中國科學院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Electron/Jet-to-Photon fake rate

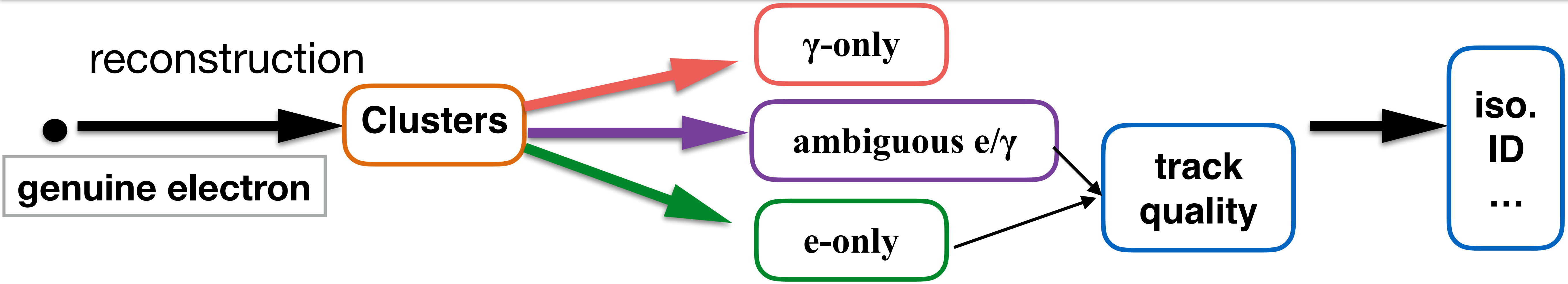
Yanping Huang[1],

[1]IHEP, CHINA

Contributions from Yifan, Shuo, Elena, Stefano, Kurt

Egamma workshop, Sheffield, 2019 Jan 29th

Electron-to-Photon Fakerate



$$\varepsilon_{\text{reco}} = \frac{N^{\text{sig}}_{\text{pass}}}{N^{\text{sig}}_{\text{pass}} + N^{\text{sig}}_{\text{fail}}} = \frac{\boxed{N^e_p - B^e_p}}{\boxed{(N^e_p - B^e_p)} + \boxed{(N^e_F - B^e_F)} + \boxed{(N^\gamma - B^\gamma)}}$$

Reconstructed candidates passing the probe track quality criteria

Reconstructed candidates failing track quality criteria

Electrons reconstructed as clusters with no associated track

$$\boxed{F_{e\rightarrow\gamma} = \frac{N^\gamma - B^\gamma}{(N^e_p - B^e_p) + (N^\gamma - B^\gamma)}}$$

- **Cluster reconstruction efficiency ~ 1**
- **Cluster types:**
  - **γ-only:** no track or TRT-only track (author=4)
  - **ambiguous e/γ:** large E/p or silicon track, no pixel (author=16)
  - **e-only:** pixel and not a good si-si track (author=1)
- Two different electron-to-photon fake-rate measurement.

Tag-Probe method

- Make use of the high statistical and pure electron via $Z\rightarrow ee$ process
- Selection：
  - **Tag electron with tight selection criteria**: tight ID/isolation and trigger match
  - **Probe electron with loose criteria (denominator of the fake rate)**: include the clusters in both the electron and photon containers, require the track quality for the electron-clusters. In practice, we can add the loose isolation cut to get pure signal (used in current measurement).
  - **Photon selection (numerator of the fake rate):** On top of the above selection for the probed electron, only require the objects from the photon container or +ambiguous e/γ depending on the overlap removal strategy. We can also use the PID requirement (used in current measurement)

$$\boxed{F_{e\rightarrow\gamma} = \frac{N_p^\gamma - B_p^\gamma}{(N_p^e - B_p^e) + (N_p^\gamma - B_p^\gamma)}}$$

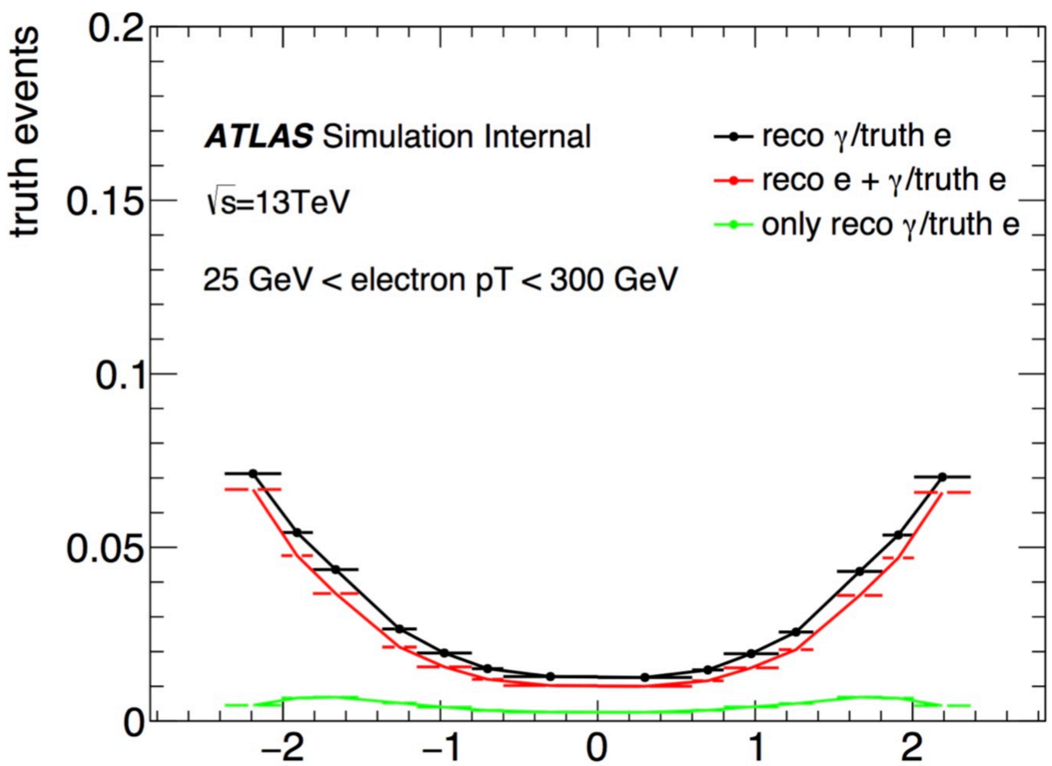
$$\Rightarrow$$

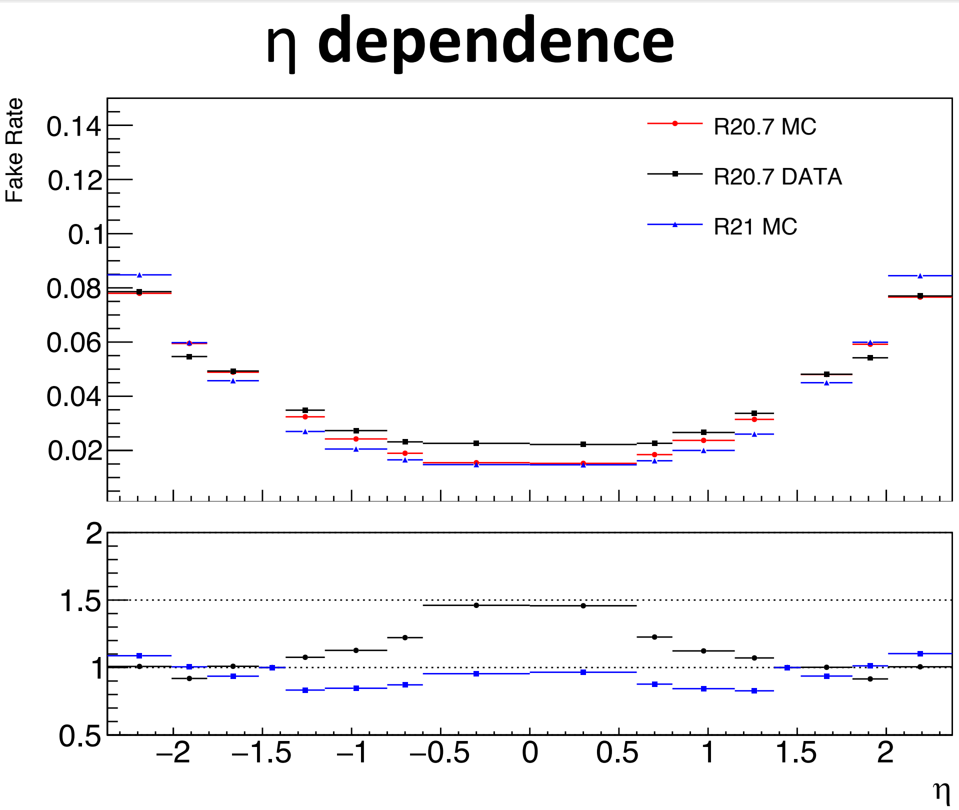
$$\boxed{F'_{e\rightarrow\gamma} = \frac{\left[N_p^\gamma - B_p^\gamma\right]\times\varepsilon_e\times\varepsilon_\gamma}{\left[(N_p^e - B_p^e) + (N_p^\gamma - B_p^\gamma)\right]\times\varepsilon_e} = F_{e\rightarrow\gamma}\frac{\varepsilon_e\times\varepsilon_\gamma}{\varepsilon_e}}$$

Tag-Probe method

- **Make use of the high statistical and pure electron via Z→ee process**
- **Selection**：
  - **Tag electron with tight selection criteria**: tight ID/isolation and trigger match
  - **Probe electron with loose criteria (denominator of the fake rate)**: include the clusters in both the electron and photon containers, require the track quality for the electron-clusters. In practice, we can add the loose isolation cut to get pure signal (used in current measurement).
  - **Photon selection (numerator of the fake rate):** On top of the above selection for the probed electron, only require the objects from the photon container or +ambiguous e/γ depending on the overlap removal strategy. We can also use the PID requirement (used in current measurement)
- **Binning**
  - 14 bins in η: ±0.0-0.6-0.8-1.15-1.37 and ±1.52-1.81-2.01-2.37;
  - 9 bins in p_T: 25-35-45-55-65-75-90-120-180-300 GeV

Fake rate estimation from MC

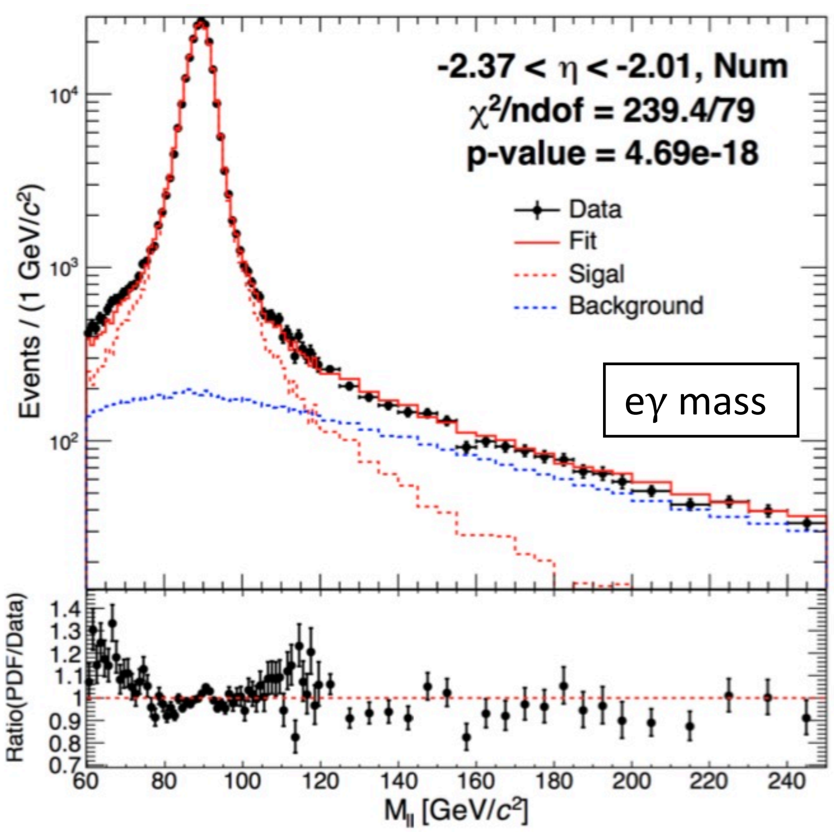




- Fakerate grows from 2% to 7% with η
- The fraction from the ambiguous e/γ cluster increases in the endcap regions.
- In the comparison between rel20.7 and rel21, there is a global fakerate decrease except a light increase in the very-endcap regions.

signal extraction

- Mass spectrum fitting in each individual η-pT bin for the signal and background separation.
- 3 different methods with different signal and background description in the mass spectrum fitting



- **Fitting strategy:**

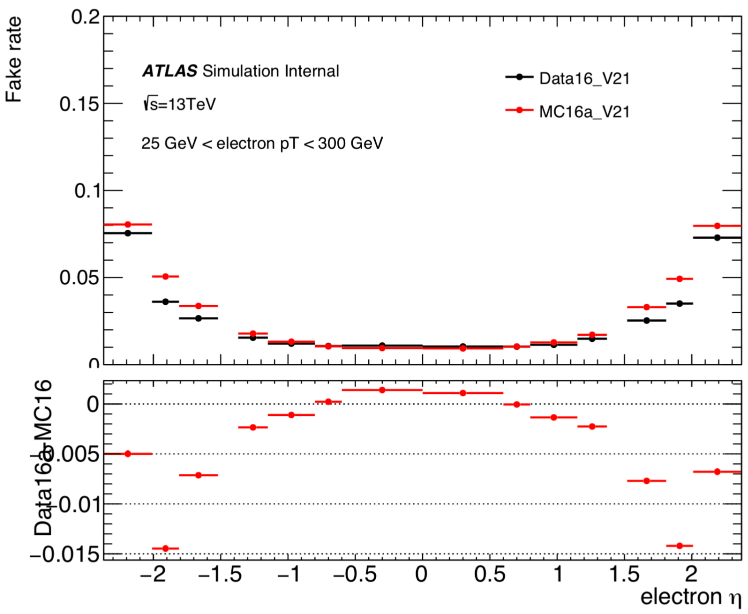
  1. Signal shape described from MC sample, background shape from the control region with signal leakage correction, and background is required to full occupy the mass region >120GeV

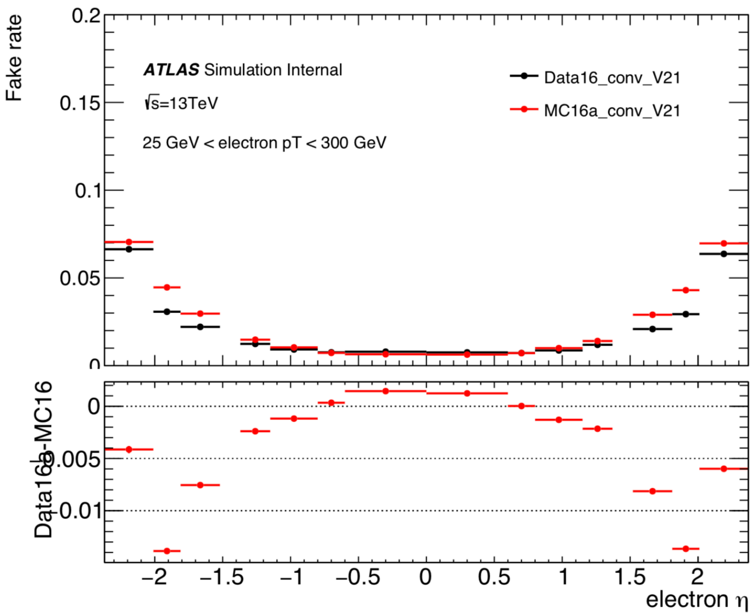
  2. Similar to the above way but float the background normalization

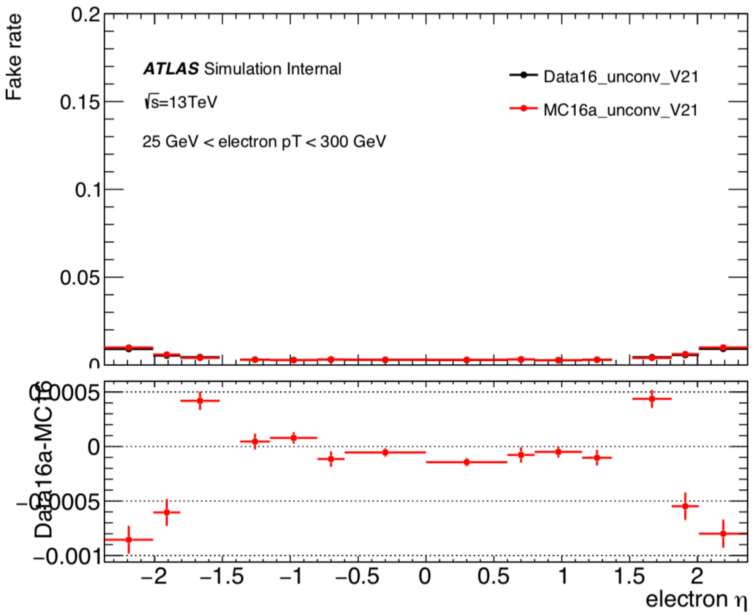
  3. Both the signal and background use analytic functions with float fractions

- Eventually, the signal yield is counted only in the peak region and small difference among the results from the above 3 methods.

Fake rate estimation from Data





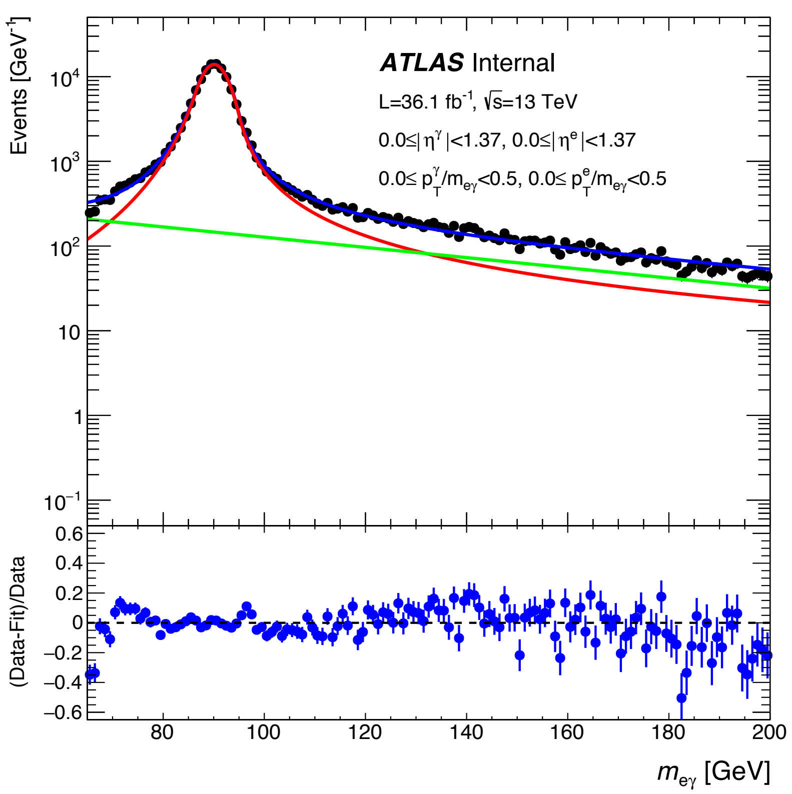


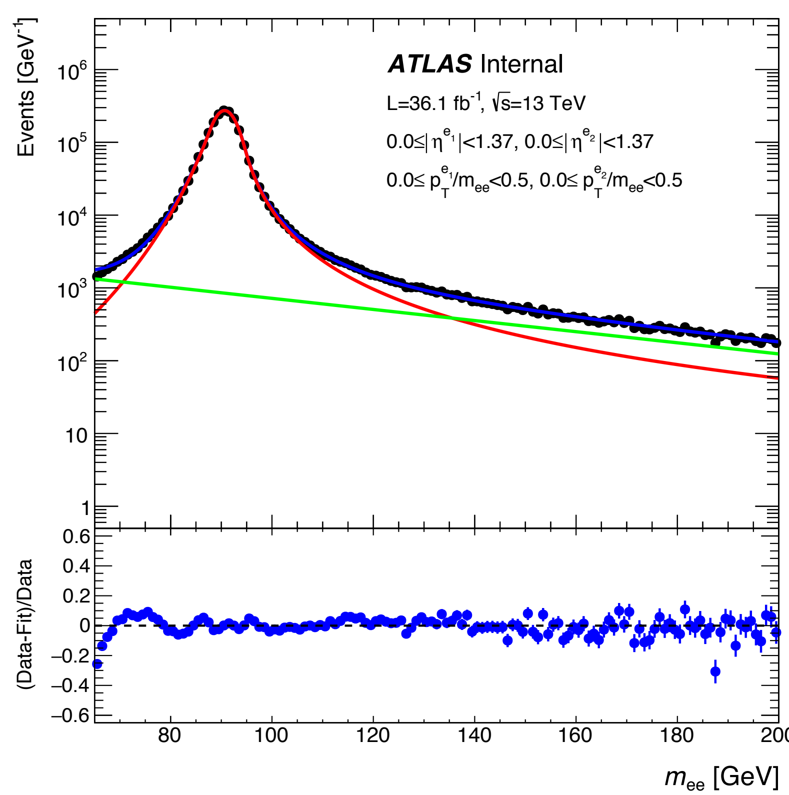
	All photons	conversion photons	unconversion photon
1st η bin:	FR=7.54%	FR=6.64%	FR=0.91%

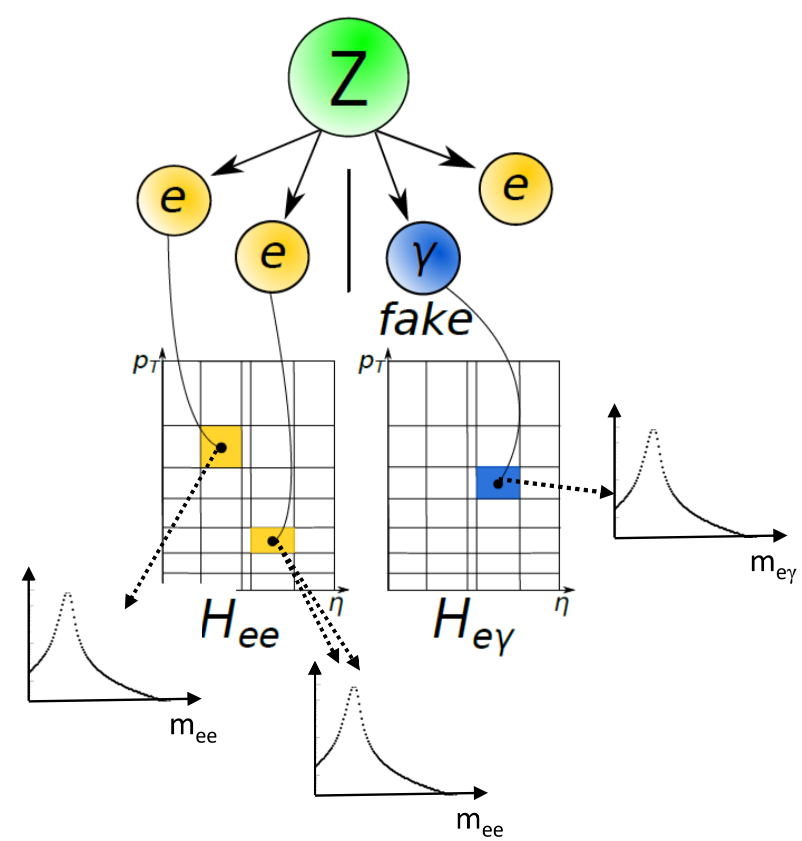
- ◈ Good agreement in the barrel, while slight large difference in the endcap. Unconverted photons give a small portion in the fake rate.
- ◈ Signal extraction is performed in each individual pt-η bin, which requires enough data statistics.
- ◈ **Plan for the next step:**
  - ✦ Complete the systematic uncertainty: mostly from the signal extraction via the mass spectrum fitting
  - ✦ Provide more general fake rate measurement: i.e. remove the isolation and ID criterial on the probed electron.
  - ✦ ……

Alternative e-γ fake rate method

- Consider the ee/eγ pair closest to the Z peak and split to coarse bin: |η|:{(0-1.37), (1.52-2.37)} and $p_T/m_{ee,e\gamma}$: {(0-0.5, 0.5-inf)} (taking into account the mass shape independence on p_T)
- Fit the mass spectrum in those coarse bin, and toss each event into eta-pT bins with the weight dependence on mass: $$\boxed{W(ee/e\gamma)=\frac{S}{S+B}(m_{ee/e\gamma})}$$
  - For ee pair, label the electron twice
  - For eγ pair, only label the photon Fit the mass spectrum in those coarse bin, and toss each event into eta-pT bins with the weight dependence on mass:
- Final fake rate: $$\boxed{F_{e\to\gamma}(\eta,p_T)=\frac{H_{e\gamma}(\eta,p_T)}{H_{ee}(\eta,p_T)}}$$







Alternative e-γ fake rate method

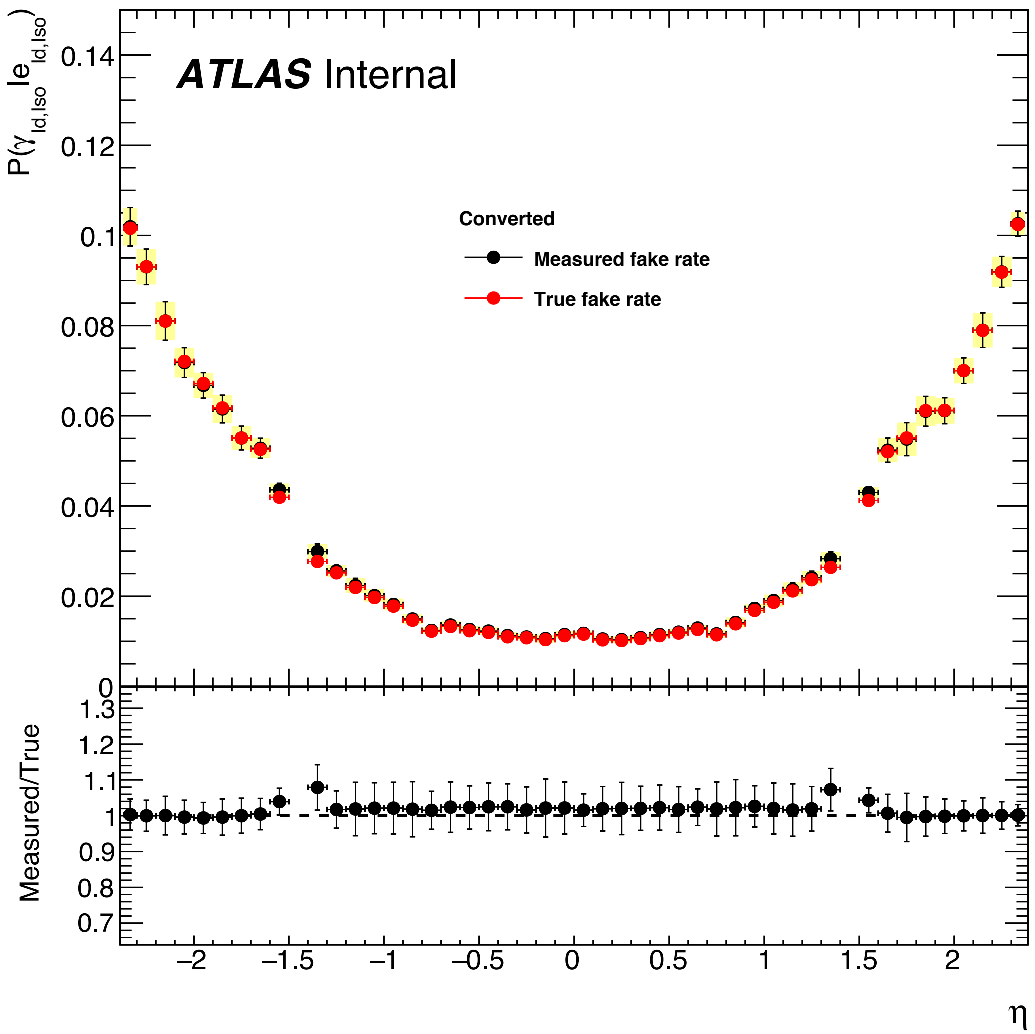
- Use coarse bins for the signal background separation: to reduce the statistical fluctuation.
- Signal-background rate is extracted in the coarse bins (**mass fitting bin**), but the fake rates are estimated in small pt-eta bins (**fake rate bin**): **it bases on the assumption that the S/B are similar or large among those bins.**
  - *If the mass fitting bins are enough as the fake rate bin, it is similar as previous method*
- Dedicated selection requirement to mimic one analysis selection:

	Electrons	**Photons**
Kinematic selection	p_T>25 GeV, |η|<1.37 or 1.52<|η|<2.37.	p_T>25 GeV, |η|<1.37 or 1.52<|η|<2.37.
Identification	Likelihood-based `Tight` working point.	`Tight` working point.
Isolation requirements	`FixedCutTight` working point.	`FixedCutLoose` working point.

$$F_{e\to\gamma} = \frac{\epsilon(e^{true}\to\gamma^{reco})}{\epsilon(e^{true}\to e^{reco})}\frac{\epsilon_\gamma}{\epsilon_e} = \frac{N(e^{true}\to\gamma^{reco})/N_e^{true}}{N(e^{true}\to e^{reco})/N_e^{true}}\frac{\epsilon_\gamma}{\epsilon_e} = \frac{\rho}{1-\rho}\frac{\epsilon_\gamma}{\epsilon_e}$$

$$F_{e\to\gamma} = \frac{\rho}{1-\rho}\frac{\epsilon_\gamma}{\epsilon_e}$$

Closure test



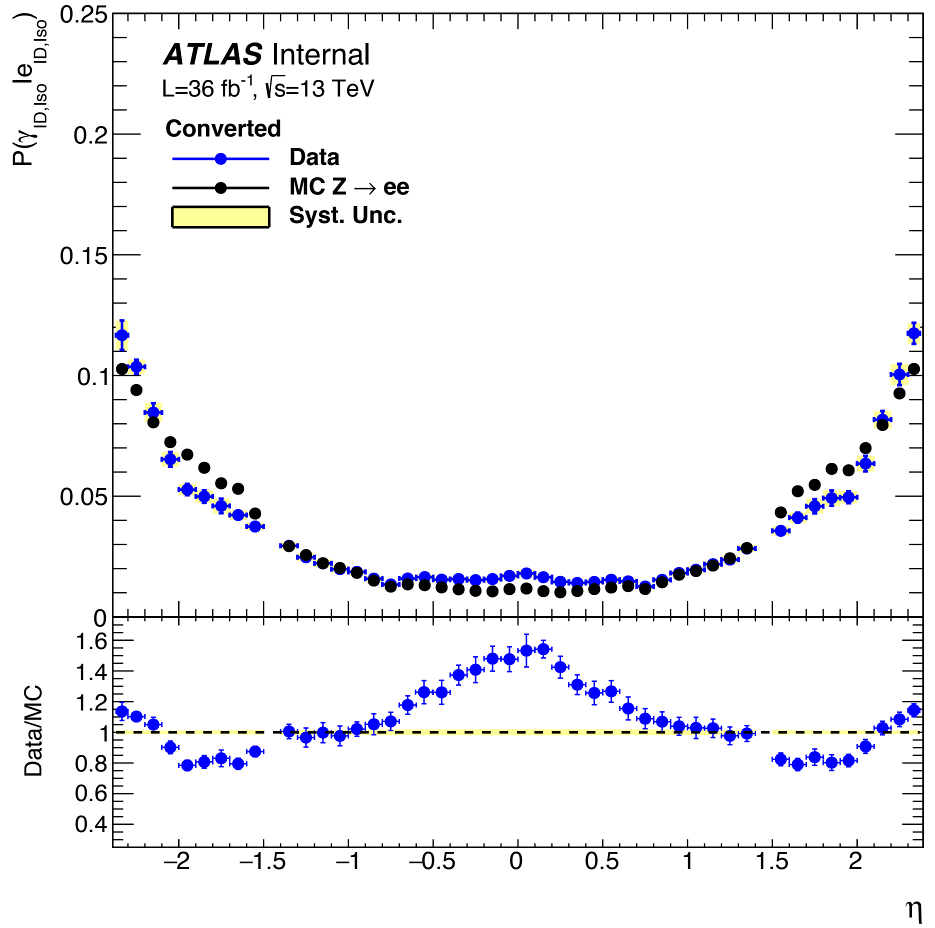
Truth matching:

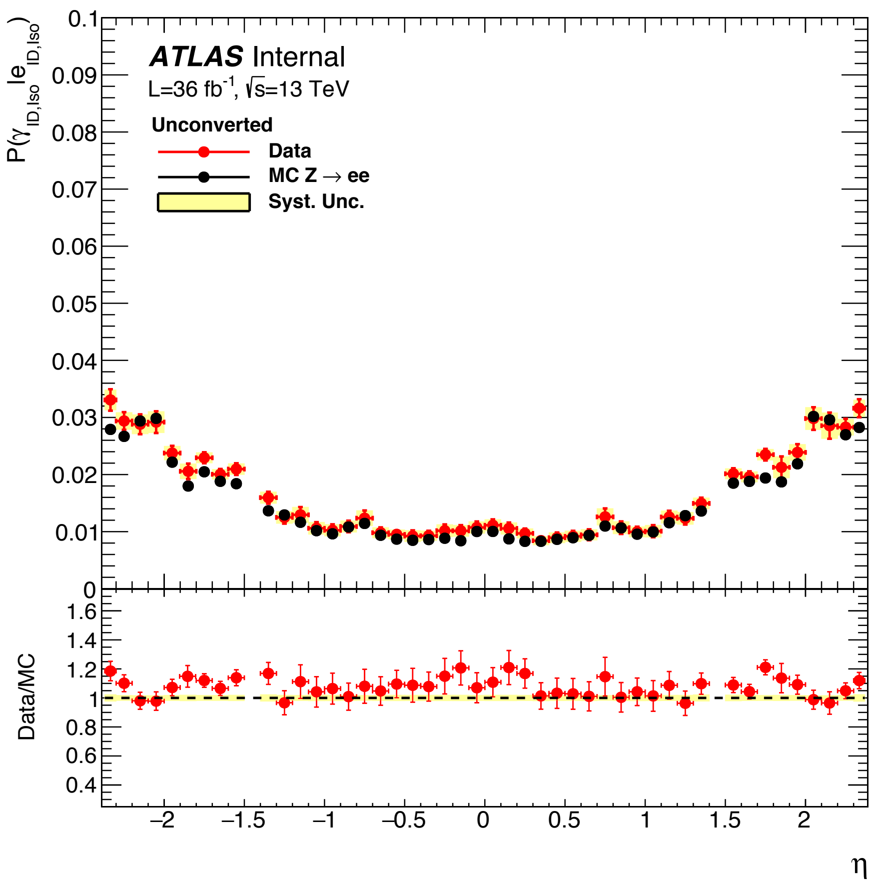
- Include the reconstructed particle to FSR photon which origins from Z
- truth matching algorithm always match the particle to the closer photon
- In this analysis, there are many soft photons collinear to electrons from Z.

◈ A good closure for the pure signal case (only include the Z→ee MC)

◈ How about the closure situation to data with both signal and background? Currently, for the fake rate estimation in data, the difference from without background subtraction is taken as the systematic uncertainty.

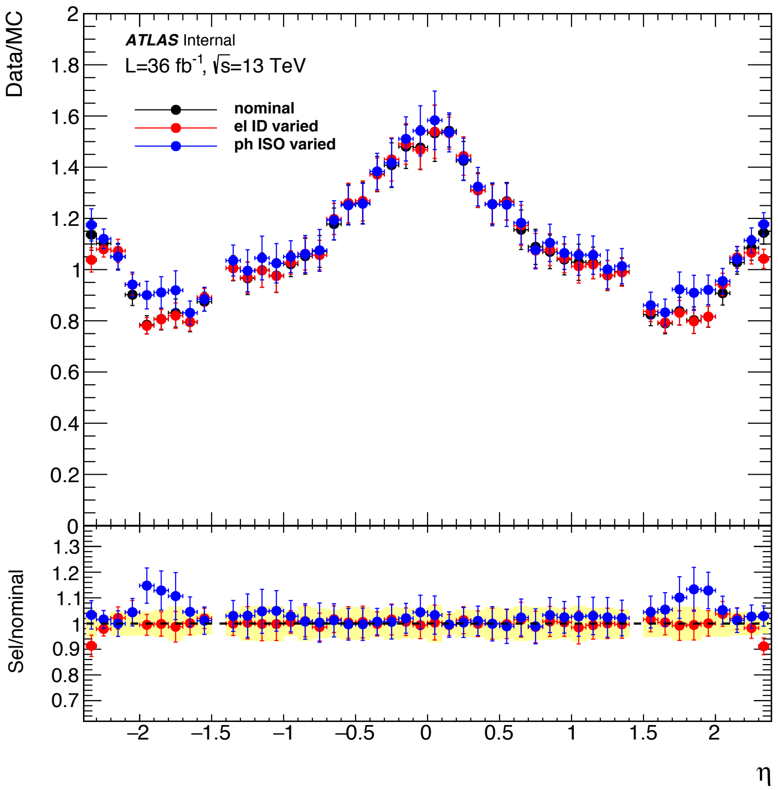
Fake Rate Measurement

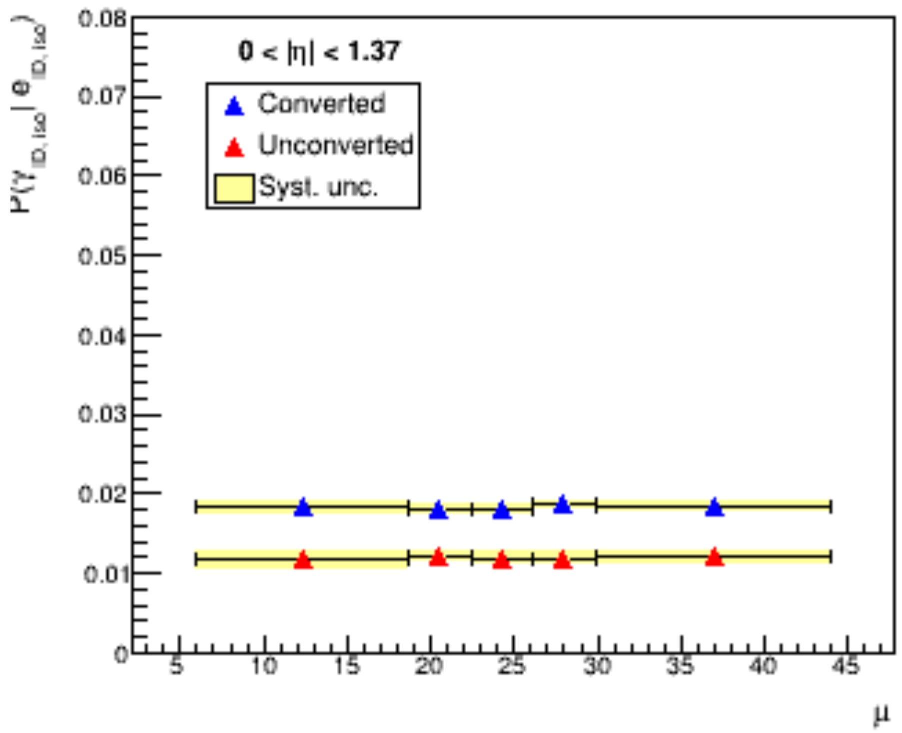


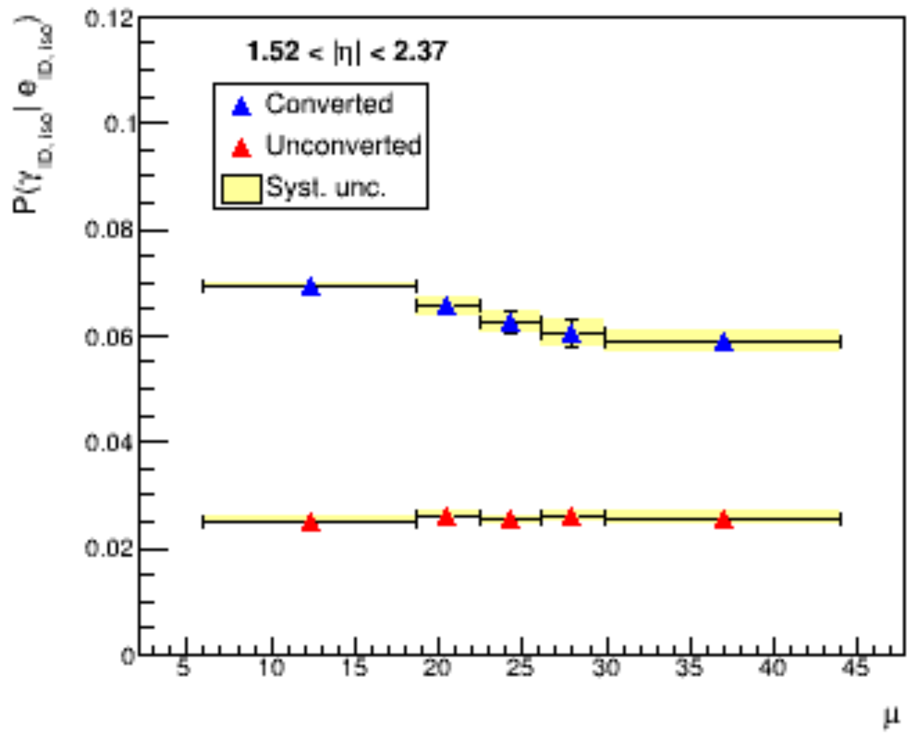


- Fake rates are measured with the dependence on η-pT bins.
- A well agreement between data and MC for the fake rate of electrons to unconverted photons.
- A visible discrepancy for the converted photons, but it is a bit different from the result obtained in the previous method.

Dependence on ID/iso/pileup







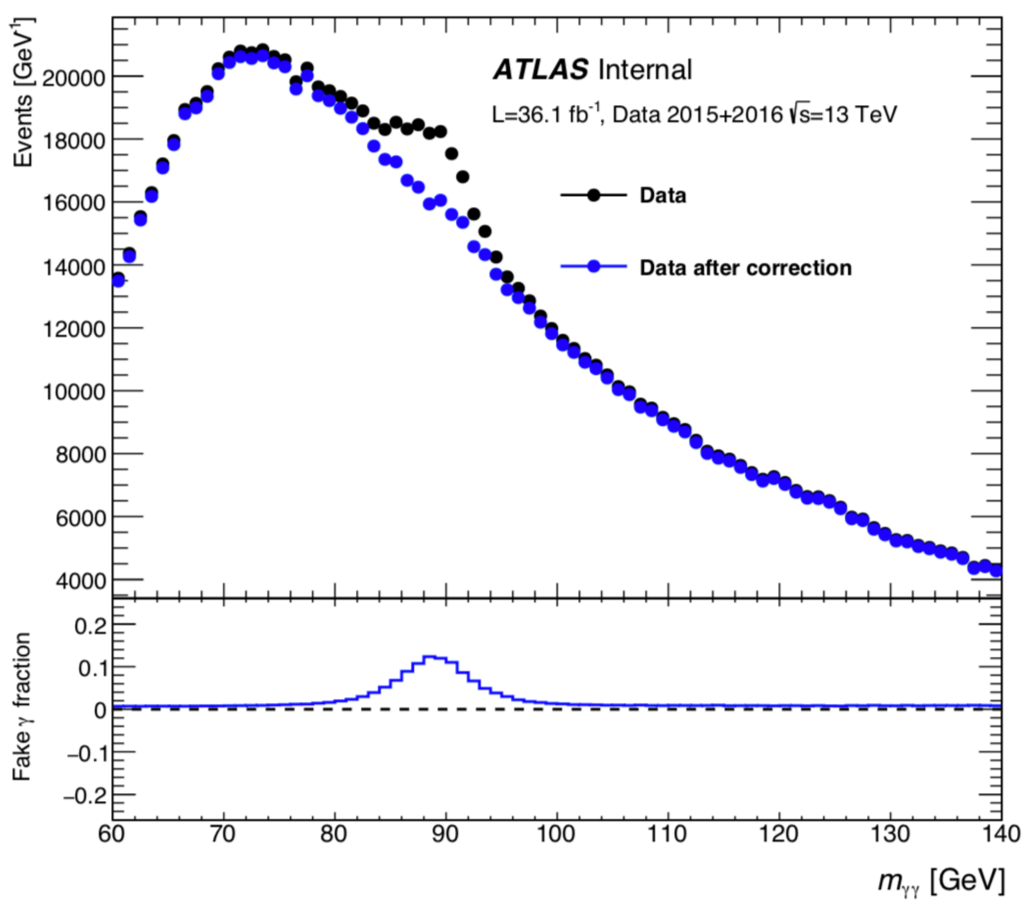
- The difference sets (e/γ ID/Iso) of scale factors are in good agreement, therefore almost independent on the analysis setup

- The fake rate is almost approximately constant with μ both in barrel/ endcap regions for both converted/unconverted photons, with a slight tendency to decrease with the increasing μ in the endcap region for converted photons.

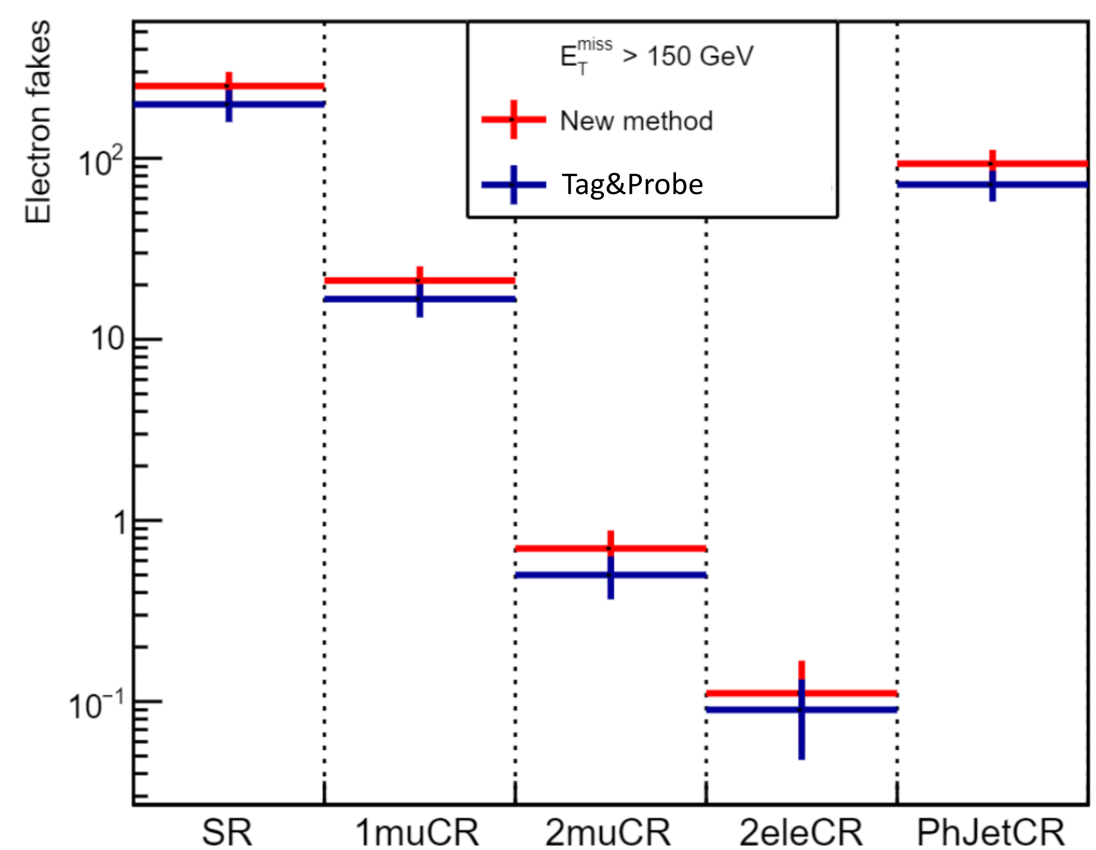
Application in analyses

Keep the photon/electron selections in the fake rate estimation are same as that in the signal region and control region in the dedicated analysis

$$F_{e\to\gamma} = \frac{\rho}{1-\rho}\frac{\epsilon_\gamma}{\epsilon_e}$$



Background estimation in di-photon analysis



Background evaluation in the Mon-photon analysis

Jet-to-photon fake rate

- Jet-to-photon fake rate definition:

$$F_{jet-\gamma} = \frac{N_{\gamma}}{N_{geni-jet}}$$

- In practice, we are concerning on the background contribution in the signal region (normalization and shape).

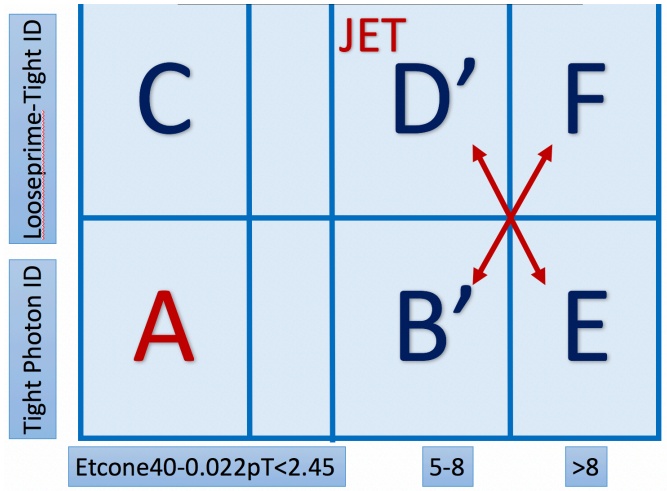
- Purity estimation with data-drive methods: ABCD method, template fit method, matrix method.

- A direct connection between fake rate and purity:

$$F_{jet-\gamma} = \frac{N_{\gamma}\varepsilon_{\gamma}}{N_{geni-jet}\varepsilon_{jet}} = \frac{N_{all}\cdot(1-P)\cdot\varepsilon_{\gamma}}{N_{geni-jet}\varepsilon_{jet}} = \frac{N_{rec-\gamma}\cdot(1-P)}{N_{rec-jet}}$$

ABCD method

- 4 different regions are defined for the purity estimation with the correction of the correlation coefficient and signal leakage.



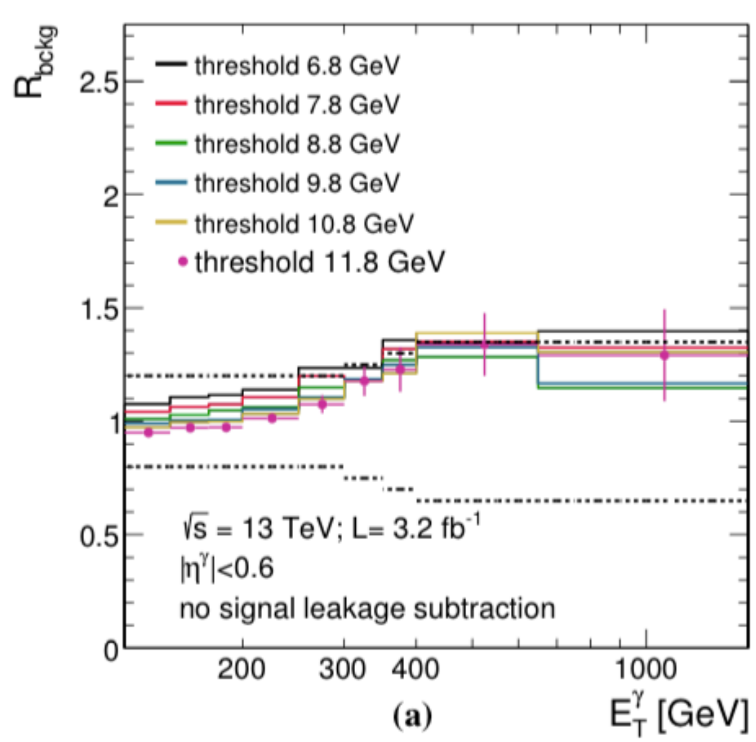
$$N_\gamma^A = N_A - \rho_{jet}\left(N_B - N_\gamma^A\, c_B\right)\frac{N_C - N_\gamma^A\, c_C}{N_D - N_\gamma^A\, c_D}$$

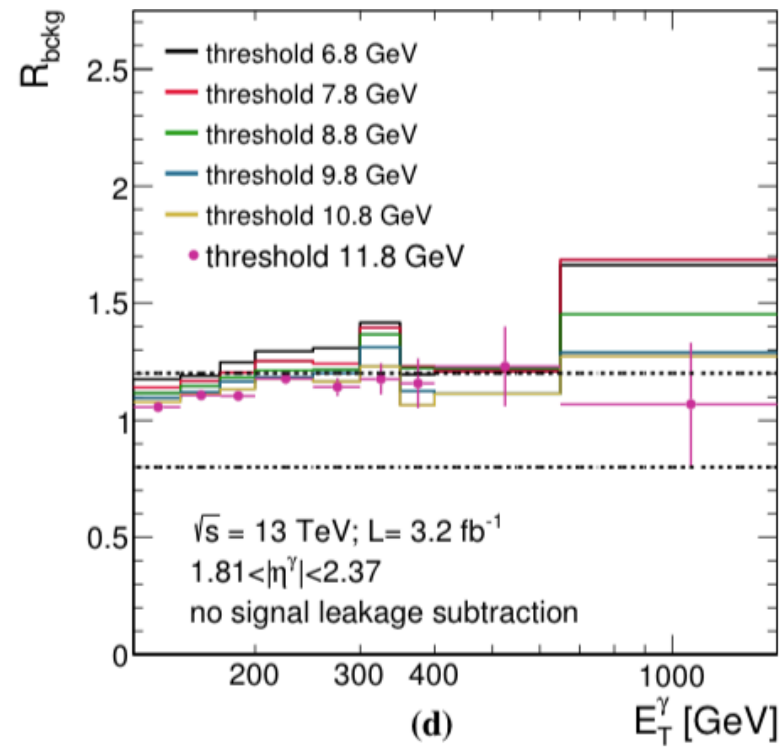
$$Purity = N_\gamma^A / N_A$$

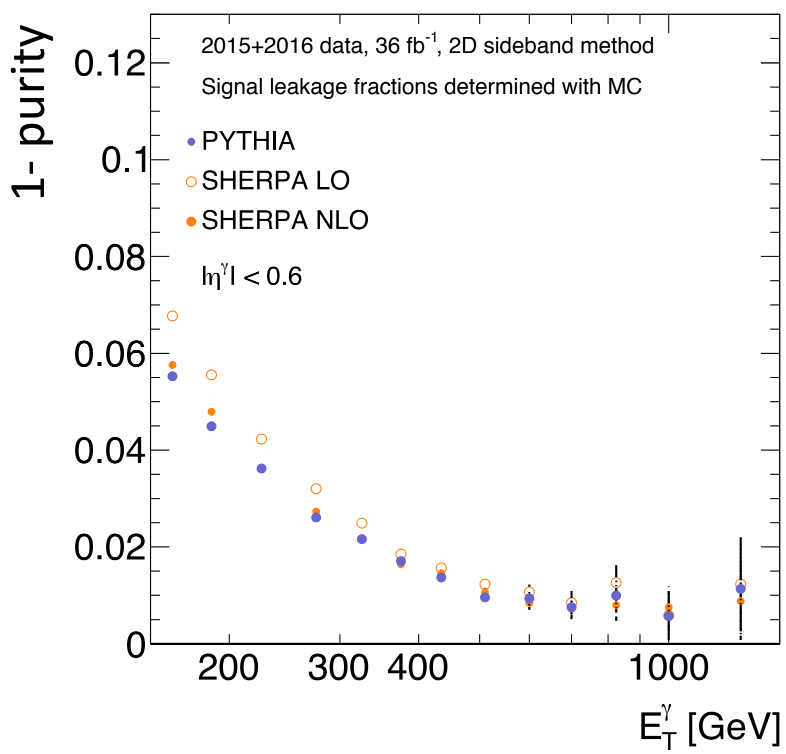
- In this method, we can only obtain the normalization information.
- How to estimate the correlation coefficient in real data?
- A closure test: pseudo-data with the mixture of single photon mc (26.2%) and di-jet MC (73.8%):

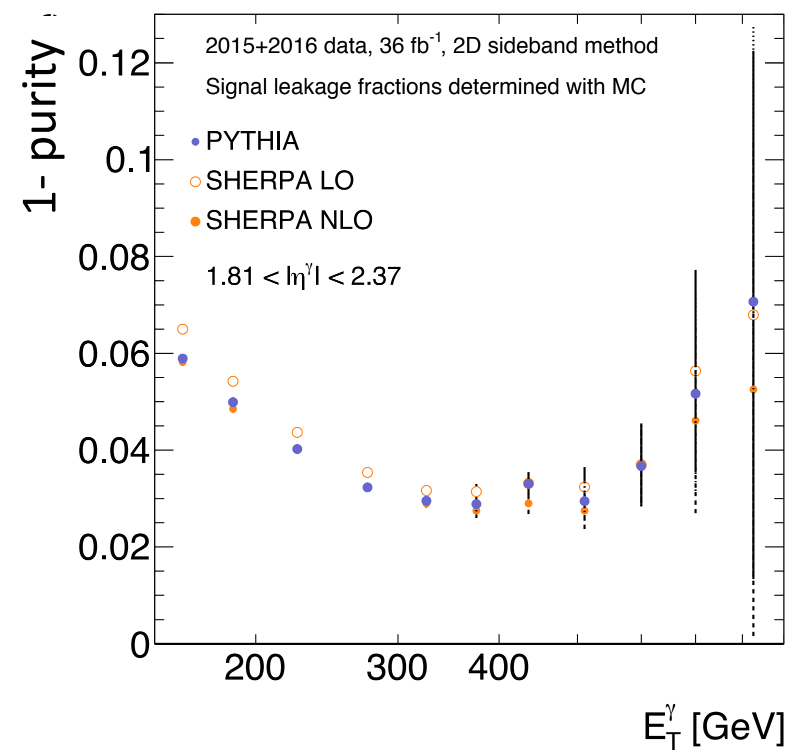
X-axis of ABCD	ρ_{jet}	How to get ρ_{jet}	Purity (mimic data)
p_T^{cone20}/p_T	1.05±0.04	B'D'EF region	28.3±2.6%
p_T^{cone20}/p_T	1.08±0.04	Real value (from MC)	26.2±3.1%
$E_T^{cone40} - 0.022p_T$	1.05±0.05	B'D'EF region	36.5±3.1%
$E_T^{cone40} - 0.022p_T$	1.20±0.05	Real value (from MC)	26.2±3.8%

Purity estimation in inclusive isolation photon analysis





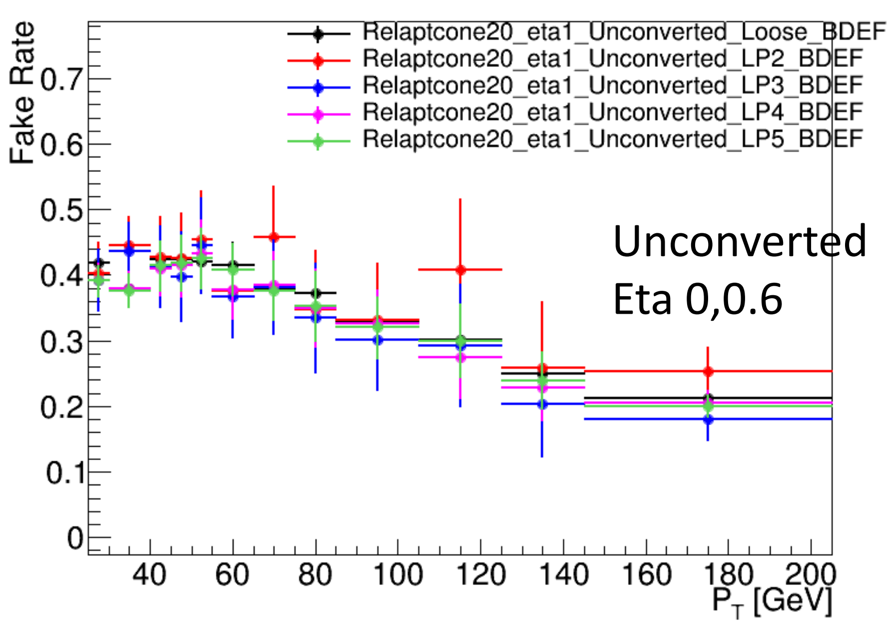




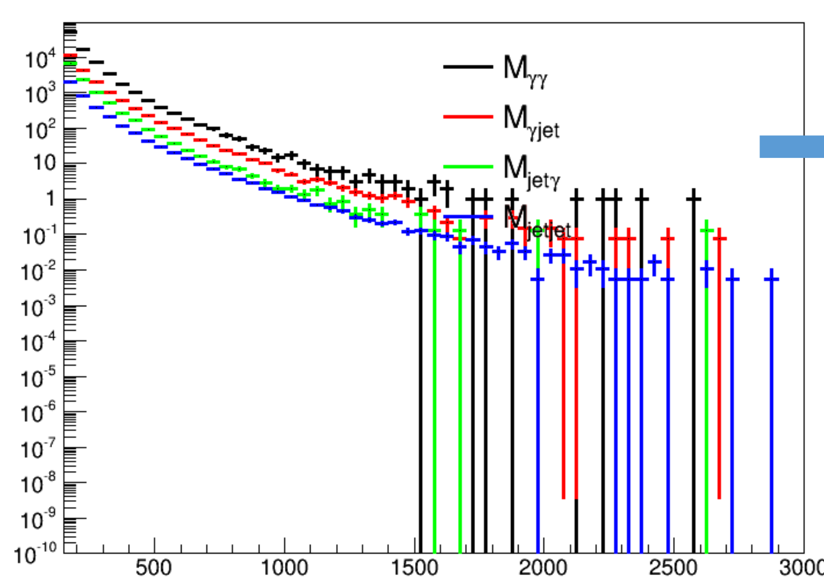
◈ 1-p: 0-0.06

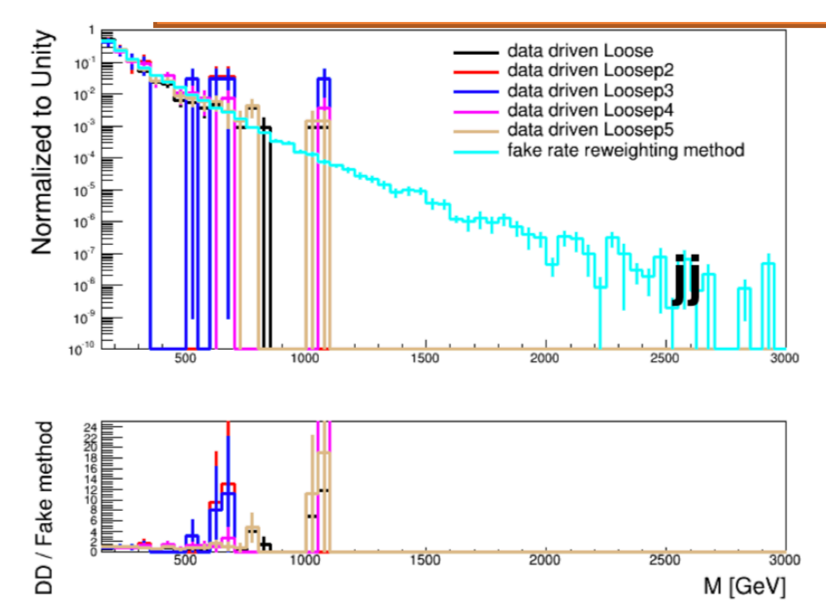
Fake rate estimation with single photon

- With high statistics single photon sample,
  - Define one CR and one SR where consistent object selection in the di-photon selection
  - we can estimate the purity dependence on individual pt-eta bin, and translate it to fake rate.
- With diphoton events:
  - apply the above fake rate on the control region to estimate the reducible background in the SR
  - The shape and normalization can be obtained



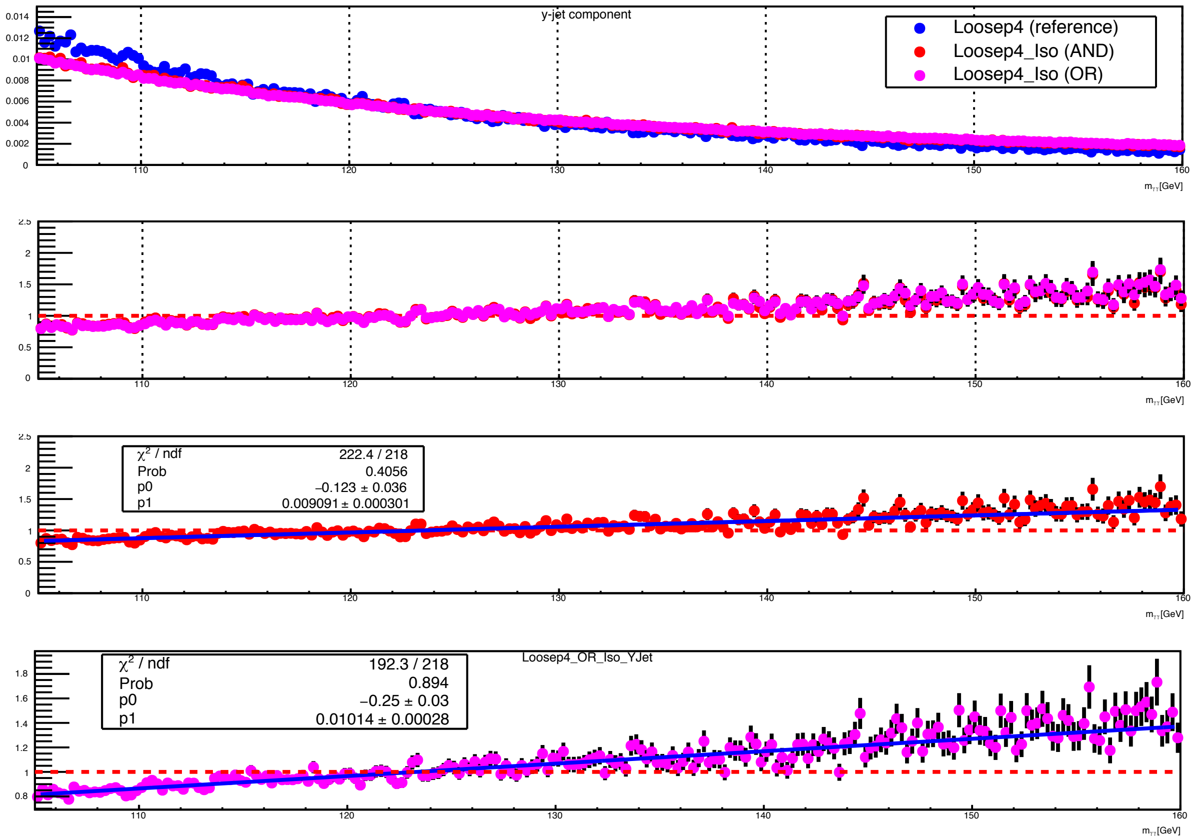
Original non-isolated shape

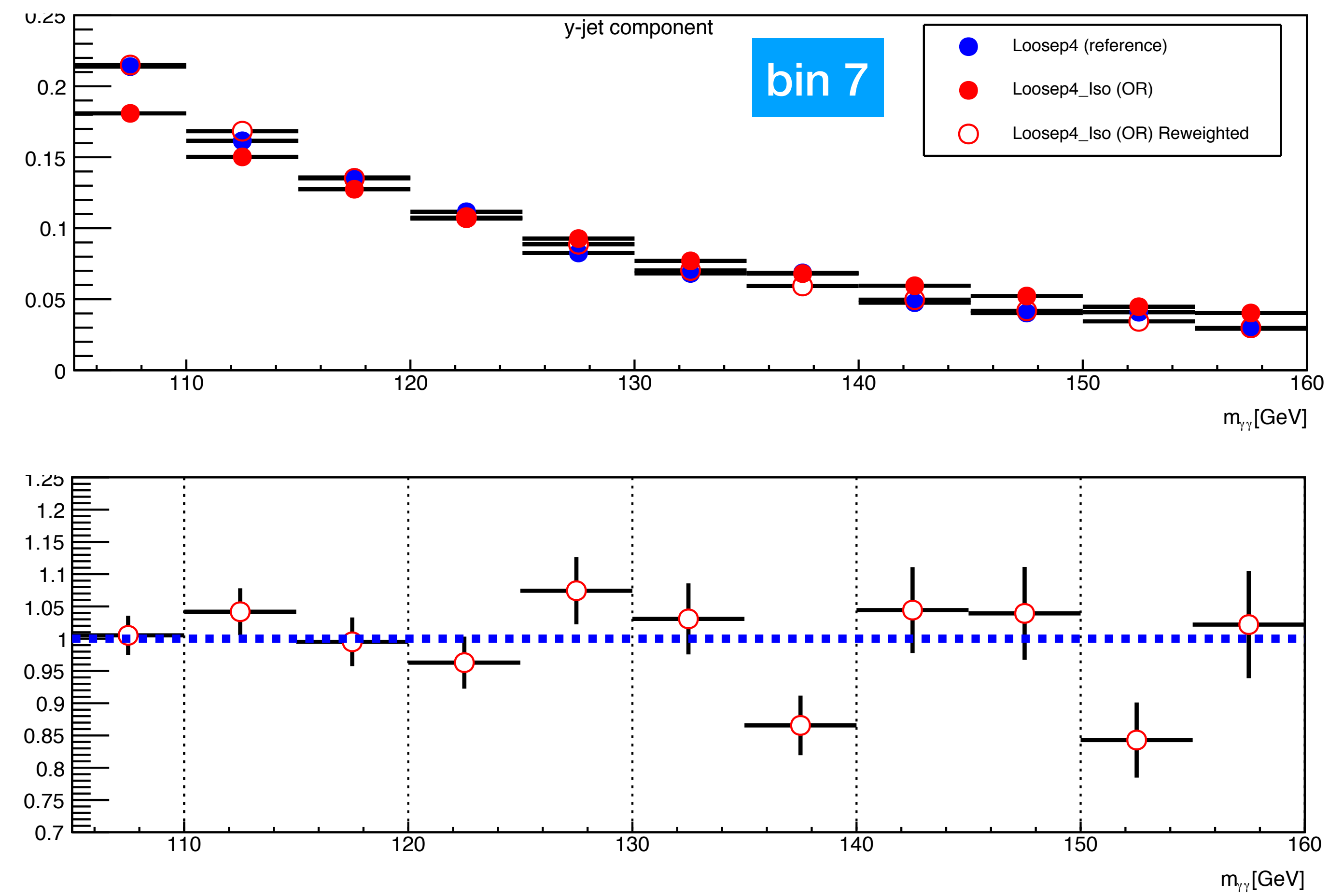




Reducible background shape estimation in di-photon analysis

Inclusive bin: y-jet component



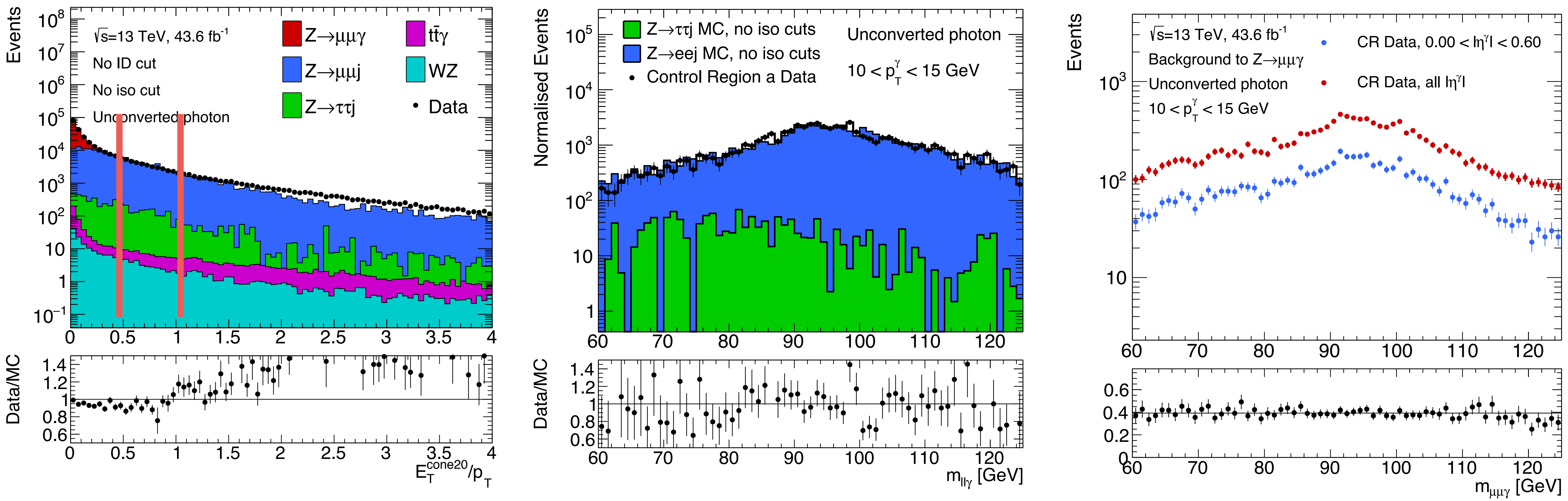


◈ In di-photon analysis, the background shape is described with CR, and purity is estimated in the signal region with data-driven method.

◈ While the statistics is not enough in some cases: loose the CR and there is a clear linear dependence on the mass shape ratio in the inclusive case. With the reweighting, we have successfully described the shape in most differential cross section bins.

◈ Possibility: This might provide an indirect way for the fake rate measurement.

Reducible background shape estimation in Z-radiative photon



- Inverse isolation cut is used to increase the CR statistics. And tighten the cut with an upper limit to avoid biasing the mass shape.
- Almost no correlation between the mass shape and the eta cuts.
- More validations are ongoing.

Summary

- ◈ Electron-to-photon fake rate:
  - ✦ 2 different methods are employed for the fake rate measurement. It is ~2%-7% dependent on eta.
  - ✦ For the data/MC scale factor, weak dependence on the ID/Iso cuts and pileup.
- ◈ Jet-to-photon fake rate:
  - ✦ In most analyses, the purity is used and measured via data-driven method, instead of direct measurement on fake rate.
  - ✦ Due to the extensive area and missing MC sample, default CR may not have enough statistics. Different CR studies are ongoing.