Limits on Physics beyond the Standard Model in the production of Z boson associated with a photon in the ATLAS experiment

Diana Pyatiizbyantseva

Supervisor: Soldatov E. Yu.

National Research Nuclear University “MEPhI”

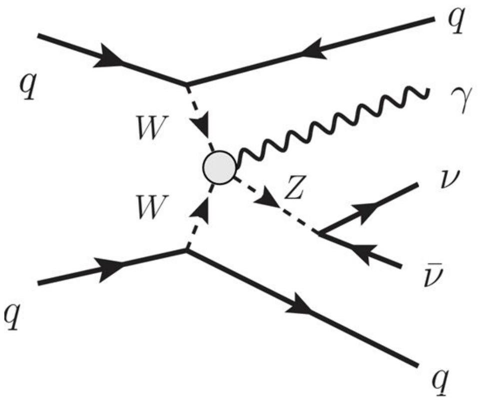
PhD Research Activities

June 28, 2021

Topology & background composition

Topology: $\gamma + p_T^{miss}$ + 2 hadronic jets. All objects are with high energy.

- The main goal of this study is to make observation or evidence of the Z(νν)γ EWK process for the first time.
- EWK Z(νν)γ production is the one of the most sensitive final states to aQGC.



Vector boson scattering

Background composition for Z(→νν)γ EWK:

Percentage (Zγ incl. region)	
48 %	Z(→νν)γ QCD — simultaneous fit to data (shape from MC)
30 %	W(→lν)γ — simultaneous fit to data (shape from MC)
6 %	ttγ — simultaneous fit to data (shape from MC)
7 %	e→γ – fake rate estimation using Z-peak (tag-n-probe) method
5 %	γ+jet – ABCD method based on E_T^{miss}-significance and soft term
2 %	jet→γ – ABCD method based on γ ID and isolation
0.7 %	$Z(\to l^+l^-)\gamma$ – via MC

Selections, signal and control regions

Photon selection:

E_T > 10 GeV, |η| < 2.37, crack region rejection, cluster quality cut, ambiguity cut, photon cleaning, Loose ID, ΔR(γ, e/μ) < 0.4

Electron selection:

p_T > 7 GeV, |η| < 2.47, crack region excluded, cluster quality cut, LooseBL ID, $|z_0 \sin\theta|$ < 0.5 mm, d_0-signif. < 5, isolation FCLoose, ΔR(e,μ) < 0.1

Muon selection:

p_T > 7 GeV, |η| < 2.47, away from bad calo region, Medium ID, $|z_0 \sin\theta|$ < 0.5 mm, d_0 signif. < 3, isolation FCLoose

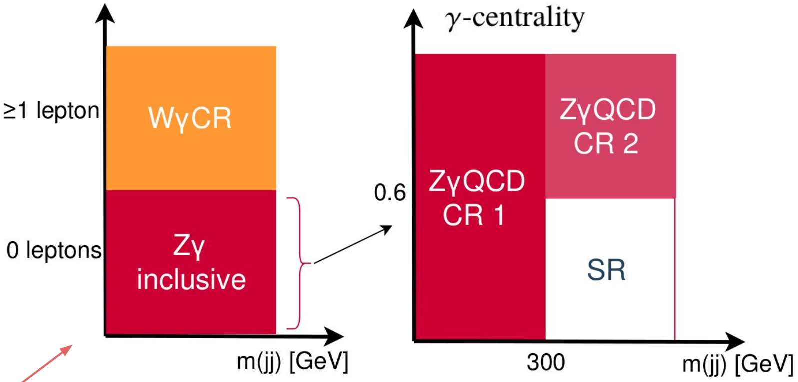
Jet selection:

E_T > 50 GeV, |η| < 4.5, AntiKt4EMTopoJets, ΔR(jet,e/μ/γ) < 0.4, JVT cut

Main event selections:

Selections	Cut Value
E_T^{miss}	> 120 GeV
E_T^{γ}	> 150 GeV
Number of tight isolated photons	$N_\gamma = 1$
Number of jets	$N_{jets} \geq 2$
Lepton veto	$N_e = 0,\ N_\mu = 0$
E_T^{miss} significance	> 12
$\lvert\Delta\phi(\gamma, \vec{p}_T^{\,miss})\rvert$	> 0.4
$\lvert\Delta\phi(j_1, \vec{p}_T^{\,miss})\rvert$	> 0.3
$\lvert\Delta\phi(j_2, \vec{p}_T^{\,miss})\rvert$	> 0.3
$p_T^{SoftTerm}$	< 16 GeV

4 regions: 1 SR and 3 CRs



γ-centrality: $$\zeta(\gamma) = \left| \frac{y(\gamma) - \frac{y(j_1)+y(j_2)}{2}}{y(j_1) - y(j_2)} \right|$$

$W\gamma$ control region	
$N_{leptons}$	≥ 1
$Z\gamma jj$ QCD control region 1	
$N_{leptons}$	$= 0$
m_{jj}	< 300 GeV
$Z\gamma jj$ QCD control region 2	
$N_{leptons}$	$= 0$
m_{jj}	> 300 GeV
γ-centrality	> 0.6
$Z\gamma jj$ EWK signal region	
$N_{leptons}$	$= 0$
m_{jj}	> 300 GeV
γ-centrality	< 0.6

Selection optimisation I: increasing statistical significance

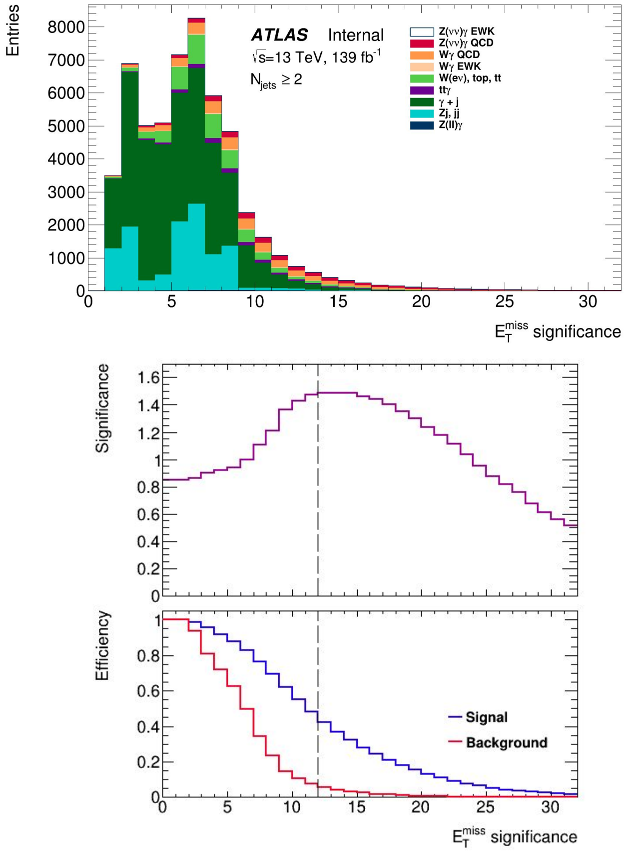
Preselection

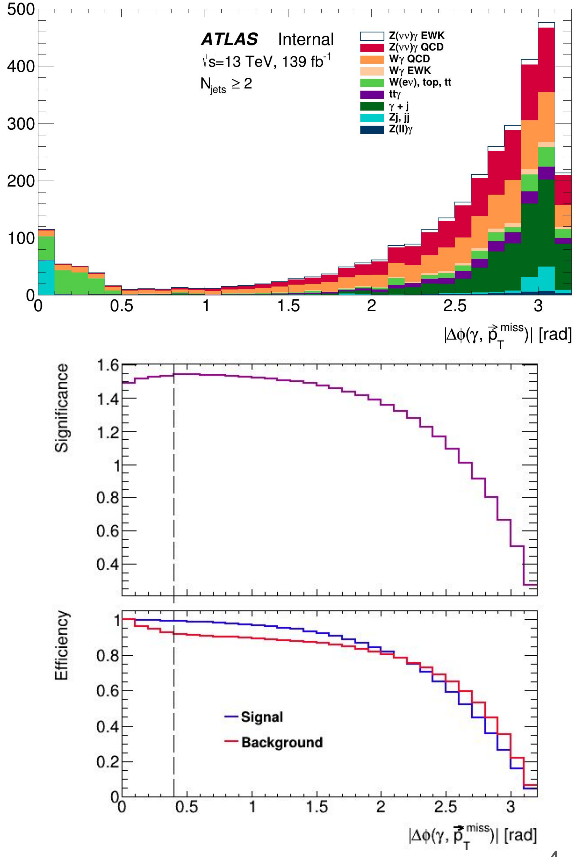
Selections	Cut Value
E_T^{miss}	> 120 GeV
E_T^{γ}	> 150 GeV
number of photons	$n_{\gamma} = 1$
lepton veto	$n_e = 0$, $n_{\mu} = 0$
|z-pointing (γ)|	< 250 mm
number of jets	$n_{jets} \geq 2$

$$S = N_{signal}/\sqrt{N_{signal} + N_{bkg}}$$

$$\varepsilon = N_{passed}/N_{all}$$

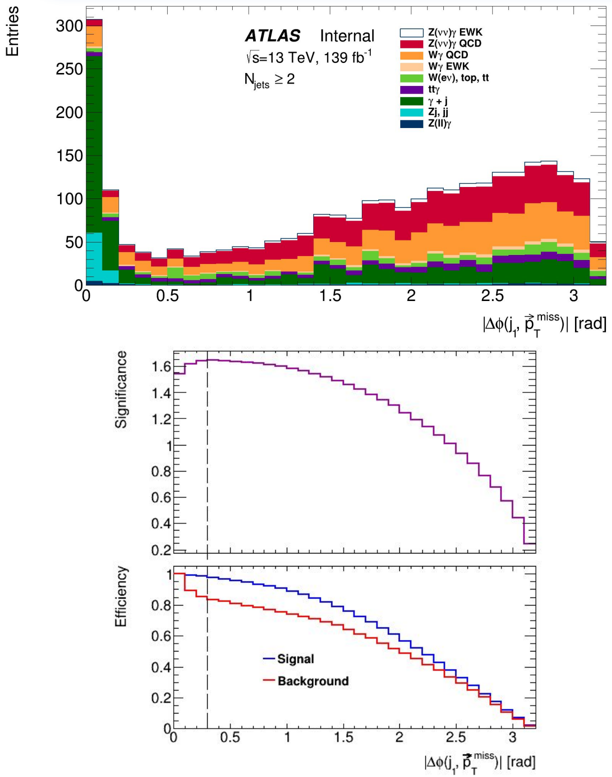
Signal: $Z(\nu\bar{\nu})\gamma$ EWK

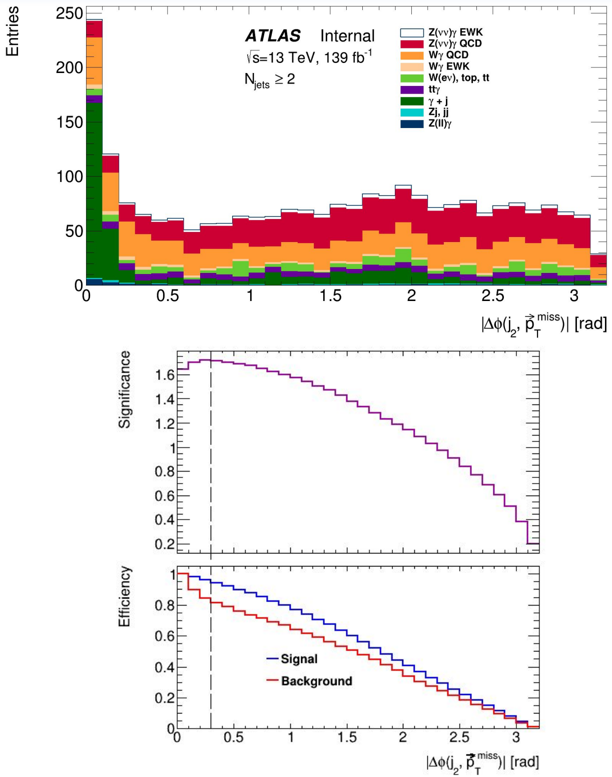


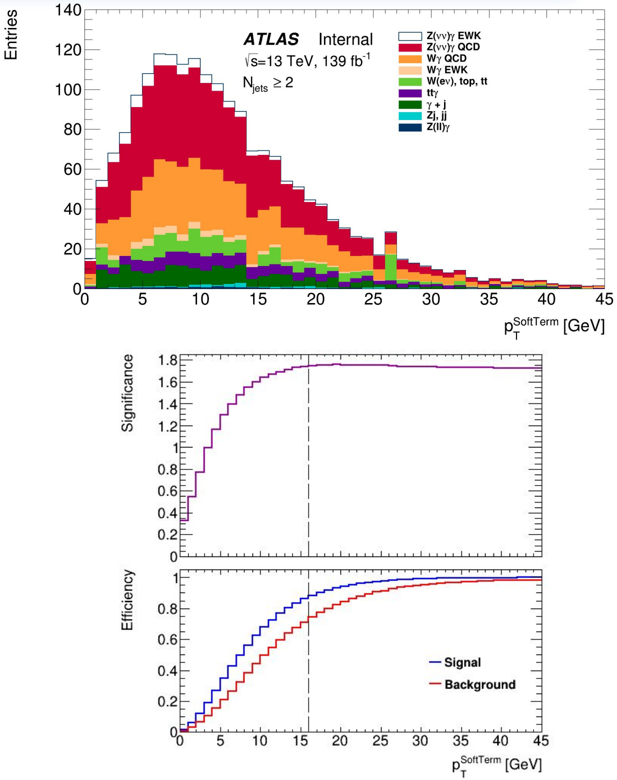


mc 16a+d+e: Preselection done → $S = 0.85 \pm 0.03$ + E_T^{miss} significance > 12 → $S = 1.490 \pm 0.013$ + $|\Delta\varphi(\gamma, E_T^{miss})| > 0.4$ → $S = 1.540 \pm 0.012$

Selection optimisation II: increasing statistical significance







+ $|\Delta\varphi(\text{jet}_{\text{lead}}, E_T^{\text{miss}})| > 0.3$ ➡ S = 1.643 ± 0.010

+ $|\Delta\varphi(\text{jet}_{\text{sublead}}, E_T^{\text{miss}})| > 0.3$ ➡ S = 1.712 ± 0.011

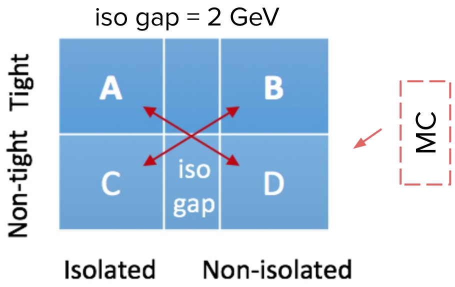
+ $p_T^{\text{soft\_term}} < 16$ GeV ➡ S = 1.739 ± 0.013

Preselection + E_T^{miss} significance > 12 + $|\Delta\varphi(\gamma, E_T^{\text{miss}})| > 0.4$ + $|\Delta\varphi(\text{jet}_{\text{lead}}, E_T^{\text{miss}})| > 0.3$ + $|\Delta\varphi(\text{jet}_{\text{sublead}}, E_T^{\text{miss}})| > 0.3$ + $p_T^{\text{soft\_term}} < 16$ GeV

the statistical significance improvement is 105%, the signal efficiency decrease is 67%

jet→γ misID background I: correlation factor

Source: Z(νν)+jets and multi-jet processes.
Background is estimated from data using **2D-sideband method**:
Photon isolation and identification variables are used to construct the sidebands.



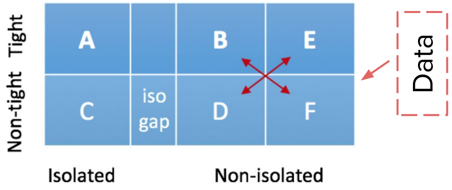
$p_T^{cone20}/p_T^{\gamma} < 0.05$

FixedCutTight:

A: tight, $E_T^{cone40} - 0.022\, p_T^{\gamma} < 2.45$ [GeV]
B: tight, $2.45 + \text{gap} < E_T^{cone40} - 0.022\, p_T^{\gamma} < 29.45$ [GeV]
C: non-tight, $E_T^{cone40} - 0.022\, p_T^{\gamma} < 2.45$ [GeV]
D: non-tight, $2.45 + \text{gap} < E_T^{cone40} - 0.022\, p_T^{\gamma} < 29.45$ [GeV]

Non-tight: at least one of the cuts on the following variables should fail in these:

- *loose'2*: w_{s3}, F_{side}
- *loose'3*: $w_{s3}, F_{side}, \Delta E$
- *loose'4*: $w_{s3}, F_{side}, \Delta E, E_{ratio}$
- *loose'5*: $w_{s3}, F_{side}, \Delta E, E_{ratio}, w_{tot}$



Isolation should not correlate with non-tight ID!

$$\frac{N_A^{jet\to\gamma}}{N_B} = \frac{N_C}{N_D}$$

Correlation is measured in data and MC by $R = \frac{N_A N_D}{N_B N_C}$.

(relaxed selections)

with upper cut

R factor	*loose'2*	*loose'3*	*loose'4*	*loose'5*
MC	1.13 ± 0.12	1.26 ± 0.13	1.33 ± 0.12	1.56 ± 0.14
Data-driven	0.98 ± 0.07	0.99 ± 0.06	0.93 ± 0.06	0.94 ± 0.06

Best agreement
Smallest correlation

without upper cut

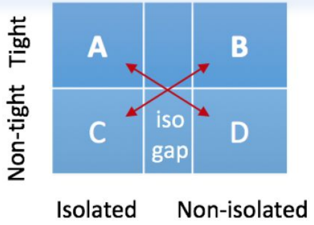
R factor	*loose'2*	*loose'3*	*loose'4*	*loose'5*
MC	1.13 ± 0.12	1.28 ± 0.12	1.34 ± 0.11	1.56 ± 0.13
Data-driven	0.79 ± 0.05	0.81 ± 0.05	0.73 ± 0.04	0.72 ± 0.04

Unstable

Used as nominal

Purity	*loose'2*	*loose'3*	*loose'4*	*loose'5*
Region C	0.20 ± 0.06	0.20 ± 0.06	0.19 ± 0.04	0.18 ± 0.04
Region D	0.36 ± 0.05	0.37 ± 0.06	0.31 ± 0.04	0.30 ± 0.04

jet→γ misID background II: track isolation inversion



Inversion of the track isolation selection in non-isolated regions

Track isolation: $p_T^{cone20}/p_T^{\gamma} < 0.05$

Region	Tight	Isolation	
		Calorimeter	Track
A	+	+	+
B	+	-	+
C	-	+	+
D	-	-	+

R factor	*loose'2*	*loose'3*	*loose'4*	*loose'5*
MC	1.13 ± 0.12	1.26 ± 0.13	1.33 ± 0.12	1.56 ± 0.14
Data-driven	0.98 ± 0.07	0.99 ± 0.06	0.93 ± 0.06	0.94 ± 0.06

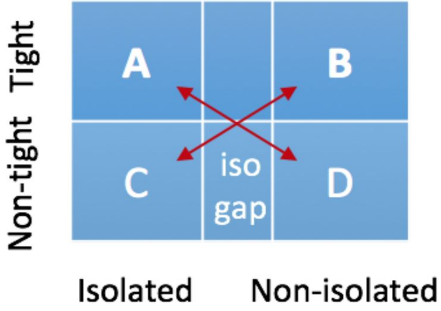
Region	Tight	Isolation	
		Calorimeter	Track
A	+	+	+
B	+	-	-
C	-	+	+
D	-	-	-

R factor	*loose'2*	*loose'3*	*loose'4*	*loose'5*
MC	1.11 ± 0.10	1.29 ± 0.10	1.42 ± 0.10	1.70 ± 0.12
Data-driven	1.21 ± 0.02	1.227 ± 0.018	1.182 ± 0.015	1.200 ± 0.015

Without track isolation inversion more stable and less correlated results are obtained.

jet→γ misID background III: estimation technique

iso gap = 2 GeV



The signal leakage parameters:

$$c_B = \frac{N_B^{Z(\nu\bar{\nu})\gamma}}{N_A^{Z(\nu\bar{\nu})\gamma}}, \quad c_C = \frac{N_C^{Z(\nu\bar{\nu})\gamma}}{N_A^{Z(\nu\bar{\nu})\gamma}}, \quad c_D = \frac{N_D^{Z(\nu\bar{\nu})\gamma}}{N_A^{Z(\nu\bar{\nu})\gamma}} \quad \Big\} \text{ MC}$$

$$N_A^{Z(\nu\bar{\nu})\gamma} = \tilde{N}_A - (\tilde{N}_B - c_B N_A^{Z(\nu\bar{\nu})\gamma}) \frac{\tilde{N}_C - c_C N_A^{Z(\nu\bar{\nu})\gamma}}{\tilde{N}_D - c_D N_A^{Z(\nu\bar{\nu})\gamma}}$$

($N\tilde{}_i$ - data N_i with subtracted N_i^{bkg})

$$a = c_D - c_B c_C;$$
$$b = \tilde{N}_D + c_D \tilde{N}_A - (c_B \tilde{N}_C + c_C \tilde{N}_B);$$
$$c = \tilde{N}_D \tilde{N}_A - \tilde{N}_C \tilde{N}_B.$$

$$N_A^{Z(\nu\bar{\nu})\gamma} = \frac{b - \sqrt{b^2 - 4ac}}{2a}$$

The number of events arising in each of the regions:

$$N_A = N_A^{Z(\nu\bar{\nu})\gamma} + N_A^{bkg} + N_A^{jet\to\gamma};$$
$$N_B = c_B N_A^{Z(\nu\bar{\nu})\gamma} + N_B^{bkg} + N_B^{jet\to\gamma};$$
$$N_C = c_C N_A^{Z(\nu\bar{\nu})\gamma} + N_C^{bkg} + N_C^{jet\to\gamma};$$
$$N_D = c_D N_A^{Z(\nu\bar{\nu})\gamma} + N_D^{bkg} + N_D^{jet\to\gamma};$$

$$\rightarrow N_A^{jet\to\gamma}$$

	Data	$Z(\nu\nu)\gamma$ QCD	$W\gamma$ QCD	$W\gamma$ EWK	$W(e\nu), top, tt$	$tt\gamma$	γ+jet	$Z(ll)\gamma$	$W(\tau\nu)$
A	blinded	576.7 ± 1.8	404 ± 6	42.3 ± 0.4	111 ± 2	88 ± 2	122 ± 11	11.1 ± 0.8	9 ± 3
B	108 ± 10	29.8 ± 0.4	16.5 ± 1.9	1.54 ± 0.08	4.33 ± 0.07	4.5 ± 0.4	2.7 ± 1.3	0.5 ± 0.2	4.4 ± 1.9
C	41 ± 6	5.17 ± 0.16	3.9 ± 0.6	0.42 ± 0.04	1.38 ± 0.02	1.1 ± 0.3	0.9 ± 0.8	0.11 ± 0.05	5.0 ± 1.9
D	37 ± 6	0.23 ± 0.03	0.14 ± 0.08	0.015 ± 0.007	0.095 ± 0.002	0.19 ± 0.09	0.3 ± 0.3	0 ± 0	8.0 ± 1.6

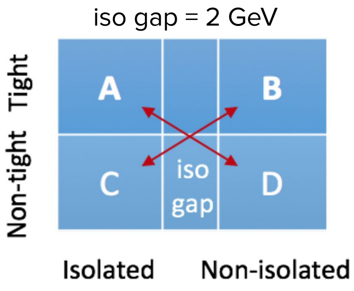
jet→γ misID background IV: uncertainties

➢ **Statistical uncertainty:**

- The event yields of four regions in data and non jet ➔ γ background are varied by ±1σ independently.
- The statistical uncertainty on the signal leakage parameters is **negligible**.

Total statistics: 52%.

Central value	33^{+15}_{-17}
Loose'3	−6
Loose'4	−1
Loose'5	−5
Isolation gap +1 GeV	+6
Isolation gap −0.6 GeV	−3



➢ **Systematic uncertainty:**

- Anti-tight definition and isolation gap choice – variations of ABCD region determination for ±1σ changes in data yield (**18%**).
- Uncertainty coming from the signal leakage parameters is obtained via using two different generators (**6%**).

	$Z\gamma jj$ EWK			$Z\gamma jj$ QCD		
Signal leakage parameters	MadGraph+Pythia8	MadGraph+Herwig7	Relative deviation	Sherpa 2.2	MadGraph+Pythia8	Relative deviation
c_B	0.0314 ± 0.0005	0.0318 ± 0.0006	1.3%	0.0517 ± 0.0007	0.047 ± 0.003	9%
c_C	0.0085 ± 0.0003	0.0088 ± 0.0003	3%	0.0090 ± 0.0003	0.0096 ± 0.0011	6%
c_D	0.00033 ± 0.00005	0.00040 ± 0.00006	18%	0.00039 ± 0.00005	0.0006 ± 0.0002	35%
$jet \to \gamma$ estimation	33^{+15}_{-17}	33^{+15}_{-17}	0%	31^{+18}_{-19}	32^{+17}_{-18}	3%

- The iso/ID uncertainty on reconstruction photon efficiency $\delta_{eff}^{iso/ID}$ (**3%**):

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

$\delta^{eff}_{iso} = 0.023$

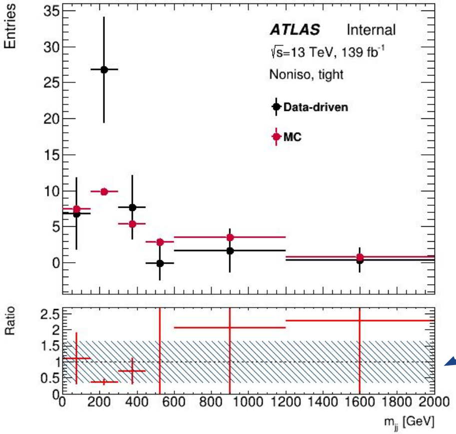
$\delta^{eff}_{iso/ID} = 0.019$

Total systematics: 19%.

★ **Resulting number** of jet➔γ events in <u>Zγ inclusive region</u> is $33^{+15}_{-17} \pm 6$. Z(νν)+jets and multi-jet MC predict 9±2 events.

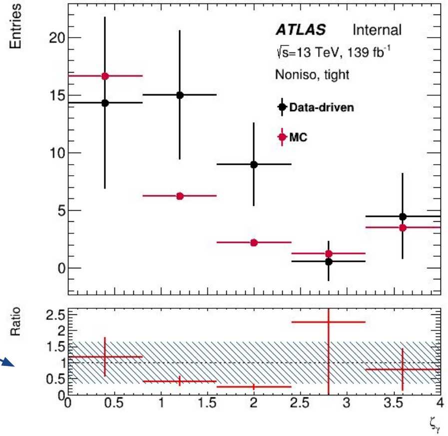
jet→γ misID background V: Zγ inclusive and signal regions

The extrapolation of jet→γ background estimation from Zγ inclusive region to <u>the signal region</u>:

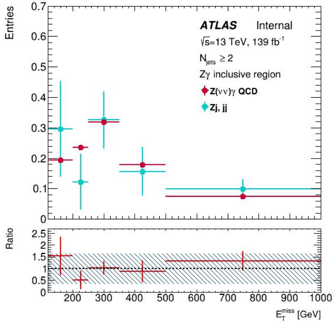


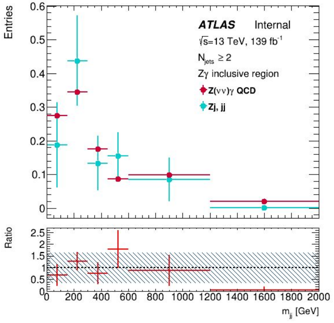
$Z\gamma jj$ EWK signal region	
N_{leptons}	$= 0$
m_{jj}	> 300 GeV
γ-centrality	< 0.6

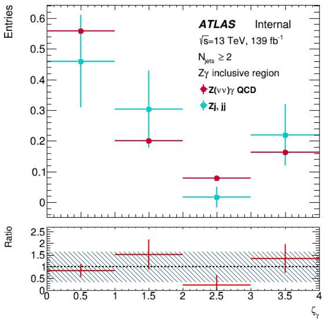
Statistical and systematic uncertainties of jet→γ background estimation



The current uncertainty for this estimation covers all differences and should not be increased.







Good agreement of the shapes.

⇩

The shape of jet → γ background for normalization is taken from Zγ QCD.

Conclusion

- Z(νν)γjj EWK analysis is almost finalized. The optimization process is finished, jet → γ background estimation with uncertainties is done.
- Z(νν)γ inclusive analysis is started (first optimizations and bkg estimations). At the moment the framework is being changed to improve the performance (use up-to-date jets topology, add b-tagging etc).
- There is other work with the student on Z(ll)Z(νν) analysis.
- TRT tracking properties optimization (qualification task is ended, added to the ATLAS author list, work continues).
- Articles and conferences.

Back-up

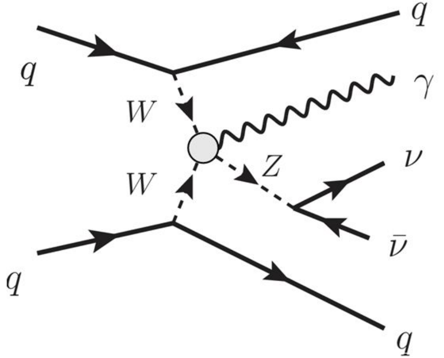
Motivation

<u>Topology</u>: $\gamma + p_T^{miss}$ + 2 hadronic jets. All objects are with high energy.

- EWK production (QCD = 0; QED ≤ 5) – aim of the study.
- QCD production (QCD = 2; QED = 2) – main irreducible background.

❖ The main goal of this study is to make observation or evidence of the $Z(\nu\nu)\gamma$ EWK process for the first time.

❖ EWK $Z(\nu\nu)\gamma$ production is the one of the most sensitive final states to aQGC (O_M and O_T operators).



Vector boson scattering

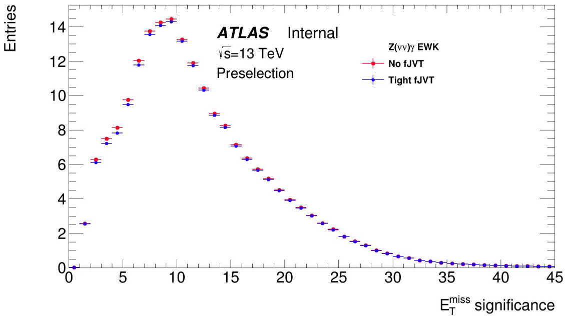
Dimension 8 operators:

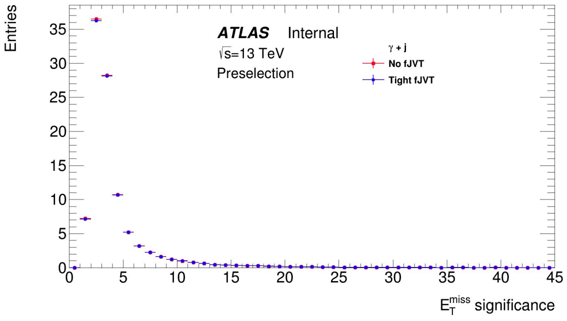
SM (WWAZ) — **Beyond SM** (ZZZA, ZZAA, ZAAA)

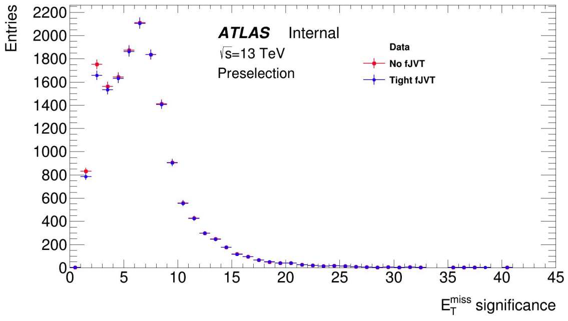
	WWWW	WWZZ	ZZZZ	WWAZ	WWAA	ZZZA	ZZAA	ZAAA	AAAA
$\mathcal{O}_{S,0/1}$	✓	✓	✓						
$\mathcal{O}_{M,0/1/6/7}$	✓	✓	✓	✓	✓	✓	✓		
$\mathcal{O}_{M,2/3/4/5}$		✓	✓	✓	✓	✓	✓		
$\mathcal{O}_{T,0/1/2}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\mathcal{O}_{T,5/6/7}$		✓	✓	✓	✓	✓	✓	✓	✓
$\mathcal{O}_{T,8/9}$			✓			✓	✓	✓	✓

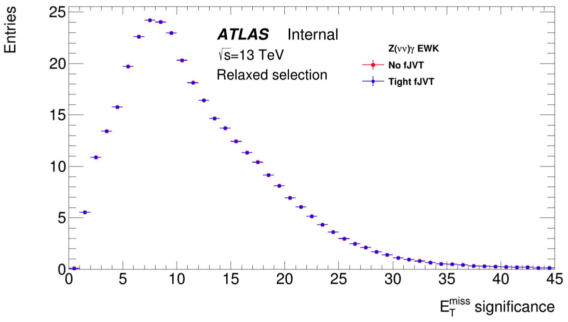
Forward jets in the final state and applying of the fJVT cut

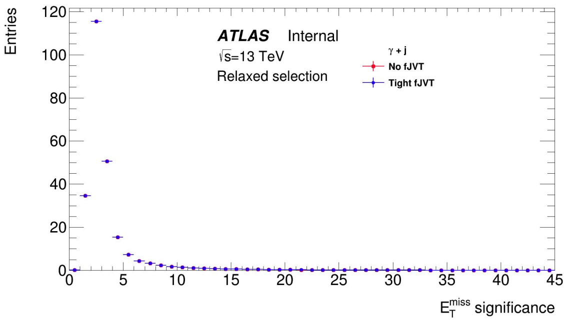
Signal Z(νν)γ EWK:	Bkg γ+j:	Data:

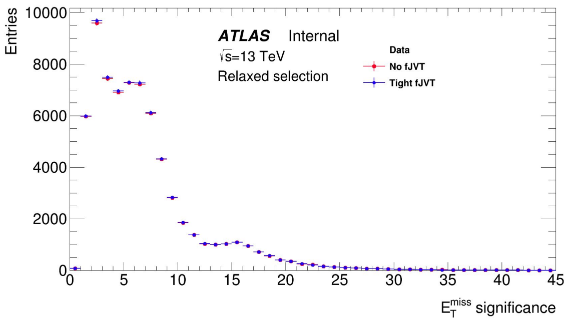








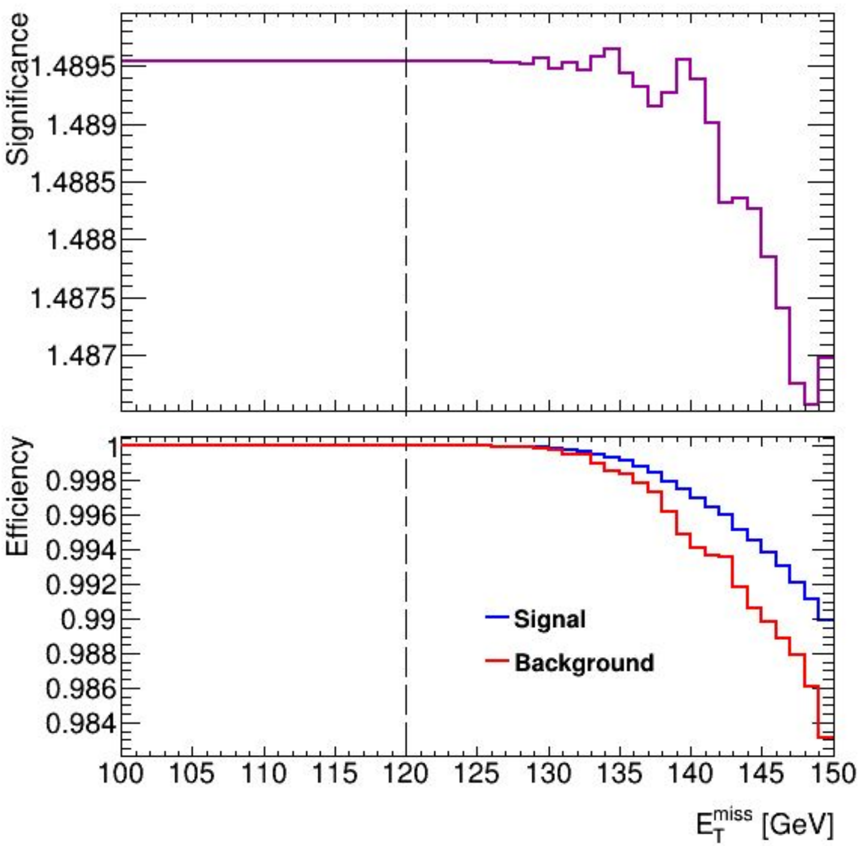




There are marginal changes in the data, signal and background after the fJVT implementation, so it was decided not to use it in the analysis.

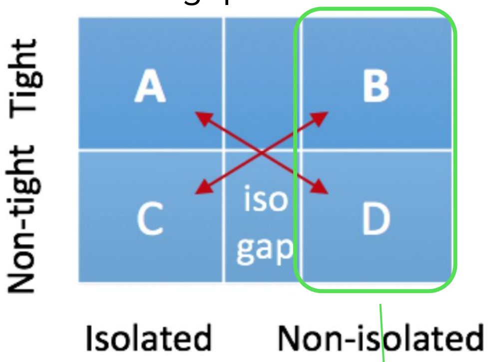
Selection optimisation: E_T^{miss} cut

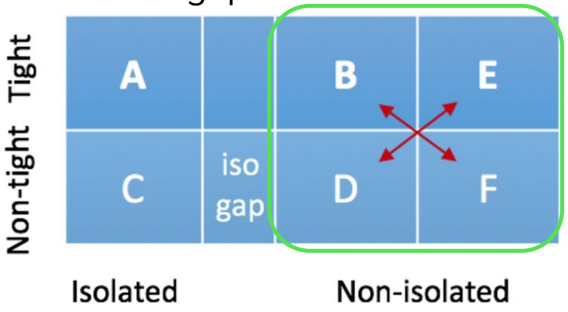
E_T^{miss} significance > 12



E_T^{miss} > 120 GeV

jet→γ misID background VI: correlation factor in data





FixedCutTight:

$p_T^{cone20}/p_T^{\gamma} < 0.05$

B-E: tight, $4.45 < E_T^{cone40} - 0.022\, p_T^{\gamma} < 11.45$ [GeV]
D-F: non-tight, $4.45 < E_T^{cone40} - 0.022\, p_T^{\gamma} < 11.45$ [GeV]
E: tight, $11.45 < E_T^{cone40} - 0.022\, p_T^{\gamma} < 29.45$ [GeV]
F: non-tight, $11.45 < E_T^{cone40} - 0.022\, p_T^{\gamma} < 29.45$ [GeV]

jet→γ misID background VI: Discriminating variables used for loose and tight photon identification

Category	Description	Name	*loose*	*tight*
Acceptance	$\|\eta\| < 2.37$, with $1.37 < \|\eta\| < 1.52$ excluded	–	✓	✓
Hadronic leakage	Ratio of E_T in the first sampling layer of the hadronic calorimeter to E_T of the EM cluster (used over the range $\|\eta\| < 0.8$ or $\|\eta\| > 1.37$)	R_{had_1}	✓	✓
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster (used over the range $0.8 < \|\eta\| < 1.37$)	R_{had}	✓	✓
EM Middle layer	Ratio of 3×7 $\eta \times \phi$ to 7×7 cell energies	R_η	✓	✓
	Lateral width of the shower	w_{η_2}	✓	✓
	Ratio of 3×3 $\eta \times \phi$ to 3×7 cell energies	R_ϕ		✓
EM Strip layer	Shower width calculated from three strips around the strip with maximum energy deposit	$w_{s\,3}$		✓
	Total lateral shower width	$w_{s\,\text{tot}}$		✓
	Energy outside the core of the three central strips but within seven strips divided by energy within the three central strips	F_{side}		✓
	Difference between the energy associated with the second maximum in the strip layer and the energy reconstructed in the strip with the minimum value found between the first and second maxima	ΔE		✓
	Ratio of the energy difference associated with the largest and second largest energy deposits to the sum of these energies	E_{ratio}		✓

jet→γ misID background VII: isolation gap

Isolation gap	c_B	c_D
0 GeV	0.0517 ± 0.0007	0.00057 ± 0.00008
1 GeV	0.0398 ± 0.0006	0.00041 ± 0.00006
2 GeV	0.0314 ± 0.0005	0.00035 ± 0.00005
3 GeV	0.0254 ± 0.0005	0.00026 ± 0.00005
4 GeV	0.0209 ± 0.0004	0.00020 ± 0.00004

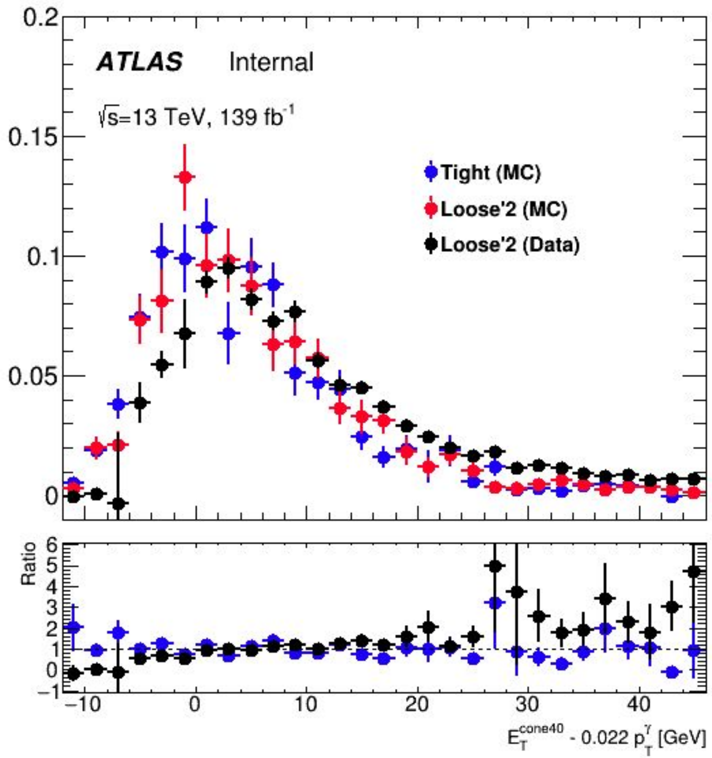
Fraction of signal leakage to control regions B and D for *loose'2* working point and different isolation gaps.

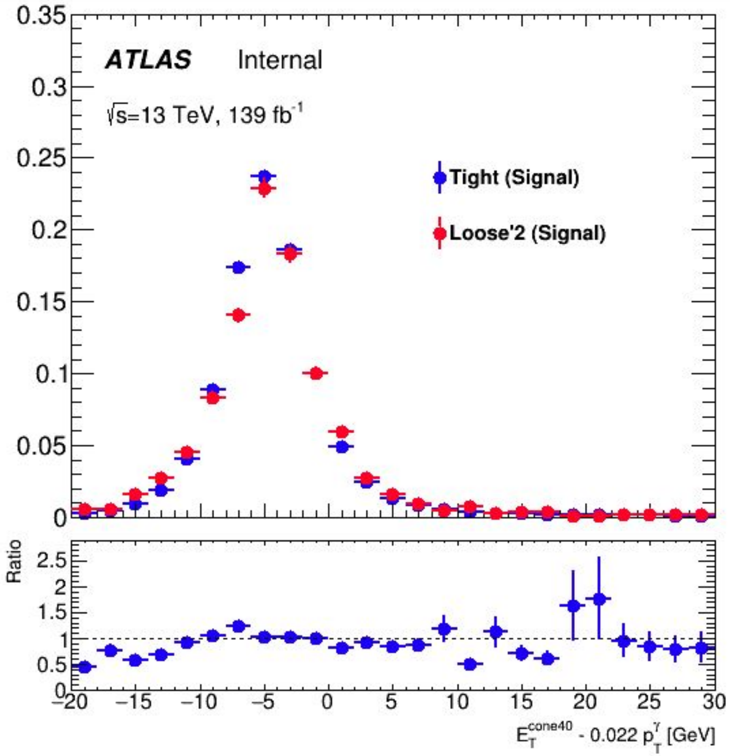
Isolation gap	$jet \rightarrow \gamma$ estimation
2 GeV	33^{+15}_{-17}
5 GeV	23^{+15}_{-19}
7 GeV	21^{+13}_{-15}

The statistical uncertainties for these estimations cover all differences.

Central values of jet → γ background number of events from data-driven estimation for *loose'2* working point and different isolation gaps.

jet→γ misID background VIII: isolation distributions



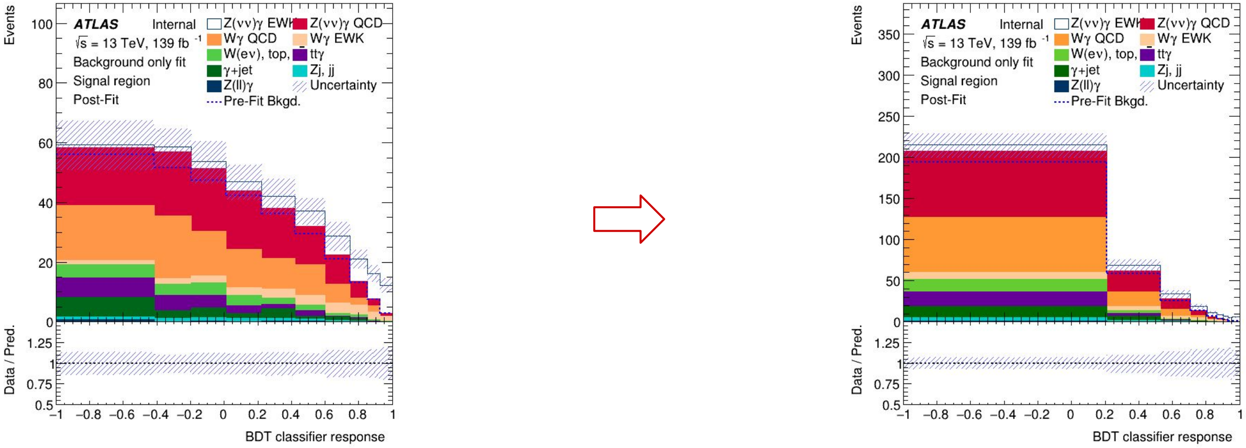


The bottom panel shows the ratio of tight photon candidates from Z+jets simulation and anti-tight photon candidates in data to the anti-tight photon candidates from Z+jets simulation.

Optimization of BDT response binning in SR

In order to increase the expected median significance, the binning of BDT response in the SR was optimized with the automatic binning algorithms (transformations) included in the TRExFitter.

For initially used TransfoD binning algorithm, the merging threshold $Z = z_b \frac{n_b}{N_b} + z_s \frac{n_s}{N_s}$.



The expected median significance can be increased by using TransfoF algorithm instead of TransfoD:

$$Z = \sqrt{z_b \frac{n_b}{N_b}} + \sqrt{z_s n_s log(1 + \frac{n_s}{n_b})}.$$

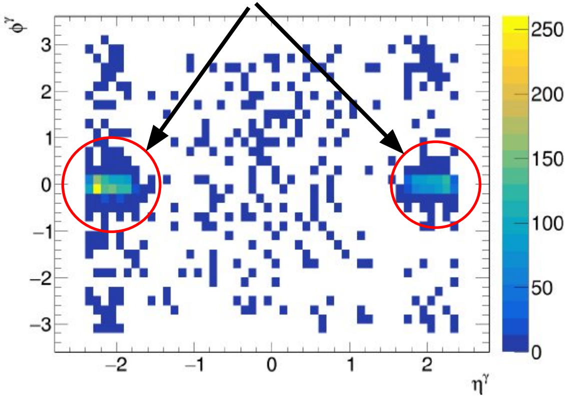
The same optimization was done for $\mu_{Z\gamma EWK}$ = 0.5 and $\mu_{Z\gamma EWK}$ = 2. It was found that the expected median significance enhancement in these cases is at the same level (9%).

Photon pointing selection

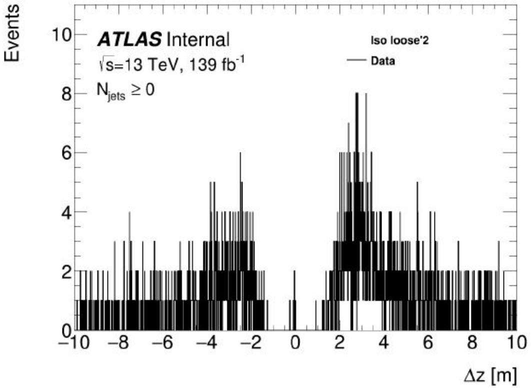
Loose isolated photons are contaminated by beam-induced background.

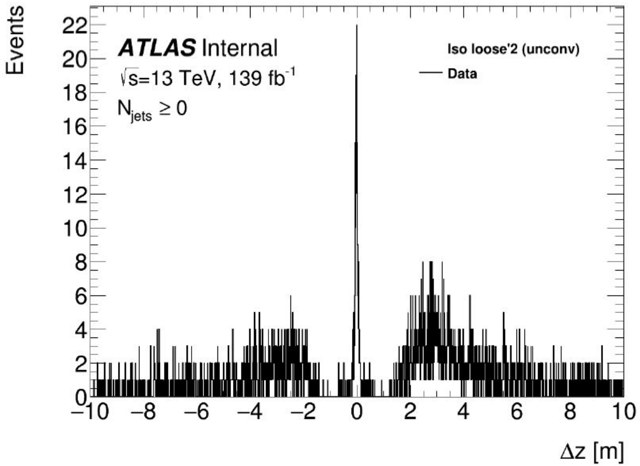
Most of background is concentrated in unconverted photon candidates.

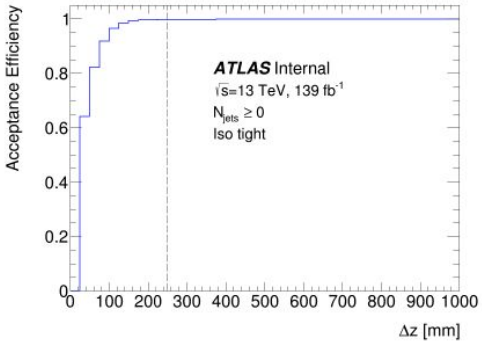
Bkg is concentrated at small φ and high η

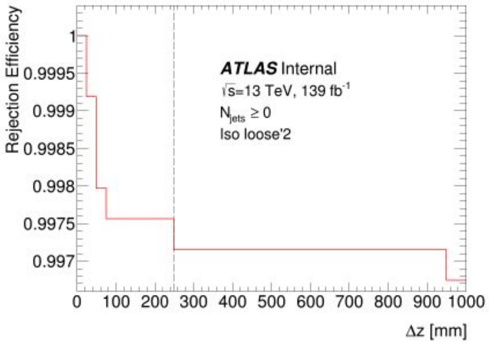


Applying |φ| < 0.2, |η| > 1.7





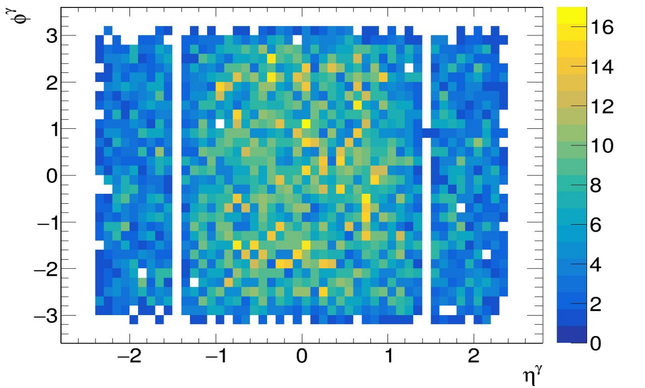




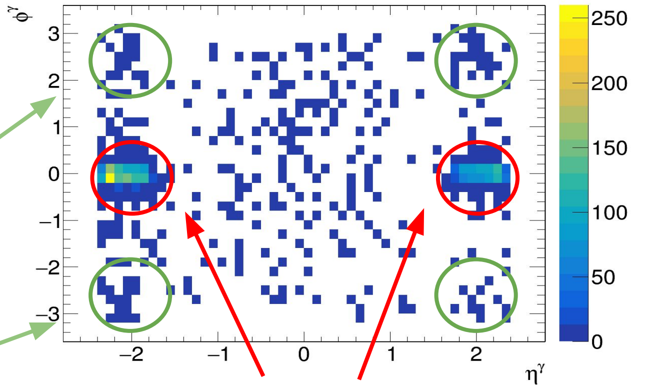
Absolute value of z coordinate pointed by the photon candidate with respect to the identified primary vertex is required to be **less than 250 mm**

Beam-induced background

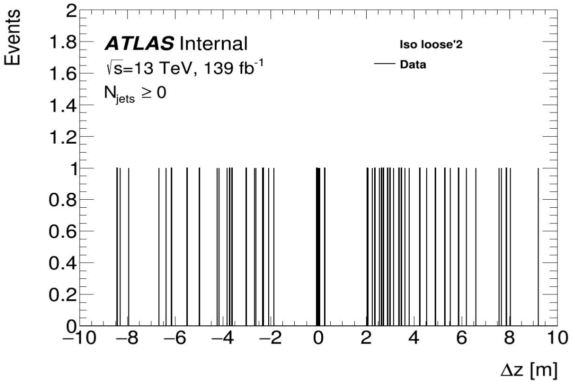
Distribution of photon φ versus photon η in the *tight* and isolated region.



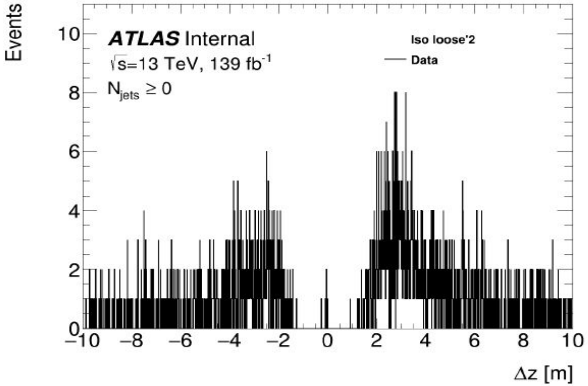
Distribution of photon φ versus photon η in the *loose'2* and isolated region.



Applying |ϕ|> 2.5, |η| > 1.7



Applying |ϕ| < 0.2, |η| > 1.7



Pile-up background

➢ In full Run2 Z(ll)γ inclusive analysis it was found that events with Z and photon from different primary vertices have non-negligible probability (up to **5%** of the total event yield)

Since our final state assumes high energetic photons, E_T(miss), probability of such events should be much smaller.

➢ Fraction of pile-up background is calculated as:

$$f_{PU} = \frac{N^{|\Delta z|>15mm}_{\text{data, 2-track Si}} - SF_1 \times SF_2 \times N^{|\Delta z|>15mm}_{\text{MC, 2-track Si}}}{N_{\text{data, 2-track Si}} \times 0.76}$$

|Δz| requirement was relaxed, because of low statistics

- SF_1 is equal to the ratio of events in data to events in Sherpa MC sample near |Δz| around zero (4.1±0.3)
- SF_2 – normalization factor taking into account the mismodelling in the tails of |Δz| distribution (was calculated for Sherpa Zγ QCD by Zγ inclusive team for us using events with FSR photons) (1.27±0.07)
- N_{data}(|z|>15mm)=11±3

f_{PU}=1.9±1.9%

✔ 1.9% global systematic uncertainty is conservatively added to take this possible background into account

✔ Δϕ distributions in CR1 are checked in order to check the impact of pile-up background on the shapes