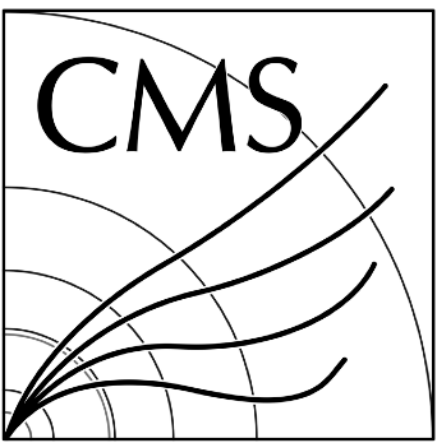




Electron and Photon Reconstruction and Identification Performance at CMS in 2022 and 2023

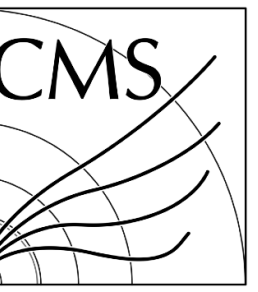
Shaowei Song, Junquan Tao, Pei-Zhu Lai, Mingshui Chen

CLHCP 2025

Oct 29 - Nov 02, 2025

Xinxiang, Henan

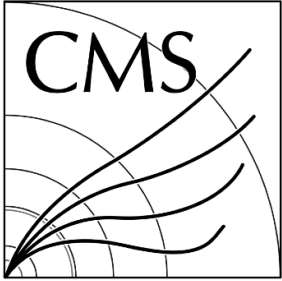
Outline



Electron & Photon

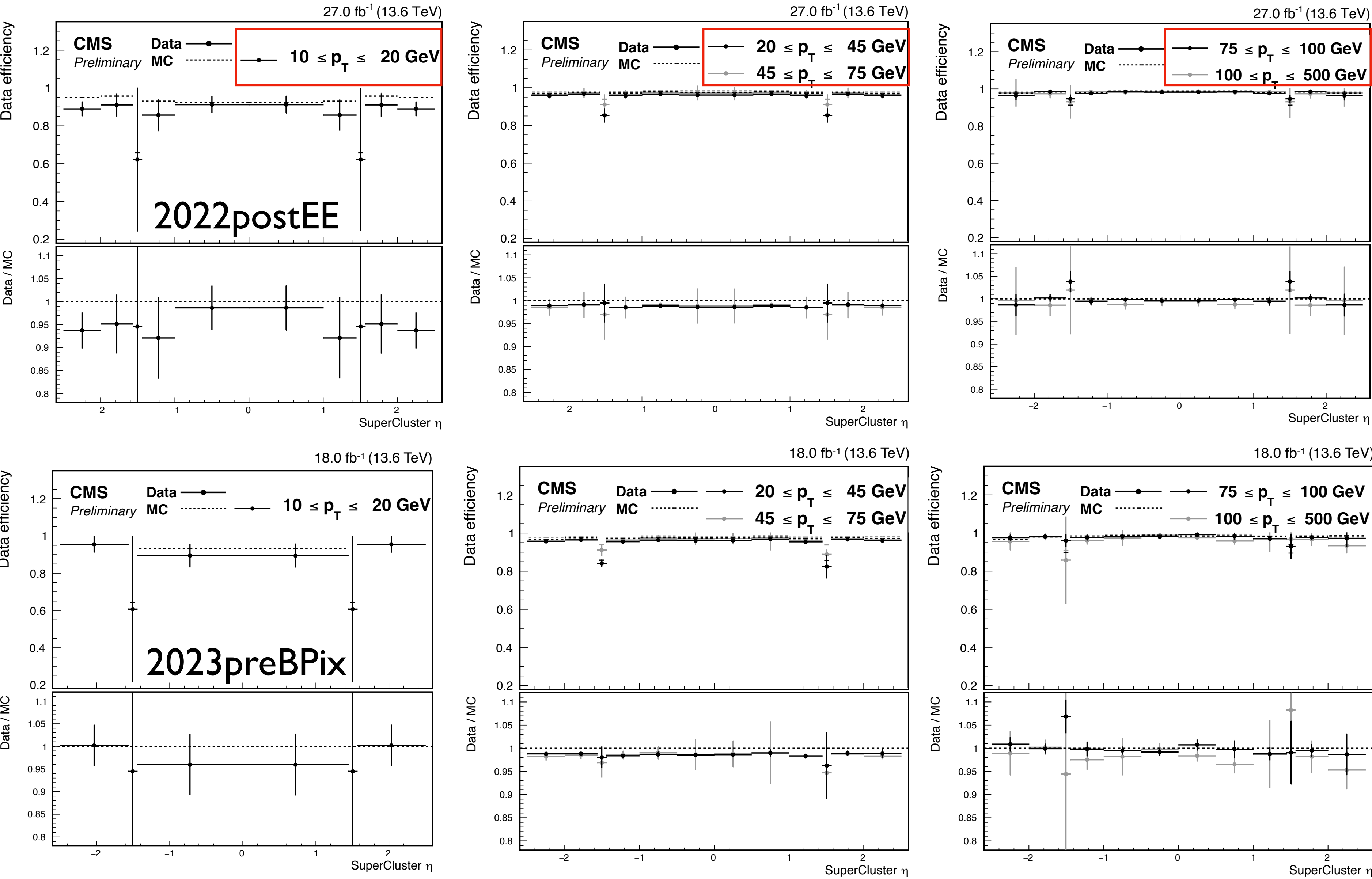
- Reconstruction Efficiency
- Identification (Cut-based / MVA)
 - Performance (Signal $\longleftrightarrow$ Background)
 - Efficiency & Scale Factors (Data $\longleftrightarrow$ MC)
- Energy Scale and Smearing Correction
- (Photon) Pixel Veto and Conversion-safe Electron Veto Scale Factors

Electron Reconstruction

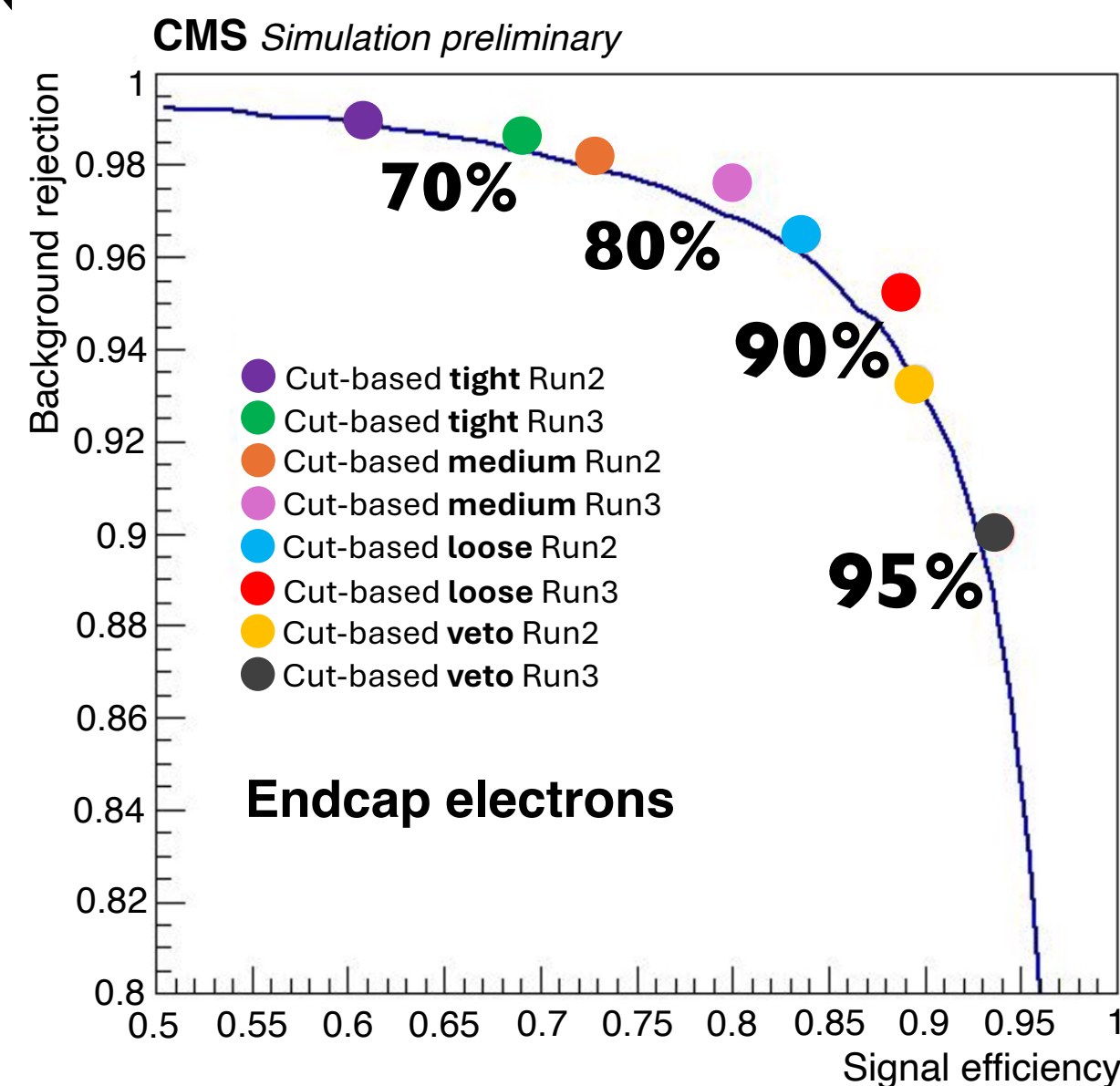
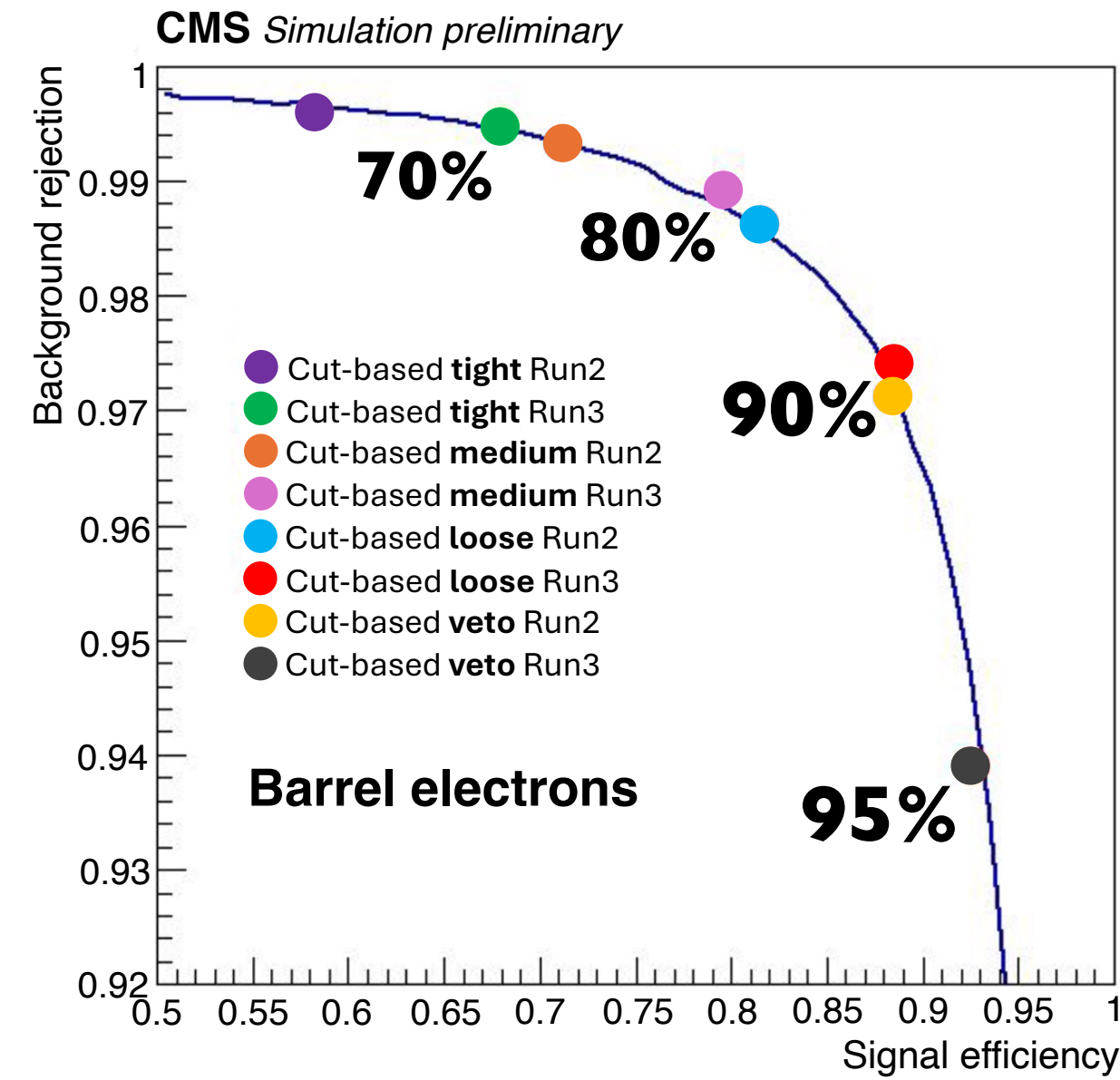
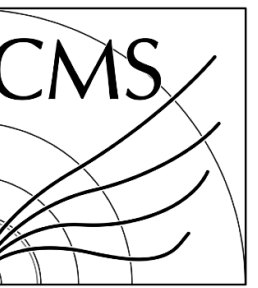


The electron reconstruction efficiency is defined as an efficiency for an ECAL supercluster to be matched with a track in the silicon tracker.

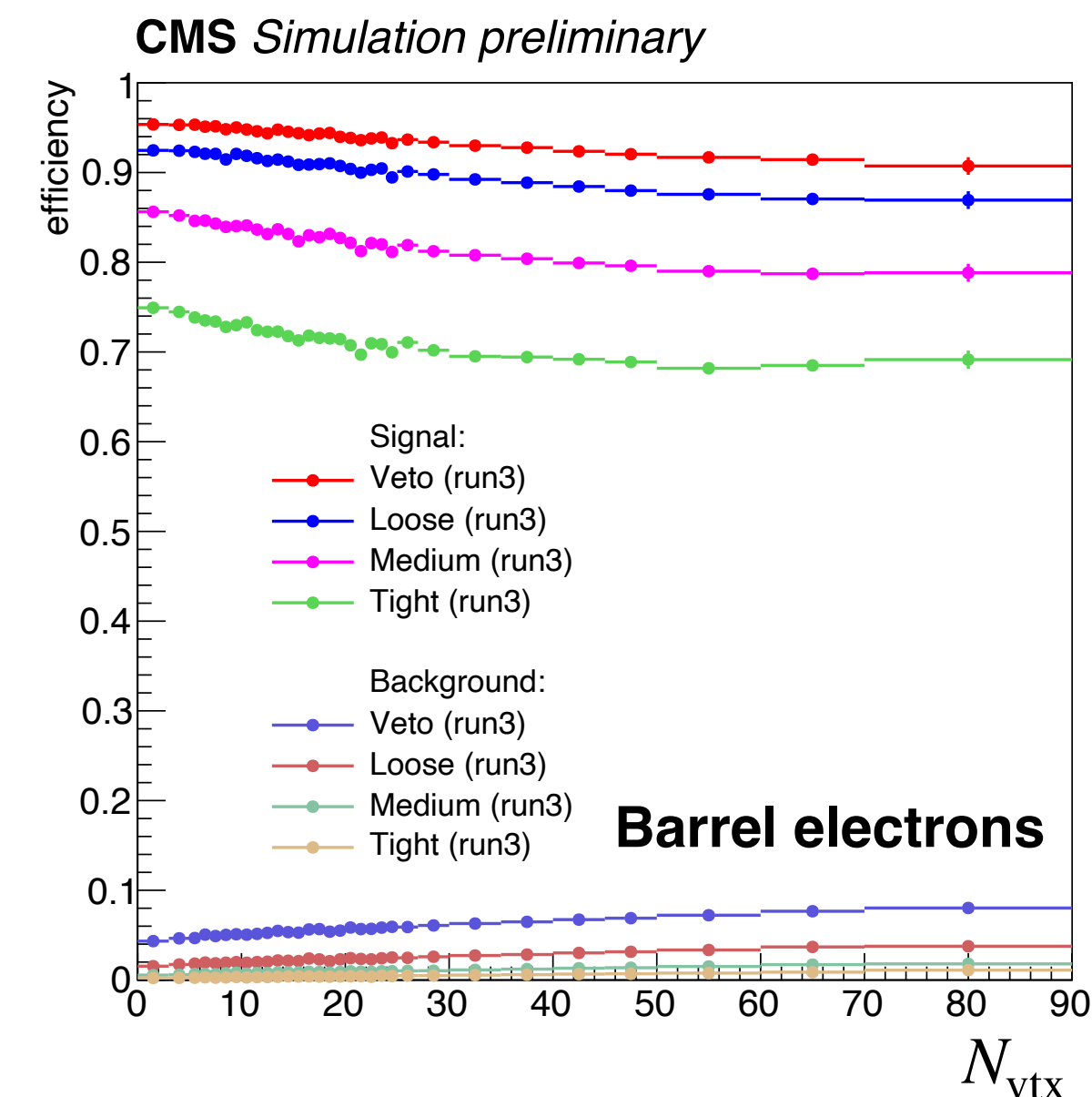
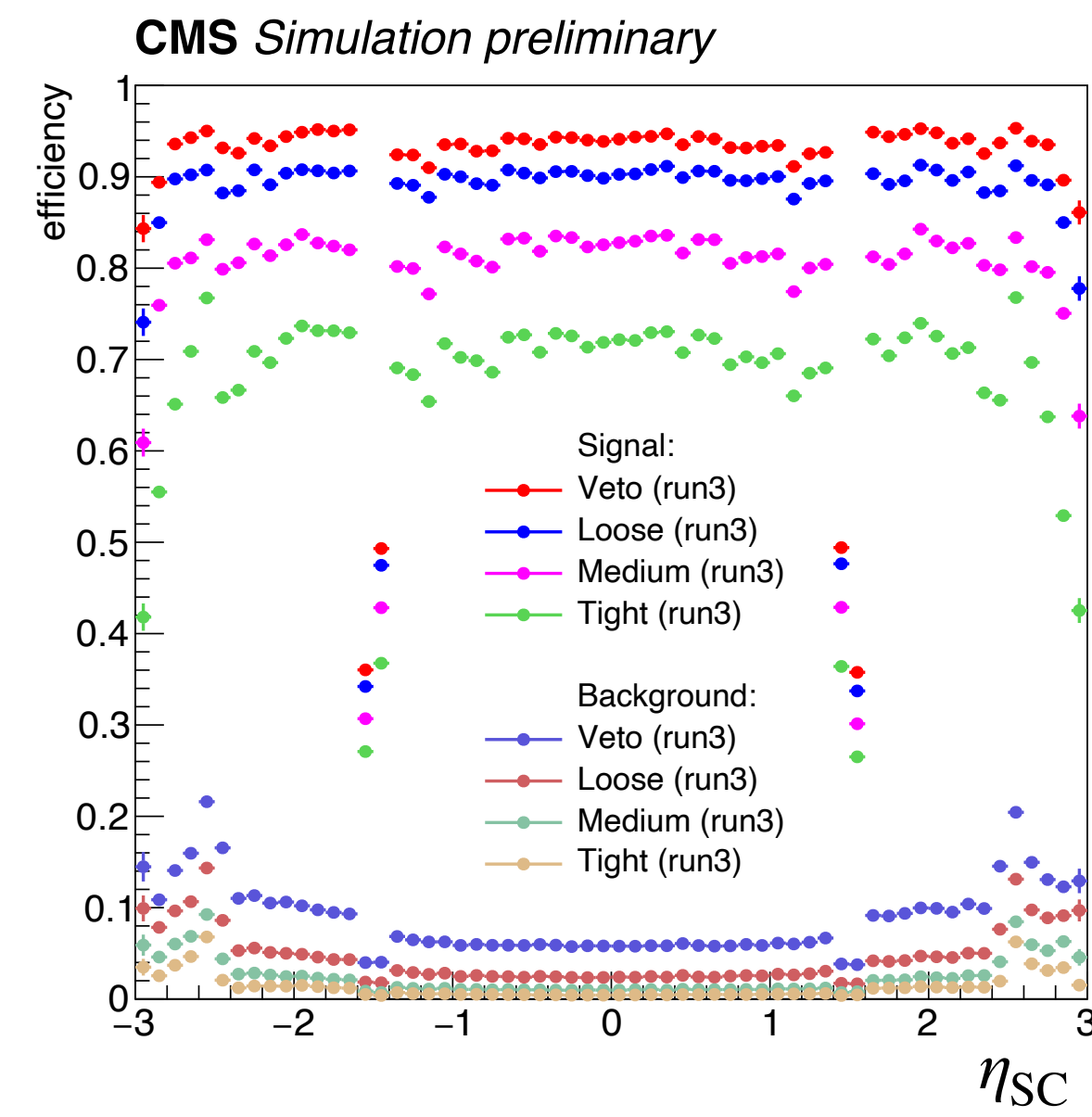
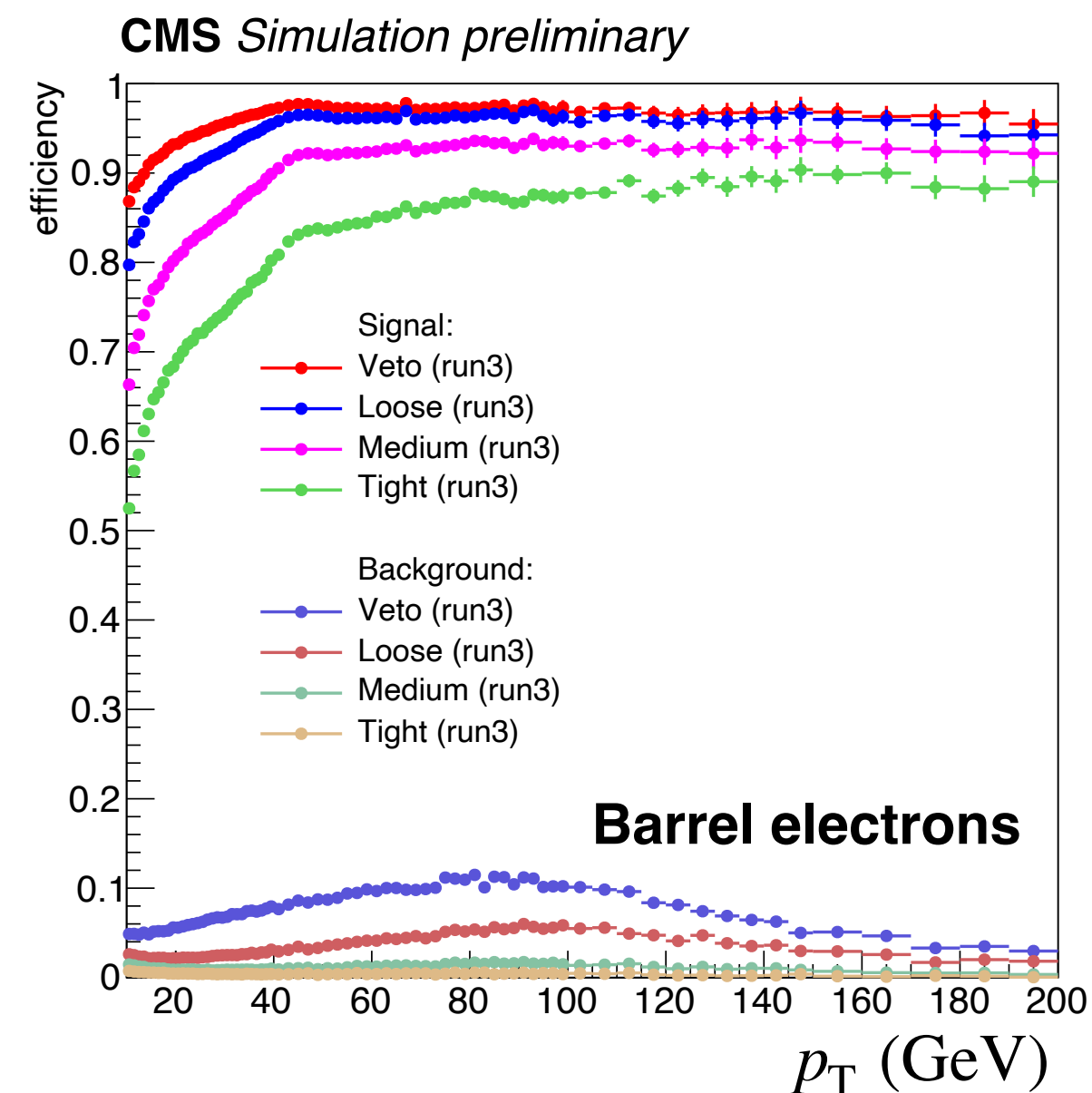
(2022postEE) prompt and reprocessed data
(2023preBPix) prompt data



Electron ID (Cut-based) Performance



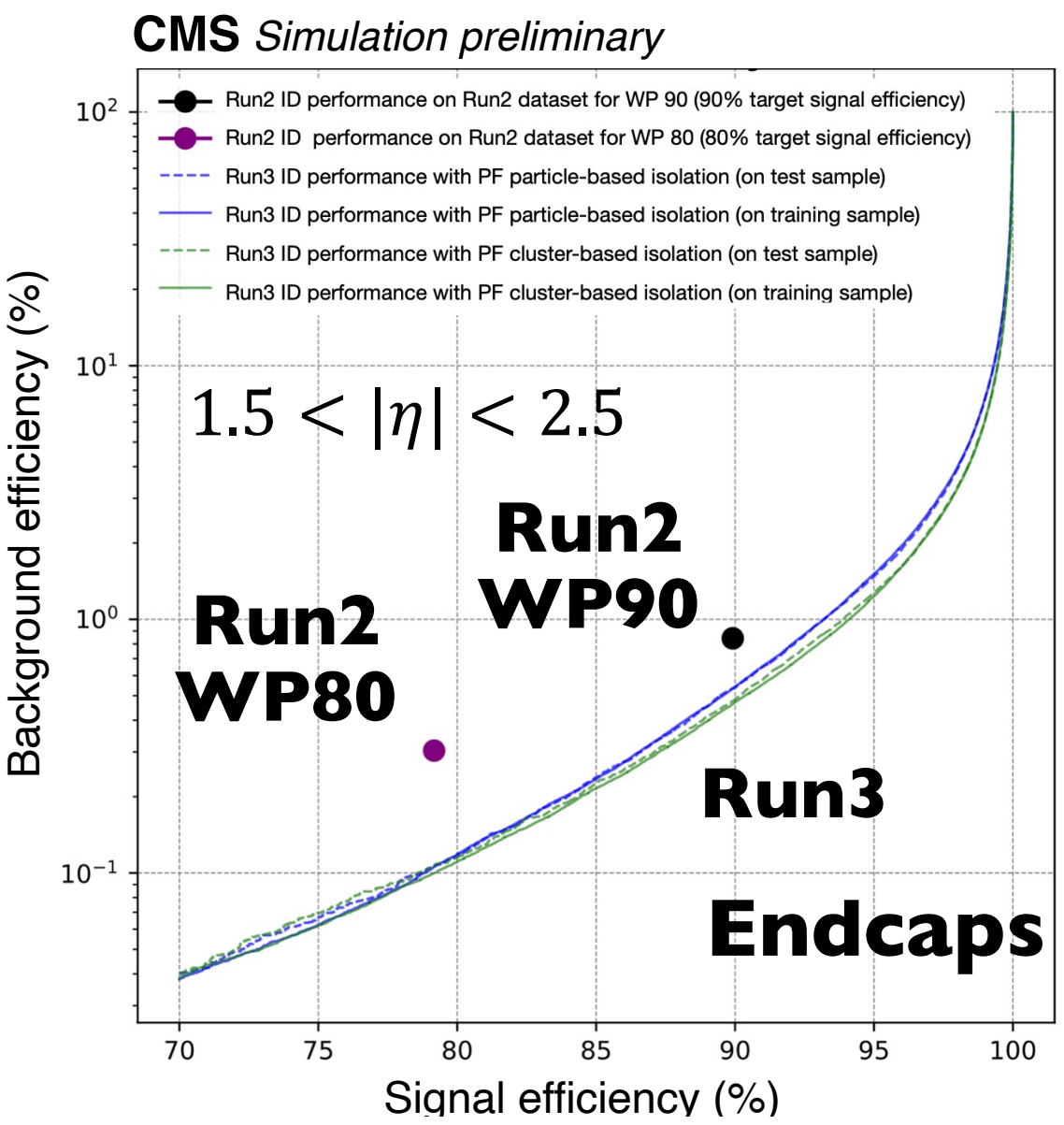
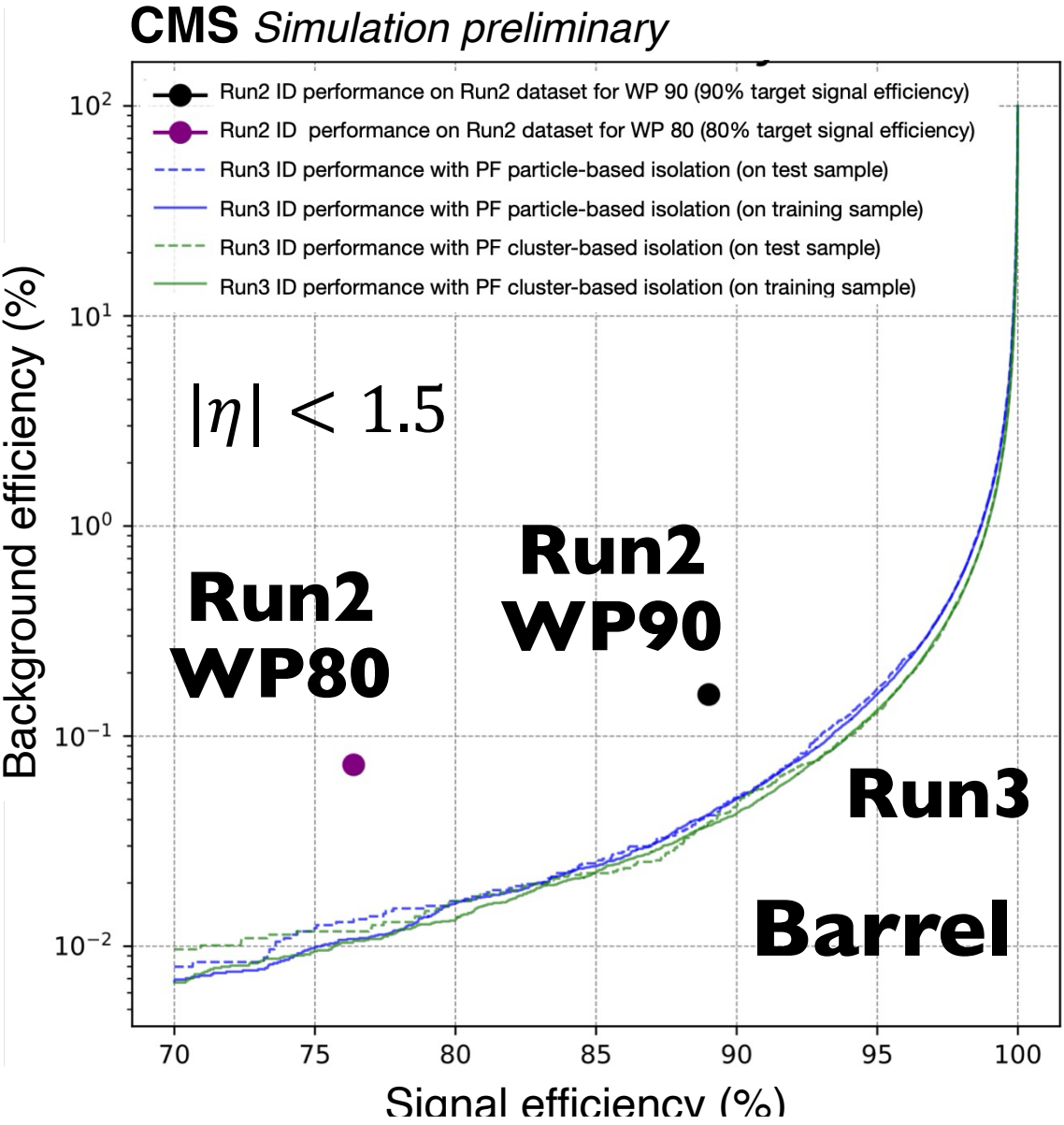
- Genetic algorithm [2][3] was used to optimize the electron cut-based ID.
- Utilized Variables: $\sigma_{in\eta}$, $|\eta_{seed} - \eta_{track}|$, $|\phi_{seed} - \phi_{track}|$, H/E , Isolation Variables, $|1/E - 1/p|$, # of missing hits, conversion veto
- The chosen working points, veto(~95%), loose(~90%), medium(~80%), and tight(~70%) were underscored. (Average Signal Efficiency over Detector Acceptance)



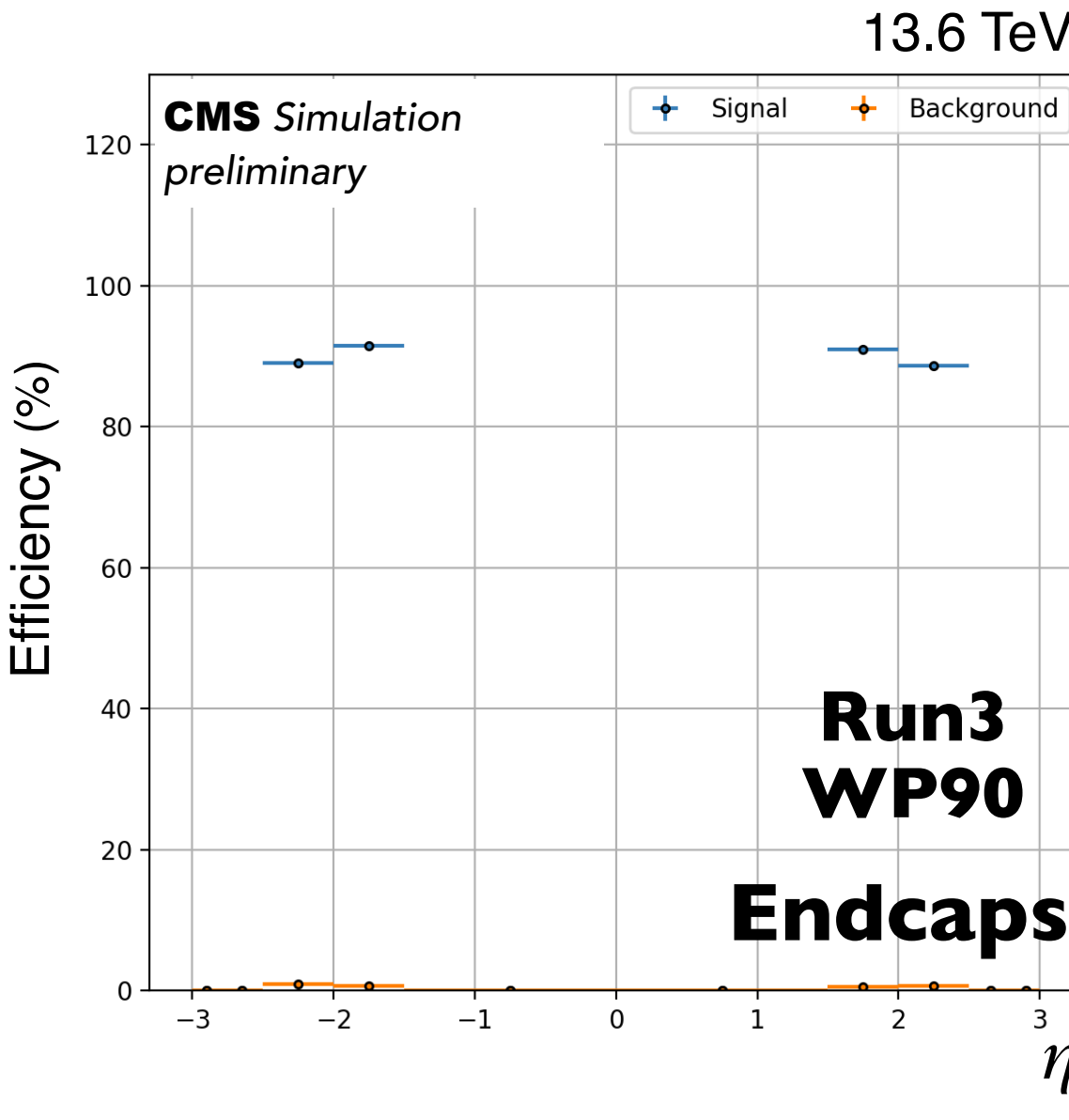
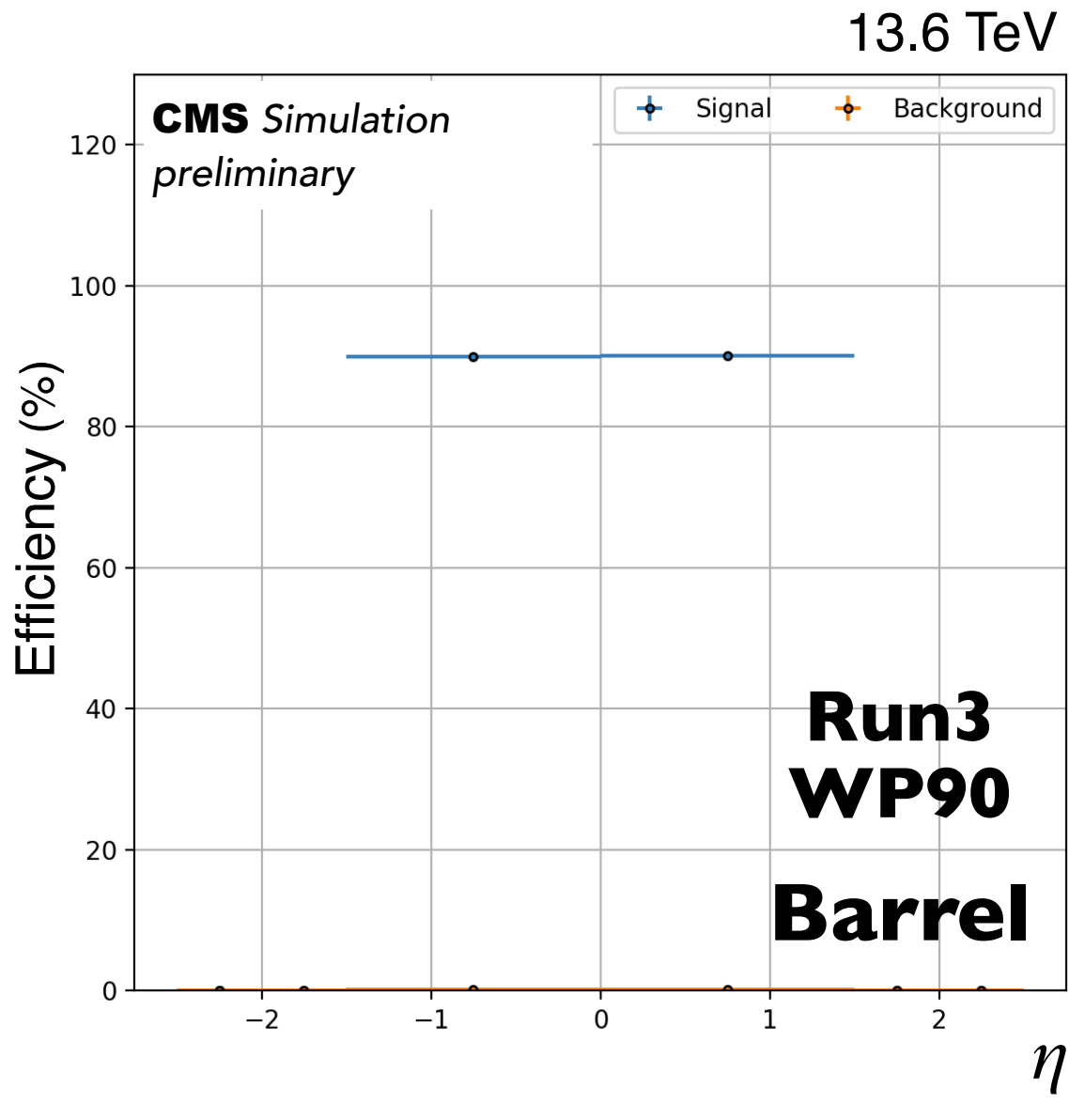
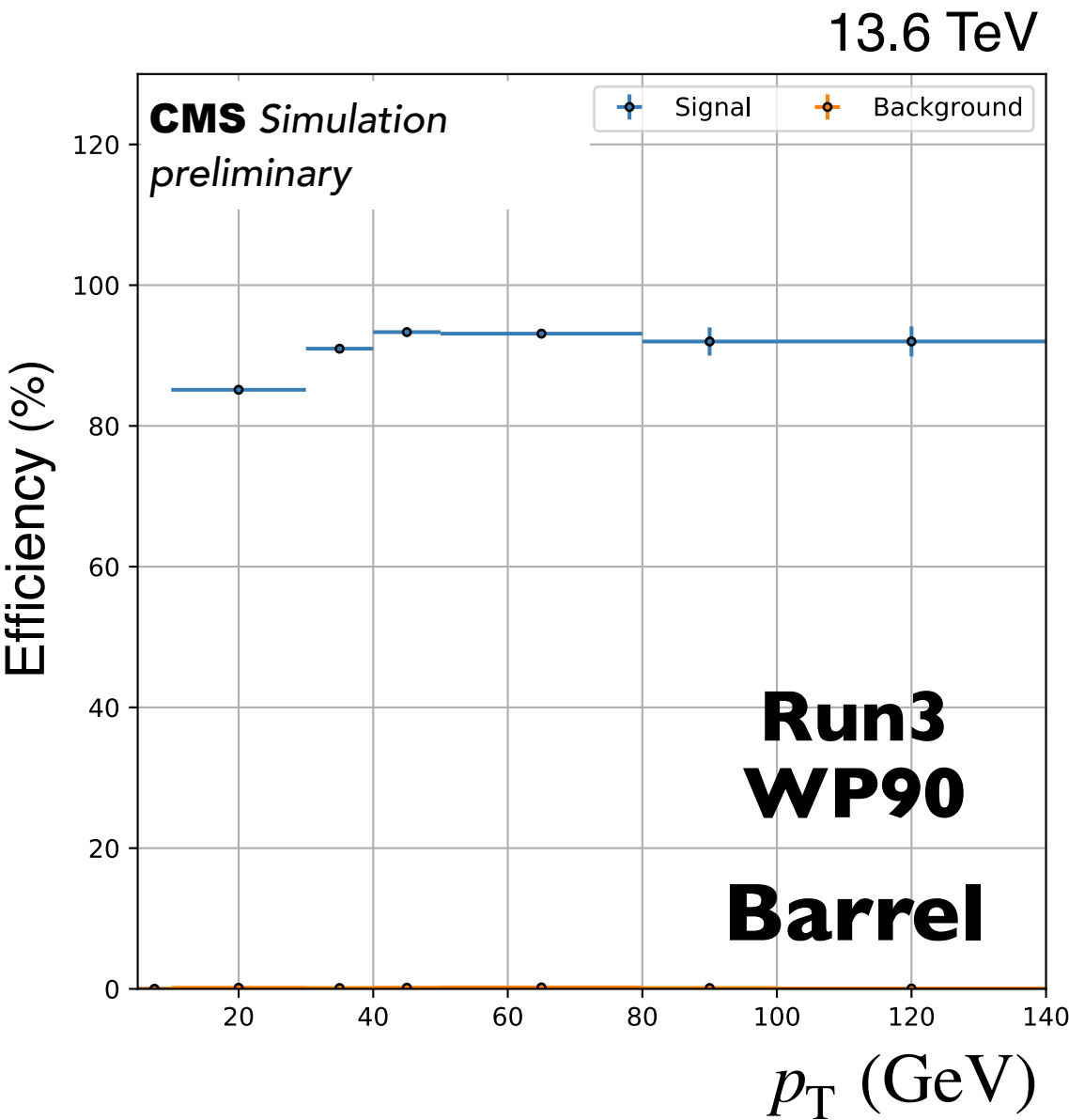
Sig Sample: DY aMC@NLO
Bkg Sample: $t\bar{t}$ semi-leptonic

Sig e: reconstructed electrons kinematically matched to prompt generator-level electrons.
Bkg e: reconstructed electrons not kinematically matched to prompt generator-level electrons.

Electron ID (MVA) Performance

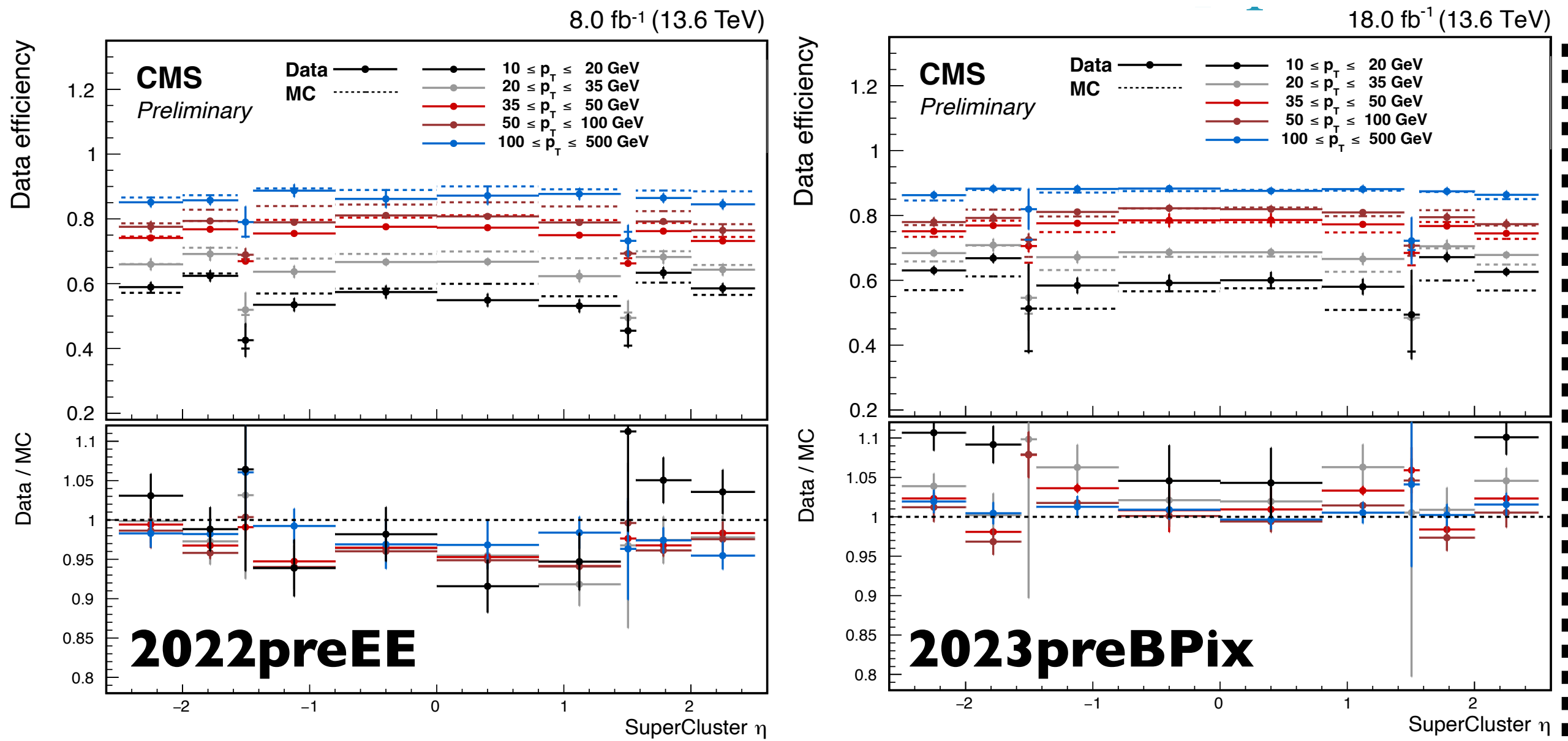


- XGBoost [2][3] was used to optimize the electron MVA ID
- Utilized Variables: $\sigma_{i\eta i\eta}$, $|\eta_{seed} - \eta_{track}|$, $|\phi_{seed} - \phi_{track}|$, H/E , Isolation Variables, $|1/E - 1/p|$, # of missing hits, conversion veto, and so on.
- Run3 electron MVA IDs perform better than run2 ones, for more training variables and state-of-the-art MVA algorithm used in run3 [5].



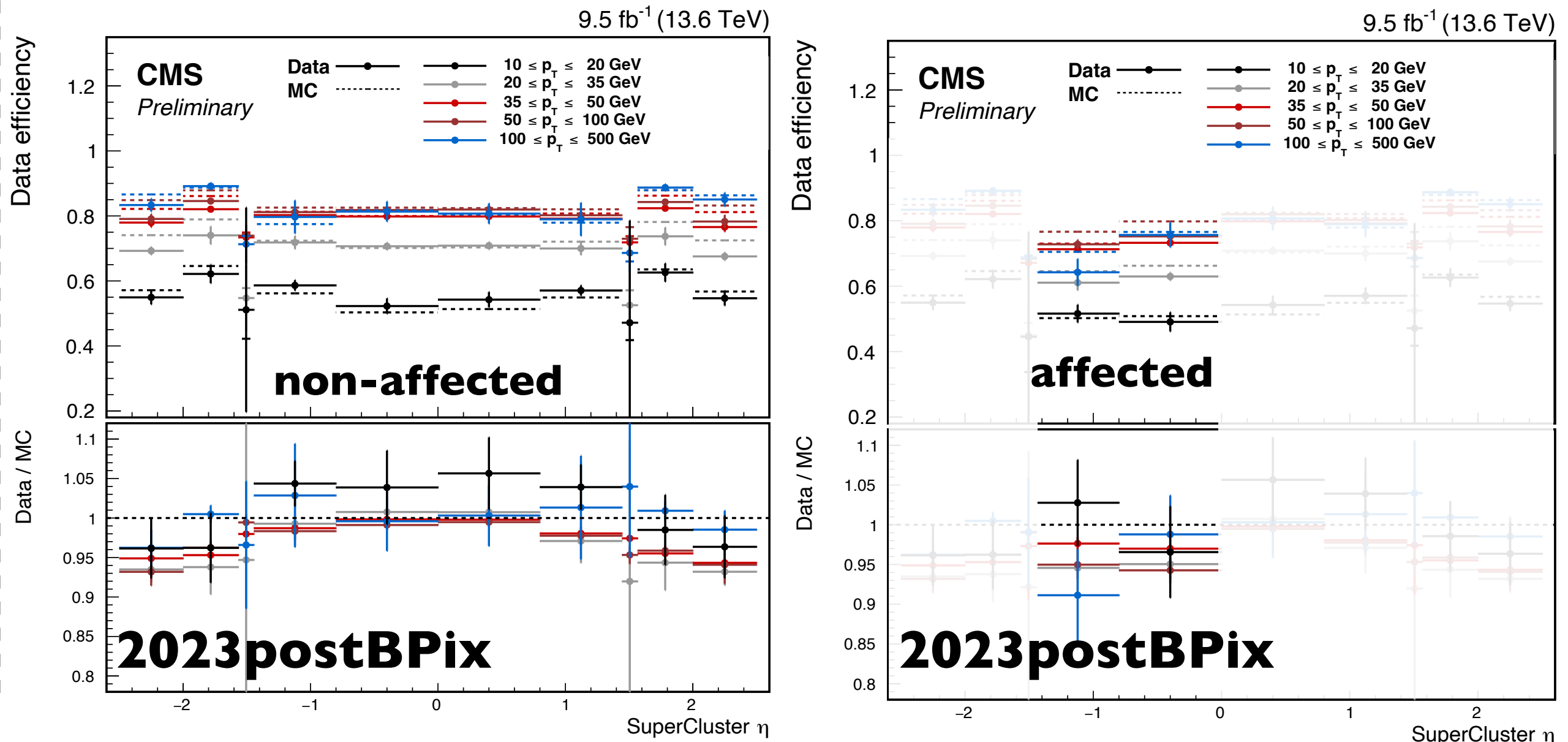
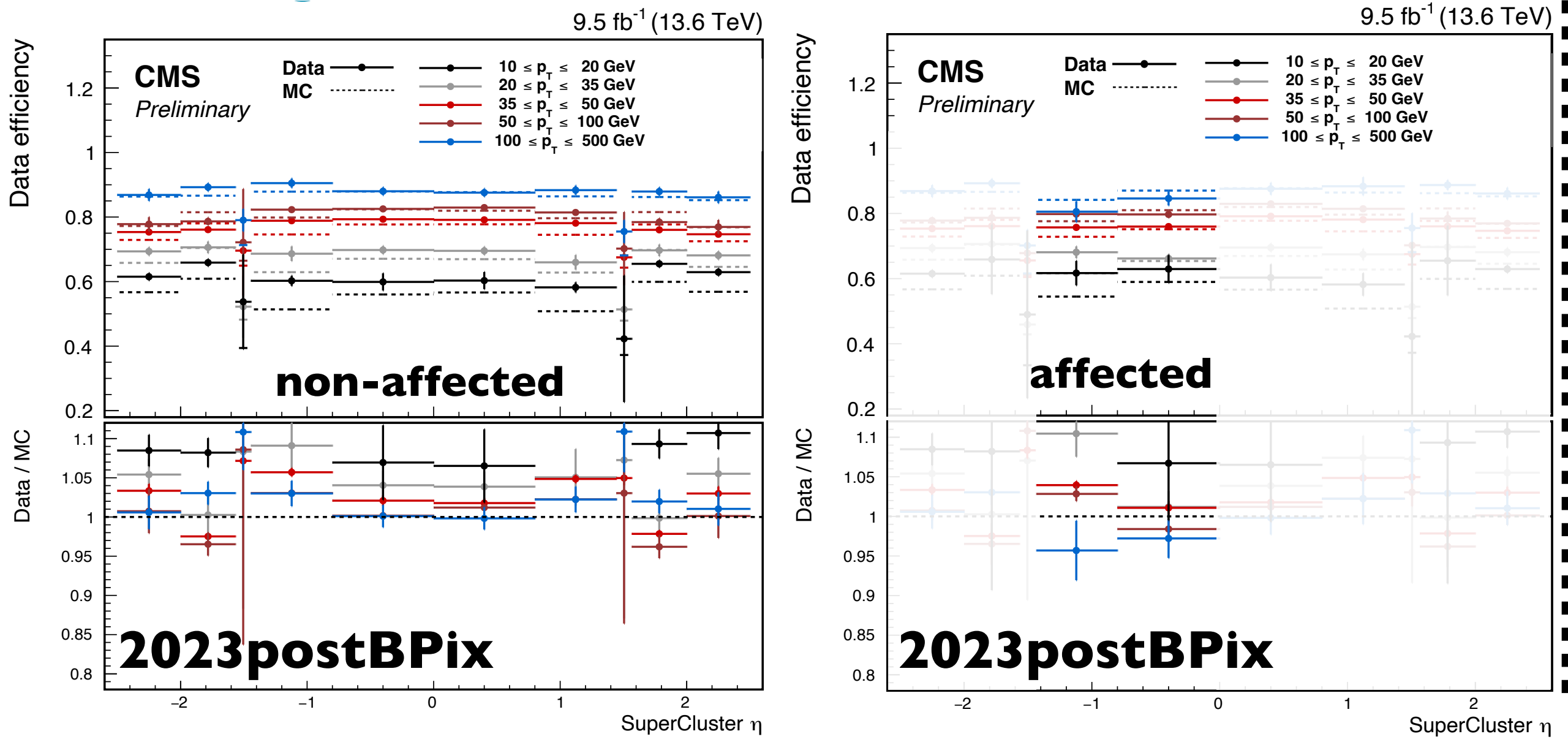
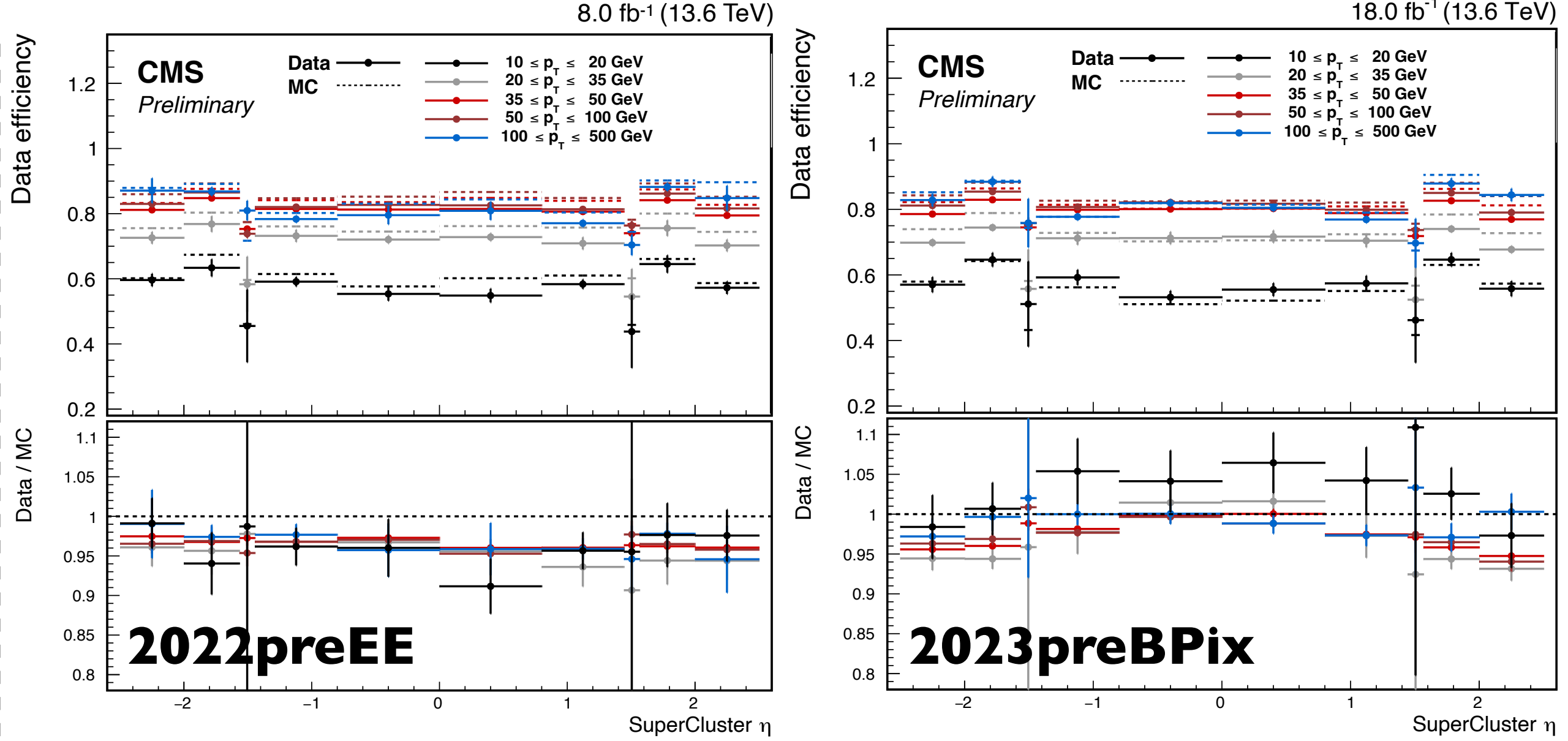
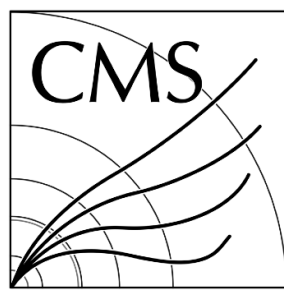
Tight Electron cut-based ID Efficiency

$10 < p_T < 500$ GeV

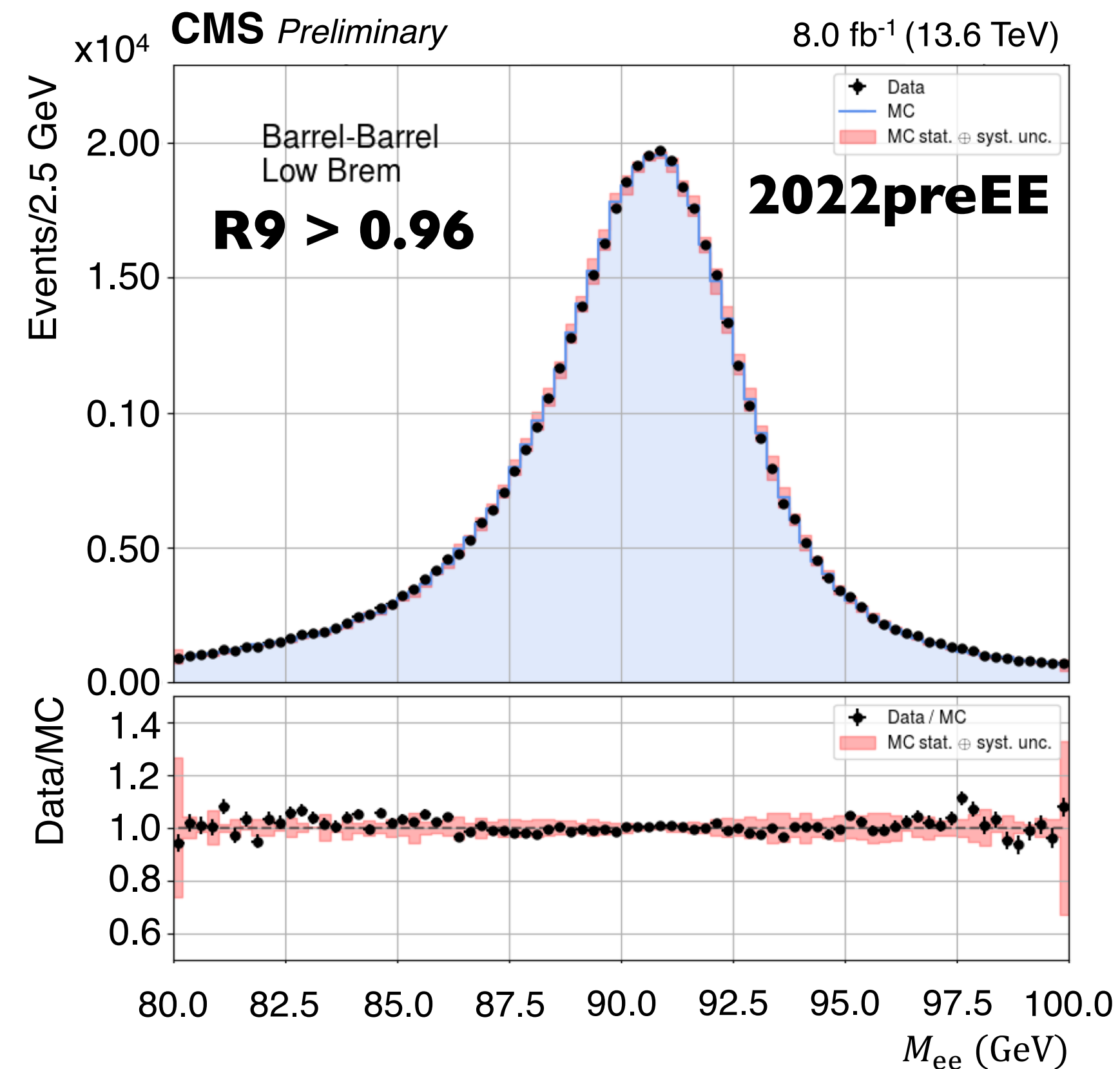
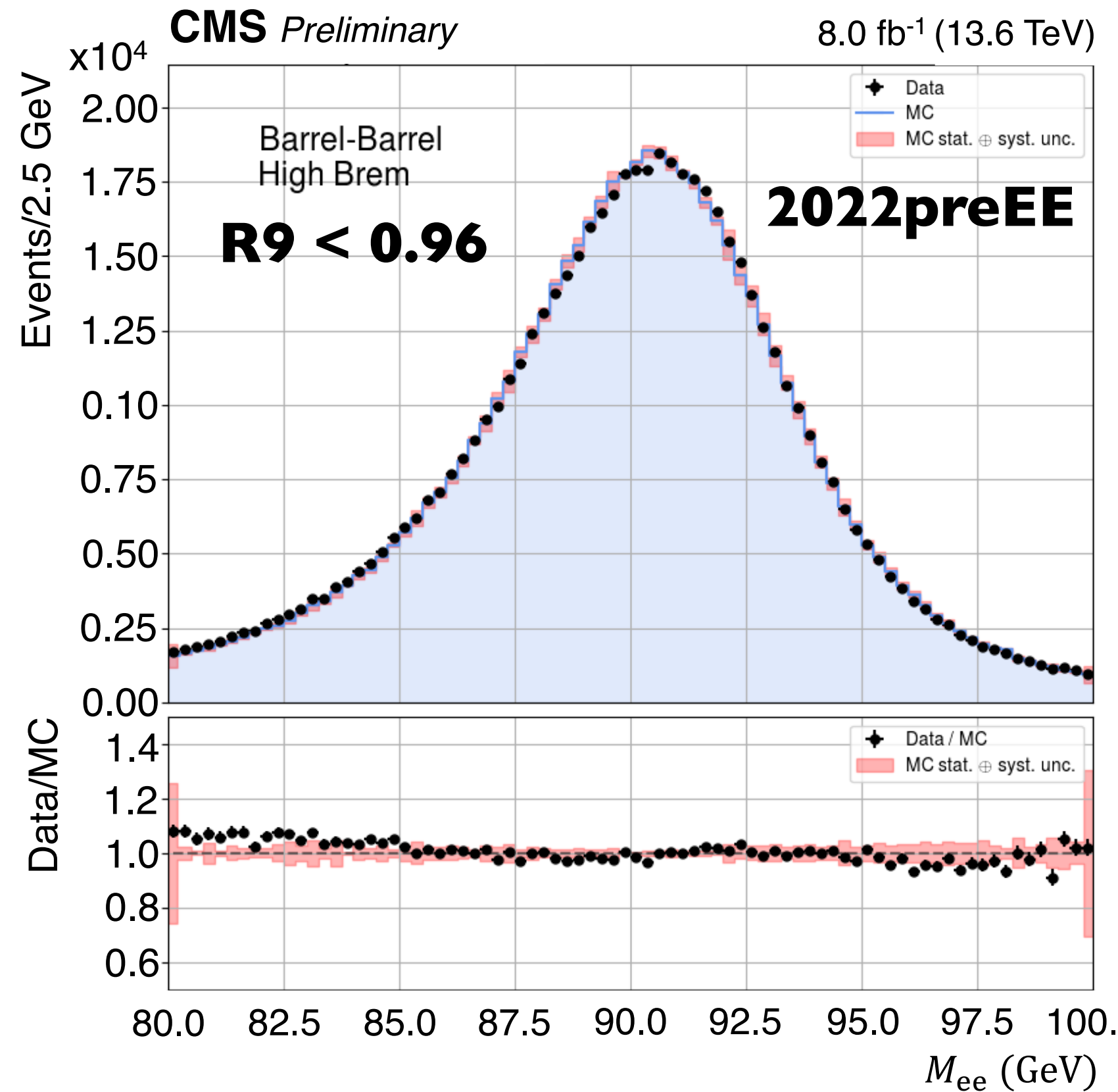
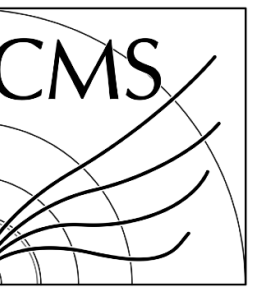


WP80 Electron MVA ID Efficiency

$10 < p_T < 500$ GeV

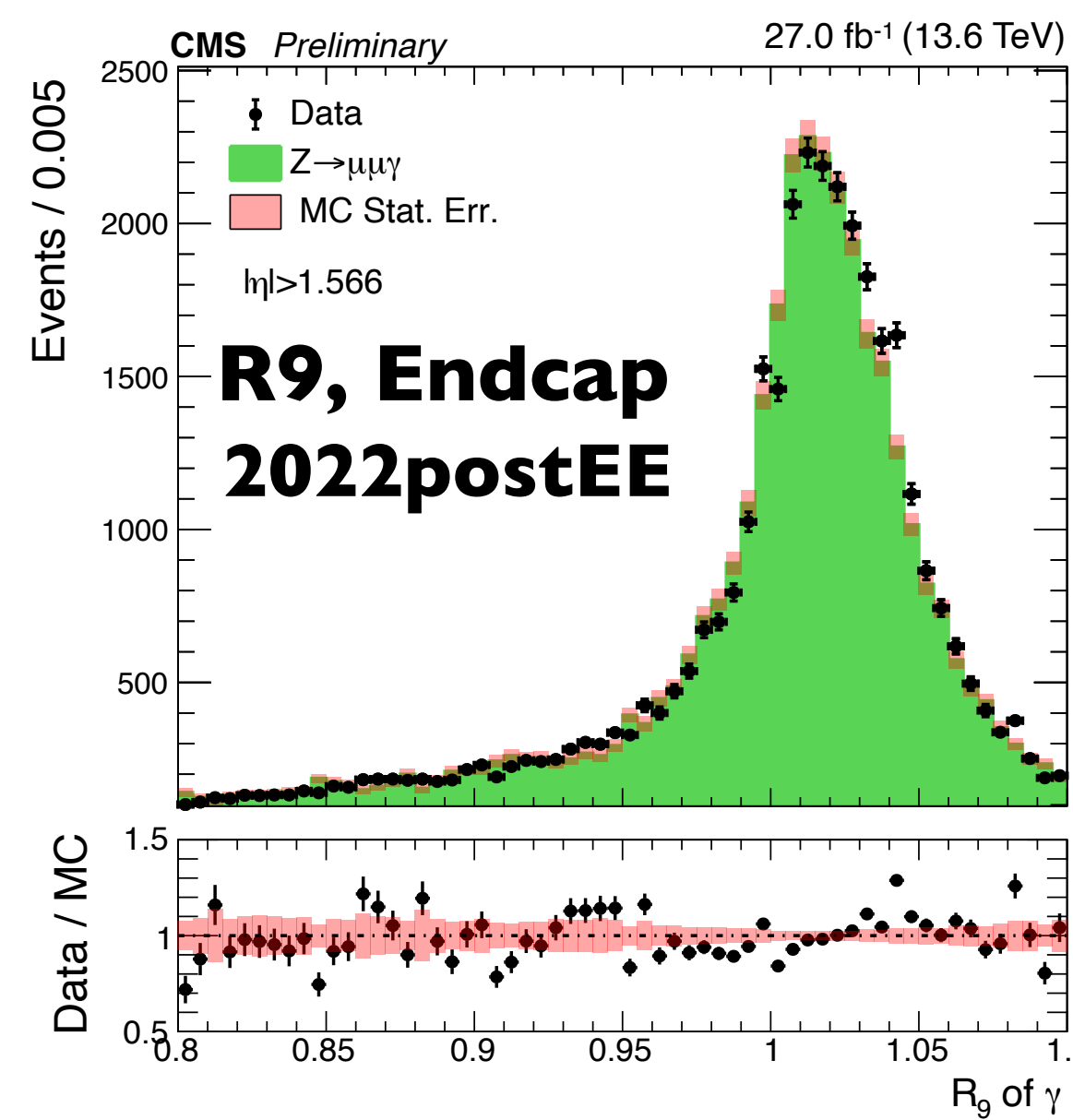
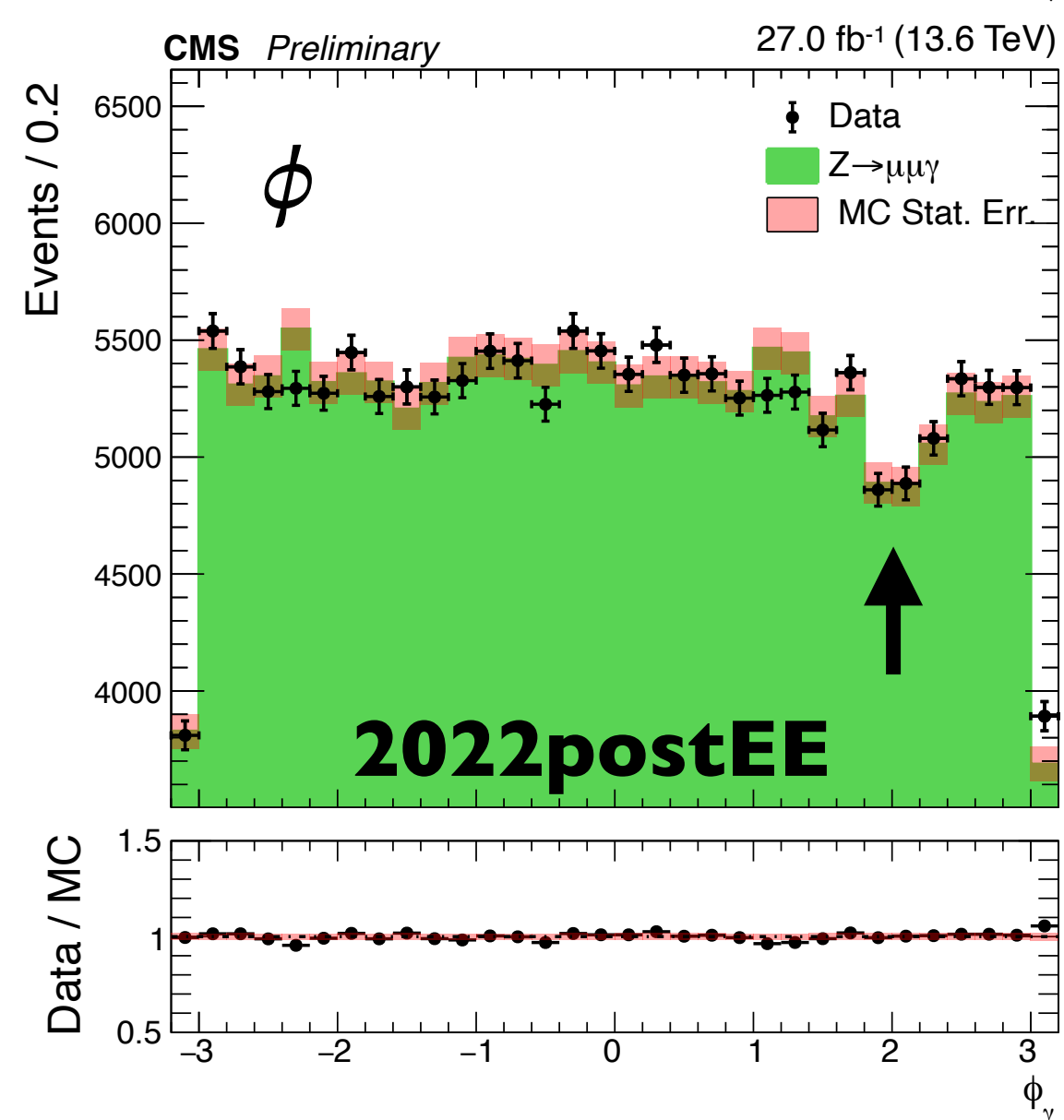
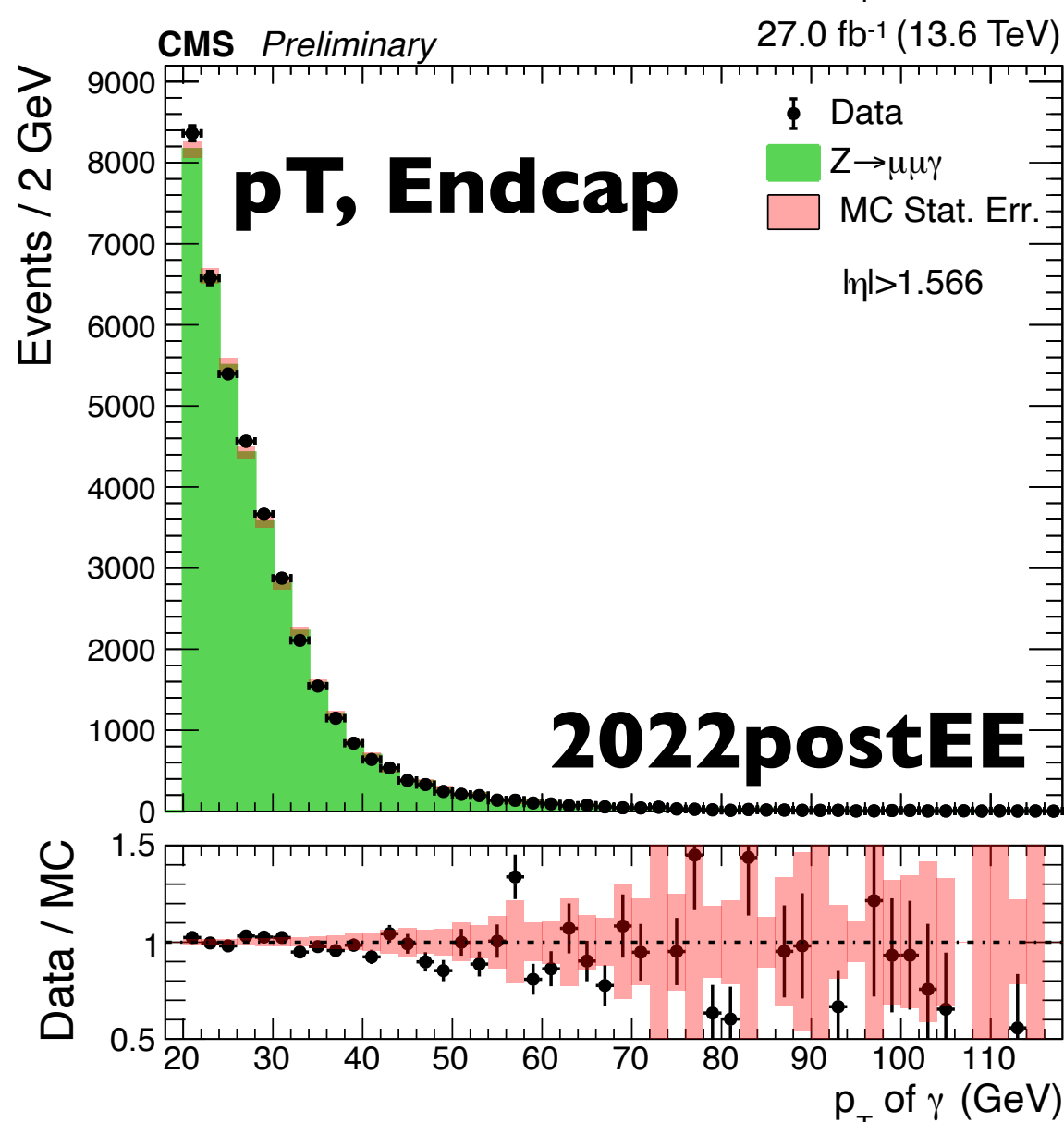
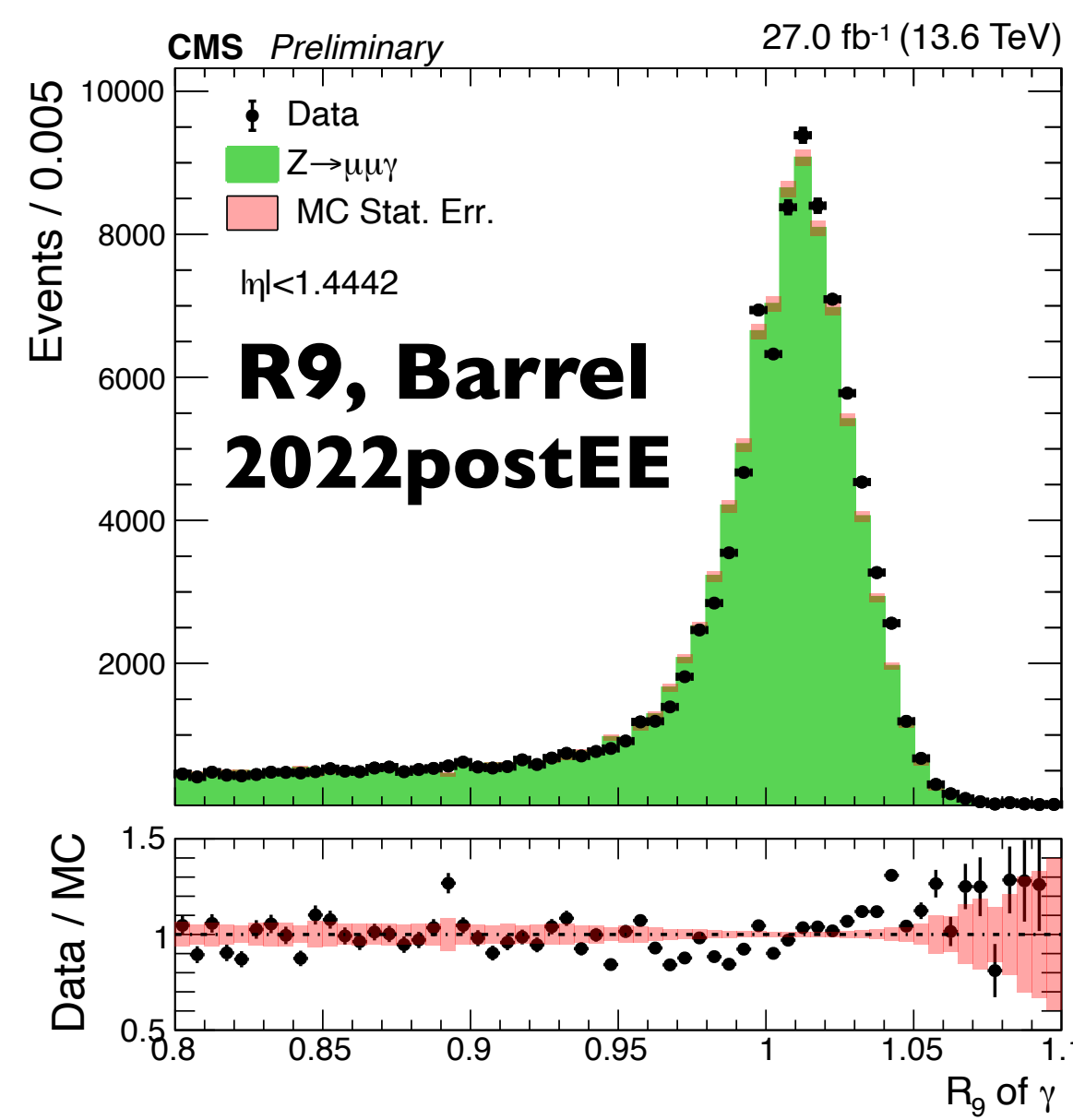
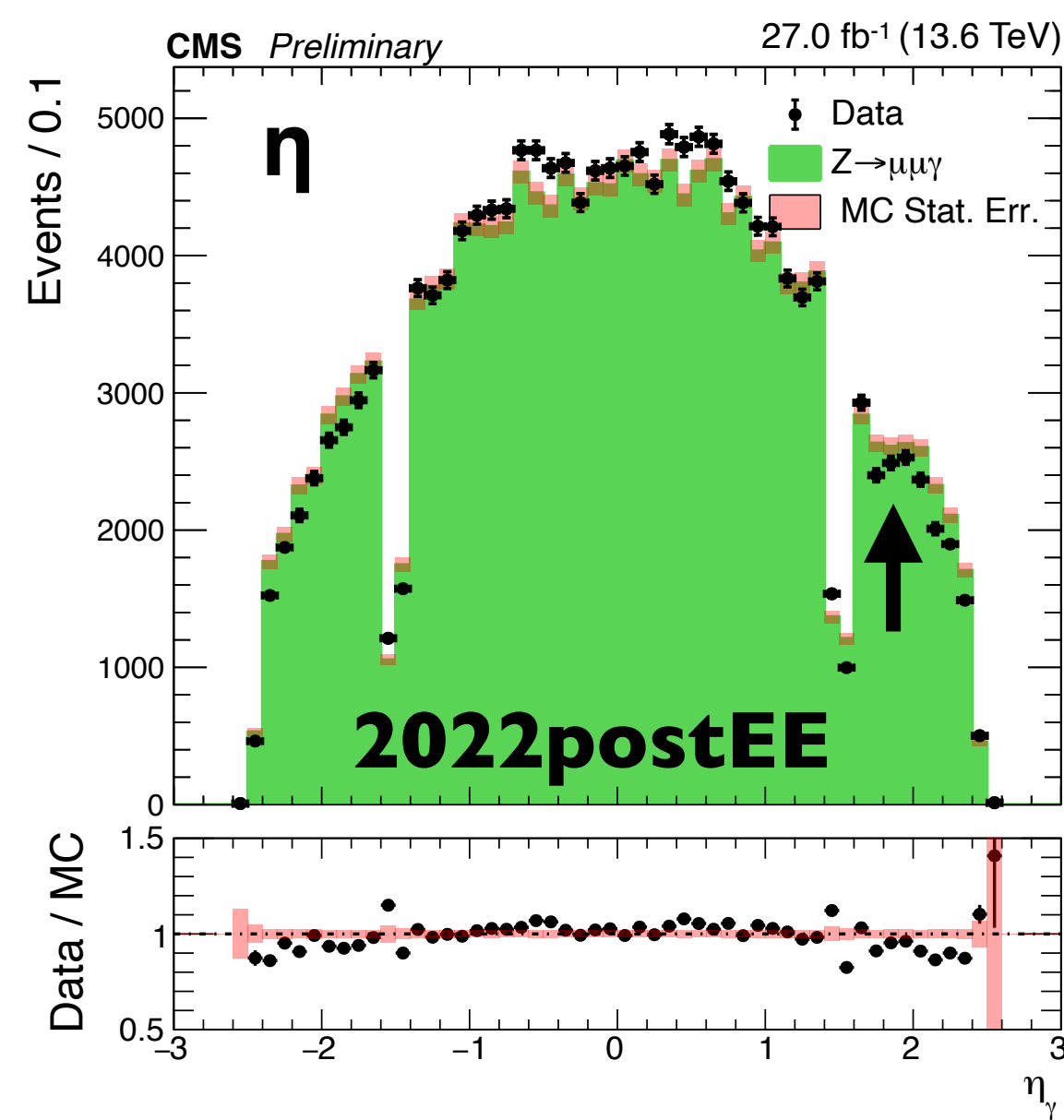
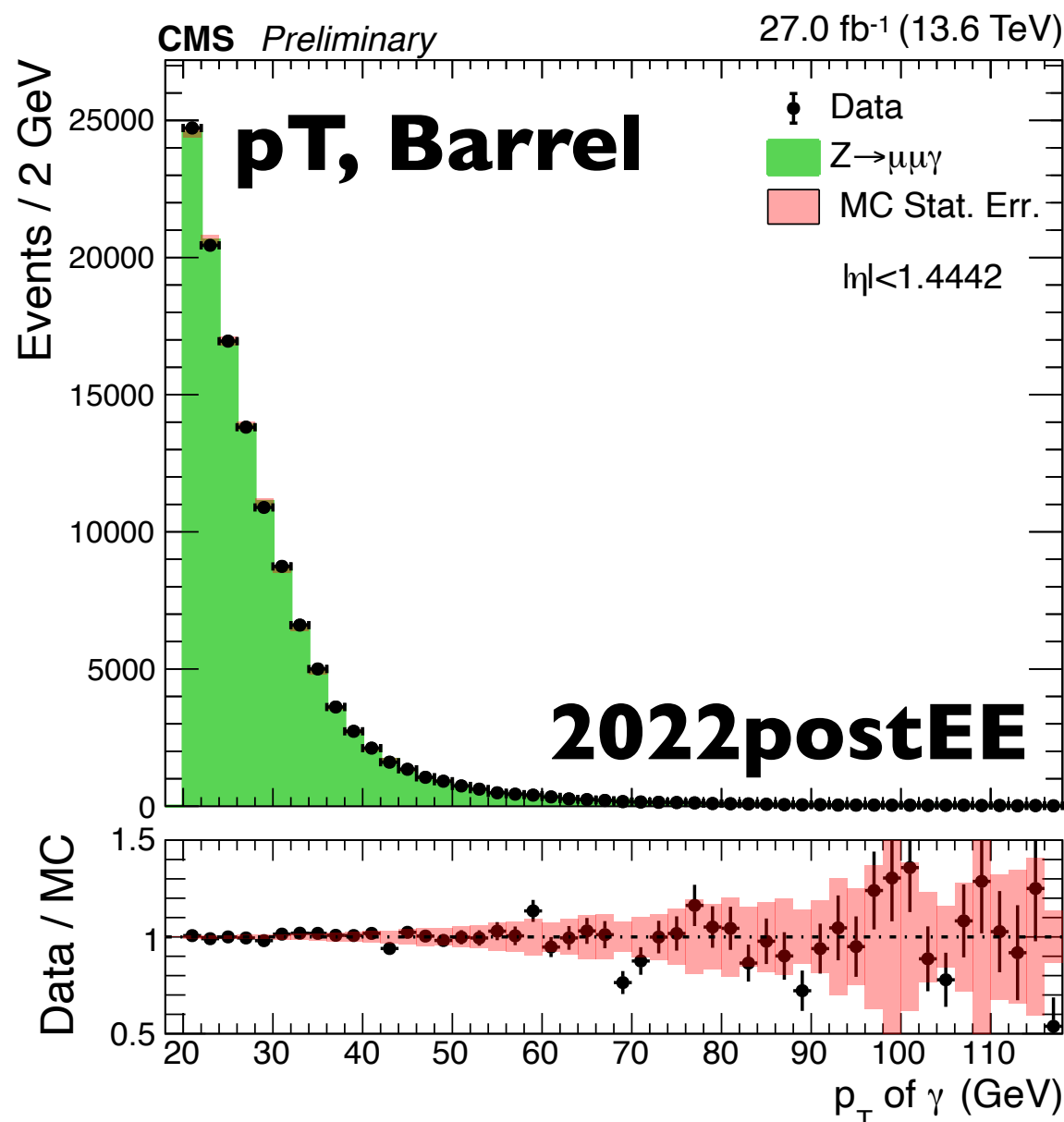
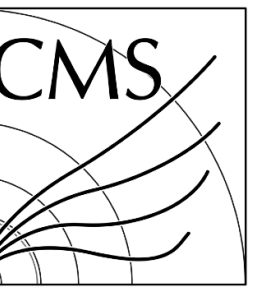


Electron Energy Scale and Smearing Corrections



- Di-electron invariant mass after applying the scale and smearing corrections
- Barrel-Barrel: both the electrons are reconstructed in the ECAL barrel.
- High Brem: high-bremsstrahlung ($R9 < 0.96$) electrons
- Low Brem: low-bremsstrahlung ($R9 > 0.96$) electrons

Photon Variables



Z → μμγ(FSR)

■ **Event Selections:**

$$60 < M_{\mu^+\mu^-\gamma} < 120 \text{ GeV}$$

$$35 \text{ GeV} < M_{\mu^+\mu^-}$$

$$M_{\mu^+\mu^-} + M_{\mu^+\mu^-\gamma} < 180 \text{ GeV}$$

$$\min(\Delta R(\mu^\pm, \gamma)) < 0.8 \text{ GeV}$$

■ **Muon Selections:**

Pass high-purity ID

Isolation from PF charged particles < 0.2

$$p_T > 20 \text{ (10) GeV}$$

$$|\eta| < 2.4$$

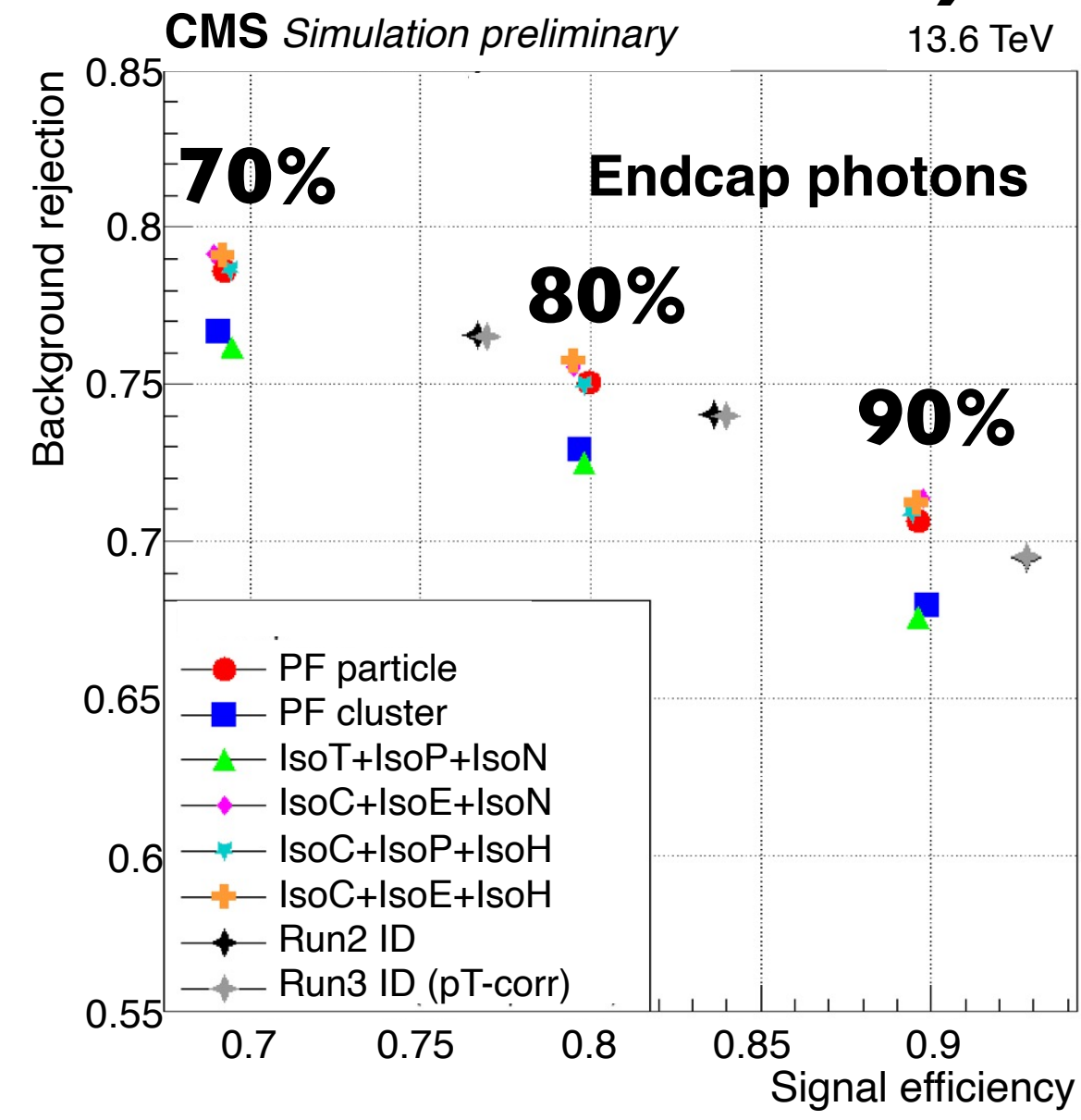
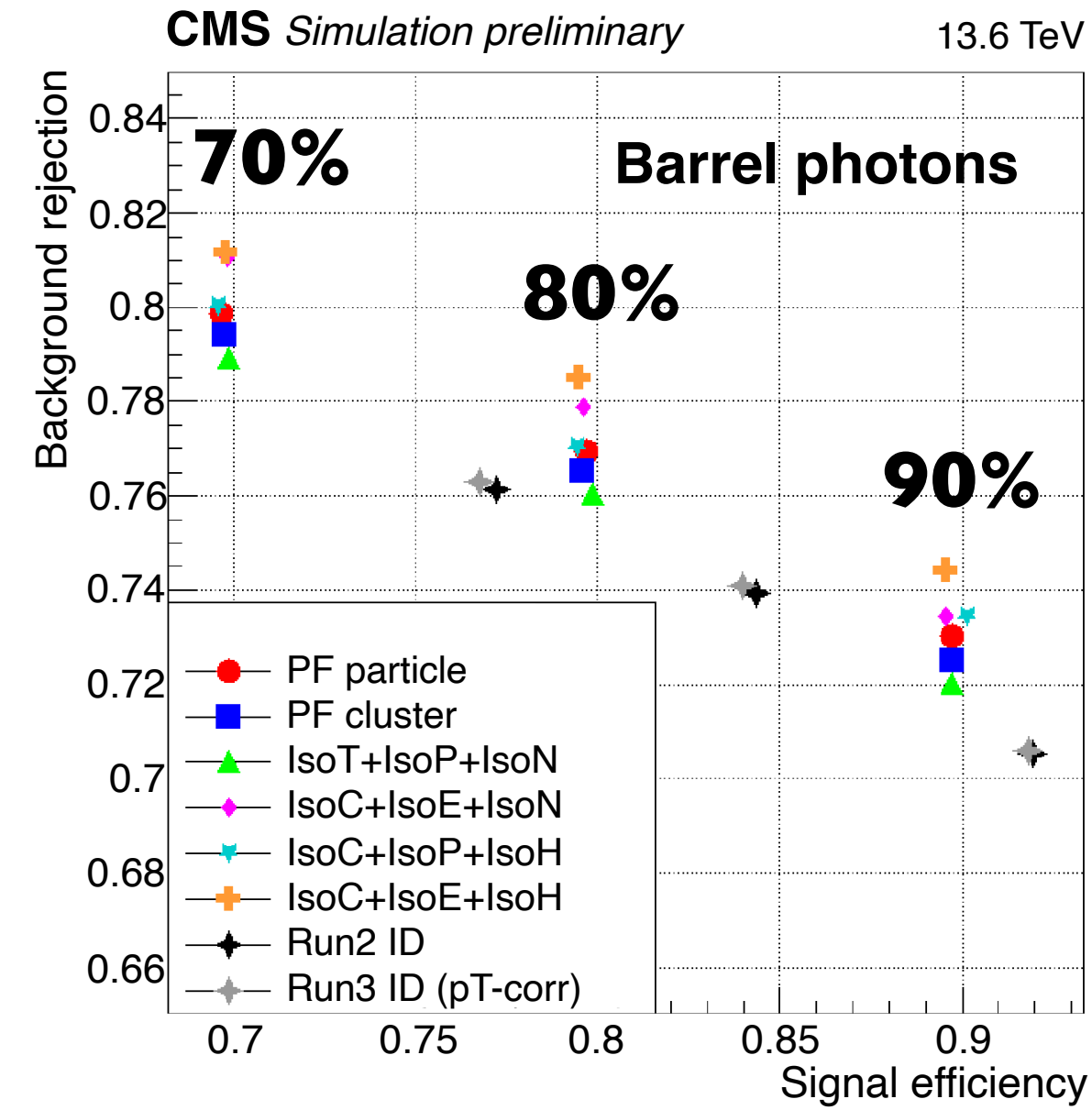
■ **Photon Selections:**

$$p_T > 20 \text{ GeV}$$

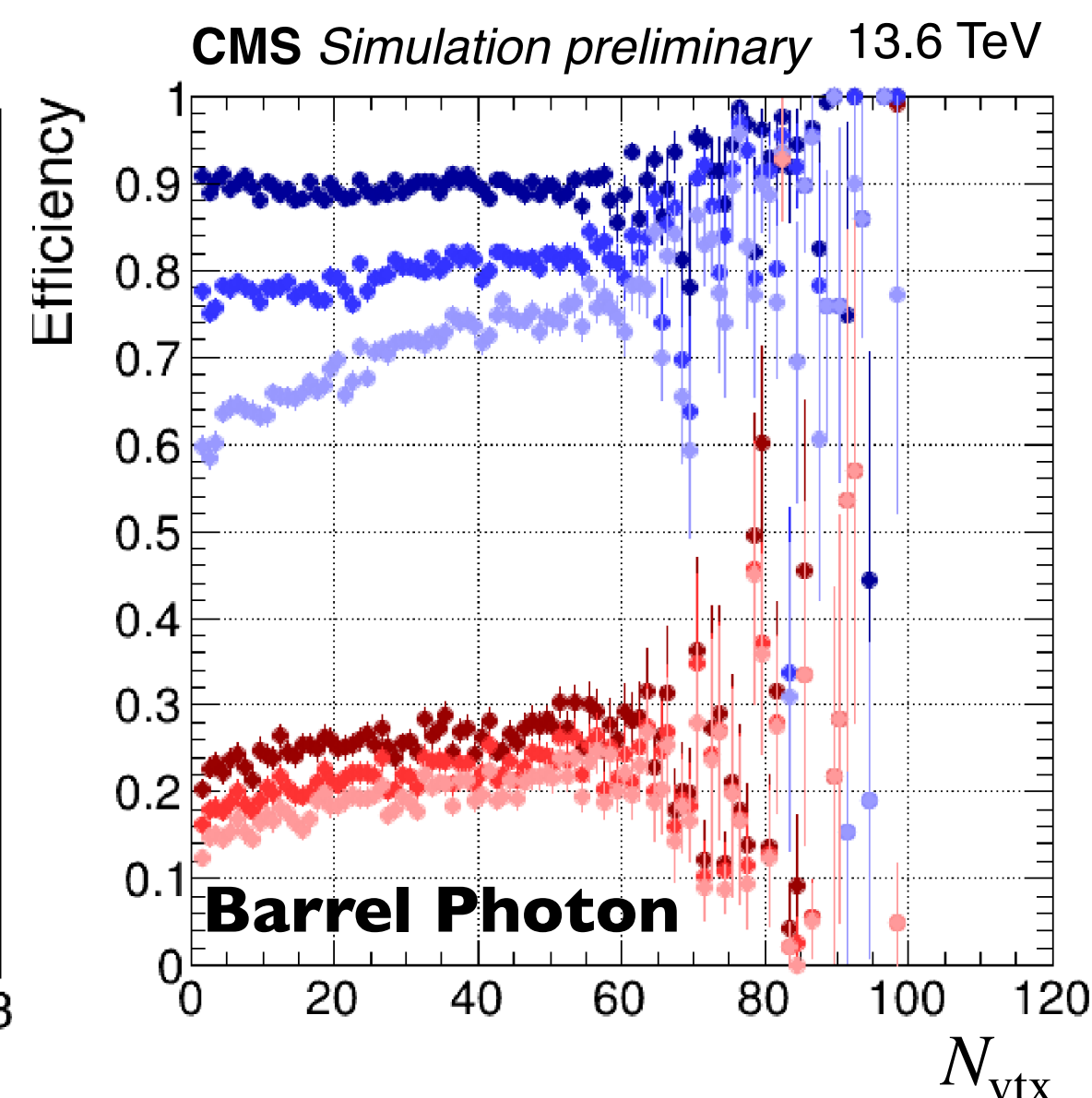
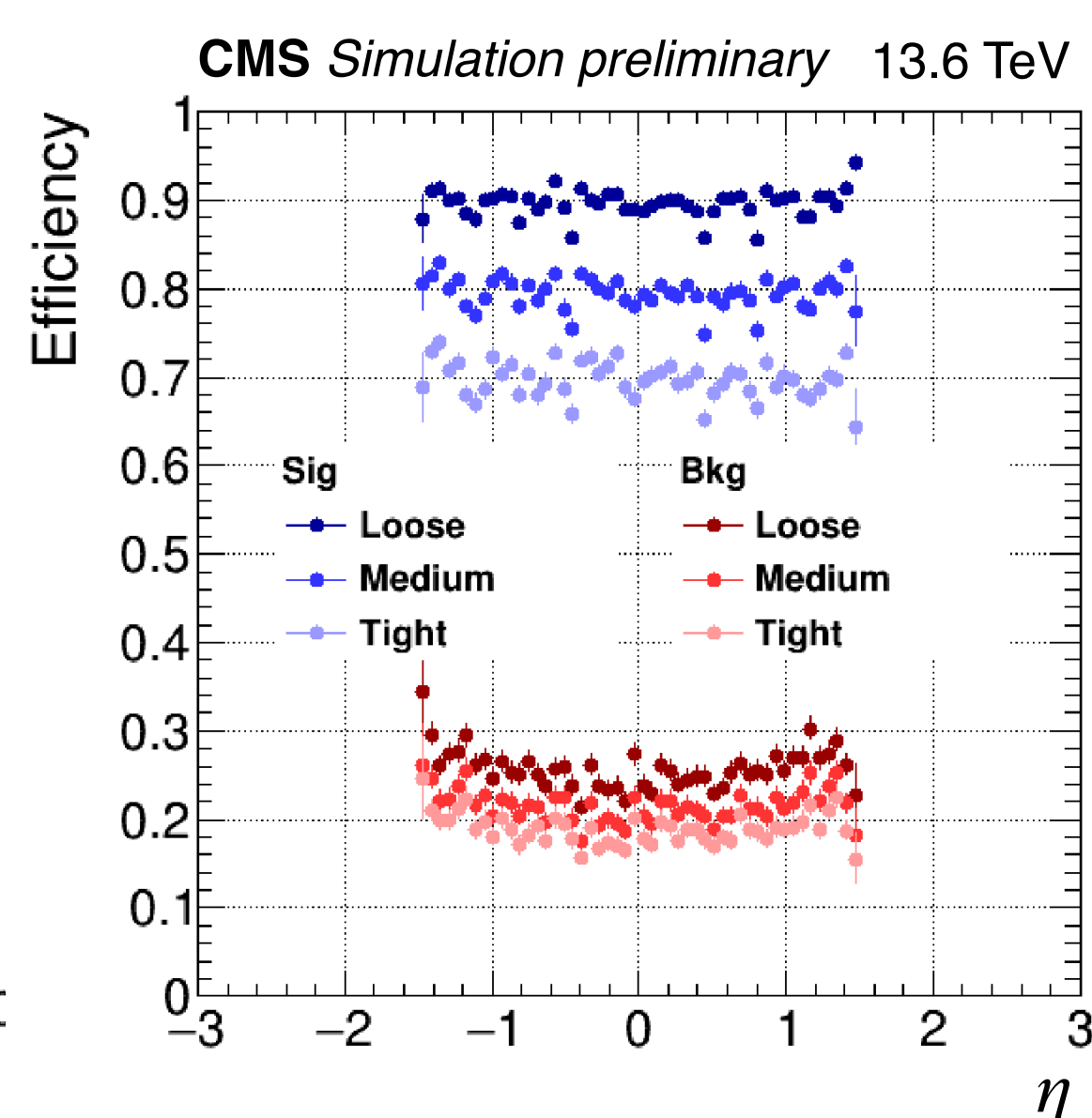
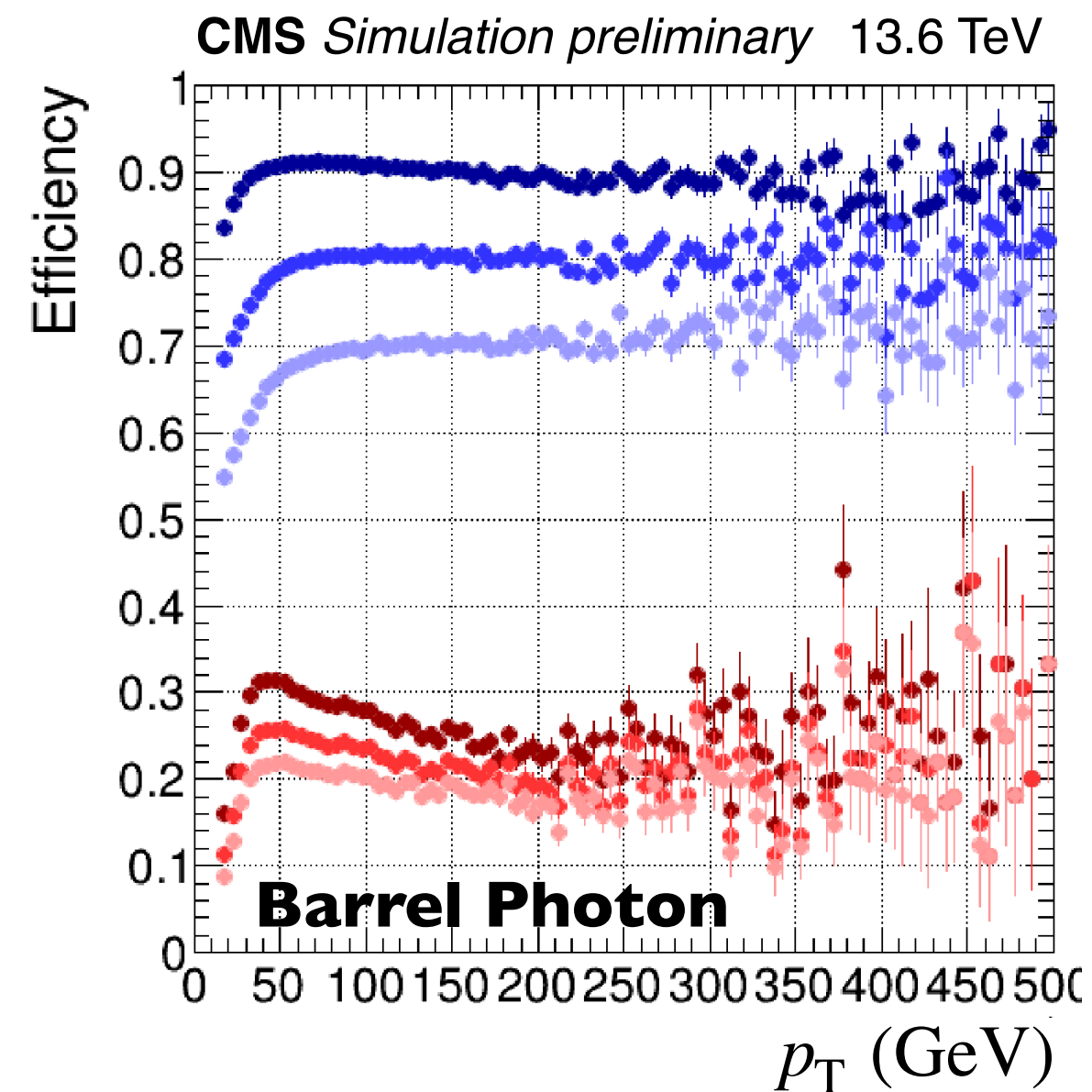
$$|\eta| < 2.5$$

$$\text{MVA ID} > -0.9$$

Photon ID (Cut-based) Performance



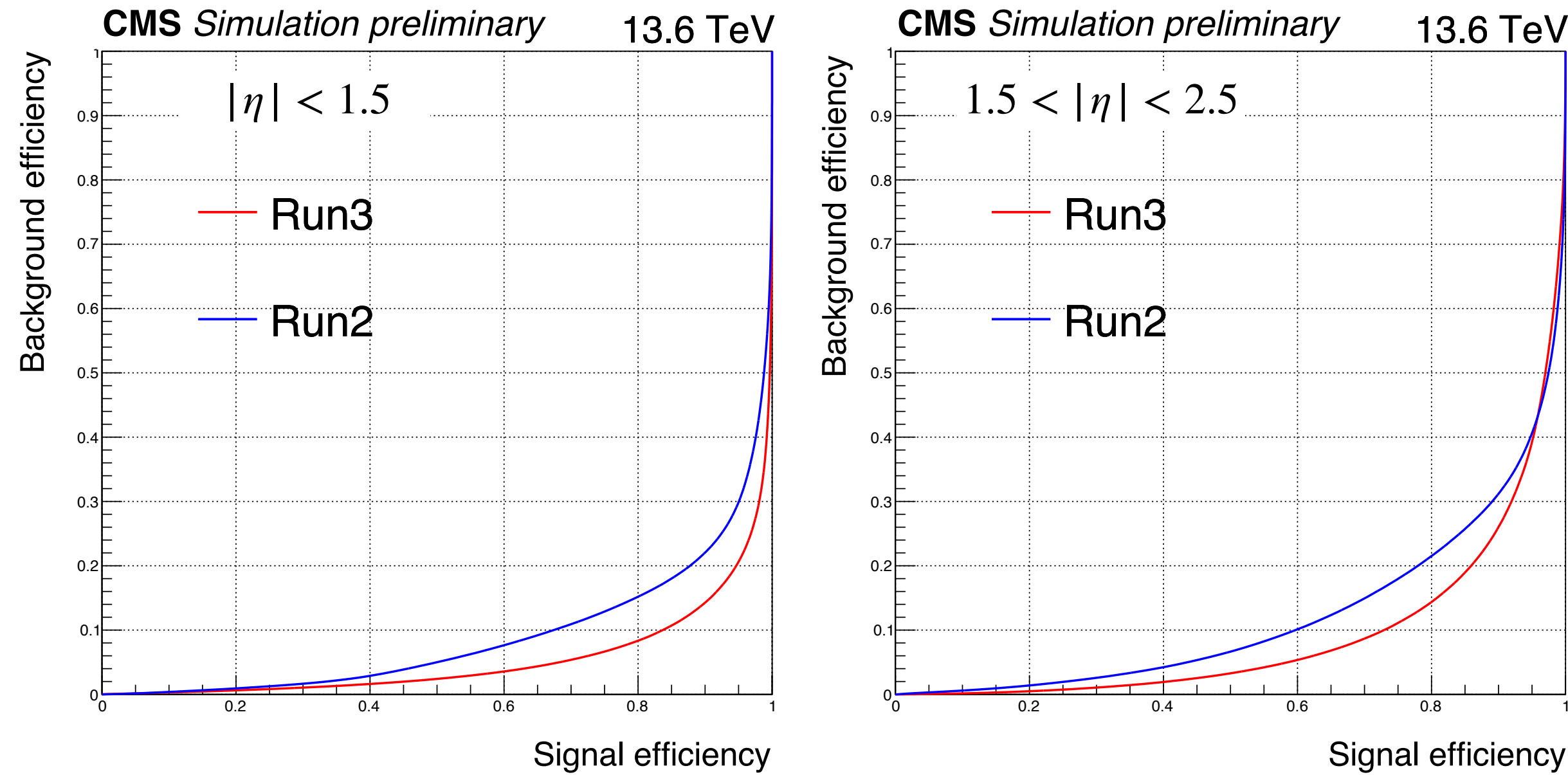
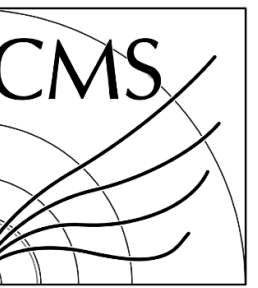
- Genetic algorithm [2][3] was used to optimize the photon cut-based ID.
- Utilized Variables: $\sigma_{i\eta i\eta}$, H/E , Isolation Variables (IsoCH, IsoP, IsoNH) (IsoT, IsoE, IsoH)
- The chosen working points, loose(~90%), medium(~80%), and tight(~70%) were underscored.
- PF particle and PF cluster refers to usage of particle- and energy cluster-type isolation only.



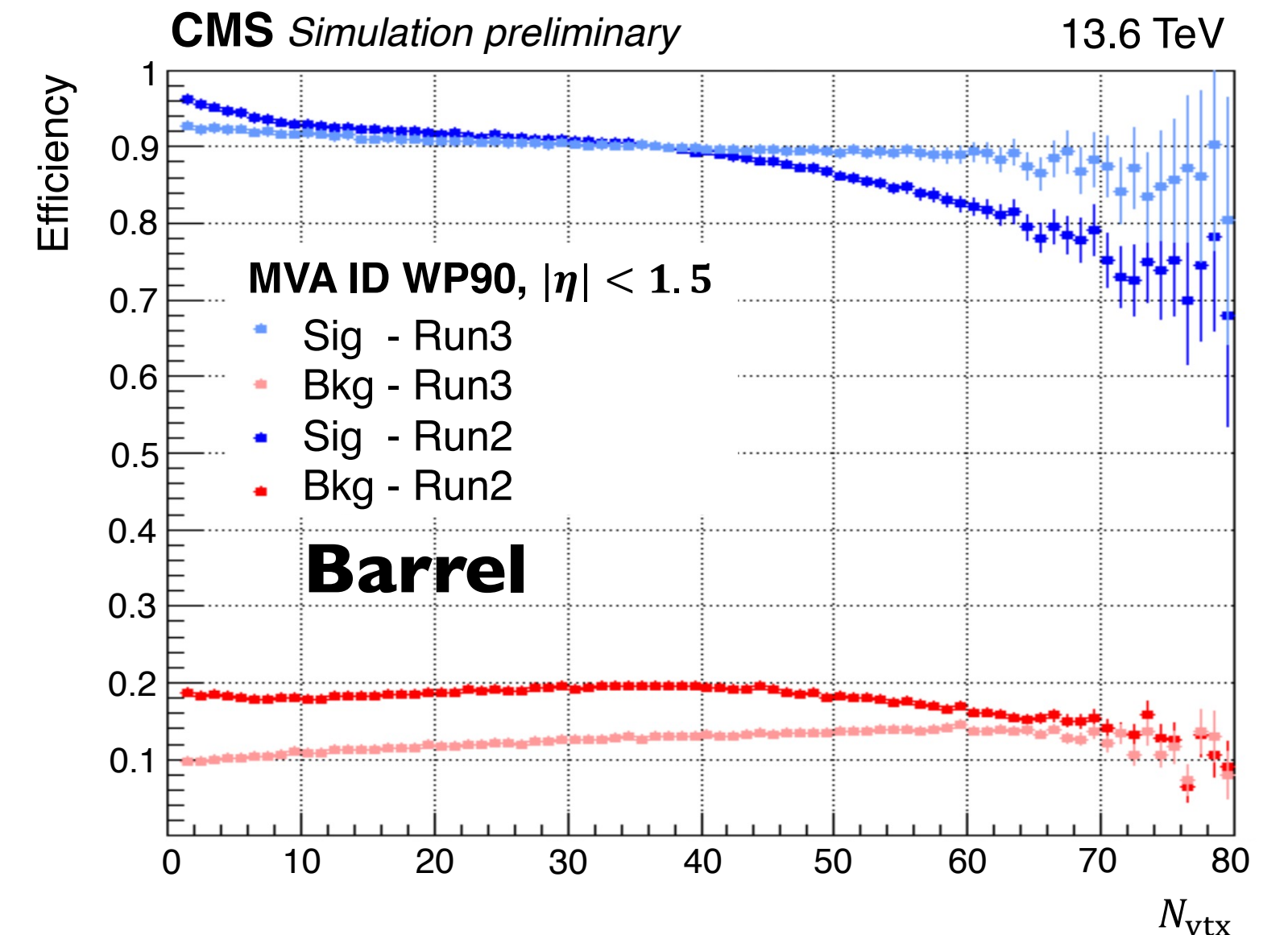
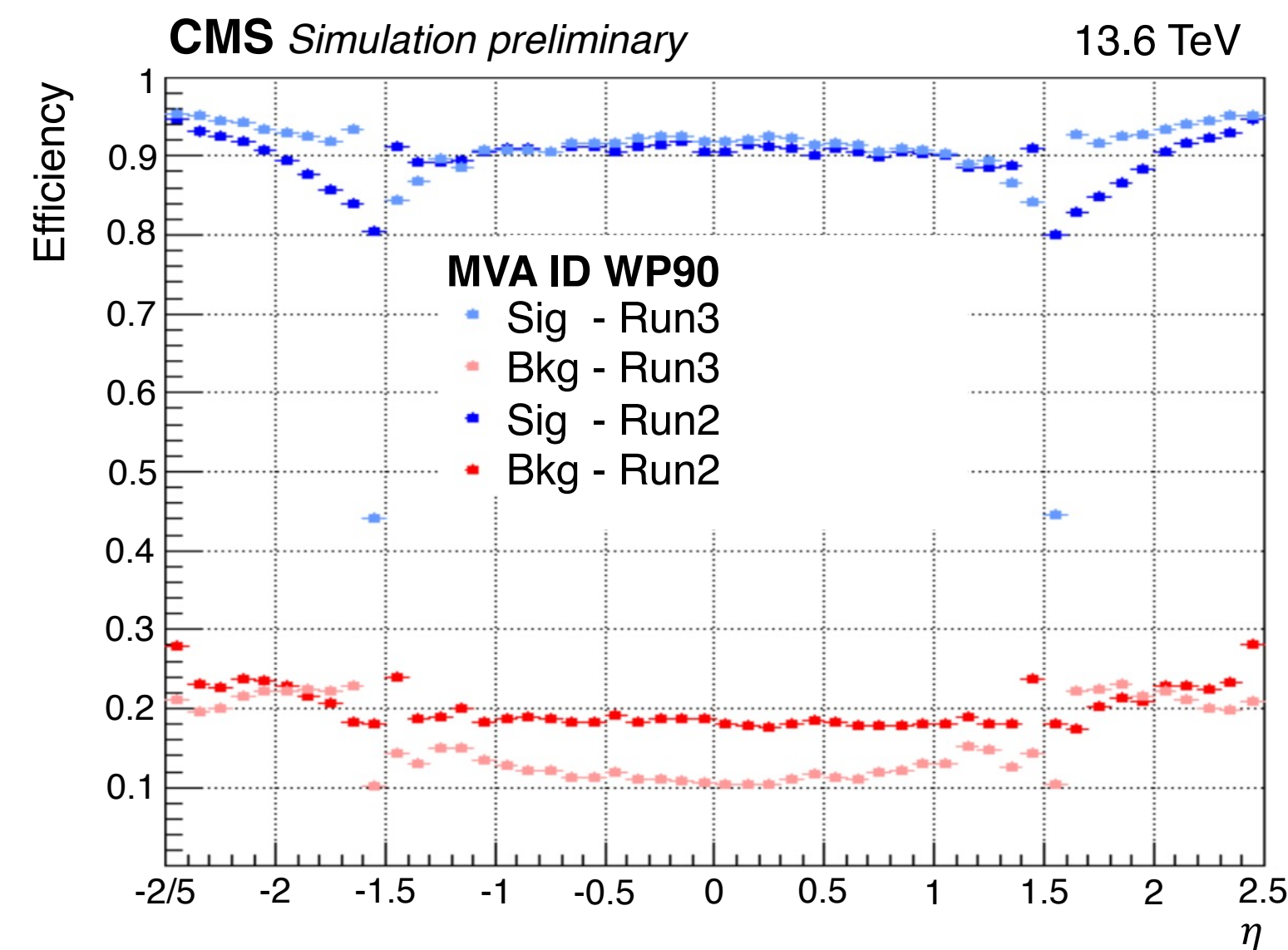
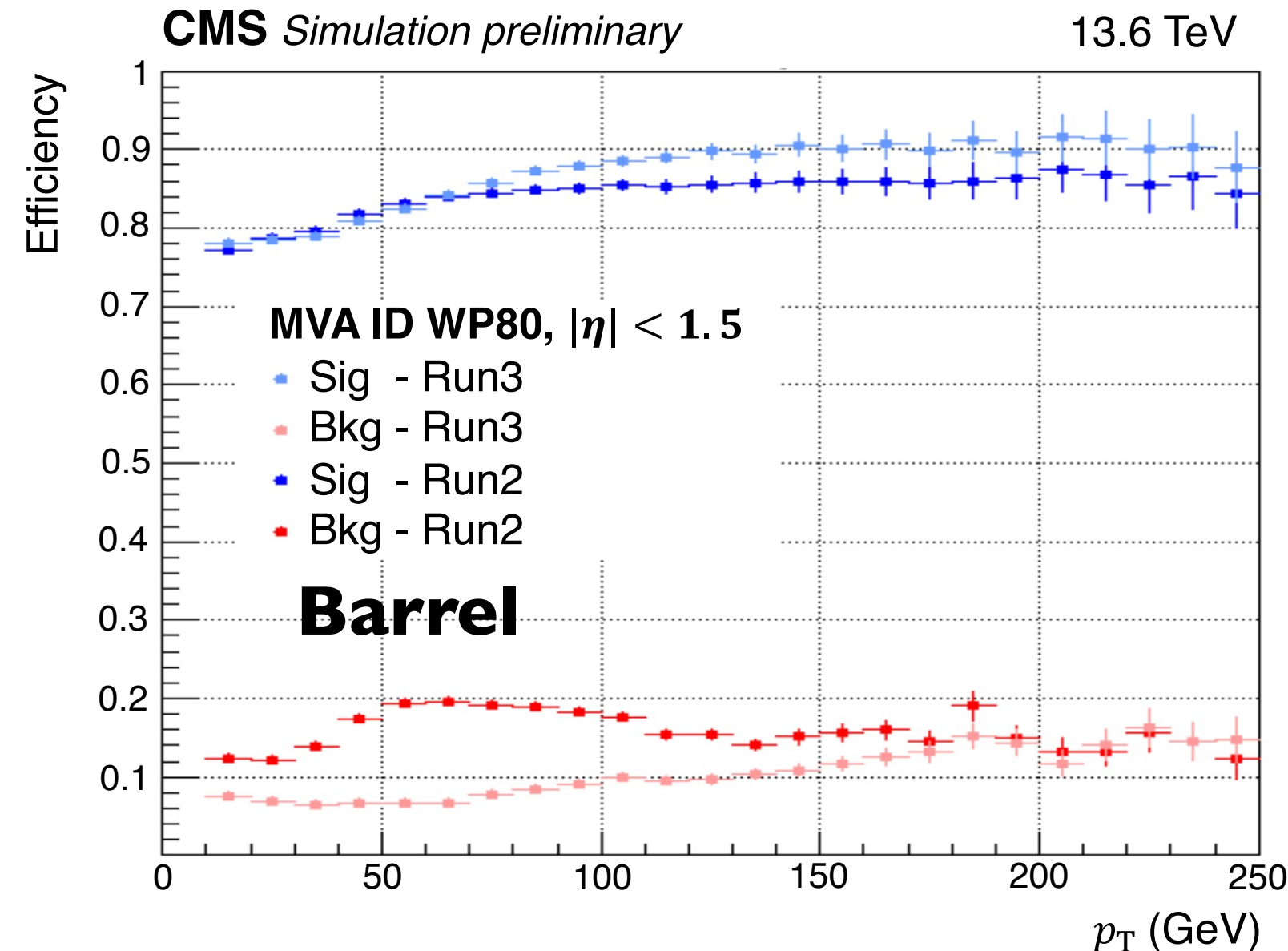
Sig & Bkg Source:
EG-enriched γ + jets

Sig γ : reco matched prompt gen
Bkg γ : jets mis-reconstructed as γ ,
not matched prompt gen γ .

Photon ID (MVA) Performance

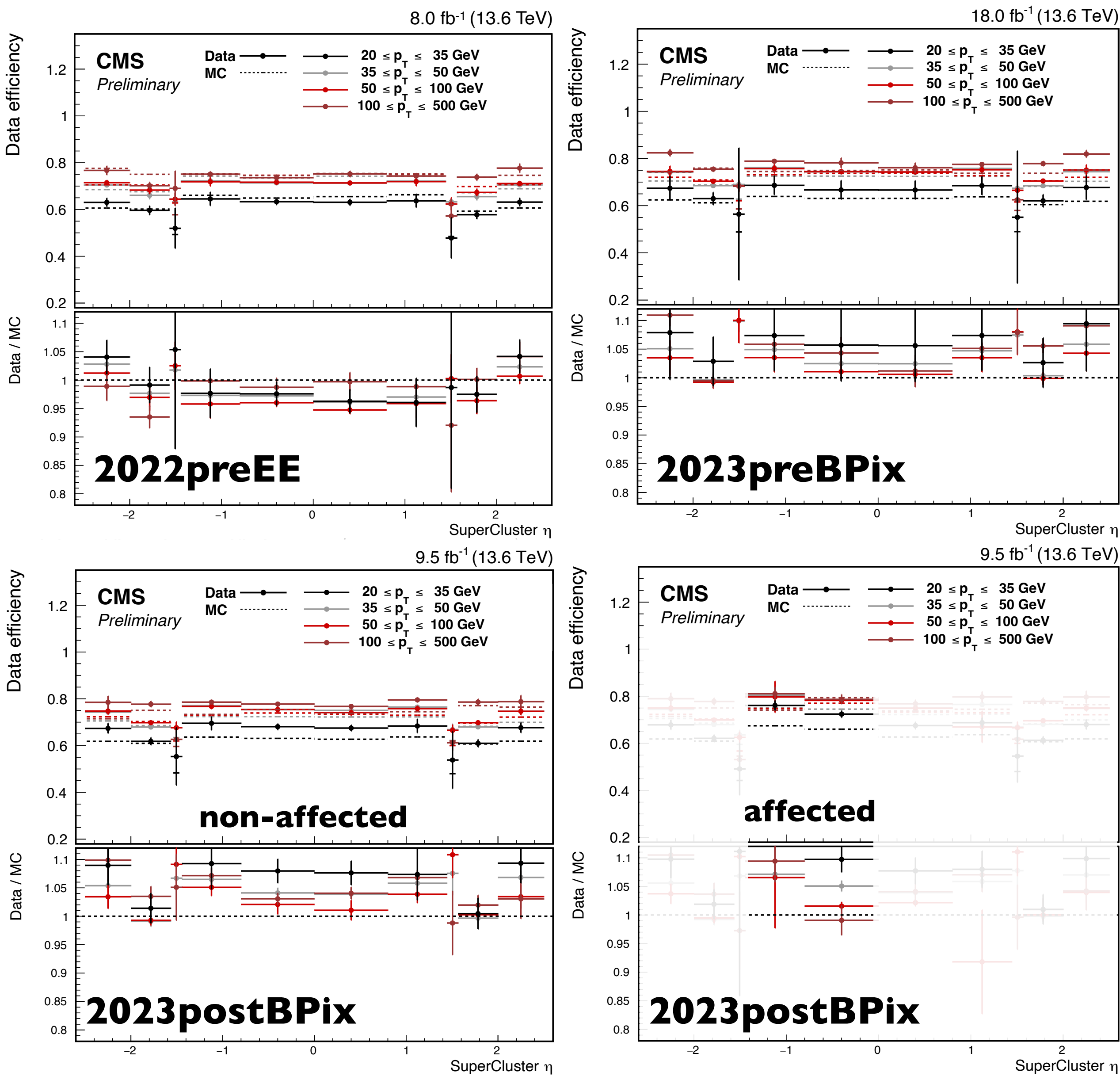


- XGBoost [2][3] was used to optimize the photon MVA ID
- Utilized Variables:
 - **Shower shapes:** $R9$, $\sigma_{i\eta i\eta}$, $\sigma_{i\eta i\phi}$, (ϕ/η) width, $\frac{E_{2\times 2}^{max}}{E_{5\times 5}}$, H/E
 - **Isolation Variables:** IsoC, IsoP, IsoN, IsoT, IsoE, IsoH,
 - **Kinematic Variables:** E_{Raw} SC η , Rho,
 - **Pre-shower:** esEffSigmaRR, esEnergyOverRawE
- Run3 photon MVA IDs perform better than run2 ones, for **more training variables** and **state-of-the-art MVA** algorithm used in run3 [6][7].



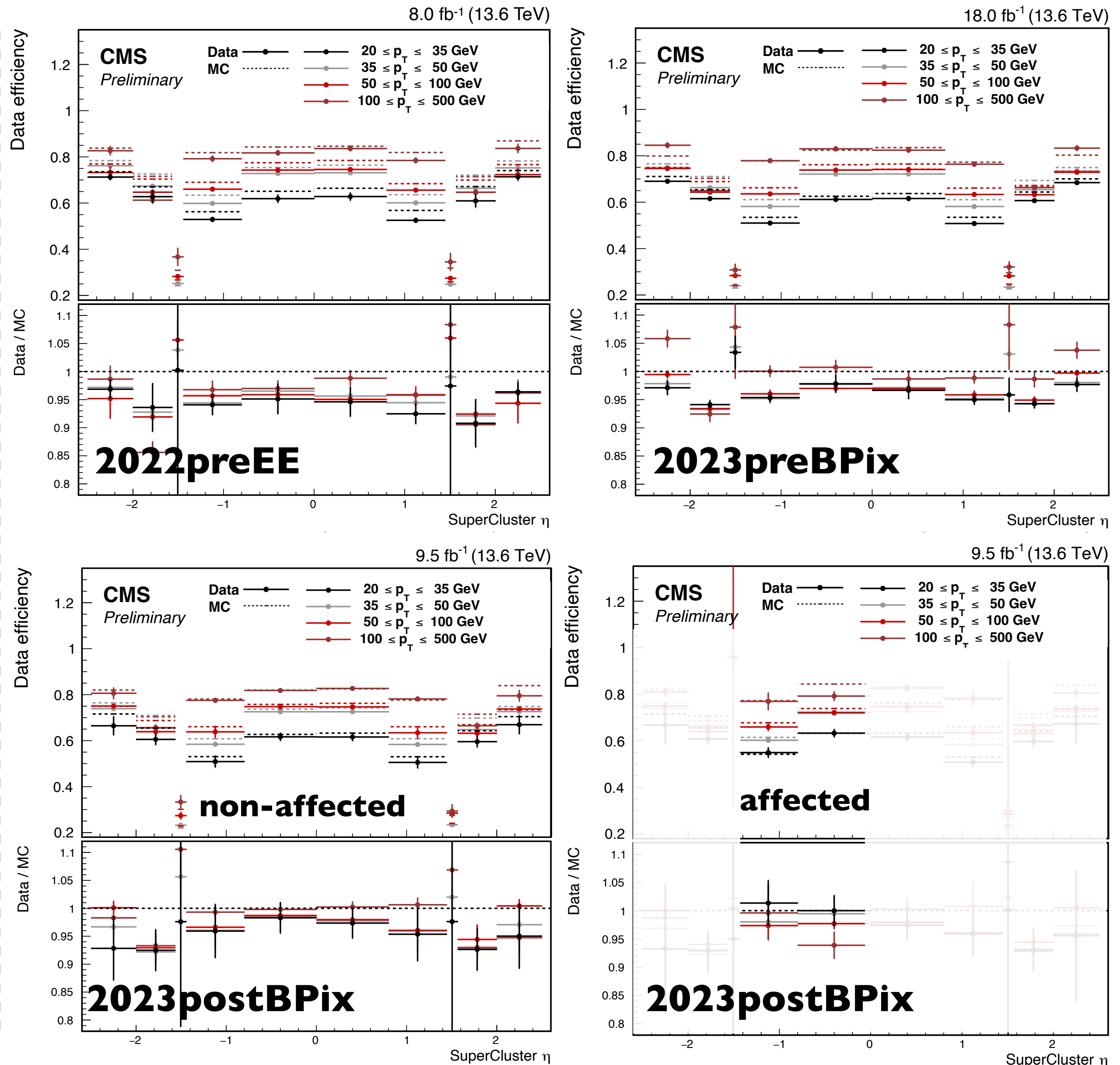
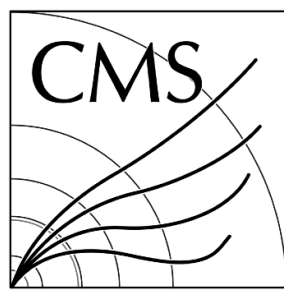
Tight Photon cut-based ID Efficiency

20<pT<500 GeV

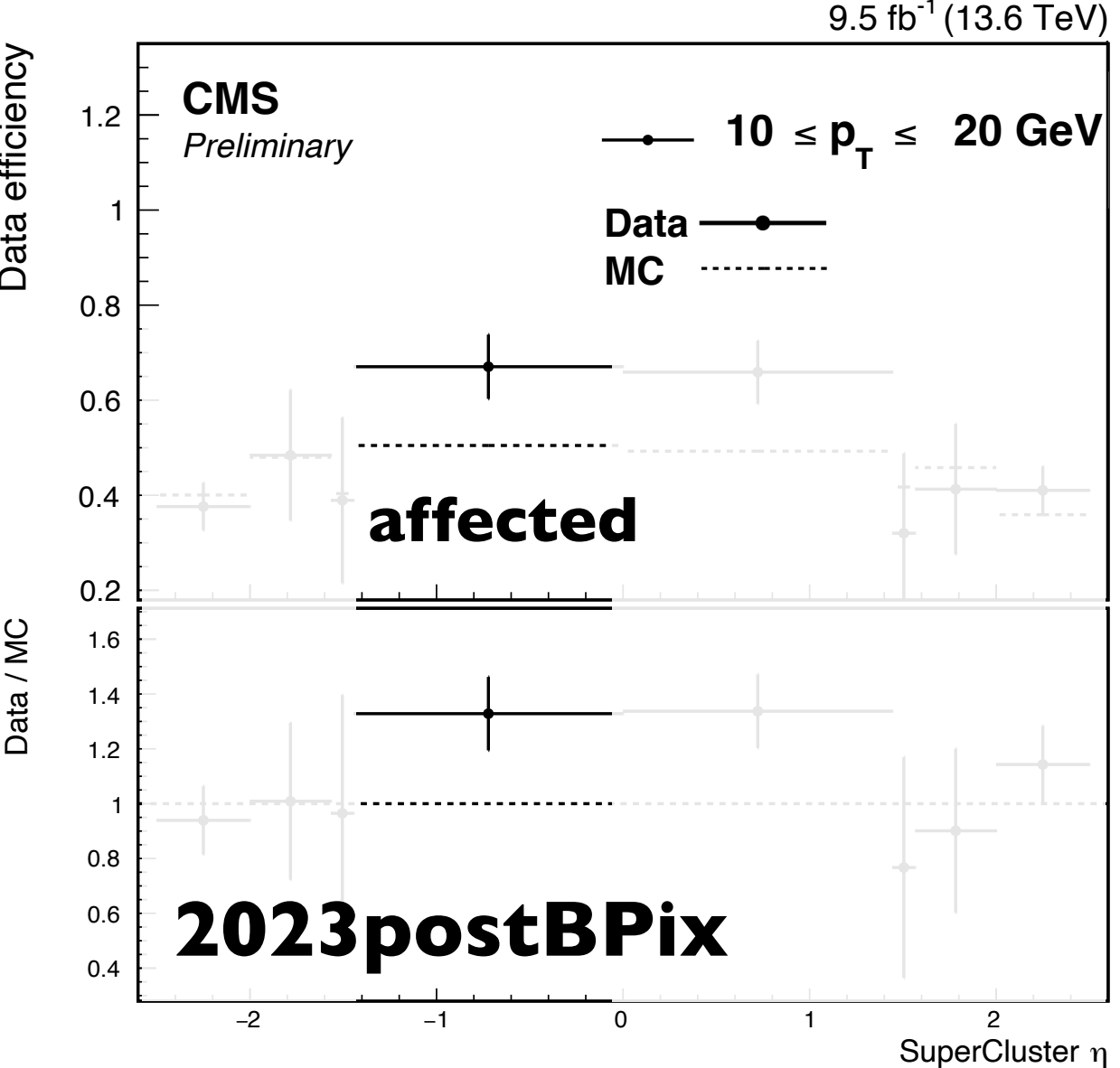
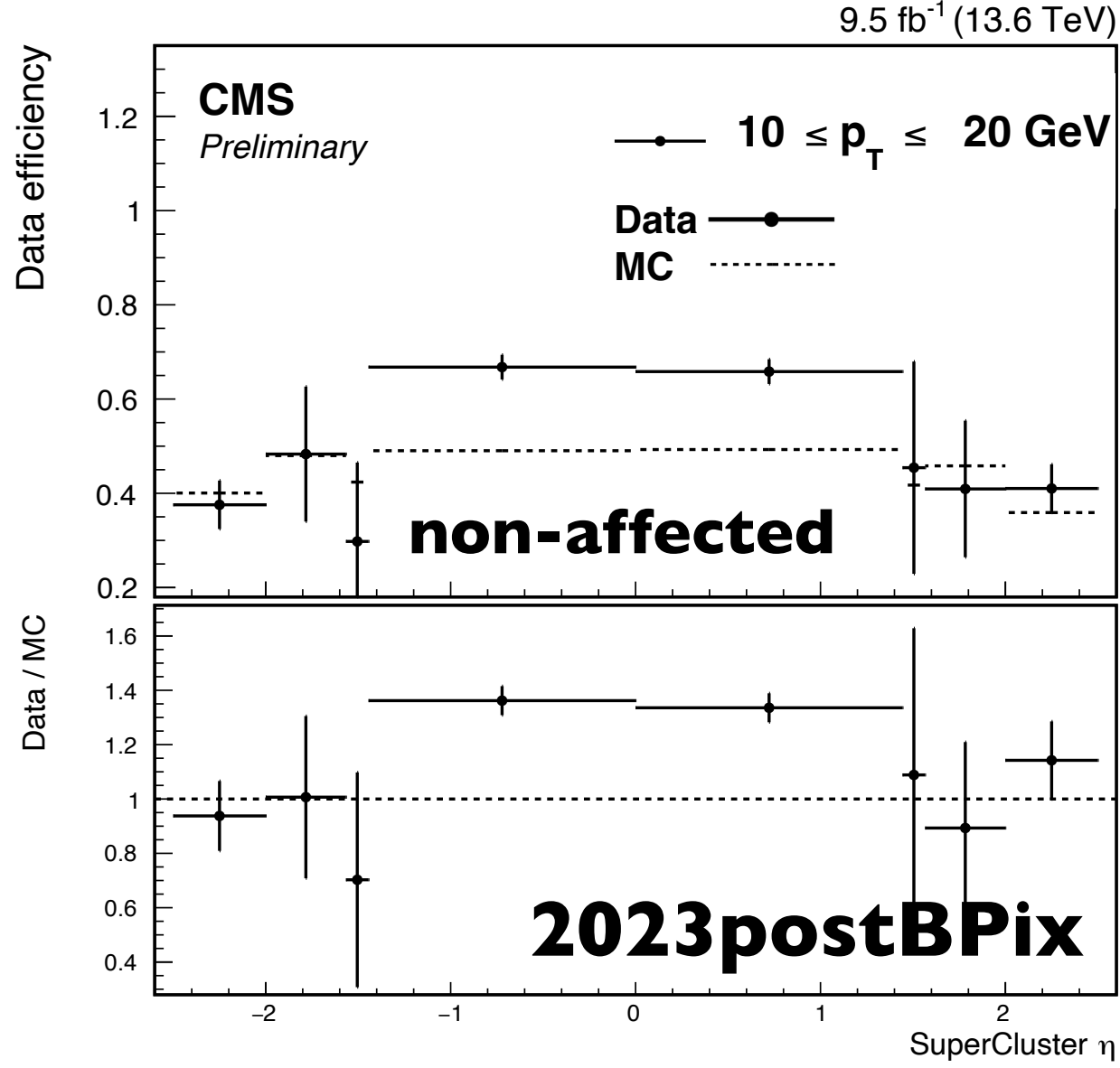
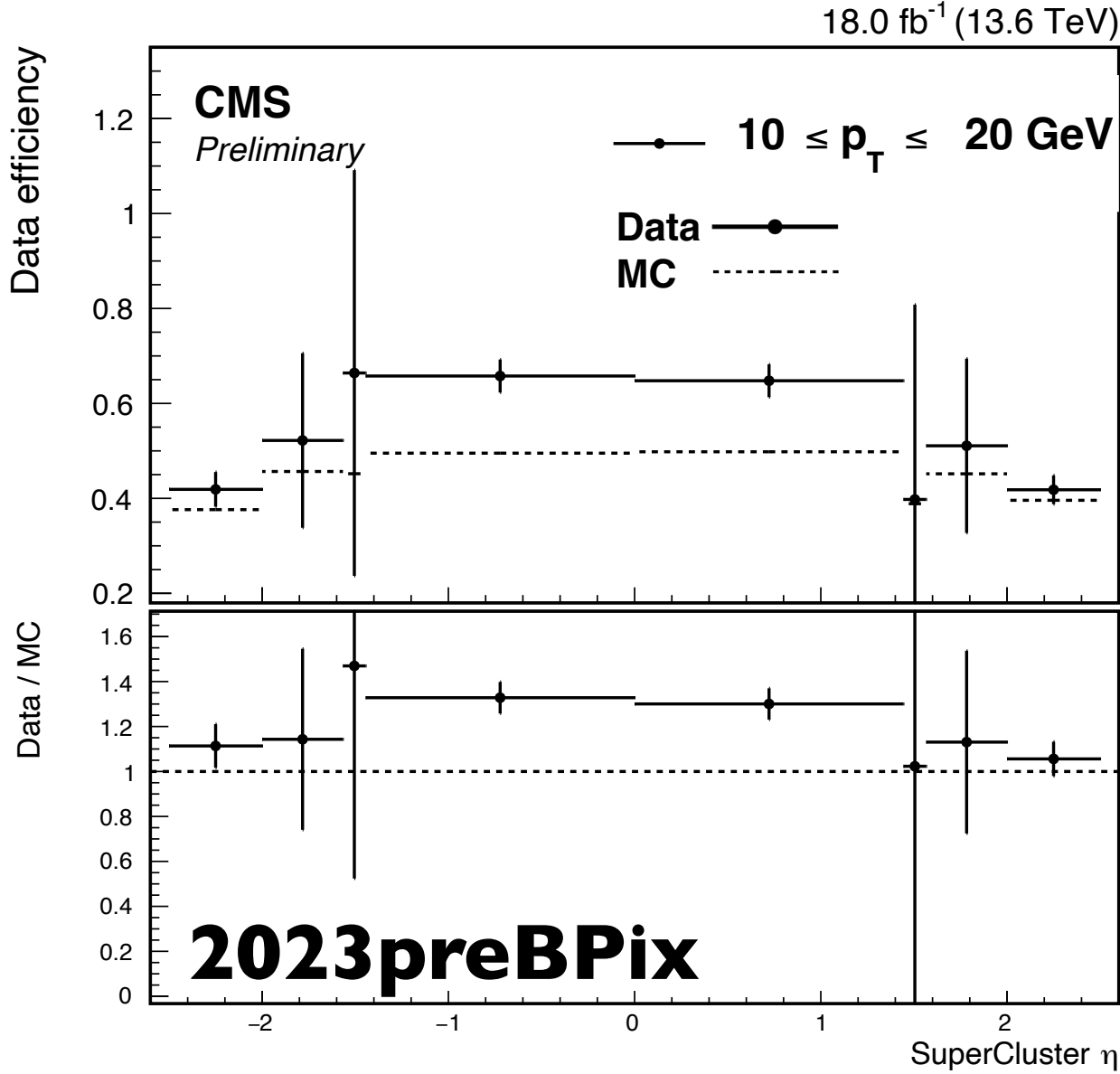


WP80 Photon MVA ID Efficiency

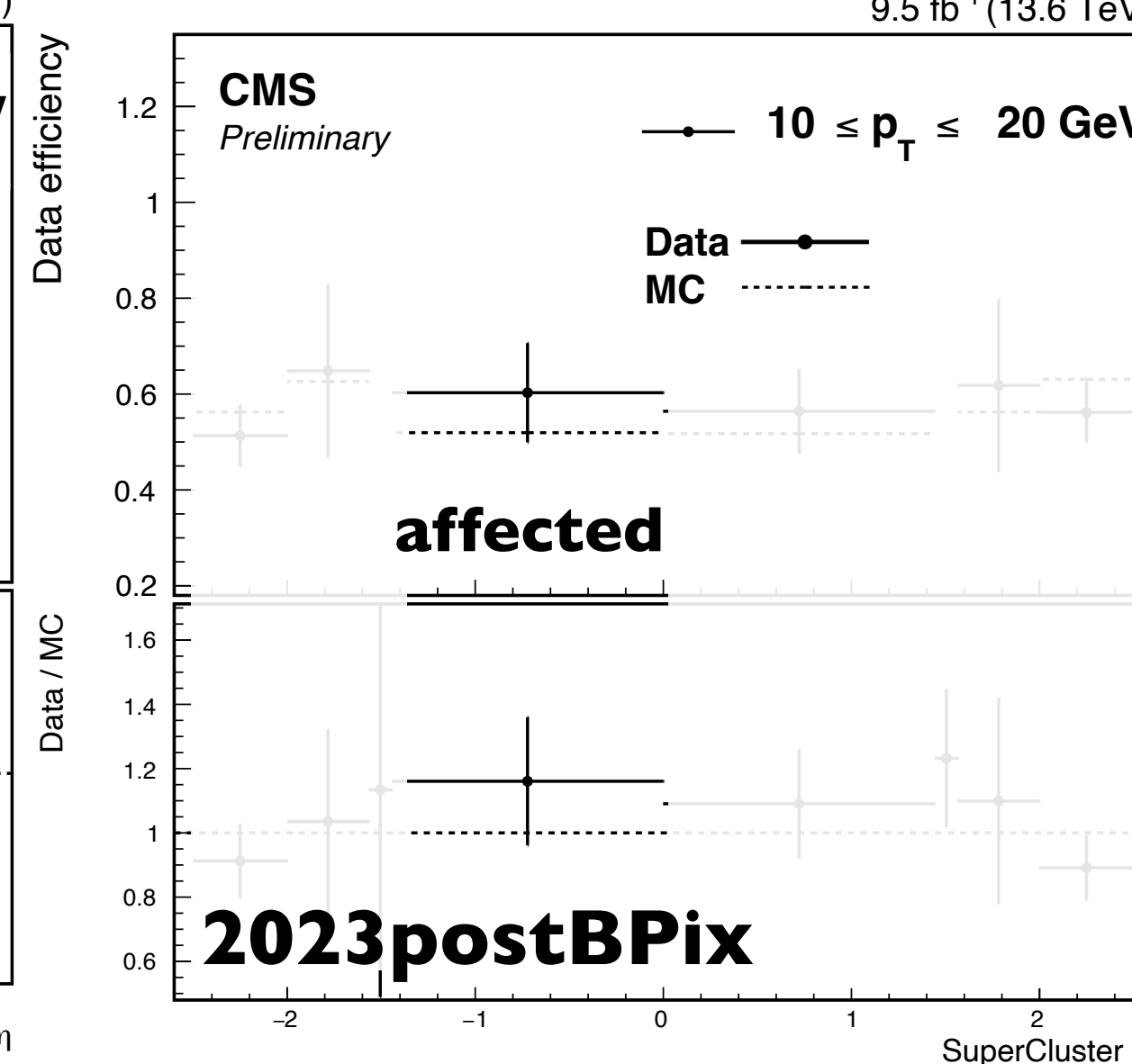
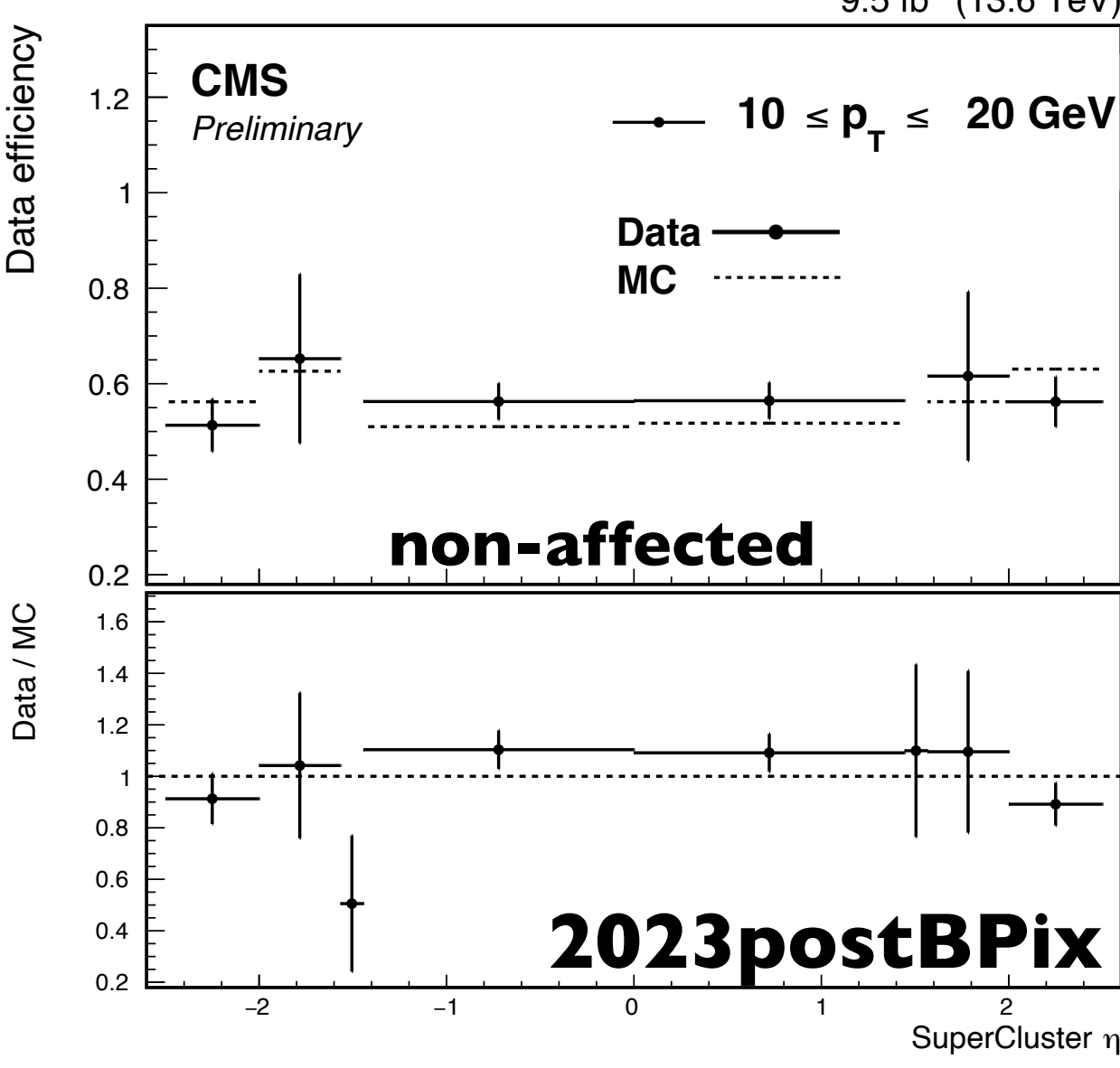
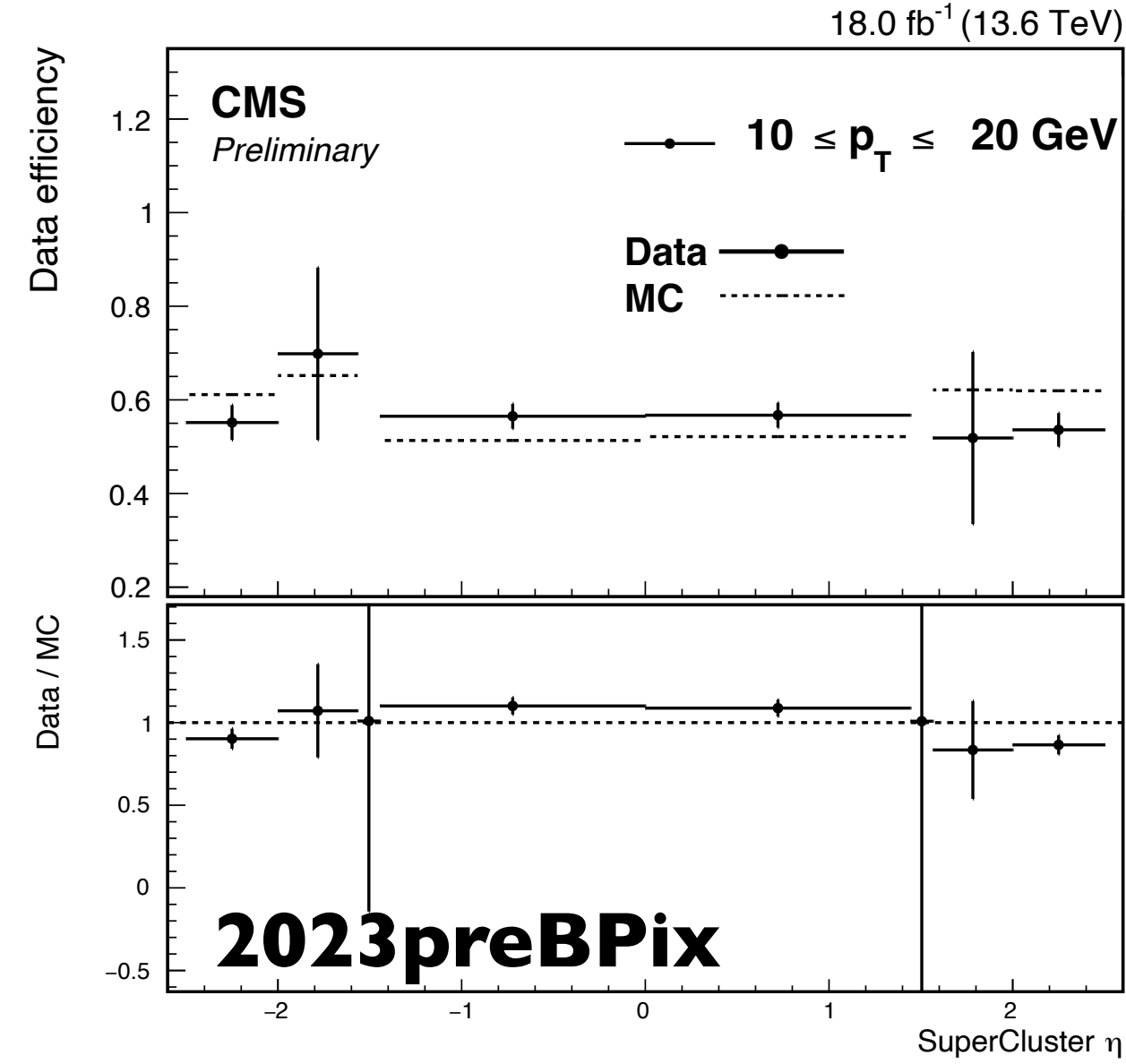
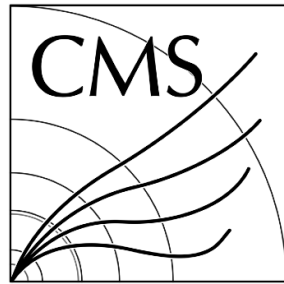
20<pT<500 GeV



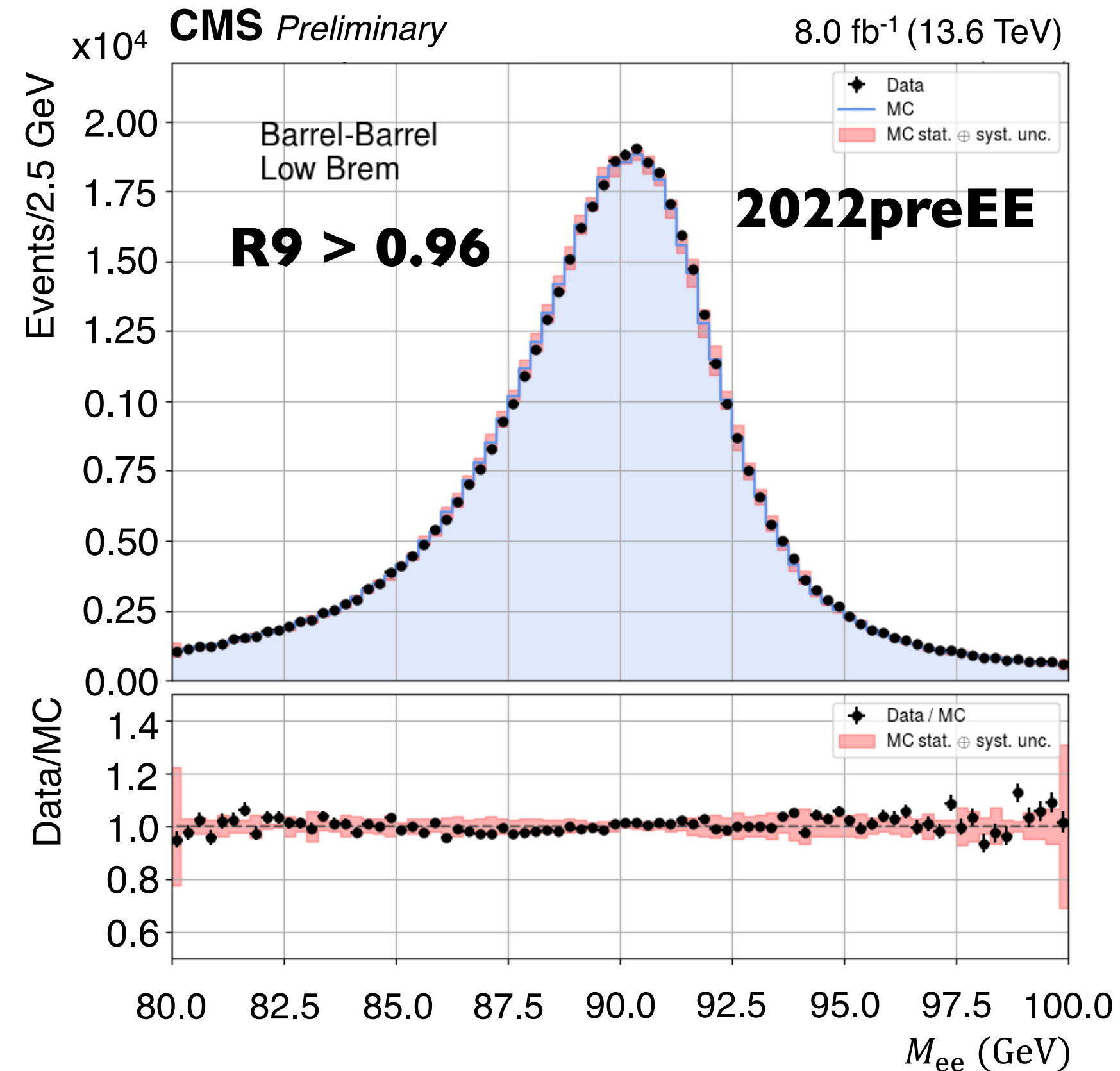
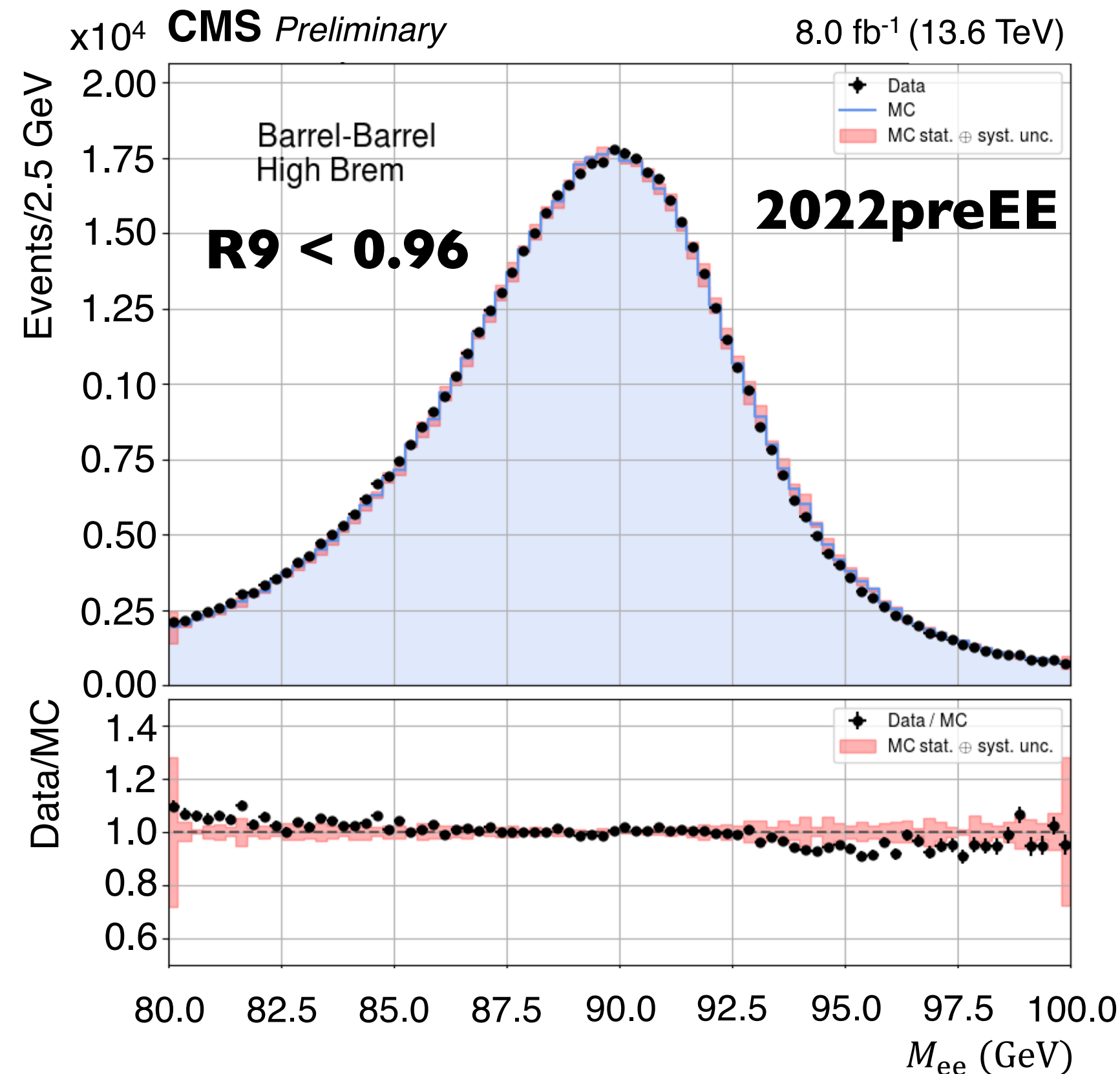
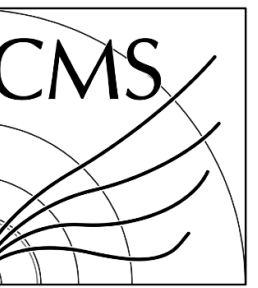
Tight Photon cut-based ID Efficiency
 $10 < p_T < 20 \text{ GeV}$



WP80 Photon MVA ID Efficiency
 $10 < p_T < 20 \text{ GeV}$

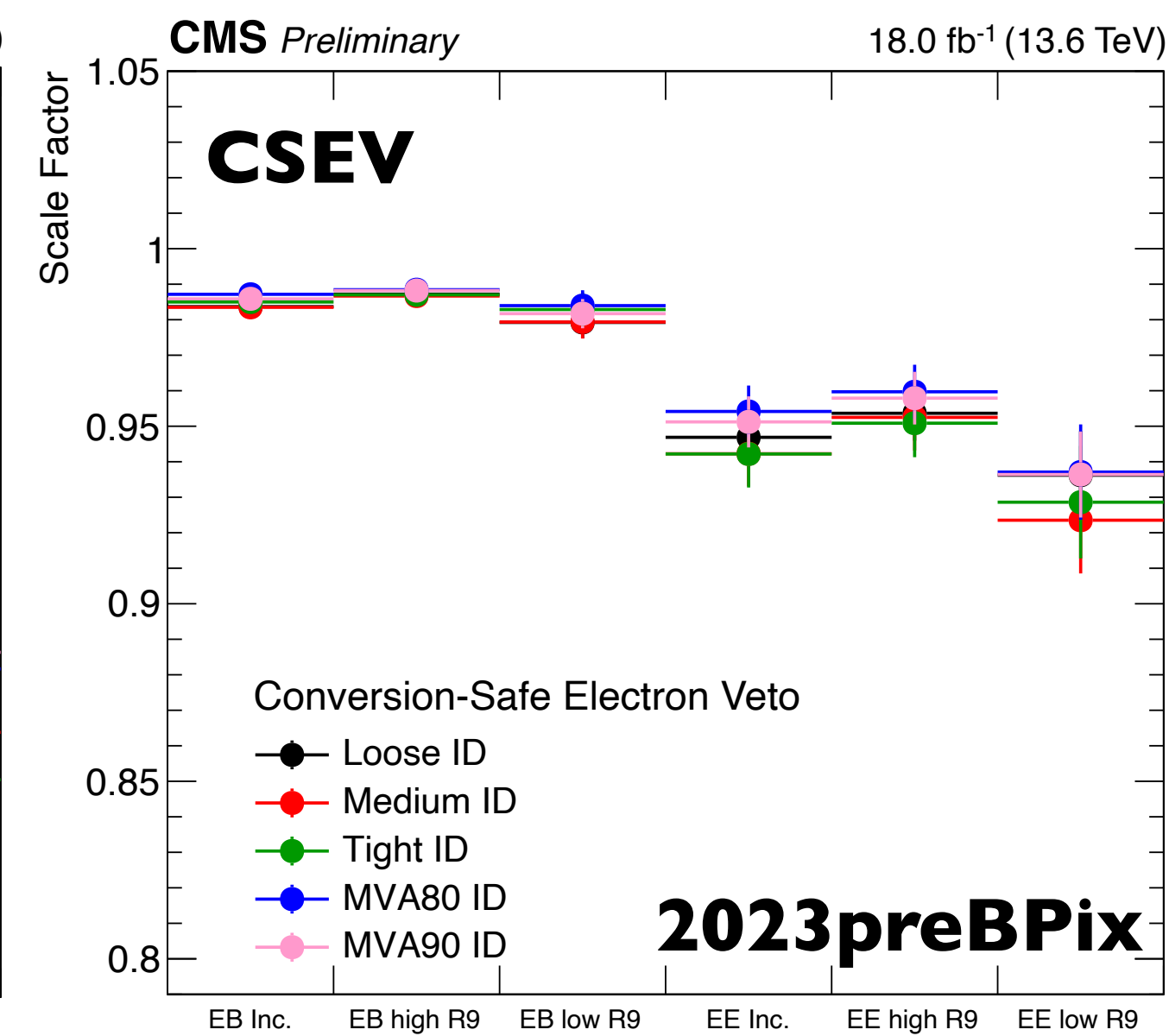
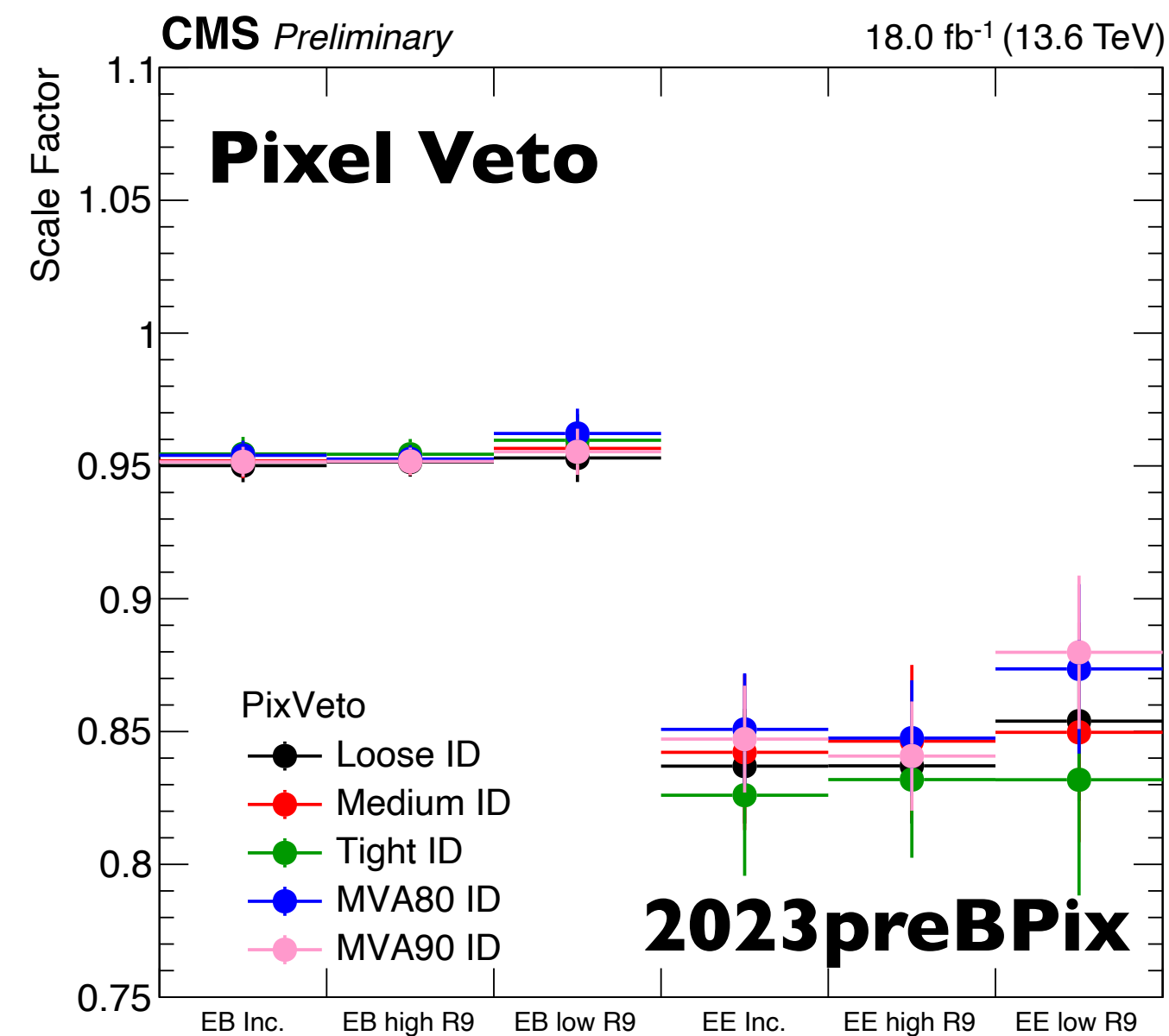
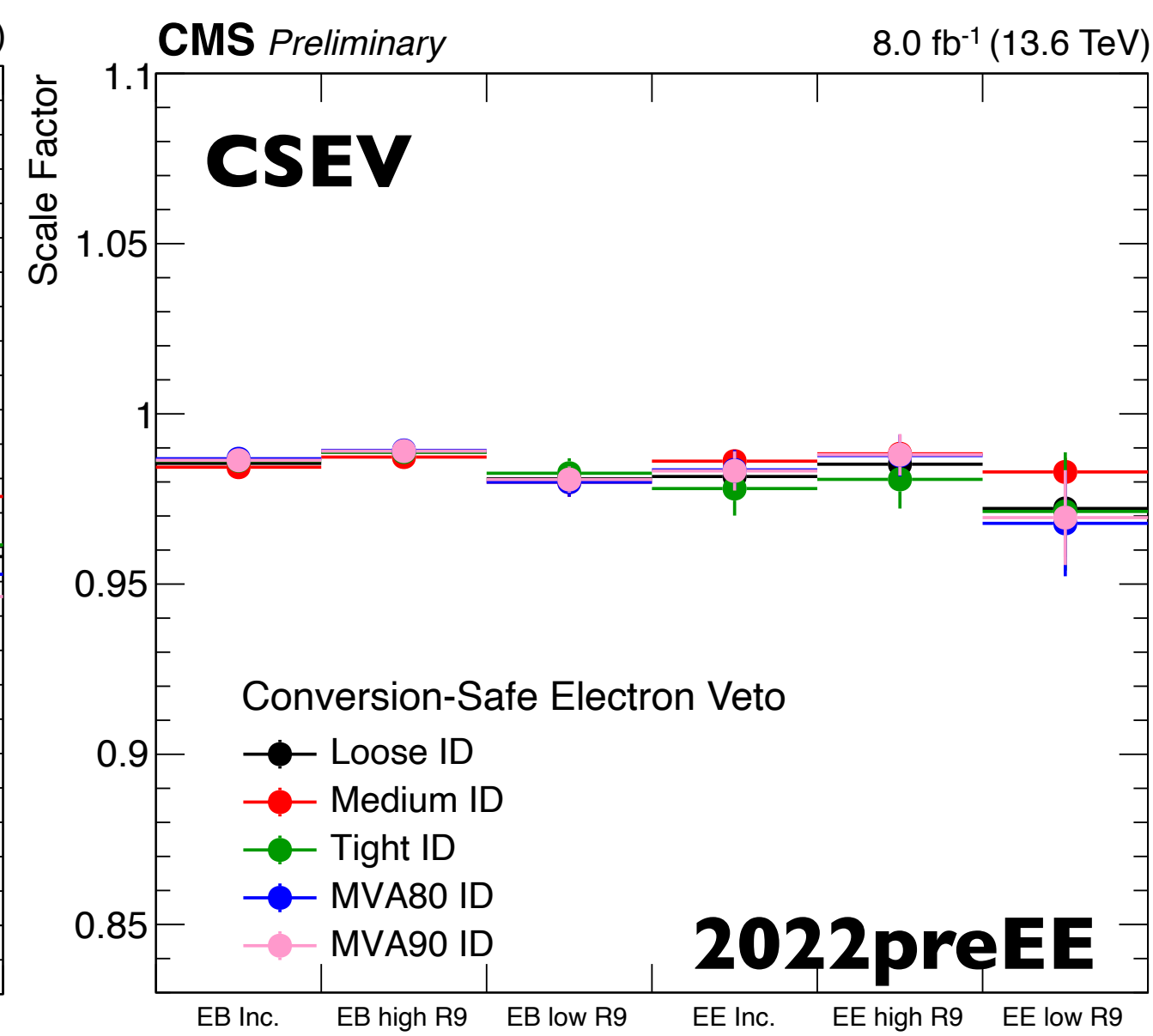
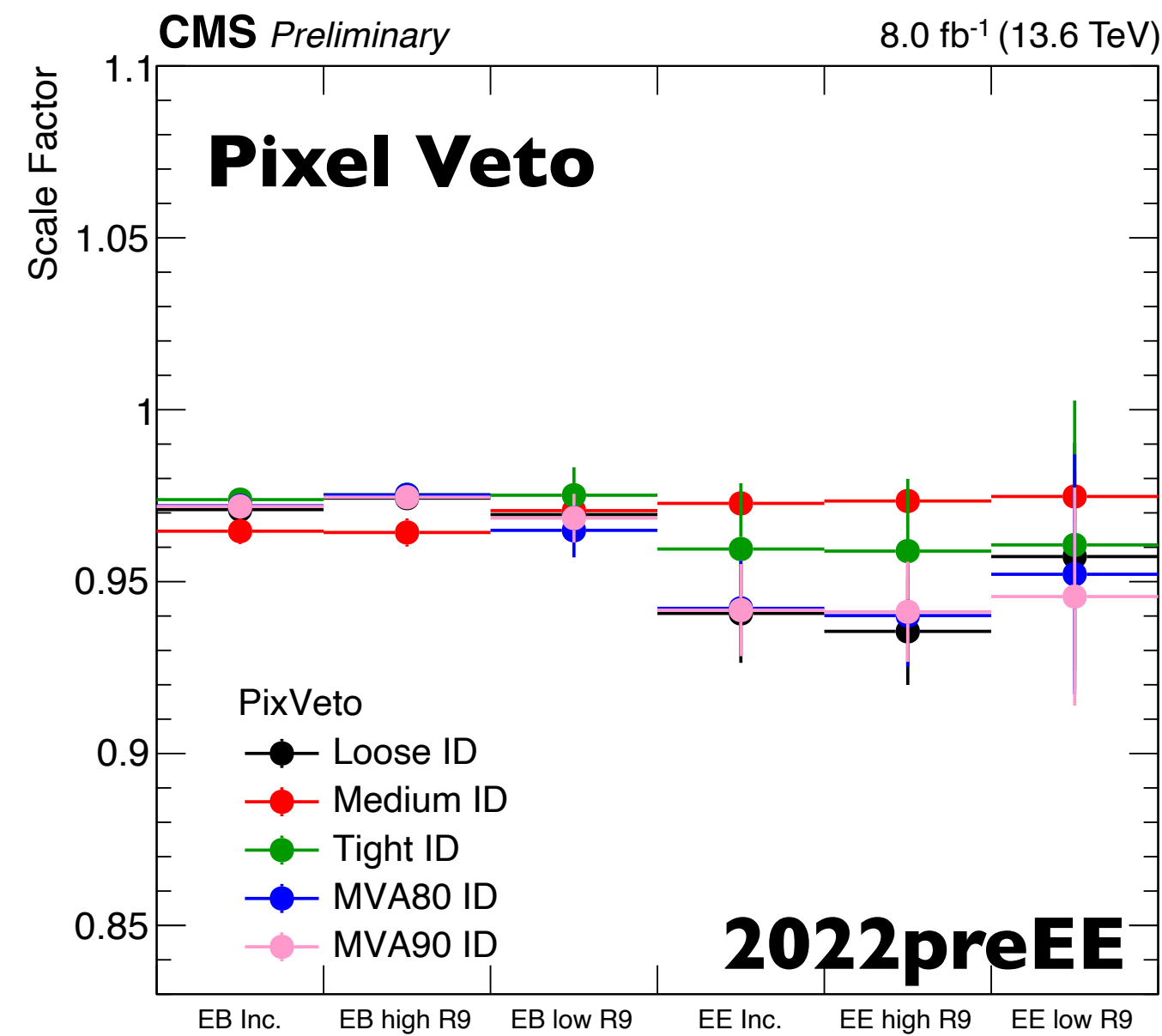
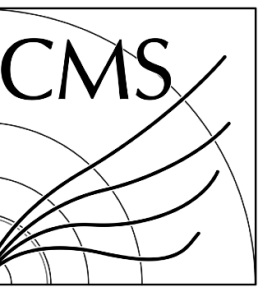


Photon Energy Scale and Smearing Corrections



- Reconstruct electrons as photons by the CMS software.
- Di-photon invariant mass after applying the scale and smearing corrections
- Barrel-Barrel: both the photons are reconstructed in the ECAL barrel.
- High Brem: high-bremsstrahlung ($R9 < 0.96$) photons
- Low Brem: low-bremsstrahlung ($R9 > 0.96$) photons

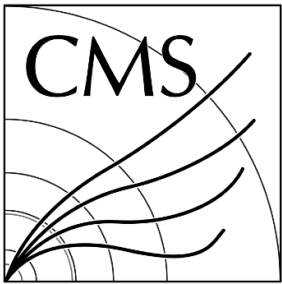
Photon Pixel Veto & CSEV Scale Factors



$Z \rightarrow \mu\mu\gamma(\text{FSR})$

- Photon cut-based ID (Loose, Medium, Tight)
- Photon MVA ID (WP80, WP90)
- EB: $|\eta| < 1.5$
- EE: $1.5 < |\eta| < 2.5$
- Inc.: Inclusive in R₉
- high R₉: R₉ > 0.96
- low R₉: R₉ < 0.96

Summary



- (Electron / Photon) (cut-based / MVA) identification, as well as corresponding selection efficiencies and data-mc scale factors, have been presented for the 2022 and 2023 datasets.
- Electron / Photon energy scale-and-smearing corrections have been centrally provided.

Run3 22 + 23

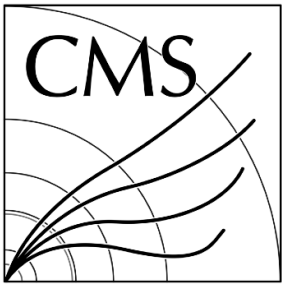
Electron

Photon

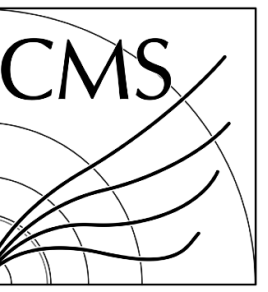
ID (cut-based)	Signal: Loose(~90%), Medium(~80%), Tight(~70%) Bkg Reject: Loose(~96%), Medium(~98%), Tight(~99%)	Signal: Loose(~90%), Medium(~80%), Tight(~70%) Bkg Reject: Loose(~74%), Medium(~78%), Tight(~81%)
ID (MVA)	Run3 ROC performs better than run2	
ID (cut-based) Efficiency	10<pT<500 GeV	(2022) 20<pT<500 GeV (2023) 10<pT<500 GeV
ID (MVA) Efficiency	10<pT<500 GeV	(2022) 20<pT<500 GeV (2023) 10<pT<500 GeV
Scale and Smearing Correction	V	V
Pixel Veto Scale Factors	X	V
Conversion-save Electron Veto Scale Factors	X	V

Thank for your attention

Back up



Reference



- [1] CMS Collaboration, “Electron and photon reconstruction and identification performance at CMS in 2022 and 2023”, CMS-DP-2024-052 ; CERN-CMS-DP-2024-052
- [2] CMS Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”, JINST 16 (2021) P05014, doi:10.1088/1748-0221/16/05/P05014
- [3] CMS Collaboration, “Cut-based photon ID tuning of CMS using Genetic Algorithm”, Springer Proc. Phys. 261 (2021) 755-760, doi:10.1007/978-981-33-4408-2_104
- [4] TMVA Collaboration, “TMVA - Toolkit for Multivariate Data Analysis”, CERN-OPEN-2007-007, e-Print: physics/0703039
- [5] <https://indico.cern.ch/event/1182858/contributions/4969321/attachments/2481556/4260121/2022-07-18.pdf>
- [6] https://indico.cern.ch/event/1182858/contributions/4969317/attachments/2481560/4260087/EGamma_photonID_Run3_Prasant_18072022.pdf
- [7] https://indico.cern.ch/event/1188686/contributions/4995761/attachments/2490690/4277953/EGamma_photonID_Run3_Update_Prasant_08082022.pdf

The 2022 data-taking period

About 7% of the channels of the positive side of the endcap Electromagnetic Calorimeter (ECAL EE+) were affected by a power cooling issue for part of the 2022 data-taking period and not active during this time. The integrated luminosities corresponding to the pre- and post-issue periods are:

- Pre ECAL EE+ issue: 8 fb^{-1}
- Post ECAL EE+ issue: 27 fb^{-1}

As a consequence of this detector problem, a deficit was observed in the number of electrons and photons reconstructed in the area corresponding to the issue.

Therefore, a dedicated simulation was produced to account for such an inefficiency. In the following slides, efficiencies and collision data-to-simulation corrections are shown for the two periods, henceforth addressed as **preEE+** and **postEE+**, separately.

Furthermore, part of the 2022 data underwent a reprocessing that used better detector calibrations.

The 2023 data-taking period

The Quartz-controlled PLL circuit connected to 28 modules of the Barrel Pixel (BPIX) detector in layers 3 and 4 had could not lock to the LHC clock and had to be excluded from a portion of the data taking. The integrated luminosities corresponding to the period before and after this issue in BPIX are:

- Pre-BPIX issue: 18 fb^{-1}
- Post-BPIX issue: 9.5 fb^{-1}

As a consequence of this detector problem, a deficit was observed in the number of electrons reconstructed in the area corresponding to the issue.

Therefore, a dedicated simulation was produced to account for such an inefficiency. In the following slides, efficiencies and collision data-to-simulation corrections are shown for the two periods, henceforth addressed as **preBPIX** and **postBPIX**, separately. The region of the detector corresponding to the faulty BPIX modules is defined using the following coordinates:

$$-1.5 < \eta_{SC} < 0, \quad -1.2 < \phi < -0.8$$

Legend of acronyms and conventions

Data early 2022: data recorded in the first part of 2022, corresponding to an integrated luminosity of 8 fb⁻¹.

Data late 2022: data recorded in the second part of 2022, corresponding to an integrated luminosity of 27 fb⁻¹.

Data early 2023: data recorded in the first part of 2023, corresponding to an integrated luminosity of 18 fb⁻¹.

Data late 2023: data recorded in the second part of 2023, corresponding to an integrated luminosity of 9.5 fb⁻¹.

Prompt data: data reconstructed offline shortly after they have been saved by the online (trigger) reconstruction.

Reprocessed data: data reconstructed offline at a later time than the prompt reconstruction. This typically allows for more refined detector calibrations to be used.

η_{SC} : “supercluster η ”, namely the pseudorapidity of the reconstructed particle calculated with respect to the center of the CMS detector.

R_9 : ratio between the energy deposited in a 3x3 matrix of ECAL crystal around the seed (the one with the largest energy deposit) and the supercluster energy. Low values of R_9 are typically associated to particles with large emissions of bremsstrahlung photons and/or electron-positron pair production.

Pixel veto: a detector-based veto used to distinguish photons from electrons based on the possibility of geometrically associating a supercluster to a track in the inner tracker.

CSEV: conversion-safe electron veto. A criterion that checks the ΔR distance between the photon direction and the nearest electron track and vetos the photon if there is an electron matching the photon supercluster, with no missing hits and having no matching reconstructed conversion.

Scale factor: a correction that is typically applied to simulated events to correct discrepancies with the collision data.

WP: working point. This generally indicates the choice of a cutoff (or a series of cutoffs) applied on one or more variables with the goal of obtaining a certain signal efficiency and/or background rejection.

Electron reconstruction efficiency - event selection

Collision data: recorded using an unprescaled single electron trigger with a 32 (30) GeV transverse momentum threshold for 2022 (2023) data.

Simulation: Drell-Yan simulated at LO and NLO in QCD using the MadGraph and amc@nlo generators, respectively.

Tag electron selection for 2022 (2023):

- $p_T > 37$ (35) GeV, $|\eta_{SC}| < 2.5$, $m_T < 60$ GeV, opposite charge with respect to the probe.
- High purity cut-based ID.
- If $p_T^{\text{probe}} < 20$ GeV, additional cutoffs are applied to suppress the background from fake electrons:
 - MVA ID score > 0.90 , supercluster relative track isolation < 0.1 , $m_T < 45$ GeV

Sources of systematic uncertainty considered:

- Alternative signal fit
- Alternative background fit
- Alternative tag selection
- Alternative MC generator

Binning:

p_T : [10, 20, 45, 75, 100, 500]

η_{SC} : [-2.5, -2.0, -1.566, -1.444, -0.8, 0, 0.8, 1.444, 1.566, 2.0, 2.5]

Electron ID tuning - event selection

The IDs to be applied to offline reconstructed electrons have been retrained and retuned in Run3 to offer a better performance with respect to the Run2 IDs in the Run3 data-taking conditions.

Goal: discrimination between signal and background electrons.

Source of signal electrons: Drell-Yan simulated at NLO using the amc@nlo generator.

Source of background electrons: $t\bar{t}$ semileptonic events simulated using the Powheg generator.

Signal electrons: reconstructed electrons kinematically matched to prompt generator-level electrons.

Background electrons: reconstructed electrons not kinematically matched to prompt generator-level electrons.

Corrections: a 2D $p_T - \eta$ reweighting to signal and background events is performed in order to avoid biases coming from the kinematics of signal and background processes.

Electron ID efficiency - event selection

Collision data: recorded using an unprescaled single electron trigger with a 32 (30) GeV transverse momentum threshold for 2022 (2023) data.

Simulation: Drell-Yan simulated at LO and NLO in QCD using the MadGraph and amc@nlo generators, respectively.

Tag electron selection for 2022 (2023):

- $p_T > 35$ (32) GeV, $|\eta_{SC}| < 2.17$, opposite charge with respect to the probe.
- High purity cut-based ID
 - If $p_T^{\text{probe}} < 20$ GeV, additional cutoffs are applied to suppress the background from fake electrons:
 - MVA ID score > 0.90

Sources of systematic uncertainty considered:

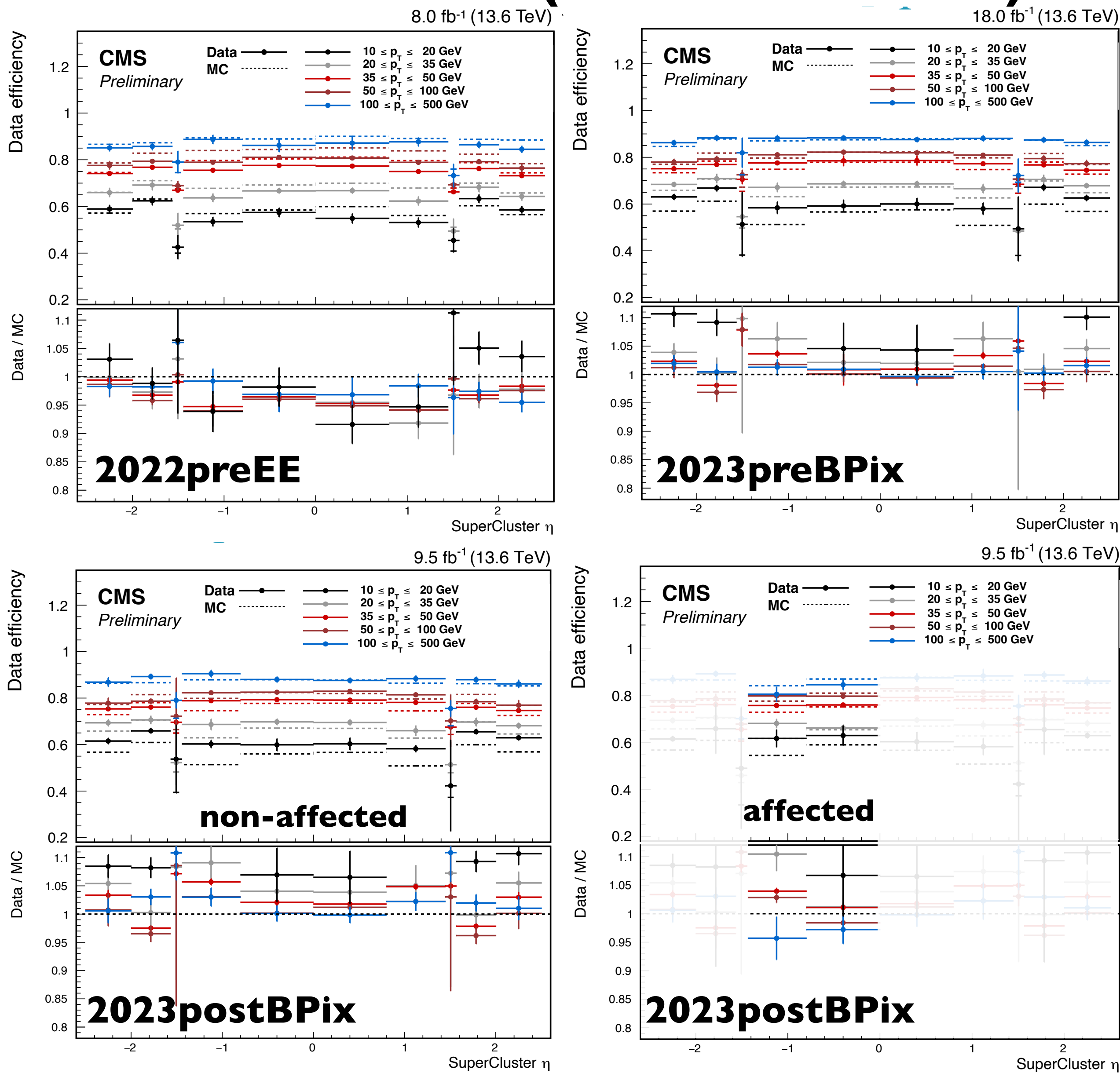
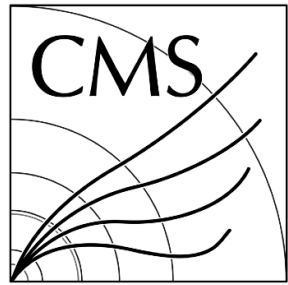
- Alternative signal fit
- Alternative background fit
- Alternative tag selection
- Alternative MC (only for 2023)

Binning:

p_T : [10, 20, 35, 50, 100, 500]

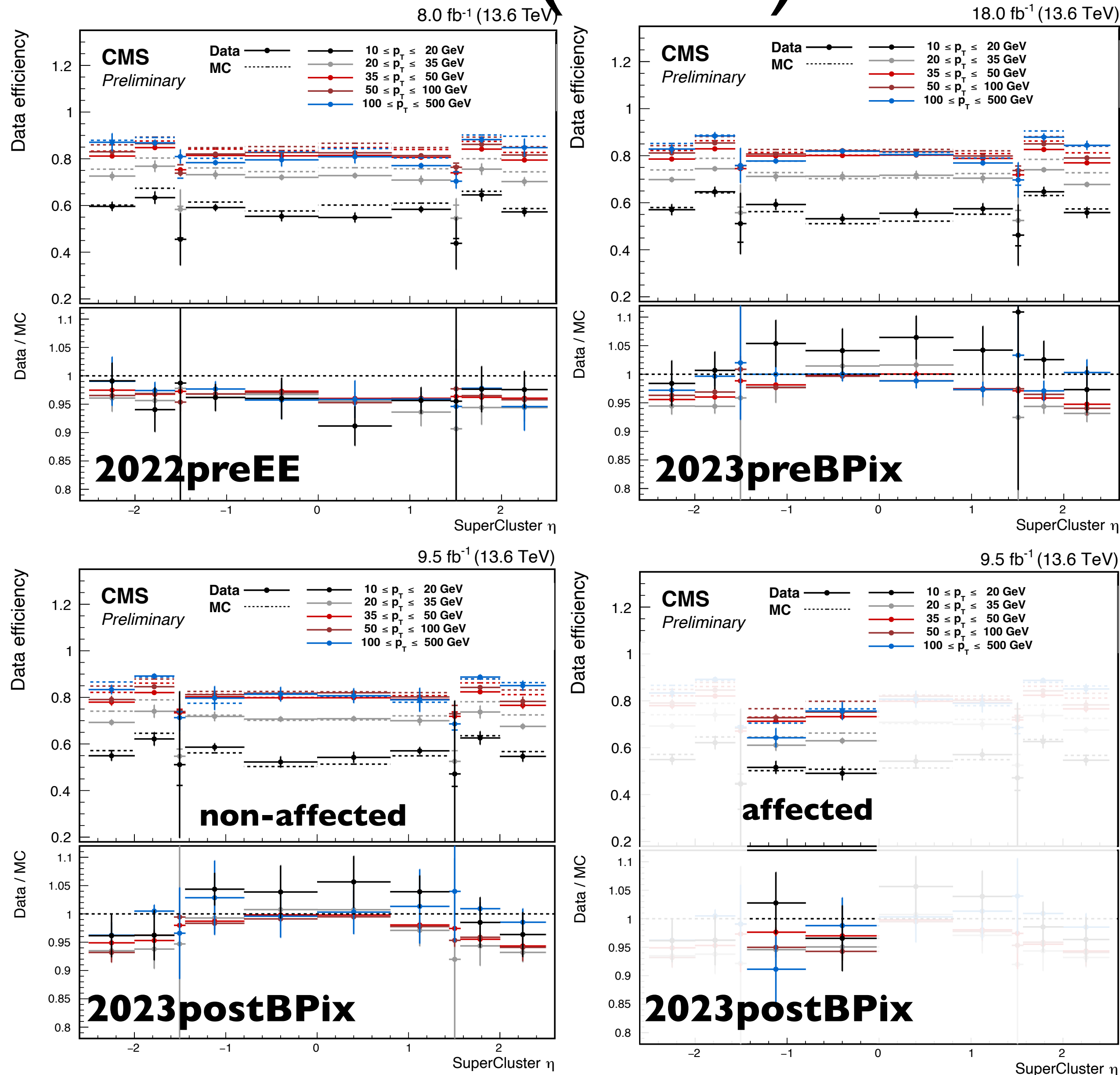
η_{SC} : [-2.5, -2.0, -1.566, -1.444, -0.8, 0, 0.8, 1.444, 1.566, 2.0, 2.5]

Electron ID (Cut-based) Efficiency ($10 < p_T < 500$)



■ **Tight** electron cut-based ID efficiency in 5 p_T ranges as a function of η_{SC}

Electron ID (MVA) Efficiency ($10 < p_T < 500$)



■ **WP80** electron MVA ID efficiency in 5 p_T ranges as a function of η_{SC}

Photon ID tuning - event selection

The IDs to be applied to offline reconstructed photons have been retrained and retuned in Run3 to offer a better performance with respect to the Run2 IDs.

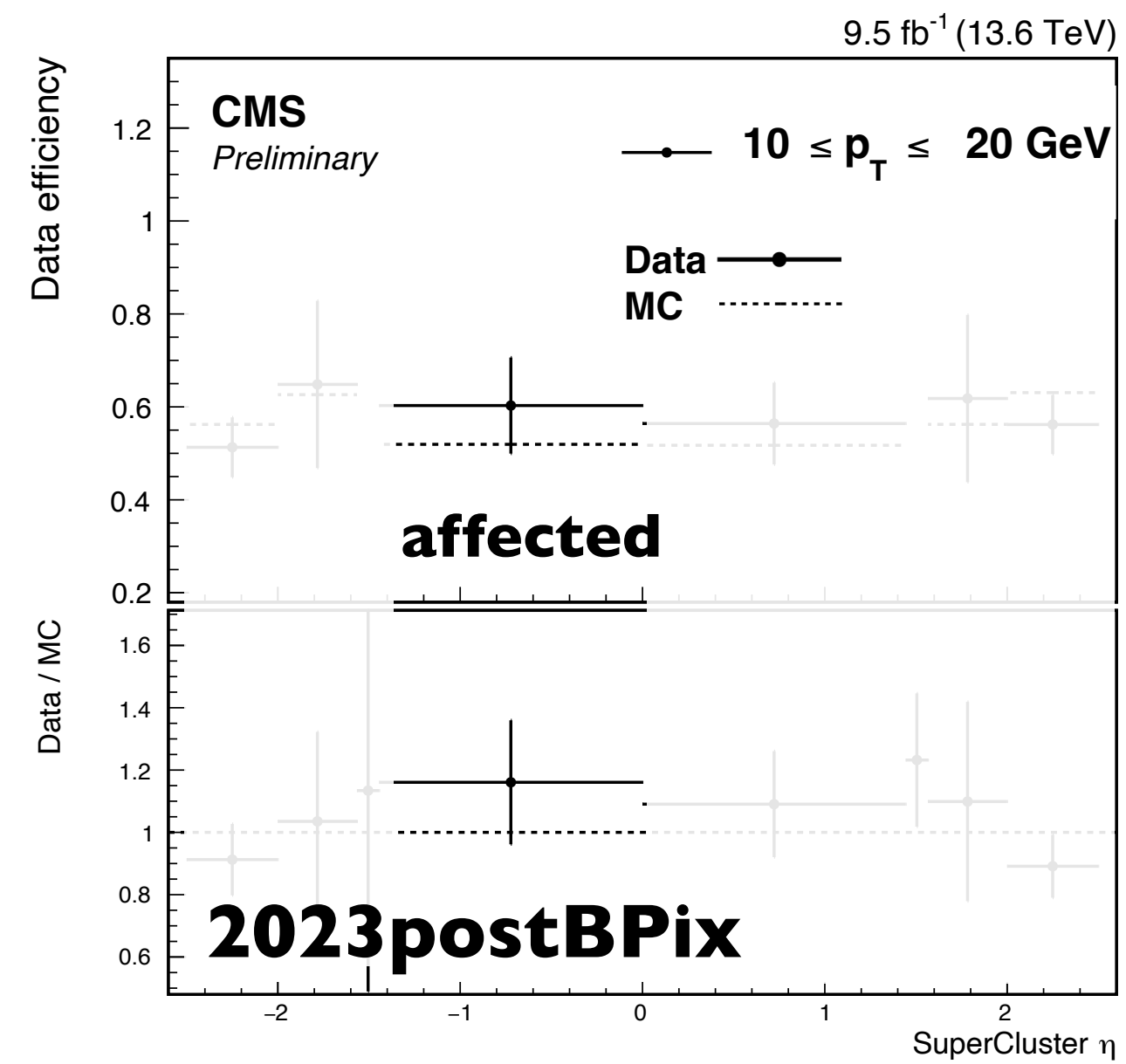
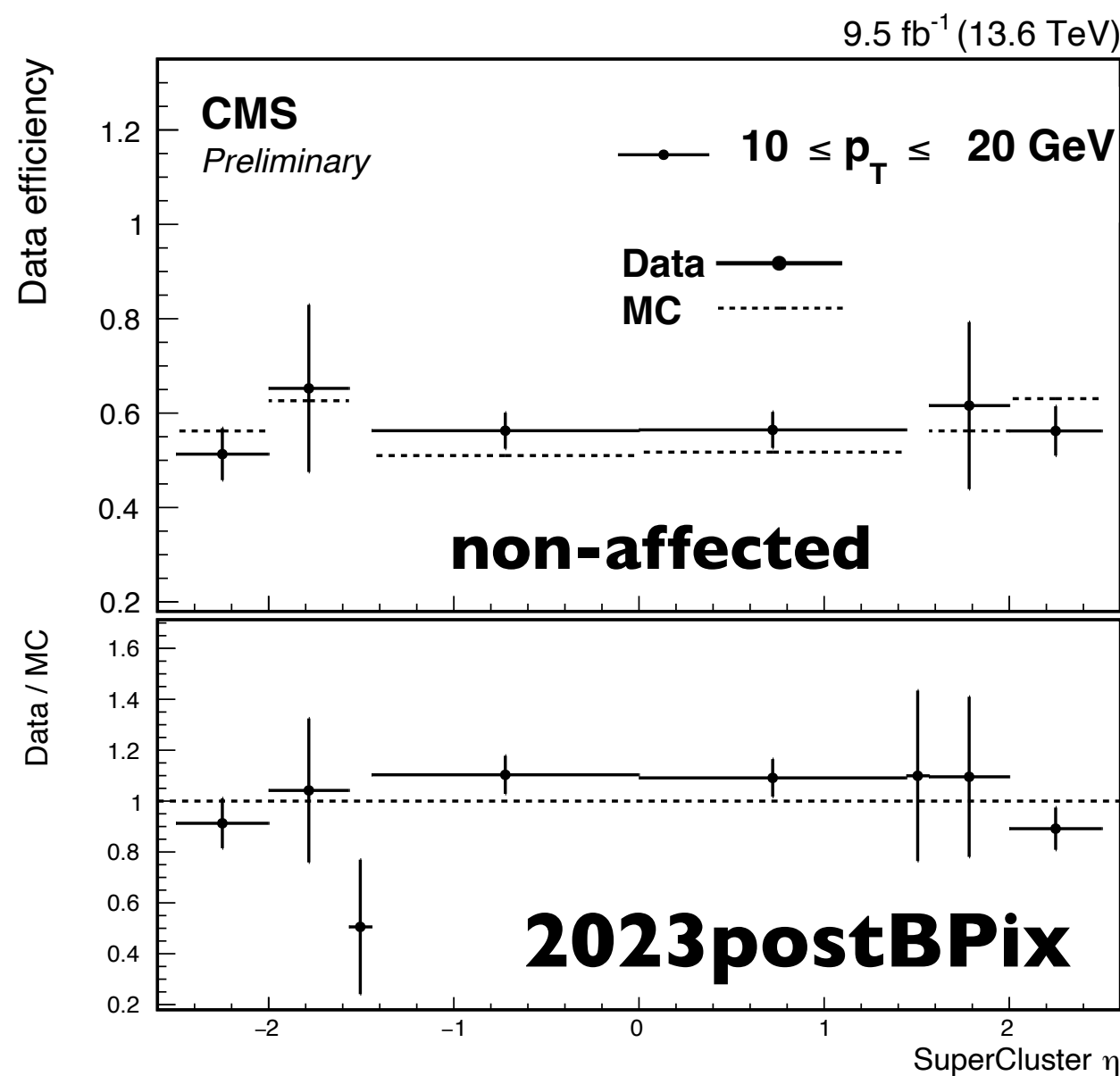
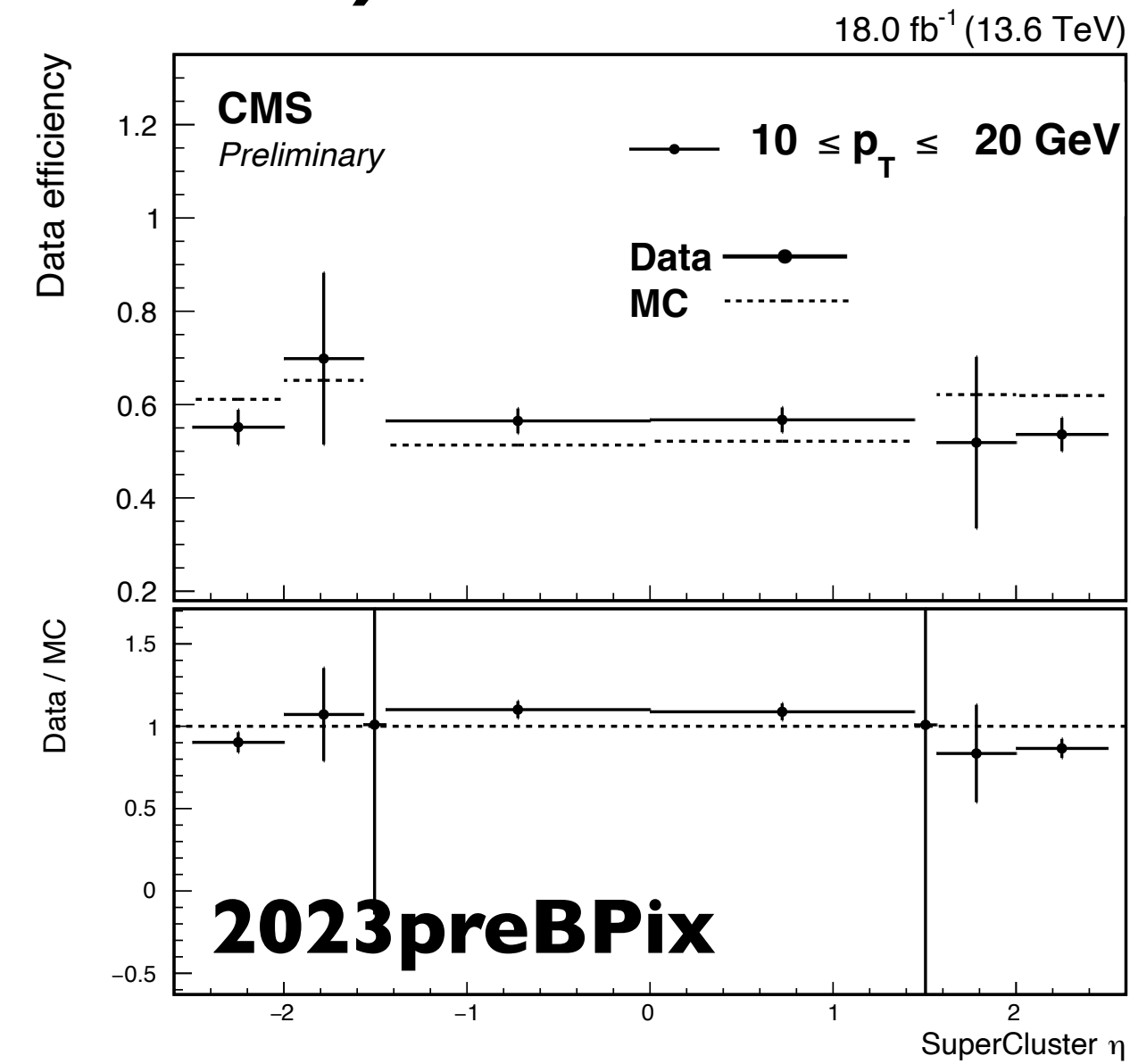
Goal: discrimination between signal and background photons.

Source of signal and background photons: EG-enriched γ + jets events simulated using the Pythia generator.

Signal photons: photons kinematically matched to prompt generator-level photons.

Background electrons: jets reconstructed as photons, not kinematically matched to prompt generator-level photons.

Photon ID (MVA) Efficiency ($10 < p_T < 20$)



■ **WP80** electron MVA ID efficiency in 5 pT ranges as a function of η_{SC}

Photon ID efficiency - event selection

Collision data: recorded using an unprescaled single electron trigger with a 32 (30) GeV transverse momentum threshold for 2022 (2023) data.

Simulation: Drell-Yan simulated at LO and NLO in QCD using the MadGraph and amc@nlo generators.

Tag electron selection for 2022 (2023):

- $p_T > 35$ (32) GeV, $|\eta_{SC}| < 2.17$.
- High purity cut-based ID
 - If $p_T^{\text{probe}} < 20$ GeV, additional cutoffs are applied to suppress the background from fake electrons:
 - MVA ID score > 0.99 , $R9 > 0.99$, $p_T > 50$ GeV

Sources of systematic uncertainty considered:

- Alternative signal fit
- Alternative background fit
- Alternative tag selection
- Alternative MC (only for 2023)

Binning:

2022 & 2023:

- p_T : [20, 35, 50, 100, 500]
- η_{SC} : [-2.5, -2.0, -1.566, -1.444, -0.8, 0, 0.8, 1.444, 1.566, 2.0, 2.5]

2023 only:

- p_T : [10, 20]
- η_{SC} : [-2.5, -2.0, -1.566, -1.4442, 0, 1.4442, 1.566, 2.0, 2.5]

Pixel veto and conversion-safe electron veto (CSEV) corrections

Data recorded with the di-muon trigger trigger.

Event selection:

- $80 < M_{\mu^+\mu^-\gamma} < 100$ GeV
- $M_{\mu^+\mu^-} + M_{\mu^+\mu^-\gamma} < 180$ GeV
- $(\Delta R(\mu_1^\pm, \gamma)) < 0.8$ or $(\Delta R(\mu_2^\pm, \gamma)) < 0.8$
- $(\Delta R(\mu_1^\pm, \gamma)) > 0.4$ and $(\Delta R(\mu_2^\pm, \gamma)) > 0.4$
- Pile up weights < 10

Muon selection:

- Muon prompt ID
- $p_T > 20$ (10) GeV
- $|\eta| < 2.4$

Sources of systematic uncertainty considered:

- Variation of the minimum bias cross section = 69200 (nominal) $\pm 3200 \mu\text{b}$
- Background systematic uncertainty.

Binning:

Number of PV: [10, 15, 20, 25, 30, 35, 40, 50, 100]

Photon p_T : [10, 15, 20, 25, 35, 50, 70, 200]

High R9 [$R9 > 0.96$], Low R9 [$R9 < 0.96$]

IDs:

Photon Cut-based: [Loose, Medium, Tight]

Photon MVA-based: [MVA80, MVA90]

The scale and smearing corrections

The goal of the scale and smearing corrections is to reduce the residual differences in the electron and photon energy scale and resolution between collision data and simulation. They are derived using electrons from $Z \rightarrow e^+e^-$ and they:

- scale the electron energy in data: $M_{ee}^{\text{scaled}} = M_{ee} \sqrt{(1 + \Delta P_{e1})(1 + \Delta P_{e2})}$, where ΔP_{ei} is the scale shift with respect to MC for electron “i”;
- smear the electron energy in the simulation: $M_{ee}^{\text{smearred}} = M_{ee} \sqrt{\text{Gaus}(1, \Delta C_{e1}) \text{Gaus}(1, \Delta P_{e2})}$, where ΔC_{ei} is the additional smearing for electron “i”.

The variables ΔP_{ei} and ΔC_{ei} are determined by minimizing a global binned negative log-likelihood of the invariant mass for each di-electron category. The categories are defined by subdividing events according to the data acquisition run number, the ECAL amplifier gain, and the electron η , p_T , and R_9 .

The derivation of scale and smearing corrections for photons uses electrons from Z boson decay reconstructed as photons by the CMS software, and to which the photon energy regression was applied.