Hunting for dark photons from Higgs boson decays with the ATLAS detector: a data-driven approach to the estimation of backgrounds in events with a photon and missing transverse momentum

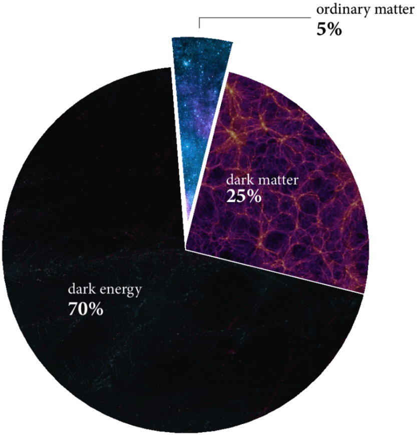
Tesi di Laurea Magistrale in Fisica di: Giulia Maineri

Relatori: Prof. Marcello Fanti, Dott.ssa Silvia Resconi, Dott.ssa Federica Piazza

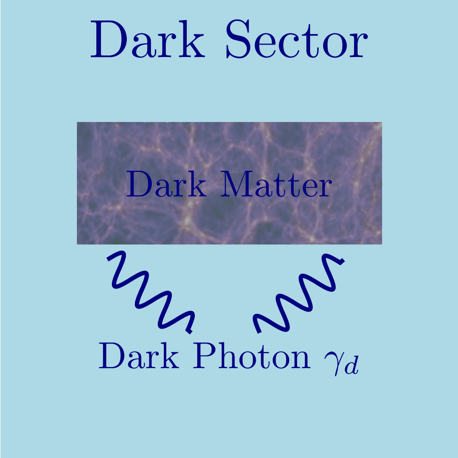
UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

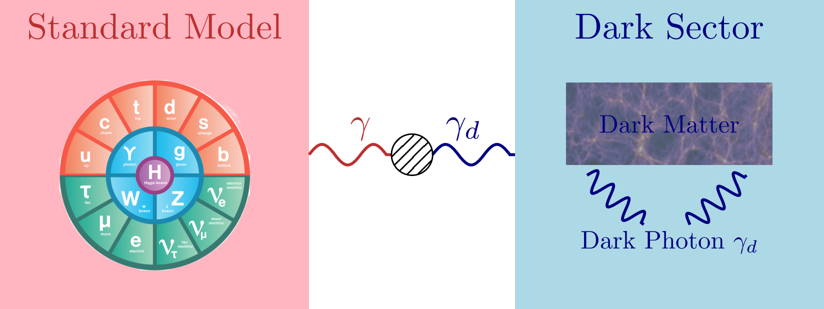
Dark Matter



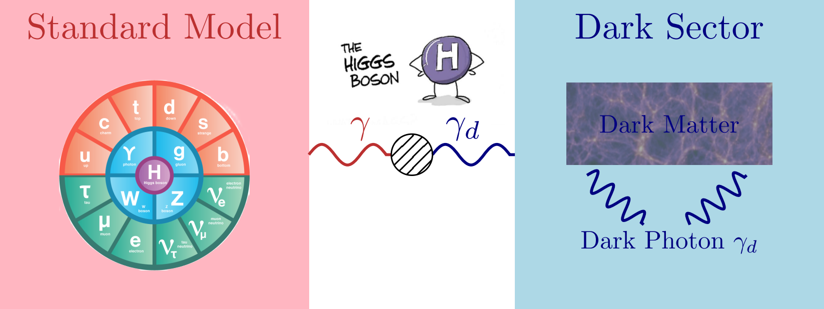
Dark sector



A portal to the dark sector

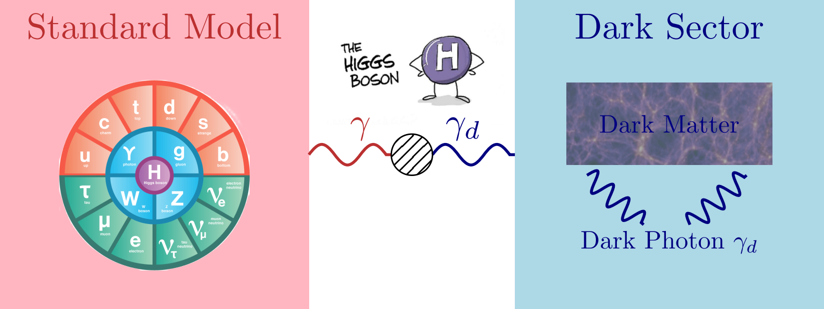


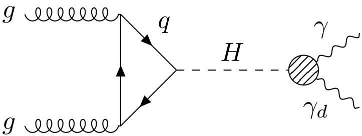
Higgs boson



Standard Model
THE HIGGS BOSON
H
γ
γ_d
Dark Sector
Dark Matter
Dark Photon γ_d

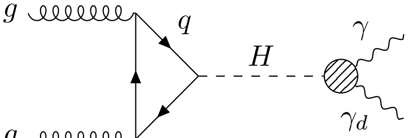
Signal

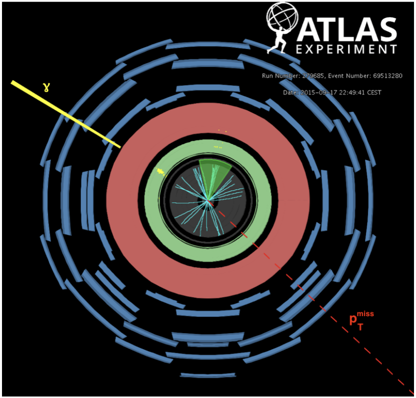




Signal: $gg \to H \to \gamma\gamma_d$

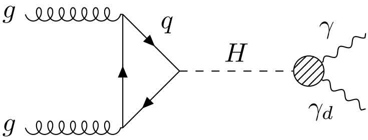
Missing transverse momentum





$$\vec{p}_T^{\,miss} = -\left[\sum_e \vec{p}_T^{\,(e)} + \sum_\mu \vec{p}_T^{\,(\mu)} + \sum_\gamma \vec{p}_T^{\,(\gamma)} + \sum_\tau \vec{p}_T^{\,(\tau)} + \sum_{jet} \vec{p}_T^{\,(jet)} + \sum_{soft} \vec{p}_T^{\,(soft)}\right]$$

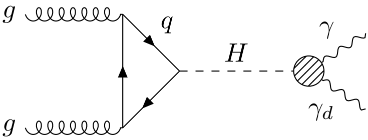
Signal Region



Signal: $gg \to H \to \gamma\gamma_d$

- 1 well-identified ("tight") isolated photon with $p_T^{\gamma} > 50$ GeV;

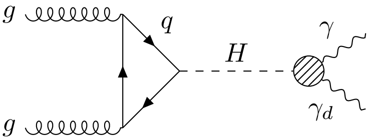
Signal Region



Signal: $gg \to H \to \gamma\gamma_d$

- 1 well-identified ("tight") isolated photon with $p_T^{\gamma} > 50$ GeV;
- missing transverse momentum $p_T^{miss} > 100$ GeV;

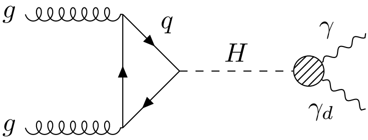
Signal Region



Signal: $gg \to H \to \gamma\gamma_d$

- 1 well-identified ("tight") isolated photon with $p_T^{\gamma} > 50$ GeV;
- missing transverse momentum $p_T^{miss} > 100$ GeV;
- leptons veto, $N_l = 0, l \in \{e, \mu, \tau\}$;

Signal Region



Signal: $gg \rightarrow H \rightarrow \gamma\gamma_d$

- 1 well-identified ("tight") isolated photon with $p_T^{\gamma} > 50$ GeV;
- missing transverse momentum $p_T^{miss} > 100$ GeV;
- leptons veto, $N_l = 0, l \in \{e, \mu, \tau\}$;
- transverse mass $m_T > 80$ GeV.

$$m_T = \sqrt{2 p_T^{miss} p_T^{\gamma} (1 - \cos \Delta\Phi(\vec{p}_T^{\gamma}, \vec{p}_T^{miss}))}$$

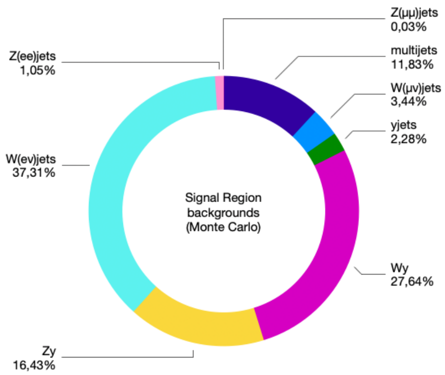
Backgrounds

Background processes:

- irreducible: $Z(\to \nu\nu)\gamma$;
- reducible:
  - $W(\to l\nu_l)\gamma$, lost lepton;
  - multijets, $W(\to \mu\nu_\mu)$jets, $Z(\to \mu\mu)$jets, jets faking photons (jet$\rightsquigarrow \gamma$);
  - γjets, true photon but fake p_T^{miss};
  - $W(\to e\nu_e)$jets, $Z(\to ee)$jets, electrons faking photons (e$\rightsquigarrow \gamma$).

Background processes:

- irreducible: $Z(\to \nu\nu)\gamma$;
- reducible:
  - $W(\to l\nu_l)\gamma$, lost lepton;
  - multijets, $W(\to \mu\nu_\mu)$jets, $Z(\to \mu\mu)$jets, jets faking photons (jet$\rightsquigarrow \gamma$);
  - γjets, true photon but fake p_T^{miss};
  - $W(\to e\nu_e)$jets, $Z(\to ee)$jets, electrons faking photons (e$\rightsquigarrow \gamma$).

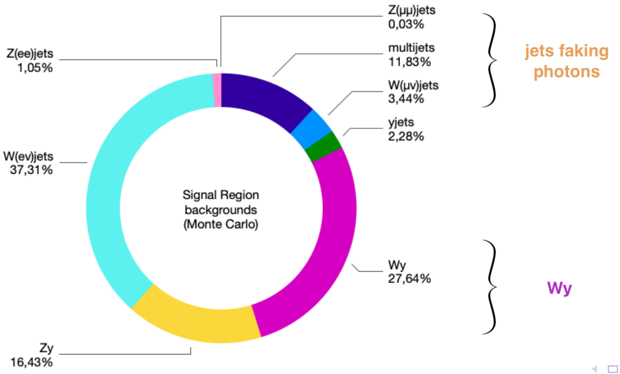


Background composition in pure Monte Carlo samples.

Backgrounds

Background processes:

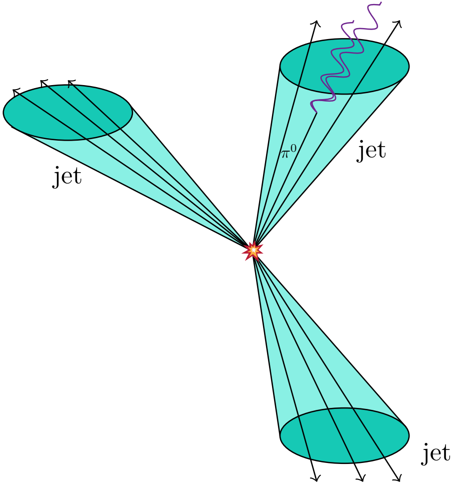
- irreducible: $Z(\to \nu\nu)\gamma$;
- reducible:
  - $W(\to l\nu_l)\gamma$, lost lepton;
  - multijets, $W(\to \mu\nu_\mu)$jets, $Z(\to \mu\mu)$jets, jets faking photons (jet$\rightsquigarrow \gamma$);
  - γjets, true photon but fake p_T^{miss};
  - $W(\to e\nu_e)$jets, $Z(\to ee)$jets, electrons faking photons (e$\rightsquigarrow \gamma$).



Jets faking photons background

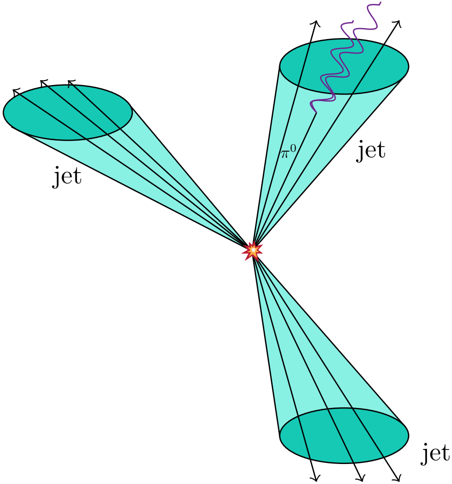
Jets faking photons

Before the reconstruction process:

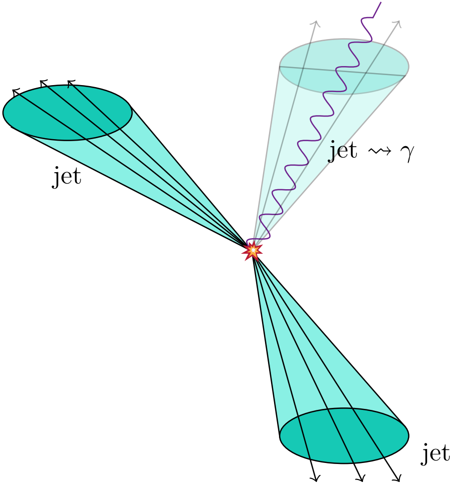


Jets faking photons

Before the reconstruction process:



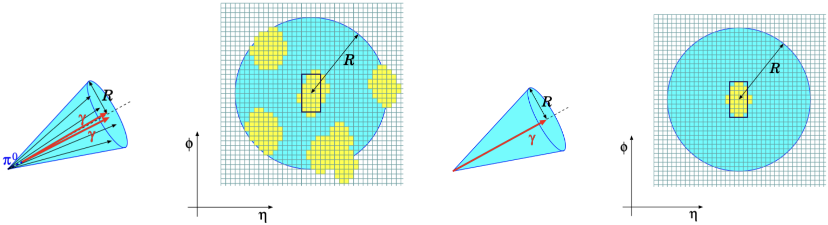
After the reconstruction process:



Isolation

How can we discriminate true and fake photons?
$\Longrightarrow$ **Isolation**

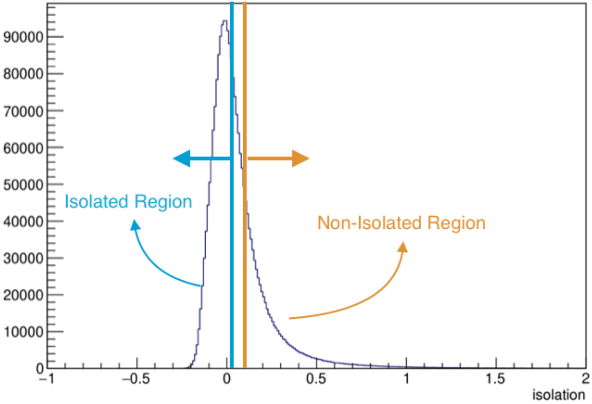
$$isol = \frac{E_T^{isolation}}{p_T^{\gamma}}$$



Isolation Regions

How can we discriminate true and fake photons?
$\Longrightarrow$ **Isolation**

$$isol = \frac{E_T^{isolation}}{p_T^{\gamma}}$$

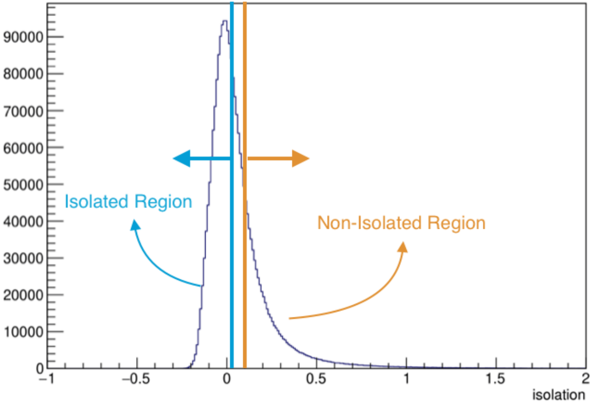


Fake factors

How can we discriminate true and fake photons?

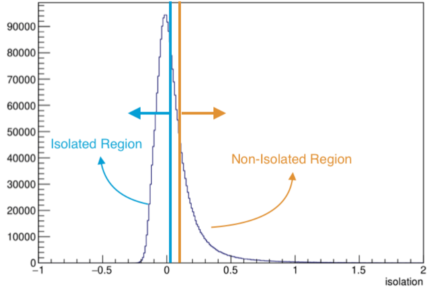
$\Longrightarrow$ **Isolation**

$$isol = \frac{E_T^{isolation}}{p_T^{\gamma}}$$



$$f = \left(\frac{N^{isol}_{j \rightsquigarrow \gamma}}{N^{non-isol}_{j \rightsquigarrow \gamma}} \right)_{tight}$$

Fake factors calculation



$$f = \left(\frac{N^{isol}_{j \rightsquigarrow \gamma}}{N^{non-isol}_{j \rightsquigarrow \gamma}} \right)_{tight}$$

Problems:

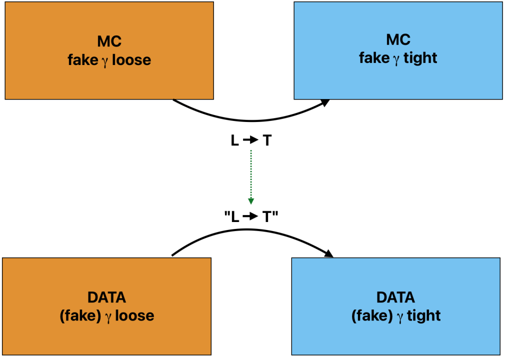
- We cannot fully trust **Monte Carlo** as isolation is typically not well modelled for jets faking photons;
- In **data**, we cannot distinguish true and fake photons!

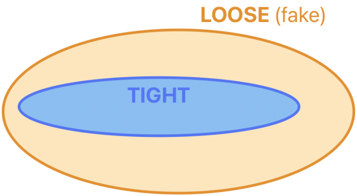
$\Longrightarrow$ we need to **extract** the tight fake photons isolation distribution in data.

Extrapolation method

How to get tight fake photons isolation distribution in data?

L: loose
T: tight

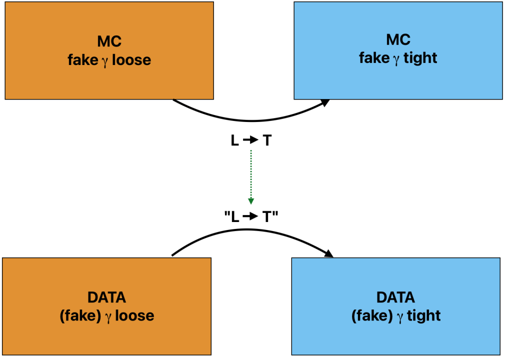


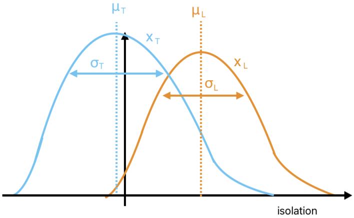


Extrapolation method

How to get tight fake photons isolation distribution in data?

L: loose
T: tight

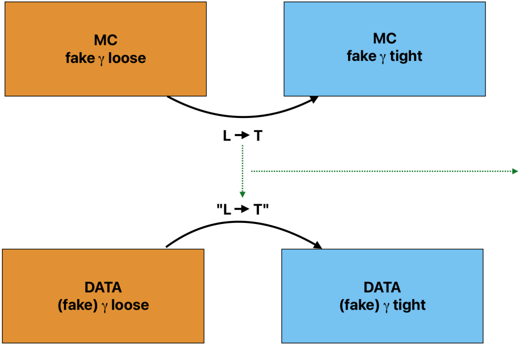




Extrapolation method

How to get tight fake photons isolation distribution in data?

L: loose
T: tight



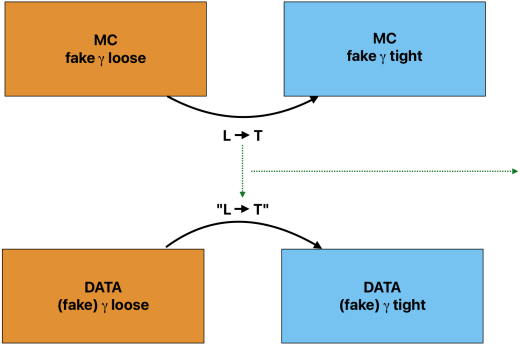
$$\frac{\sigma_T^{data}}{\sigma_L^{data}} = \frac{\sigma_T^{MC}}{\sigma_L^{MC}}$$

$$\frac{\mu_T^{data} - \mu_L^{data}}{\sigma_L^{data}} = \frac{\mu_T^{MC} - \mu_L^{MC}}{\sigma_L^{MC}}$$

Extrapolation method

How to get tight fake photons isolation distribution in data?

L: loose
T: tight



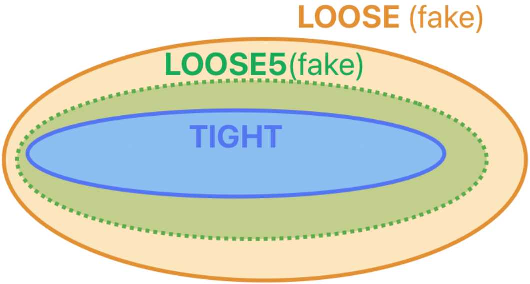
$$\frac{\sigma_T^{data}}{\sigma_L^{data}} = \frac{\sigma_T^{MC}}{\sigma_L^{MC}}$$

$$\frac{\mu_T^{data} - \mu_L^{data}}{\sigma_L^{data}} = \frac{\mu_T^{MC} - \mu_L^{MC}}{\sigma_L^{MC}}$$

→to be validated

Validation of the hypotheses

L: loose
T: tight
L5: loose5



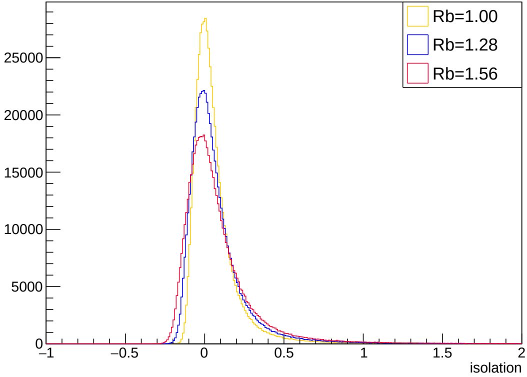
$$\frac{\sigma_{L5}^{data}}{\sigma_{L}^{data}} = R_b \frac{\sigma_{L5}^{MC}}{\sigma_{L}^{MC}}$$

$$\frac{\mu_{L5}^{data} - \mu_{L}^{data}}{\sigma_{L}^{data}} = R \frac{\mu_{L5}^{MC} - \mu_{L}^{MC}}{\sigma_{L}^{MC}}$$

→validated on another photon sample similar to loose (loose5)

Extrapolated distributions

Extrapolated isolation distribution of fake photons passing tight identification

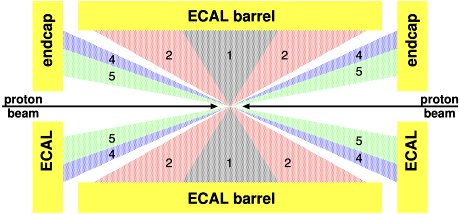


Uncertainties on R, R_b have been assumed to be equal to $|R-1|, |R_b-1|$.

Fake factors

Fake factors have been computed in different **geometric regions** of the detector...

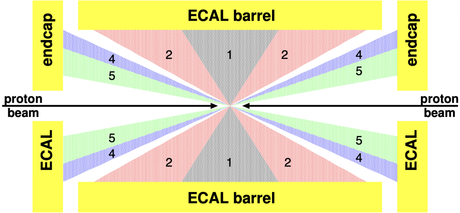
Region	f	σ_f^{sys}	%
1	1.51	0.18	11.7 %
2	2.03	0.35	17.1 %
4	1.95	0.34	17.2 %
5	1.70	0.27	15.7 %



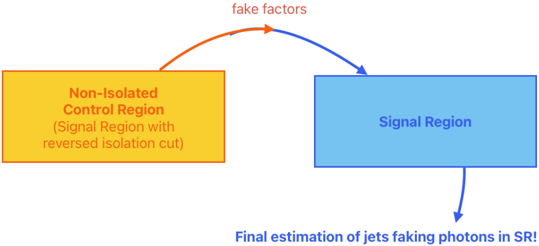
Jets faking photons final estimation

Fake factors have been computed in different **geometric regions** of the detector...

Region	f	σ_f^{sys}	%
1	1.51	0.18	11.7 %
2	2.03	0.35	17.1 %
4	1.95	0.34	17.2 %
5	1.70	0.27	15.7 %

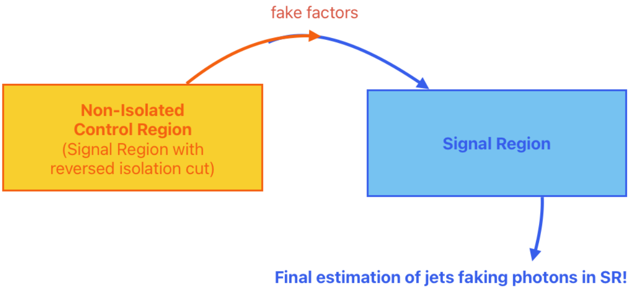


...and then **applied** to the Non-Isolated Control Region.



Jets faking photons final estimation

...and then **applied** to the Non-Isolated Control Region.

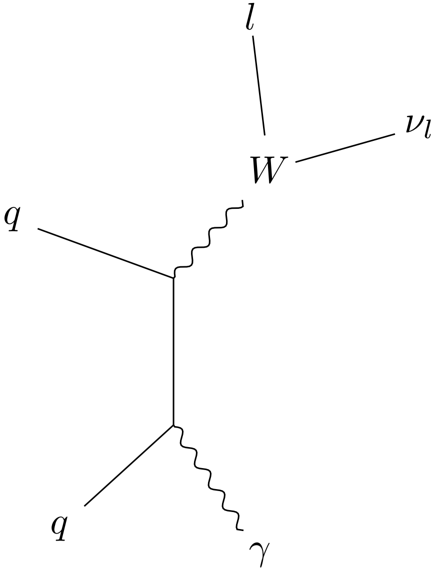


Comparison with the yield in Monte Carlo:

$$N^{MC}_{j \rightsquigarrow \gamma} = 228 \pm 70 \qquad N^{data-driven}_{j \rightsquigarrow \gamma} = 718 \pm 116$$

Wγ background

$W\gamma$ background



$W(\rightarrow l\nu_l)\gamma$

- 1 photon
- genuine p_T^{miss}

$\rightarrow$ when the lepton gets lost, this is a background for our analysis

K-factors

We need to extract **K-factors** to be used to correct the cross-section approximation of Monte Carlo simulations.

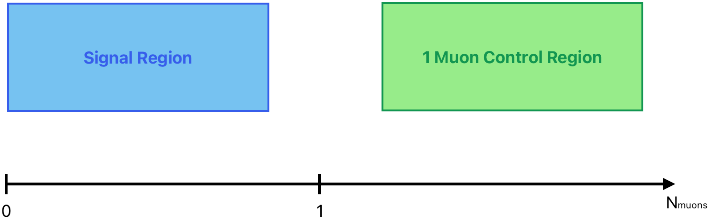
$$K = \left(\frac{N_{W\gamma}^{data}}{N_{W\gamma}^{MC}} \right)_{1\mu CR}$$

K-factors and 1 Muon Control Region

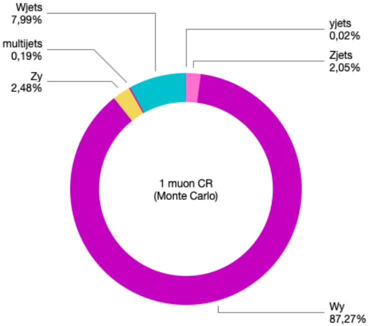
We need to extract **K-factors** to be used to correct the cross-section approximation of Monte Carlo simulations.

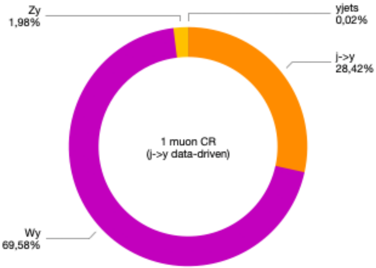
$$K = \left(\frac{N^{data}_{W\gamma}}{N^{MC}_{W\gamma}} \right)_{1\mu CR}$$

We construct a **1 Muon Control Region** enriched of $W\gamma$ events, where the muon is treated as **invisible**.



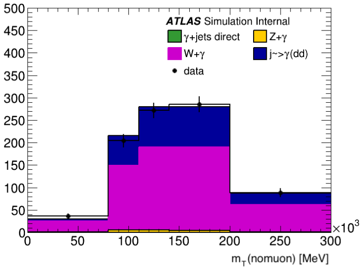
The jets faking photons contribution to the 1 Muon Control Region is estimated using the **fake factors** calculated in the previous part of the work.
$\rightarrow$ Jets faking photons data-driven estimation is much higher than Monte Carlo!





K-factors calculation

K-factors have been computed in different **transverse mass** bins...

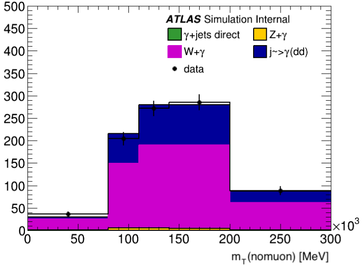


m_T without muon contribution in the 1 Muon CR for $W\gamma$, $Z\gamma$, γjets and data.

K-factors calculation

K-factors have been computed in different **transverse mass** bins...

$$K = \left(\frac{N^{data}_{W\gamma}}{N^{MC}_{W\gamma}} \right)_{1\mu CR} = \left(\frac{N^{data} - N_{bkg \neq W\gamma}}{N^{MC}_{W\gamma}} \right)_{1\mu CR}$$

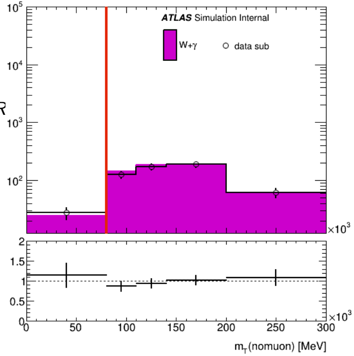


m_T without muon contribution in the 1 Muon CR for $W\gamma$, $Z\gamma$, γjets and data.

K-factors calculation

K-factors have been computed in different **transverse mass** bins...

$$K = \left(\frac{N^{data}_{W\gamma}}{N^{MC}_{W\gamma}} \right)_{1\mu CR} = \left(\frac{N^{data} - N_{bkg \neq W\gamma}}{N^{MC}_{W\gamma}} \right)_{1\mu CR}$$

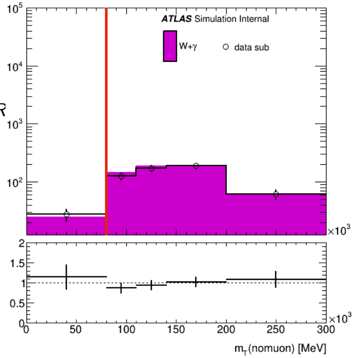


m_T without muon contribution in the 1 Muon CR for $W\gamma$ and subtracted data.

K-factors calculation

K-factors have been computed in different **transverse mass** bins...

$$K = \left(\frac{N^{data}_{W\gamma}}{N^{MC}_{W\gamma}}\right)_{1\mu CR} = \left(\frac{N^{data} - N_{bkg\neq W\gamma}}{N^{MC}_{W\gamma}}\right)_{1\mu CR}$$



m_T (GeV)	K	σ_K^{stat}	σ_K^{sys}	σ_K^{tot}
80-110	0.869	0.131	0.074	0.150
110-140	0.939	0.119	0.076	0.141
140-200	1.023	0.117	0.073	0.138
>200	1.089	0.197	0.067	0.208

where

$$\sigma_K^{sys} = \frac{K(f - \sigma_f) - K(f + \sigma_f)}{2}$$

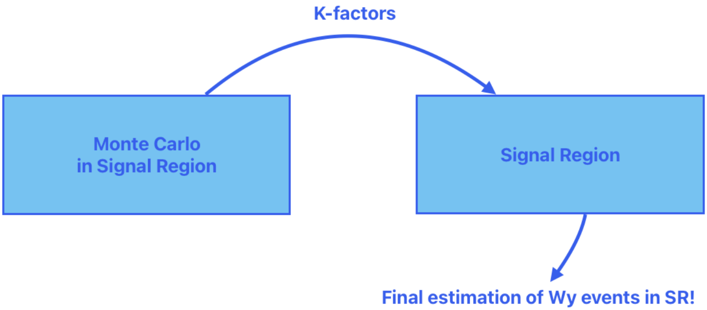
m_T without muon contribution in the 1 Muon CR for $W\gamma$ and subtracted data.

K-factors application

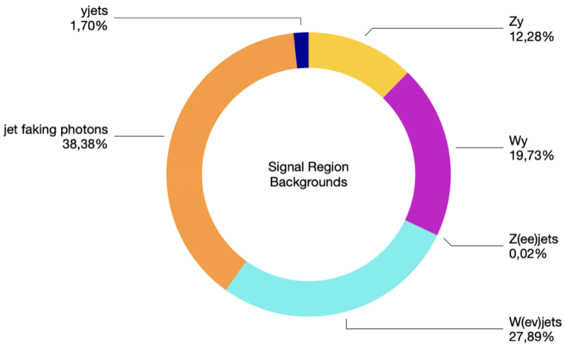
K-factors have been computed in different **transverse mass** bins...

m_T (GeV)	K	σ_K^{tot}	%
80-110	0.869	0.150	17.3 %
110-140	0.939	0.141	15.0 %
140-200	1.023	0.138	13.5 %
>200	1.089	0.208	19.1 %

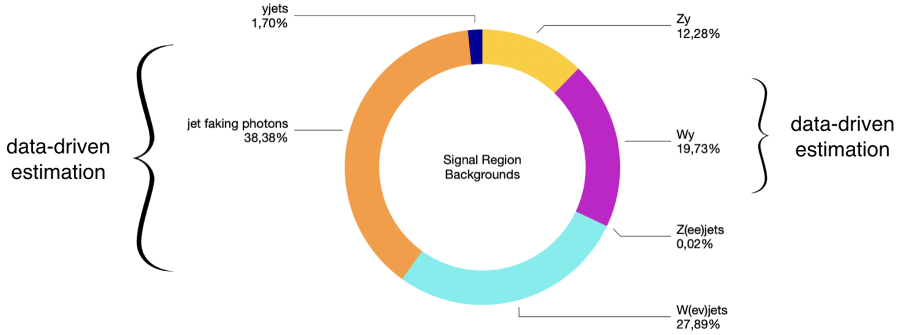
..and then **applied** to Monte Carlo in Signal Region.



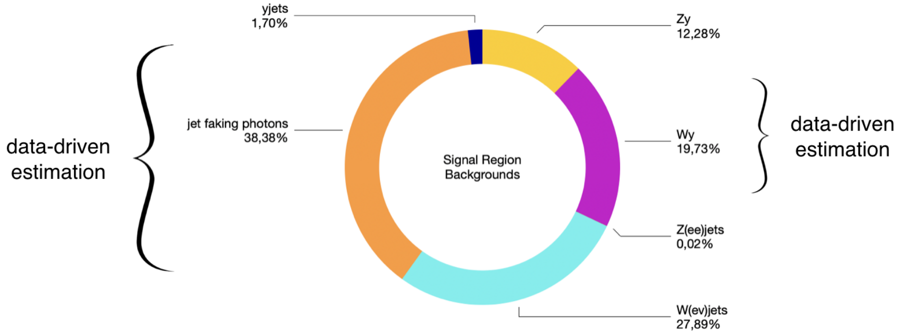
Backgrounds in Signal Region



Backgrounds in Signal Region

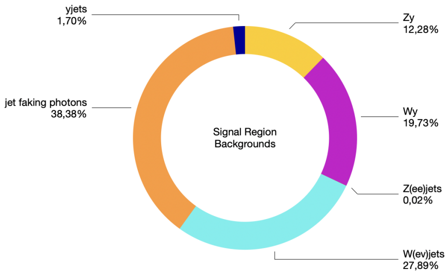


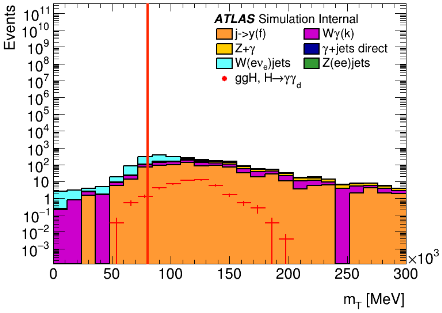
Backgrounds in Signal Region



Jets faking photons and $W\gamma$ constitute $\sim 60\%$ of the total background in Signal Region.

Backgrounds in Signal Region

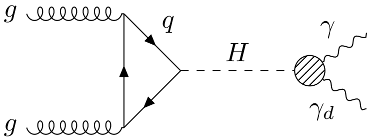




Jets faking photons and $W\gamma$ constitute $\sim 60\%$ of the total background in Signal Region.

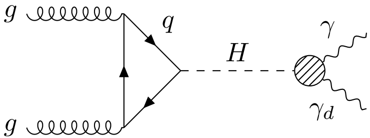
Conclusions

- I contributed to the backgrounds estimation in the ATLAS search for dark photons;



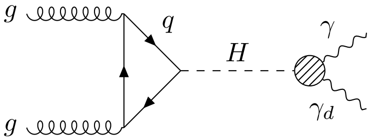
Conclusions

- I contributed to the backgrounds estimation in the ATLAS search for dark photons;
- I developed a new data-driven technique for the jets faking photons estimation;



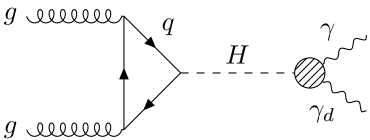
Conclusions

- I contributed to the backgrounds estimation in the ATLAS search for dark photons;
- I developed a new data-driven technique for the jets faking photons estimation;
- I estimated the $W\gamma$ background, using the k-factors method;



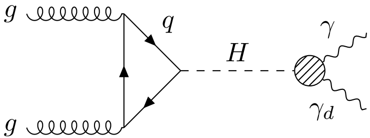
Conclusions

- I contributed to the backgrounds estimation in the ATLAS search for dark photons;
- I developed a new data-driven technique for the jets faking photons estimation;
- I estimated the $W\gamma$ background, using the k-factors method;
- Jets faking photons and $W\gamma$ constitute $\sim$ 60% of the total background in Signal Region;



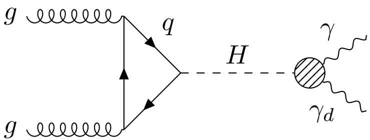
Conclusions

- I contributed to the backgrounds estimation in the ATLAS search for dark photons;
- I developed a new data-driven technique for the jets faking photons estimation;
- I estimated the $W\gamma$ background, using the k-factors method;
- Jets faking photons and $W\gamma$ constitute $\sim 60\%$ of the total background in Signal Region;
- The analysis $gg \to H \to \gamma\gamma_d$ is on-going;



Conclusions

- I contributed to the backgrounds estimation in the ATLAS search for dark photons;
- I developed a new data-driven technique for the jets faking photons estimation;
- I estimated the $W\gamma$ background, using the k-factors method;
- Jets faking photons and $W\gamma$ constitute $\sim 60\%$ of the total background in Signal Region;
- The analysis $gg \rightarrow H \rightarrow \gamma\gamma_d$ is on-going;
- These results will enter the official ATLAS analysis publication.



Backup

Dark Matter, Dark Sector, Dark Photon

DM candidates

DM candidates should be:

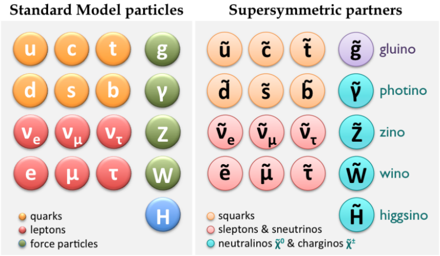
- neutral;
- cold, non-relativistic at the time of CMB formation;
- stable or at least with lifetime longer than the age of the Universe;
- weakly interacting with themselves and with ordinary matter.

Some candidates:

- WIMPs, Weakly Interacting Massive Particles, e.g. SUSY;
- sterile neutrinos, RH neutrinos with low mixing constant with ordinary neutrinos;
- many others...

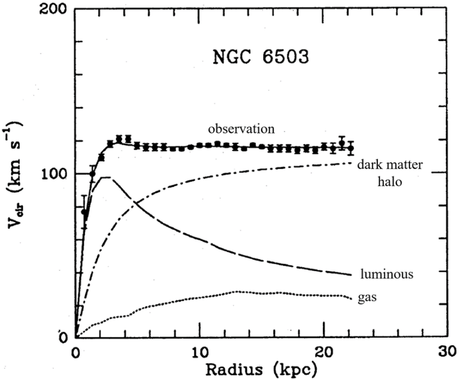
SUSY

- SUSY theory introduced to explation the difference between the measured value of the Higgs boson mass and the one predicted by the first order calculation, including the top annihilation term;
- particles with spin differing by half a unit with respect to the SM;
- s-top annihilation term would compensate the top annihilation term;
- viable DM candidates: neutralinos.

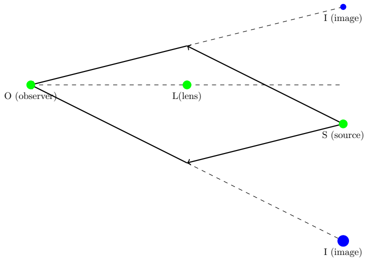


DM evidences

- Rotational curves of galaxies



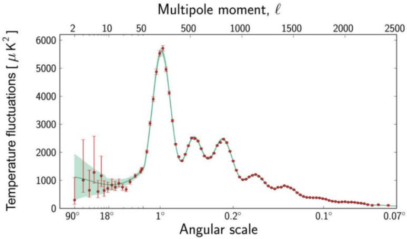
- Gravitational lensing



DM evidences

- Bullet cluster

- CMB spectrum



A portal to the dark sector

The portal and can take various forms:

- vector portal $\Longrightarrow$ massive dark photon

$$\mathcal{L}_{kin.mix.} = \frac{1}{2}\varepsilon F_{\mu\nu} F'^{\mu\nu}$$

  where F, F' are field strength tensors of the SM U(1) and the dark U(1)$_D$. For a massless dark photon, the direct kinetic mixing is not possible. There should be "something" in between.

- scalar portal;
- neutrino portal.

Analysis

Luminosity

Data used in this work of thesis are normalized at:

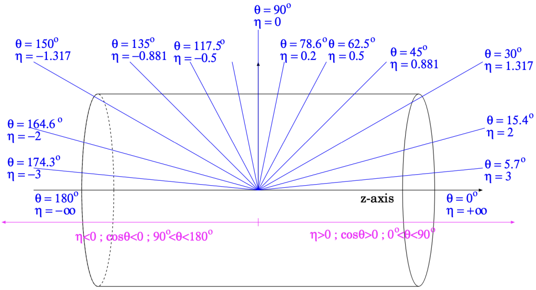
$$\mathcal{L} = 25{,}76\,\mathrm{fb}^{-1}$$

Normalizing data with full Run3 luminosity (expected $\sim 300\,\mathrm{fb}^{-1}$) will increase the statistic and reduce statistical uncertainties.

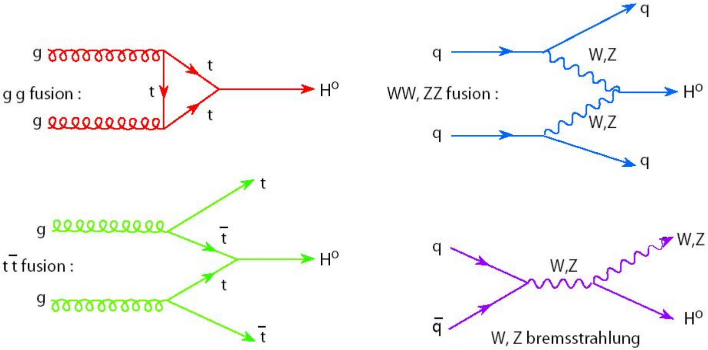
Pseudorapidity

$$\eta = -\ln\left[\tan\frac{\theta}{2}\right]$$

where θ is the polar angle. This formula set a one-to-one correspondence between the θ coordinate of a polar system and η, moving domain from $(0, \pi)$ to $(-\infty, +\infty)$.
$\Delta\eta$ is invariant under Lorentzian boosts along beam axis; this becomes important as the reference frame of the center-of-mass of the interaction is unknown.



Higgs production channels



Signal Region

All the cuts defining Signal Region

- $n_e = 0$, electrons veto;
- $n_\mu = 0$, muon veto;
- $n_\tau = 0$, tau lepton veto;
- $p_T^{miss} > 100\,\text{GeV}$;
- $n_\gamma^{isol} = 1$, one isolated photon;
- $p_T^\gamma > 50\,\text{GeV}$;
- $m_T > 80\,\text{GeV}$;
- $n_{jet} \leq 3$, maximum 3 jets;
- $\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_\gamma) \geq 1.25$, γ, γ_d in the transverse region;
- $S_{p_T^{miss}} > 6$, in order to remove fake p_T^{miss};
- $\Delta p_T^{miss} > -10\,\text{GeV}$
- $|\eta_\gamma| < 1.75$, γ, γ_d in the transverse region;;
- $\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_j) \leq 0.75$, the Higgs boson should scatter on the jets;
- $\Delta\Phi(\vec{p}_T^{j1}, \vec{p}_T^{j2}) \leq 2.5$, in order to remove dijets.

Δp_T^{miss} variable

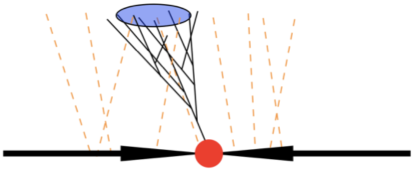
$$\Delta|\vec{p}_T^{miss}| = |[\vec{p}_T^{miss}]_{noJVT}| - |\vec{p}_T^{miss}|$$

This cut targets γ+jets background events.
As the **hard scattering event vertex** is chosen by picking the one with the highest scalar sum of the momenta of all the tracks produced in it, in a γ+jets event, where a big portion of momentum is carried away by the photon, there is a non-negligible probability to elect a pile-up vertex as hard scattering vertex.
If JVT cut is applied in such a case, this will lead to **exclude the real jet** from $\vec{p}_T^{miss}$ calculation, hence resulting in a large fake missing transverse momentum in the final state.
Events where the p_T^{miss} calculated with JVT is much higher than the p_T^{miss} calculated with JVT are in most of the cases events with a mis-reconstructed primary vertex and can be then **excluded**.

Jet Vertex Tag (JVT)

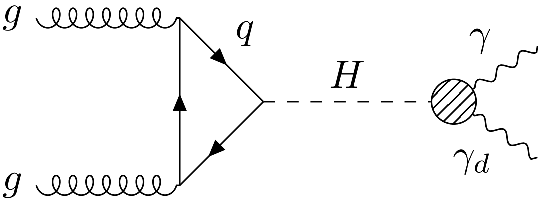
$$JVT = \frac{\sum_{j \in hard\ scattering} p_T^j}{\sum_j p_T^j}$$

$\Longrightarrow$ reject the jet if JVT is under a certain threshold!



Analysis Trigger

- 1 tight photon, $N_\gamma = 1$;
- photon transverse momentum $|\vec{p}_T^{\gamma}| > 50\,\text{GeV}$;
- missing transverse momentum $|\vec{p}_T^{miss}| > 40\,\text{GeV}$ with calculation based on cells and $|\vec{p}_T^{miss}| > 70\,\text{GeV}$ with calculation including tracks;
- transverse mass $m_T > 80\,\text{GeV}$.



Backgrounds estimation strategy

- irreducible background: $Z(\to \nu\nu)\gamma$;
- lost lepton: $W(\to l\nu_l)\gamma$;
- jets faking photons: multijets, Zjets, Wjets;
- electrons faking photons: $W(\to e\nu_e)$jets, $Z(\to ee)$jets;
- fake p_T^{miss}: multijets, γjets.

Strategy 1

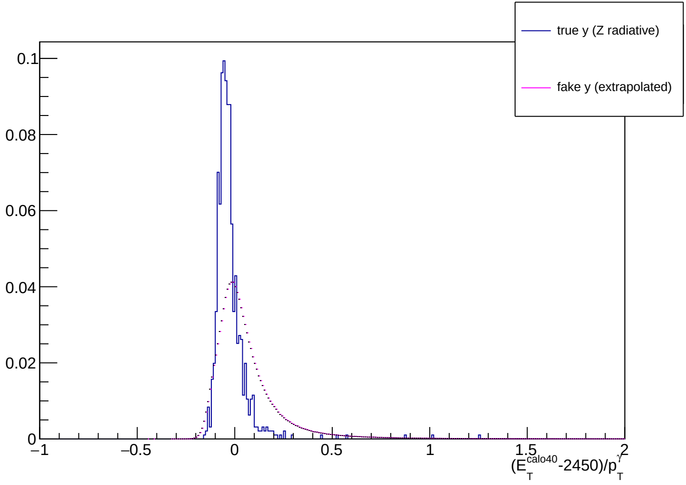
- electrons faking photons data-driven
- jets faking photons data-driven
- $W\gamma, Z\gamma$ from leptons CR
- γjets Monte Carlo

Strategy 2

- electrons faking photons data-driven
- fake p_T^{miss} data-driven
- $W\gamma + W$jets, $Z\gamma + Z$jets from leptons CR

Jets faking photons

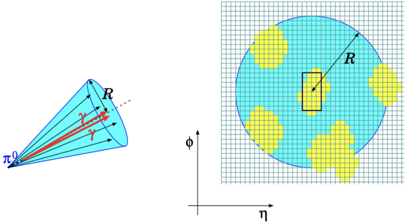
True and fake photons isolation comparison

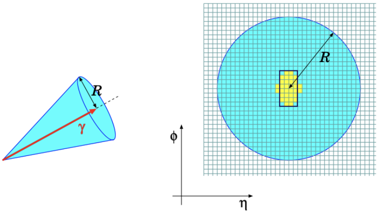


Calorimeter relative isolation

$$isol^{rel}_{calo} = \frac{E_T^{calo40} - 2450}{p_T^{\gamma}}$$

where
p_T^{γ} is the photon transverse momentum;
E_T^{calo40} is the energy not belonging to the photon measured in a cone with radius $\Delta R = 0.4$ around the photon;
2,45 GeV is a pedestal factor.





Purity

The fraction of true photons in the Non-Isolated Region is given by the **purity** P.

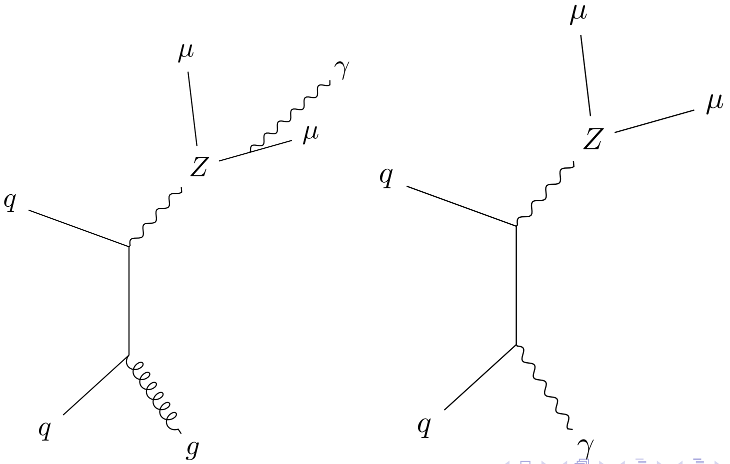
$$P = \left(\frac{N_{\gamma}^{non-isol}}{N^{non-isol}}\right)_{tight} = \left(\frac{N^{non-isol} - N_{j \rightsquigarrow \gamma}^{non-isol}}{N^{non-isol}}\right)_{tight} = \left(1 - \frac{N_{j \rightsquigarrow \gamma}^{non-isol}}{N^{non-isol}}\right)_{tight}$$

depends on the number of non isolated photons produced in the event.

It is not an intrinsic property of how we "see" jets!
We would like to have $P \sim 0$, i.e. no contamination of true photons in the Non-Isolated Region.

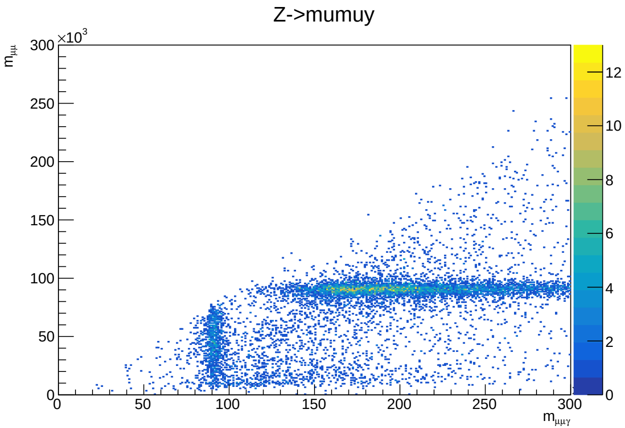
Non-Isolated Region definition

Let's define the Non-Isolated Region such to have $P \sim 0$, i.e. no contamination of true photons in the Non-Isolated Region.
Let's look at a **pure** sample of photons, that can be obtained selecting $Z(\to \mu\mu\gamma)$ events.



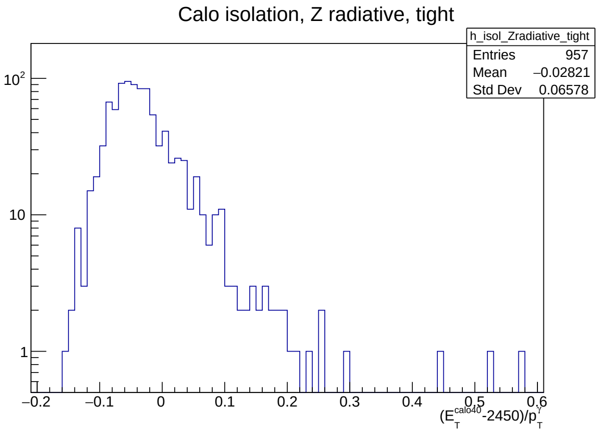
Non-Isolated Region definition

In order to select $Z(\rightarrow \mu\mu\gamma)$ events, let's consider events in the $\mu\mu\gamma$ sample with a tight photon and $80\,\text{GeV} < m_{\mu\mu\gamma} < 100\,\text{GeV}$.



Non-Isolated Region definition

Looking at the calorimeter relative isolation of these events, we decide to put the cut of the Non-Isolated Region at 0.1.



Photons ID

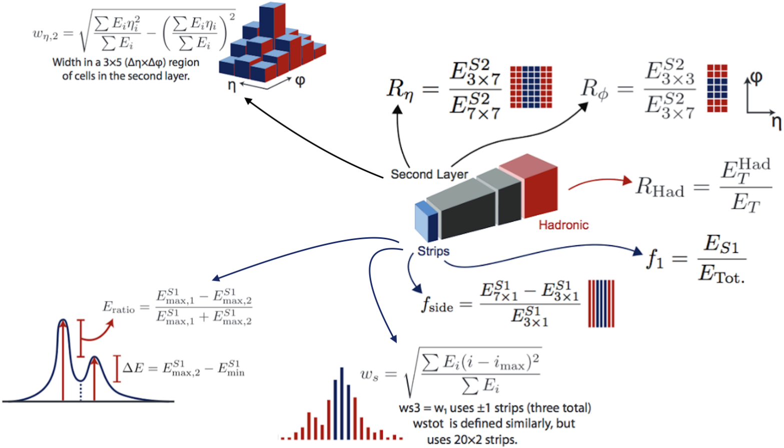
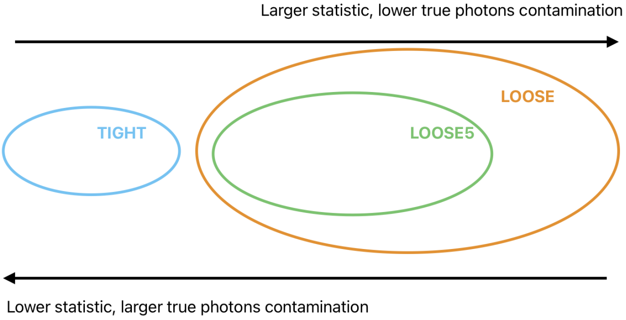


Figure 3: Discriminant Variables (DVs) describing shower shapes, energy ratios and width of the energy deposit

Photons ID

Different possibile **ID selection**:

- **tight**, passing tight cuts on all the DVs;
- **loose**, if they pass looser cuts on some DVs (R_η, R_{had} and $w_{\eta,2}$) but not the tight ones;
- **loose5**, if they are loose and pass tight cuts on more DVs (R_η, R_{had}, $w_{\eta,2}$ and R_ϕ);



Loose photons

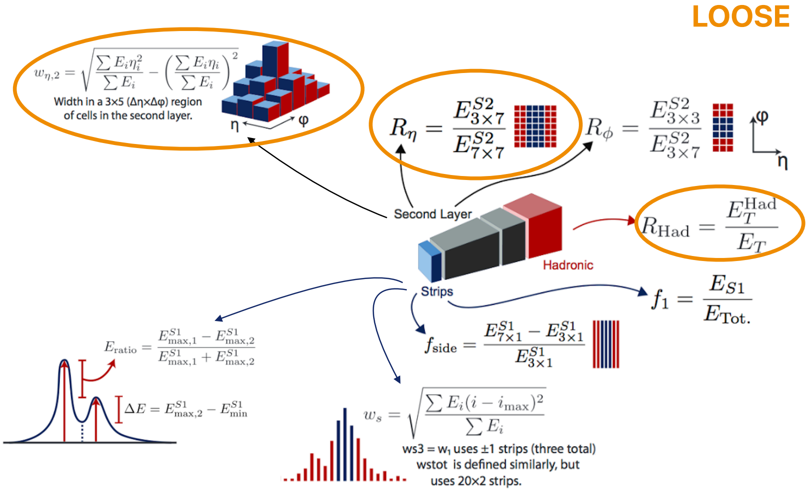


Figure 4: Discriminant Variables (DVs) describing shower shapes, energy ratios and width of the energy deposit (loose)

Loose5 photons

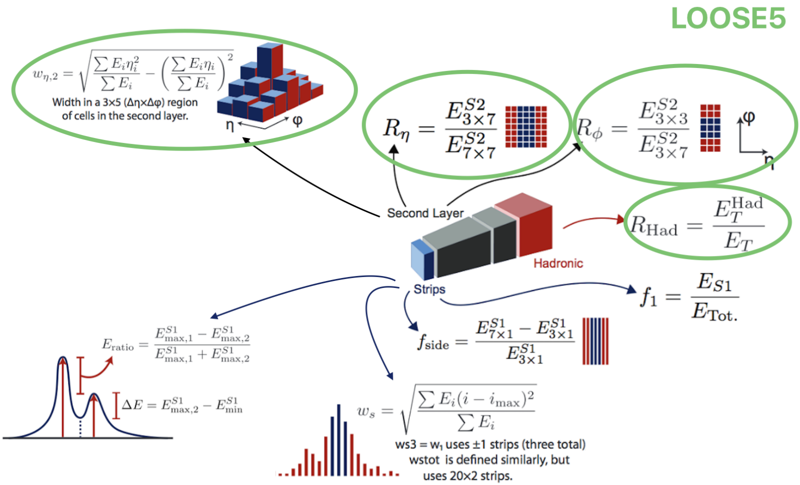


Figure 5: Discriminant Variables (DVs) describing shower shapes, energy ratios and width of the energy deposit (loose5)

Step 1: get L->T transformation from MC

Let's assume that tight $isol_T^{MC}$ and loose $isol_L^{MC}$ distributions in MC are linked by an affine transformation.

μ: **median**
σ: **width**

$$isol_T^{MC} = a + b\ isol_L^{MC}$$

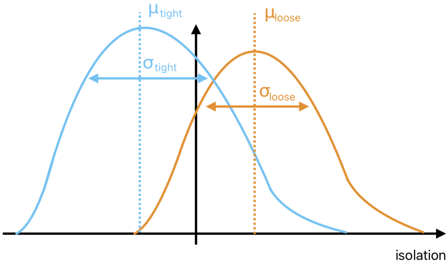
We want to find a, b such that:

$$isol_T^{MC} = \mu_T^{MC} + \frac{\sigma_T^{MC}}{\sigma_L^{MC}}(isol_L^{MC} - \mu_L^{MC})$$

so

$$a = \mu_T^{MC} - \frac{\sigma_T^{MC}}{\sigma_L^{MC}}\mu_L^{MC}$$

$$b = \frac{\sigma_T^{MC}}{\sigma_L^{MC}}$$



Step 1: get L->T transformation from MC

Let's assume:

μ: **median**
σ: **width**

- the scale factor b stays the same in MC and data;

$$\frac{\sigma_T^{data}}{\sigma_L^{data}} = \frac{\sigma_T^{MC}}{\sigma_L^{MC}}$$

- the offset a in data should depend on $\sigma_L^{data}, \sigma_T^{data}$, which is known, and on μ_T^{data}, which is unknown. So we assume the shift of the average going from loose to tight is proportional to the rms in both data and MC.

$$a = \mu_T^{MC} - \frac{\sigma_T^{MC}}{\sigma_L^{MC}} \mu_L^{MC} \longrightarrow a = \mu_T^{DATA} - \frac{\sigma_T^{DATA}}{\sigma_L^{DATA}} \mu_L^{DATA}$$

known

b

$$\frac{\mu_T^{data} - \mu_L^{data}}{\sigma_L^{data}} = \frac{\mu_T^{MC} - \mu_L^{MC}}{\sigma_L^{MC}}$$

$$\mu_T^{data} = \mu_L^{data} + \frac{\sigma_L^{data}}{\sigma_L^{MC}} (\mu_T^{MC} - \mu_L^{MC})$$

Step 2: apply L->T transformation to DATA

μ: **median**
σ: **width**

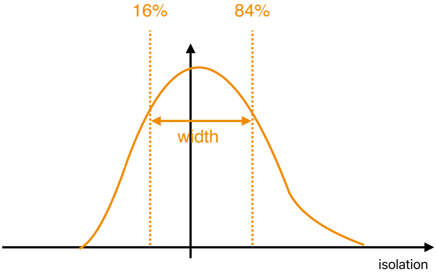
Putting all together, the transformation for DATA is:

$$isol_T^{data} = \mu_L^{data} + \frac{\sigma_L^{data}}{\sigma_L^{MC}}(\mu_T^{MC} - \mu_L^{MC}) + \frac{\sigma_T^{MC}}{\sigma_L^{MC}}(isol_L^{data} - \mu_L^{data})$$

$\rightarrow$ We obtain **tight fake photons** distributions in DATA.

Width

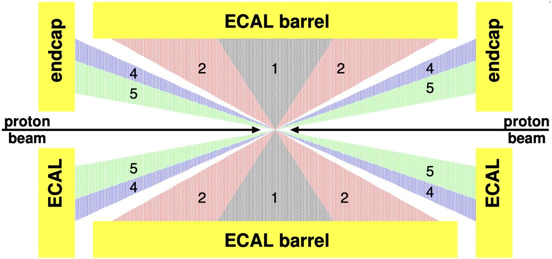
Width is calculated as difference between the 16th and 84th percentile.



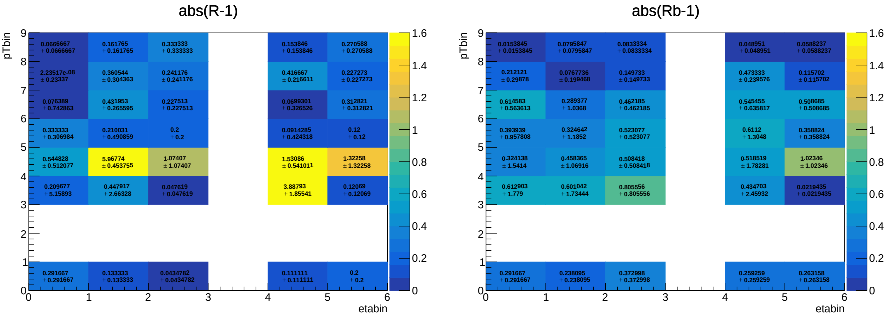
Binning in η

Pseudorapidity binning is chosen considering the detector geometry:

- Region 1: $[0; 0.6]$, the upper limit $\eta = 0.6$ is the point after which the material in front of ECAL increases a lot;
- Region 2: $[0.6; 1.37]$, the upper limit is defined by the beginning of the crack region;
- Crack region: $[1.37; 1.52]$, not used due to low reconstruction performance;
- Region 4: $[1.52; 1.81]$, the upper limit is the point where the presampler ends;
- Region 5: $[1.81; 2.37]$.



Analysis Trigger, R, R_b



(a) $R(\eta, p_T)$ (b) $R_b(\eta, p_T)$

Figure 6: $R(\eta, p_T)$ and $R_b(\eta, p_T)$ computed with $(\mu_{med}, \sigma_{q16})$ using Analysis Trigger.

Median and width

We decide to use (median&width) instead of (mean&sigma) for the extrapolation: R, R_b values in different η, p_T are indeed less spread.

Trigger	Ratio	Used variables	RMS in η, p_T bins
Analysis	R	μ, σ	0.19
		μ_{med}, σ_{q16}	0.14
	R_b	μ, σ	0.30
		μ_{med}, σ_{q16}	0.28
p_T^{miss}	R	μ, σ	0.27
		μ_{med}, σ_{q16}	0.10
	R_b	μ, σ	0.22
		μ_{med}, σ_{q16}	0.13
Leptonic	R	μ, σ	0.54
		μ_{med}, σ_{q16}	0.18
	R_b	μ, σ	0.51
		μ_{med}, σ_{q16}	0.36

Inclusive Region and Trigger

Problem: R and R_b are quite unstable. It is not possible to perform the extrapolation in exclusive regions in p_T, η.

Solution: Let's be either inclusive in p_T or in η.

Inclusive Region and Trigger

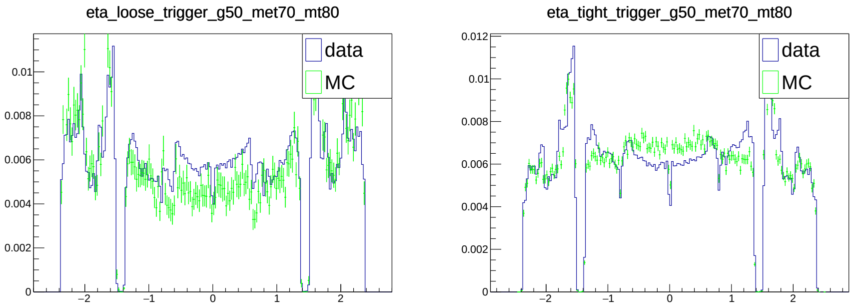
Problem: How to choose the Inclusive Region and the Trigger to be used?

Trigger → / Inclusivity ↓	Analysis	p_T^{miss}	Leptonic
p_T			
η			

Solution: Let's choose the configuration satisfying the requirements:

- the spectrum of the inclusive variable in data and MC should be similar;
- the configuration should ensure the lowest uncertainties on the fake factors.

η, p_T spectra in data and MC

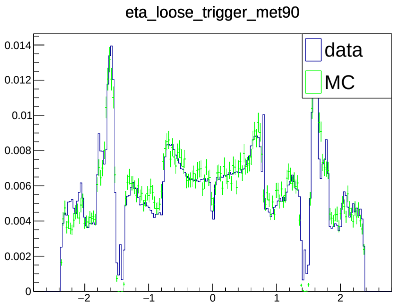


(a) Loose (b) Tight

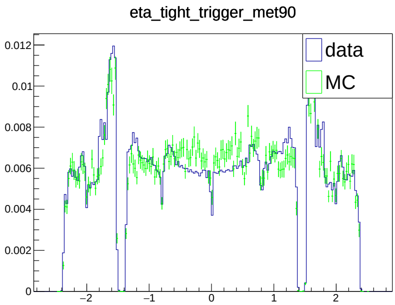
Figure 7: η distribution for loose and tight γ in data and MC samples (Analysis Trigger).

Trigger / Inclusivity	Analysis	p_T^{miss}	Leptonic
p_T			
η			

$\implies$ good agreement ✓



(a) Loose



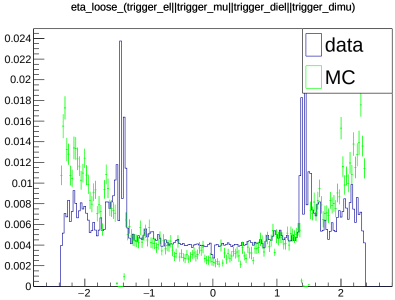
(b) Tight

Figure 8: η distribution for loose and tight γ in data and MC samples (p_T^{miss} Trigger).

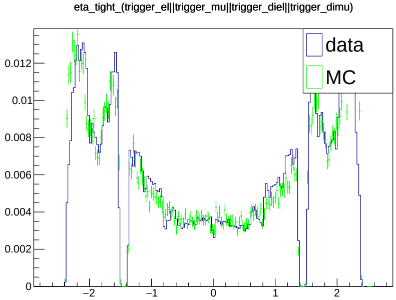
Trigger → / Inclusivity ↓	Analysis	p_T^{miss}	Leptonic
p_T			
η		■	

$\Longrightarrow$ good agreement ✔

η, p_T spectra in data and MC



(a) Loose



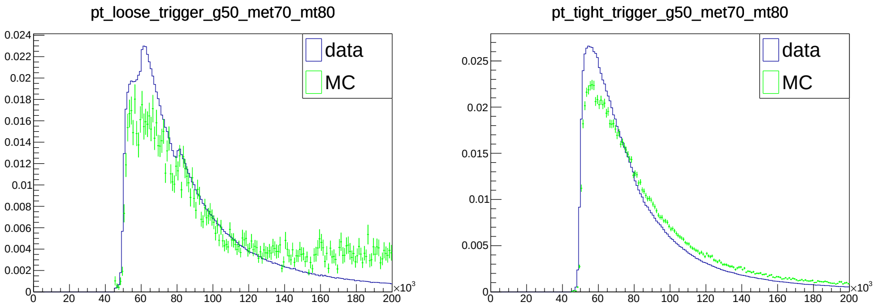
(b) Tight

Figure 9: η distribution for loose and tight γ in data and MC samples (Leptonic Trigger).

Trigger → / Inclusivity ↓	Analysis	p_T^{miss}	Leptonic
p_T			
η			

$\Longrightarrow$ bad agreement ✘

η, p_T spectra in data and MC

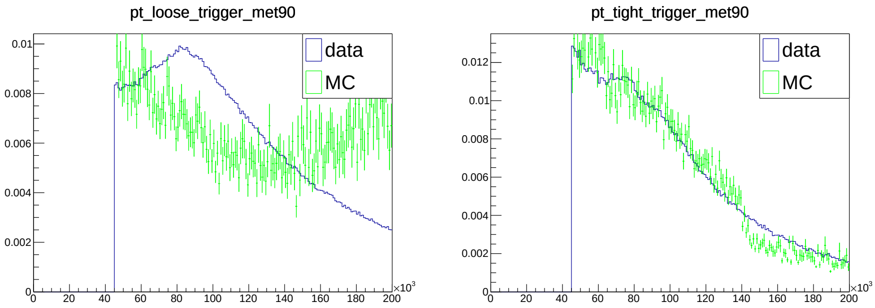


(a) Loose (b) Tight

Figure 10: p_T distribution for loose and tight γ in data and MC (Analysis Trigger).

Trigger → / ↓ Inclusivity	Analysis	p_T^{miss}	Leptonic
p_T			
η			

$\Longrightarrow$ good agreement ✔



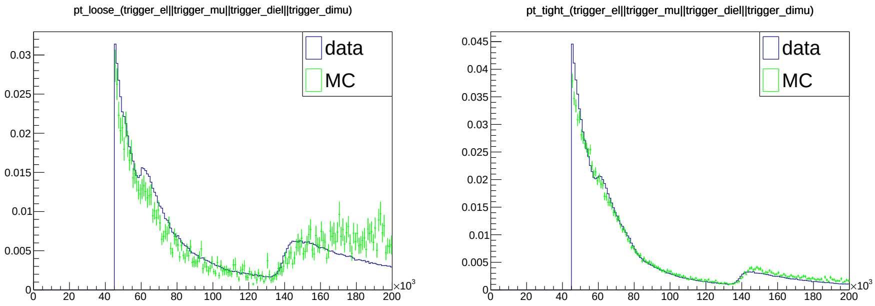
(a) Loose (b) Tight

Figure 11: p_T distribution for loose and tight γ in data and MC (p_T^{miss} Trigger).

Trigger → / ↓ Inclusivity	Analysis	p_T^{miss}	Leptonic
p_T		■	
η			

$\Longrightarrow$ bad agreement ✘

η, p_T spectra in data and MC

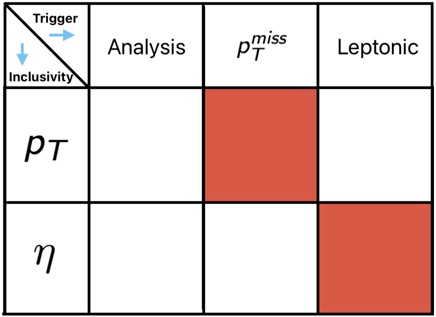


(a) Loose (b) Tight

Figure 12: p_T distribution for loose and tight γ in data and MC (Leptonic Trigger).

Trigger → / Inclusivity ↓	Analysis	p_T^{miss}	Leptonic
p_T			
η			

$\Longrightarrow$ good agreement ✔


Trigger
Inclusivity
Analysis
p_T^{miss}
Leptonic
p_T
η

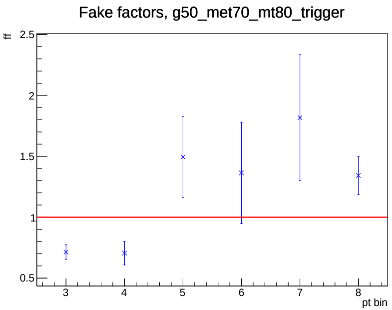
R, R_b in different configurations

Trigger	Incl. Region	R_{nom}	R_{up}	R_{down}	$R_{b,nom}$	$R_{b,up}$	$R_{b,down}$
analysis	η	1.10	1.20	1.00	1.36	1.72	1.00
	p_T	1.06	1.12	1.00	1.28	1.56	1.00
p_T^{miss}	η	0.82	1.00	0.64	1.06	1.12	1.00
leptonic	p_T	0.79	1.00	0.58	1.02	1.04	1.00

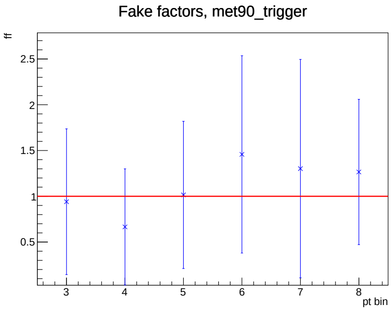
Table 1: Nominal and varied values of R and R_b for the different configurations of inclusive region and trigger possible.

Let's now extrapolate the tight fake photons isolation distributions in data using these R, R_b (nominal and varied) for each configuration.

Fake factors in different configurations



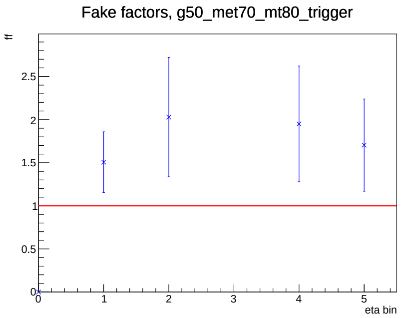
(a) Analysis Trigger, η inclusive region.



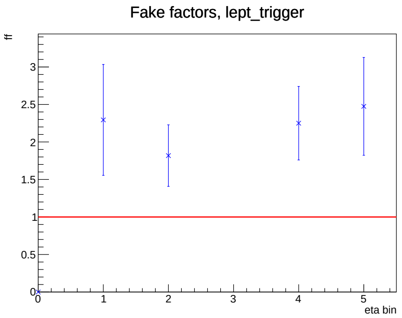
(b) p_T^{miss} Trigger, η inclusive region.

Figure 13: Fake factors stability check for p_T^{miss} Trigger in the η inclusive region (right) and Analysis Trigger in the η inclusive region (left) using nominal R, R_b.

Fake factors in different configurations



(a) Analysis Trigger, p_T inclusive region.



(b) Leptonic Trigger, p_T inclusive region.

Figure 14: Fake factors stability check for Leptonic Trigger in the p_T inclusive region (right) and Analysis Trigger in the p_T inclusive region (left) using nominal R, R_b.

Uncertainties on fake factors

We can now calculate the uncertainties on these fake factors as:

$$\sigma_{ff} = \sqrt{\sigma_{R_b}^2 + \sigma_R^2}$$

where

$$\sigma_{R_b} = \left(\frac{ff(R_{nom}, R_{b,up}) - ff(R_{nom}, R_{b,down})}{2}\right)$$

$$\sigma_R = \left(\frac{ff(R_{up}, R_{nom}) - ff(R_{down}, R_{nom})}{2}\right)$$

Fake factors in different configurations

η bin	ff	σ_{ff}	%
1	1.51	0.35	23.4 %
2	2.03	0.69	34.1 %
4	1.95	0.67	34.4 %
5	1.70	0.54	31.4 %

η bin	ff	σ_{ff}	%
1	2.29	1.48	64.4 %
2	1.82	0.82	45.1 %
4	2.25	0.98	43.6 %
5	2.47	1.30	52.7 %

Table 2: Fake factors in p_T incl region using Analysis(l) and Leptonic Trigger(r).

p_T bin	ff	σ_{ff}	%
3	0.71	0.12	17.4 %
4	0.71	0.19	27.6 %
5	1.50	0.67	44.6 %
6	1.36	0.83	61.1 %
7	1.82	1.03	56.9 %
8	1.34	0.31	23.4 %

p_T bin	ff	σ_{ff}	%
3	0.94	0.80	84.6 %
4	0.67	0.63	95.3 %
5	1.02	0.80	79.2 %
6	1.46	1.08	73.9 %
7	1.30	1.19	91.6 %
8	1.27	0.79	62.7 %

Table 3: Fake factors in the η incl region using Analysis(l) and p_T^{miss} Trigger(r).

Applying fake factors to CR

We extrapolate the number of jets faking photons in SR $N^{SR}_{\text{ext}}(ff)$ and compute its uncertainty from the fake factors uncertainty σ_N

$$\sigma_N = \frac{N(ff + \sigma_{ff}) - N(ff - \sigma_{ff})}{2}$$

Applying fake factors to CR

Jet faking photons extrapolated with nominal/varied fake factors

	$N^{SR}_{\text{ext}}(ff) \pm \sigma_N$	$N^{SR}_{\text{ext}}(ff - \sigma_{ff})$	$N^{SR}_{\text{ext}}(ff + \sigma_{ff})$
all	9.2e+06 $\pm 1.5e+06$	1.0745e+07	7.7145e+06
$n_e = 0$	9.2e+06 $\pm 1.5e+06$	1.0745e+07	7.7145e+06
$n_\mu = 0$	9.2e+06 $\pm 1.5e+06$	1.0745e+07	7.7145e+06
$p_T^{miss} > 100$ GeV	1.4e+06 $\pm 2.2e+05$	1.6335e+06	1.1838e+06
$n_\gamma^{nonisol}$	2.7e+05$\pm 4.2e+04$	3.1206e+05	2.2676e+05
$p_T^\gamma > 50$ Gev	2.6e+05 $\pm 4.2e+04$	3.0792e+05	2.2377e+05
n_{jet}	2.1e+05 $\pm 3.3e+04$	2.4302e+05	1.7653e+05
$m_T > 80$ GeV	2.1e+05 $\pm 3.3e+04$	2.4175e+05	1.756e+05
$\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_\gamma) \geq 1.25$	8146 ± 1291	9440	6857
$S_{p_T^{miss}} > 6$	1695 ± 267	1963	1429
$\Delta p_T^{miss} > -10$ GeV	1311 ± 206	1518	1105
$\lvert\eta_\gamma\rvert < \leq 1.75$	1022 ± 160	1183	863
$\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_j) \leq 0.75$	926 ± 145	1073	782
$\Delta\Phi(\vec{p}_T^{j1}, \vec{p}_T^{j2}) \leq 2.5$	775 ± 116	897	655

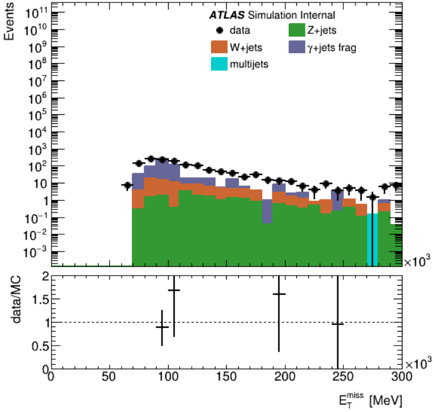
Applying fake factors to CR

Monte Carlo samples in Signal Region

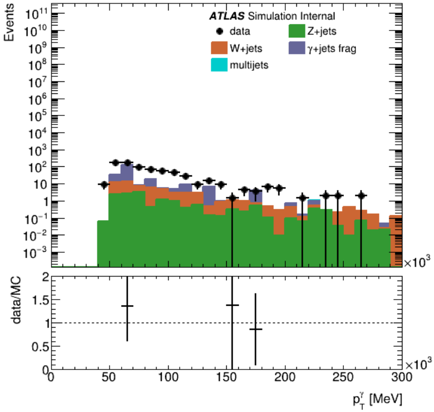
	multijets	γ+jets(f)	W+jets	Z+jets	sum
all	6.2709e+08	1.9313e+06	21836	13044	6.2906e+08
$n_{el} = 0$	6.2709e+08	1.9313e+06	21836	13044	6.2906e+08
$n_{\mu} = 0$	6.2709e+08	1.9313e+06	21836	13044	6.2906e+08
$p_T^{miss} > 100$ GeV	3.691e+08	2.3888e+05	11572	8612	3.6936e+08
$n_{\gamma}^{isol} = 1$	5.3566e+07	1.1368e+05	1282	701.51	5.3682e+07
$p_T^{\gamma} > 50$ GeV	5.3566e+07	1.1248e+05	1276	699.92	5.3681e+07
$n_{jet} < 4$	3.8745e+05	82875	1112	630.72	4.7207e+05
$m_T > 80$ GeV	3.8736e+05	82259	1061	626.18	4.713e+05
$\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_{\gamma}) \geq 1.25$	2552.8	5230	133.89	43.494	7959.9
$S_{p_T^{miss}} > 6$	344.28	1001	104.31	36.287	1486.2
$\Delta p_T^{miss} > -10$ GeV	215.74	591.05	95.261	34.069	936.12
$\lvert\eta_{\gamma}\rvert < 1.75$	165.45	383.81	71.445	28.204	648.91
$\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_{\text{jet}}) \leq 0.75$	137.74	338.96	63.101	23.481	563.28
$\Delta\Phi(p_T^{j1}, p_T^{j2}) \leq 2.5$	0.6921	166.28	56.848	22.082	245.91

Applying fake factors to CR

Comparison between MC and data-driven estimation

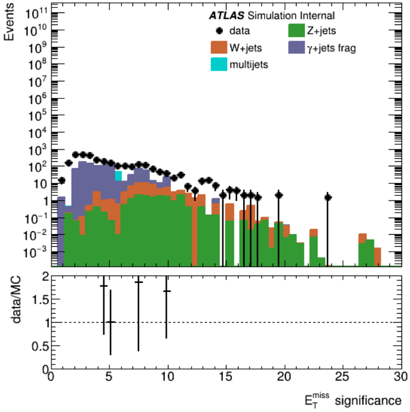


(a) p_T^{miss}

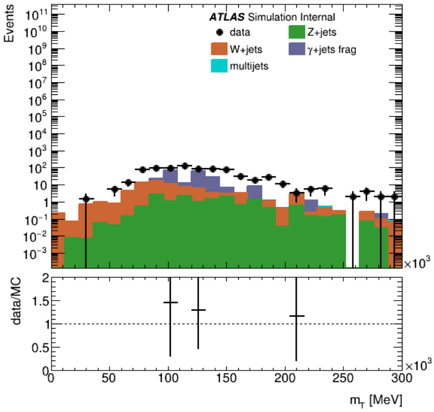


(b) p_T^{γ}

Comparison between MC and data-driven estimation

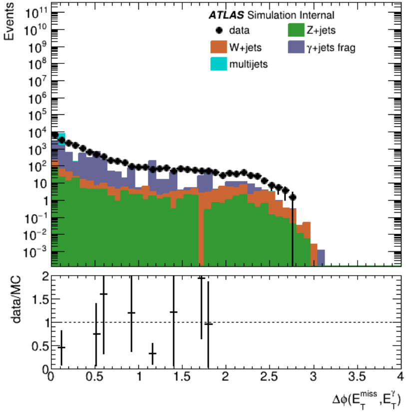


(a) $S_{p_T^{miss}}$

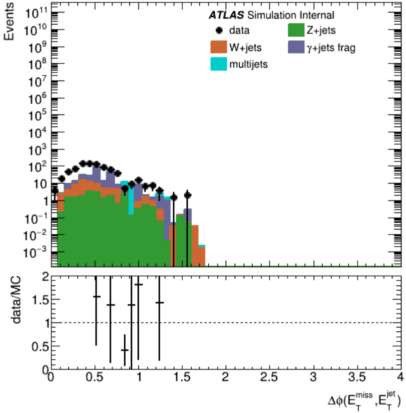


(b) m_T

Comparison between MC and data-driven estimation

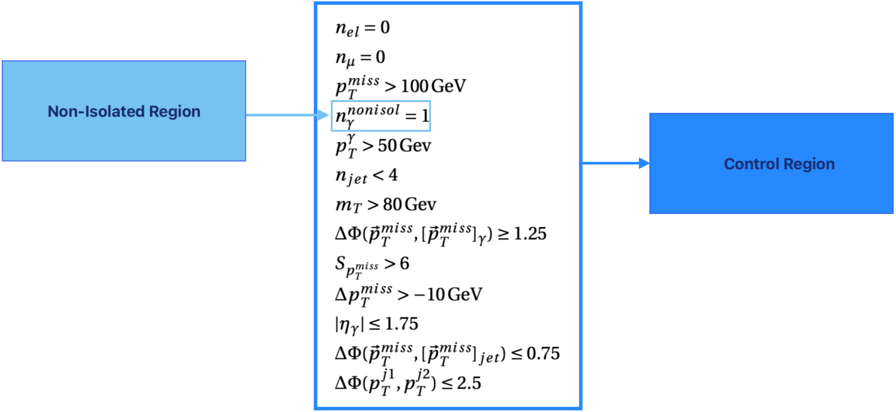


(a) $\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_\gamma)$

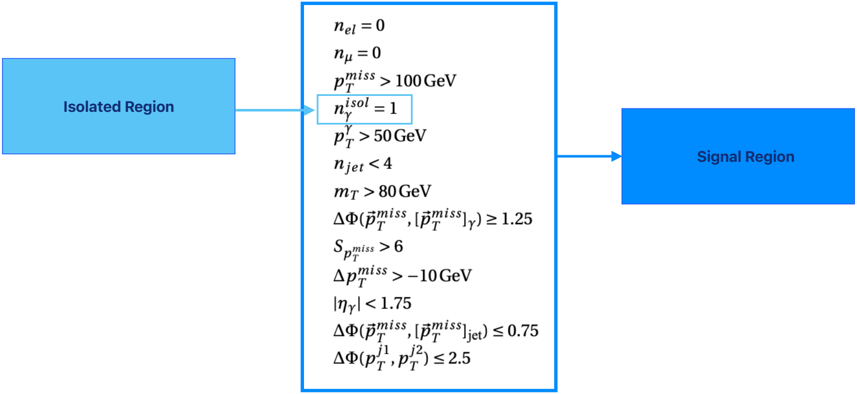


(b) $\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_{jet})$

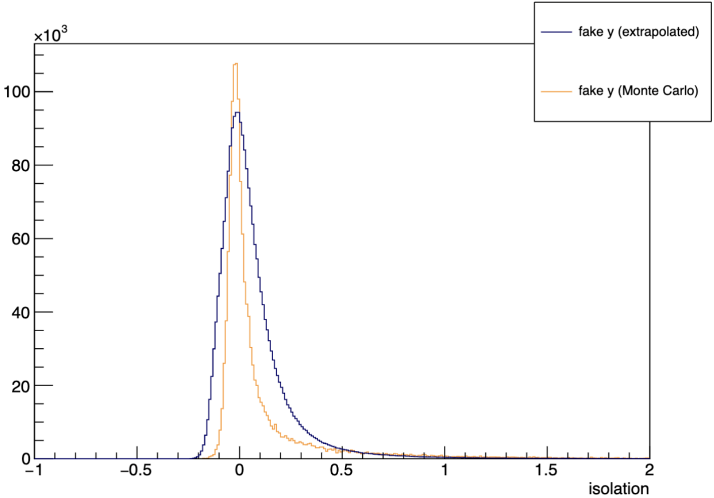
Non-Isolated Control Region



Signal Region



Comparison between extrapolated isolation distribution and MC isolation distribution



Wγ

1μ Control Region

Signal Region

- $N_\gamma^{isol} = 1$;
- $N_e = 0$;
- $N_\mu = 0$;
- $|\vec{p}_T^{miss}| > 100$ GeV;
- $|\vec{p}_T^{\gamma}| > 50$ GeV;
- $N_{jets} \leq 3$;
- $m_T > 80$ GeV;
- $\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_\gamma) \geq 1.25$;
- $S_{p_T^{miss}} > 6$;
- $\Delta|\vec{p}_T^{miss}| > -10$ GeV;
- $\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_{jet}) \leq 0.75$;
- $\Delta\Phi(\vec{p}_T^{j1}, \vec{p}_T^{j2}) \leq 2.5$.

1μ Control Region

- $N_\gamma^{isol} = 1$;
- $N_e = 0$;
- $N_\mu = 1$;
- $|[\vec{p}_T^{miss}]_{no\ \mu}| > 100$ GeV;
- $|\vec{p}_T^{\gamma}| > 50$ GeV;
- $N_{jets} \leq 3$;
- $[m_T]_{no\ \mu} > 80$ GeV;
- $\Delta\Phi([\vec{p}_T^{miss}]_{no\ \mu}, [\vec{p}_T^{miss}]_\gamma) \geq 1.25$;
- $S_{[p_T^{miss}]_{no\ \mu}} > 6$;
- $\Delta|[\vec{p}_T^{miss}]_{no\ \mu}| > -10$ GeV;
- $\Delta\Phi([\vec{p}_T^{miss}]_{no\ \mu}, [\vec{p}_T^{miss}]_{jet}) \leq 0.75$;
- $\Delta\Phi(\vec{p}_T^{j1}, \vec{p}_T^{j2}) \leq 2.5$.

K-factors calculation

We would like to calculate a K-factor as ratio between the number of events in data and Monte Carlo in the Control Region, to **correct** the imperfections of the simulations.

K-factors calculation

We would like to calculate a K-factor as ratio between the number of events in data and Monte Carlo in the Control Region, to **correct** the imperfections of the simulations.

The K-factor will be later applied to Monte-Carlo in the Signal Region to get the **final estimation** of the background yield in Signal Region.

K-factors calculation

We would like to calculate a K-factor as ratio between the number of events in data and Monte Carlo in the Control Region, to **correct** the imperfections of the simulations.

The K-factor will be later applied to Monte-Carlo in the Signal Region to get the **final estimation** of the background yield in Signal Region.

We calculated K-factors in 3 **different ways**:

- for all the Monte Carlo samples (approach 1);
- for $W\gamma$+Wjets sample (approach 2);
- for $W\gamma$ sample only, using a data-driven estimation of Wjets events (approach 3).
  $\rightarrow$ Fake factors computed as described in the last meeting (link[1]) were used!

[1] https://indico.cern.ch/event/1420907/contributions/5975099/attachments/2864837/5013945/JetsFakingPhotons.pdf

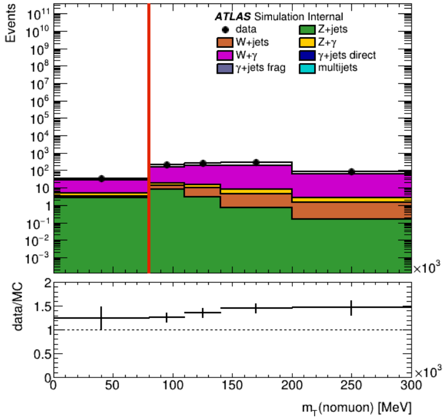
K-factors calculation

K-factors are calculated in bins of transverse mass m_T, a good candidate to be a discriminating variable.

- bin 1: $[m_T]_{no\mu} \in [0, 80]$GeV $\rightarrow$ cut on m_T defining the CR;
- bin 2: $[m_T]_{no\mu} \in [80, 110]$GeV;
- bin 3: $[m_T]_{no\mu} \in [110, 140]$GeV $\rightarrow$ centred on m_H, expected for signal;
- bin 4: $[m_T]_{no\mu} \in [140, 200]$GeV;
- bin 5: $[m_T]_{no\mu} > 200$ GeV.

Approach 1: K-factors as data/MC

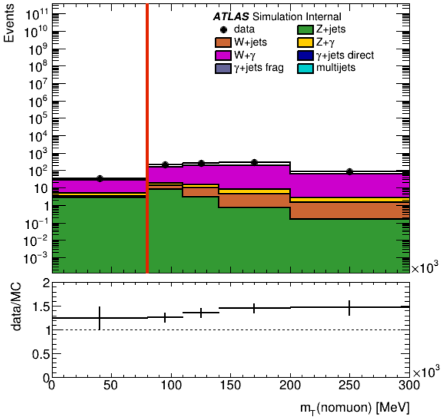
$$K(m_T) = \frac{N^{1\mu CR}_{data}(m_T)}{N^{1\mu CR}_{MC}(m_T)}$$



Approach 1: K-factors as data/MC

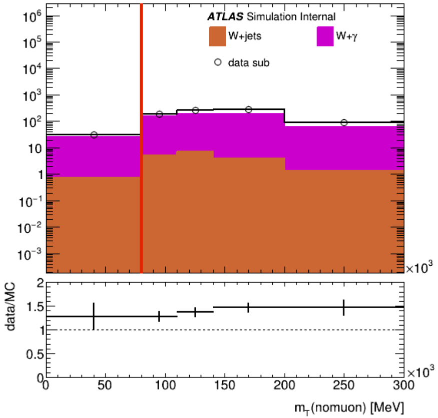
$$K(m_T) = \frac{N^{1\mu CR}_{data}(m_T)}{N^{1\mu CR}_{MC}(m_T)}$$

m_T (GeV)	K	σ_K^{stat}
0-80	1.240	0.244
80-110	1.259	0.100
110-140	1.361	0.097
140-200	1.459	0.096
>200	1.465	0.156



Approach 2: K-factors for $W\gamma + W$jets

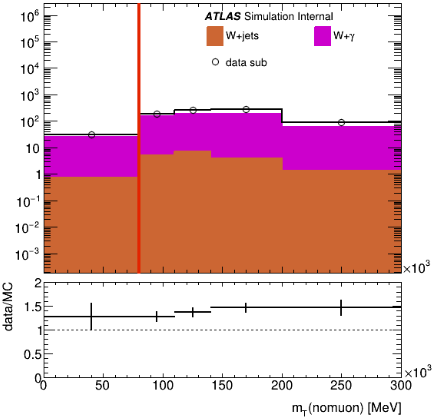
$$K(m_T) = \frac{[N^{1\mu CR}_{data} - N^{1\mu CR}_{Zjets} - N^{1\mu CR}_{Z\gamma} - N^{1\mu CR}_{\gamma jets} - N^{1\mu CR}_{multijets}](m_T)}{[N^{1\mu CR}_{W\gamma} + N^{1\mu CR}_{Wjets}](m_T)}$$



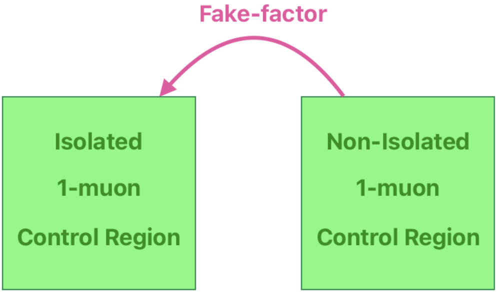
Approach 2: K-factors for $W\gamma$+Wjets

$$K(m_T) = \frac{[N^{1\mu CR}_{data} - N^{1\mu CR}_{Zjets} - N^{1\mu CR}_{Z\gamma} - N^{1\mu CR}_{\gamma jets} - N^{1\mu CR}_{multijets}](m_T)}{[N^{1\mu CR}_{W\gamma} + N^{1\mu CR}_{Wjets}](m_T)}$$

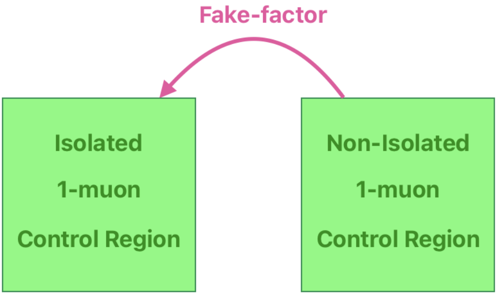
m_T (GeV)	K	σ_K^{stat}
0-80	1.279	0.287
80-110	1.282	0.109
110-140	1.376	0.101
140-200	1.473	0.099
>200	1.476	0.160



Jets faking photons in the 1 μ Control Region



Jets faking photons in the 1 μ Control Region



$$N_{dd}^{1\mu CR} = 276 \pm 84$$

Applying K-factors to MC in Signal Region

cut	$N^{SR}_{W\gamma}(K)$	$N^{SR}_{W\gamma}(K+\sigma_K)$	$N^{SR}_{W\gamma}(K-\sigma_K)$
all	28543	22835	34250
$n_e = 0$	28543	22835	34250
$n_\mu = 0$	28543	22835	34250
$n_\tau = 0$	23343	18681	28003
$p_T^{miss} > 100$ GeV	10143	8062.1	12223
n_γ^{isol}	7408.6	5882.7	8934.3
$p_T^\gamma > 50$ Gev	7387.8	5866.1	8909.5
n_{jet}	6332.2	5029.4	7635.1
$m_T > 80$ GeV	6254.6	4972.9	7536.4
$\Delta\Phi(\vec{p}_T^{miss}, [\vec{p}_T^{miss}]_\gamma) \geq 1.25$	822.81	655.64	989.99
$S_{p_T^{miss}} > 6$	711.51	566.92	856.09
$\Delta p_T^{miss} > -10$ GeV	650.97	518.67	783.28
$\|\eta_\gamma\| <\leq 1.75$	490.74	391.01	590.47
$\Delta\Phi(\vec{p}_T^{miss}), [\vec{p}_T^{miss}]_j) \leq 0.75$	425.12	338.69	511.55
$\Delta\Phi(\vec{p}_T^{j1}, \vec{p}_T^{j2}) \leq 2.5$	369.38	294.26	444.5