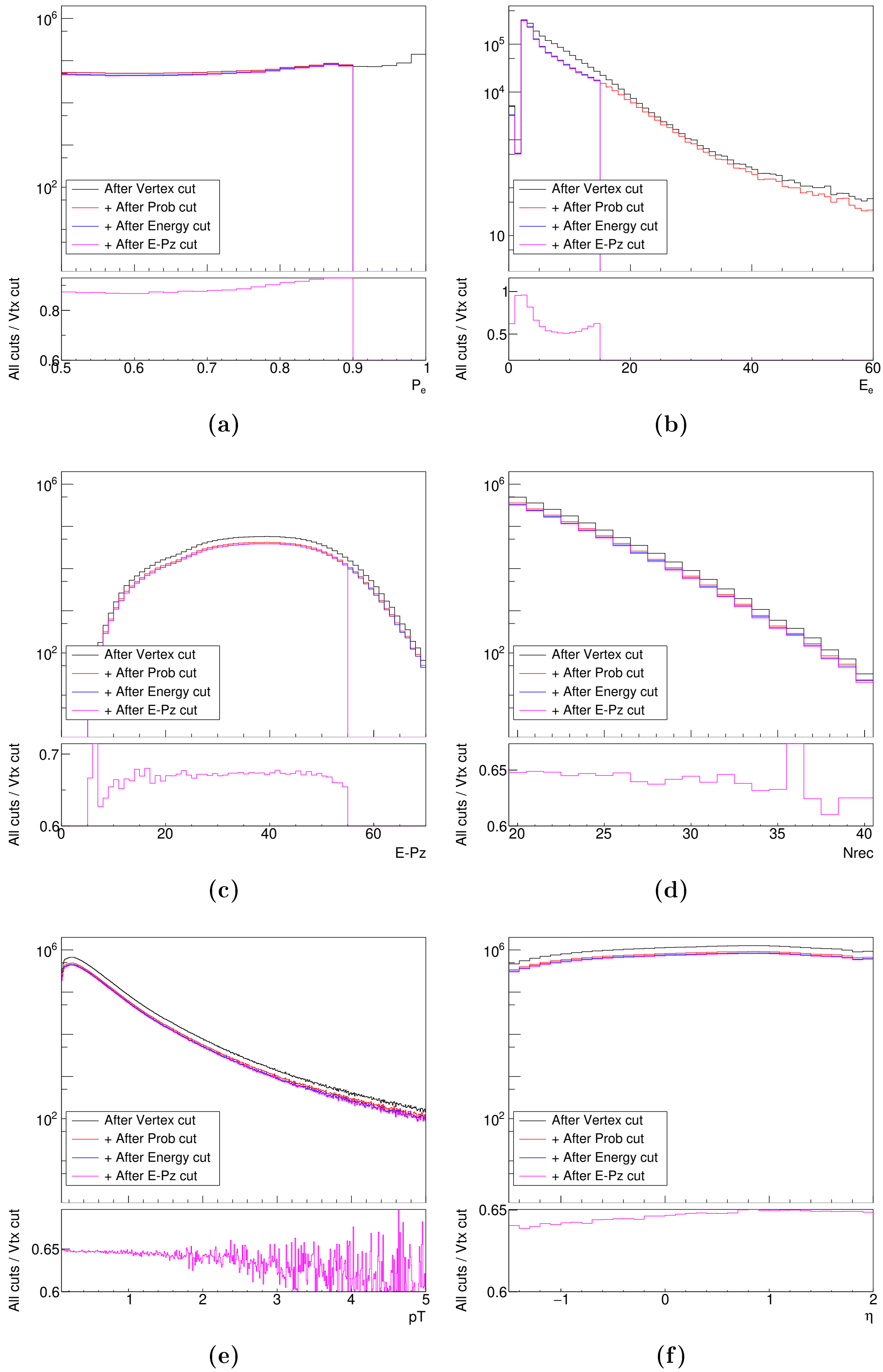


Figure 30: PhP lfjet distributions after sequential PHP offline cuts. $N_{\text{rec}} > 20$. The lower sub-panels represent the ratio of vertex + all offline cuts to vertex only cuts.

11 Direct versus Resolved PhP

Shown in Fig .31 is a comparison of N_{ch} distribution for the direct and resolved components of PhP. We see that at high multiplicity, the resolved component clearly dominates the

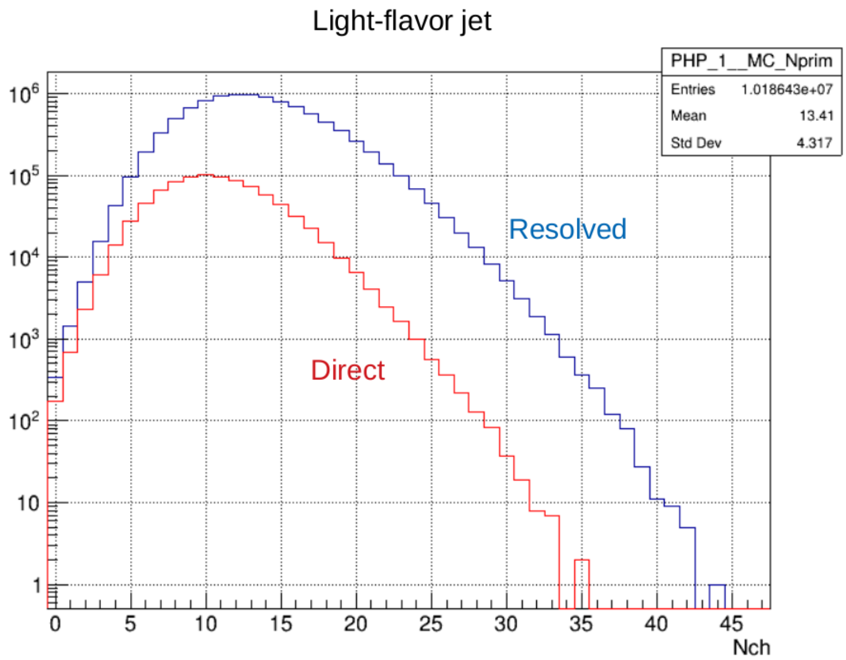


Figure 31: Comparison of N_{ch} distributions in resolved and direct PhP (light-flavor jet MC). All trigger and offline cuts applied.

spectrum.

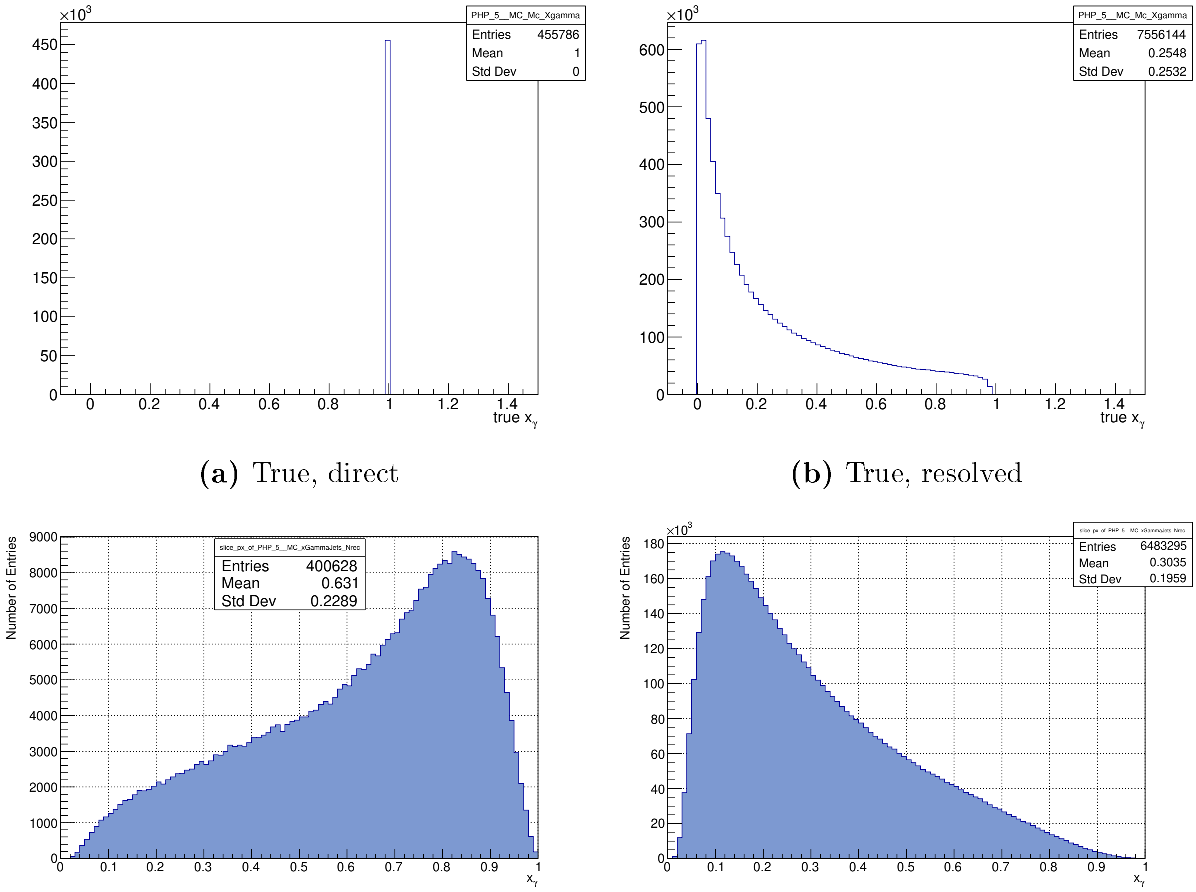
11.1 x_γ

The x_γ variable can be used to help distinguish direct from resolved PhP events. That is, events where the exchange photon struck a quark in the proton and was fully absorbed (direct) versus events where the photon fluctuates into hadronic matter, part of which may scatter off of the proton (resolved). It is 1.0 for direct events and typically much smaller in resolved events. We define x_γ in the following way:

$$\begin{aligned} x_\gamma &= \frac{E^{jet1} + E^{jet2} - p_Z^{jet1}, -p_Z^{jet2}}{E^{all} - p_Z^{all}}, \\ E &= E_T \cosh\eta, \\ p &= \sqrt{E^2 - m^2}, \\ p_Z &= p\tanh\eta, \end{aligned} \tag{12}$$

using the two leading jets of the event. The orange block “ktJETS_A" is used, which corresponds to: “Zufos, without removal of electron candidate, are used as input for jet algorithm. Massive jets are reconstructed with E-scheme and in inclusive mode. Dead material corrections are applied. Information about 10 jets in laboratory frame is saved. Jets are required to have transverse energy greater than 2.5 GeV and pseudorapidity in range from -2.5 to 2.5."

Figure 32 demonstrates how the true x_γ in MC is smeared out after reconstruction from the two leading jets.



(a) True, direct

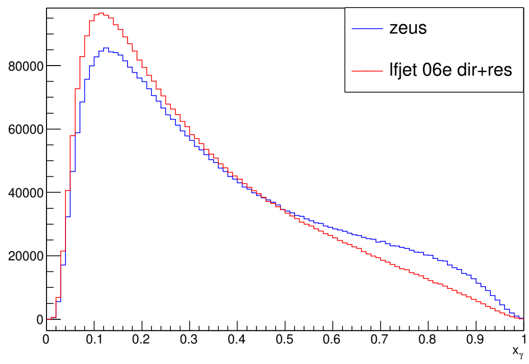
(b) True, resolved

(c) Reconstructed, direct

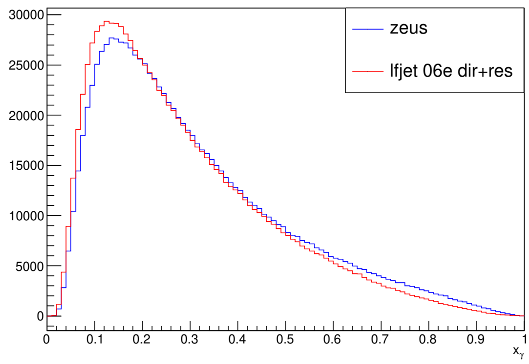
(d) Reconstructed, resolved

Figure 32: True and reconstructed x_γ.

A comparison of x_γ in PHP data and lfjet MC at mid and high multiplicity is shown in Fig. 33. At high multiplicity, lfjet MC matches the data reasonably well. Figure 33 also demonstrates that the resolved component of photoproduction dominates at high multiplicities where we perform our analysis.



(a) $10 < N_{\mathrm{rec}} < 20$.



(b) $20 < N_{\mathrm{rec}} < 50$.

Figure 33: x_γ in PhP MC and ZEUS data.

12 MC studies

12.1 Efficiency corrections

12.1.1 Single-, Pair-, and Quadruplet-weights

Tracking inefficiencies cause distortions to the distributions measured in this analysis. We extract efficiency corrections from light-flavor jet MC for the PhP analysis and from Ariadne/Lepto for the DIS analysis.

The efficiency correction factors (weights) for 1-, 2-, and 4-particle distributions are generically defined as:

$$w^{(n)}(\vec{x}) = \frac{N^n_{gen}(\vec{x})}{N^n_{rec}(\vec{x})}, \tag{13}$$

which is a ratio of generated to reconstructed distributions of order n and calculated differently in $\vec{x}$. The dimensions of $\vec{x}$ are described in Tab. 12.

dimension of $\vec{x}$	One-particle (n=1)	Two-particle (n=2)	Four-particle (n=4)
x_1	φ	$\varphi_1 - \varphi_2$	$\varphi_1 + \varphi_2 - \varphi_3 - \varphi_4$
x_2	η	$\langle \eta_i - \langle \eta \rangle \rangle$	$\langle \eta_i - \langle \eta \rangle \rangle$
x_3	p_{T}	$\langle p_{T,i} - \langle p_T \rangle \rangle$	$\langle p_{T,i} - \langle p_T \rangle \rangle$
x_4 (charge)	q	$\lvert q_1 + q_2 \rvert$	$\lvert q_1 + q_2 + q_3 + q_4 \rvert / 2$
x_5	-	N_{rec}	N_{rec}

Table 12: *Dimensions of the efficiency corrections for single-particle weights (n=1), pair-weights (n=2), and quadruplet-weights (n=4). The weights are also calculated separately for each data-taking period.*

The weights are also calculated separately for each data-taking period.

12.1.2 Performance of single-particle efficiency corrections

The performance of the efficiency corrections for the p_{T} and η distributions are shown in Fig. 34

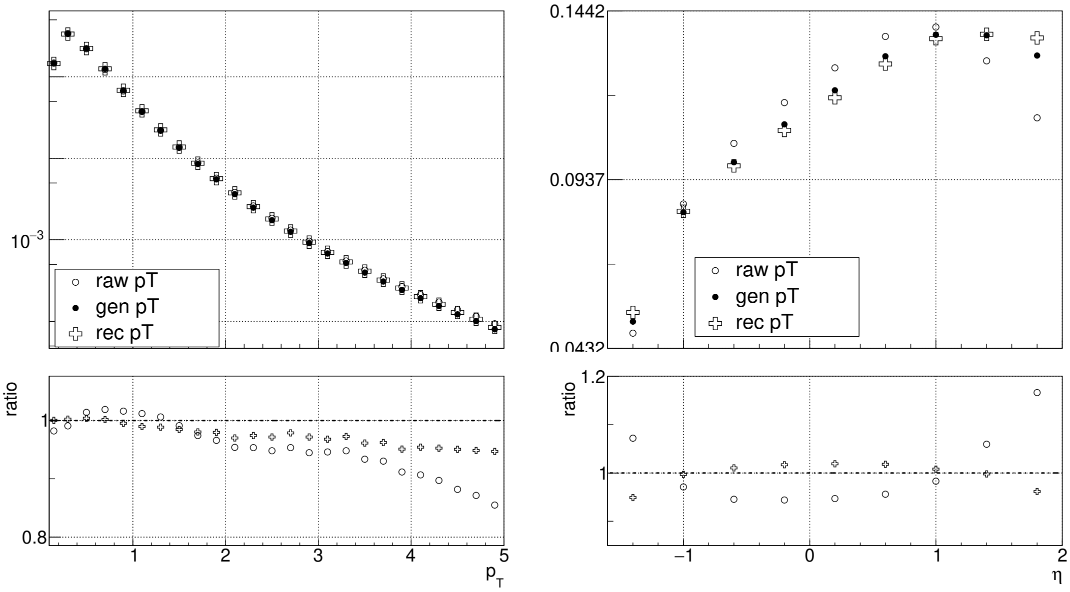


Figure 34: Performance of single-particle efficiency corrections. Residual differences are used as a systematic uncertainty.

12.1.3 N_{ch} estimation

The N_{ch} distribution is extracted using unfolding as explained in Sec. 12.5. Event-by-event, N_{ch} is determined using $w^{(1)}$:

$$N_{\mathrm{ch}} = \sum_{i}^{N_{\mathrm{rec}}} w^{(1)} \tag{14}$$

, and is used to select high multiplicity events for the analysis.

12.1.4 Low p_{T} efficiency overestimation in ZEUS MC

The studies from Vladyslav Libov (DESY-THESIS-2013-030) and Olena Bachynska (DESY-THESIS-2012-045), it is understood that the ZEUS tracking efficiency at low p_{T} is overestimated in the existing Monte-Carlo data. Both authors have provided a correction factor for this effect which is applied as a correction to our extracted efficiencies.

Libov's correction is given by the difference of Eq 5.4 in his thesis from unity:

$$f_{Libov} = \frac{1 - \epsilon\, p_{hadr,MC}}{1 - p_{hadr,MC}} \tag{15}$$

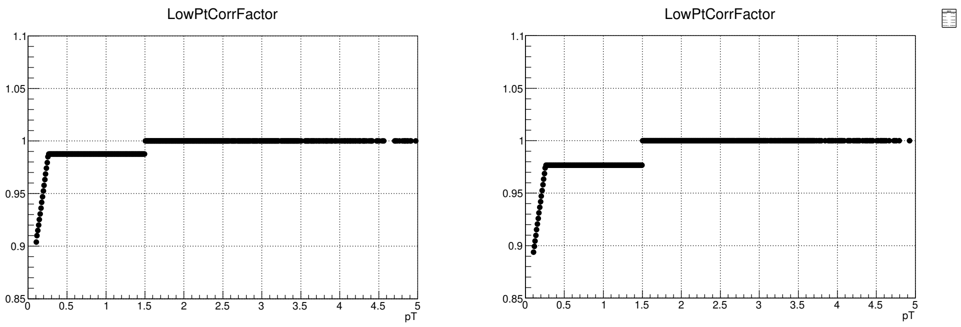
The hadronic interaction probability in MC is given by $p_{hadr,MC}$ and is found to be dependent on the track's θ angle as shown in 5.9(a) in his thesis. $p_{hadr,MC}$ varies between 0.03 at mid-rapidity to 0.07 near $|\eta| = 2$. The ϵ factor is the correction for the overestimation of the hadronic interaction rate in ZEUS MC and is taken to be 1.4 for $p_{\mathrm{T}} < 1.5$ (GeV/c) and unity for higher p_{T}.

Bachynska's correction is given by Eq 7.4 in her thesis:

$$f_{Bachynska} = 1 + 0.548 \times (p_{\mathrm{T}} - 0.26). \tag{16}$$

The equation is expected to be relevant for $p_{\mathrm{T}} < 0.26$ and is only applied there.

Both Libov and Bachynska's correction factors are applied as a weight to “scale-down" the reconstructed particle yield in ZEUS MC (weight $= f_{Libov} \times f_{Bachynska}$). The factors therefore lower the estimated tracking efficiency. The weight factors are used to reject tracks in a probabilistic manner for the published N_{rec} distribution while they are used to weight down the contribution from a track for all other parts of this analysis. These two procedures are equivalent for the correction of $c_n\{2\}$. They affect both the correlation functions as well as the N_{ch} estimation. The product of both factors at mid-rapdity and forward-rapidity is shown in Fig 35.



(a) $0 < \eta < 0.1$: lowest detector material traversed.

(b) $1.9 < \eta < 2$: largest detector material traversed.

Figure 35: $f_{Libov} \times f_{Bachynska}$

12.2 Trigger bias correction

As there is no inclusive PhP trigger(s) available, whatever triggers we choose to employ will introduce some bias to our measurements. To minimize this bias, we chose to focus the analysis on high multiplicity with the understanding that such PhP events are more likely to fire at least one trigger. From the mentioned trigger studies, we found a set of HFL triggers which biased the generator level MC correlations the least. To correct for the remaining bias, we form the ratio:

$$\frac{D_{lfjet}}{D_{lfjet}^{\mathrm{Trigger}}}, \tag{17}$$

$D \in \{N_{\mathrm{ch}}, \frac{dN}{dp_{\mathrm{T}}}, \frac{dN}{d\eta}, c_n\{2\}, c_n\{4\}\}$. The D distributions represent generator level quantities, not reconstructed. Comparisons of generator level and triggered distributions are shown in Fig. 36

12.2.1 Performance of trigger simulation in MC

Here we demonstrate the application of the trigger bias correction on ZEUS data for two choices of triggers: HFL cocktail (default) and HFL 28 alone. After the trigger bias correction, both choices should in principle yield the same result. Remaining discrepancies are treated as a systematic uncertainty and are understood to reflect the quality of the trigger simulation in MC. Figure 37 illustrates this for the distributions which were most biased by the trigger selection.

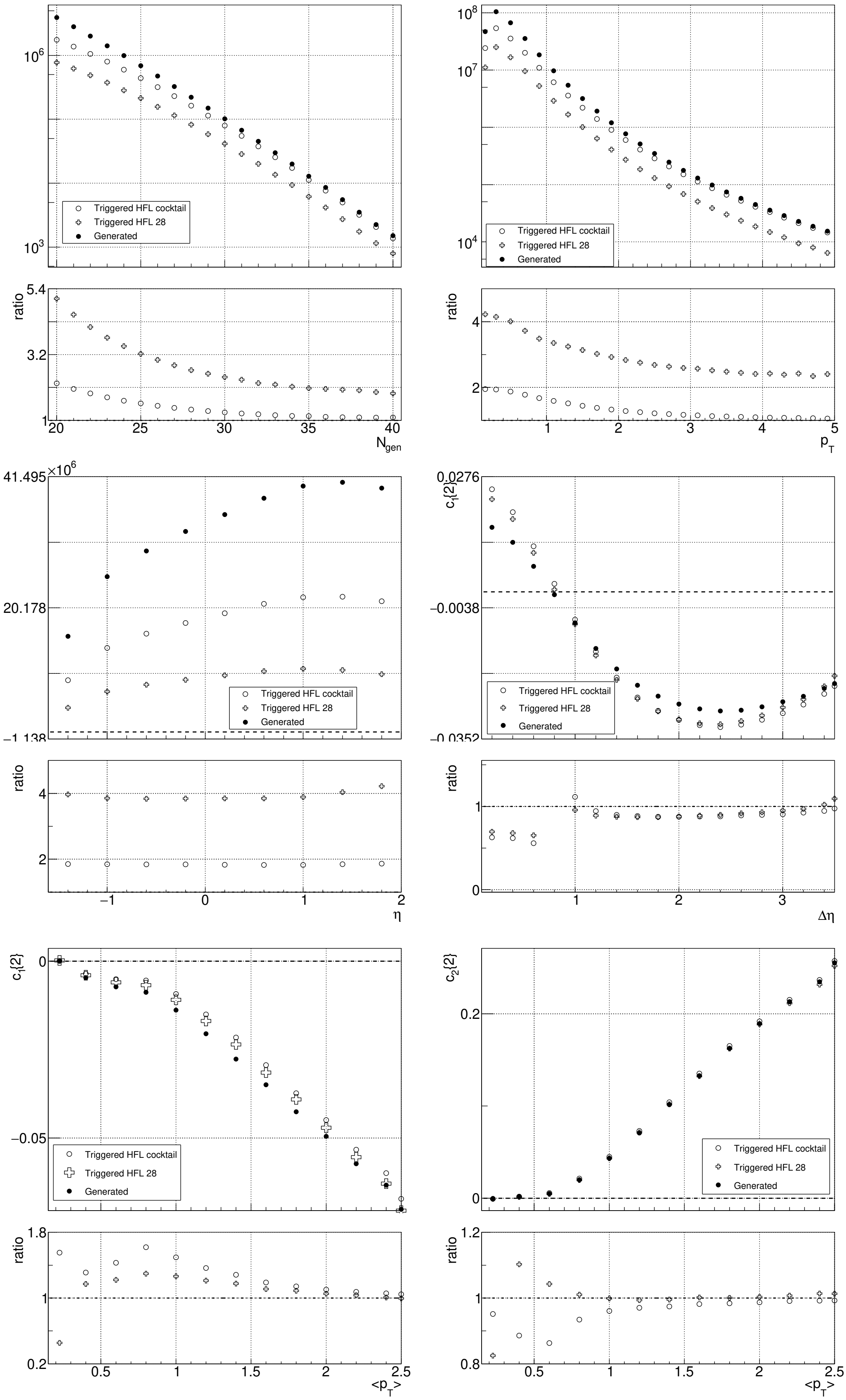
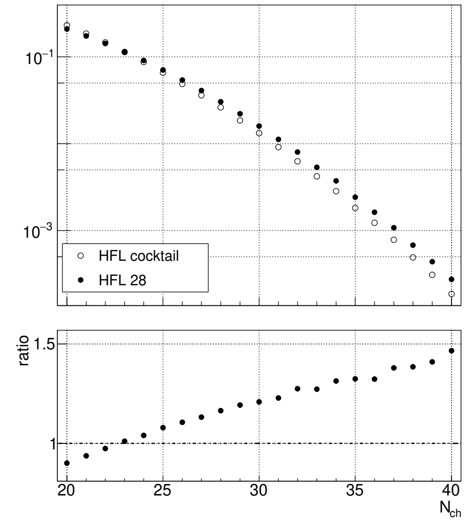
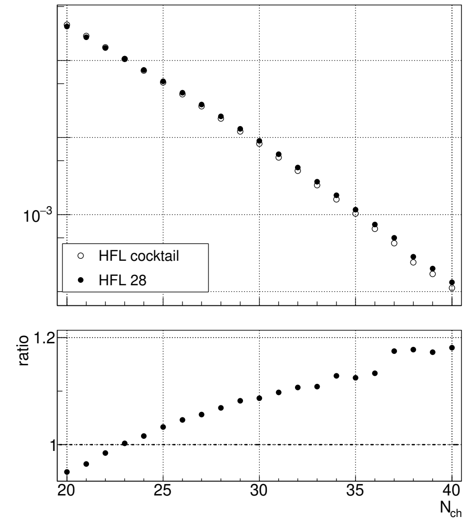


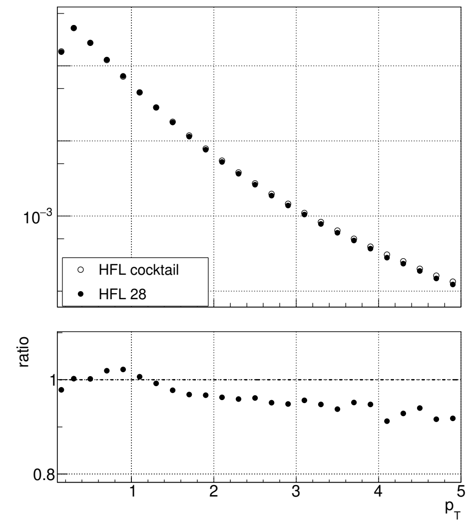
Figure 36: Trigger biases to various distributions. Ratios in bottom panel represent the correction factors to be applied to data.



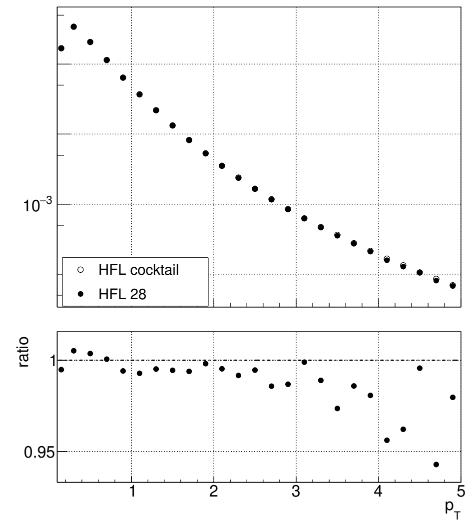
(a) Before trigger correction



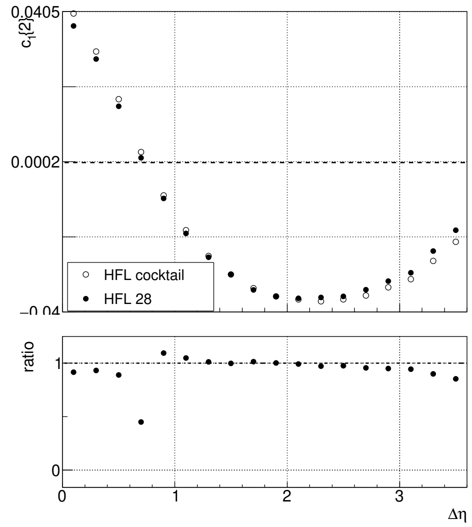
(b) After trigger correction



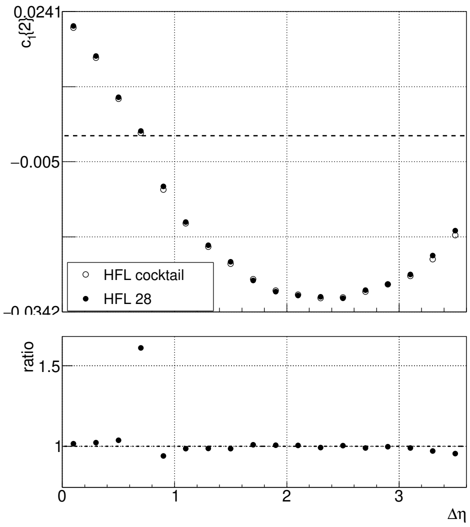
(c) Before trigger correction



(d) After trigger correction



(e) Before trigger correction



(f) After trigger correction

Figure 37: Comparison of distributions in ZEUS data before and after trigger bias correction. The distributions obtained from default trigger (HFL cocktail) are compared to those from HFL 28 alone. Residual differences after the trigger bias correction reflect the performance of trigger simulations in MC.

12.2.2 Light-flavor jet bias

The trigger bias was estimated using the pythia light-flavor jet MC, which is itself biased because of the jet requirement. We estimate and correct for this additional bias by generating inclusive PhP pythia as well as a jet-biased sample. Jet reconstruction is chosen to match that used in the light-flavor jet MC. Pythia provides built-in jet reconstruction algorithms of which we use fast-jet with the following constraints:

1. kT algorithm
2. $\Delta R = 1$
3. $p_{\mathrm{T}min} = 3.0$ GeV
4. $-3 < \eta < 3$
5. Observable massive final-state particles used for jet clustering

The ratio of inclusive to jet-biases pythia will form a correction factor to the previously mentioned trigger bias factor. This jet-bias correction factor will also be assigned as a systematic uncertainty. The full trigger bias correction factor is then:

$$\frac{D_{PhP}}{D_{PhP}^{\geq 1jet}} \frac{D_{lfjet}}{D_{lfjet}^{\mathrm{Trigger}}}, \tag{18}$$

where the first term represents a correction for the jet bias in our main MC sample and the second term is the trigger bias.

Figure 38 illustrates the effect of a jet requirement in pythia PhP for the distributions which were most biased by the trigger.

12.3 Application of correction factors.

There are two types of correction factors used in this analysis. First, there is the correction for tracking inefficiencies given by $w^{(n)}$, or the response matrix in the case of the N_{ch} distribution. Second, there is the trigger bias correction. The trigger bias correction itself contains a correction for the jet bias of the available ZEUS MC samples. Mathematically the correction procedure is:

$$N_{\mathrm{ch}} = \frac{N_{gen,PhP}}{N_{gen,PhP}^{\geq 1jet}} \frac{N_{gen,lfjet}}{N_{gen,lfjet}^{\mathrm{Trigger}}} [N_{rec}]_{\mathrm{unfolding}} \tag{19}$$

$$D_{corrected} = \frac{D_{PhP}}{D_{PhP}^{\geq 1jet}} \frac{D_{lfjet}}{D_{lfjet}^{\mathrm{Trigger}}} w^{(n)} D_{rec}. \tag{20}$$

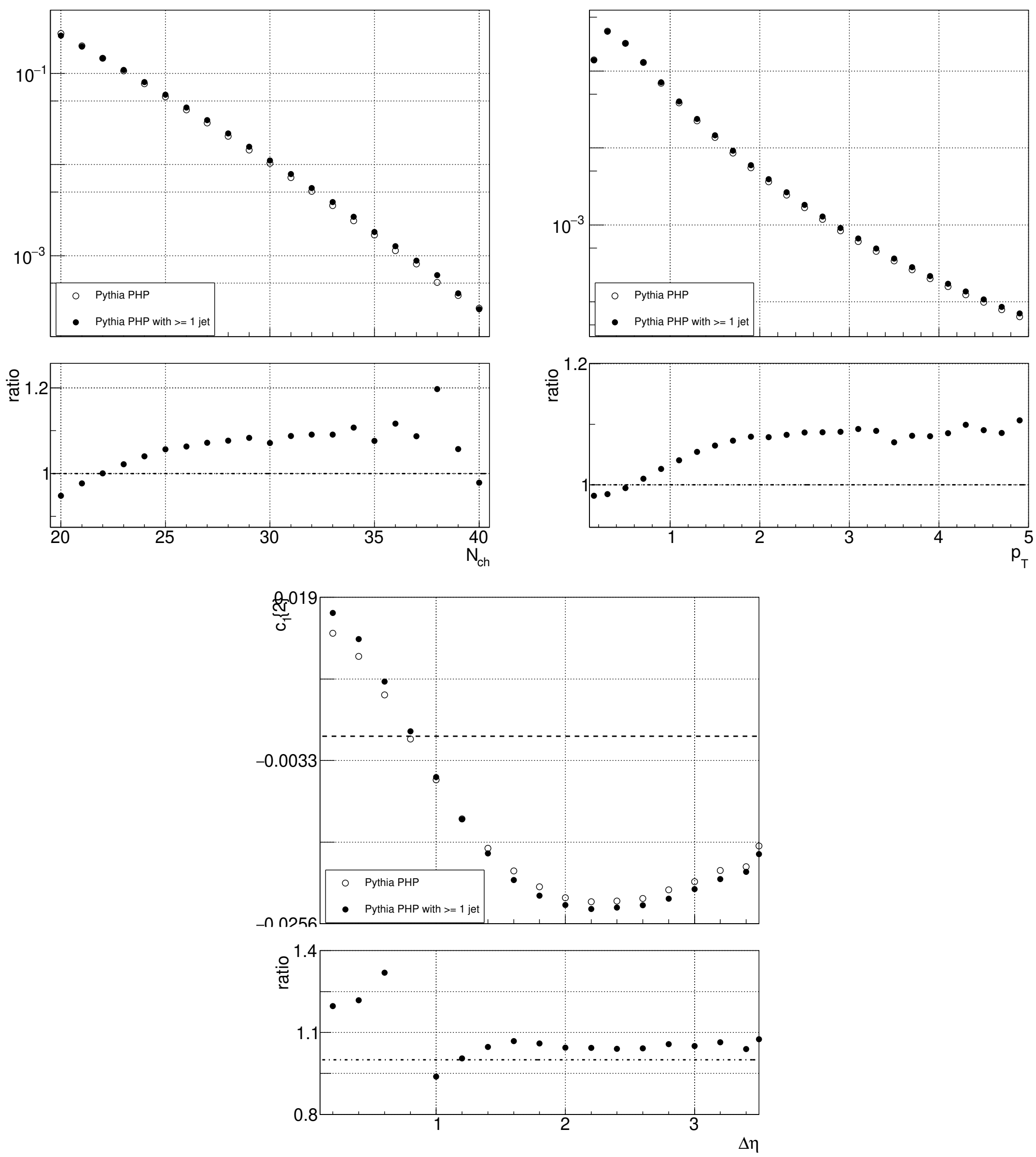


Figure 38: Comparison of inclusive to jet-biased pythia. The bias is not that large mainly due to the multiplicity cut of this analysis ($N_{\text{gen}} \geq 20$). High multiplicity events often have reconstructable jets.

For the DIS part of this analysis, there is no significant trigger bias and so no trigger bias correction.

12.4 ϕ resolution

Finite track ϕ resolution one source of distortions to the correlation functions. It is not correctable using single-particle weights but is, in principle, using multiparticle-weights as in Eq. 13. The ϕ resolution is shown in Fig. 39.

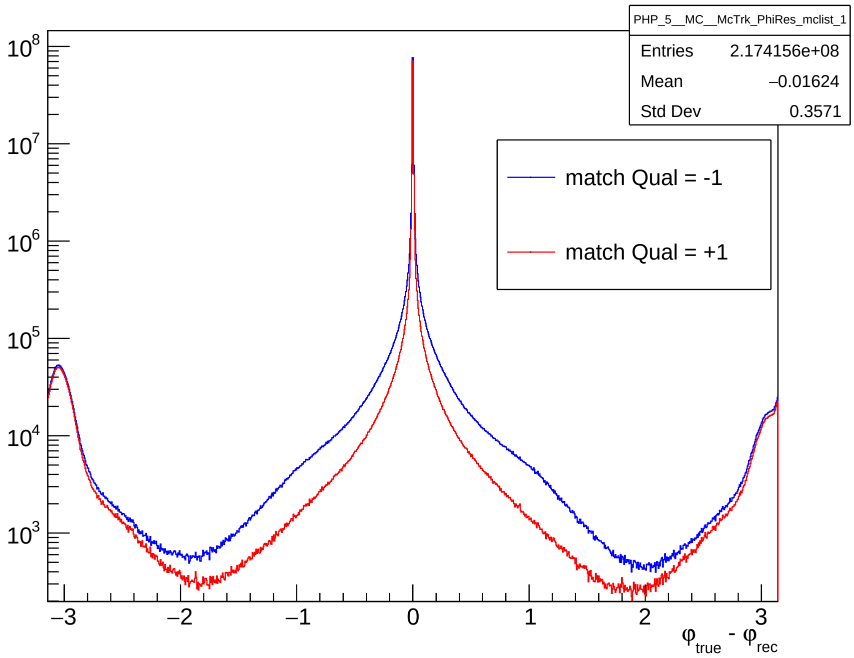


Figure 39: Resolution of φ obtained from MC pythia PhP lfjet 06e. The wings are caused by charge flipping for low-p_{T} tracks (<0.3 GeV).

12.5 Unfolding

Part of the MC nonclosure stems from Nch bin migration. This occurs when true correlation belongs to a particular $N_{\rm ch}$ bin but gets inserted at a different $N_{\rm rec}$ bin due to detection inefficiencies. This part of the MC nonclosure has nothing to do with the smearing of $c_n\{2\}$ due to single-particle and pair reconstruction inefficiencies.

The effect of 1D smearing (folding) can be formulated as

$$f_{rec}(x_i) = \sum_j R(x_i, y_j)\, f_{true}(y_j), \tag{21}$$

where $R(x, y)$ is the response matrix and f is a probability distribution function. The goal of unfolding is essentially to invert the response matrix and obtain $f_{true}(y)$. If one seeks to unfold a multiplicity distribution, f_{rec} and f_{true} represent the $N_{\rm rec}$ and $N_{\rm ch}$ distributions, respectively.

To demonstrate this effect we consider generator level correlations in MC binned either by the true MC $N_{\rm ch}$ and by $N_{\rm rec}$. We use the RooUnfold package using the Bayesian approach with the regularization parameter, iterations=4. The response matrix is the histogram of $N_{\rm rec}$ versus $N_{\rm ch}$ obtained from MC and is shown in Fig. 40

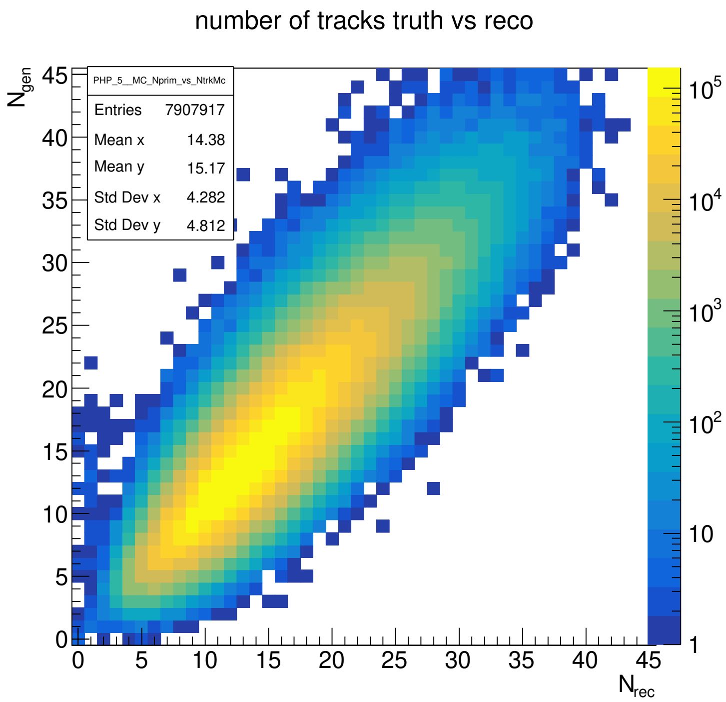


Figure 40: Response histogram used for unfolding.

Figure 41 demonstrates the performance of RooUnfoldBayes on the multiplicity distribution.

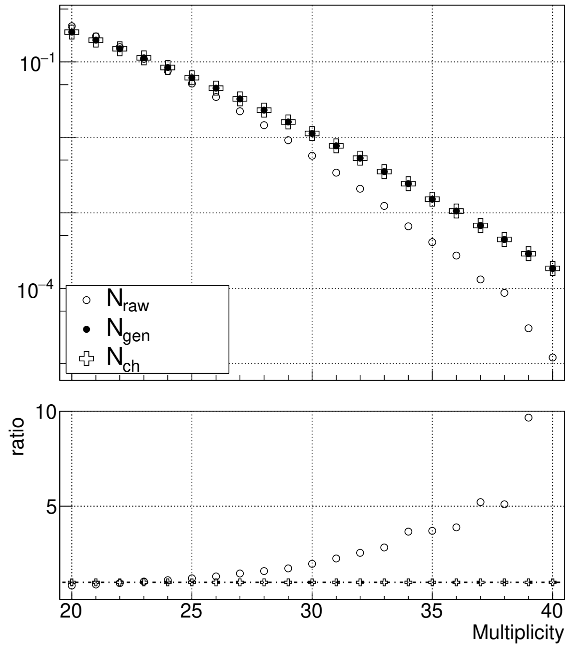


Figure 41: Unfolded multiplicity distribution in Ariadne 0607p.

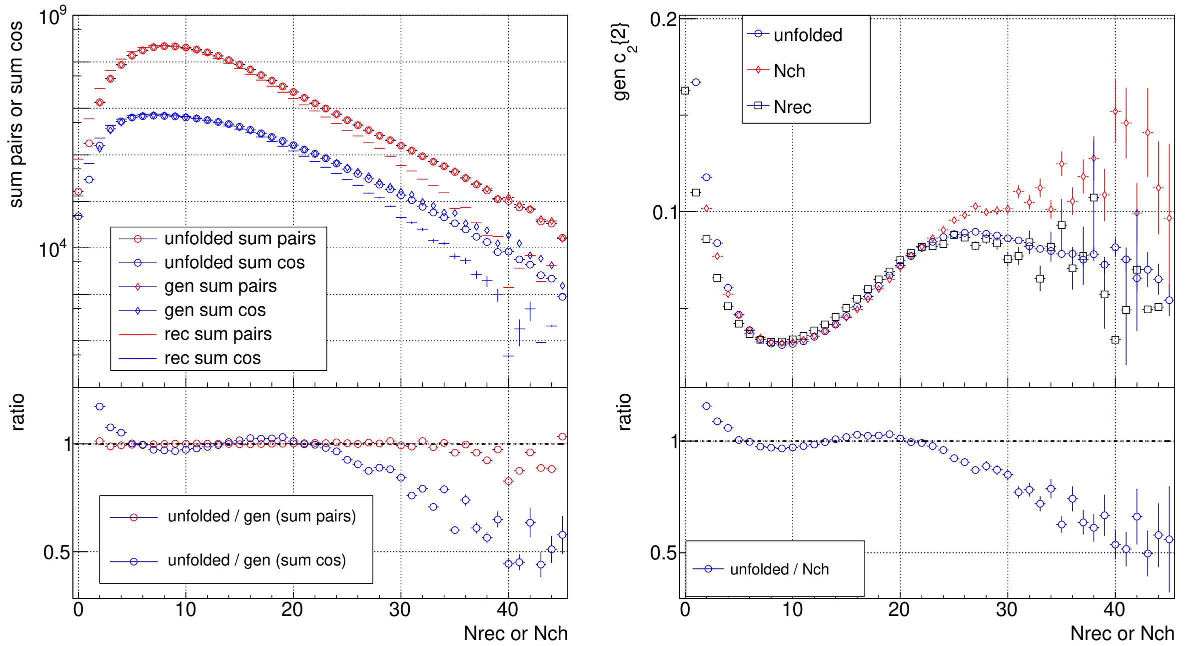
Unfolding of the $c_n\{2\}$ correlation function is more complicated because it is inherently a multidimensional problem. An event may be characterized by several parameters which can be grouped into a vector: $\mathbf{y} = (y_1, y_2, ..., y_N)$. These may be quantities like $N_{\rm rec}$, sum over particle $p_{\rm T}$, sum over over particle η, etc. The effect of multidimensional bin migration can then be formulated as

$$f_{meas}(x) = \int d\mathbf{y}\, R(x|\mathbf{y})\, f_{true}(\mathbf{y}) \tag{22}$$

To unfold $c_n\{2\}$, it is important to take into account the number of pairs, which may be different in an event described by $\mathbf{y}_1$ from another event described by $\mathbf{y}_2$. For that reason we split $c_n\{2\}$ into its numerator (sum of cosines) and denominator (sum of pairs). These components are to be unfolded separately and then reassembled to obtain the unfolded $c_n\{2\}$.

Figure 42 shows the 1D unfolding process attempted for the correlation function. We see from the left side of 42 that denominator of the $c_n\{2\}$ can be reliably unfolded while the numerator cannot. The reason for the failure of the latter is due to the 1D treatment of the unfolding. Events with the same pairing of $N_{\rm rec}$ and $N_{\rm ch}$ can have very different $c_n\{2\}$ due to different topologies. This can only be overcome by increasing the dimensionality of $\mathbf{y}$ in Eq. 22. To gauge how well and unfolding procedure for $c_n\{2\}$ will work for a given dimensionality of $\mathbf{y}$, one can define a variance for event topology $\mathbf{y}$:

$$\sigma(\mathbf{y}) = \frac{1}{N}\sum_{i}^{N} \left[c_n\{2\}(\mathbf{y})_i - \langle c_n\{2\}(\mathbf{y})\rangle\right]^2, \tag{23}$$



(a) Components of generated $c_2\{2\}$ separately unfolded.

(b) Reassembled $c_2\{2\}$ after separately unfolding its components.

Figure 42: Attempt to unfold generated $c_2\{2\}$ in Ariadne 0607p. Numerator of correlation function is the sum of cosines. Denominator is the sum of pairs.

where the sum is over all N events and the second term is the event average. As the dimensionality is increased, $\sigma \to 0$. Sufficient dimensionality is therefore achieved for sufficiently small σ.

13 Systematic uncertainties

The considered systematic uncertainties for PhP and DIS are shown in Table 13

Type of Systematic	Reference (default)	Variation(s)
MC closure	generator level distributions	efficiency corrected distributions
Track DCA variation *	DCA_xy,z $<$ 2.0 cm	DCA_xy,z $<$ 1.0 cm
Efficiency correction	PHP: direct + resolved, DIS: Ariadne	PHP: resolved only, DIS: Lepto
Z-vertex interval **	$\|Vz\| < 30$ cm	Separate intervals: $[-30, 0]$, $[0, +30]$
Tracking efficiency overestimation at low p_{T}	Using Libov and Bachynska's correction factors	1/2 correction factor
Data taking conditions * **	All	individual periods
PhP: MC light-flavor jet bias	Ratio of inclusive to jet-biased PhP Pythia	Ratio set to unity
PHP: trigger variation *	Trigger cocktail: HFL 1 || 5 || 21 || 28	HFL 5 alone, HFL 28 alone
PhP: Offline cuts	$P_e < 0.9$ && $E - Pz < 55$ && $E_e < 15$	$P_e < 0.98$ && $E - Pz < 65$ && $E_e < 30$

Table 13: *Systematic variations for the DIS and PhP analysis. Variations specific only to one of the analyses are specified as such in the first column. * means that the systematic was symmetrized. ** means that each variation was weighted by their relative contribution.*

For the Z-vertex and data taking condition systematics, the variations of each sub-dataset wrt the full dataset were scaled by the appropriate statistical sample size and added together in quadrature to form the total uncertainty. For instance:

$$\delta_{Z-vtx} = \sqrt{(0.5 \times (c_n\{2\}_{total} - c_n\{2\}_{left}))^2 + (0.5 \times (c_n\{2\}_{total} - c_n\{2\}_{right}))^2}, \qquad (24)$$

where left and right refer to $-30 < V_z < 0$ and $0 < V_z < 30$ cm, respectively. Each contain about 50% of the data-sample size.

Whenever the systematic variation was smaller than the propagated statistical uncertainty, the systematic uncertainty is set to zero. We also apply a smoothing procedure which replaces the systematic uncertainty of a given point to the average of that point plus its left and right neighbor. For instance, $\delta_n \to (\delta_{n-1} + \delta_n + \delta_{n+1})/3$.

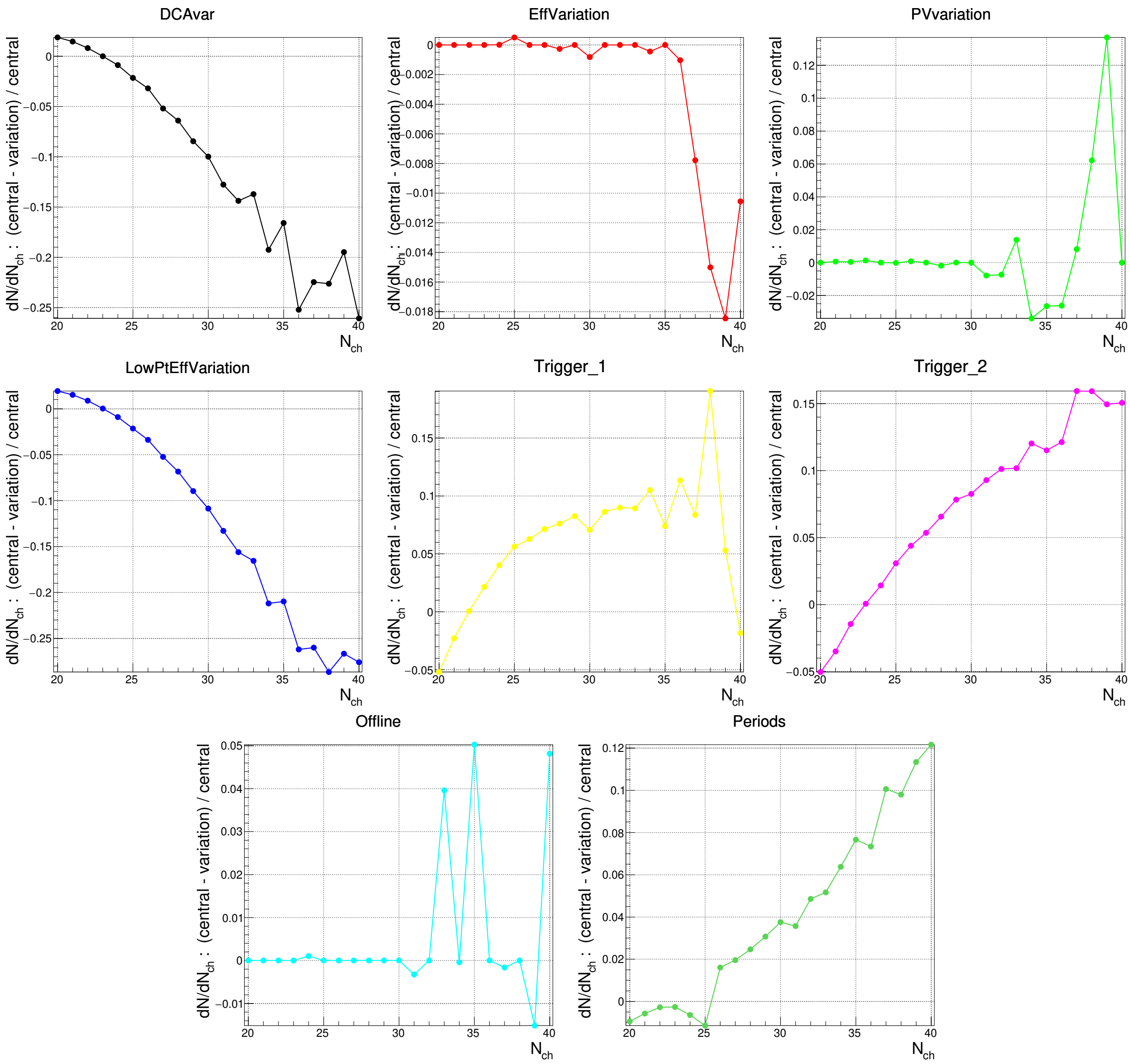


Figure 43: Separated systematics: dN/dN_{ch} in PhP.

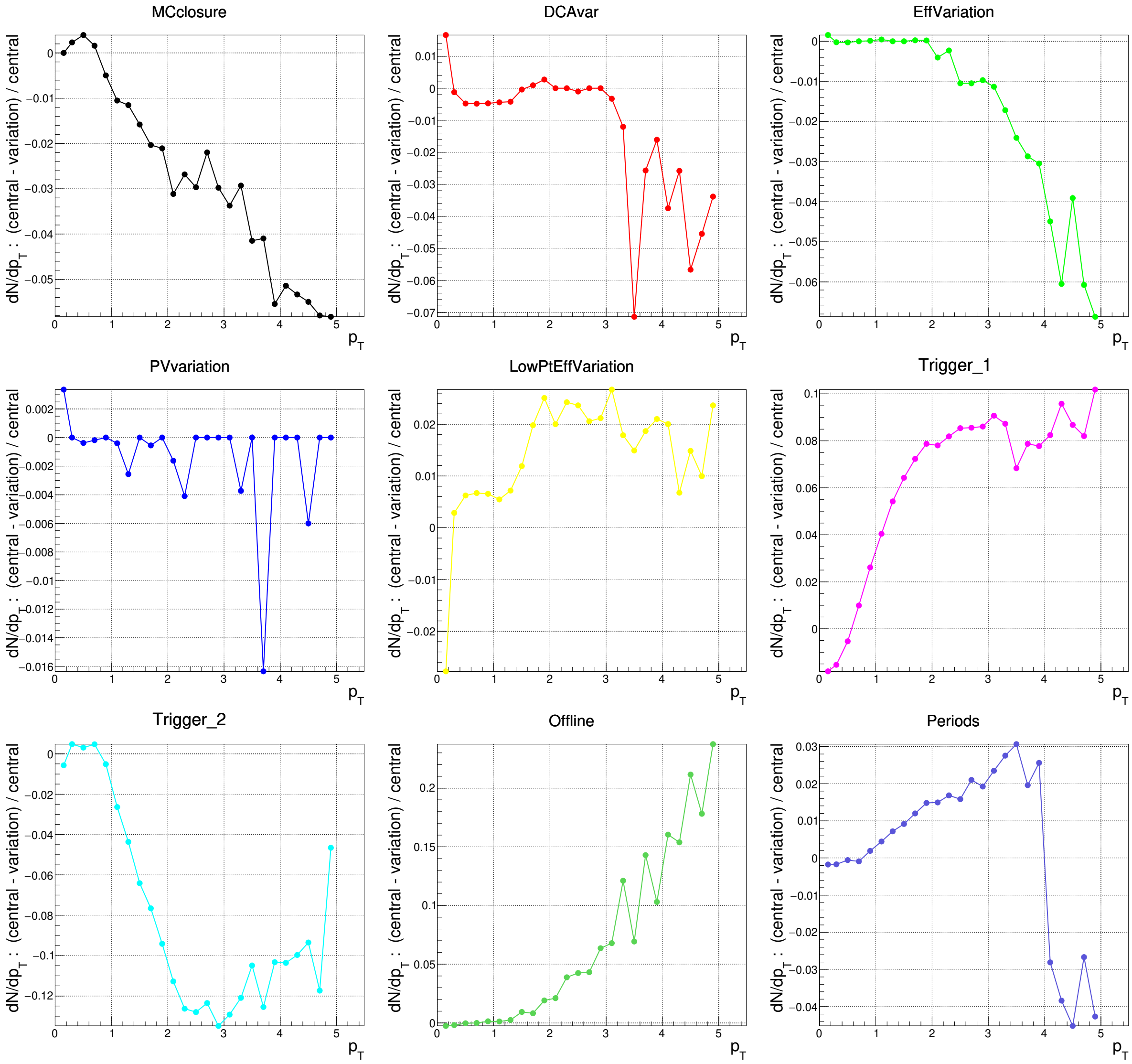


Figure 44: Separated systematics: dN/dp_{T} in PhP.

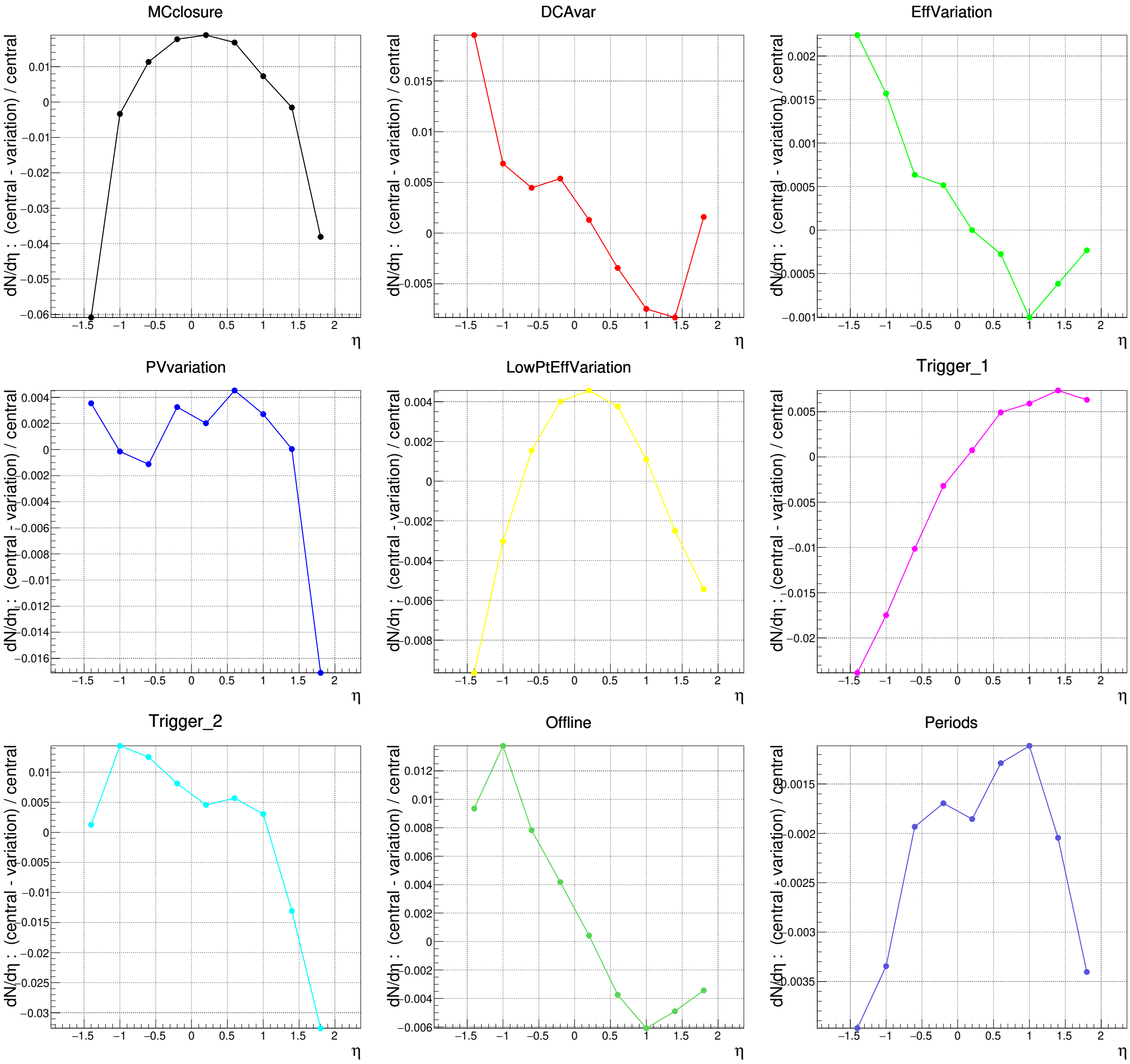


Figure 45: Separated systematics: $dN/d\eta$ in PhP.

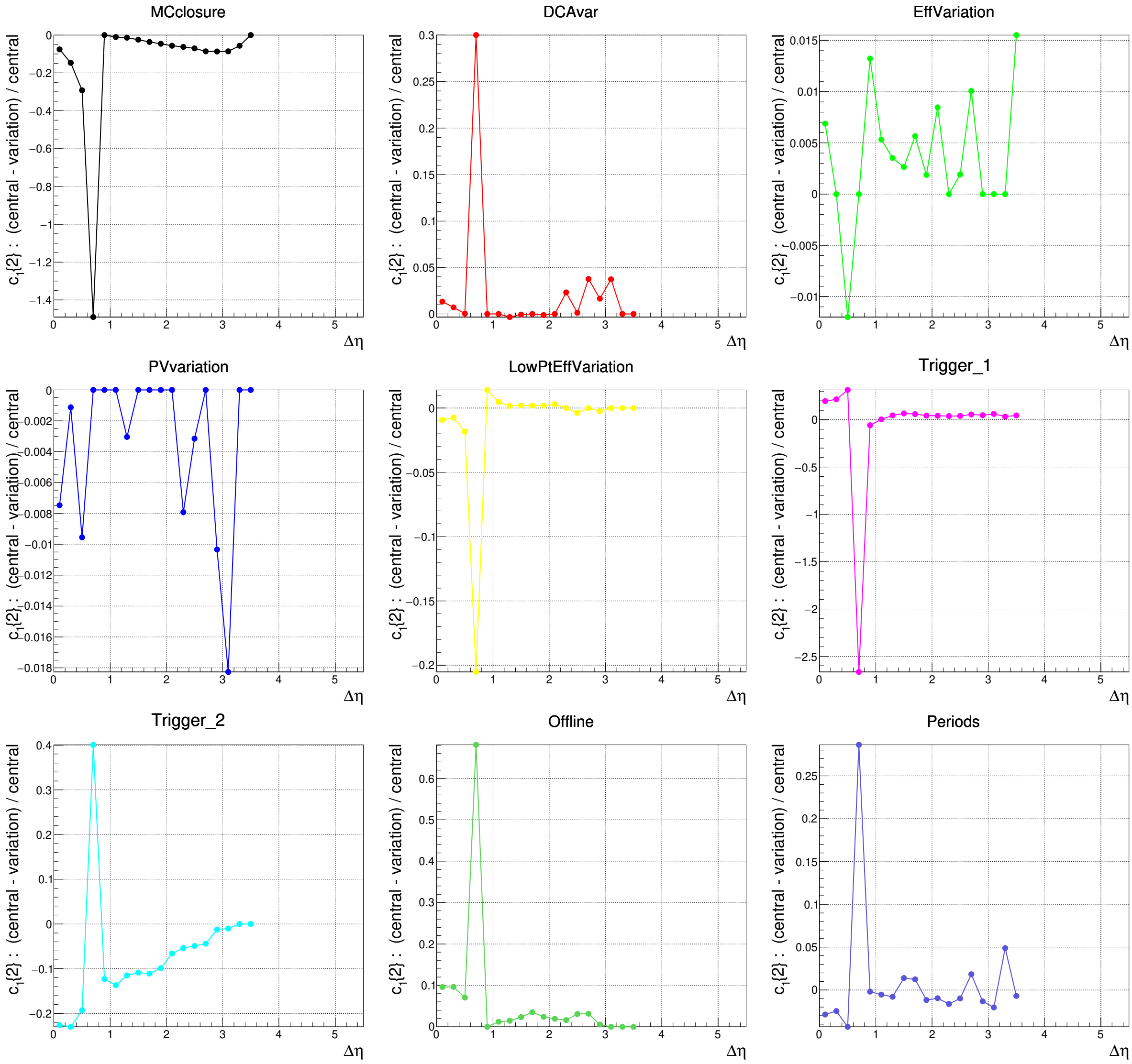


Figure 46: Separated systematics: $c_1\{2\}$ versus $\Delta\eta$ in PhP.

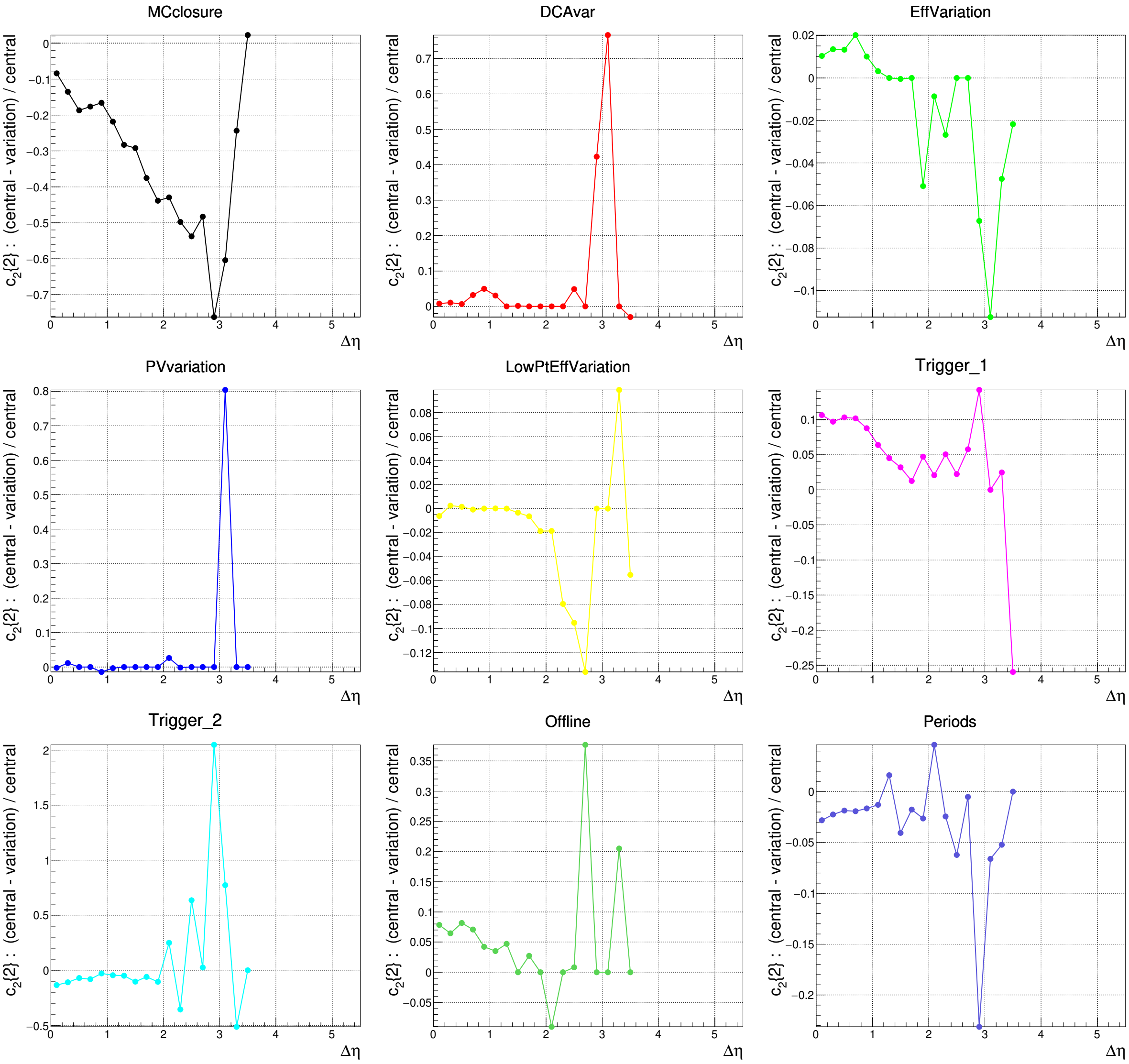


Figure 47: Separated systematics: $c_2\{2\}$ versus $\Delta\eta$ in PhP.

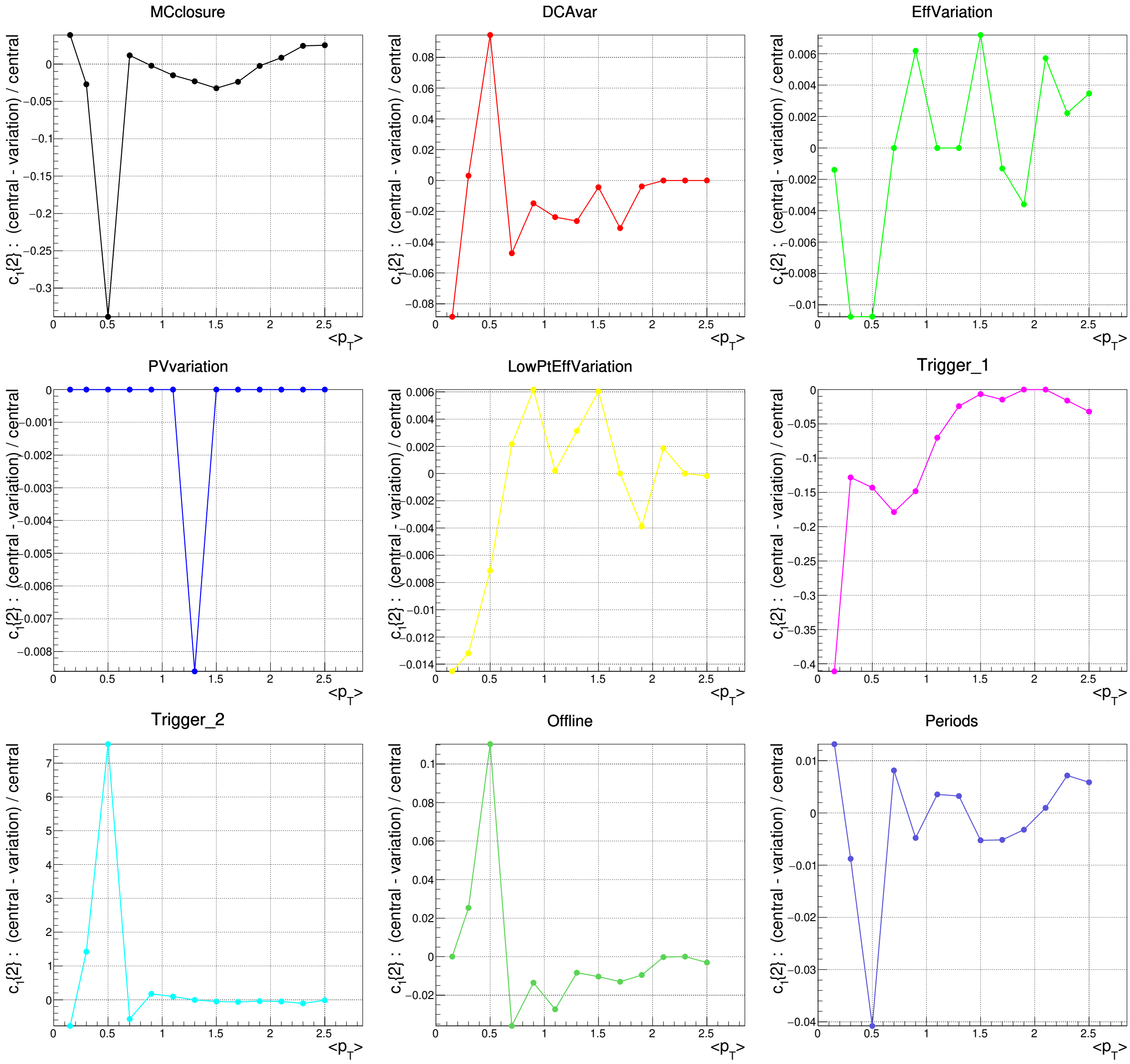


Figure 48: Separated systematics: $c_1\{2\}$ versus $\langle p_{\mathrm{T}} \rangle$ in PhP.

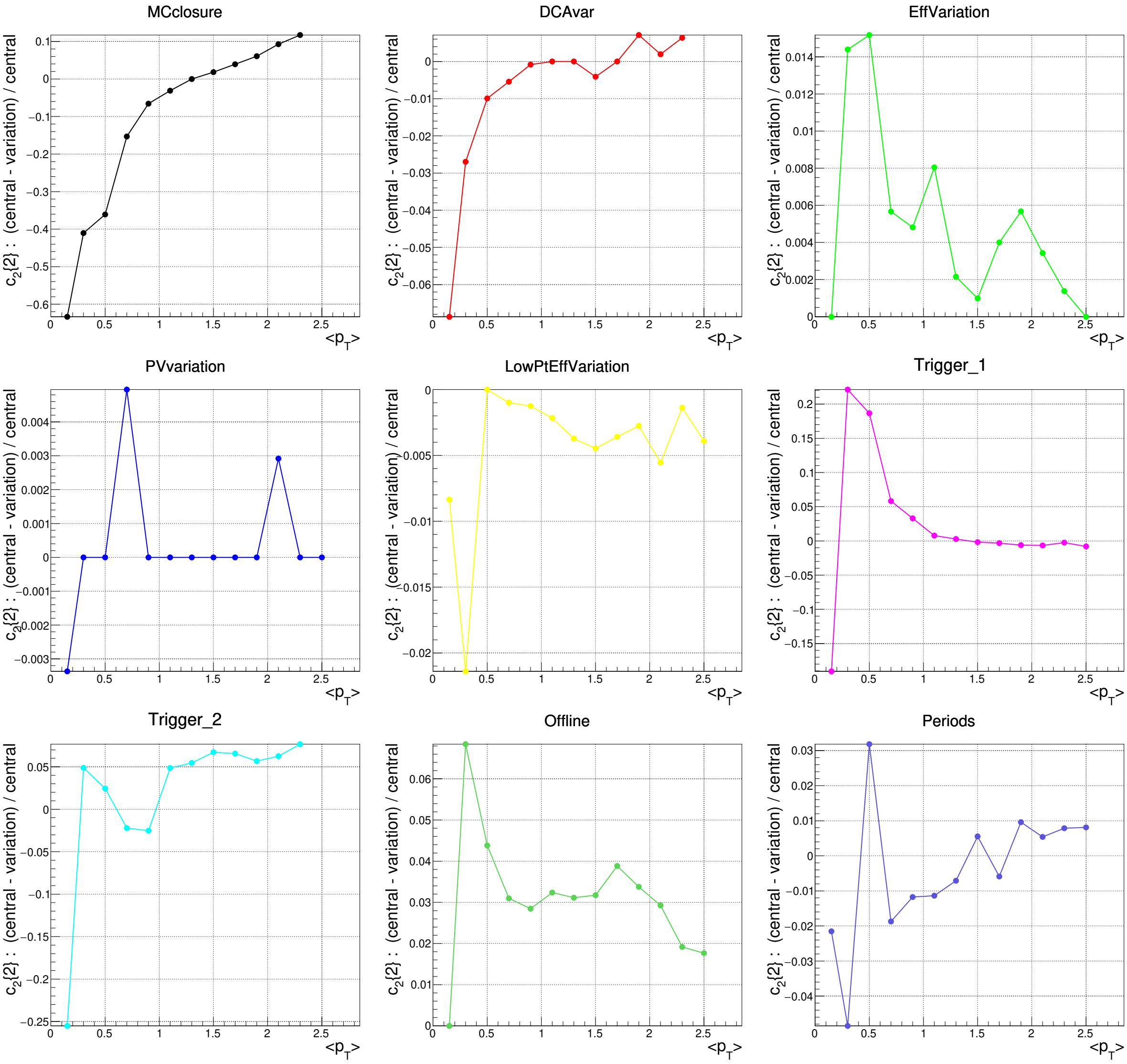


Figure 49: Separated systematics: $c_2\{2\}$ versus $\langle p_\mathrm{T} \rangle$ in PhP.

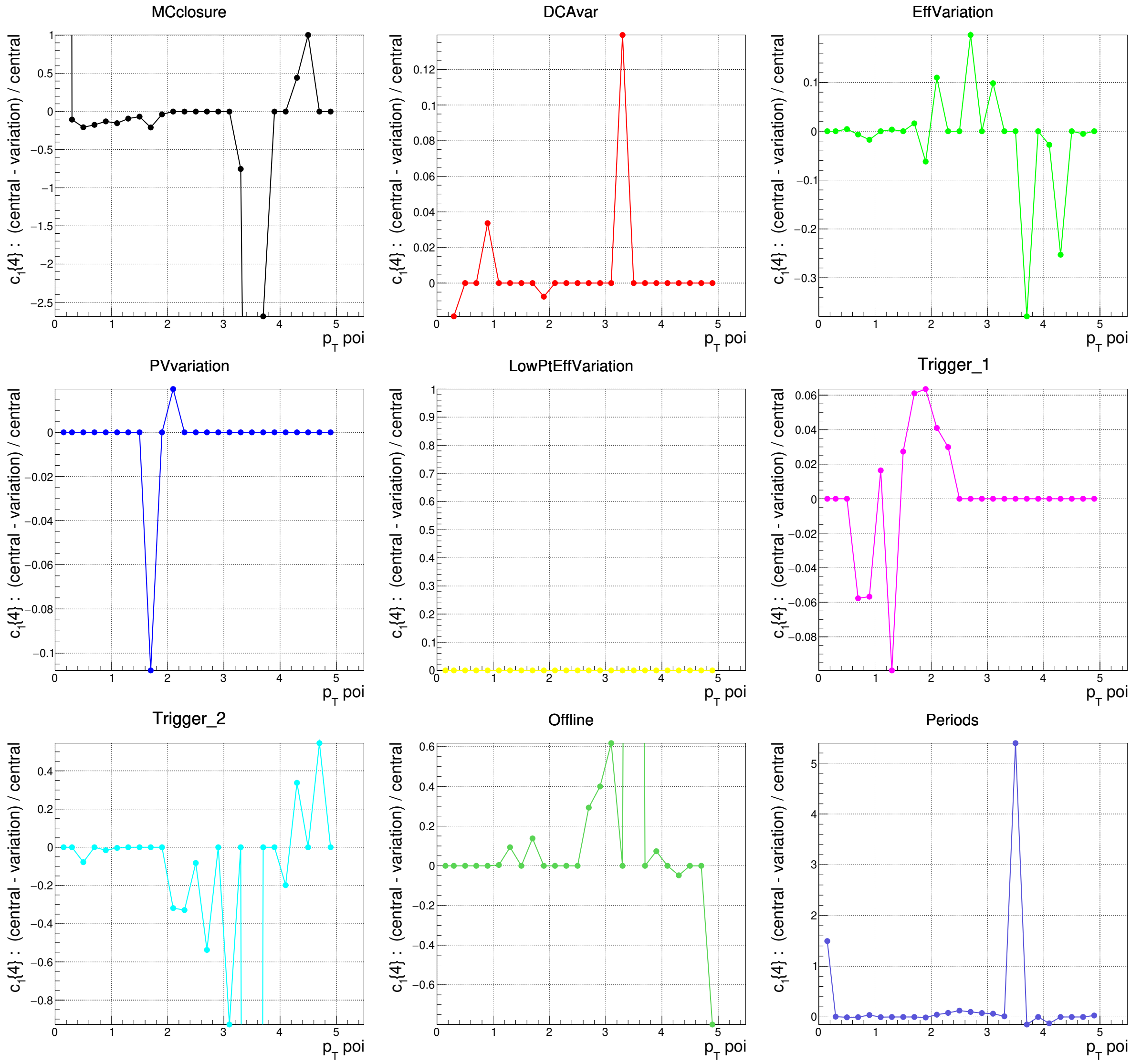


Figure 50: Separated systematics: $c_1\{4\}$ versus p_{T} poi in PhP.

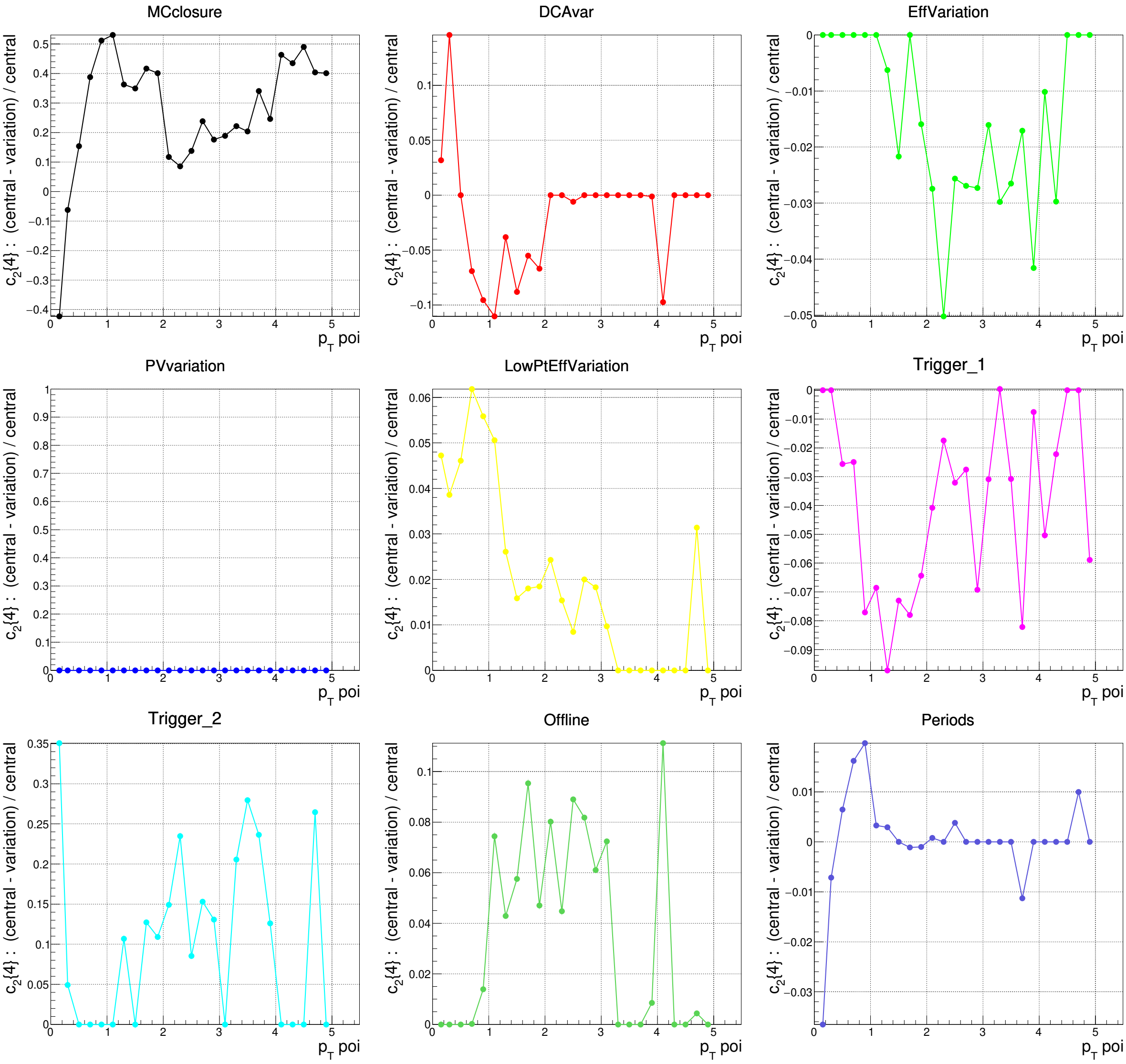


Figure 51: Separated systematics: $c_2\{4\}$ versus p_{T} poi in PhP.

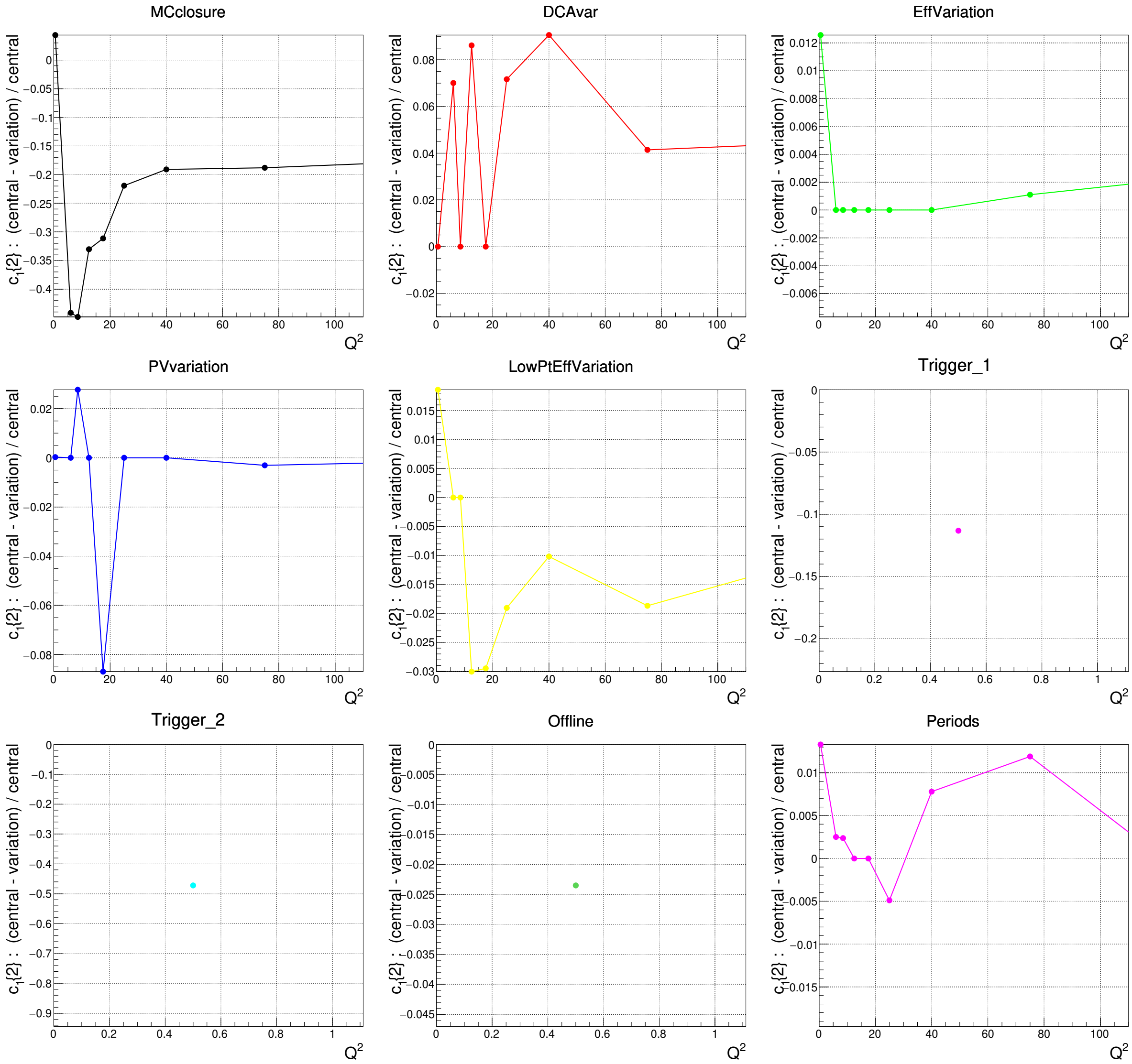


Figure 52: Separated systematics: $c_1\{2\}$ versus Q^2. Note, no DIS systematics for Trigger and Offline categories.

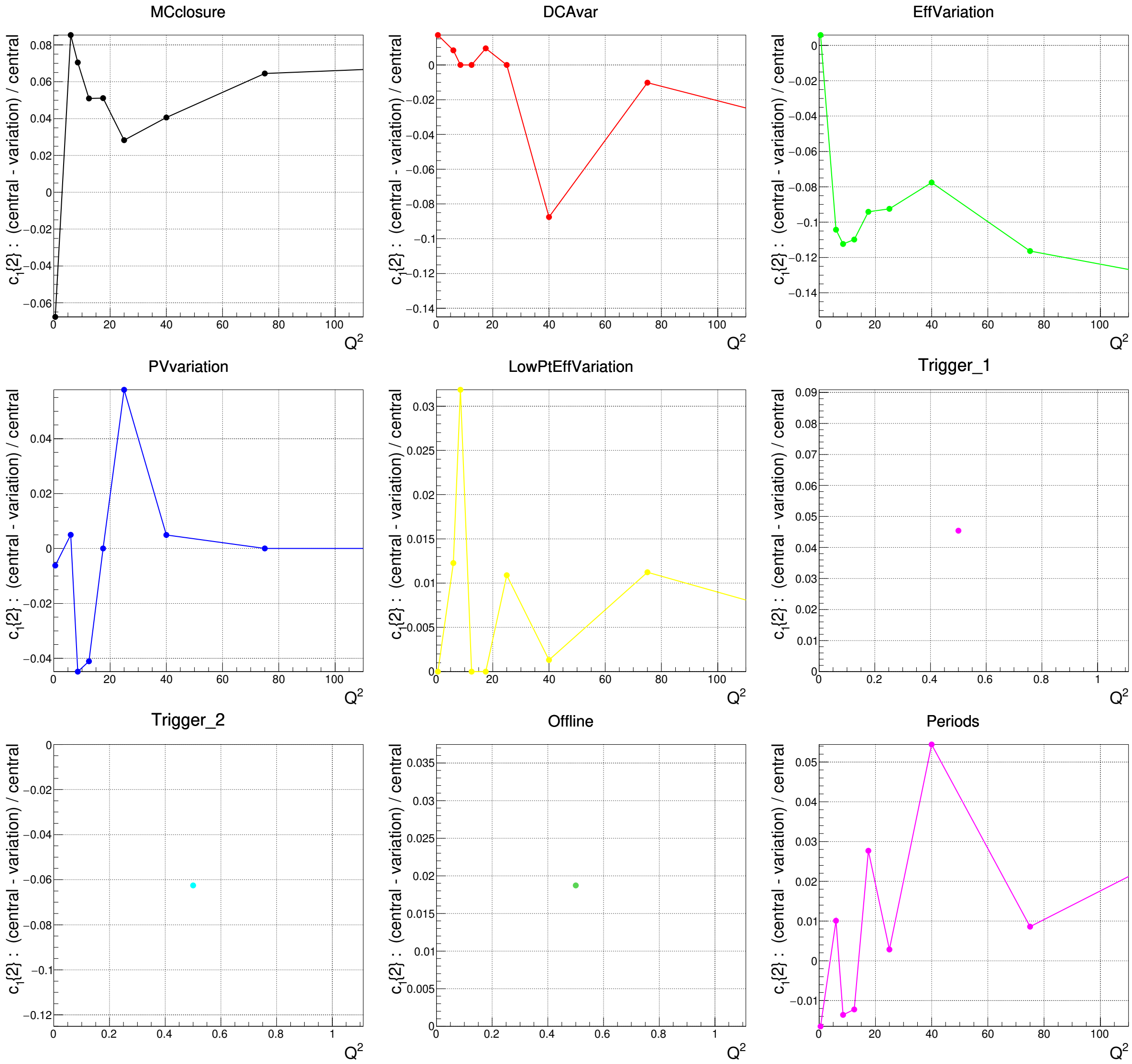


Figure 53: Separated systematics: $c_1\{2\}$, $\Delta\eta > 2.0$, versus Q^2. Note, no DIS systematics for Trigger and Offline categories.

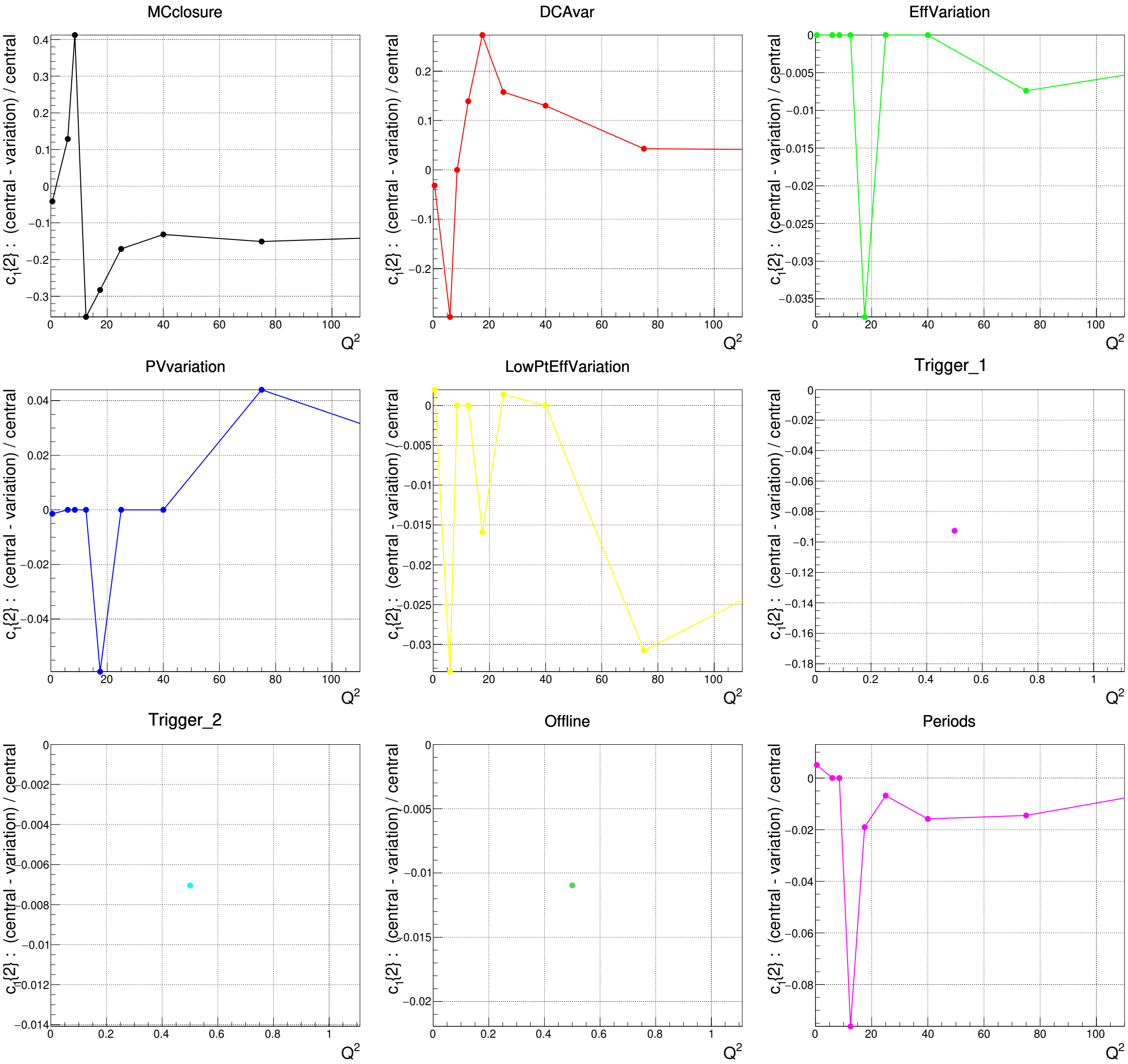


Figure 54: Separated systematics: $c_1\{2\}$, $p_{\mathrm{T}} > 0.5$, versus Q^2. Note, no DIS systematics for Trigger and Offline categories.

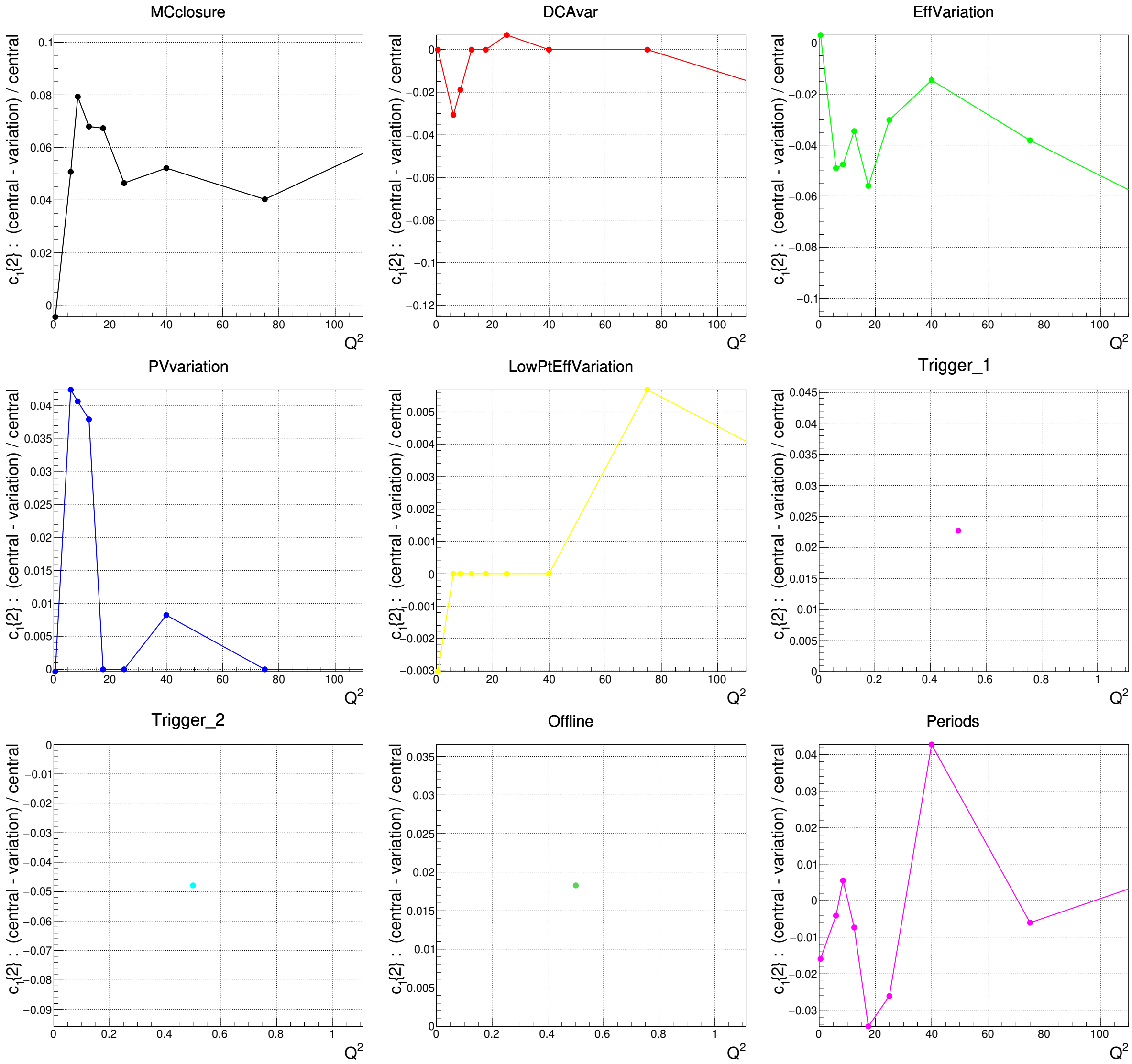


Figure 55: Separated systematics: $c_1\{2\}$, $p_\mathrm{T} > 0.5$, $\Delta\eta > 2.0$, versus Q^2. Note, no DIS systematics for Trigger and Offline categories.

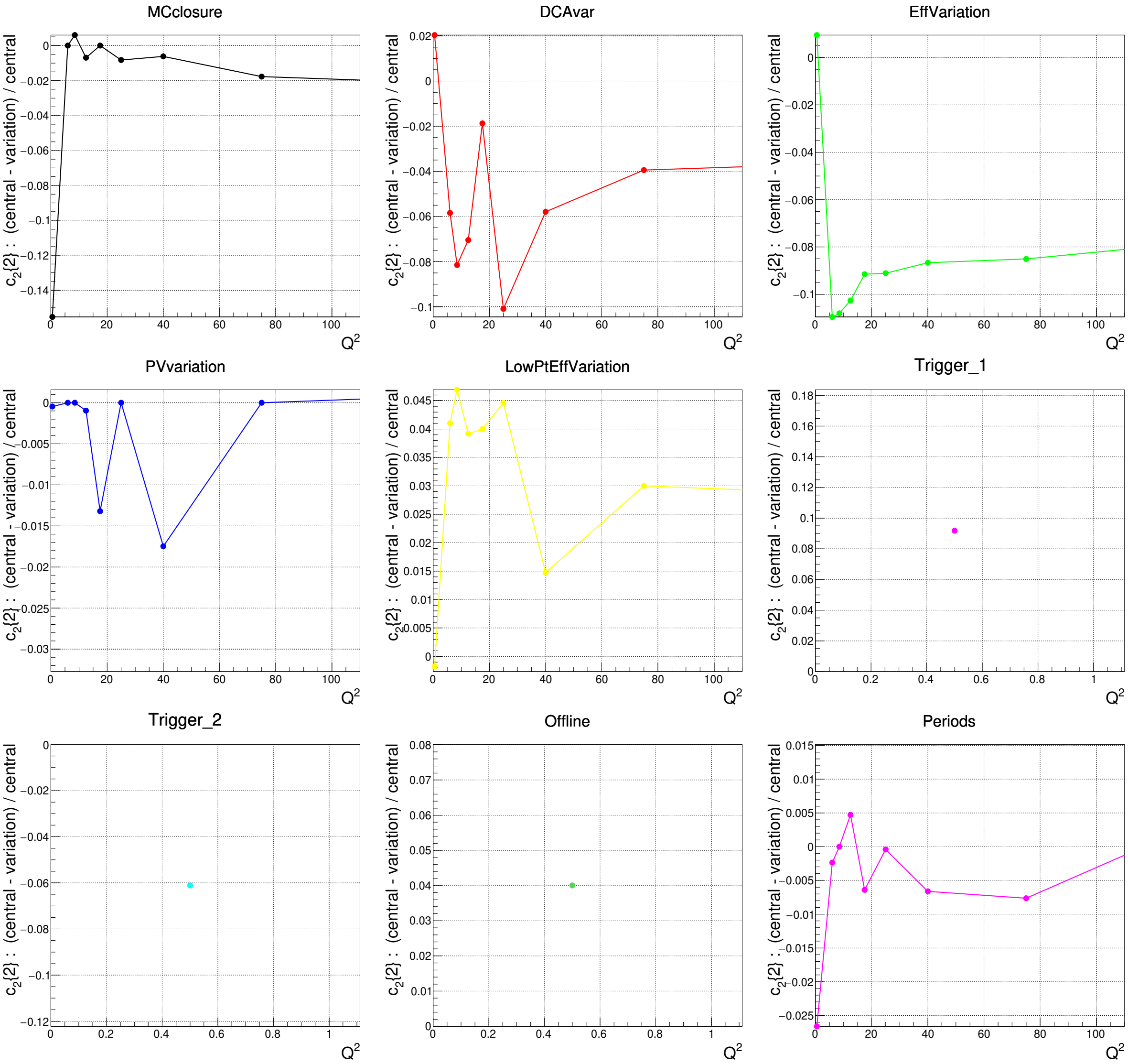


Figure 56: Separated systematics: $c_2\{2\}$ versus Q^2. Note, no DIS systematics for Trigger and Offline categories.

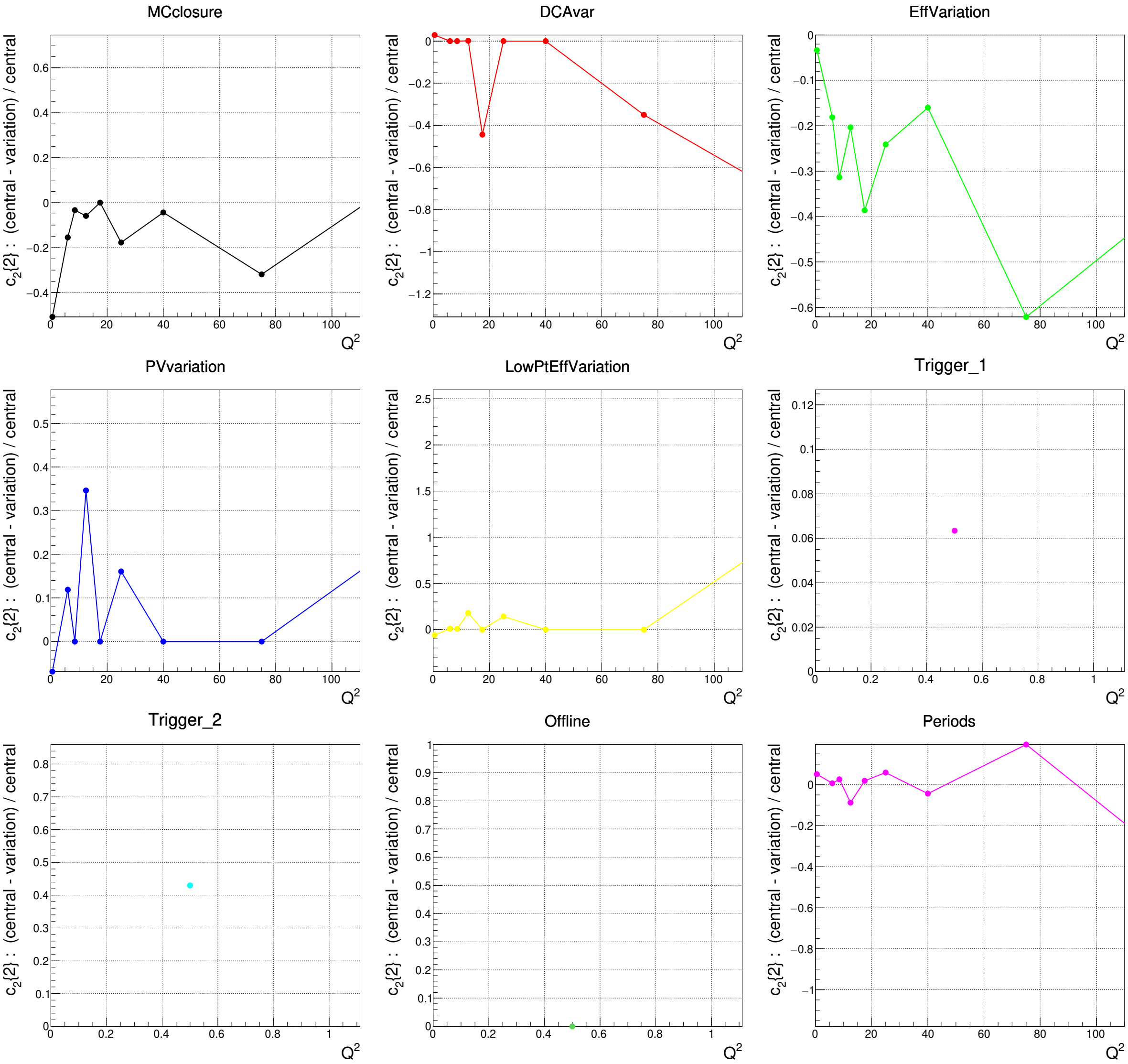


Figure 57: Separated systematics: $c_2\{2\}$, $\Delta\eta > 2.0$, versus Q^2. Note, no DIS systematics for Trigger and Offline categories.

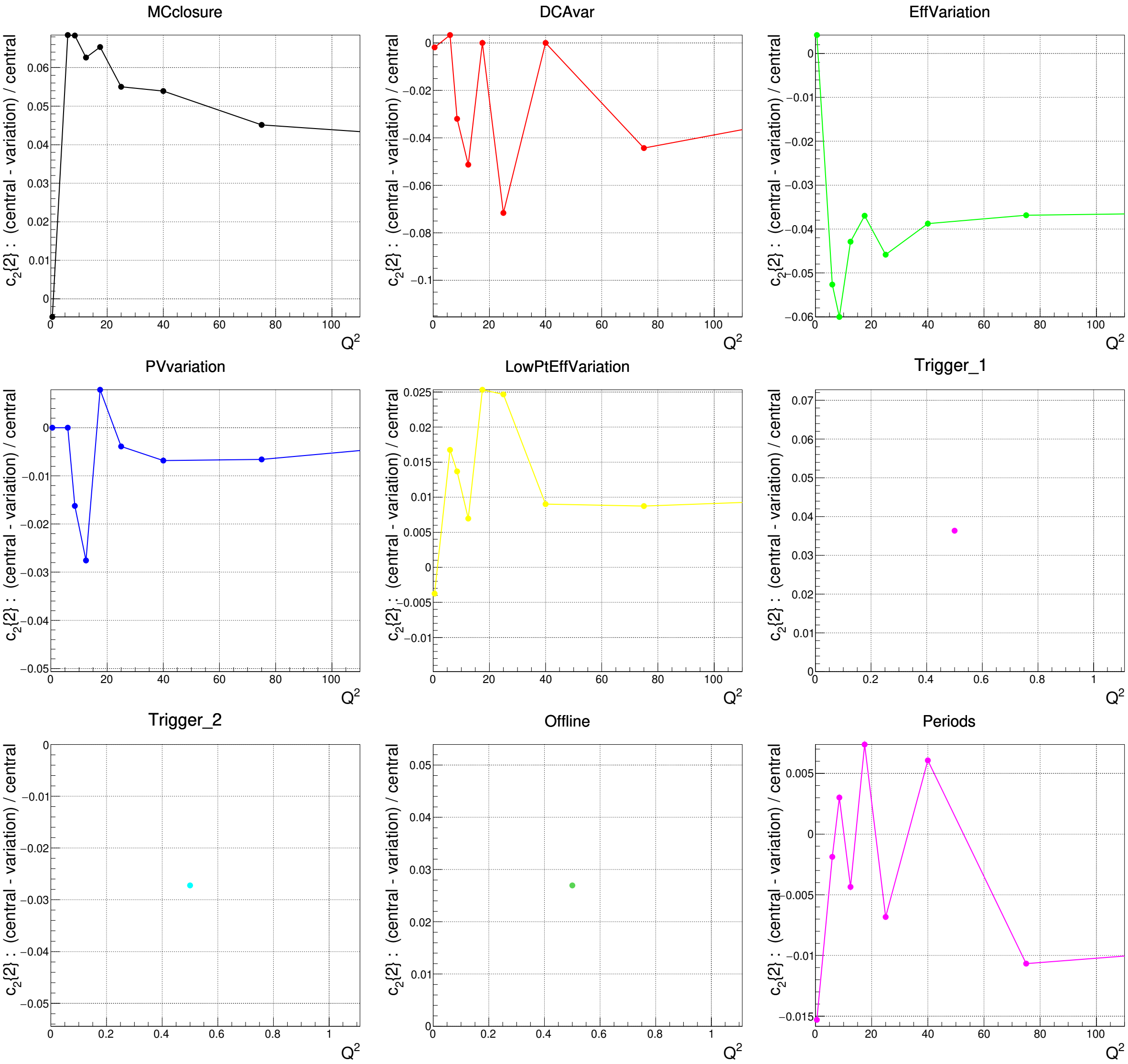


Figure 58: Separated systematics: $c_2\{2\}$, $p_{\mathrm{T}} > 0.5$, versus Q^2. Note, no DIS systematics for Trigger and Offline categories.

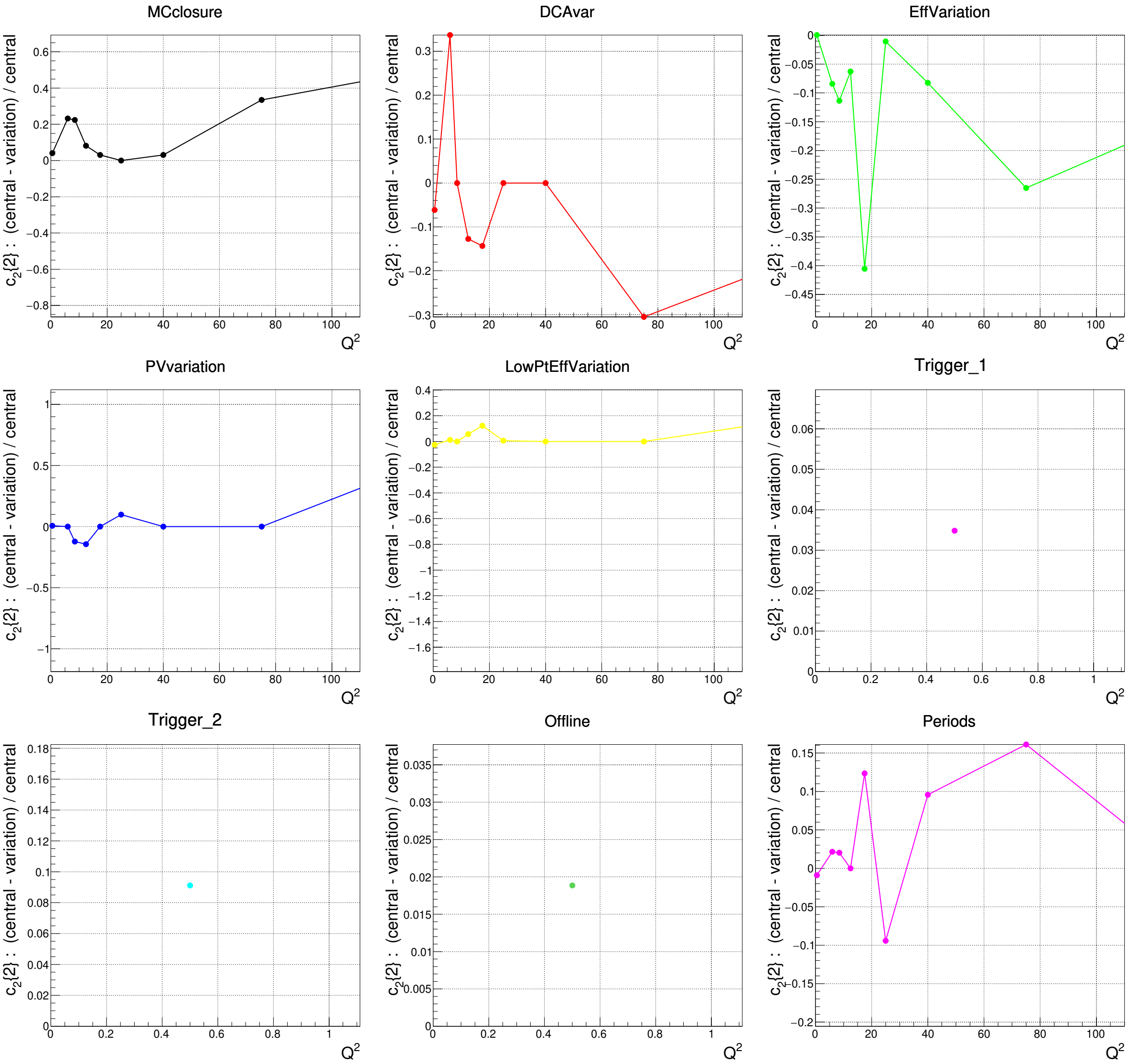


Figure 59: Separated systematics: $c_2\{2\}$, $p_{\mathrm{T}} > 0.5$, $\Delta\eta > 2.0$, versus Q^2. Note, no DIS systematics for Trigger and Offline categories.

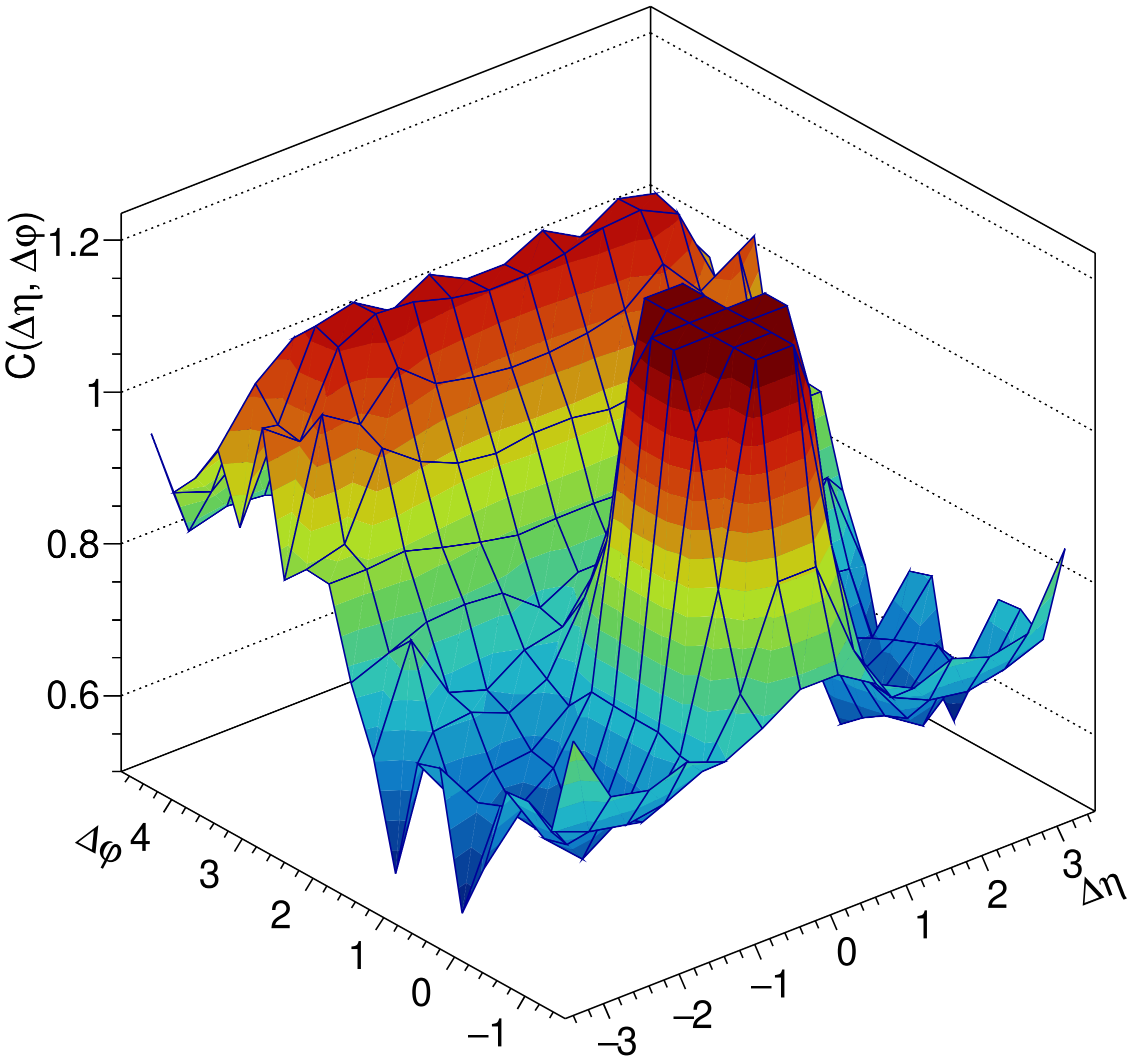


Figure 60: Ridge plot in DIS for $Q^2 > 20$, $N_{\mathrm{ch}} \geq 20$.

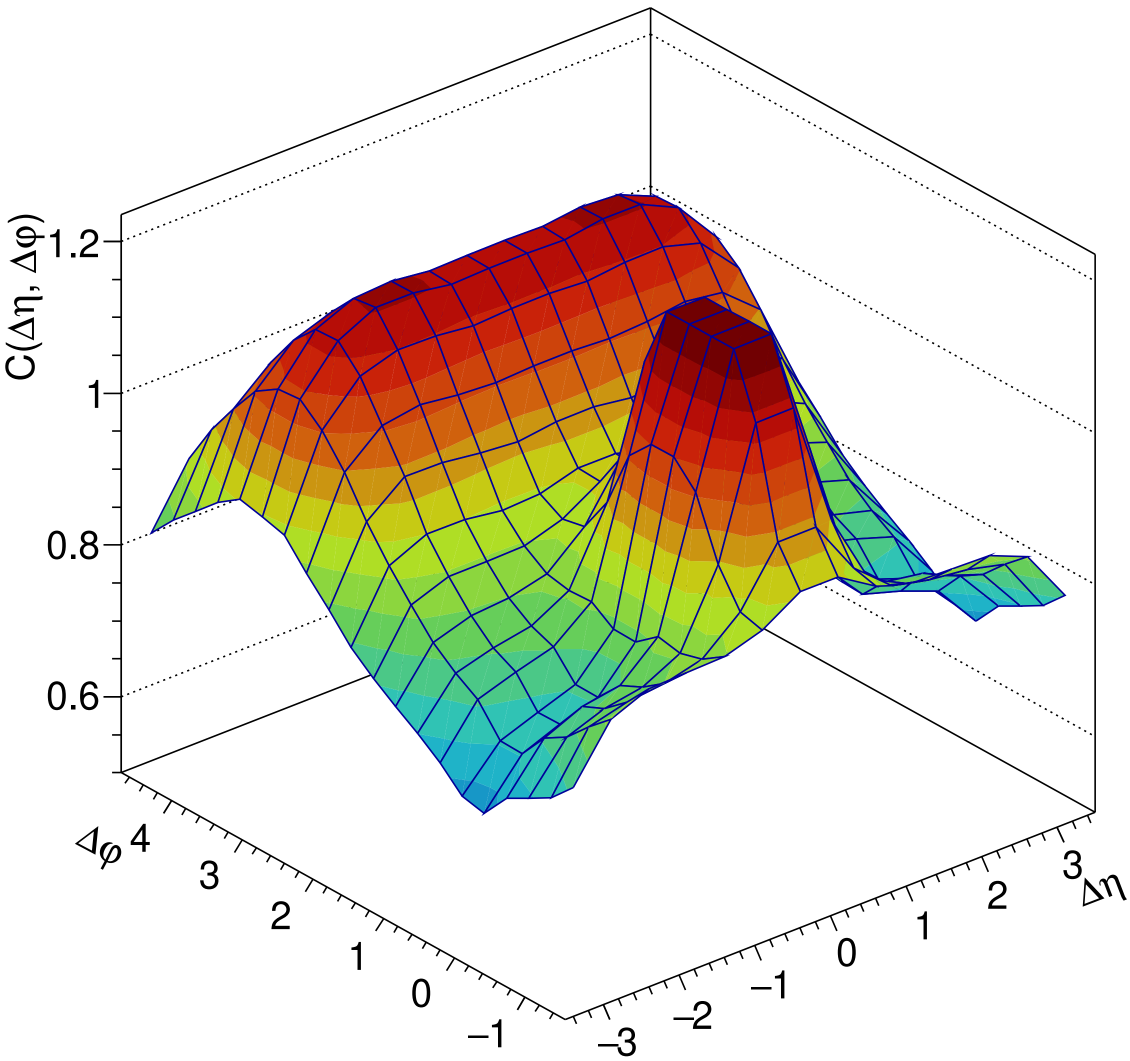


Figure 61: Ridge plot in PHP, $N_{\mathrm{ch}} \geq 20$.

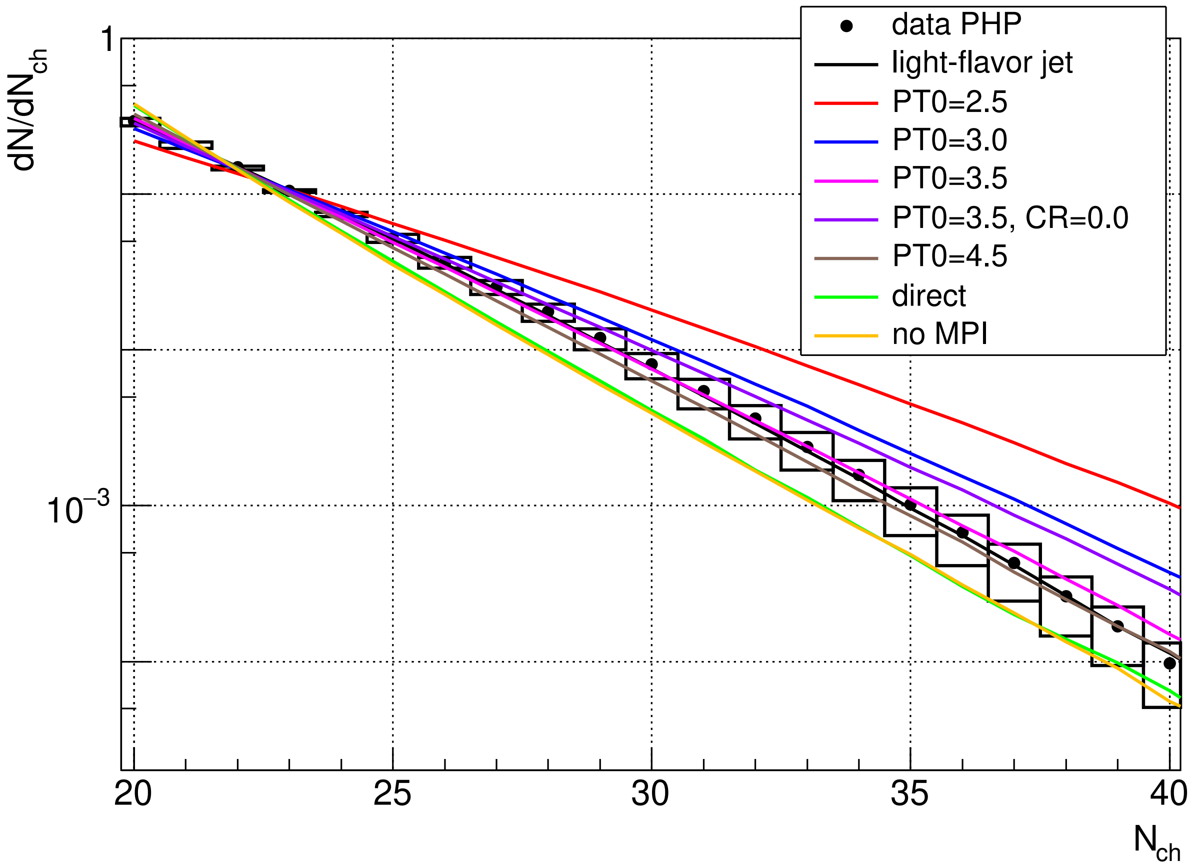


Figure 62: Unfolded multiplicity distribution.

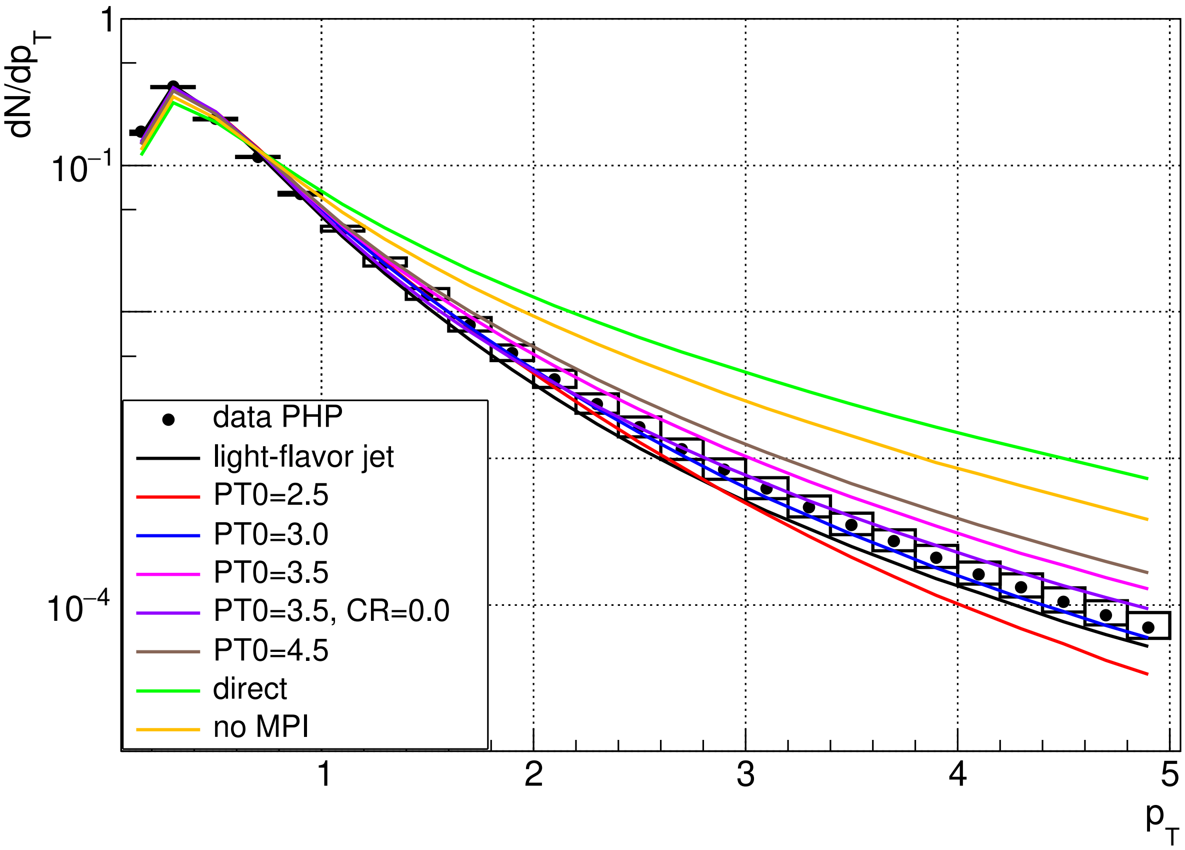


Figure 63: dN/dp_{T}

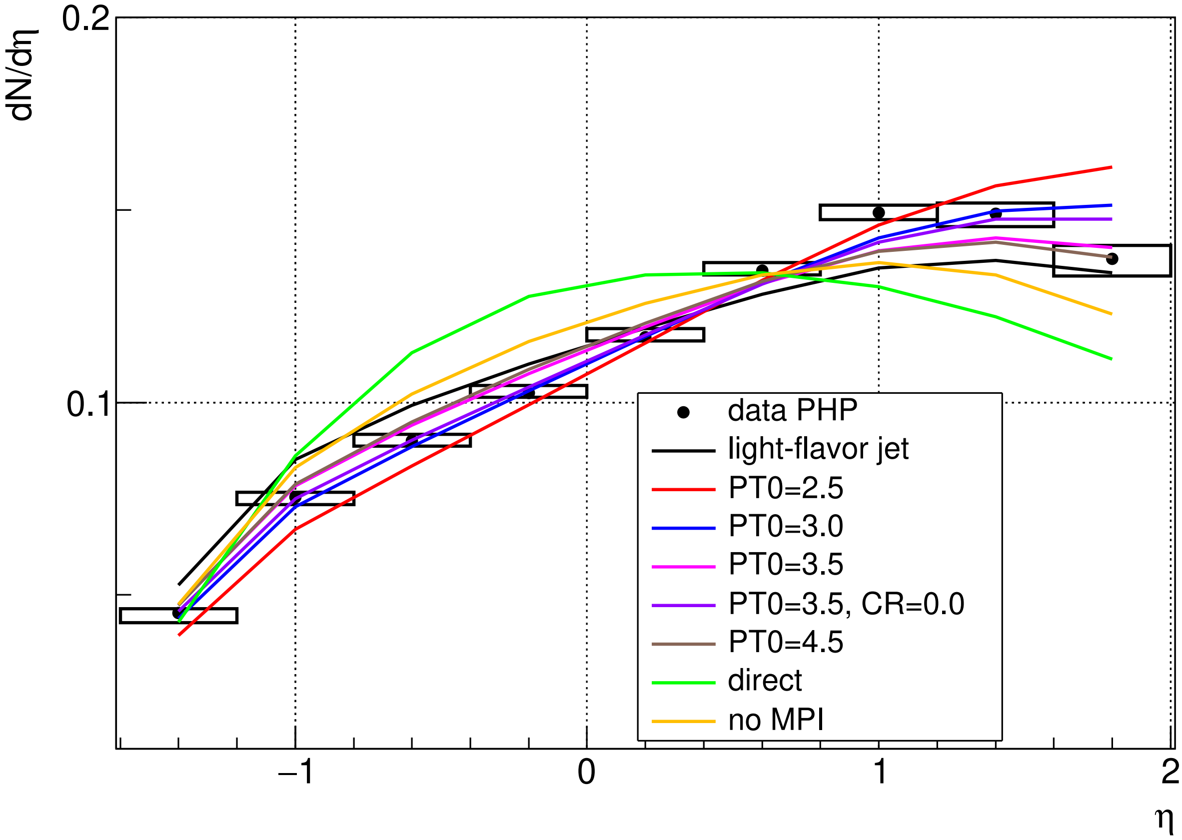


Figure 64: $dN/d\eta$

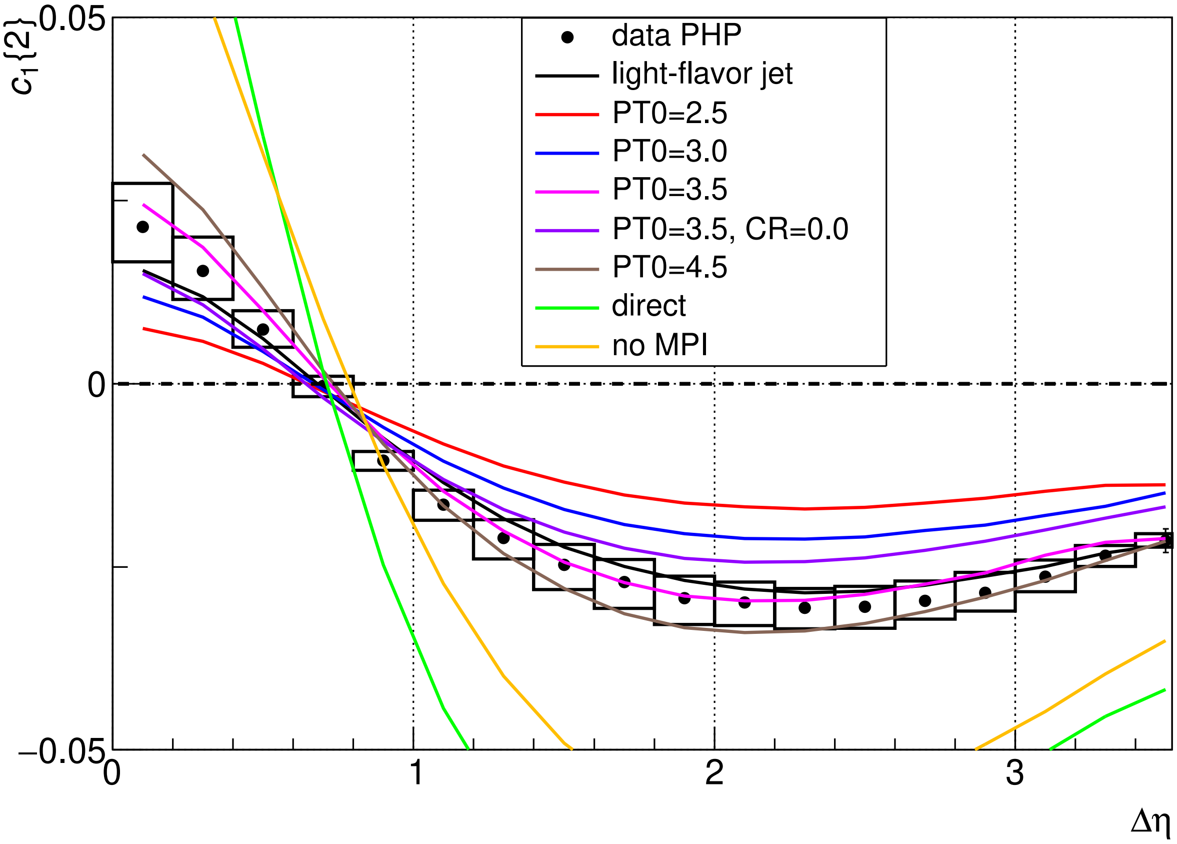


Figure 65: $c_1\{2\}$ versus $\Delta\eta$.

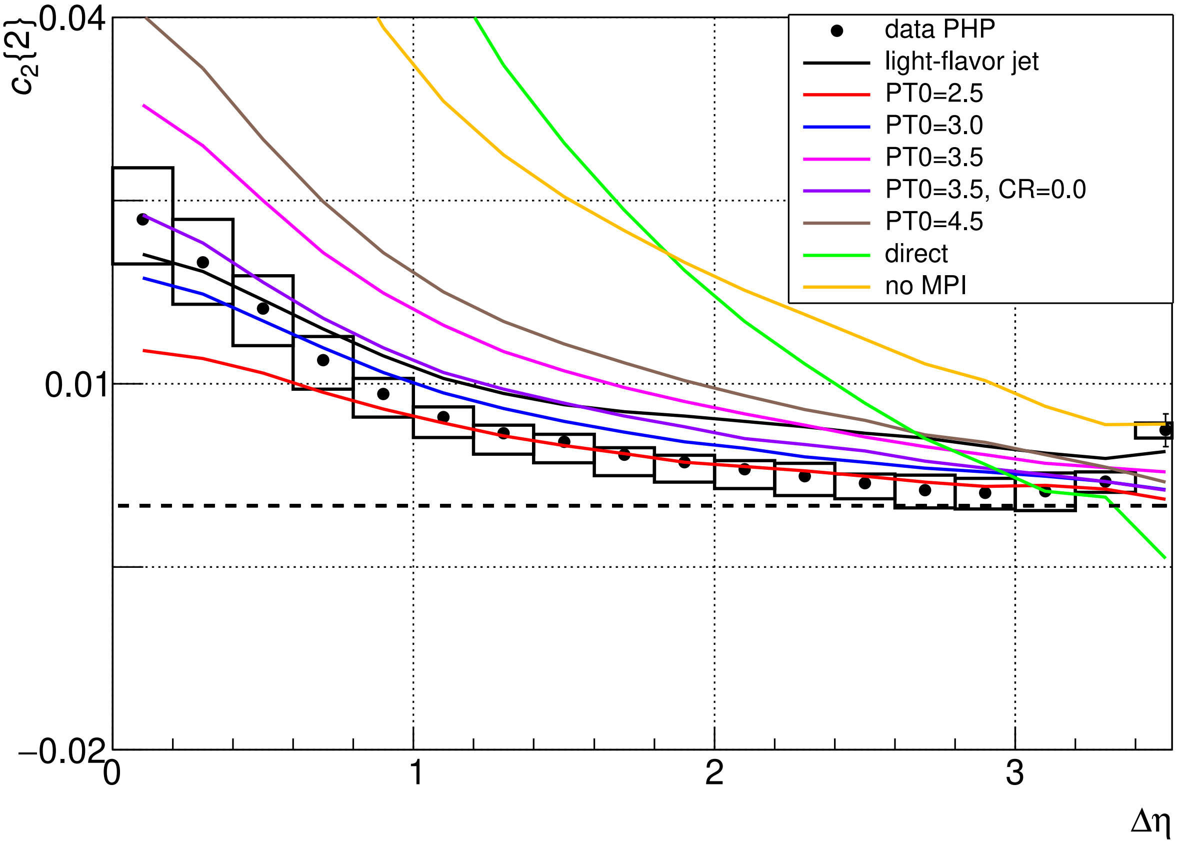


Figure 66: $c_2\{2\}$ versus $\Delta\eta$.

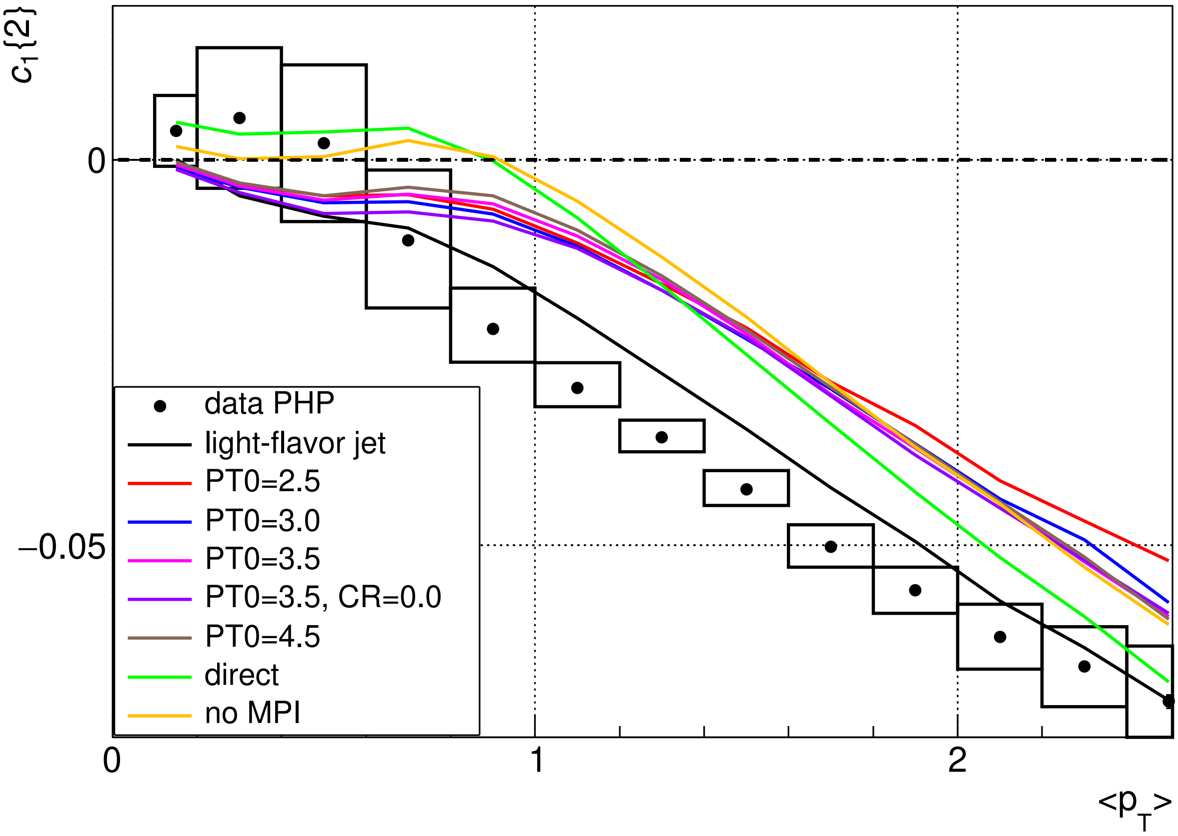


Figure 67: $c_1\{2\}$ versus $\langle p_\mathrm{T} \rangle$.

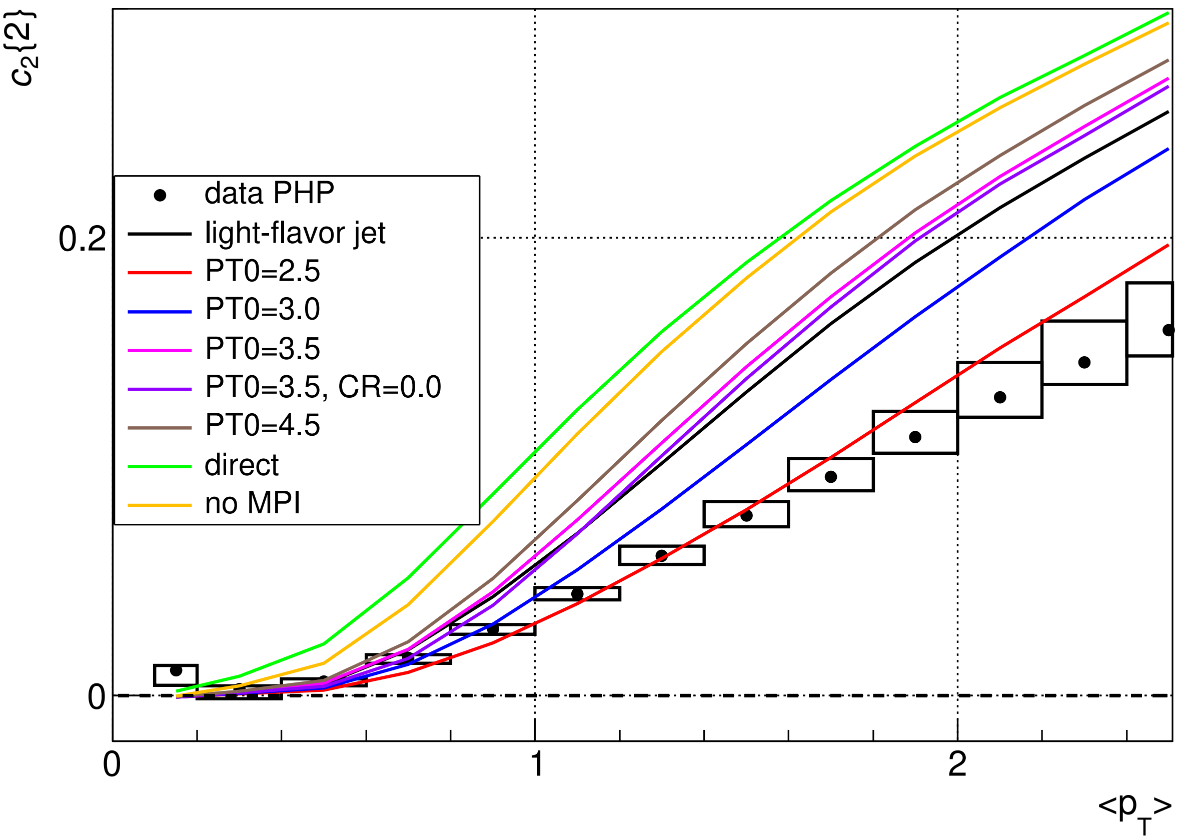


Figure 68: $c_2\{2\}$ versus $\langle p_{\mathrm{T}} \rangle$.

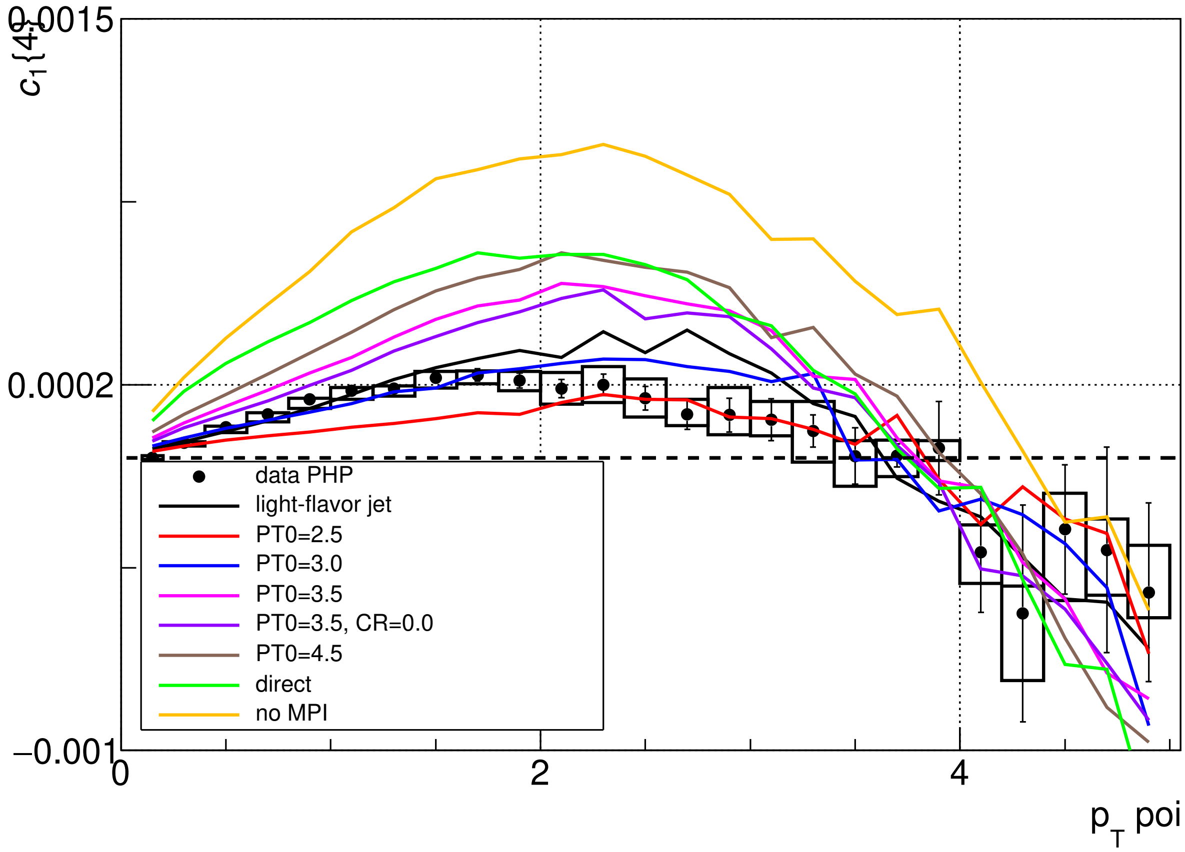


Figure 69: $c_1\{4\}$ versus p_{T} poi.

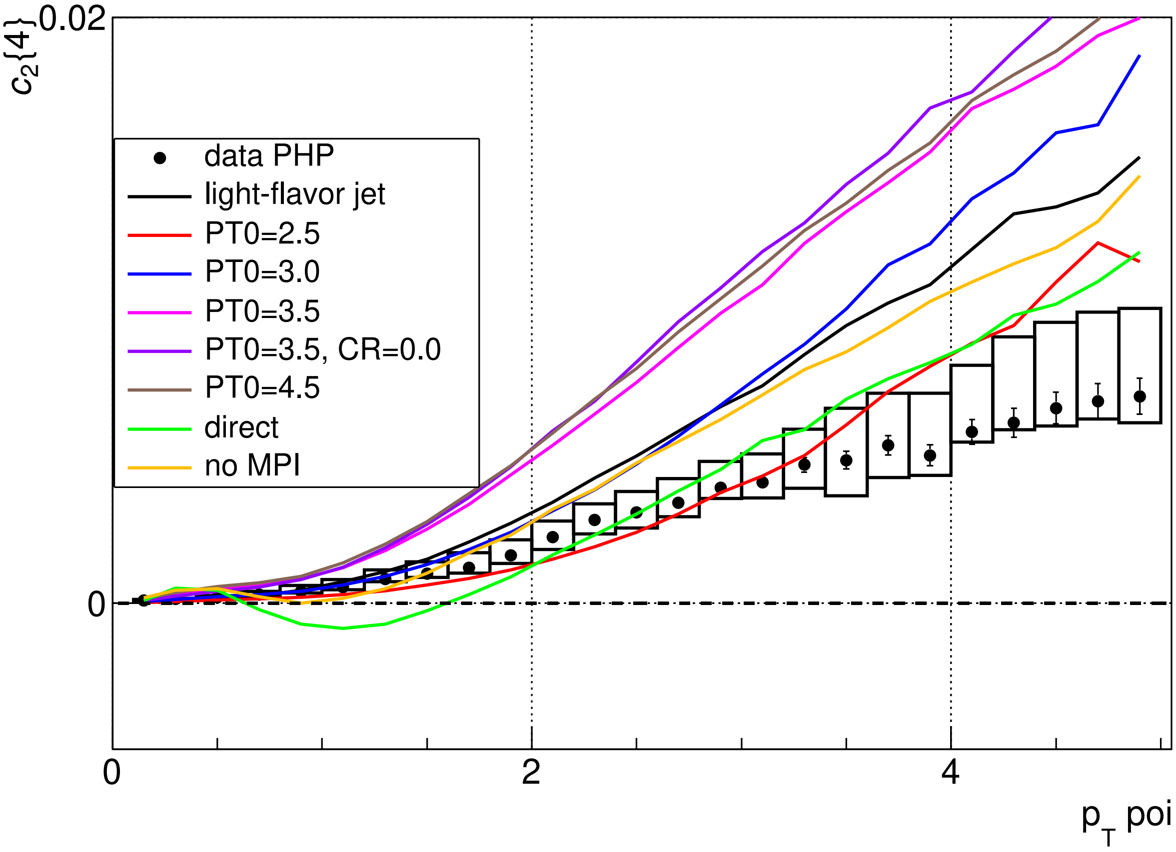


Figure 70: $c_2\{4\}$ versus p_T poi.

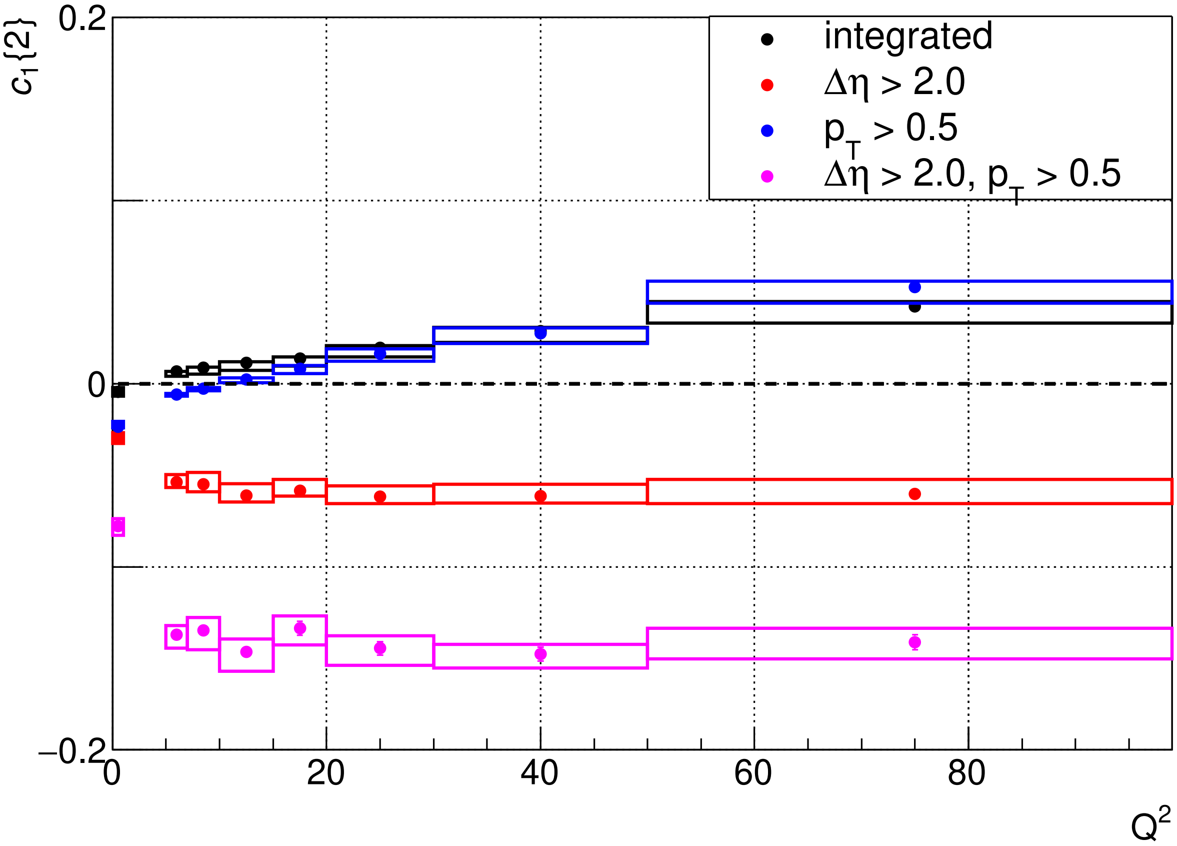


Figure 71: $c_1\{2\}$ versus Q^2.

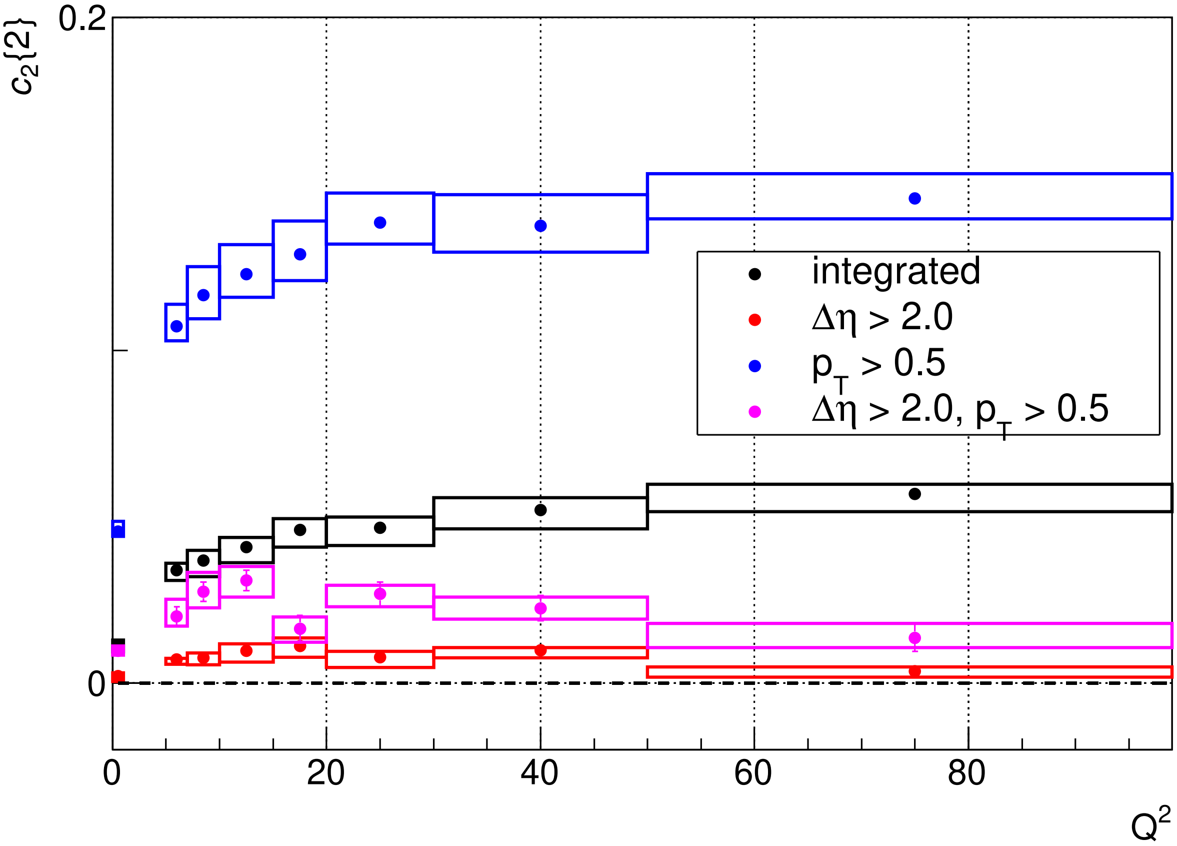


Figure 72: $c_2\{2\}$ versus Q^2.

15 Summary and outlook

To be written for the paper and copied here...

A Instructions to perform the analysis

The analysis code is compiled and run on the NAF batch farm at desy (ssh username@naf-zeus12.desy.de).

Analysis code:

- main.c
  Runs the code (compile with build.sh)
- configTask.C
  Configuration of analysis task parameters.
- cumulantAnalysis.C
  main analysis file with nested for-loops
- cumulants.h
  Definition of cumulant formulas which are used to calculate 2- and 4-particle cumulants without nested for-loops
- histogramManager.cxx
  Defines event and track class histograms.
- orangeAnalyser.cxx
  MakeClass of the orange trees, only branches used in analysis are activated and orangeAnalyserFriend is added as friend class.
- orangeAnalyserFriend.cxx
  Contains getters for tree variables, as well as functions to define the trigger selection, event quality selection, track selection, MC weights and various MC helper functions.
- QCumulants.cxx
  The file where the most correlations are calculated and histograms filled. Ridge histograms are filled in cumulantAnalysis.C
- build.sh
  Compiles the code

- submitJobsCondor.py
  Submits jobs to new cluster: python submitJobsCondor.py <number of files per job> <maximum number of jobs>
  Configure the data to be run over, the output folder and files to be copied.
- nafgo2.sh
  Executes the code on the cluster.

  Calculate weight histograms:
  efficiency/EffCorrections.sh Executes calculation of weight histograms efficiency/EffCorrections.C
  Change the timestamps for the data/mc histograms to be used for corrections

After building the code with build.sh, test code with

```
./readTupleexecutable.o <"DATA"/"MC"> <version> <data sample string>
<number of files per job> <number of jobs> <output directory>
```

For example:

```
./readTupleexecutable.o "MC" "v08b" "ari_incl_nc_DIS_lowQ2_06e" 14 1 $PWD
```

After running code on the farm, check, merge and clean up output with:
python manageJobs.py <timestamp>

A.1 Process analysis output

- produceComponentPlots.C
  Contains the main processing/reduction of histograms. Output all in one file results.root
- runsystematics.C
  Calculates the systematic uncertainties, outputs the histograms in resultsys.root
- Plot_figures.C
  Draws the figures

A.2 Example walk-through

- Checkout git repository on nafhh-x1: git clone https://gitlab.com/ISOQUANT/ZEUS.git. Switch to naf-zeus11. Navigate to analysisCode/.

- Compile makelist.c: g++ makelist.c ‘root-config –glibs –cflags‘ -g

- Build code with ./build.sh

- For tracking efficiency we have to run on MC, select "MC" and a full dataset in submitJobsCondor.py. Modify output path in the file as well, e.g. /nfs/dust/zeus/group/<username> Submit with "python submitJobsCondor.py 25".

- Monitor status of jobs with "condor_q <username>"

- Merge output with "python manageJobs.py <timestamp>", in the directory where all the folders with timestamps are written

- Go to zeus/analysis/efficiency/EffCorrections.C and fill in the appropriate “eventclass" and “AllperiodsMC". run ’source EffCorrections.sh <timestamp>

- In configTask.sh, set the desired weight file and set SetUseEffWeights to kTRUE

- In submitJobsCondor.py, select "DATA" and dataset and submit again "python submitJobsCondor.py 10"

- Merge output with "python manageJobs.py <timestamp>"

B External links

- Github webpage: code
- Description of ZEUS data/MC trees: ZEUS root variables
- MC ISTHEP number: ISTHEP

- ALICE definition of primary particles: ALICE primaries.

References

[1] S. Voloshin and Y. Zhang. Flow study in relativistic nuclear collisions by Fourier expansion of Azimuthal particle distributions. *Z. Phys.*, C 70:665–672, 1996.

[2] Jean-Yves Ollitrault. Determination of the reaction plane in ultrarelativistic nuclear collisions. *Phys. Rev.*, D48:1132–1139, 1993.

[3] Guang-You Qin, Hannah Petersen, Steffen A. Bass, and Berndt Muller. Translation of collision geometry fluctuations into momentum anisotropies in relativistic heavy-ion collisions. *Phys. Rev.*, C82:064903, 2010.

[4] Derek Teaney and Li Yan. Triangularity and Dipole Asymmetry in Heavy Ion Collisions. *Phys. Rev.*, C83:064904, 2011.

[5] Zhi Qiu and Ulrich W. Heinz. Event-by-event shape and flow fluctuations of relativistic heavy-ion collision fireballs. *Phys. Rev.*, C84:024911, 2011.

[6] N. Borghini et al.. Flow analysis from multiparticle azimuthal correlations. *Phys. Rev.*, C 64:054901, 2001.

[7] R. Snellings A. Bilandzic and S. Voloshin. Flow analysis with cumulants: Direct calculations. *Phys. Rev.*, C 83:044913, 2011.