The estimation methods of the background induced by the misidentification of a jet as a photon in *pp* collisions at $\sqrt{s}$ = 13 TeV with the ATLAS Detector

K. Kazakova, E. Soldatov, D. Pyatiizbyantseva

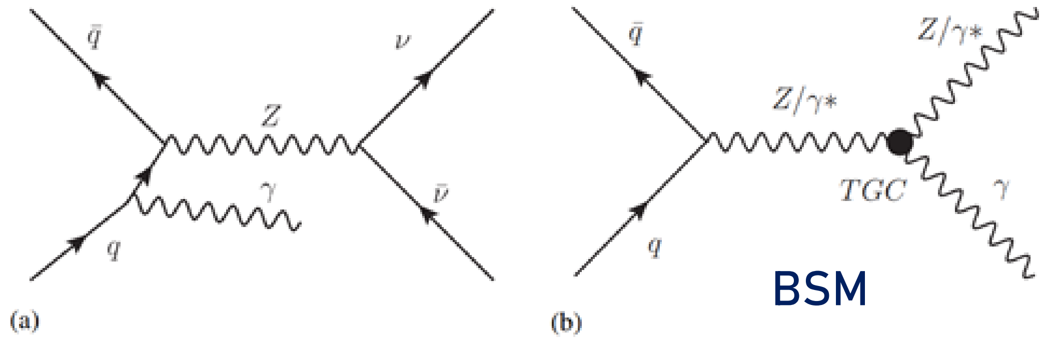
on behalf of the ZnunuGamma group

MEPhI@Atlas meeting
16/12/2022

Motivation and goals

Motivation:

- To measure the parameters of the Standard Model (SM) to very high precision;
- The search of new physics predicted by the beyond SM (BSM) theories;
- Precise measurements of triple and quartic gauge couplings sensitive to BSM physics. One of the sensitive processes is Z(νν)γ process.



Goals:

- To calculate integral and differential in E_T^{γ}, N_{jets}, p_T^{miss}, $\Delta\phi(\gamma, p_T^{\text{miss}})$, $p_T(Z\gamma)$, η_γ. cross-sections and compare the results with the theory predictions;
- To obtain the strongest up-to-date limits on anomalous neutral triple gauge-boson couplings (aTGCs).

⇨ We want to estimate backgrounds as accurate as possible but background processes emerging from object misidentification are not well-modeled in Monte-Carlo. All analyses at the LHC experiments use data-driven methods to solve this issue.

The backgrounds and the phase space definition

Signal: Z(νν)γ

Backgrounds:

- 35% • γ + jets – via MC → ABCD method based on $E_{\mathrm{T}}^{\mathrm{miss}}$ significance and additional variable (or slice method?);
- 26% • W(→lν)γ – fit to data in additional CR based on N_{lep} (shape from MC);
- 20% • $e \rightarrow \gamma$ – fake-rate estimation using Z-peak (tag-n-probe) method;
- 14% • $jet \rightarrow \gamma$ – ABCD method based on photon ID and isolation and slice method;
- 1.9% • Z(ll)γ – via MC;
- 1.6% • ttγ – fit to data in additional CR based on N_{lep} (shape from MC).

- FixedCutLoose isolation working point is chosen.

Preselections

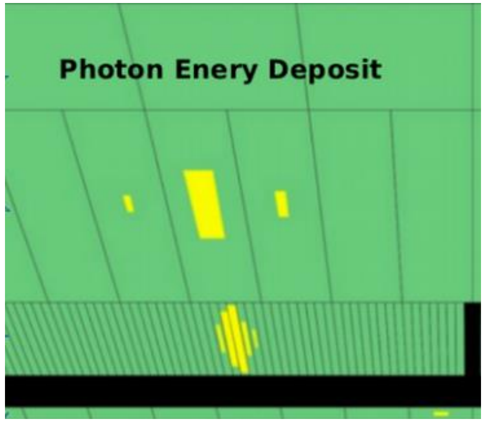
Preselections	Cut value
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 130 GeV
E_{T}^{γ}	> 150 GeV
Number of photons	$N_{\gamma} = 1$
Lepton veto	$N_e = 0,\ N_{\mu} = 0$

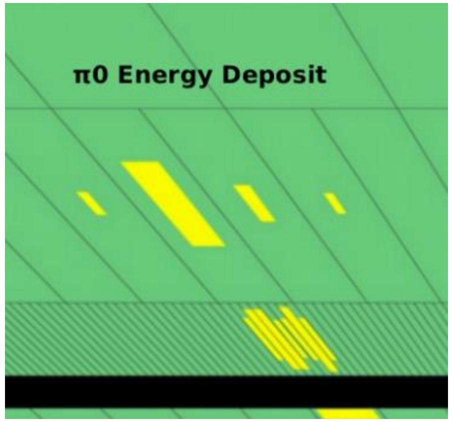
Selections

Selections	Cut value
$E_{\mathrm{T}}^{\mathrm{miss}}$ significance	> 11
$\lvert\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, \gamma)\rvert$	> 0.7
$\lvert\Delta\phi(E_{\mathrm{T}}^{\mathrm{miss}}, j_1)\rvert$	> 0.4

jet → γ background

- The background induced by the misidentification of a jet as a photon is studied in this analysis.



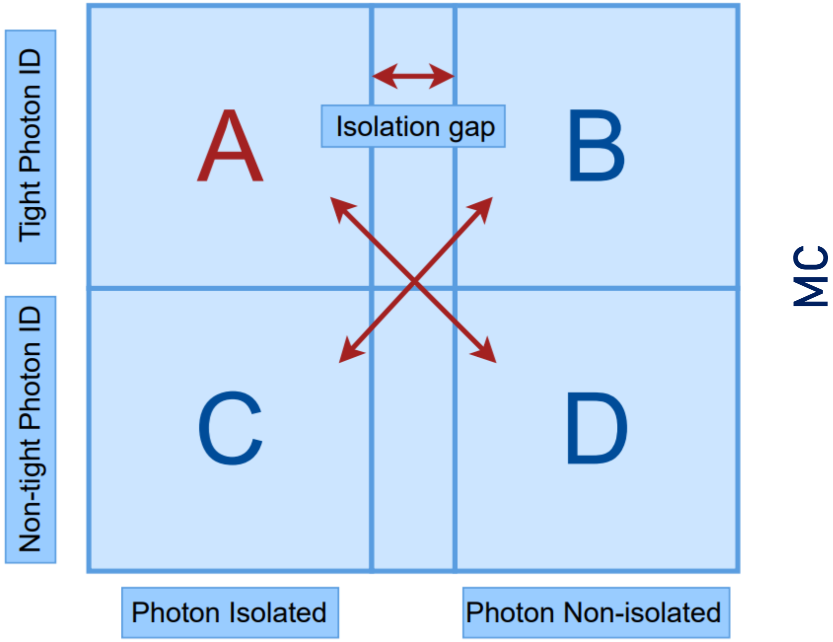


Hadronic jets in which neutral mesons carry a significant fraction of energy may be misidentified as isolated photons.

⇨ the SR will be contaminated with $jet \rightarrow \gamma$

ABCD method for $jet \rightarrow \gamma$:

- the phase space is splitted into 4 regions based on the identification (*tight* or *loose'*) and isolation (*isolated* or *non-isolated*) criteria for photons;
- the main assumption is the absence of correlation between identification and isolation criteria.



The estimate of $jet \rightarrow \gamma$ events in signal region A derived by ABCD method is $2100 \pm 100 \pm 300$

⇨ But using ABCD method we cannot get the shape of the distribution. Thus, we have a motivation to estimate $jet \rightarrow \gamma$ with other methods.

More details in back-up

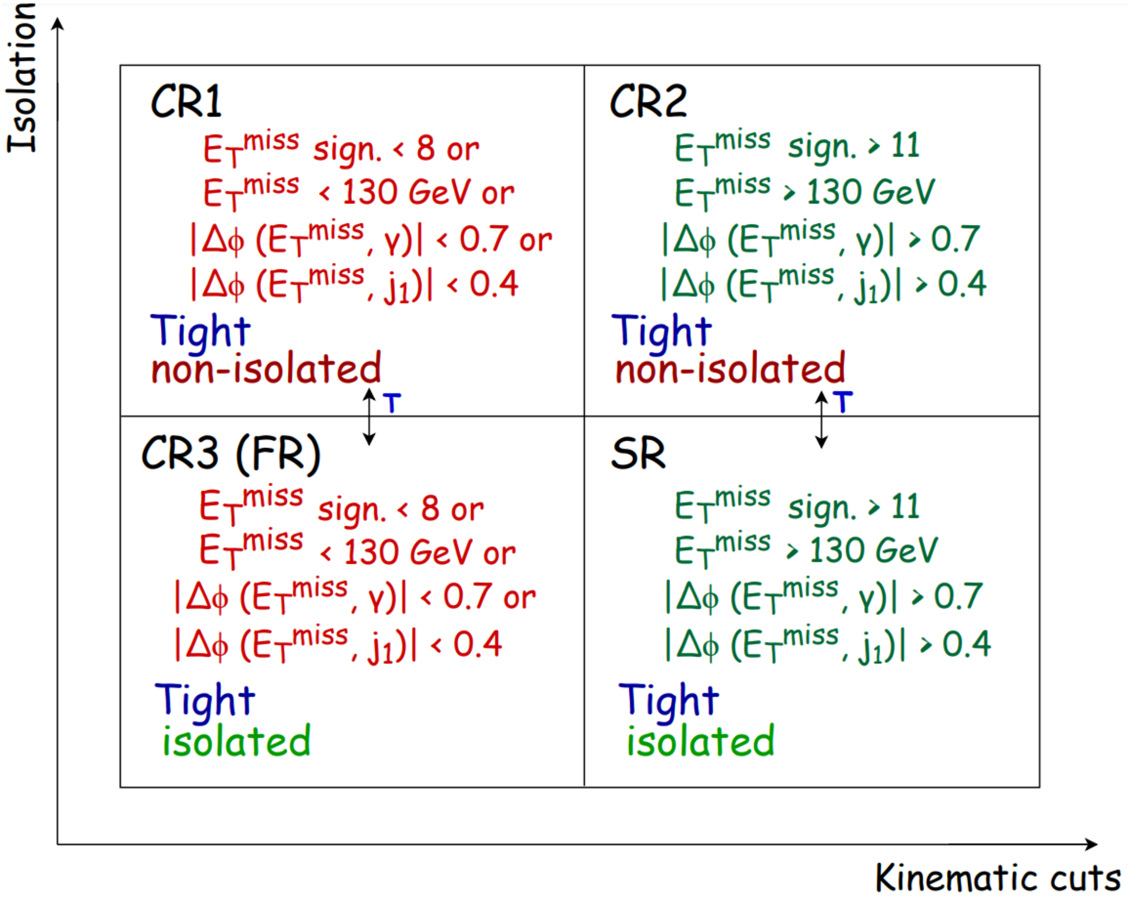
Estimation techniques of the slice method I

Strategy:

1. To split the phase space into 4 orthogonal regions based on kinematic cuts and isolation. The fit region (FR) is the kinematically inverted signal region (SR). Events in the FR have a leading photon candidate that is isolated. Events in the SR pass all signal kinematic selections.

2. The CR2 is a region, where events have a leading photon candidate that is not isolated. Events in the CR2 pass all signal kinematic selections. The CR1 is the kinematically inverted CR2.

3. Photons in all four regions pass the tight selection criteria.



4. The normalized fit is performed in the FR, where the $jet \rightarrow \gamma$ process used for the fit is derived from CR1.

5. Photon is required to pass $p_T^{cone20}/p_T^{\gamma} < 0.05$ track isolation in isolated regions. To increase the statistics in non-isolated regions the inverted track isolation $p_T^{cone20}/p_T^{\gamma} > 0.05$ is applied.

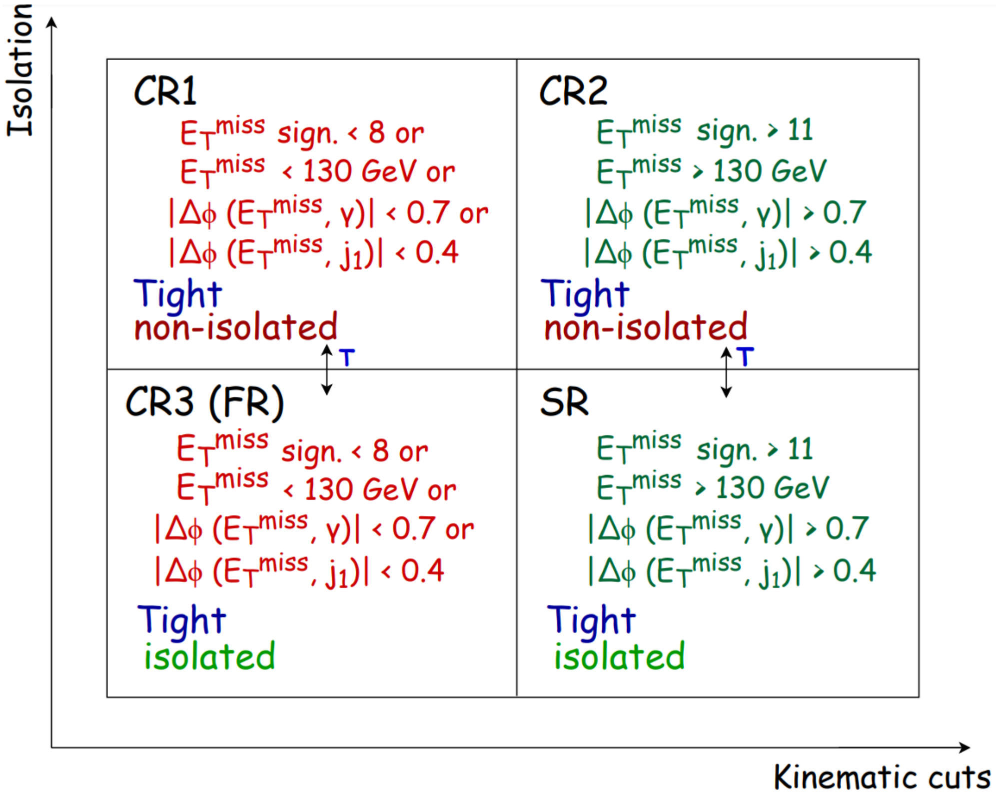
Estimation techniques of the slice method II

6. The normalized fit can be performed for different variables in the phase-space region with inverted cuts on these variables.

7. To study the dependence of the result on the isolation criteria, control regions CR1 and CR2 are split into successive intervals by the isolation variable, instead of a single integrated anti-isolated region.

8. In this way, the number of $jet \to \gamma$ background events for a given isolation slice i can be estimated as follows:

$$N^{jet\to\gamma}_{\mathrm{CR1\,(i)}} = N^{\mathrm{data}}_{\mathrm{CR1\,(i)}} - N^{\mathrm{Z}(\nu\bar{\nu})\gamma}_{\mathrm{CR1\,(i)}} - N^{\mathrm{bkg}}_{\mathrm{CR1\,(i)}}$$



9. The normalized fit is performed in the FR. Thus, the total number of events in the FR estimated from non-isolated slice of the CR1 is given by:

$$N^{\mathrm{data}}_{\mathrm{FR\,(i)}} = \alpha \cdot (N^{\mathrm{Z}(\nu\bar{\nu})\gamma}_{\mathrm{FR\,(i)}} + N^{\mathrm{bkg}}_{\mathrm{FR\,(i)}}) + N^{jet\to\gamma}_{\mathrm{FR\,(i)}}$$

10. The fitting parameter $T_{\mathrm{(i)}}$ gives the estimated number of $jet \to \gamma$ events in the FR: $N^{jet\to\gamma}_{\mathrm{FR\,(i)}} \approx T_{(i)} \cdot N^{jet\to\gamma}_{\mathrm{CR1\,(i)}}$

Estimation techniques of the slice method III

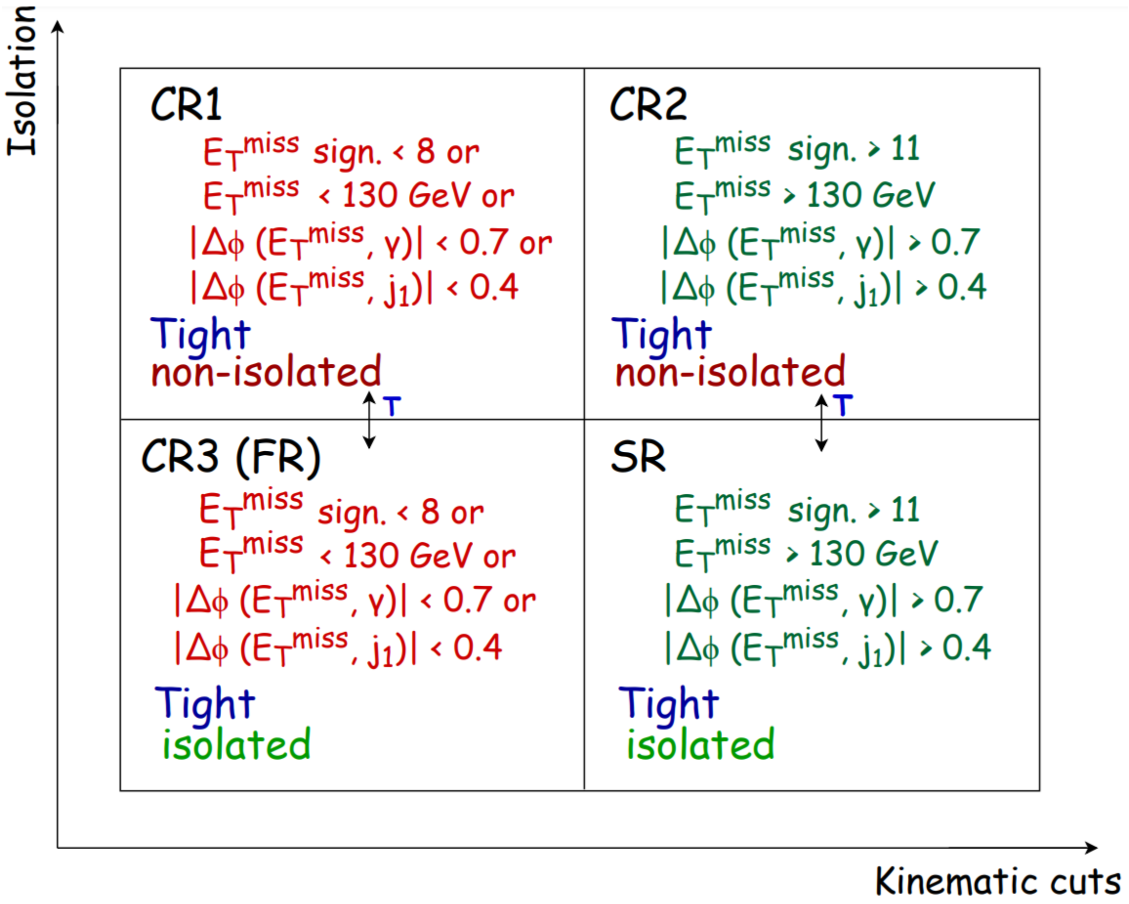
11. In this study, a parameter α is taken to be equal to 1. The fit parameter $T_{(i)}$ is derived for each slice and kinematic variable.

12. Finally, the fitted *jet* $\rightarrow \gamma$ yield is extrapolated to the SR. The estimate for each slice and kinematic variable is determined by the equation:

$$N^{jet\rightarrow\gamma}_{\mathrm{SR(i)}} = T_{(i)} \cdot (N^{\mathrm{data}}_{\mathrm{CR2(i)}} - N^{\mathrm{Z}(\nu\bar{\nu})\gamma}_{\mathrm{CR2(i)}} - N^{\mathrm{bkg}}_{\mathrm{CR2(i)}})$$

- FixedCutLoose isolation working point is chosen. Isolation working point is defined as:

$$E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}}^{\gamma} < 0.065$$

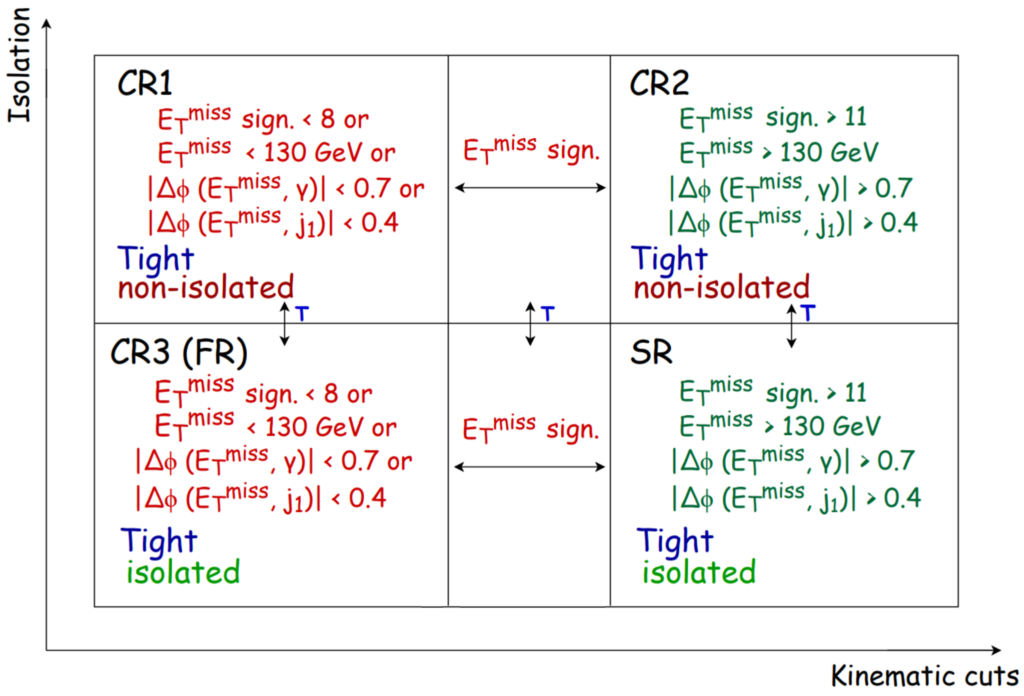


⇨ Five isolation slices are chosen: [0.065, 0.08, 0.095, 0.115, 0.14]

T factor estimation via MC and data

T factor in data in the Etmiss signif. gap:

Slice	Data	Bkg. (excl. $jet \to \gamma$)	Signal	$jet \to \gamma$	T_i
0	6599 ± 80	4384 ± 39	1005 ± 3	1210 ± 70	-
1	379 ± 19	8.5 ± 1.2	4.90 ± 0.17	366 ± 19	3.3 ± 0.3
2	281 ± 71	5.5 ± 1.0	4.17 ± 0.14	271 ± 17	4.5 ± 0.4
3	273 ± 17	7.4 ± 1.6	4.54 ± 0.15	261 ± 17	4.6 ± 0.4
4	201 ± 14	5.4 ± 0.6	4.35 ± 0.14	191 ± 14	6.3 ± 0.6
5	271 ± 16	24.3 ± 1.6	20.8 ± 0.3	256 ± 16	4.7 ± 0.4



T factor in MC in the Etmiss signif. gap:

Slice	$jet \to \gamma$	T_i
0	268 ± 89	-
1	98 ± 48	2.7 ± 1.6
2	88 ± 48	3.0 ± 1.9
3	101 ± 48	2.7 ± 1.5
4	49 ± 34	5 ± 4
5	77 ± 31	3.5 ± 1.8

T factor in data in the FR and CR1:

Slice	Data	Bkg. (excl. $jet \to \gamma$)	Signal	$jet \to \gamma$	T_i
0	60391 ± 246	45293 ± 128	2072 ± 3	13027 ± 210	-
1	3730 ± 61	55 ± 5	5.70 ± 0.15	3669 ± 61	3.55 ± 0.08
2	3158 ± 56	34 ± 3	4.93 ± 0.13	3119 ± 56	4.18 ± 0.10
3	3083 ± 56	55 ± 4	5.81 ± 0.14	3022 ± 56	4.31 ± 0.11
4	2492 ± 50	30 ± 3	5.77 ± 0.15	2456 ± 50	5.30 ± 0.14
5	6930 ± 83	169 ± 6	40.3 ± 0.4	6721 ± 83	1.94 ± 0.04

- Events in the slice 0 are events in the FR

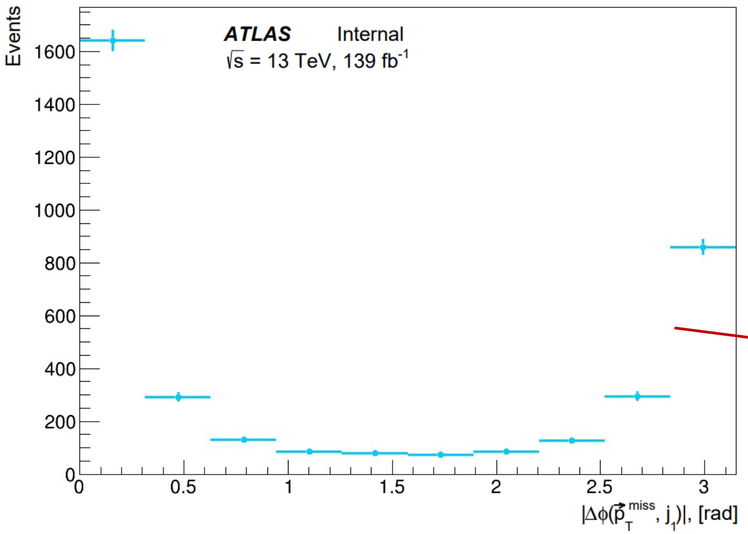
Normalization process

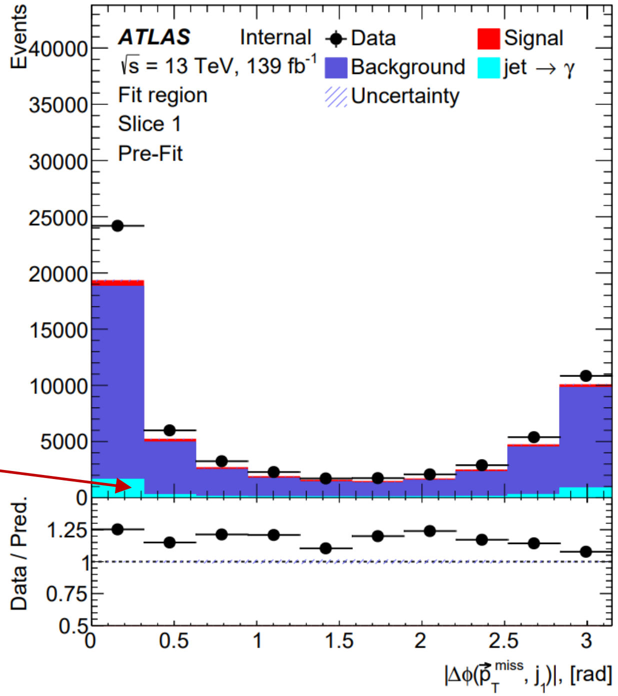
- The normalization was performed for 4 variables: E_T^{miss}, E_T^{miss} significance, $|\Delta\phi(\gamma, \vec{p}_T^{\,miss})|$ and $|\Delta\phi(j_1, \vec{p}_T^{\,miss})|$
- The normalization parameter T is derived for each slice and variable.

Strategy:

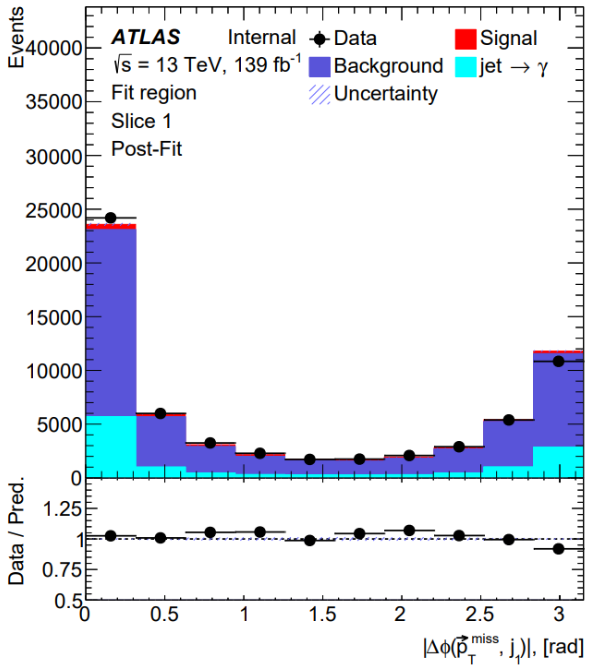
1) To derive the distribution of the *jet* $\rightarrow \gamma$ in the CR1 from [data – other bkg. – signal] in this region;
2) To add derived *jet* $\rightarrow \gamma$ distribution to the FR. Thus, in the FR we have data, signal and other bkg., that are derived in FR, and *jet* $\rightarrow \gamma$, which is derived in the CR1
3) To perform the normalization

jet $\rightarrow \gamma$ in the CR1:





Normalization



To derive the normalization parameters T for each slice

Events in the regions

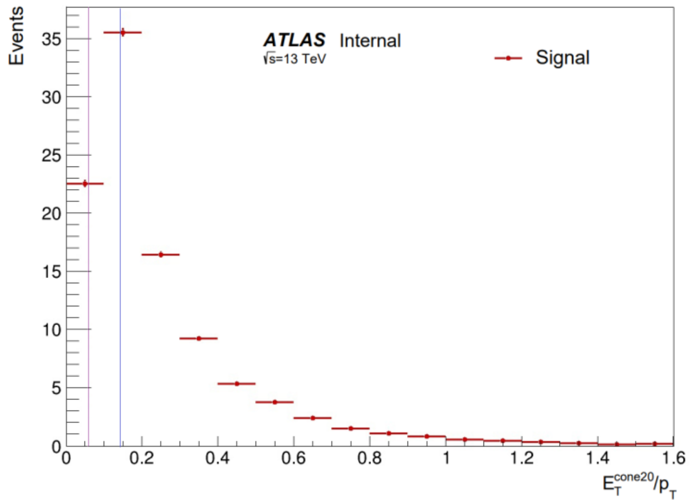
- Events in the FR and in the CR1 for each slice:

Slice	Data	Bkg. (excl. $jet \to \gamma$)	Signal	$jet \to \gamma$	T_i
0	60391 ± 246	45293 ± 128	2072 ± 3	13027 ± 210	-
1	3730 ± 61	55 ± 5	5.70 ± 0.15	3669 ± 61	3.55 ± 0.08
2	3158 ± 56	34 ± 3	4.93 ± 0.13	3119 ± 56	4.18 ± 0.10
3	3083 ± 56	55 ± 4	5.81 ± 0.14	3022 ± 56	4.31 ± 0.11
4	2492 ± 50	30 ± 3	5.77 ± 0.15	2456 ± 50	5.30 ± 0.14
5	6930 ± 83	169 ± 6	40.3 ± 0.4	6721 ± 83	1.94 ± 0.04

"Theoretical" T (w/o normalization)

- Events the CR2 for each slice:

Slice	Data	Bkg. (excl. $jet \to \gamma$)	Signal	$jet \to \gamma$
1	463 ± 22	8.4 ± 0.9	10.1 ± 0.3	444 ± 22
2	337 ± 18	8 ± 3	9.3 ± 0.3	320 ± 19
3	286 ± 17	11.6 ± 0.9	9.7 ± 0.2	256 ± 17
4	223 ± 15	5.5 ± 0.9	10.9 ± 0.3	207 ± 15
5	471 ± 22	41 ± 2	67.0 ± 0.6	363 ± 22



- Events in the slice 0 are events in the FR

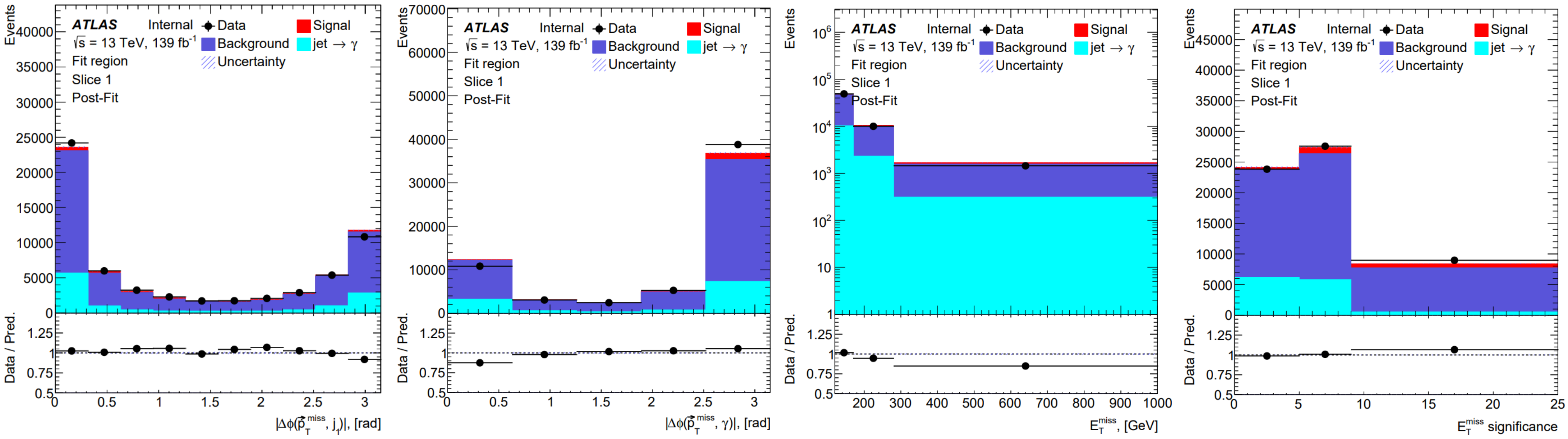
Normalized process

- The normalization was performed for 4 variables: $E_{\mathrm{T}}^{\mathrm{miss}}$, $E_{\mathrm{T}}^{\mathrm{miss}}$ significance, $|\Delta\phi(\gamma, \vec{p}_{\mathrm{T}}^{\;\mathrm{miss}})|$ and $|\Delta\phi(j_1, \vec{p}_{\mathrm{T}}^{\;\mathrm{miss}})|$
- The normalization parameter T is derived for each slice and variable.

The likelihood function:

$$\mathcal{L}(N_i^{\mathrm{data}}|T) = \prod_{i=1}^{N_{\mathrm{bins}}} \mathrm{Pois}(N_i^{\mathrm{data}}|N_i^{\mathrm{sig}} + N_i^{\mathrm{bkg}} + T \cdot N_i^{jet\to\gamma})$$

The results of the fit for slice 1 [0.065, 0.08]:



Pre-fits and post-fits for different variables are performed in back-up

The results of the normalization

Result of the fit for Zγ QCD Sherpa generator:

Slice	T_1, E_T^{miss}	T_2, E_T^{miss} sign.	T_3, $\|\Delta(E_T^{miss}, j_1)\|$	T_4, $\|\Delta(E_T^{miss}, \gamma)\|$
1	3.50 ± 0.08	3.42 ± 0.08	3.42 ± 0.08	3.33 ± 0.07
2	4.14 ± 0.09	3.94 ± 0.09	3.89 ± 0.09	3.76 ± 0.08
3	4.30 ± 0.10	4.04 ± 0.09	3.99 ± 0.09	3.82 ± 0.09
4	5.24 ± 0.12	4.97 ± 0.12	4.82 ± 0.11	4.48 ± 0.10
5	1.90 ± 0.04	1.77 ± 0.04	1.62 ± 0.04	1.44 ± 0.04

The fit parameters $T_{(i)}$

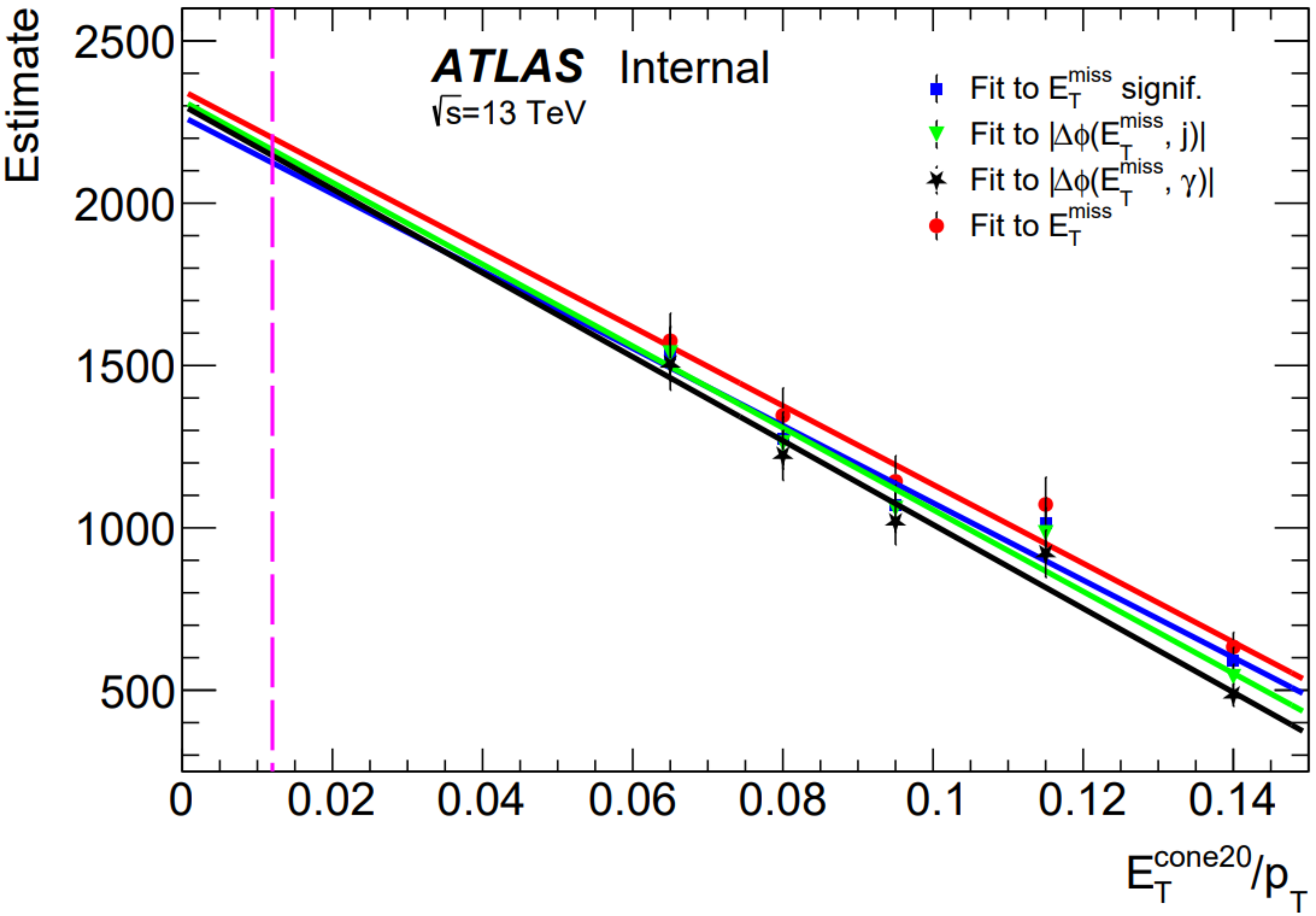
Slice	Observed $N_{CR2(i)}^{jet\to\gamma}$
1	444 ± 22
2	320 ± 19
3	265 ± 17
4	207 ± 15
5	363 ± 22

Observed *jet* → *γ* events in the CR2

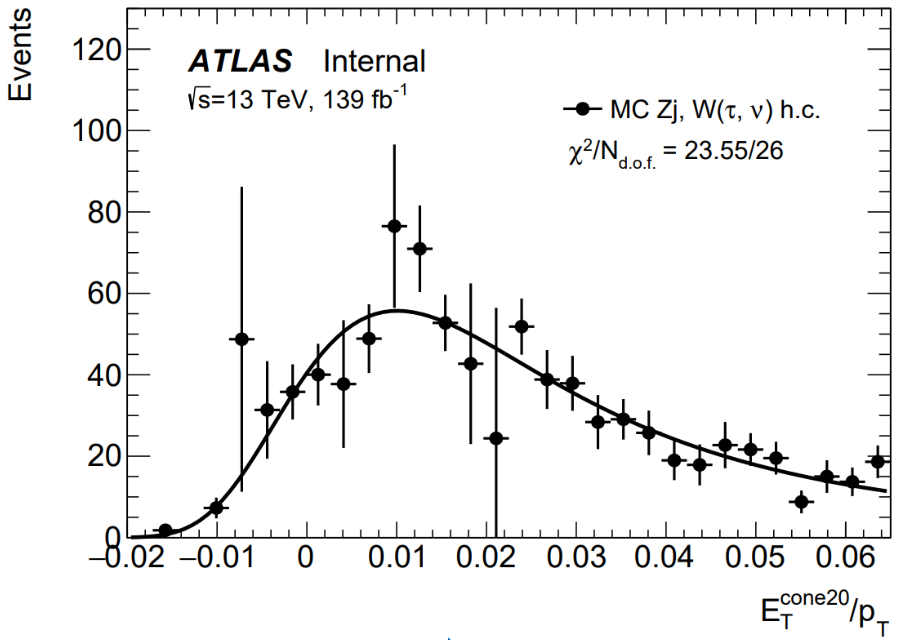
Slice	$N_{SR(i)}^{jet\to\gamma}$, E_T^{miss}	$N_{SR(i)}^{jet\to\gamma}$, E_T^{miss} sign.	$N_{SR(i)}^{jet\to\gamma}$, $\|\Delta(E_T^{miss}, j_1)\|$	$N_{SR(i)}^{jet\to\gamma}$, $\|\Delta(E_T^{miss}, \gamma)\|$
1	1555 ± 83	1518 ± 82	1521 ± 82	1484 ± 78
2	1323 ± 83	1258 ± 79	1242 ± 78	1201 ± 75
3	1137 ± 77	1068 ± 73	1056 ± 71	1010 ± 68
4	1084 ± 82	1027 ± 78	996 ± 75	926 ± 70
5	688 ± 44	643 ± 42	588 ± 38	524 ± 34

jet → *γ* events in the SR for each slice

Linear extrapolation



Isolation distribution for *jet* → *γ* MC samples



Landau fit ⇨ X = 0.012 ± 0.010

$\chi^2/N_{d.o.f.}$ for Gaus fit is 92.78/26

Variable	Estimate in x = 0.012
E_T^{miss}	2110 ± 120
E_T^{miss} sign.	2040 ± 120
$\lvert\Delta(E_T^{miss}, j_1)\rvert$	2080 ± 110
$\lvert\Delta(E_T^{miss}, \gamma)\rvert$	2070 ± 110

$$\overline{X} = \frac{\sum \frac{x_i}{\sigma_i^2}}{\sum \frac{1}{\sigma_i^2}}$$

- The estimate of *jet* → *γ* background is 2070 ± 60 events.

The estimate of *jet*→*γ* events in signal region A derived by ABCD method is $2100 \pm 100 \pm 300$

The sources of the systematics

Systematic uncertainties come from:

- The uncertainty in the choice of the extrapolation target for the isolation scan, estimated by changing the isolation target by ±1σ;
- The uncertainty comes from different generators.
- The uncertainty comes from the choice of the variable. (34 events)

Variable	Estimate in x = 0.002
$E_{\mathrm{T}}^{\mathrm{miss}}$	2220 ± 120
$E_{\mathrm{T}}^{\mathrm{miss}}$ sign.	2150 ± 120
$\lvert\Delta(E_{\mathrm{T}}^{\mathrm{miss}}, j_1)\rvert$	2200 ± 110
$\lvert\Delta(E_{\mathrm{T}}^{\mathrm{miss}}, \gamma)\rvert$	2190 ± 110

⇨ 2188 events, δ = 114 events

Variable	$Z\gamma$ QCD MadGraph
$E_{\mathrm{T}}^{\mathrm{miss}}$	2200 ± 120
$E_{\mathrm{T}}^{\mathrm{miss}}$ sign.	2130 ± 120
$\lvert\Delta(E_{\mathrm{T}}^{\mathrm{miss}}, j_1)\rvert$	2170 ± 120
$\lvert\Delta(E_{\mathrm{T}}^{\mathrm{miss}}, \gamma)\rvert$	2150 ± 110

⇨ 2159 events, δ = 85 events

- Total systematic uncertainty is 150 events.
- Thus, the estimate of $jet \to \gamma$ events in signal region A by slice method is 2070 ± 60 ± 150.
- The estimate of $jet \to \gamma$ events in signal region A derived by ABCD method is $2100 \pm 100 \pm 300$.
- The final estimates for different methods coincide within the uncertainty.

Likelihood-based approach I

- The main idea: to fit signal and other backgrounds distributions except *jet* $\rightarrow \gamma$ to data in all ABCD regions

The essence of the method is to perform a fit of the likelihood function, which is defined as:

$$L(N_{ji}|f_{F_{ji}}, f_{N_j}) = \prod_{j=A}^{B,C,D} \prod_{i=1}^{N_{bins}} \mathrm{Pois}(N_{ji}|\nu_{b_{ji}} + \nu_{\gamma_{ji}} f_{F_{ji}} + \nu_{s_{ji}} f_{N_j})$$

where model parameters are defined as:

- N_{ji} – the number of the data events in each region and bin;
- f_{N_j} – varying parameter for signal in each region;
- $f_{F_{ji}}$ – varying parameter for estimated background in each region and bin;
- $\nu_{b_{ji}}$ – the number of events in MC backgrounds (excl. *jet* $\rightarrow \gamma$);
- $\nu_{s_{ji}}$ – the number of signal events;
- $\nu_{\gamma_{ji}}$ – the number of estimated background (*jet* $\rightarrow \gamma$) events.

Likelihood-based approach II

- Likelihood based approach is constructed with the assumption that R = 1 for each bin in the distribution for $jet \to \gamma$ background:

$$1 = \frac{\nu_{\gamma Ai} f_{FAi} \cdot \nu_{\gamma Di} f_{FDi}}{\nu_{\gamma Bi} f_{FBi} \cdot \nu_{\gamma Ci} f_{FCi}}$$

- To avoid the redundancy of the model the following limitation is applied: $f_{F_{Bi}} = f_{F_{Di}}$

The search of maximum of likelihood function is performed with RooFit toolkit:

$$\frac{\partial L}{\partial f_{F_{ji}}} = 0, \quad \frac{\partial L}{\partial f_{N_j}} = 0$$

This way the number of $jet \to \gamma$ events in SR: $N_A^{jet \to \gamma} = \nu_{\gamma Ai} f_{FAi}$

- The proposed method significantly reduces the number of steps to be done to obtain the estimate compared to ABCD-method

MC samples

- The likelihood-based approach is applied to associated Zγ production with Z-boson decaying into neutrinos (Z → νν). One of the backgrounds comes from γ+j events. Zj events come from jet → γ misidentification
- The processes considered in the analysis are generated in MadGraph5 MC event generator using pp collisions with √s = 13 TeV and the integrated luminosity of 139 fb^{-1}
- Pythia8 is used for parton showering and hadronization, Delphes is used for detector simulation.

Selection	Cut value
E_T^{miss}	> 130 GeV
E_T^{γ}	> 150 GeV
Number of tight photons	$N_\gamma = 1$
Lepton veto	$N_e = 0, N_\mu = 0$

Event selection criteria for Zγ candidate events

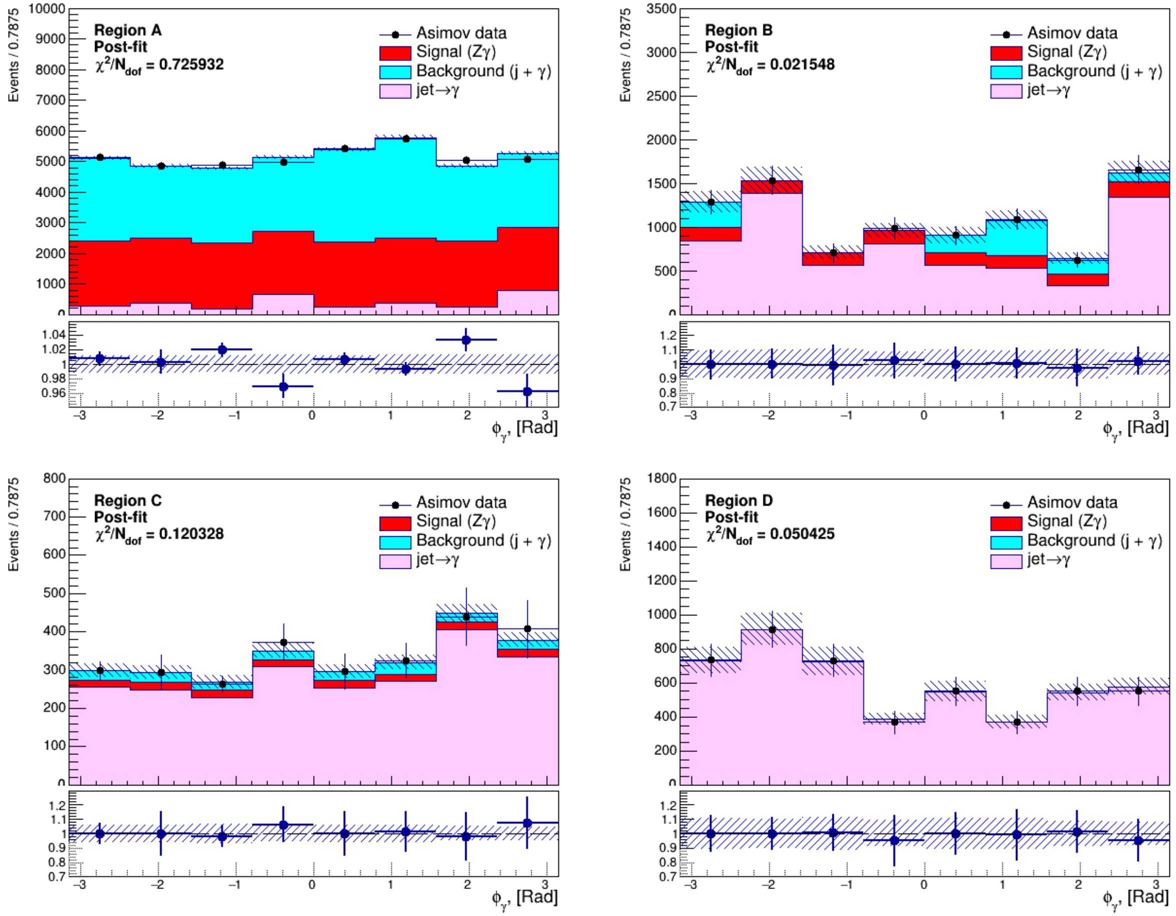
Thus the study uses Asimov data which is not real data but the sum of MC generated processes, the likelihood-based estimate of jet → γ background and MC prediction should coincide. It is so-called «closure test».

The results of the fit

The fit was performed for ϕ_γ and η_γ :

The final estimate is chosen based on the $\chi^2/N_{d.o.f.}$ value in the SR and R-factor.

N_{bins}	ϕ_γ			η_γ		
	Estimate	R-factor	$\chi^2/N_{d.o.f.}$	Estimate	R-factor	$\chi^2/N_{d.o.f.}$
6	3255^{+111}_{-106}	1.04 ± 0.03	0.45	3238^{+129}_{-125}	1.03 ± 0.03	0.39
7	2906^{+110}_{-108}	0.94 ± 0.03	0.73	3243^{+126}_{-122}	1.04 ± 0.02	0.55
8	3179^{+117}_{-108}	1.04 ± 0.03	0.73	3276^{+141}_{-137}	1.04 ± 0.02	0.26
9	3119^{+130}_{-127}	1.01 ± 0.03	0.62	3251^{+133}_{-130}	1.05 ± 0.02	0.50



- The systematic uncertainties were derived by variating the value of isolation gap by ±σ in non-isolated control regions.
- The estimate of jet → γ events in SR obtained by likelihood method is $N_A^{jet\to\gamma} = 3179^{+117}_{-108} \pm 69$ for ϕ_γ and $N_A^{jet\to\gamma} = 3243^{+126}_{-122} \pm 48$ for η_γ
- The MC prediction is $N_A^{jet\to\gamma} = 3093 \pm 178$ events

Summary

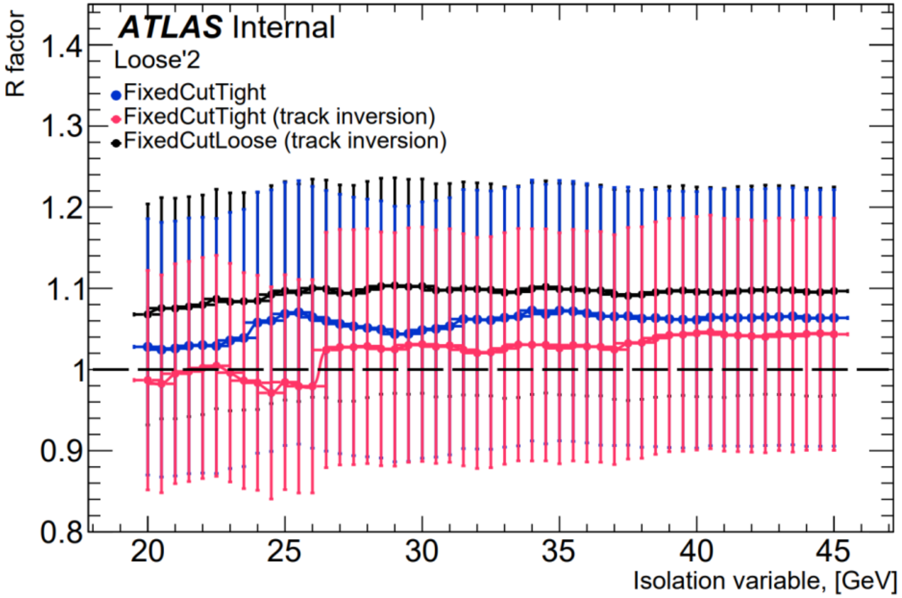
- The estimate of $jet \to \gamma$ events in signal region A is derived by ABCD method. The estimate is $2100 \pm 100 \pm 300$ events.
- The alternative slice method is performed for $jet \to \gamma$ estimation process. The estimate of $jet \to \gamma$ events in signal region A derived by slice method is 2070 ± 60 ± 150. The final estimates for the methods coincide within the uncertainty.
- The alternative likelihood-based method of estimation of $jet \to \gamma$ events was developed. It uses the information about the shape of the distributions in the regions and provides a much simpler way to obtain the estimate of the number of background events.

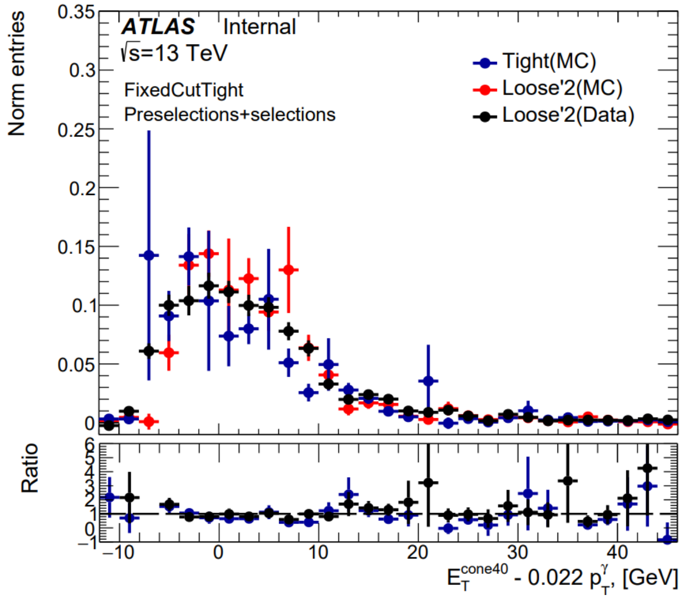
Thank you for your attention!

BACK-UP

R factor Zj and W(τν) in MC

Working point	loose'2	loose'3	loose'4	loose'5
FCTight	1.06 ± 0.16	1.15 ± 0.16	1.17 ± 0.15	1.30 ± 0.17
FCTight (inversion)	1.05 ± 0.14	1.21 ± 0.15	1.38 ± 0.16	1.65 ± 0.19
FCTCaloOnly	1.18 ± 0.13	1.31 ± 0.13	1.37 ± 0.13	1.54 ± 0.14
FCLoose	1.0 ± 0.2	1.0 ± 0.2	1.0 ± 0.2	1.3 ± 0.2
FCLoose (inversion)	1.11 ± 0.13	1.23 ± 0.12	1.34 ± 0.12	1.60 ± 0.13



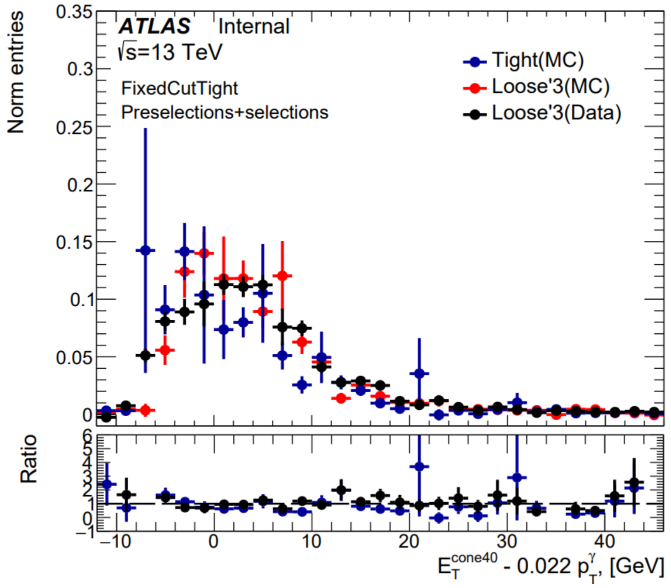


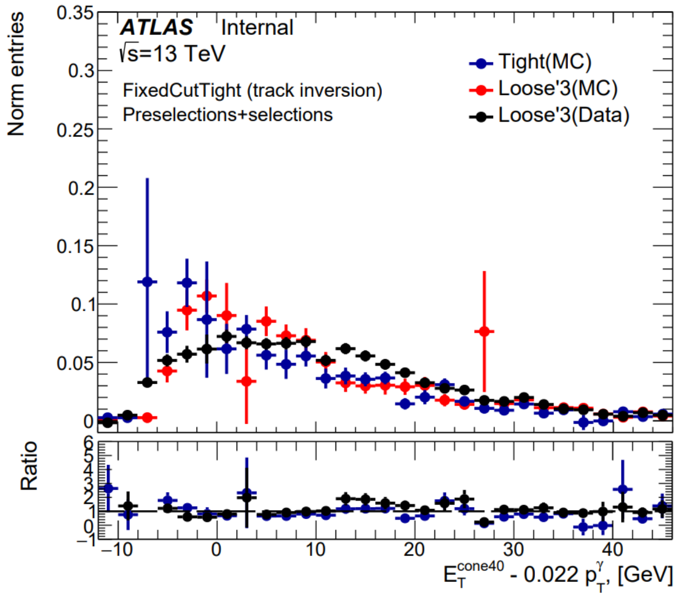
$$E_T^{\text{miss}} \text{ significance } = |\vec{E}_T^{\text{miss}}|^2 / (\sigma_L^2 (1 - \rho_{LT}^2))$$

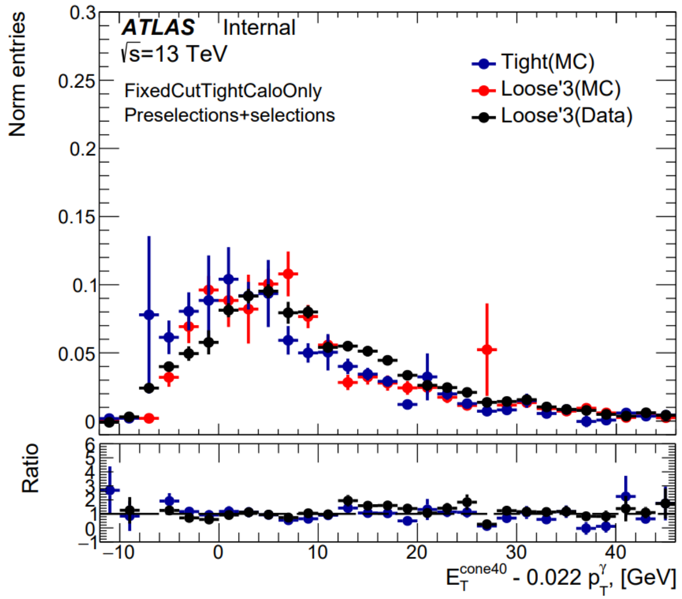
σ_L is the total variance in the longitudinal direction to the E_T^{miss}

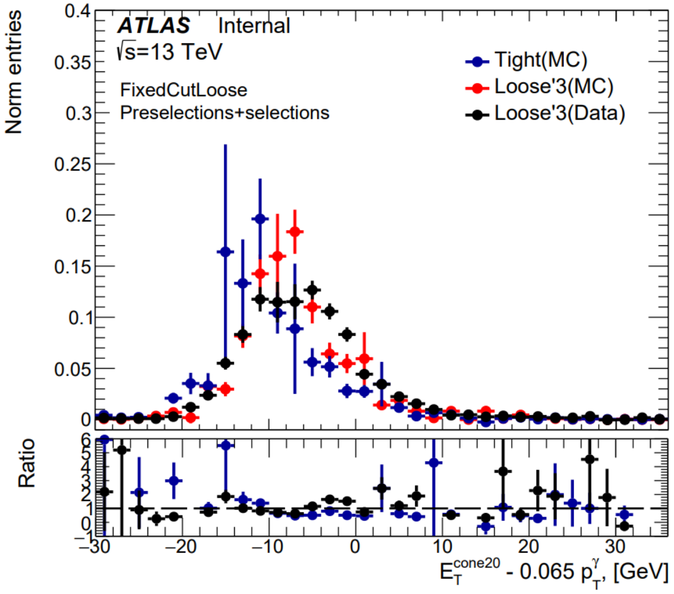
ρ_{LT} is the correlation factor of the longitudinal L and transverse T measurement

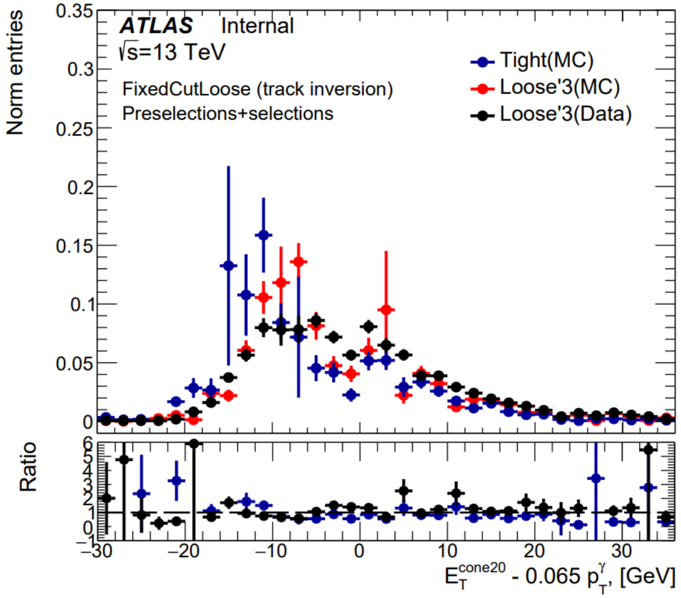
Isolation distributions (loose'3)

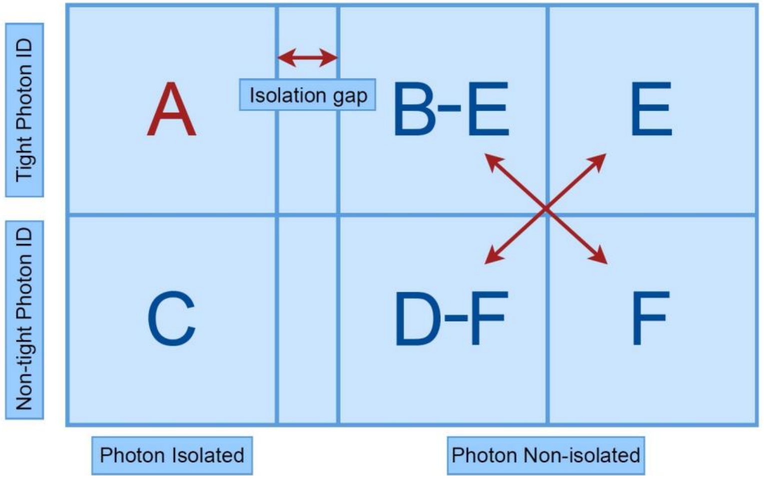












R factor in data

FixedCutLoose (inverted), w/o upper cut

MC

	loose'2	loose'3	loose'4	loose'5
R-factor	1.11 ± 0.13	1.23 ± 0.12	1.34 ± 0.12	1.60 ± 0.13

Data-driven

Cut	loose'2	loose'3	loose'4	loose'5
4.5	0.97 ± 0.10	1.05 ± 0.10	1.05 ± 0.09	1.06 ± 0.08
4.6	1.00 ± 0.10	1.08 ± 0.10	1.06 ± 0.09	1.07 ± 0.08
4.75	1.03 ± 0.10	1.05 ± 0.10	1.07 ± 0.09	1.09 ± 0.08
9.5	1.04 ± 0.09	1.03 ± 0.08	0.98 ± 0.07	0.97 ± 0.07
10.0	1.04 ± 0.09	1.03 ± 0.08	0.98 ± 0.07	0.98 ± 0.07
10.5	1.02 ± 0.09	1.02 ± 0.08	0.95 ± 0.07	0.96 ± 0.07
11.0	1.06 ± 0.09	1.02 ± 0.08	0.97 ± 0.07	0.96 ± 0.07

FixedCutTightCaloOnly, w/o upper cut

MC

	loose'2	loose'3	loose'4	loose'5
R-factor	1.18 ± 0.13	1.31 ± 0.13	1.37 ± 0.13	1.54 ± 0.14

Data-driven

Cut	loose'2	loose'3	loose'4	loose'5
9.45	1.15 ± 0.07	1.21 ± 0.06	1.20 ± 0.06	1.23 ± 0.06
9.95	1.14 ± 0.06	1.20 ± 0.06	1.19 ± 0.06	1.22 ± 0.06
10.45	1.15 ± 0.06	1.20 ± 0.06	1.19 ± 0.05	1.21 ± 0.05
10.45	1.21 ± 0.07	1.26 ± 0.06	1.24 ± 0.06	1.26 ± 0.06

FixedCutTight (inverted), w/o upper cut

MC

	loose'2	loose'3	loose'4	loose'5
R-factor	1.05 ± 0.14	1.21 ± 0.15	1.38 ± 0.16	1.65 ± 0.19

Data-driven

Cut	loose'2	loose'3	loose'4	loose'5
9.45	1.10 ± 0.08	1.15 ± 0.07	1.11 ± 0.06	1.16 ± 0.06
9.95	1.09 ± 0.07	1.15 ± 0.07	1.12 ± 0.06	1.16 ± 0.06
10.20	1.08 ± 0.07	1.14 ± 0.07	1.11 ± 0.06	1.15 ± 0.06
10.45	1.10 ± 0.07	1.15 ± 0.07	1.13 ± 0.06	1.17 ± 0.06

FixedCutTight, upper cut = 25.45 GeV

MC

	loose'2	loose'3	loose'4	loose'5
R-factor	1.07 ± 0.16	1.17 ± 0.17	1.18 ± 0.16	1.31 ± 0.17

Data-driven

Cut	loose'2	loose'3	loose'4	loose'5
8.45	1.15 ± 0.13	1.16 ± 0.12	1.16 ± 0.11	1.21 ± 0.11
8.95	1.11 ± 0.13	1.11 ± 0.12	1.14 ± 0.11	1.17 ± 0.11
9.45	1.19 ± 0.14	1.22 ± 0.13	1.27 ± 0.13	1.30 ± 0.12
9.95	1.16 ± 0.14	1.17 ± 0.13	1.23 ± 0.12	1.28 ± 0.12
10.45	1.19 ± 0.14	1.20 ± 0.14	1.22 ± 0.12	1.26 ± 0.12

FixedCutLoose was chosen. In order to decrease syst. uncert. the loose'3 was chosen

jet → γ background estimation (loose'3)

Event yields for the data and non-jet → γ background processes considered in the ABCD method

	Data	Wγ QCD	Wγ EWK	e → γ	$tt\gamma$	γ + jet	$Z(ll)\gamma$
A	26523 ± 163	3936 ± 23	136.3 ± 0.7	3039 ± 12	234 ± 3	5262 ± 53	285 ± 5
B	1475 ± 38	52 ± 4	1.86 ± 0.08	8.95 ± 0.03	1.3 ± 0.2	0.6 ± 0.4	1.0 ± 0.6
C	2568 ± 51	60 ± 2	2.16 ± 0.09	61.4 ± 0.2	4.2 ± 0.4	76 ± 6	4.8 ± 0.5
D	1443 ± 38	2.7 ± 0.6	0.17 ± 0.02	0.0715 ± 0.0002	0.35 ± 0.13	0 ± 0	0 ± 0

$$N_{\rm A}^{\rm sig} = \widetilde{N}_{\rm A} - R(\widetilde{N}_{\rm B} - c_{\rm B}N_{\rm A}^{\rm sig})\frac{\widetilde{N}_{\rm C} - c_{\rm C}N_{\rm A}^{\rm sig}}{\widetilde{N}_{\rm D} - c_{\rm D}N_{\rm A}^{\rm sig}}$$

$$N_{\rm A}^{\rm sig} = \frac{b - \sqrt{b^2 - 4ac}}{2a}$$

$$\begin{cases} a = c_D - Rc_Bc_C; \\ b = \widetilde{N}_{\rm D} + c_D\widetilde{N}_{\rm A} - R(c_B\widetilde{N}_{\rm C} + c_C\widetilde{N}_{\rm B}); \\ c = \widetilde{N}_{\rm D}\widetilde{N}_{\rm A} - R\widetilde{N}_{\rm C}\widetilde{N}_{\rm B}. \end{cases}$$

The signal leakage parameters:

(Isolation gap = 2GeV)	
c_B	0.00939 ± 0.00007
c_C	0.01536 ± 0.00010
c_D	0.00051 ± 0.00002

Event yields signal:

	$Z(\nu\bar{\nu})\gamma$ QCD	$Z(\nu\bar{\nu})\gamma$ EWK
A	10513 ± 8	152.1 ± 0.3
B	98.0 ± 0.8	2.14 ± 0.04
C	161.5 ± 1.0	2.31 ± 0.04
D	5.3 ± 0.2	0.135 ± 0.009

With R by data-driven ⇒

$$N_A^{jet\to\gamma} = \mathbf{2078^{+100}_{-97}}$$

Systematic uncertainty I

Systematic uncertainties come from:

- non-tight definition and isolation gap choice. Variation for ±1σ changes in data yield
- different generators
- imperfect photon iso/ID modeling

Different loose prime and isolation gap

Central value (with R_{data})	2078
loose'2	+327
loose'4	-111
loose'5	-173
Iso gap +0.25 GeV	+48
Iso gap −0.35 GeV	+29

$$R_{data}^{iso\ gap\ +0.25\ GeV} = 1.07 \pm 0.11$$

$$R_{data}^{iso\ gap\ -0.35\ GeV} = 1.06 \pm 0.09$$

Iso gap, GeV	N_B	N_D
-0.40	1524 ± 39	1488 ± 39
-0.35	1518 ± 39	1482 ± 38
-0.30	1513 ± 39	1477 ± 38
-0.25	1503 ± 39	1474 ± 38
-0.20	1497 ± 39	1468 ± 38
2.0	1475 ± 38	1443 ± 38
+0.15	1448 ± 38	1416 ± 38
+0.20	1443 ± 38	1404 ± 37
+0.25	1437 ± 38	1398 ± 37

⇒ δ = 16%

The choice of loose prime 3 reduced the systematic uncertainty from 32% to 16%

Systematic uncertainty II

Different generators:

Signal leakage parameters	Different generators		
	MadGraph+Pythia8, Sherpa 2.2	MadGraph+Pythia8, MadGraph+Pythia8	Relative deviation
c_B	0.00939 ± 0.00007	0.0155 ± 0.0004	39%
c_C	0.01536 ± 0.00010	0.0156 ± 0.0004	1.5%
c_D	0.00051 ± 0.000028	0.00077 ± 0.00009	34%
$jet \to \gamma$ est. (with R_{data})	2078	2061	0.8%

- Uncertainty coming from signal leakage is obtained δ = 0.8%

$$R_{data}^{diff.gen.} = 1.10 \pm 0.10$$

Systematic uncertainty come from imperfect photon iso/ID modeling:

- $\sigma_{iso}^{c_B} = \delta_{iso}^{eff} \cdot (c_B + 1)/c_B$
- $\sigma_{ID}^{c_C} = \delta_{ID}^{eff} \cdot (c_C + 1)/c_C$
- $\sigma_{iso}^{c_D} = \delta_{iso}^{eff} \cdot (c_B + 1)/c_B$
- $\sigma_{ID}^{c_D} = \delta_{ID}^{eff} \cdot (c_C + 1)/c_C$

δ^{eff}_{iso} = 0.013
$\delta^{eff}_{iso/ID}$ = 0.013

⇒ $\delta_{eff}^{iso/ID}$ = 1.3%

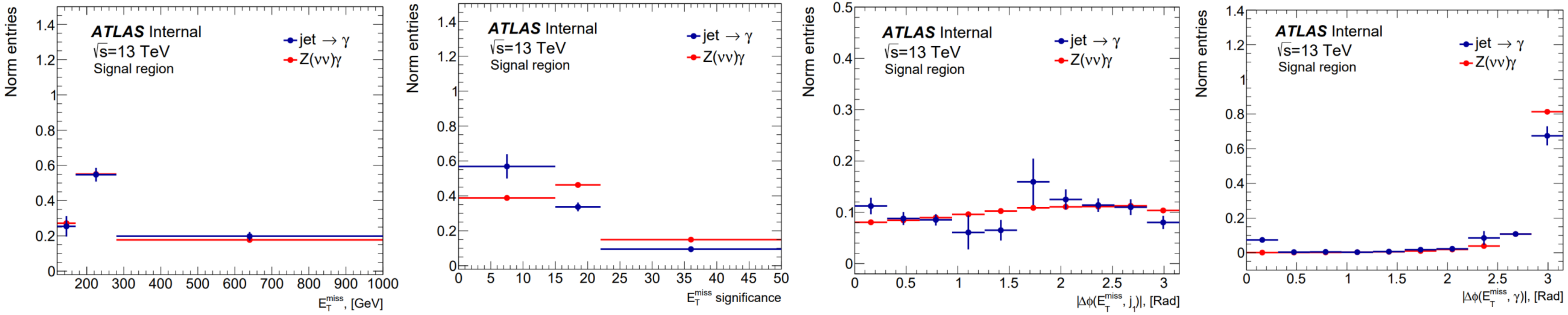
Estimate

with R_{Data}: 2100^{+100}_{-100}(stat.) $\pm$ 300(syst.)

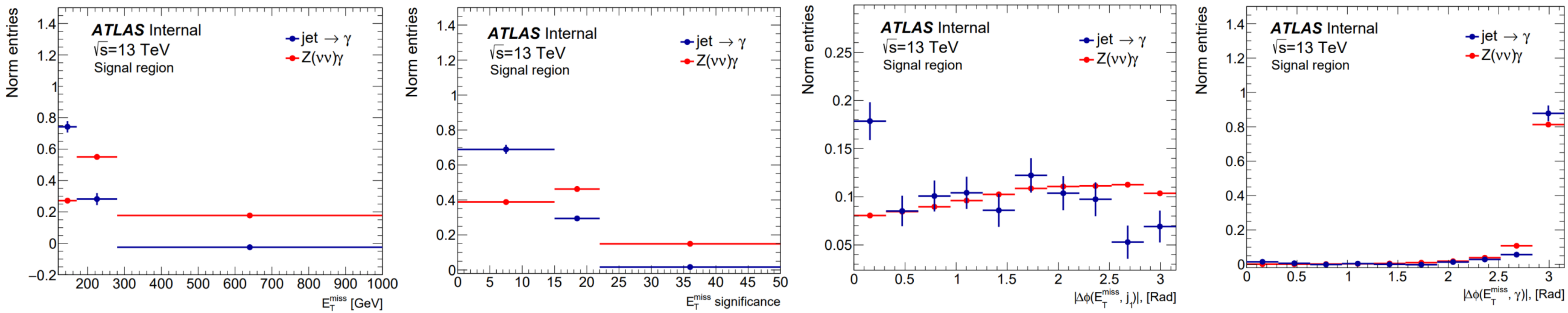
- Total systematics: δ_{Data} = 16%

jet $\rightarrow \gamma$ and Zγ comparison in the SR

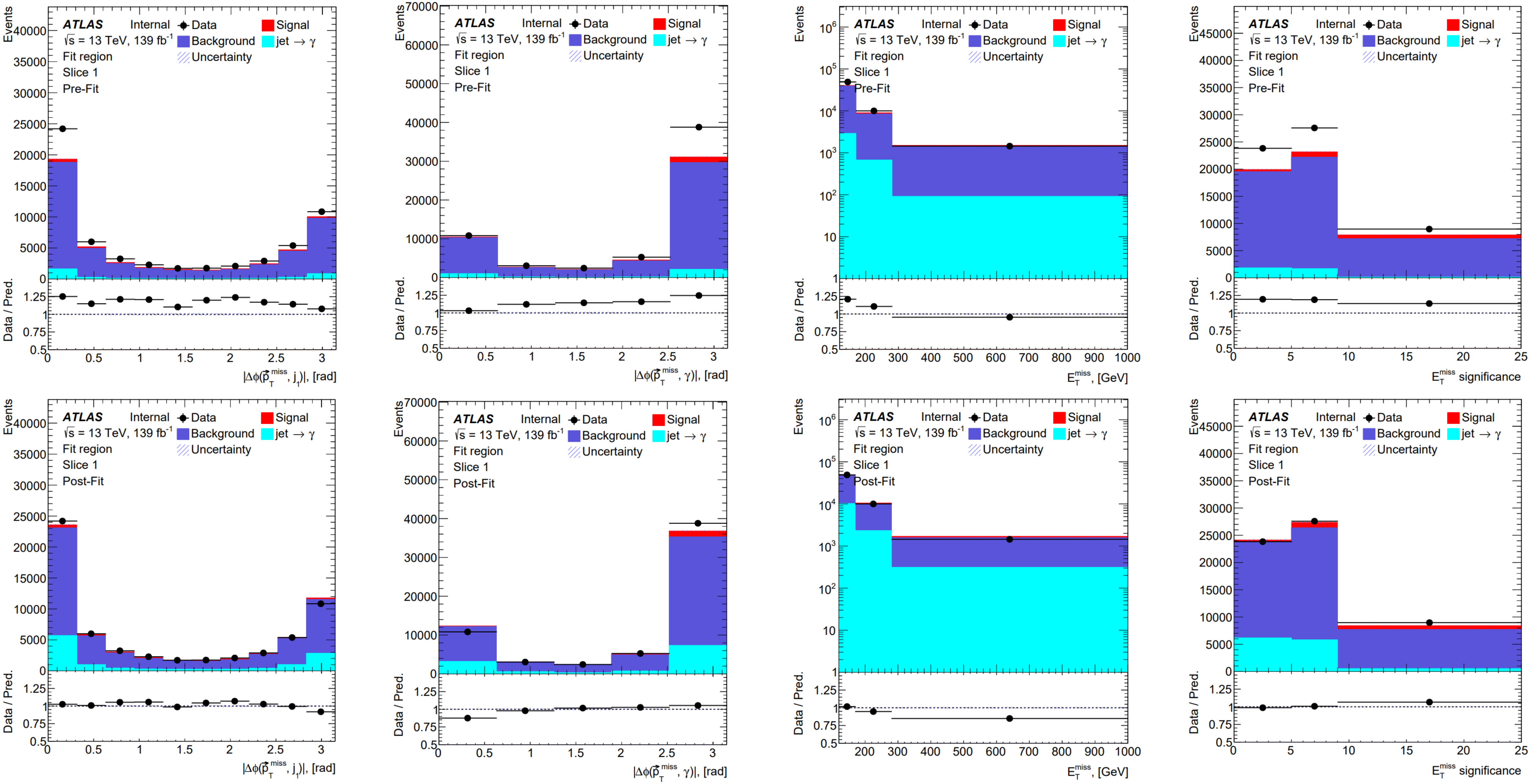
MC *jet* $\rightarrow \gamma$:



Observed *jet* $\rightarrow \gamma$:



Normalized fit. Slice 1



The results of the normalization

- Events in FR:

Data	Background (excl. $jet \to \gamma$)	Signal
60391 ± 246	45292 ± 128	2072 ± 3

- Events in CR1 for each slice:

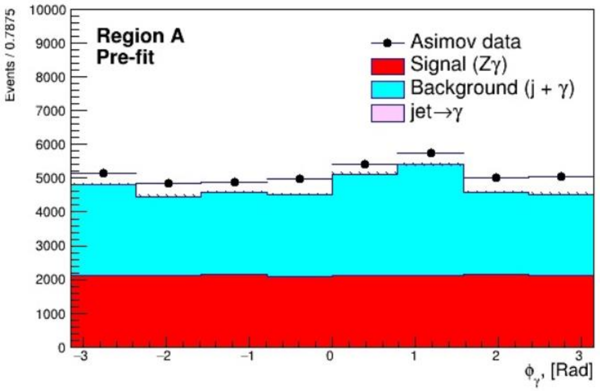
Slice	Data	Background (excl. $jet \to \gamma$)	Signal (Sherpa)	Signal (MadGraph)
1	3730 ± 61	55 ± 5	5.70 ± 0.15	3.9 ± 0.5
2	3158 ± 56	34 ± 3	4.93 ± 0.13	3.0 ± 0.4
3	3083 ± 56	55 ± 4	5.81 ± 0.14	4.3 ± 0.5
4	2492 ± 50	30 ± 3	5.77 ± 0.15	3.4 ± 0.4
5	6930 ± 83	169 ± 6	40.3 ± 0.4	27.2 ± 1.2

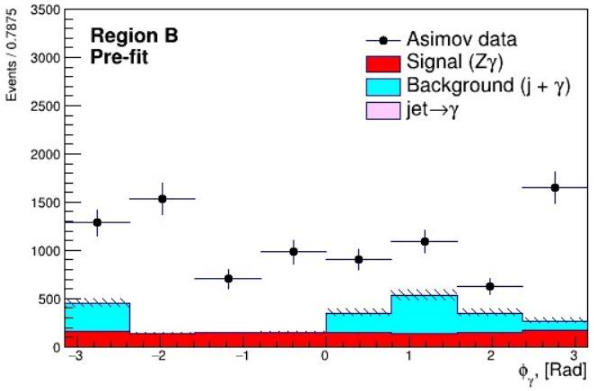
- Events in CR2 for each slice:

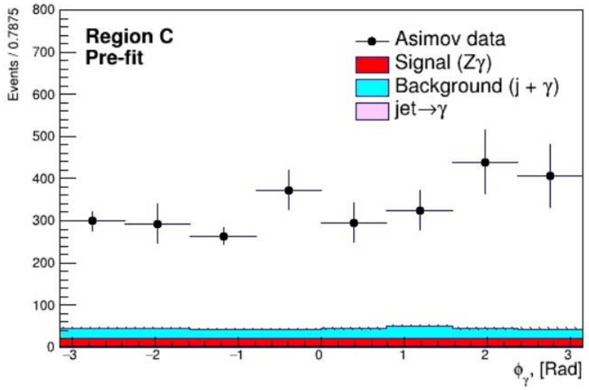
Slice	Data	Background (excl. $jet \to \gamma$)	Signal (Sherpa)	Signal (MadGraph)
1	463 ± 22	8.4 ± 0.9	10.1 ± 0.3	16.7 ± 1.3
2	337 ± 18	8 ± 3	9.3 ± 0.3	12.7 ± 1.2
3	286 ± 17	11.6 ± 0.9	9.7 ± 0.2	15.5 ± 1.3
4	223 ± 15	5.5 ± 0.9	10.9 ± 0.3	18.6 ± 1.3
5	471 ± 22	41 ± 2	67.0 ± 0.6	105 ± 3

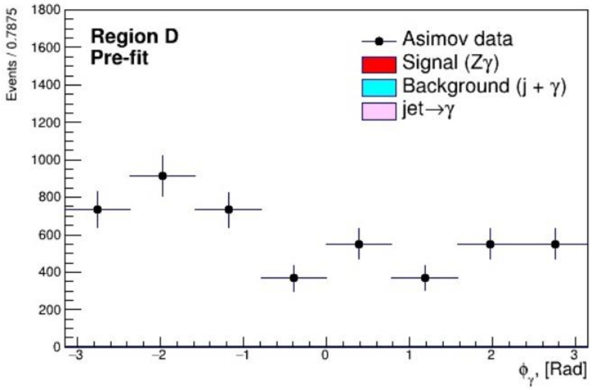
The results of the fit

Pre-fit for ϕ_γ

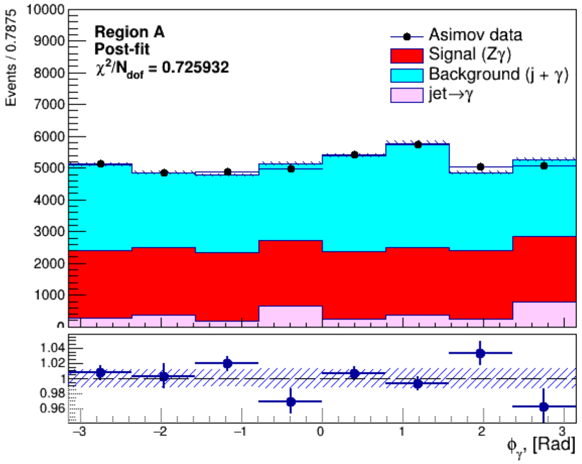


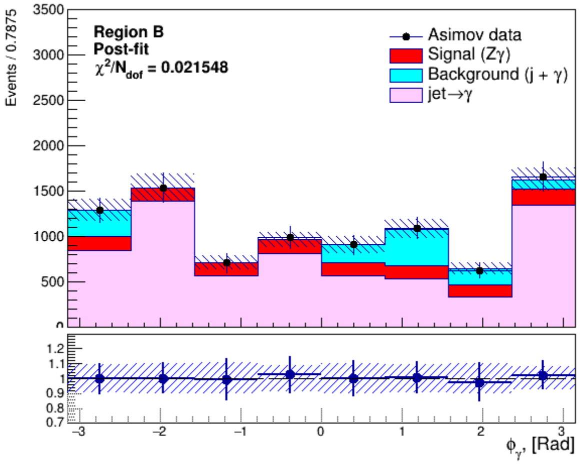


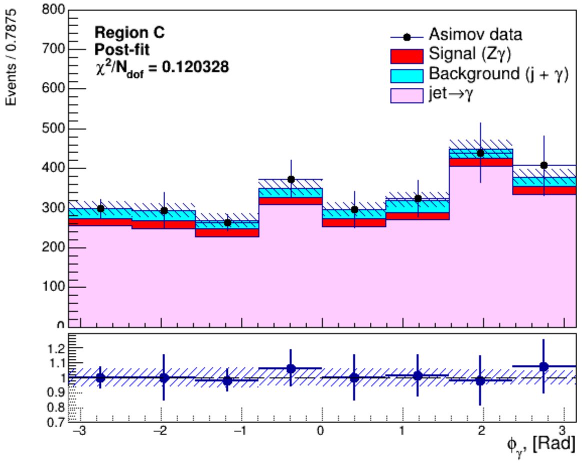


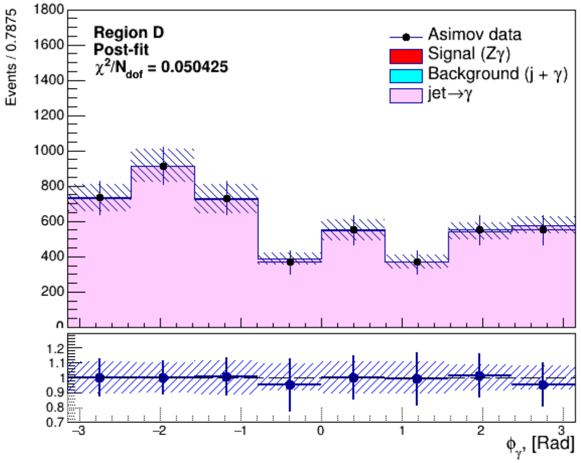


Post-fit for ϕ_γ



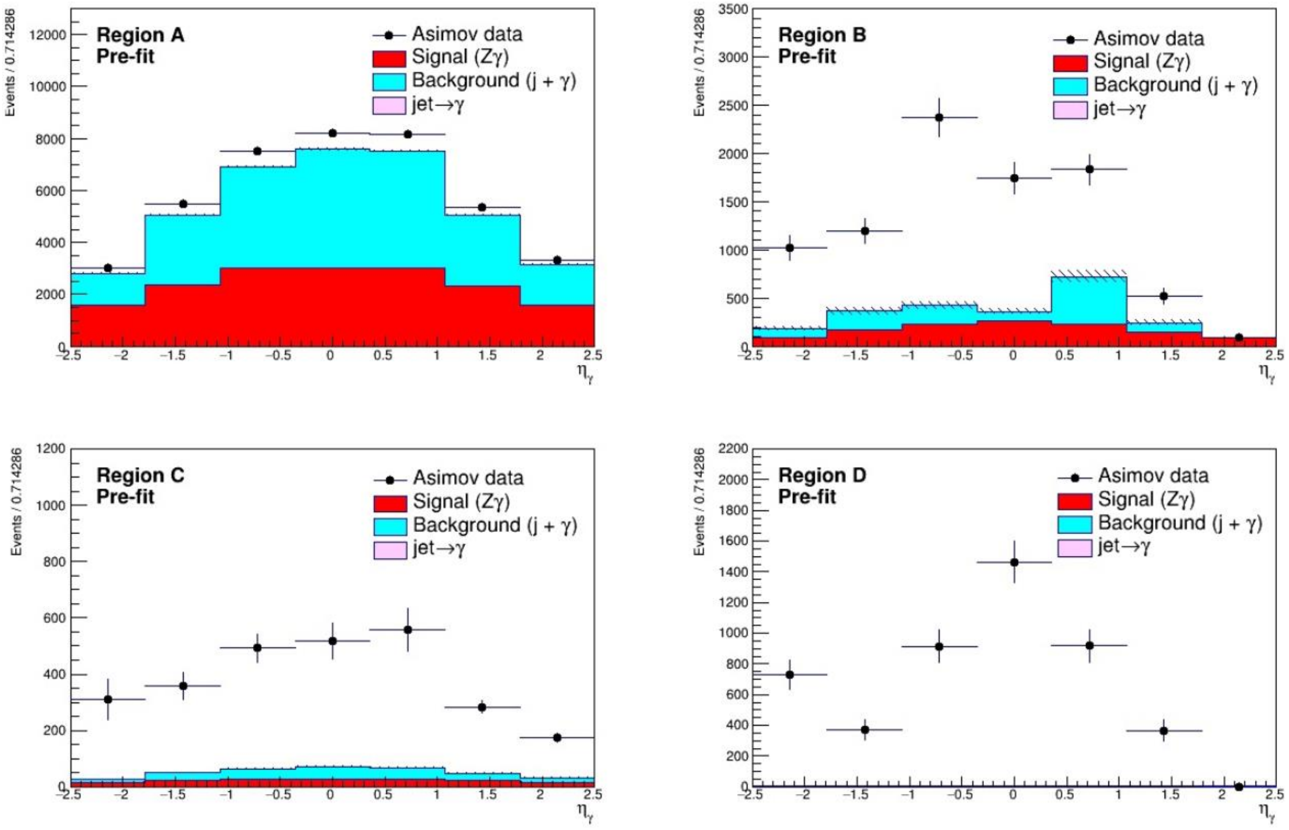






The results of the fit

Pre-fit for η_γ



Post-fit for η_γ