Measurement of isolated-prompt photons in $p+p$ collisions at $\sqrt{s} = 200$ GeV with the sPHENIX detector

Yeonju Go, Shuhang Li, Blair Seidlitz, Justin Bennett, Muhammad Elsayed

Abstract

This note details the measurement of the isolated-prompt photon cross-section as a function of transverse energy (E_T^γ) in proton-proton ($p+p$) collisions at $\sqrt{s}$ = 200 GeV taken with the sPHENIX detector in 2024 at the Relativistic Heavy Ion Collider. Photons are identified using the sPHENIX Electro-Magnetic Calorimeter (EMCal) and the isolation selection utilizes both the EMCal and the Hadronic Calorimeters. Kinematic selections on the photons include $|\eta_\gamma| < 0.7$ and $10 < E_\mathrm{T}^\gamma < 30$ GeV, and an isolation energy requirement $E_\mathrm{T}^{\mathrm{iso,truth}} < 4$ GeV within an isolation cone of $\Delta R = 0.3$. Final cross-sections are unfolded for detector effects and are compared to theoretical calculations from next-to-leading-order pQCD and Monte Carlo generators PYTHIA-8 and JETPHOX.

Contents

1   Introduction

Prompt photons are photons produced either directly from parton-parton scattering, so-called *direct photons*, or from the collinear fragmentation of a final-state parton, so-called *fragmentation photons*. As shown in Figure 1, at leading order (LO), direct photons are produced predominantly from quark–gluon Compton scattering and quark–antiquark annihilation processes. At next-to-leading order (NLO), additional contributions come from final-state fragmentation and radiation into photons.

Because their production can be calculated within the framework of perturbative Quantum Chromodynamics (pQCD), prompt photons provide a stringent test of pQCD predictions. Furthermore, prompt photon production in proton–proton (p+p) collisions is particularly sensitive to the gluon parton distribution function (PDF) in the proton when the quark–gluon Compton process is dominant, providing insight into gluon dynamics in hadronic collisions.

To reduce the contribution of fragmentation photons, an isolation requirement can be imposed. This involves restricting the amount of hadronic energy surrounding the photon within a fixed cone in pseudorapidity (η)–azimuth (ϕ) space. As shown in Figure 2, utilizing PYTHIA-8 simulations, requiring an isolation energy of less than 4 GeV within a specific cone size substantially suppresses
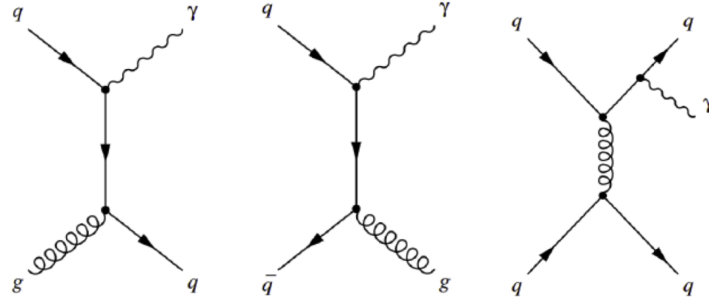
Figure 1: Example Feynman diagrams of (left) quark-gluon Compton scattering, (middle) quark-antiquark annihilation, (right) fragmentation photon

the fragmentation component, thereby enhancing the fraction of direct photons by 75-85% for a range of photon transverse energy (E_T^γ) from 10 to 35 GeV.
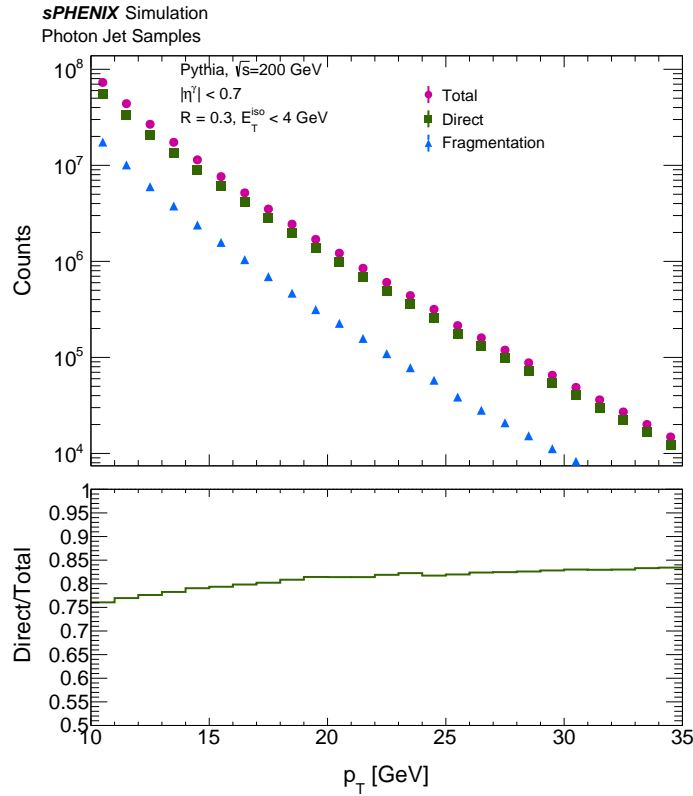
Figure 2: Top panel: photon E_T^γ distributions for direct (green), fragmentation (blue) photons, and total (pink) in PYTHIA-8. Bottom panel: fraction of direct photons.

In this note, we present the measurement of isolated prompt photons in $p+p$ collisions at $\sqrt{s}$ = 200 GeV, using data collected in 2024 with an integrated luminosity of 14.8 pb^{-1}. Section 2 describes the data and MC samples used in this analysis and trigger and event selections. Section 3 describes the procedures for reconstructing and identifying photons. Section 4 details the methods employed to subtract the remaining backgrounds, unfolding, and efficiency correction. Section 5

details the systematic uncertainty evaluation. Finally, in Section 6, the final physics results are presented and compared with both PYTHIA-8 and JETPHOX NLO predictions, as well as with previously published measurements from the PHENIX experiment [1].

2 Data and MC Samples

2.1 Data samples

The $p+p$ Run 2024 data is used for this analysis and summarized in Table 1. Runs are selected from a centrally produced calorimeter "golden run" list, which is selected based on the subsystem validation offline triage database and described in Section 2.3.

Run Numbers	Production tag	CDB tag	DST type
47289-53880	ana462	2024p010_v001	DST_JETCALO

Table 1: List of data sample used in this analysis with production and CDB tag information.

2.2 Monte Carlo Samples

PYTHIA-8 Monte Carlo (MC) [2] events are generated for signal and background photons, respectively, and simulated through GEANT-4 [3] with the sPHENIX detector geometry. In PYTHIA-8 configuration files:

https://github.com/sPHENIX-Collaboration/calibrations/blob/master/Generators/JetStructure_TG/

both *HardQCD:all = on* and *PromptPhoton:all = on* processes are turned on for both signal (labeled as "photon") and background (labeled as "jet") samples. After events are generated from PYTHIA-8, there is a truth-level event filtering: jet samples require a truth jet within $|\eta| < 1.5$ for p_T^{jet} greater than a threshold (e.g., 10, 20, and 30 GeV), while photon samples require a truth photon within $|\eta| < 1.5$ with p_T^{γ} greater than a threshold (e.g., 5, 10, and 20 GeV). Multiple samples are generated with different minimum p_T thresholds to ensure high statistics in the high-p_T region. Each sample contains 10 million processed events. Table 2 shows the list of MC samples used in this analysis. Samples with different p_T thresholds are combined with each sample's effective cross-section. The effective cross sections are calculated using the following equation:

$$\sigma_{\text{eff}} = \frac{\sigma_{\text{generator}} N_{\text{accepted}}}{N_{\text{generated}}} \tag{1}$$

where $\sigma_{\text{generator}}$ is the cross-section for generating events given the generator configurations ($\hat{p}_{T,\text{min}}$), $N_{\text{generated}}$ is the number of events generated by the generator, and N_{accepted} is the number of events that pass the truth-level event filtering. Figure 3 shows the ratio between PYTHIA-8 photon samples normalized by the effective cross-section, the leading photon E_T^{γ} energy cut-off for each sample are chose to be at where the sample is fully efficient.
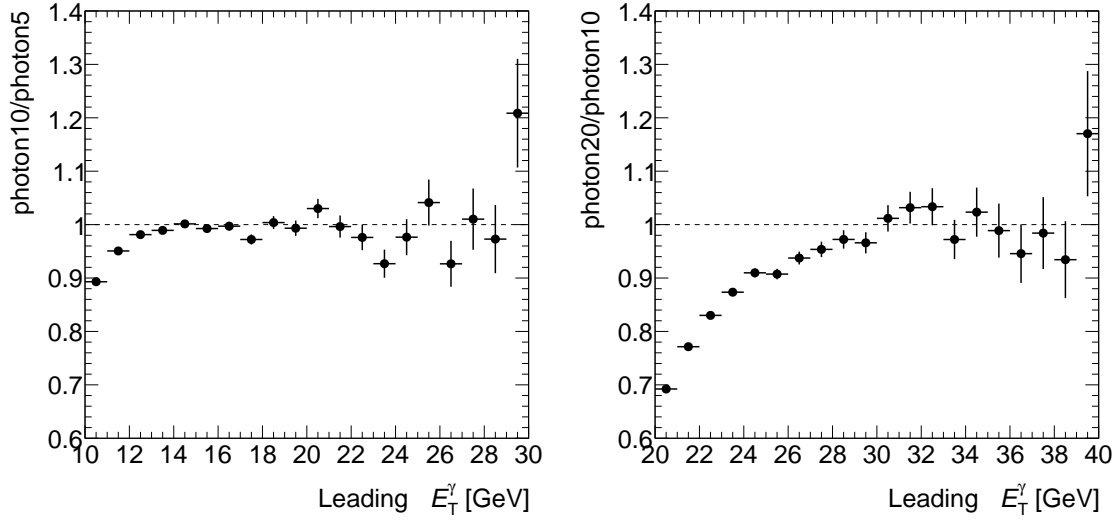
Figure 3: Leading photon E_T^γ distributions from PYTHIA-8 photon 10 GeV(20 GeV) normalized by the effective cross-section over photon 5 GeV(10 GeV) samples. Showing the photon 10 GeV(20 GeV) sample is fully efficient at 14 GeV(30 GeV) leading photon E_T^γ.

| Generator	Production tag	CDB tag
PYTHIA-8-8	run 21 type jet10	MDC2
PYTHIA-8-8	run 21 type jet20	MDC2
PYTHIA-8-8	run 21 type jet30	MDC2
PYTHIA-8-8	run 21 type photon5	MDC2
PYTHIA-8-8	run 21 type photon10	MDC2
PYTHIA-8-8	run 21 type photon20	MDC2

Table 2: Summary of MC samples used.

Figure 4 shows the per-event leading photon E_T^γ distribution after the simulation sample combining process. The smooth transition around each threshold confirms the validity of this methodology. The same combining strategy is applied to the jet samples as well.

In this note, distributions or results constructed using PYTHIA-8 photons with truth-level signal photon selections (see details in Section 4) are referred to as "signal MC". Similarly, those using PYTHIA-8 jet samples with anti-selection against truth-level signal photons are referred to as "background MC". The PYTHIA-8 jet samples without any selections are referred to as "inclusive MC".

2.3 Trigger

All runs in this analysis were taken after the Level-1 trigger firmware update at run 47289 are subjected to basic calorimeter quality assurance (QA) and required to have more than one million recorded events, a runtime of at least five minutes, and an MBD NS ≥ 1 live time above 80%. Among those runs, the active triggers are identified by checking if specific triggers of interest

6
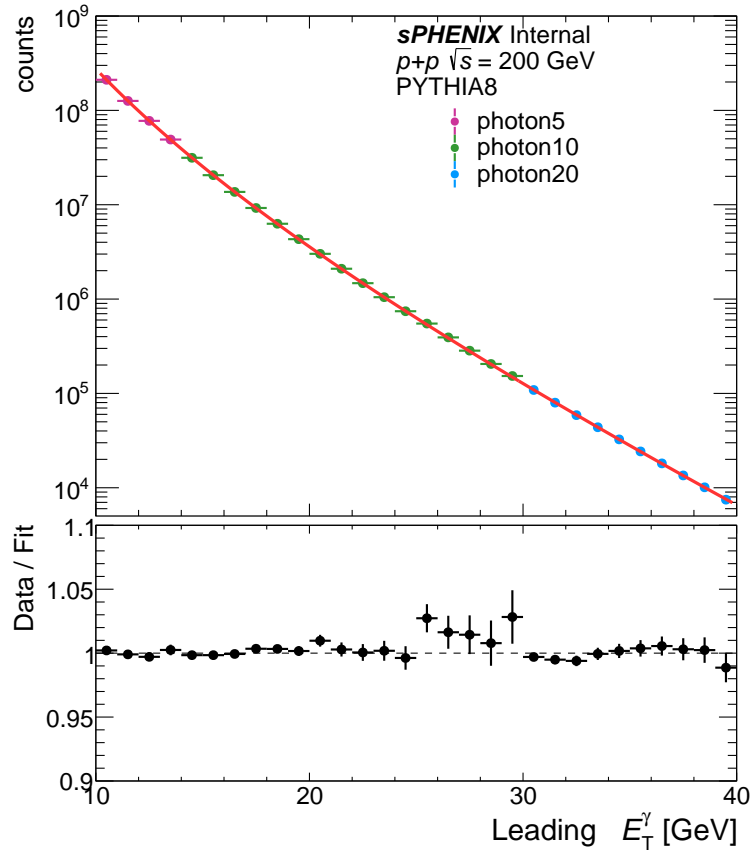
Figure 4: Leading photon E_T^γ distributions from PYTHIA-8 photon 5, 10, 20 GeV samples normalized by the effective cross-section. The combined spectrum is fitted with a modified power law function [4], the MC/fit ratio is shown in lower panel.

have a prescale $\neq -1$, including MBD NS ≥ 1 alone (at least one PMT firing in the north and south MBD) and each "Photon N GeV + MBD NS ≥ 1" variant (for $N = 3, 4, 5$). In each qualifying run, a maximum cluster energy histogram is filled via combining the MBD NS ≥ 1 (MBD NS) raw bits with the live bits of the Photon $3, 4, 5$ GeV + MBD NS ≥ 1 triggers. Specifically, each event is first required to fire the MBD NS raw bit, after which the maximum cluster energy is recorded in a histogram associated with that trigger. In the same event, the live bits of each photon trigger are checked; if any are active, the corresponding photon-trigger histograms are also filled. This method circumvents the need for run-by-run prescale corrections that would otherwise be necessary when using only scaled trigger bits. The MBD NS ≥ 1 trigger typically has a prescale, or ratio of live-to-scaled trigger counts, of about 2000, whereas the photon triggers have prescales closer to 1–4, where the highest photon trigger have no scale-down for most of the time.

When all histograms from these runs are combined, the efficiency of a photon trigger is determined on a bin-by-bin basis by dividing its scaled histogram by the scaled MBD NS ≥ 1 histogram, thus revealing how effectively each photon threshold selects events relative to the minimum-bias sample. Figure 5 shows the leading electromagnetic calorimeter (EMCal) cluster distributions and

efficiencies of photon triggers. These distributions are fit with the logistic function

$$f(x) = \frac{p_0}{1 + e^{(-p_1(x - p_2))}} \qquad (2)$$

where p_0 is the amplitude (asymptotic plateau), p_1 is the slope, and p_2 is the horizontal offset. In this analysis, the "Photon 4 GeV + NBD NS $\geq$ 1" trigger (trigger bit 26) is used. This analysis reports photon E_T^γ above 10 GeV (unfolding underflow bin starting from 8 GeV) where the trigger efficiency is at a plateau about 99.1%.
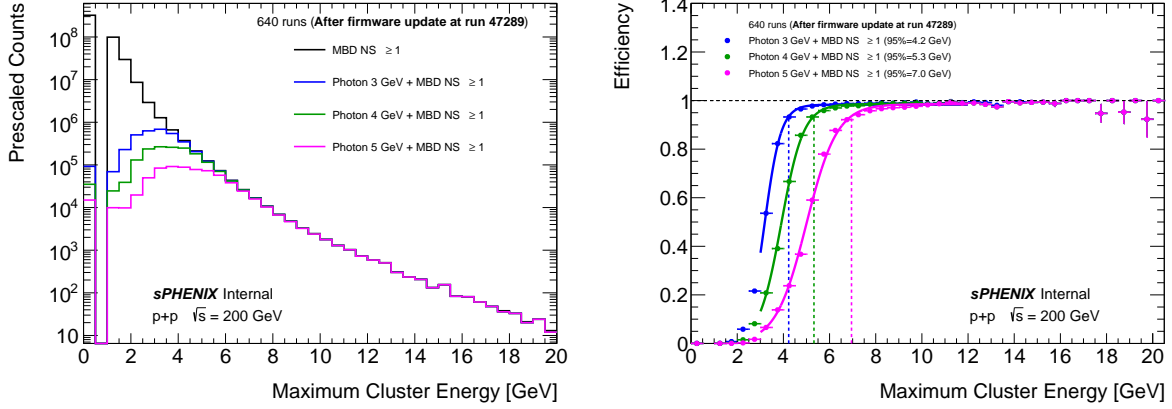
Figure 5: (Left) Prescaled maximum cluster energy distributions for runs passing the stated criteria, illustrating MBD NS $\geq$ 1 and various photon triggers. (Right) The corresponding photon trigger efficiencies computed by dividing each photon histogram by MBD NS $\geq$ 1 on a bin-by-bin basis, where binomial errors are applied to the ratio in each energy bin.

As an additional cross-check of the converged amplitude and how it reflects the Photon 4 + MBD NS $\geq$ 1 trigger efficiency above 10 GeV, the trigger's fit is isolated and refit with a constant function in the $10 - 20$ GeV range. The resulting comparison is shown in Fig. 6.

2.4 Event Selection

Triggered events are further required to pass the event selection criteria as follows:

- $|z_{\text{vtx}}^{\text{MBD}}| < 30$ cm
- MBD hits north $\geq$ 1 and MBD hits south $\geq$ 1

The luminosity passes the trigger and event selection is $14.8^{+0.6}_{-2.1}$ pb^{-1}
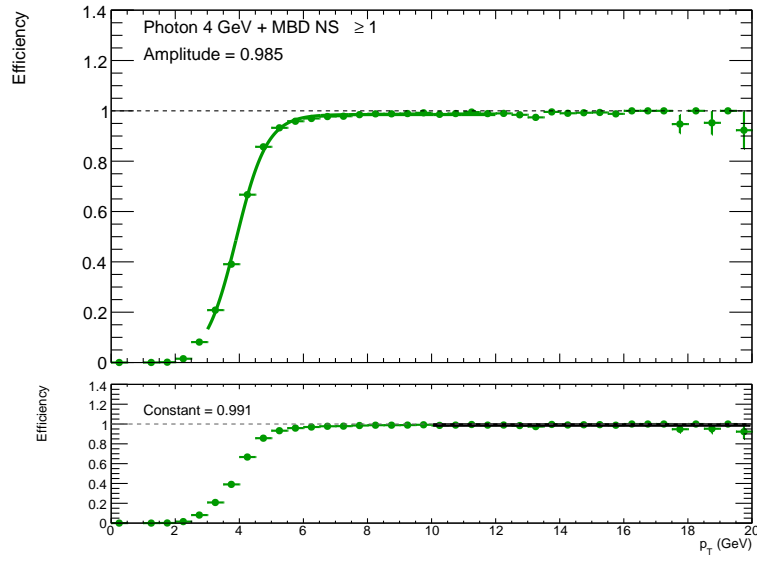
Figure 6: Photon 4 GeV trigger efficiency with respect to the MBD NS ≥ 1 trigger as a function of cluster E_T is shown. The fit of the efficiency is shown for turn-on region (top panel) and for the plateau region (bottom panel). Above 10 GeV, where this analysis takes, the constant fit of the efficiency is at 99.1%

3 Photon Reconstruction

3.1 Clustering Algorithm

A simple clustering routine is used to aggregate EMCal towers into clusters. It is implemented in the RawClusterBuilderTemplate class. Adjacent towers (sharing an edge) above a single threshold (default at 70 MeV) and passing the *isGood* tower quality cuts are clustered together. These clusters grow until no adjacent towers are above the threshold. No cluster splitting is performed. The cluster kinematics are determined in the standard way implemented in RawClusterBuilderTemplate, which relies on tower position in x-y-z space and corrections for shower depth. The output x-y-z is then used to calculate the η and ϕ positions relative to the collision vertex, which is reconstructed with the Minimum Bias Detector (MBD). The reconstructed energy is determined by summing all of the individual tower energies in the cluster.

3.2 Photon Energy Response

Photon energy scale and resolution are estimated using MC. Figure 7 shows the ratio of reconstructed cluster E_T to truth-level photon E_T^{truth} as a function of E_T^{truth}. The ratio of E_T^{reco} to E_T^{truth} in each E_T^{truth} bin is fitted with a Gaussian function – see examples in Figure 8. The long tail at the lower E_T^{reco}/E_T^{truth} region comes from the missing position-dependent energy calibration, which will be addressed in the future. Figure 9 shows the extracted photon energy scale and resolution as a function of E_T^{truth}. Unfolding is performed to account for the remaining resolution and scale (discussed in Section 4.5).
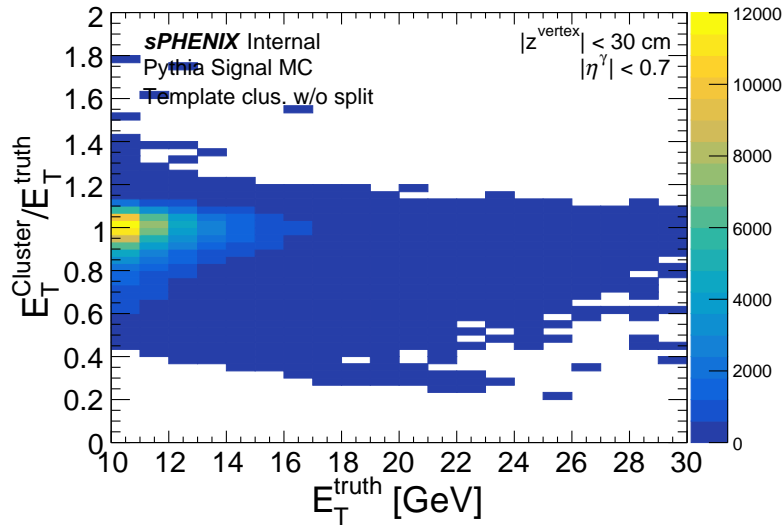
Figure 7: Photon energy response as a function of E_T^{truth} in signal PYTHIA-8.
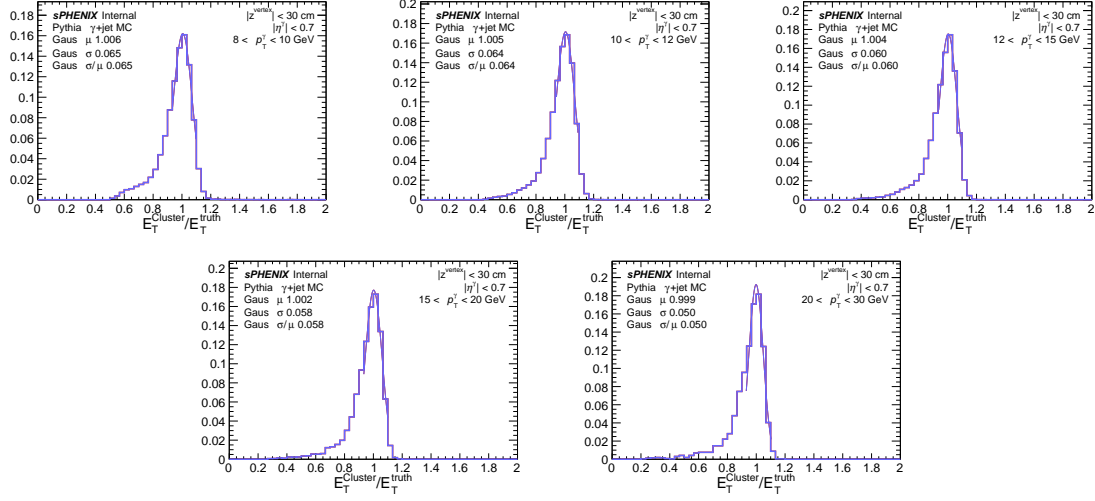
Figure 8: The ratio of E_T^{reco} to E_T^{truth} distributions for different E_T^{truth} bins in different panels. The Gaussian fits are overlaid around the peaks.
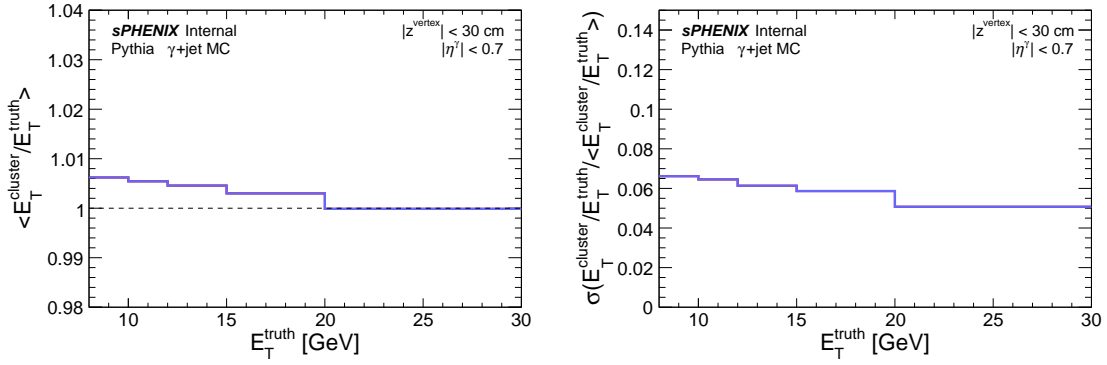
Figure 9: Photon energy scale (left) and resolution (right) as a function of E_T^{truth} in signal PYTHIA-8.
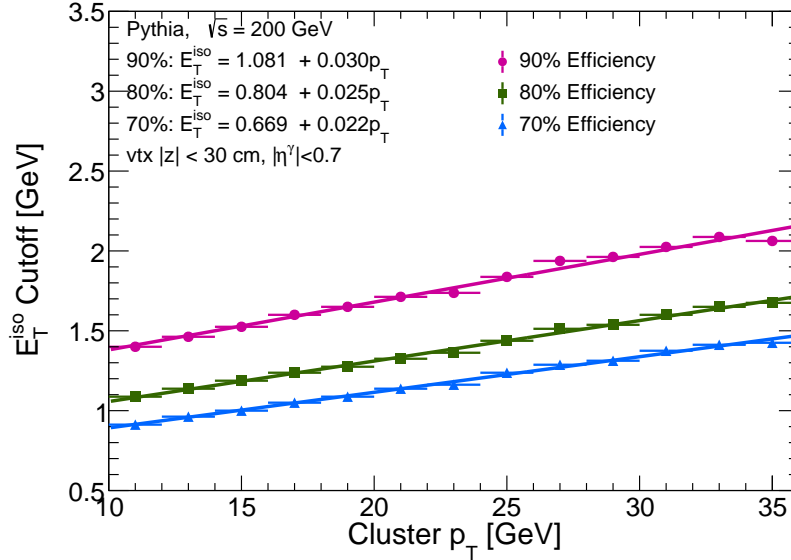
3.3 Isolated Photons

The cluster isolation transverse energy (E_T^{iso}) is calculated by summing the E_T of all EMCal, inner and outer hadronic calorimeter (HCal) towers that pass the *isGood* tower quality selection and within an "isolation cone" of radius $\Delta R = 0.3$, but excluding the E_T^{reco} of the cluster of interest. Because the clustering algorithms used in this analysis disable sub-cluster splitting, large superclusters of decay photons (e.g., from π^0 decays) may naturally exhibit lower isolation energy than split clusters. However, these superclusters can still be distinguished from prompt photons through shower shape variables.

Using signal photon MC, the E_T^{iso} cuts for 70%, 80%, 90% isolation cut efficiency are calculated as a function of E_T^γ, as shown in Figure 10. Each selection is fit to a simple linear function. The fit results from the 90%-efficiency selection is used as the default in this analysis:

$$E_T^{\text{iso}} < 1.08128 + 0.0299107 \cdot E_T^{\text{reco}}$$

Figure 10: Cut value of the isolation energy E_T^{iso} as a function of the cluster E_T^{reco} to obtain 70%, 80%, and 90% prompt-isolated photon efficiency. Each selection is fitted with a linear function.

The non-isolated criteria is chosen to compromise between signal leakage into the non-isolated sideband regions and statistics in the non-isolated sideband regions. A larger gap in E_T^{iso} between isolated and non-isolated reduces leakage but reduces the statistics of the sideband region. To balance the statistics and signal leakage, the non-isolated criteria is chosen to be:

$$E_T^{iso} > 1 + 1.08128 + 0.0299107 \cdot E_T^{reco}$$

Different non-isolated criteria are used as systematic uncertainty for purity.

3.4 Signal and Background Identification

To distinguish signal photons from primarily high-p_T π^0 backgrounds ($\pi^0 \to \gamma + \gamma$) — as well as other neutral meson decays, high-p_T charged particles, and non-physics backgrounds — a set of moments and energy ratios are calculated from the EMCal towers in the cluster. These moments and ratios are collectively referred to as "shower shapes". The definition of the shower shapes used in this analysis is given in Table 3. In general, the decay photons of high-p_T π^0 have a ΔR less than a tower size; thus, they are reconstructed in one cluster. However, the energy may be more spread out in ϕ or η, or the energy of the leading tower may be a smaller fraction of the total cluster energy. The shower shapes encode this information. The shower shape distributions for two example cluster-p_T selections are shown in Figure 46 and Figure 47 for all clusters reconstructed in this analysis. The distributions are compared to the signal MC (see Section 2.2) and background MC. The data more closely follow the background MC, which is expected due to the low signal-to-background ratio without selection. There are significant deviations from either MC, which result from non-physics backgrounds and are further discussed
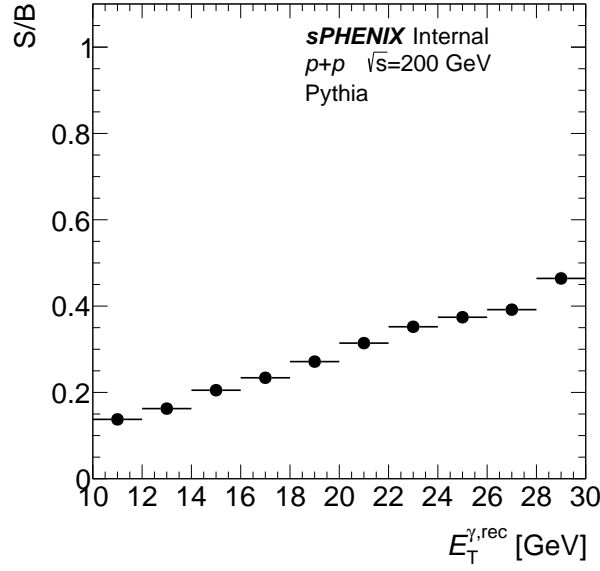
Figure 11: Signal-background ratio as a function of E_T^{reco} derived from PYTHIA

Symbol	Definition
E3x2/E3x5	Ratio of energy in regions of 3 towers wide in η and 2 towers wide in ϕ to the energy in a region 3 towers wide in η and 5 towers wide in ϕ. The towers closest to the center-of-gravity are chosen.
E1x1/E	Center of energy tower divided by cluster energy.
$w^{\text{COGX}}_{\eta(\phi)}$	energy weighted second moment of $\eta(\phi)$ from all towers in a given cluster excluding the tower containing the center of gravity.
$w72$	$\sum E_i(\eta_i - \eta_i^{max})^2 / \sum E_i$ for $(\eta_i, \phi_i) = (7, 2)$ tower area
R_{had}	$E_T^{HCal} / (E_T^{EMCal} + E_T^{HCal})$
E_{t1}	$(E_1 + E_2 + E_3 + E_4)/E_{tot}$
E_{t2}	$(E_1 + E_2 - E_3 - E_4)/E_{tot}$
E_{t3}	$(E_1 - E_2 - E_3 + E_4)/E_{tot}$
E_{t4}	E_3/E_{tot}

Table 3: Definitions of various shower shape variables.

in Section 3.5. The result of this discussion is a set of shower shape cuts designed to remove this non-physics background.

Figure 11 shows the signal background ratio as a function of E_T^{reco} without any shower shape selections, the signal background ratio at 10 GeV is about 15% and increase to about 50% around 30 GeV. Shower shape selections can be used to enhance the fraction of either background or signal. The shower shape criteria used to enhance the fraction of the signal process is referred to as the *tight* selection criteria, and clusters passing these selections are referred to as *tight clusters*. The criteria used to enhance the fraction of the background are referred to as the *non-tight* selection

199 and *non-tight clusters*.

200 The data-driven physics background removal procedure, discussed in Section 4.3, relies heavily on
201 the independence of the E_T^{iso} and the shower shape selections for signal and background. Put more
202 directly, the background physics process in the non-tight control region must have a similar E_T^{iso}
203 distribution as the background contamination in the tight control region. To check this and inform
204 the development of the selection criteria, Figure 48-49 displays the average E_T^{iso}, as a function of
205 shower shape value, for various shower shapes. Shower shape criteria that can improve the signal
206 purity but are highly correlated with the E_T^{iso} must be applied to both tight and non-tight clusters.

207 The shower shapes are shown again in Figure 12 and Figure 13, with a pre-selection, listed in
208 Table 4. With these selections, the non-physics background is greatly reduced, which allows for
209 more detailed comparisons of data and MC (for which the pre-selection is applied to both) - please
210 see the following Section 3.5 for more detail.

211 Using this information, the tight and non-tight selections are defined in Table 4. Figure 14
212 and Figure 15 display the shower shape distributions after the tight and non-tight selections,
213 respectively.

Pre-selection
E11/E33 < 0.98
0.6 < et1 < 1.0
0.8 < E32/E35 < 1.0
weta_cogx < 0.6
Tight
0 < weta_cogx < $0.15 + 0.006 \cdot E_T^{\gamma}$
0 < wphi_cogx < $0.15 + 0.006 \cdot E_T^{\gamma}$
0.4 < E11/E33 < 0.98
0.9 < et1 < 1.0
0.92 < E32/E35 < 1.0
Non-tight
Pass the pre-selections, Fails at least two of the five tight requirements

Table 4: Photon identification criteria
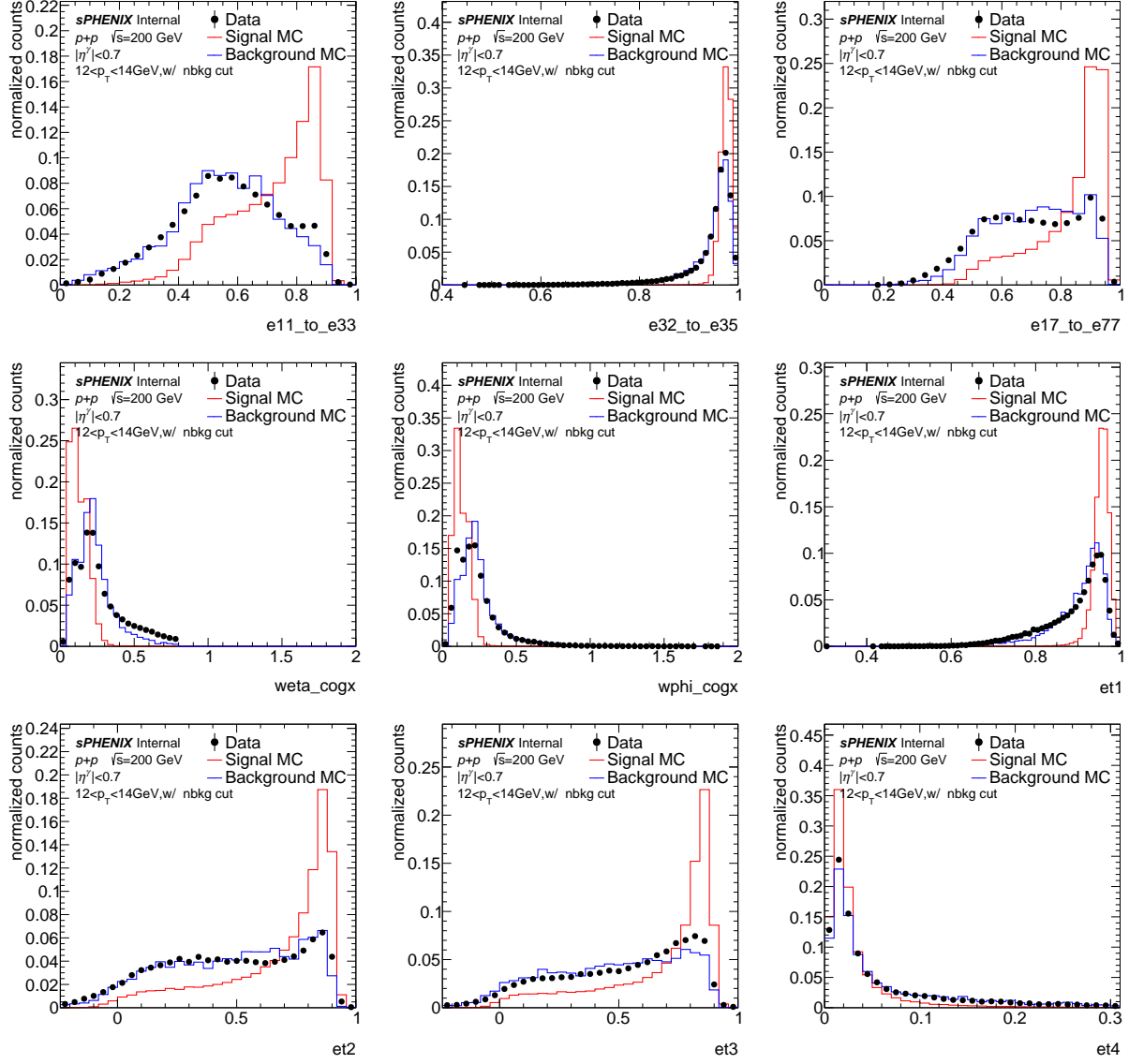
Figure 12: Shower shape distributions with non-physical background removal cut at lower p_T (12-14 GeV).
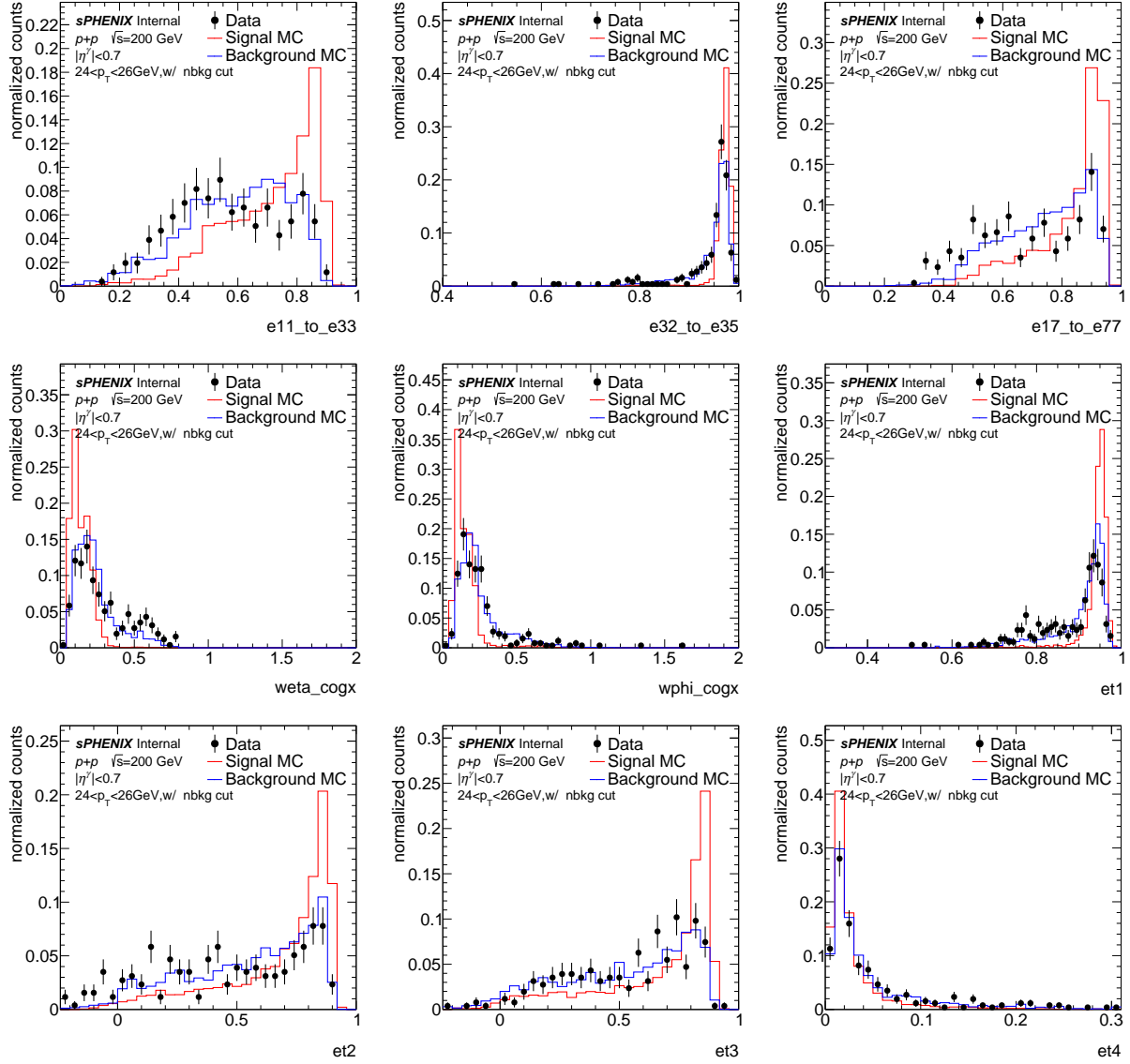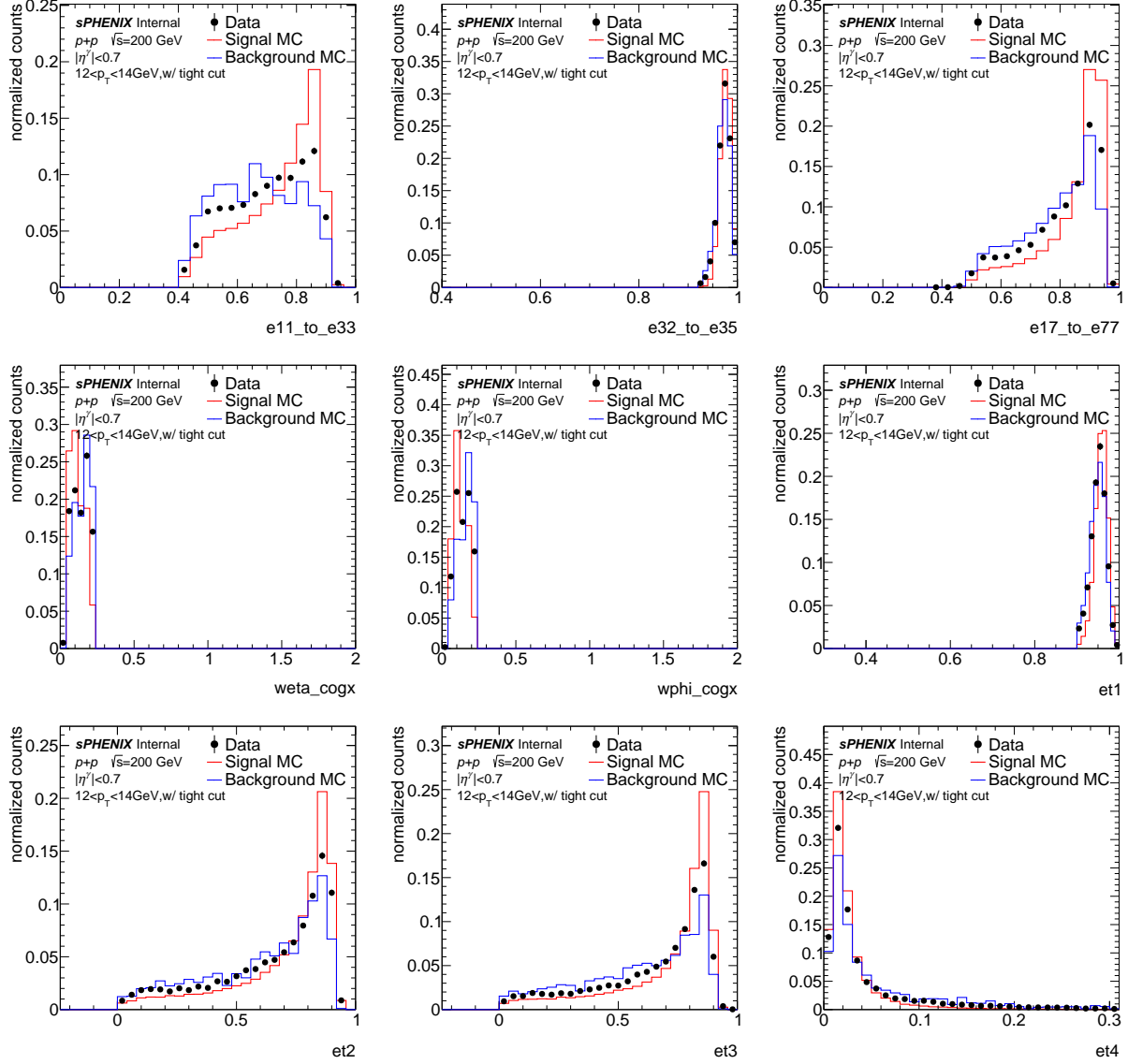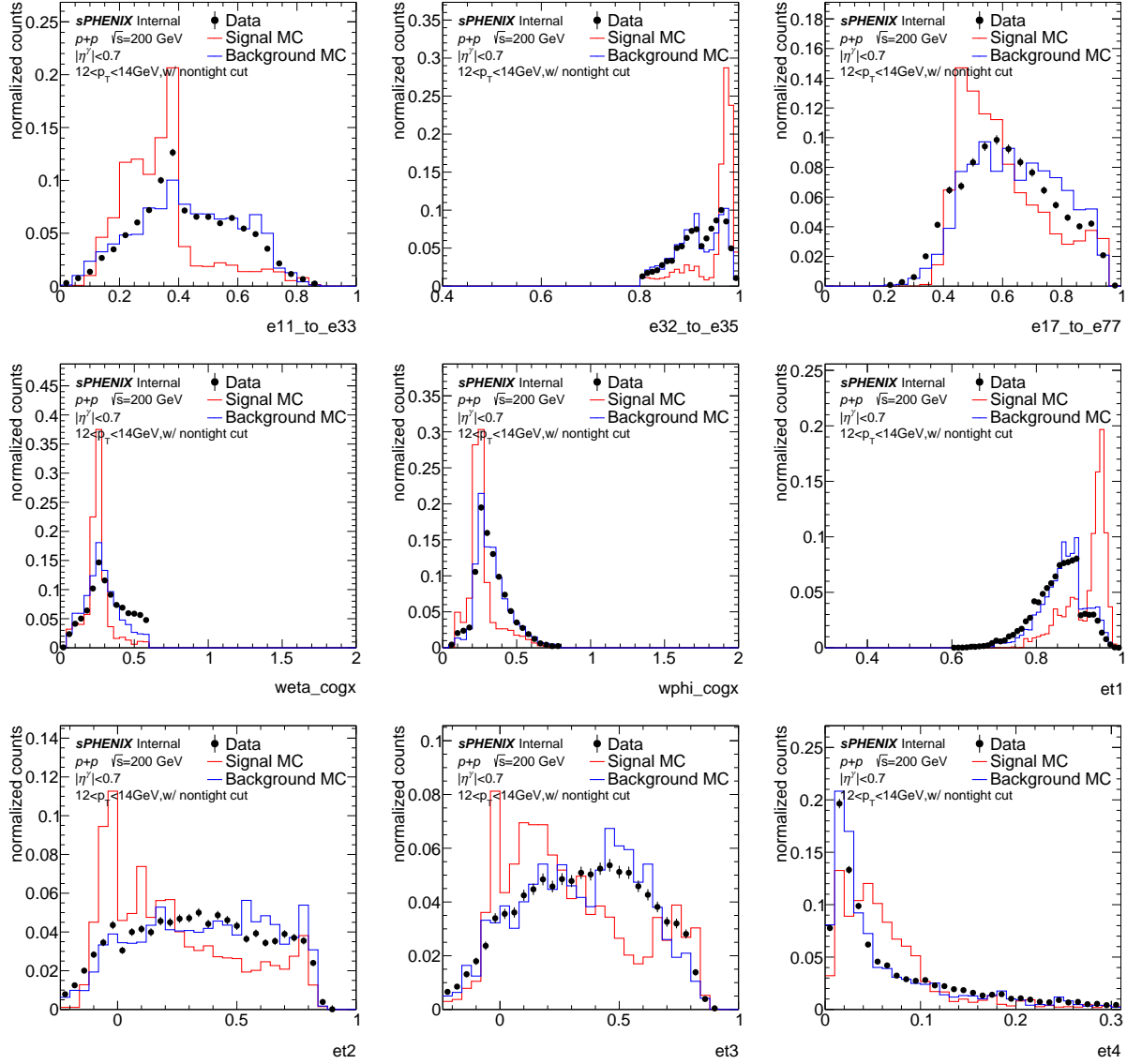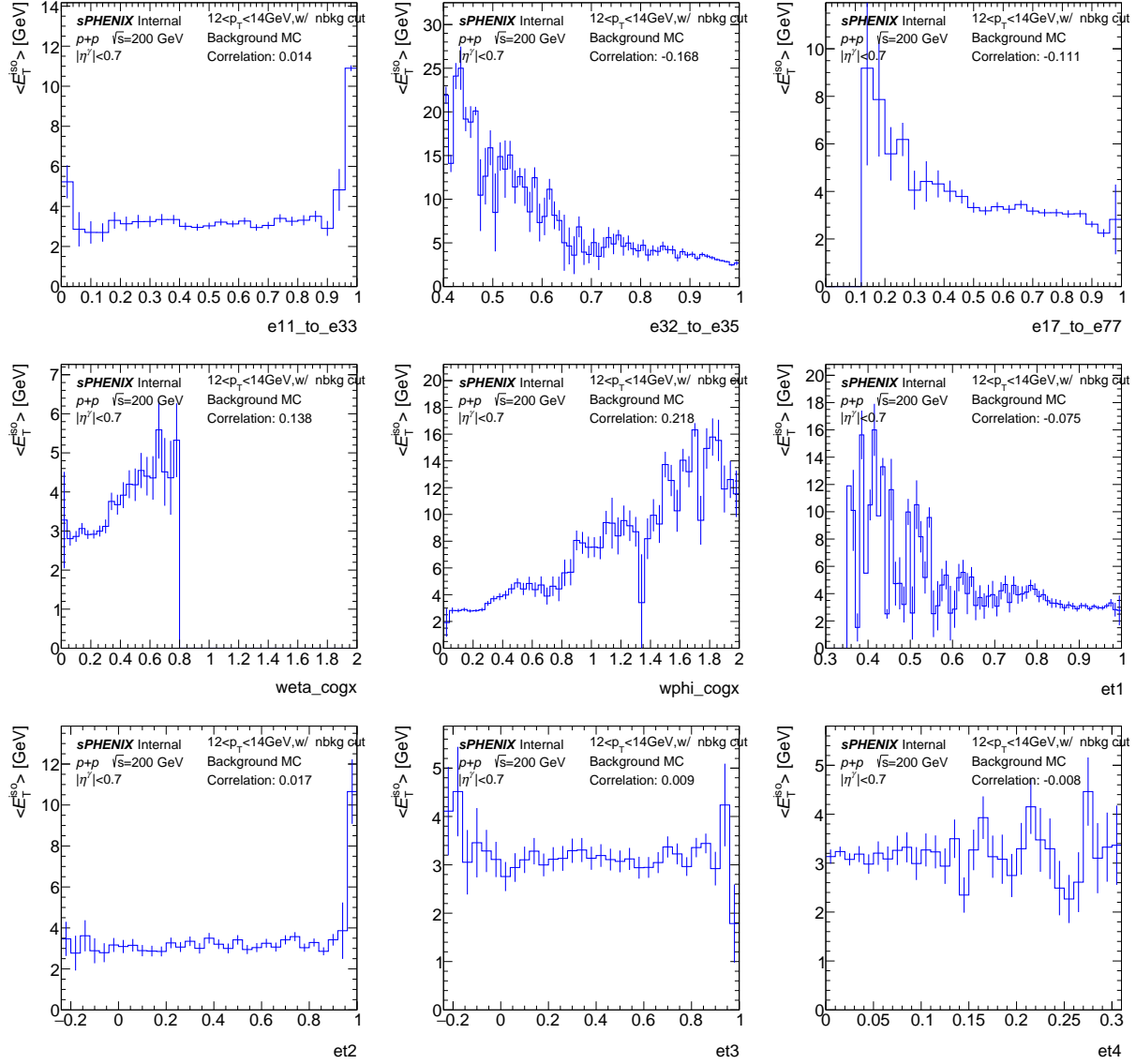
Figure 13: Shower shape distributions with pre-selection at high p_T (24-26 GeV).

**Figure 14:** Shower shape distributions with **tight selection** at lower p_T (12-14 GeV).

Figure 15: Shower shape distributions with **non-tight selection** at lower p_T (12-14 GeV).

Figure 16: The average E_T^{iso} as a function of shower shape value for various shower shapes with non physical background removal cuts at lower p_T (12-14 GeV).

3.5 Non-physics Background Removal

It is possible for beam backgrounds to enter the EMCal and fire the photon triggers. These backgrounds occur at a high enough rate that a coincident physics collision (or other background) also triggering the MBD N&S trigger is non-negligible. These types of events result in a high-energy EMCal cluster that can pass either the photon tight or non-tight selections. In this section, we discuss the nature of this background and derive a cut to remove it.
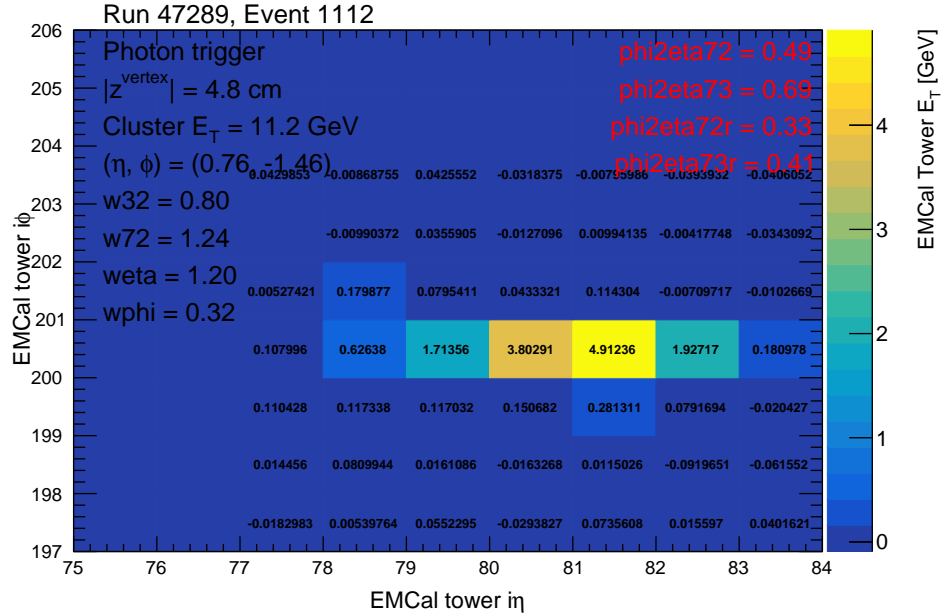
Figure 17: Example tower η-ϕ distribution of a non-physics background cluster (streak event). Tower energies are shown for each tower.

Based on jet background studies [5], it has been found that beam backgrounds traveling parallel with the beam, but at larger radii, deposit significant energy in z-adjacent calorimeter towers. An example of such a background event in the EMCal is shown in Figure 17. Due to their generally extended energy deposit in η, these clusters are discernible from isolated photons. These features are measured with shower shapes, namely weta_cogx and wphi_cogx/weta_cogx=wr_cogx (see Table 3 for definitions). The non-physics backgrounds have a larger spread in η and thus a characteristically large weta_cogx and a small wr_cogx. Because signal photons and high-p_T π^0s are often accompanied by a high-p_T recoil jet, it can be confirmed that these clusters, like the one seen in Figure 17, come from non-physics processes. Figure 18 displays 2D scatter plots of weta_cogx and wr_cogx. If the bottom two panels are considered, the enhancement of low wr_cogx and high weta_cogx seen in events without a balancing recoil jet (right panel) are not present in the events with a recoil jet (left pannel). This demonstrates that this excess is from non-physics backgrounds and a cut on weta_cogx is imposed in the pre-selection, noted in Table 4. In the top panels of Figure 18, one can observe the lack of excess from the non-physics backgrounds in the MC and the high efficiency for signal and background MC events for the weta_cogx imposed in the pre-selection.
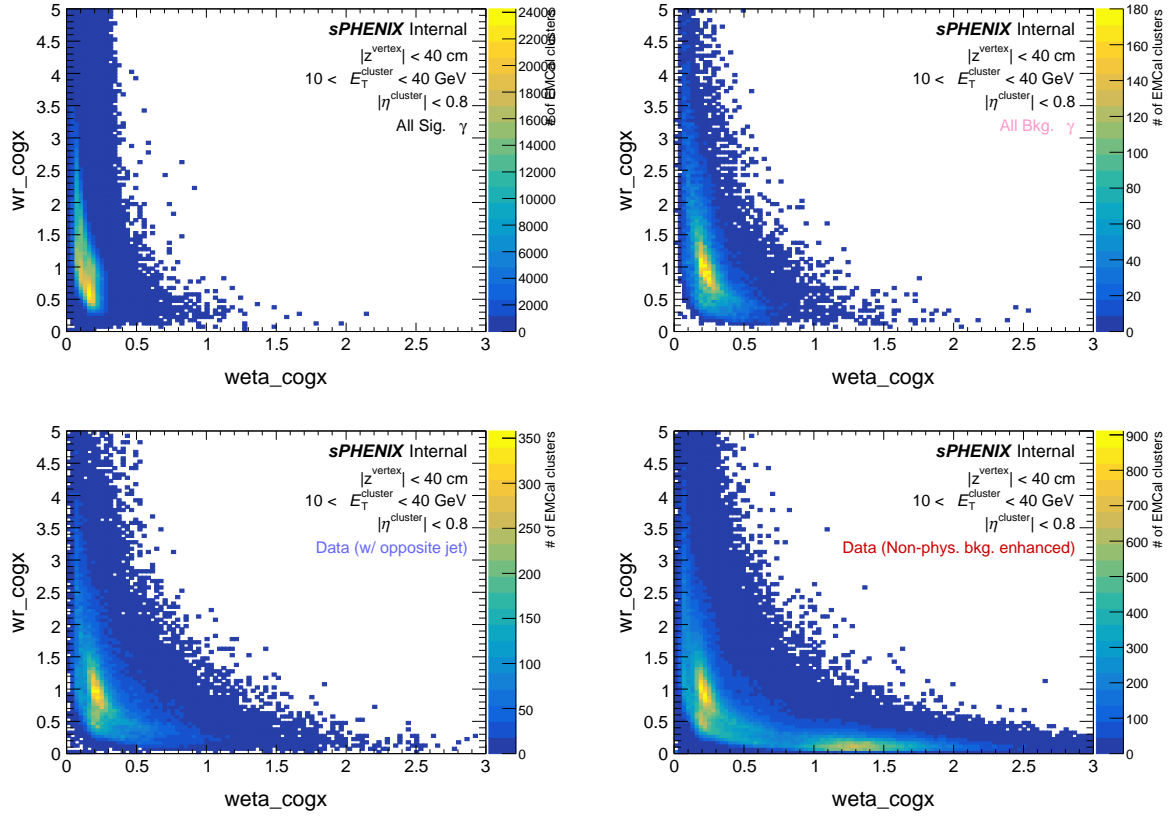
Figure 18: 2D scatter plots of wr_cogx vs. weta_cogx for signal MC (top, left), background MC (top, right), signal-enhanced data by requiring jets in the opposite ϕ direction (bottom, left), background-enhanced data by requiring no jets in the opposite ϕ direction (bottom, right).

4 Analysis Procedure

4.1 Analysis Workflow Summary

The analysis workflow is summarized as following:

1. Cluster EMCal towers using RawClusterBuilder without sub-cluster splitting.

2. Apply photon tight identification cut based on shower shape variables and isolation energy E_T^{iso} cut using an isolation cone of $R = 0.3$ around the cluster.

3. Estimate purity using the data-driven double sideband method and correct for it on a bin-by-bin basis.

4. Unfold the raw (purity-corrected) p_T spectrum using the D'Agostini Bayesian method with RooUnfold.

5. Correct for photon reconstruction, identification and isolation efficiency as a function of truth photon p_T on a bin-by-bin basis.

6. Correct for MBD trigger and event selection efficiency as a function of truth photon p_T on a bin-by-bin basis.

7. Scale by luminosity to obtain the cross section of isolated prompt photons.

4.2 Truth-level Isolated Prompt Photon Definition

Truth-level prompt photons are defined as follows; firstly, final-state particles with pid = 22 in PYTHIA-8 are considered as photons. The final-state photons are also required to be either direct photons or fragmentation photons. Direct photons are defined as photons generated directly from the initial 2-to-2 hard scattering process. Fragmentation photons are defined as photons fragmented from a final-state quark involved in primary hard scattering. The classification of the photon type is achieved by tracking the parton history in the HEPMC record.

The truth-level isolation energy ($E_T^{iso,truth}$) is defined as the sum of the transverse energies of all final-state particles within a cone of $R = 0.3$ in $\eta - \phi$ space around the photon of interest, excluding the photon energy itself. This isolation condition suppresses the contribution of fragmentation photons as well as high-p_T π within jets. Figure 19 (20) shows the fraction of fragmentation (direct) photons as a function of the maximum $E_T^{iso,truth}$ threshold for different photon E_T^γ. At $E_T^{iso,truth}$ threshold of 4 GeV, direct photons at all E_T^γ ranges are fully efficient while the fragmentation photons are suppressed. Hence, in this analysis, isolated prompt photons at the truth-level is defined as:

- $|\eta^\gamma| < 0.7$

- Direct or fragmentation photon with $E_T^{iso,truth} < 4$ GeV, with isolation cone of $\Delta R = 0.3$.

For truth-reconstructed photon matching, a truth photon are required to be matched with a reconstructed cluster within $|\eta| < 0.7$ with $p_T^{reco} > 5$ GeV within $\Delta R < 0.05$, and being the best match particle based on CaloRawClusterEval module.
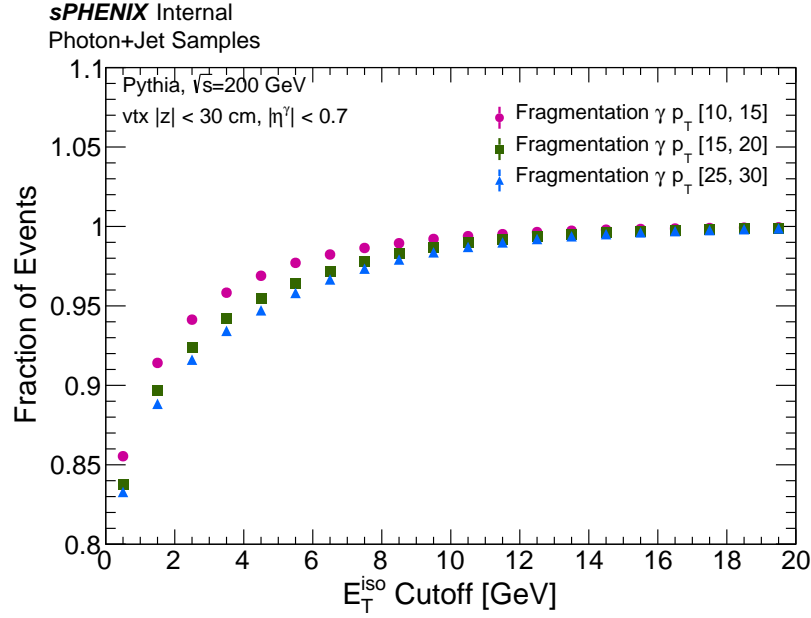
Figure 19: Fraction of fragmentation photons passing the truth-level E_T^{iso} cut as a function of the cut threshold for three different truth photon E_T^γ ranges.
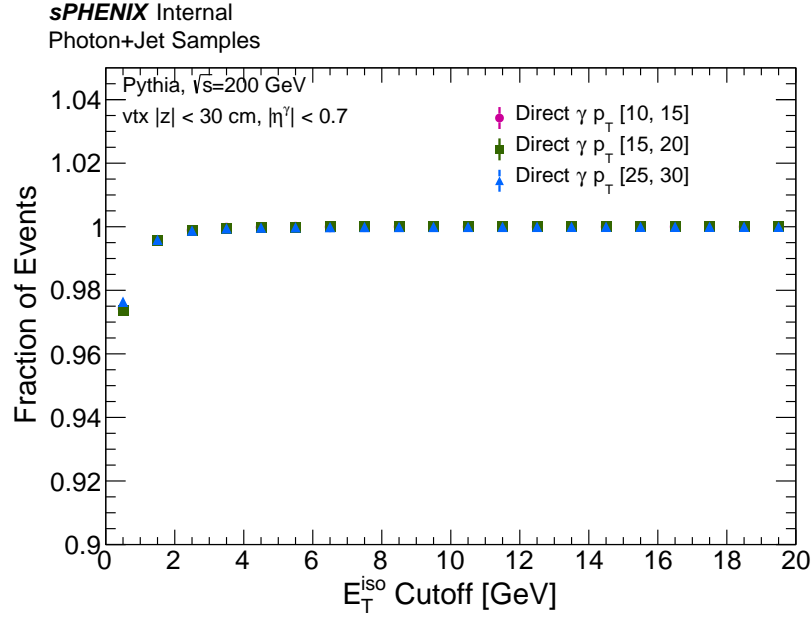
Figure 20: Fraction of direct photons passing the truth-level E_T^{iso} cut as a function of the cut threshold for three different truth photon E_T^γ ranges

4.3 Purity and Background Subtraction

4.3.1 Purity Estimation

Even after applying the tight identification and isolation requirements on EMCal clusters, there are significant contributions from background photons from high-p_{T} neutral mesons decaying to two photons and reconstructed as one EMCal cluster. The remaining background in our signal criteria is estimated and corrected for. The purity ($\mathcal{P}$) is defined as the fraction of signal photons (1−fraction of background photons), and is estimated through a data-driven double-sideband method. In this method, 4 regions (1 signal region, 3 sideband regions) are defined as:

- A: Signal region, clusters passing tight identification cut and isolation $E_{\mathrm{T}}^{\mathrm{iso}}$ cut
- B: Control region or sideband region, clusters passing tight identification cut and failing $E_{\mathrm{T}}^{\mathrm{iso}}$ cut
- C: Control region or sideband region, clusters passing non-tight identification cut and passing $E_{\mathrm{T}}^{\mathrm{iso}}$ cut
- D: Control region or sideband region, clusters passing non-tight identification cut and failing $E_{\mathrm{T}}^{\mathrm{iso}}$ cut

and illustrated as a diagram shown in Figure 21.
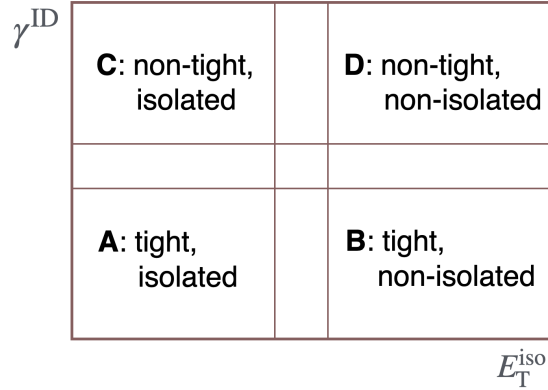
Figure 21: Cartoon of the sideband region definitions.

Figure 22 shows the $E_{\mathrm{T}}^{\mathrm{iso}}$ distributions for data with tight and non-tight selections, compared to signal MC. The difference in $E_{\mathrm{T}}^{\mathrm{iso}}$ distributions between tight and non-tight selected data is expected since the tight selection is supposed to include more signal photons while the non-tight selects more background photons. Figure 23 compares the $E_{\mathrm{T}}^{\mathrm{iso}}$ for tight and non-tight selected data to those in the inclusive MC sample, the general feature that there are higher $E_{\mathrm{T}}^{\mathrm{iso}}$ clusters with the non-tight selections agrees between data and MC.

Under the assumption that the isolation $E_{\mathrm{T}}^{\mathrm{iso}}$ is largely uncorrelated with the shower shape variables used in the tight identification criteria, the ratio of background clusters in region C over D should be similar to the ratio of background in region A over D. Therefore the amount of signal
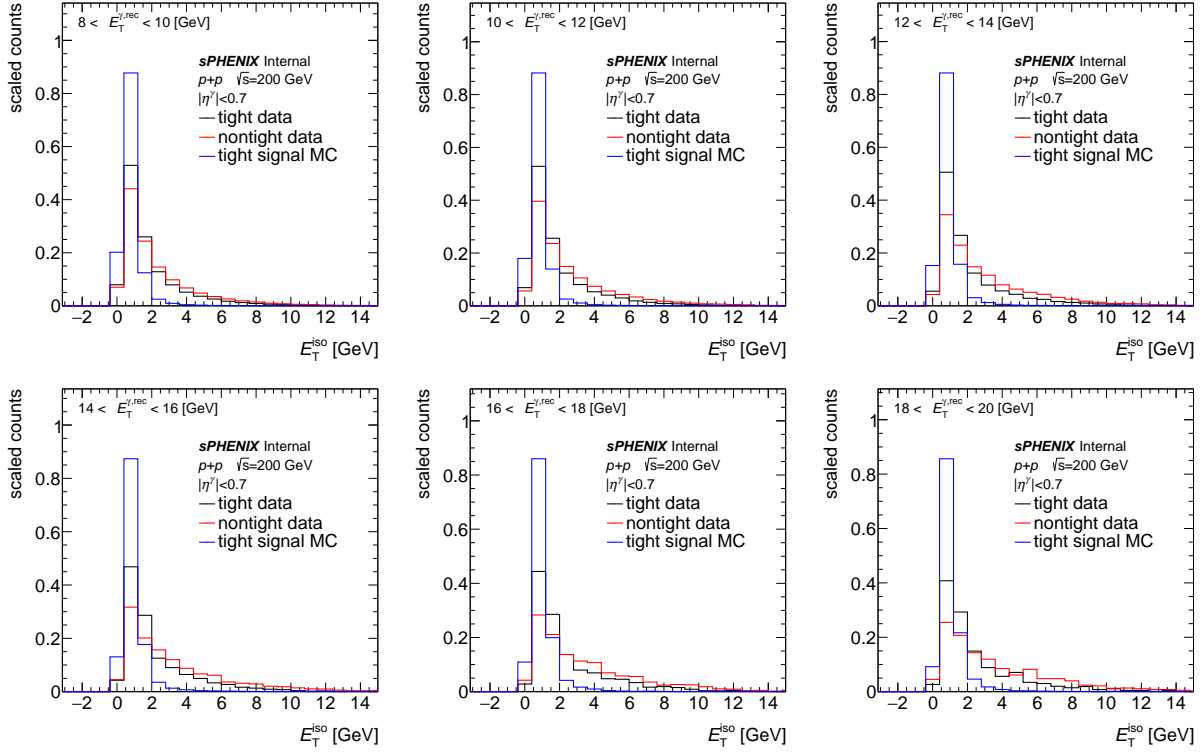
Figure 22: Distributions of E_T^{iso} for different E_T^{γ} bins in different panels, respectively. Photons with tight ID (black), with non-tight ID (red) in data and signal MC with tight ID (blue) are overlaid.

photons in A can be calculated with:

$$N_{\text{signal}}^{A,\text{data}} = N_{\text{raw}}^{A,\text{data}} - N_{\text{raw}}^{B,\text{data}} \cdot \frac{N_{\text{raw}}^{C,\text{data}}}{N_{\text{raw}}^{D,\text{data}}} \tag{3}$$

where $N_{\text{signal}}^{A,\text{data}}$ is the number of reconstructed signal clusters in region A, $N_{\text{raw}}^{X,\text{data}}$ is the number of clusters in region X. The number of clusters in each sideband region for each p_T bin and their ratio with respect to the signal region A are shown in Figure 24.

4.3.2 Signal Leakage Correction

Equation 3 is valid only if the signal photons do not contribute to the controlled region. However, in real applications, signal photons leak into the controlled region and it needs to be corrected. When accounting for the signal leakage into the controlled region, the sideband subtraction calculation becomes:

$$N_{\text{signal}}^{A,\text{data}} = N_{\text{raw}}^{A,\text{data}} - \left[\left(N_{\text{raw}}^{B,\text{data}} - f^{B,\text{MC}} N_{\text{signal}}^{A,\text{data}} \right) \cdot \frac{\left(N_{\text{raw}}^{C,\text{data}} - f^{C,\text{MC}} N_{\text{signal}}^{A,\text{data}} \right)}{\left(N_{\text{raw}}^{D,\text{data}} - f^{D,\text{MC}} N_{\text{signal}}^{A,\text{data}} \right)} \right] \tag{4}$$

where $f^{X,\text{MC}} = \frac{N_{\text{signal}}^{X,\text{MC}}}{N_{\text{signal}}^{A,\text{MC}}}$ is the ratio of signal clusters in region X over region A, derived with signal MC, shown in Figure 25. The statistical uncertainty on the signal in region A after subtraction is
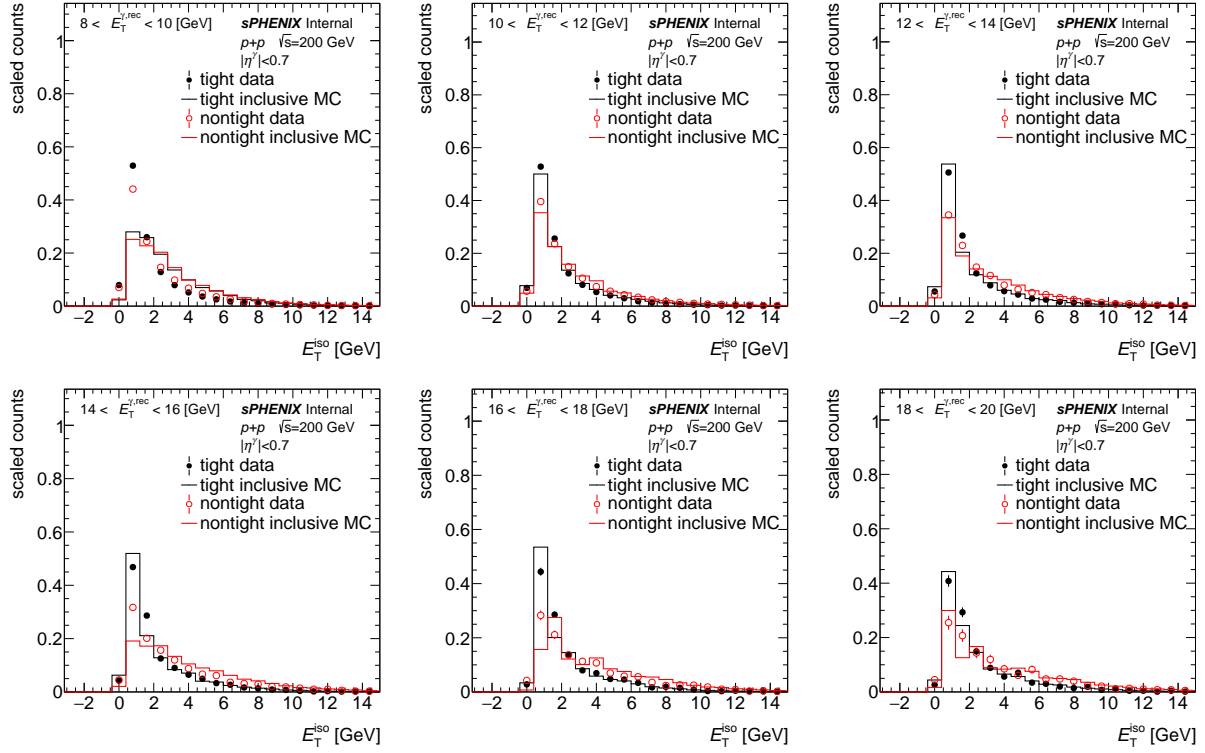
Figure 23: Distributions of E_T^{iso} for different E_T^γ bins in different panels, respectively. Photons in data with tight ID (black closed circle) and non-tight ID (red open circle) are compared to photons in inclusive MC with tight ID (black histogram) and and non-tight ID (red histograms).
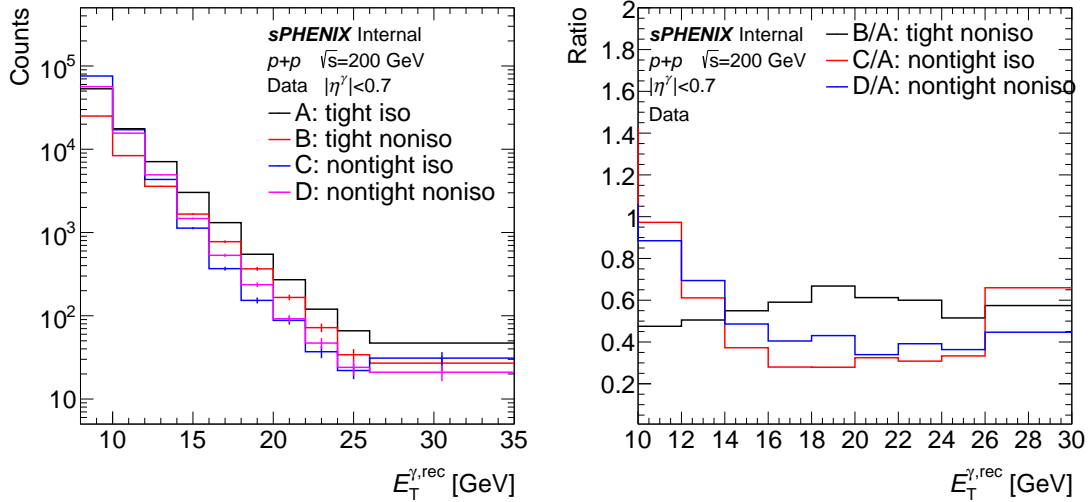
Figure 24: Yield of photons in signal and sideband regions (left) and ratio of yield in each sideband region to the signal region (right) as a function of E_T^γ in data.

estimated using a bootstrapping method by resampling $N_{\text{raw}}^{X,\text{data}}$, assuming Poisson statistics, with 10,000 iterations.

The purity in the signal region is defined as the fraction of signal photons relative to the total photon candidates in the signal region A (tight, iso):

$$\text{Purity}(\mathcal{P}) = \frac{N_A^{sig}}{N_A}. \tag{5}$$

Figure 26 shows the estimated purity in data both with and without corrections for signal leakage. Purity is found to be around 50% at 10 GeV and increases with E_T^{γ} to be about 70% at 20 GeV.

As a closure test, the same purity estimation procedure was performed with inclusive MC samples and compared with truth-level purity in MC calculated using truth-level information (ground-truth signal composition). Figure 27 shows the closure test performed with inclusive MC. Purity values are estimated in inclusive MC following the same procedure as data, and compared with the truth-level purity, which shows good agreement within uncertainties.

The inclusive MC sample is not fully efficient below 14 GeV leading to the non-accurate signal background ratio. More MC samples are currently being produced and the results will be updated with more statistics in the future.
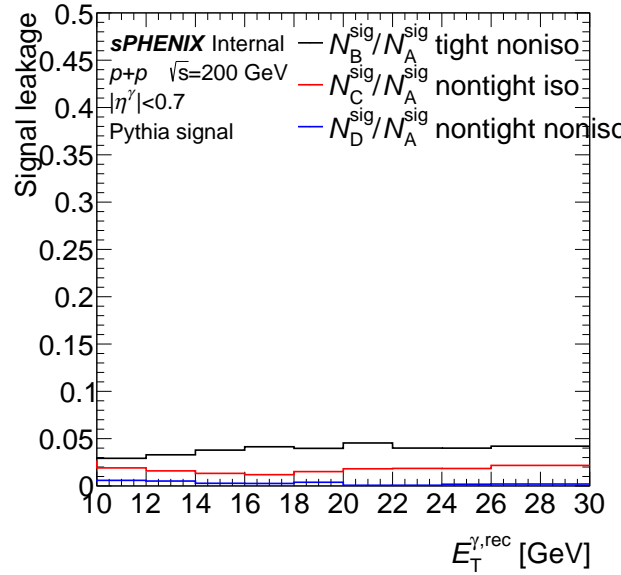
Figure 25: The fraction of signal photons in different sideband regions in MC.
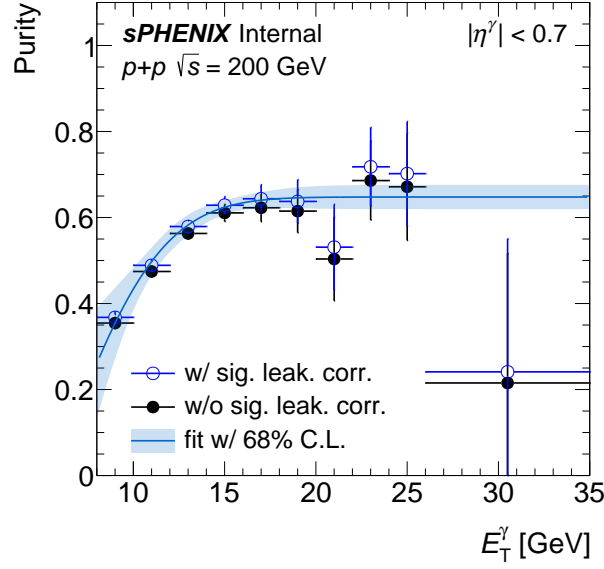
Figure 26: Purity as a function of E_T^γ with (black closed circle) and without (blue open circle) signal leakage correction in data. The purity with leakage correction is fitted by an error function, the shaded area shows the 68.3% confidence interval of the fit.
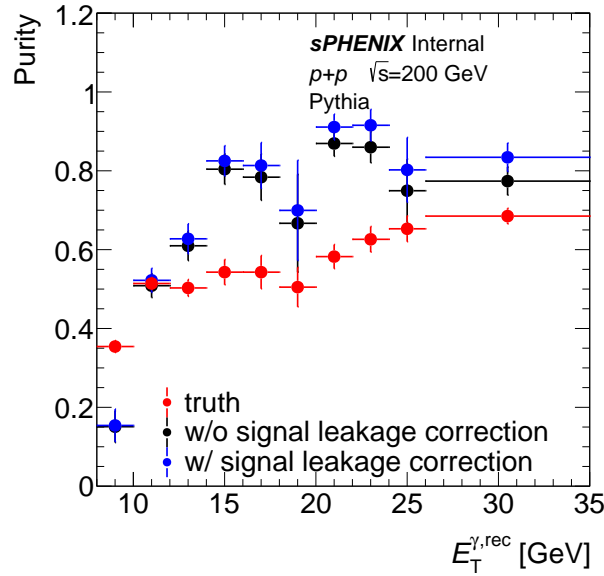

Figure 27: Purity as a function of E_T^γ in inclusive MC as a closure test. The truth-level purity (red) is compared to the purity values, estimated using the same method as data, with (black) and without (blue) signal leakage correction.
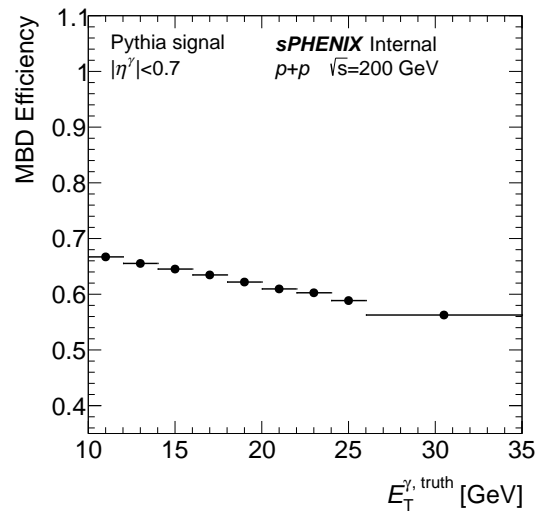
4.4 Efficiency

The photon reconstruction, identification, isolation efficiencies, and event-level selection efficiencies are estimated as functions of the truth-level photon E_T^γ. The truth-reco matching requirement is described in Section 4.2. Each efficiency definition is listed:

- **Reconstruction efficiency**: The fraction of truth-level signal photons (defined in Section 4.2) that are matched to a reconstructed cluster.

- **Identification efficiency**: The fraction of truth-level signal photons passing both the pre-selection criteria and the tight photon identification defined in Section 3.4 out of all truth-level signal photons that are matched to reconstructed photons.

- **Isolation efficiency**: The fraction of truth-level signal photons satisfying the isolation requirements outlined in Section 3.3 out of all truth-level signal photons that are matched to reconstructed photons.

- **MBD coincidence and event selection efficiency**: The fraction of truth-level signal photons that pass the MBD coincidence cut and event selection requirements detailed in Section 2.4 out of all truth-level signal photons that are matched to reconstructed photons.

Figure 28 shows photon reconstruction, identification, and isolation efficiencies, respectively, as well as the combined efficiency. The combined efficiency of MBD trigger and event selection requirement is shown in Figure 29. These efficiencies are corrected on a bin-by-bin basis to the spectrum after unfolding.
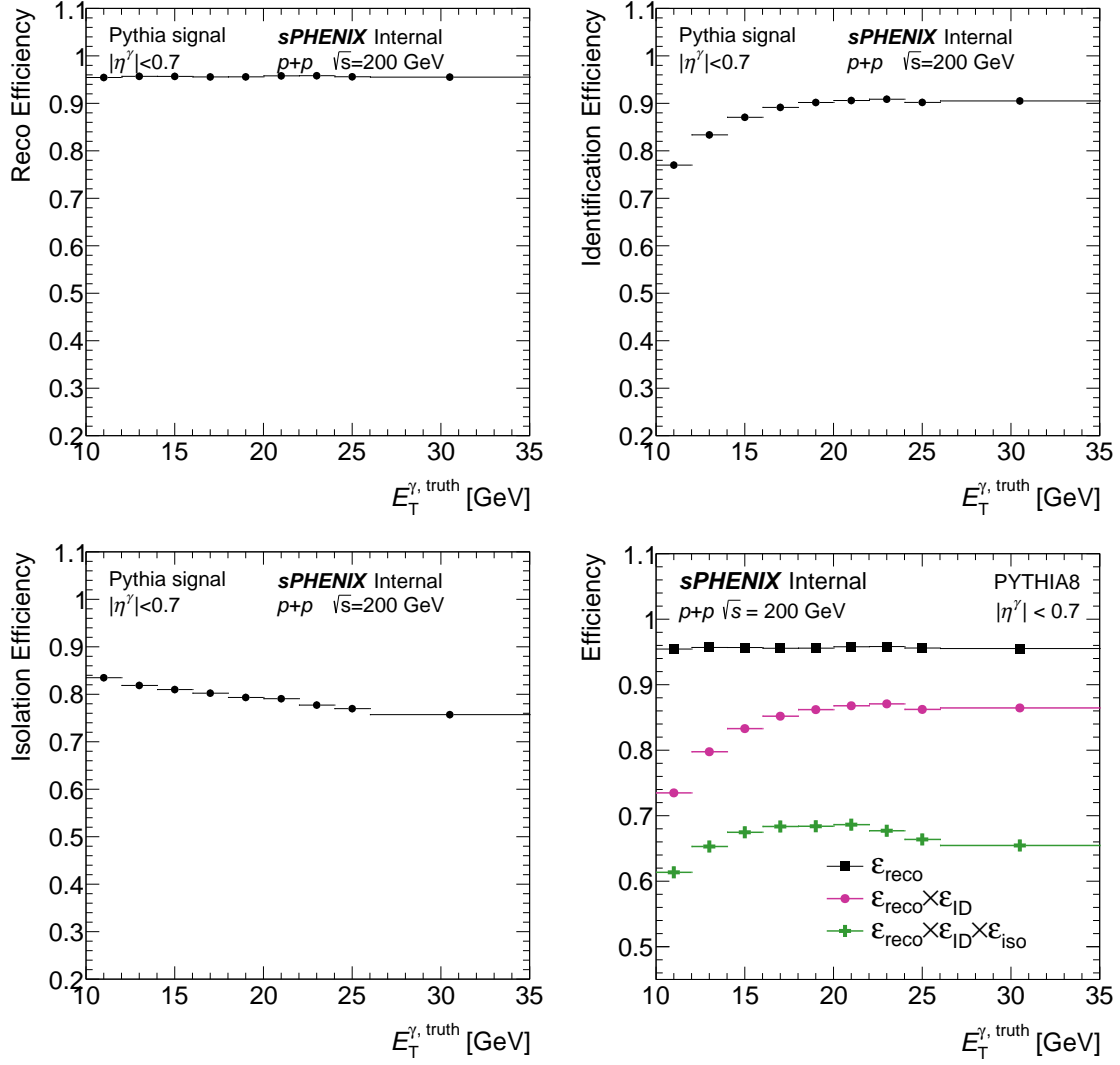
Figure 28: Photon efficiencies for reconstruction (top left), identification (top right), isolation requirement (bottom left), and convoluted step-by-step efficiencies (bottom left) as a function of truth photon $E_T^{\gamma,\,\text{truth}}$.

Figure 29: MBD coincidence and selection efficiencies as a function of truth photon $E_T^{\gamma, \text{truth}}$.

4.5 Unfolding

The unfolding procedure employs the D'Agostini Bayesian iterative method [6] using RooUnfold version 3.0.5 [7]. The fiducial binning for the final results is [10, 12, 14, 16, 18, 20, 22, 24, 26, 35] GeV, with an underflow bin of [8, 10] GeV and an overflow bin of [35, 40] GeV. For truth photon $E_T^{\gamma,\,\text{truth}}$, there is an additional underflow bin of [5, 8] GeV to account for lower-$E_T^{\gamma,\,\text{truth}}$ photons that are reconstructed at higher reconstructed photon E_T^{γ}.

To make the prior of E_T^{γ} distribution in MC similar to that in data, the response matrix is reweighted using the ratio of purity-corrected yield in data to signal photon yield in MC as shown in Figure 30. Figure 31 shows the reweighted response matrix, constructed from signal MC samples containing truth-level photons matched to reconstructed clusters that pass the isolation and tight selection criteria. The data to MC ratio is fitted by a third polynomial function from 8 to 30, the reweighting factor based on the fit function is applied photon-to-photon based on E_T^{γ} while populating the response matrix.

Due to the relatively small photon energy resolution, the impact of unfolding on the spectrum is minor compared to the statistical uncertainties. As shown in Figure 32, the relative deviation between successive unfolding iterations becomes smaller than the statistical uncertainty after the first iteration, supporting the choice of a minimal number of iterations. For the nominal result, the data-reweighted response matrix with 2 iterations is selected. This choice balances stability (minimizing iteration-dependent biases) and statistical precision, as justified by the closure tests. Figure 33 shows the photon yield as a function of E_T^{γ}, evaluated at each stage: after the pre-selection cut, tight ID and isolation requirements, purity correction, unfolding, and efficiency correction. The ratio of the yield after unfolding to that before unfolding is found to be 2–10%, depending on the E_T^{γ} bin.

### 4.5.1 Unfolding Closure Test

To validate the unfolding procedure, two closure tests were performed:

- Full closure test: The response matrix is constructed using the entire events in the MC sample, and the same MC events are unfolded using this matrix.
- Half closure test: The response matrix is built with half of the events in the MC sample, and the unfolding is applied to the independent remaining half.

Figure 34 shows the closure test results: the full closure test reproduces the truth-level spectrum within 1%. The half closure test achieves agreement within 2%, consistent with statistical uncertainties.

4.6 Luminosity

The luminosity is calculated by counting minimum bias collisions for which the trigger system was *live* and and multiplied by a per-minimum-bias-event measured cross section. The luminosity calculated is a prescale-corrected luminosity with the reconstructed vertex criteria and the calculation is documented in the sPHENIX IAN [5]. The result is 16.8468 pb^{-1}.
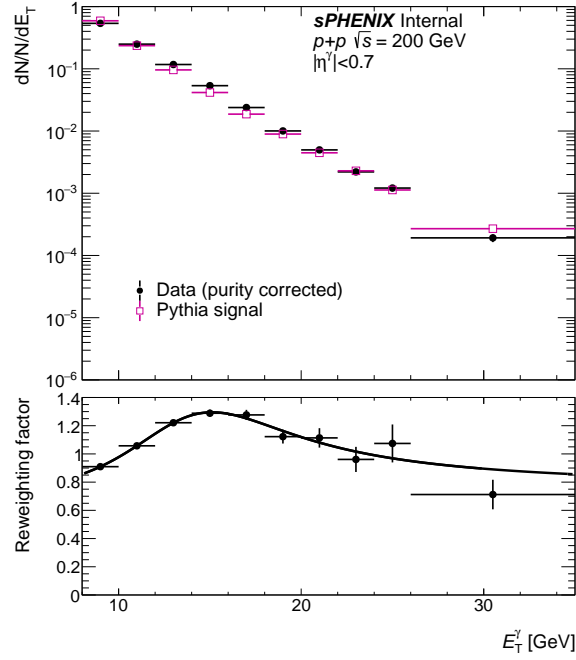
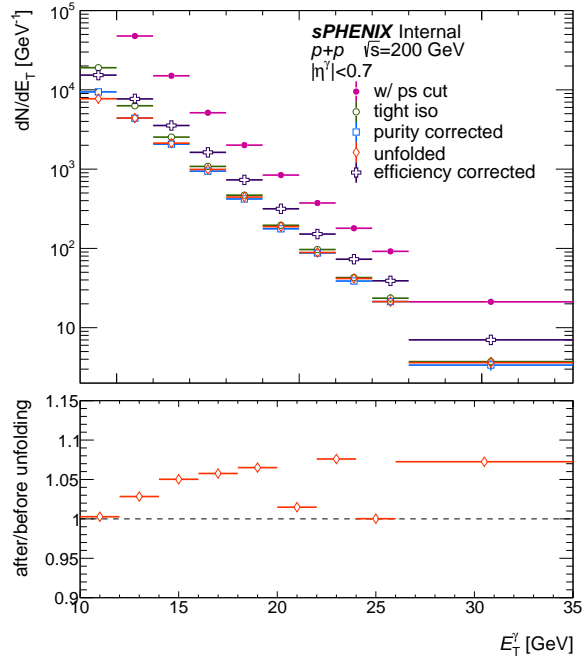**Figure 30:** Top panel: photon E_T^γ distributions of purity-corrected yield in data (black) and signal yield in PYTHIA-8 MC (red). Bottom panel: the ratio of data to MC, which is fitted by an order [2/2] Padé approximant from 8 to 35 GeV.
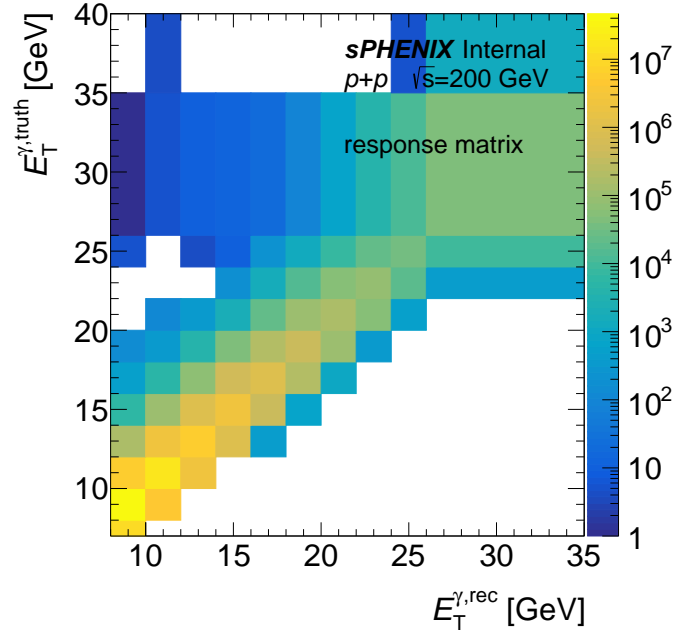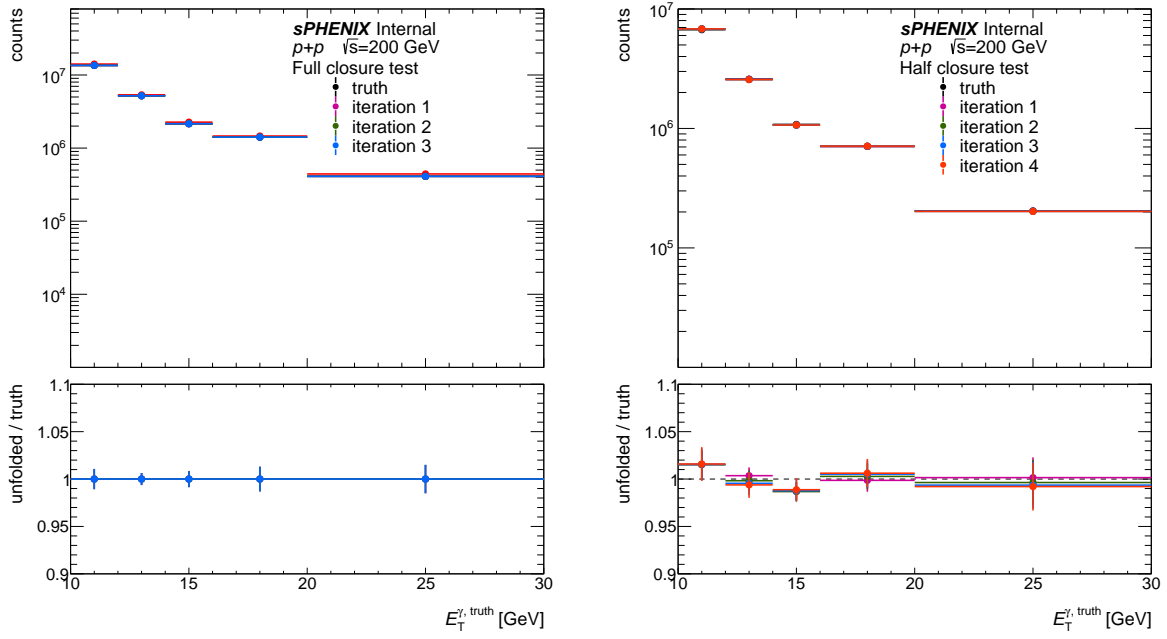

Figure 31: Response matrix after reweighting.
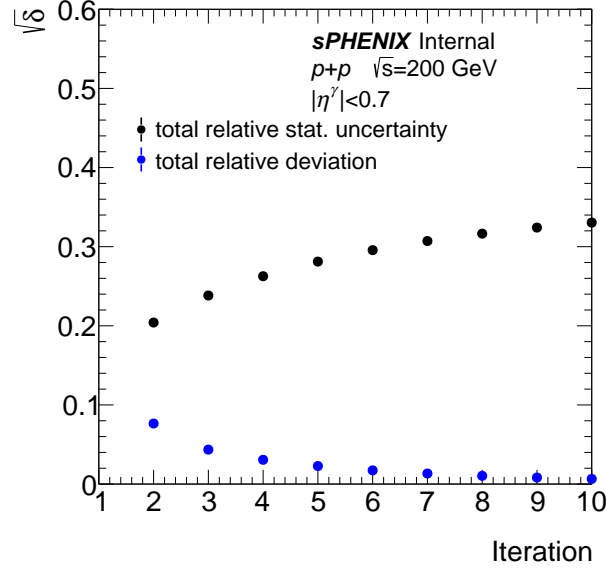
Figure 32: Unfolding relative change and statistic efficiency for each unfolding iteration

Figure 33: Top panel: yield of photons as a function of E_T^γ at different stages of the analysis. Bottom panel: the ratio after and before unfolding

Figure 34: Unfolding full (left) and half (right) closure test result, with truth and unfolded spectrum as a function of E_T^{γ}, the lower panels of each plot shows the ratio between unfolded spectrum and truth.

5 Systematic Uncertainties

Systematic uncertainty sources of this analysis are listed:

- Photon Energy Scale and Resolution - (Subsection 5.1)
- Efficiency (tight ID variation) - (Subsection 5.2)
- Photon Purity (non-tight ID, non-E_T^{iso} variations, Purity fitting) (Subsection 5.3)
- Unfolding (Subsection 5.5)
- Luminosity (Subsection 5.6)
- MBD Trigger Efficiency (Subsection 5.7)

The systematic uncertainty for each source is calculated by repeating the whole analysis procedure with the corresponding variations. For the sources that have distinct "up" and "down" variations, the largest absolute change of the two is taken as the variation for each E_T^γ bin. For sources where there is only one variation (the large majority), its absolute magnitude is taken and symmetrized. The relative differences between with and without each systematic variation in each E_T^γ bin are then added in quadrature to produce the total uncertainty.

5.1 Photon Energy Scale and Resolution

5.1.1 Photon Energy Scale

The uncertainty from the energy scale calibration is determined by shifting the cluster E_T^{reco} by 2.6% in the MC as suggested by the *Calorimeter Calibration* Working Group. The difference from this variation is symmetrized to determine the systematic uncertainty bin-by-bin. Figure 35 shows that this variation leads to a systematic uncertainty of 8–26% relative to the nominal.
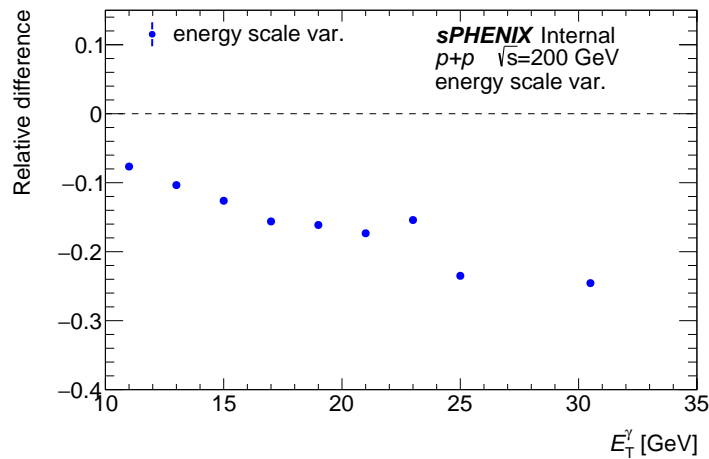
Figure 35: Systematic uncertainty from photon energy scale variation as a function of E_T^γ.

5.1.2 Photon Energy Resolution

The uncertainty from the mismodeling the photon energy resolution and detector response is conservatively estimated by smearing the cluster E_T^{reco} by 5% in the signal MC as suggested by the *Calorimeter Calibration* Working Group. Figure 36 shows that this variation leads to a systematic uncertainty of less than 5% relative to the nominal.
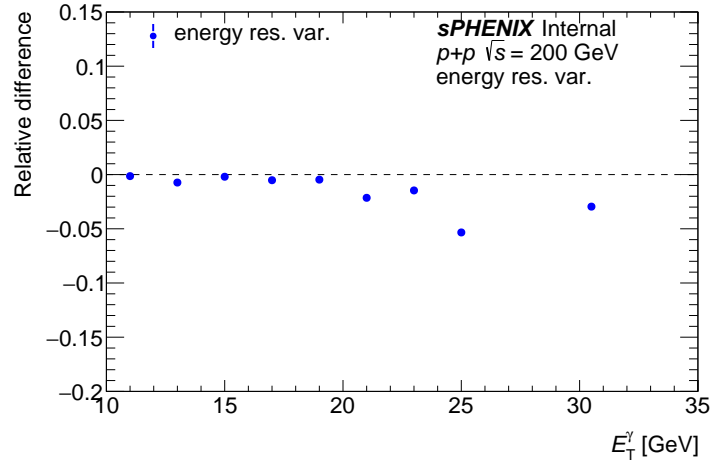
Figure 36: Systematic uncertainty from photon energy resolution variation as a function of E_T^γ.
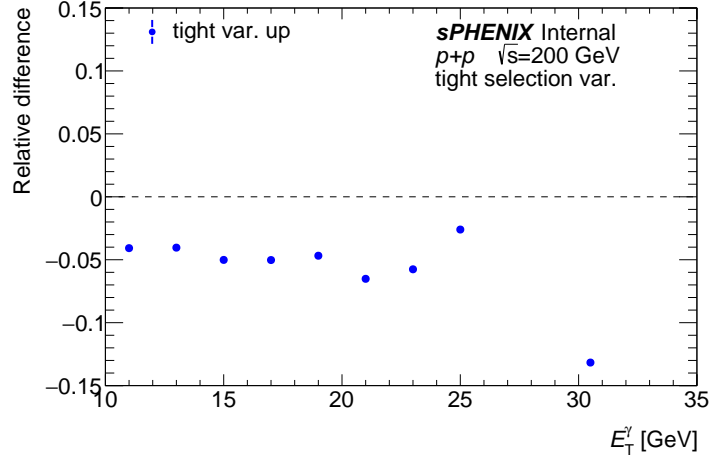
5.2 Efficiency

5.2.1 Photon Identification

The choice of tight ID criteria affects both efficiency and purity calculations. Because the shower shape variable weta_cogx provides the greatest signal-to-background separation in the tight selection, we assess the systematic variation by changing its requirement from the default value shown in Table 3 to $0 <$ weta_cogx < 0.2. Difference between this variation is symmetrized to determine the systematic uncertainty bin by bin. Figure 37 shows that this variation leads to a systematic uncertainty of 5–10% relative to the nominal tight ID selection.

5.3 Photon purity

5.3.1 Non-tight Selection

The purity calculation is sensitive to the non-tight selection criteria. To quantify this effect, we vary the requirements for non-tight identification (ID) by changing the number of allowed failed tight cuts from the nominal value of 2 to 1 and 3. The resulting deviations in purity are evaluated, and the maximum difference between these variations is symmetrized to determine the systematic uncertainty bin by bin. As shown in Figure 38, this uncertainty reaches up to 15% relative to the nominal selection, reflecting the impact of non-tight ID criteria on the analysis.

Figure 37: Systematic uncertainty for efficiency as a function of E_T^γ.
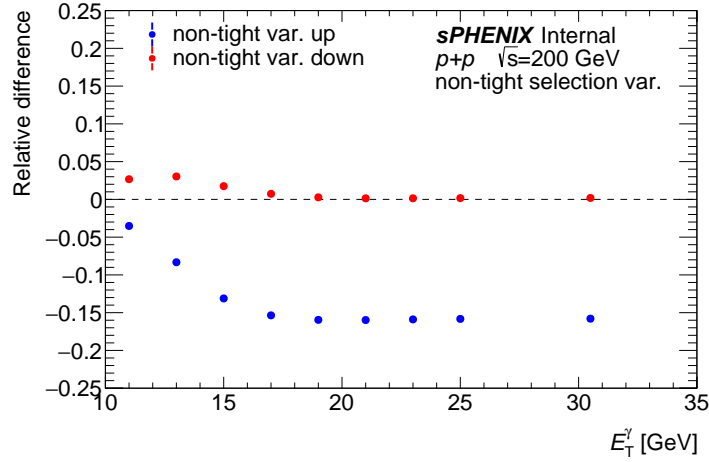
Figure 38: Systematic uncertainty for non-tight selection variation as a function of E_T^γ.

### 5.3.2 Sideband Isolation Requirement

The choice on the non-isolation criteria for the sideband region will have influence on the purity calculation. The upper systematic variation for the non-isolation criteria is chosen to be:

$$E_T^{iso} > 2 + 1.08128 + 0.0299107 \cdot E_T^{reco}$$

The lower systematic variation for the non-isolation criteria is chosen to be:

$$E_T^{iso} > 0.5 + 1.08128 + 0.0299107 \cdot E_T^{reco}$$

The differences between these variations are used to determine the systematic uncertainty bin by bin. Figure 39, shows deviations between 4% to 20% from the nominal with this selected variation.
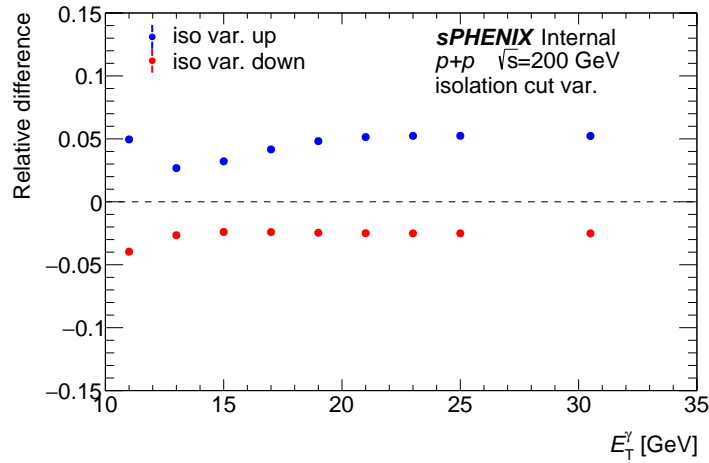
Figure 39: relative deviation for isolation selection systematic variation as a function of E_T^γ

5.4 Purity Fitting

The systematic uncertainties associated with fitting the purity distribution arise from two sources: (1) statistical uncertainties in the fit parameters and (2) the choice of fitting function. These are evaluated as follows:

Fit parameter uncertainties: The upper and lower fluctuations due to fit uncertainties are determined using the 68.3% confidence interval (1σ) of the fitted parameters, derived from the 68.3% quantiles of the parameter posterior distributions.

Fitting function variation: To assess the dependence on the functional form, the purity distribution is refit using a [1/1] Padé approximant (a rational function with first-degree numerator and denominator). The bin-by-bin deviations between the nominal fit and the Padé-based fit are computed.

The total systematic uncertainty is determined by combining these two components. For each bin, the maximum deviation between the nominal fit uncertainty (from quantiles) and the Padé-based variation is identified. This maximum deviation is then symmetrized (i.e., assigned as $\pm$ the larger of the upward/downward deviations) to conservatively represent the systematic uncertainty in each bin. Figure 40 shows that this variation leads to a 10% to 20% deviation from the nominal result.

5.5 Unfolding

Since the response matrix in the nominal analysis is reweighted to match the data spectrum, the systematic variation is chosen to unfold without reweighting the response matrix. The deviation is used to determine the systematic bin by bin. Figure 41 shows that this variation leads to a less than a 10% deviation from the nominal result.
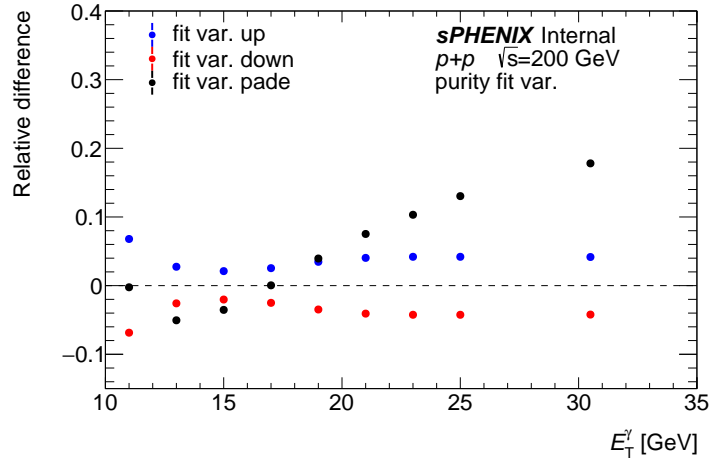
**Figure 40:** relative deviation for purity fitting systematic variation as a function of E_T^γ
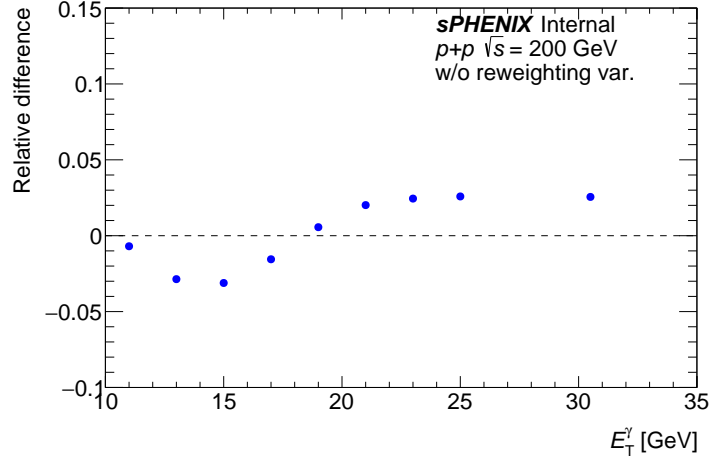
Figure 41: Systematic uncertainty for unfolding variation as a function of E_T^γ.

5.6 Luminosity

The luminosity systematic uncertainty is propagated directly from the minimum-bias dimuon (MBD) trigger cross-section measurement of $26.1^{+4.3}_{-1.1}$ mb [cite]. This corresponds to an asymmetric uncertainty of $^{+16.5\%}_{-4.2\%}$ from the integrated luminosity. The uncertainty is applied uniformly across all E_T^γ bins without E_T^γ-dependent variations.

5.7 MBD Trigger Efficiency

Currently we are using a conservative estimation for the systematic uncertainty from the MBD trigger efficiency by varying the MC driven efficiency up and down by 5%. Figure 42 shows that this variation leads to a less than a 10% deviation from the nominal result.
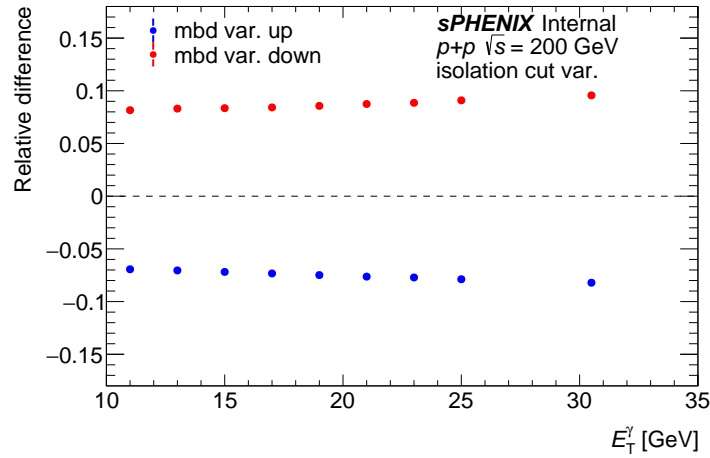
Figure 42: Systematic uncertainty for MBD trigger efficiency as a function of E_T^γ.

5.8 Total Systematic Uncertainties

Figure 43 summarizes the contributions to the systematic uncertainty from each variation for systematic uncertainties. The total systematic uncertainty is obtained by taking the quadrature sum of systematic uncertainties from all independent variations. At lower E_T^γ, the dominant source of systematic uncertainty (20%) comes from the energy scale.
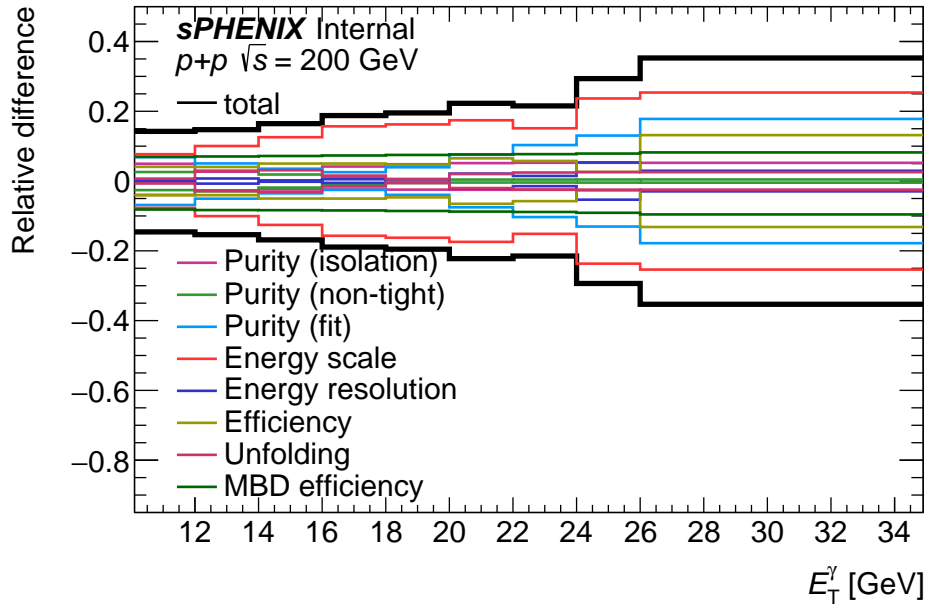
Figure 43: Summary of systematic uncertainties for different variations as a function of E_T^γ.

6 Results

In this Section, the final results for the isolated-prompt photon cross section as a function of p_T in $p+p$ collisions at $\sqrt{s}$ =200 GeV are shown. The raw yield of EMCal clusters passing the tight shower shape criteria and the isolation energy requirement $N^{\text{tight,iso}\,\gamma}$ is corrected purity $\mathcal{P}$ and reconstruction effects through unfolding according to:

$$Y^{\text{rec}}(E_T^\gamma) = \text{Unfolded} \left[N^{\text{tight,iso}\,\gamma}(E_T^\gamma) \times \mathcal{P}(E_T^\gamma) \right] \tag{6}$$

This yield is then corrected for the signal efficiency $\mathcal{E}$ and divided by the luminosity $\mathscr{L}$, to calculate the measured differential cross section $d\sigma/(dE_T^\gamma d\eta^\gamma)$,

$$\frac{d\sigma}{dE_T^\gamma d\eta} = \frac{1}{\mathscr{L}} \frac{Y^{\text{rec}}}{\mathcal{E} \Delta E_T^\gamma \Delta \eta^\gamma}. \tag{7}$$

Figure 44 shows the fully-corrected differential cross-section. The results are then compared to theoretical calculations from MC generators of PYTHIA-8 and JETPHOX. The PYTHIA-8 is version 8.307 with the Detroit tune [8]. The JETPHOX MC generator (v1.3.1_4) calculates cross sections for both direct and fragmentation photons at next-to-leading order (NLO). In the JETPHOX calculations, the CT14LO parton distribution functions are employed, and the BFG setII [9] is used for the fragmentation functions. The renormalization (μ_R), factorization (μ_F), and fragmentation (μ_f) scales are all set to E_T^γ. Both PYTHIA-8 and JETPHOX require the same truth-level isolation criteria as used in the data. Both the JETPHOX and the PYTHIA-8 predictions agree with the measurements within experimental uncertainties. JETPHOX will be updated with better statistics and with systematic uncertainties by varying μ_R, μ_F and μ_f scales. The result is also compared to NLO pQCD calculations of prompt photons, provided by Werner Volgesang. The calculation does not require E_T^{iso} condition.

The results are also compared with previous measurements reported by the PHENIX experiment [1], as shown in Figure 45. The PHENIX measurements report direct photons, whereas sPHENIX measures both direct and fragmentation photons. Additionally, the PHENIX measurement does not require isolation of direct photons, whereas our results do. Furthermore, the PHENIX data were collected in $|\eta| < 0.25$ and then scaled to match the double-differential cross section, $d\sigma/(d\eta dE_T^\gamma)$. Despite these differences, both measurements are consistent within their respective uncertainties.
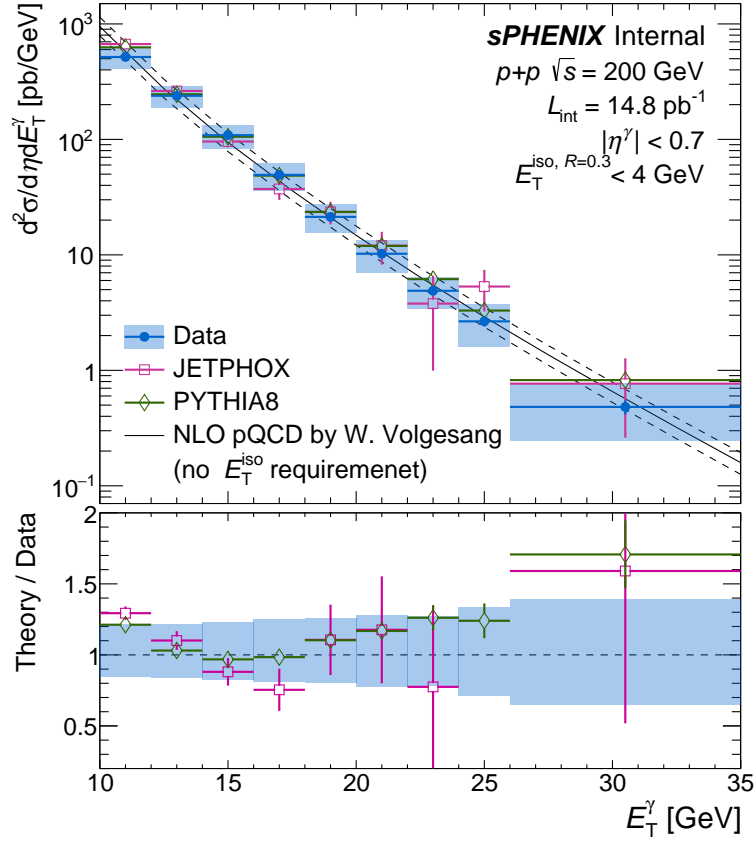
Figure 44: The differential cross section of isolated prompt photons as a function of E_T^γ is compared to theoretical predictions of PYTHIA-8 (green), JETPHOX (pink) and NLO pQCD calculations by Werner Volgesang. The statistical uncertainties are plotted as vertical lines and the systematic uncertainties are plotted as shaded bands. The lower panel shows a theory-to-data ratio to this analysis, where the experimental systematic uncertainties are shown as shaded bands around unity. The theory and experimental statistical uncertainties are combined on the theory points.
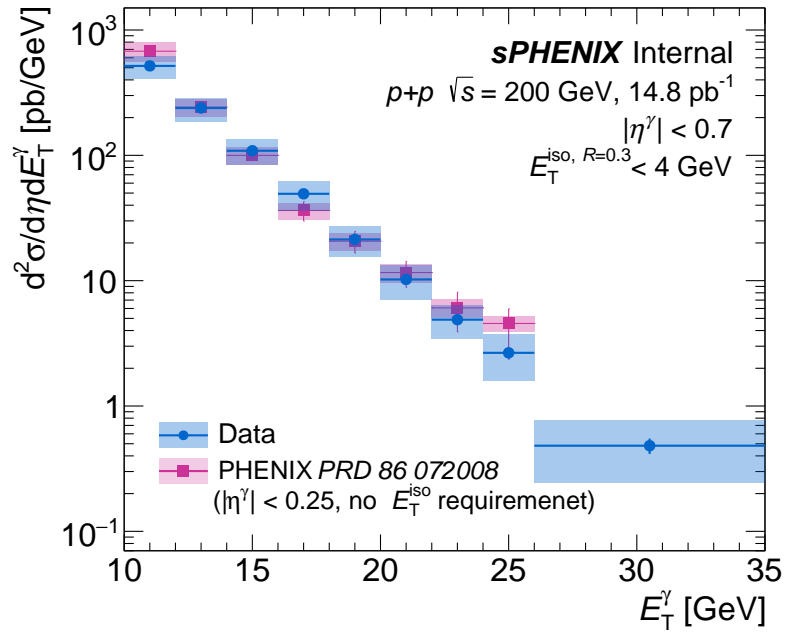
Figure 45: The differential cross section of isolated prompt photons as a function of E_T^γ is compared to the PHENIX measurements [1] of direct photons. The statistical uncertainties are plotted as vertical lines and the systematic uncertainties are plotted as shaded bands.

7 Conclusion

The differential cross section of isolated prompt photons is measured in proton-proton collisions at $\sqrt{s} = 200$ GeV using the data taken during Run 24. The results are reported for photon E_T^γ of 10–30 GeV within $|\eta^\gamma| < 0.7$. The data is compared to theoretical predictions at next-to-leading order and past experiential results of direct photons. This analysis utilizes a data-driven purity estimation approach and is fully unfolded to account for all reconstruction effects.

References

[1] PHENIX Collaboration. Direct-Photon Production in $p + p$ Collisions at $\sqrt{s} = 200$ GeV at Midrapidity. *Phys. Rev. D*, 86:072008, 2012. `arXiv:1205.5533`, `doi:10.1103/PhysRevD.86.072008`. 1, 6, 45

[2] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015. `arXiv:1410.3012`, `doi:10.1016/j.cpc.2015.01.024`. 2.2

[3] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. `doi:10.1016/S0168-9002(03)01368-8`. 2.2

[4] Martin Spousta and Brian Cole. Interpreting single jet measurements in pb+pb collisions at the lhc. *The European Physical Journal C*, 76(2), January 2016. URL: `http://dx.doi.org/10.1140/epjc/s10052-016-3896-0`, `doi:10.1140/epjc/s10052-016-3896-0`. 4

[5] Jiang H., Narde A., James J., Brahma A., Clement J., and Applegate N. Hodges A. Inclusive jet cross-section in $\sqrt{s} = 200$ GeV p+p collisions collected in run 2024. *sPHENIX-Invenio*, 2025. URL: `https://sphenix-invenio.sdcc.bnl.gov/me/requests/ca04e089-3024-4267-aafd-81f081b4af6c`. 3.5, 4.6

[6] G. D'Agostini. Improved iterative bayesian unfolding, 2010. URL: `https://arxiv.org/abs/1010.0632`, `arXiv:1010.0632`. 4.5

[7] Tim Adye. Unfolding algorithms and tests using RooUnfold. In *PHYSTAT 2011*, pages 313–318, Geneva, 2011. CERN. `arXiv:1105.1160`, `doi:10.5170/CERN-2011-006.313`. 4.5

[8] Manny Rosales Aguilar, Zilong Chang, Raghav Kunnawalkam Elayavalli, Renee Fatemi, Yang He, Yuanjing Ji, Dmitry Kalinkin, Matthew Kelsey, Isaac Mooney, and Veronica Verkest. pythia8 underlying event tune for RHIC energies. *Phys. Rev. D*, 105(1):016011, 2022. `arXiv:2110.09447`, `doi:10.1103/PhysRevD.105.016011`. 6

[9] L. Bourhis, M. Fontannaz, and J. P. Guillet. Quark and gluon fragmentation functions into photons. *Eur. Phys. J.*, 2(3):529–537, 1998. `arXiv:hep-ph/9704447`, `doi:10.1007/s100529800708`. 6

Appendix

8 Shower Shape Variables

Figure 46 and Figure 47 are shower shape distributions for all clusters in the analysis. The data generally follows the background MC (inclusive jet sample with truth background selection) with the exception of specific excesses induced by non-physics backgrounds. One can see that after the pre-selection, in Figure 12 and Figure 13 the non-physics backgrounds are significantly reduced but not completely removed.
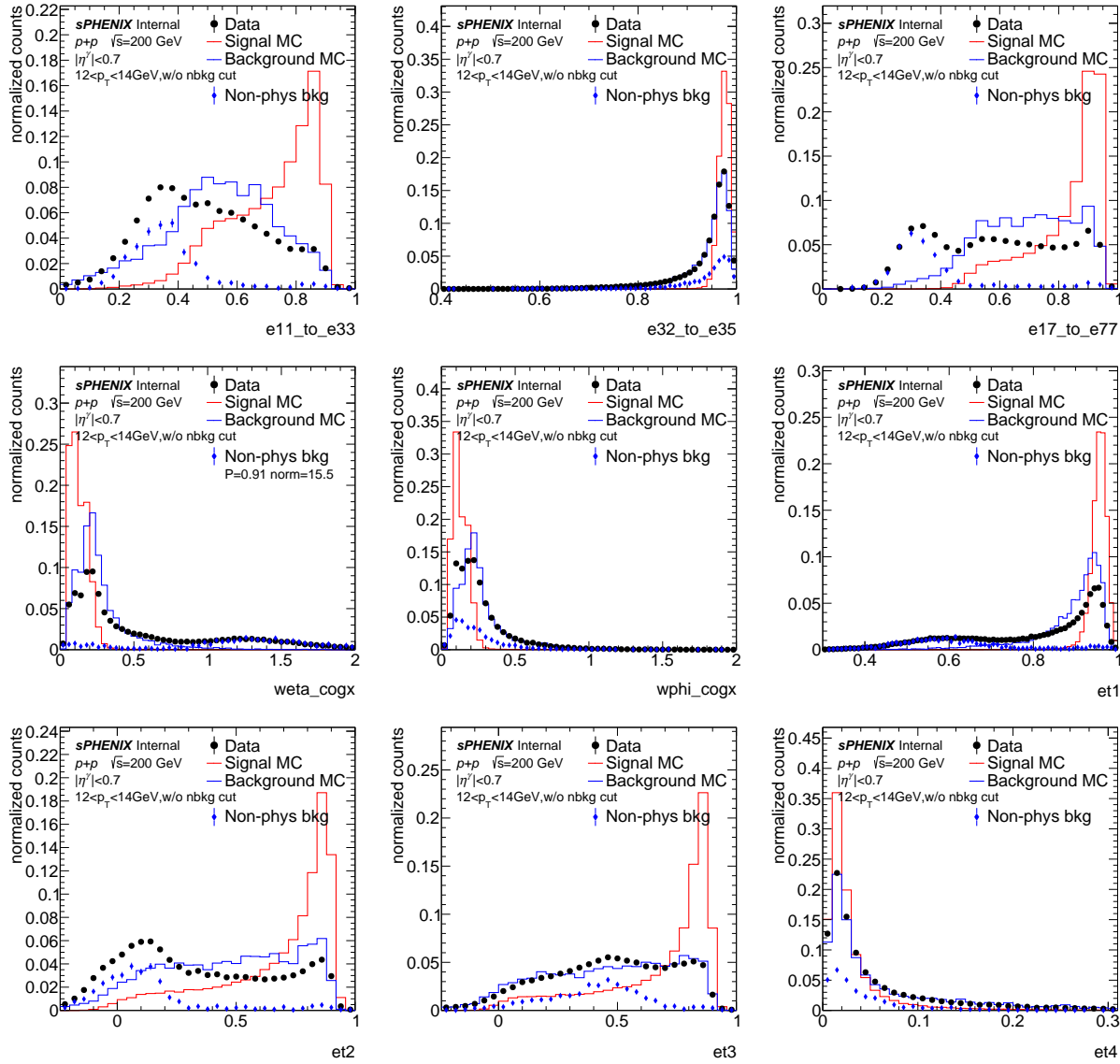
**Figure 46:** Shower shape distributions with no pre-selection at lower p_T (12-14 GeV).

It is important to observe the correlation between E_T^{iso} and the shower shapes without any pre-
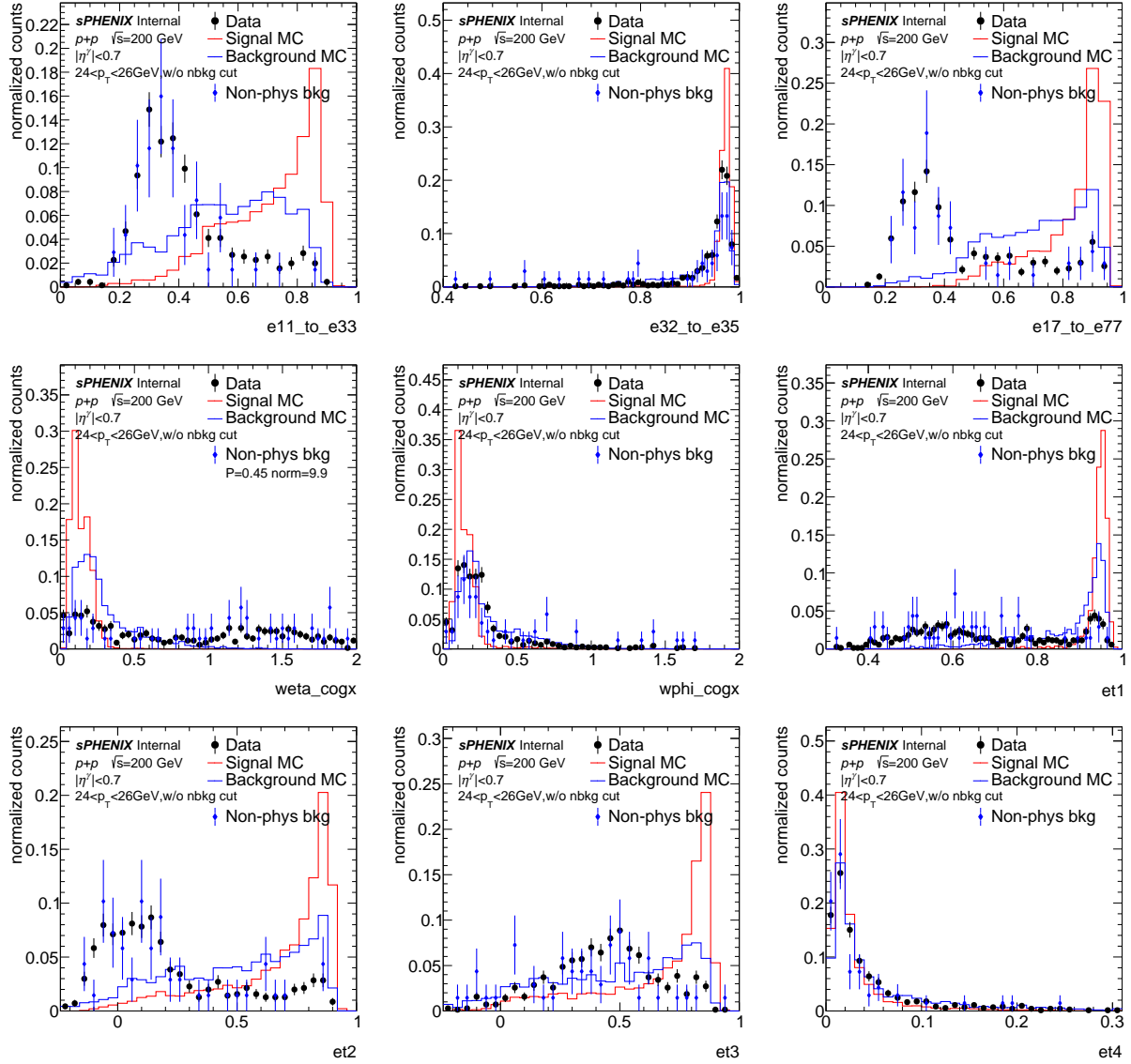
Figure 47: Shower shape distributions with no pre-selection higher p_T (24-26 GeV).

selection. In Figure 48 and Figure 48 the average E_T^{iso} with shower shape value are show for many shower shapes. It can be seen that some variable are more correlated than others.
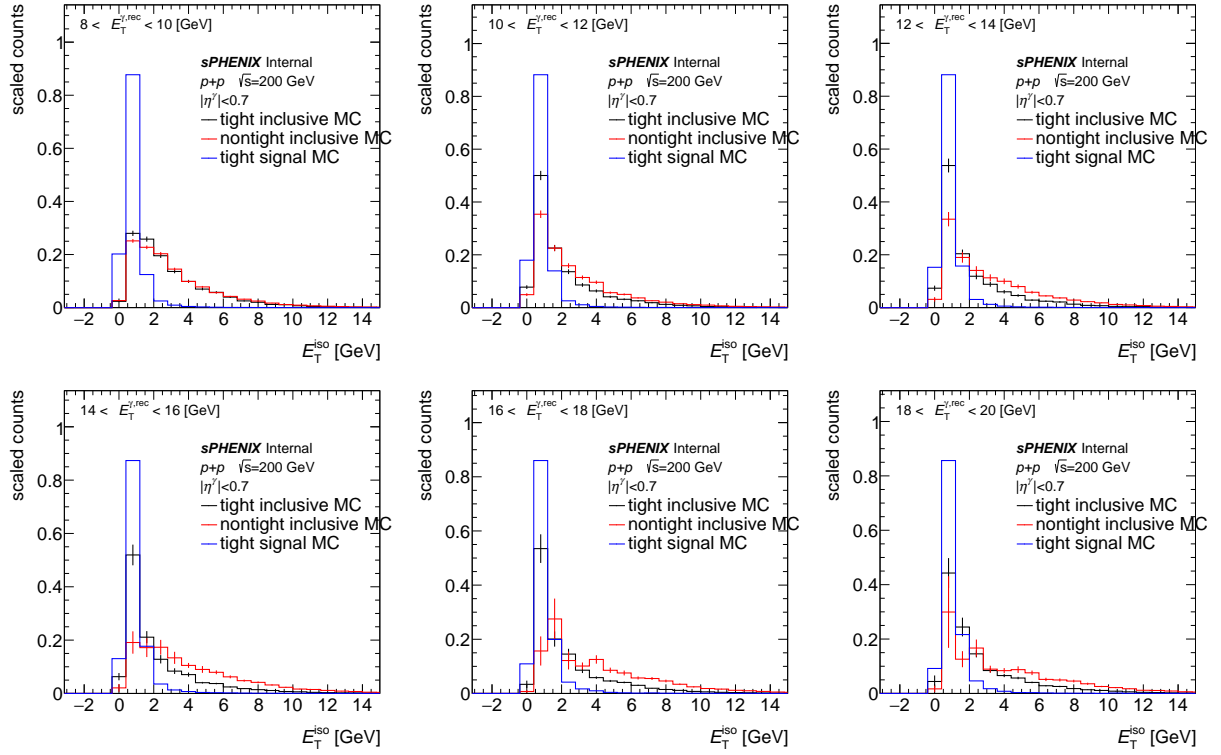
9 Isolation Energy Distributions
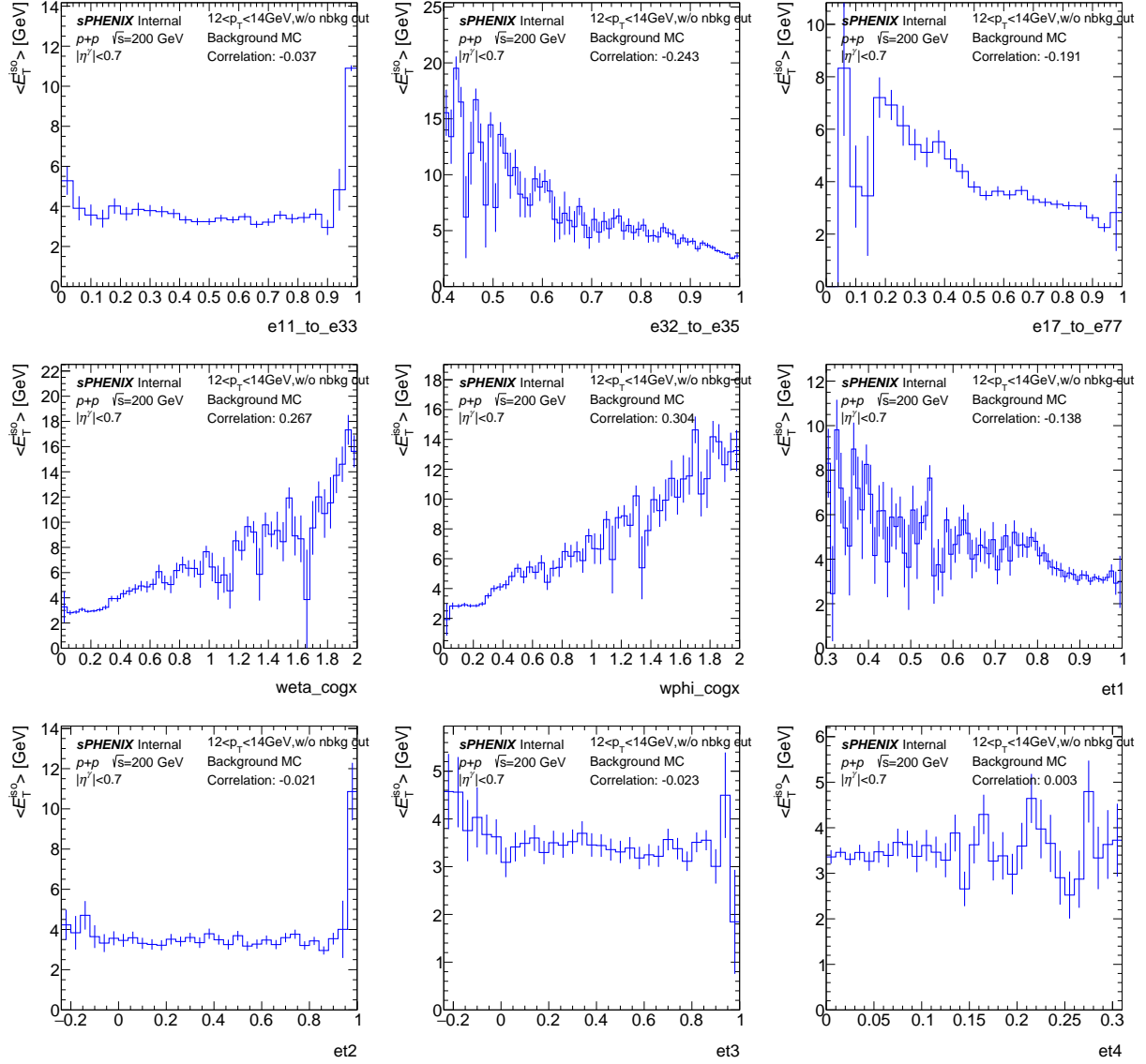
Figure 48: The average E_T^{iso} as a function of shower shape value for various shower shapes with no selections for lower p_T (12-14 GeV).
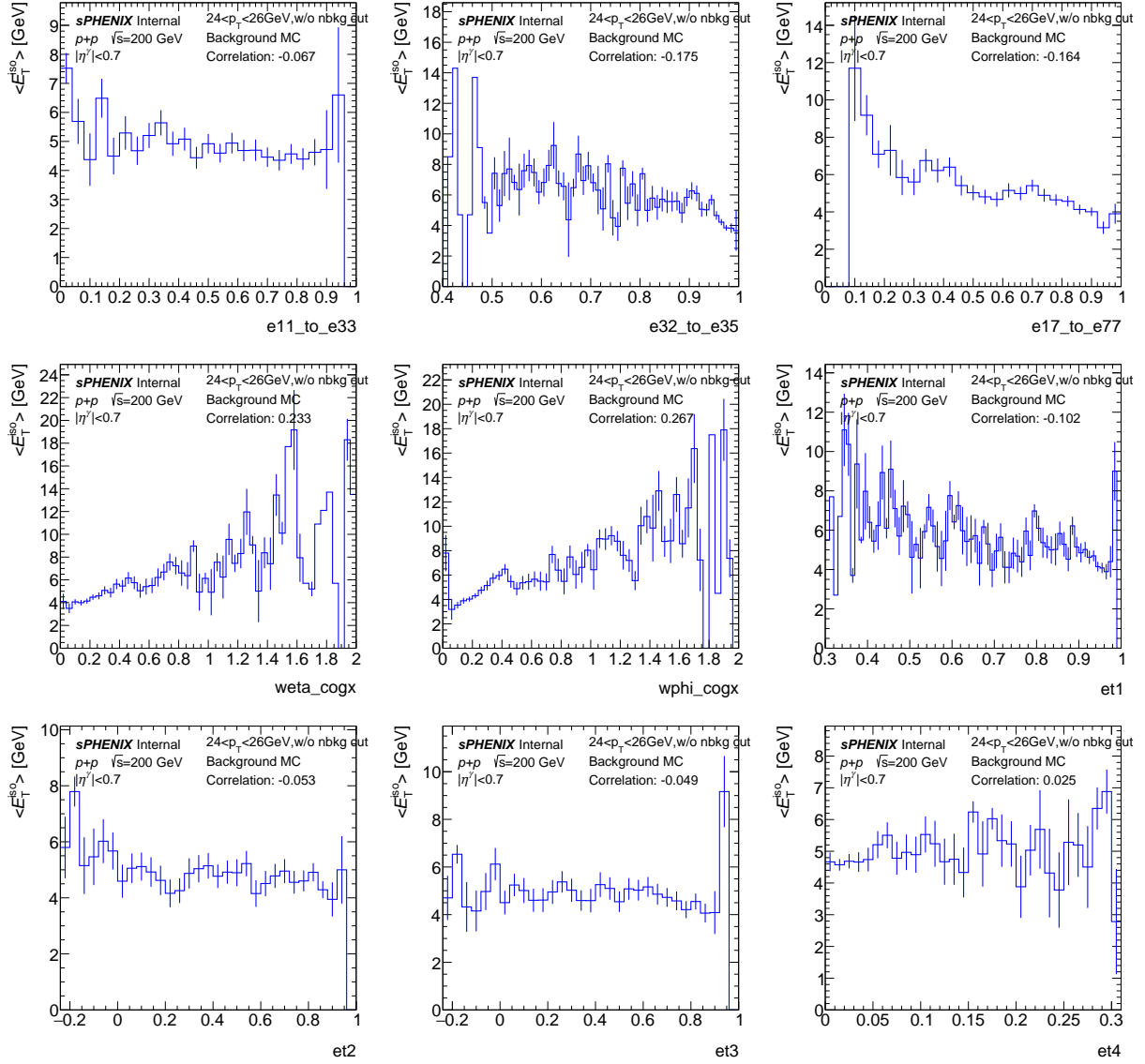
Figure 49: The average E_T^{iso} as a function of shower shape value for various shower shapes with no selections for higher p_T (24-26 GeV).

Figure 50: Distributions of E_T^{iso} for different E_T^γ bins in different panels, respectively. Photons with tight ID (black), non-tight ID (red) in inclusive MC and photons with tight ID in signal MC (blue) are overlaid.