Studying jets faking photons

Giulia Maineri, Marcello Fanti, Silvia Resconi, Federica Piazza

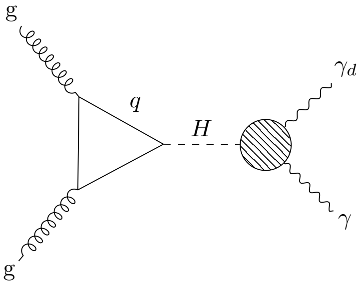
University of Milan

4th April 2024

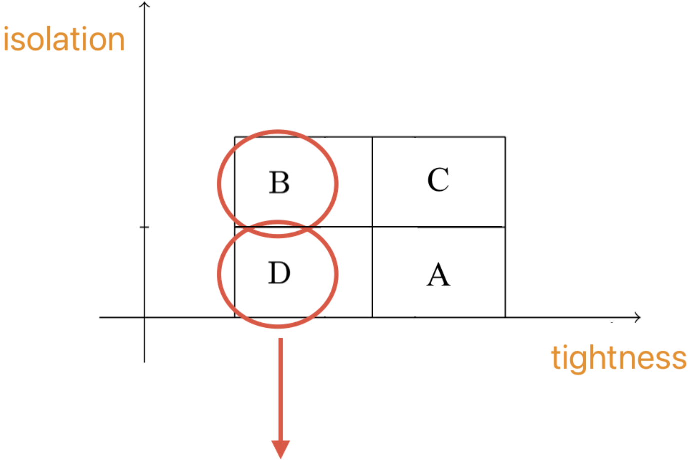
Why don't we use ABCD method?

Topological Trigger

- $N_{\gamma^{tight}} = 1 \implies$ almost no loose photons!
- $|\vec{p}_T^{\gamma}| > 50$ GeV
- $|\vec{p}_T^{miss}| > 70$ GeV
- $m_T > 80$ GeV



Why don't we use ABCD method?



Problem: we don't have loose photons

Relative isolation variables

Problem: We can't use loose photons, because our trigger selects only Tight Photons.

Solution: define Non-Isol Region with cuts on relative isolation variables, in tracker and calorimeter.

$$isol^{rel}_{track} = \frac{p_T^{track20}}{p_T^{\gamma}}$$

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

Non-Isolated Region

We first based this study on **Monte Carlo** simulations, including dijets and γ+jets.

We want to find the best **Non-Isolated Region**, i.e. the region richest in jet faking photons.

Fake factor

This Non-Isol Region will be used to estimate **fake factor** f.

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

where
$N_{j\gamma}^{isol}$ is the number of jet faking photons in the Isolated Region
$N_{j\gamma}^{non-isol}$ is the number of jet faking photons in the Non-Isolated Region

Purity

We also need to take care of true photons in the Non-Isolated Region: this fraction is given by **purity** P.

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

i.e. the fraction of true photons in the Non-Isolated Region.

f and P will be used to estimate the number of jet faking photons in the Signal Region.

In this study the quantity P' was analized instead of P:

$$P' = \frac{N_{\gamma}^{non-isol}}{N_{j\gamma}^{non-isol}}$$

This choice doesn't affect the study, because minimizing P' means minimizing P:

$$\frac{1}{P} = 1 + \frac{1}{P'}$$

Color map - Fake and true photons

Calo relative isolation vs track relative isolation

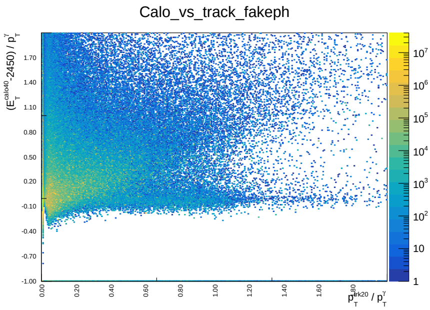


Figure 1: Jet faking photons distribution

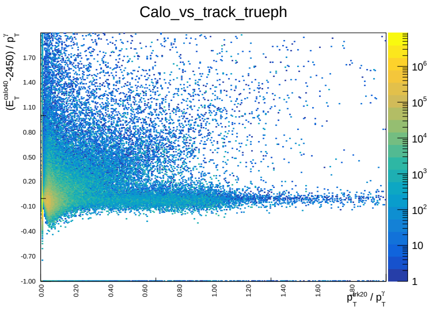


Figure 2: True photons distribution

Fake factor and purity

Calo relative isolation vs track relative isolation

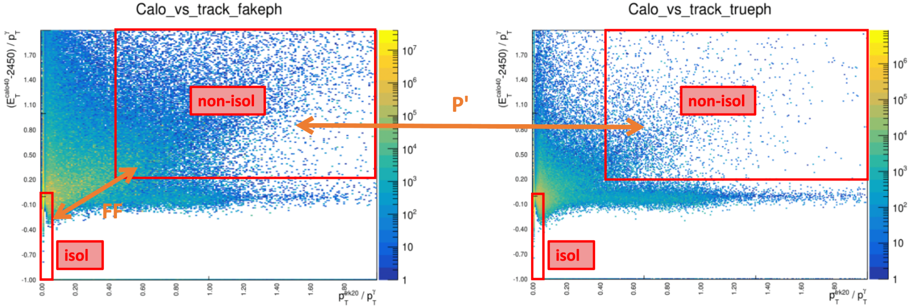


Figure 3: Fake factors method sketch

We would like to have P' **as small as possible**, and at the same time f **reasonably small**: few fake photons in the Isolated Region and few true photons in the Non-Isolated Region.

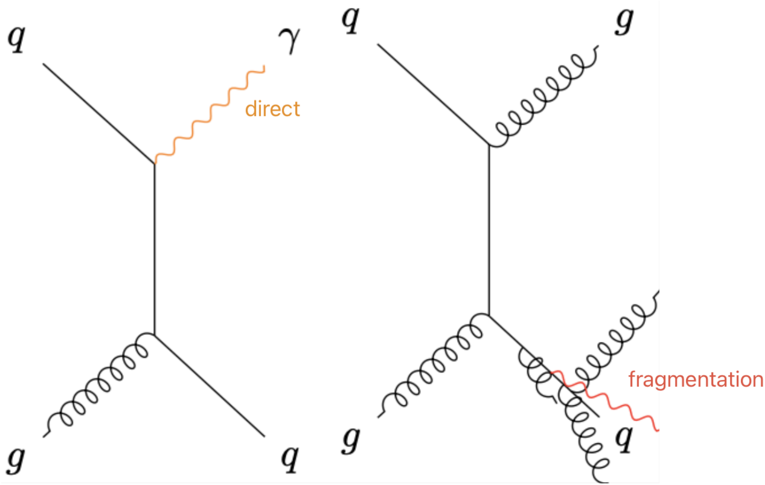
Which photons are "true" and which are "fake"?

True/Fake	Name	MC Truth Classifier Type
True	IsoPhotons	14
Fake	NonIsoPhotons	15
	UnknownPhotons	13
	BkgPhotons	16
	everything else that is not γ neither e	...

Fragmentation photons (MC Truth Classifier Types 15, 13, 16)

Fragmentation photons are photons emerging from parton showers. The photons we are looking for, from $H \to \gamma\gamma_d$, are much more similar to **direct** photons.

Which photons are "true" and which are "fake"?



Fragmentation photons are in any case prompt photons so in inclusive photon analyses they are considered part of the signal"

Figure of merit

The first approach consisted in defining a **figure of merit** M to elect the best Non-isolated Region.

$$M = \left(\frac{\sigma_b}{b}\right)^2 = \frac{f^2}{N^{CR}} + \left(\frac{\sigma_f}{f}\right)^2 + \left(\frac{\sigma_P}{1-P}\right)^2 \approx f^2 + \left(\frac{\sigma_f}{f}\right)^2 + \left(\frac{P}{1-P}\right)^2$$

- f^2: error on the estimation of the number of fake photons in SR
- $\left(\frac{\sigma_f}{f}\right)^2$: error on the fake factor estimation
- $\left(\frac{P}{1-P}\right)^2$: error on the purity estimation

Isolated Region
Fixed Cut Tight

$isol^{rel}_{track} < 0.05$

$isol^{rel}_{calo} < 0.022$

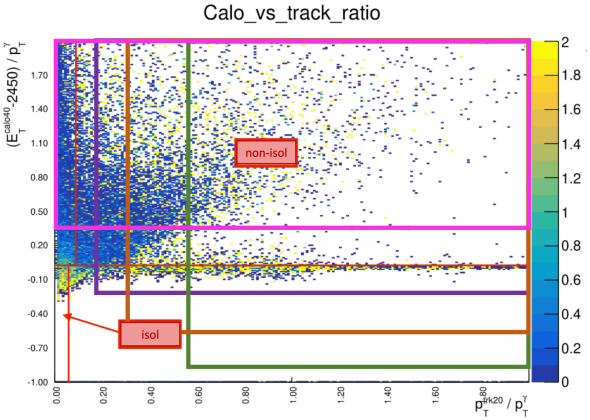


Figure 4: Search for a Non-Isolated Region ortogonal to the Isolated Region

Search for a Non-Isolated Region

The best Non-Isolated Region is defined by:

$$isol^{rel}_{calo} = \frac{E_T^{calo40} - 2450}{p_T^{\gamma}} > 0.022$$

$$\forall \quad isol^{rel}_{track} = \frac{p_T^{track20}}{p_T^{\gamma}}$$

Fake factor and purity in this Non-Isolated Region are:

$$f = 0.83 \pm 0.10$$

$$P = 0.0864761$$

Elected Non-Isolated Region

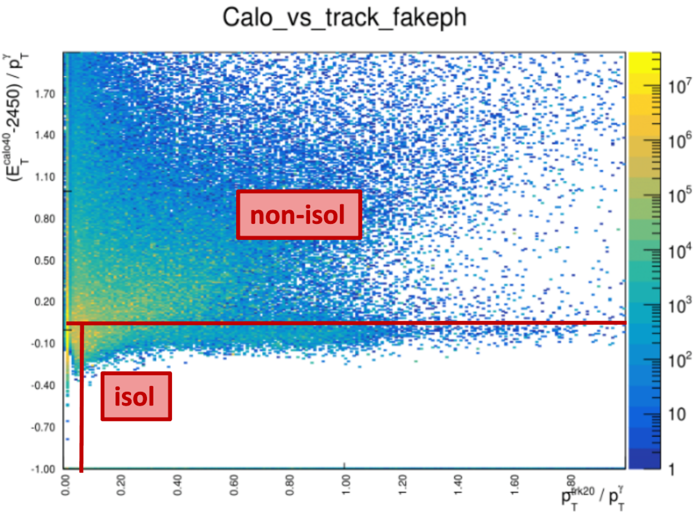
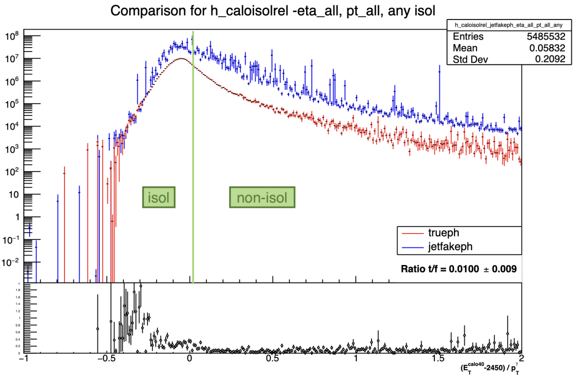


Figure 5: Pseudo-purity P' in fine bins of relative isolation in tracker and calo

Elected Non-Isolated Region

We then abandoned relative isolation in the tracker: it is not recommended as working point.
This doesn't affect our choice for the Non-Isolated Region.



Fixed Cut Tight Calo Only

Figure 6: Relative isolation in calo for tight photons MC.

Should we trust f and P from Monte Carlo?

Isolation is typically **not** well modelled by MC for jet faking photons.

We need to check our results on **data**.

Loose photons

How to get jet faking photons in data?

We have different possibile **ID selection**:

- **loose**, if they satisfy loose criteria but not the tight ones;
- **loose5**, if they satisfy the loose criteria and pass tight cuts on all egamma shower shapes of HCAL and ECAL Middle layer;
- **loose4**, if they are loose5 and pass tight cuts on W_{stot} shower shape of the ECAL Strips.

Loose photons

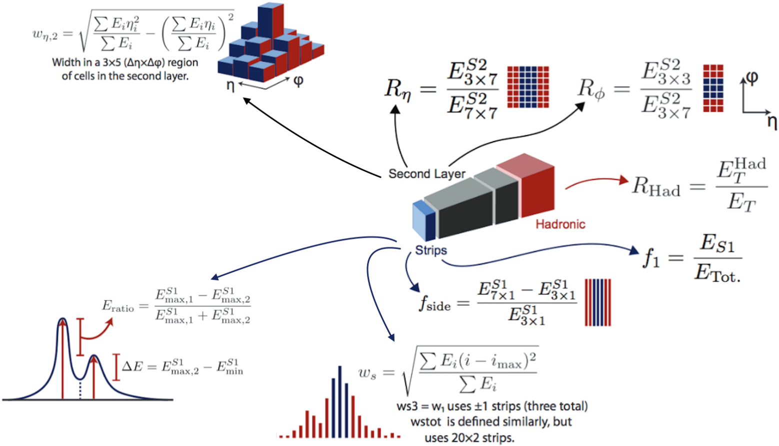


Figure 7: Variables describing shower shapes, energy ratios and width of the energy deposit

Loose photons

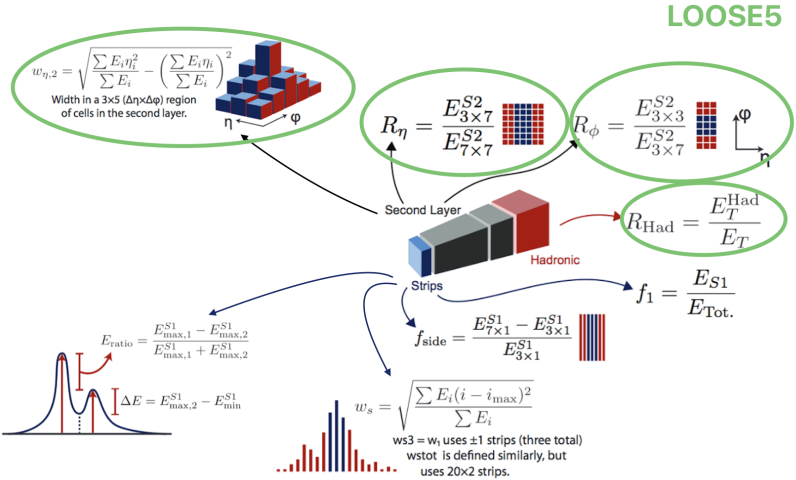


Figure 8: Variables describing shower shapes, energy ratios and width of the energy deposit (loose5)

Loose photons

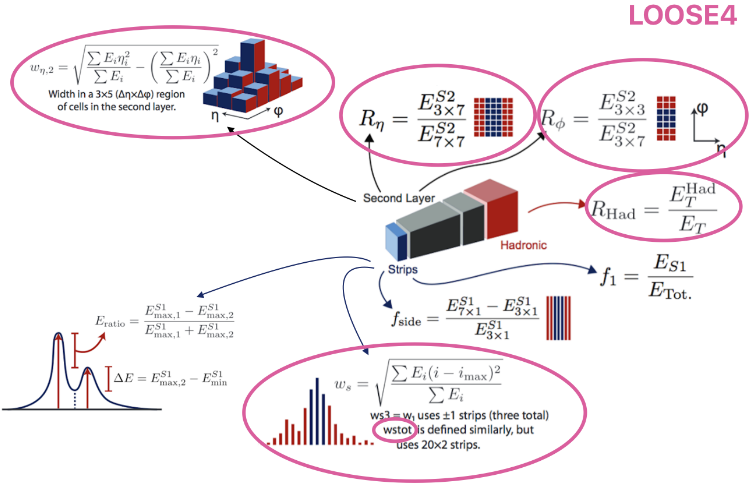
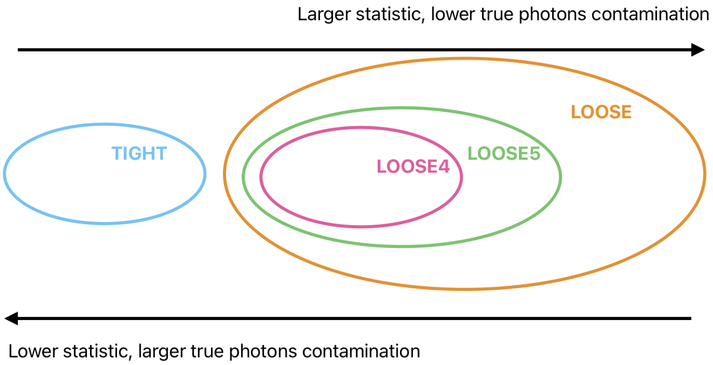


Figure 9: Variables describing shower shapes, energy ratios and width of the energy deposit (loose4)

Loose photons



Isolation comparison in MC: tight, loose, loose4, loose5

Isolation for tight, loose, loose4 and loose5 fake photons in MC is different

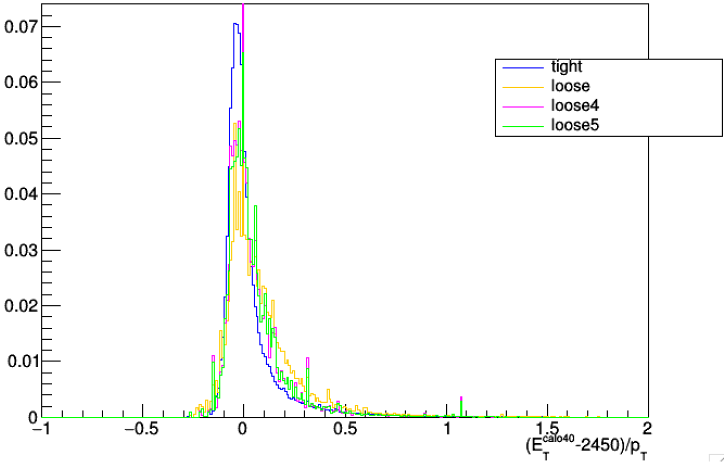


Figure 10: MC isolation for fake photons with different ID

Isolation comparison (fake)γ in DATA: loose, loose4, loose5

Isolation for loose, loose4 and loose5 photons in DATA is different

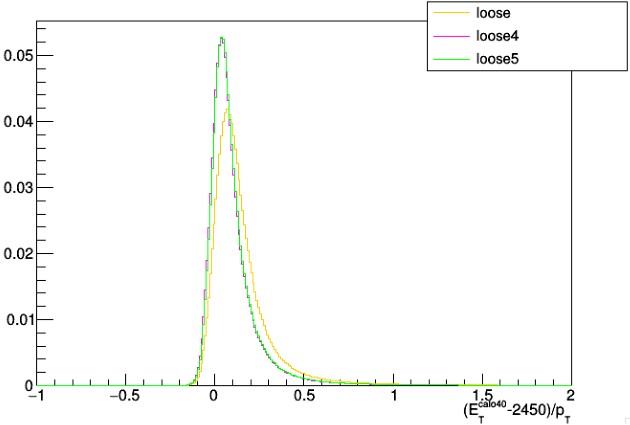


Figure 11: DATA isolation for (fake) photons with different ID

Isolation comparison for loose fake γ: MC and data

MC and DATA isolations for a same ID are different

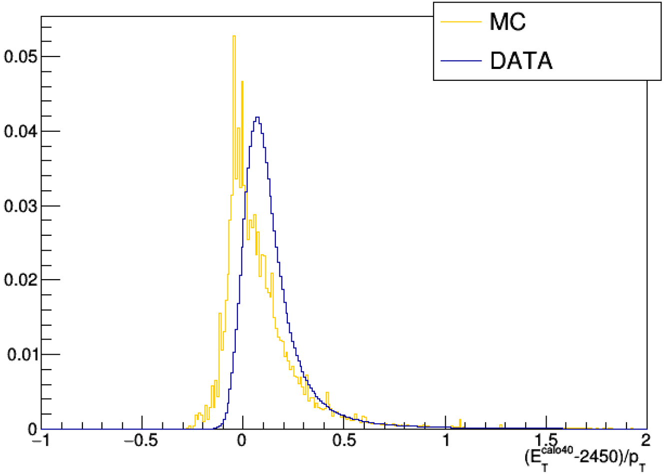


Figure 12: Calo isolation in MC and DATA for **loose** photons

Isolation comparison for loose5 fake γ: MC and data

MC and DATA isolations for a same ID are different

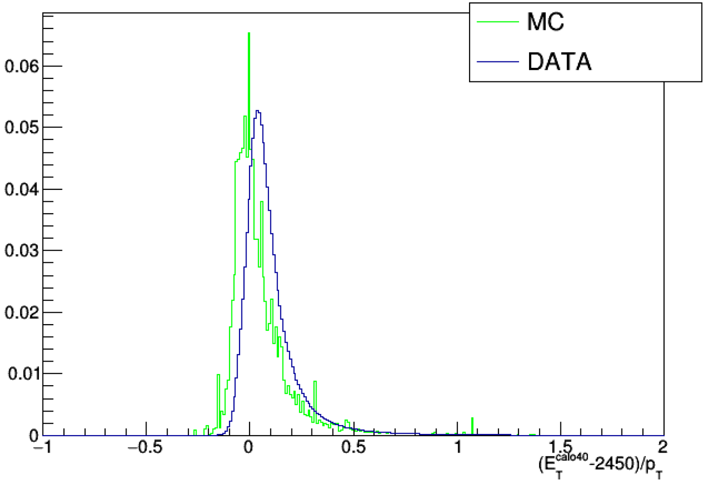


Figure 13: Calo isolation in MC and DATA for **loose5** photons

Isolation comparison for loose4 fake γ: MC and data

MC and DATA isolations for a same ID are different

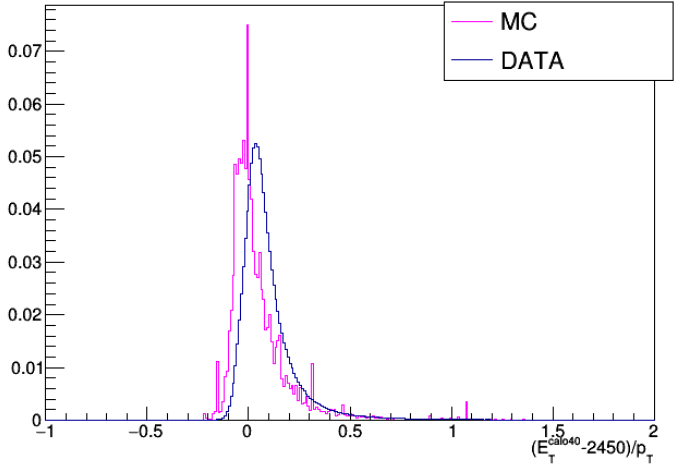


Figure 14: Calo isolation in MC and DATA for **loose4** photons

Extrapolation L->T

We would like to calculate fake factors using tight fake photons in data, to check our results obtained with MC.

Problem: In data, we cannot know which tight photons are true and which fake.

Solution: We can try to extrapolate the tight fake photons distribution in data from the loose photons distribution in data, assuming that:

- loose photons in data are mostly fake;
- the transformation that links tight and loose distributions in MC is "somehow" related to the one that links tight and loose in data.

Extrapolation L->T

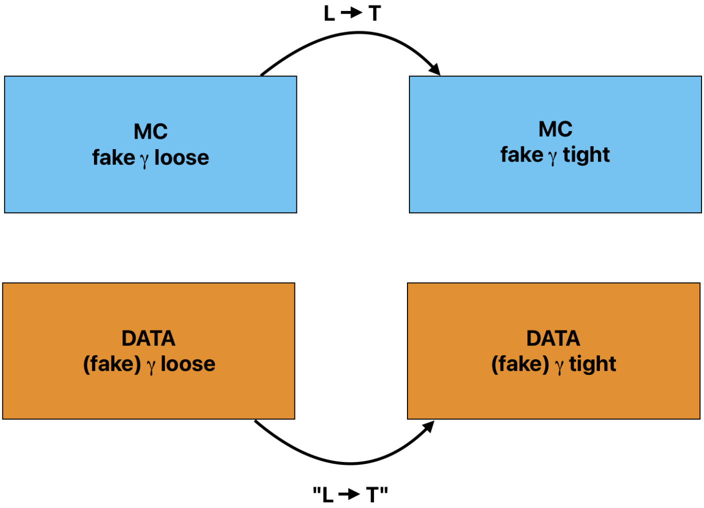


Figure 15: L $\rightarrow$T transformation

Step 1: get L->T transformation from MC

Let's assume that tight x_T^{MC} and loose x_L^{MC} distributions in MC are linked by an affine transformation, the easiest transformation to reproduce mean and standard deviation of the start distribution:

$$x_T^{MC} = a + b x_L^{MC}$$

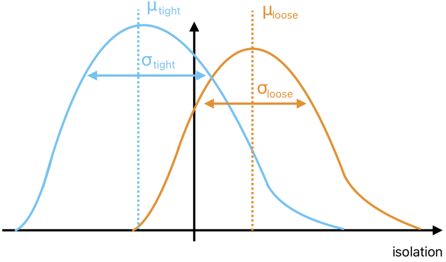
We want to find a, b such that:

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

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:

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

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

b

known

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

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

Step 2: apply L->T transformation to DATA

It's all set, we can now apply the transformation to DATA:

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

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

Step 3: obtained distributions

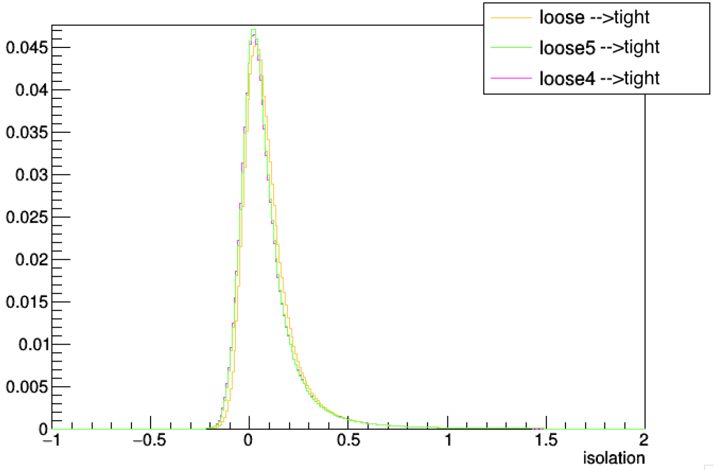


Figure 16: Distributions of tight fake photons (DATA) isolation, extrapolated from loose (DATA)

Data and Monte Carlo comparison

Problem: we need to introduce a trigger to be able to compare data and Monte Carlo.
Solution: Ok, but which one?

Analysis Trigger

- $N_{\gamma^{tight}} = 1$
- $|\vec{p}_T^{\gamma}| > 50$ GeV
- $|\vec{p}_T^{miss}| > 70$ GeV
- $m_T > 80$ GeV

Leptonic Trigger

- $N_{el} = 1$ or $N_{el} = 2$;
- $N_{\mu} = 1$ or $N_{\mu} = 2$;

MET Trigger

- $|\vec{p}_T^{miss}| > 90$ GeV

Validation on loose5

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

Let's validate our hypothesis on loose5 photons, which we can compare in data and MC assuming they are mostly fake photons.
We introduce a factor R; let's see when R is close to 1.

$$\frac{\mu_{L5}^{data} - \mu_L^{data}}{\sigma_L^{data}} = R \, \frac{\mu_{L5}^{MC} - \mu_L^{MC}}{\sigma_L^{MC}}$$

Binning in η

Pseudorapidity binning is chosen considering the detector geometry:

- etabin00 represents the inclusive region;
- etabin01: $[0; 0.6]$, the upper limit $\eta = 0.6$ is the point after which the material in front of ECAL increases a lot;
- etabin02: $[0.6; 1.37]$, the upper limit is defined by the beginning of the crack region;
- etabin03: $[1.37; 1.52]$, corresponds to the crack region;
- etabin04: $[1.52; 1.81]$, the upper limit is the point where the presampler ends;
- etabin05: $[1.81; 2.37]$.

Binning in p_T

Transverse momentum binning is chosen as follows:

- ptbin00 represents the inclusive region;
- ptbin01: $[25, 35]$GeV;
- ptbin02: $[35, 45]$GeV;
- ptbin03: $[45, 55]$GeV;
- ptbin04: $[55, 65]$GeV;
- ptbin05: $[65, 75]$GeV;
- ptbin06: $[75, 100]$GeV;
- ptbin07: $[100, 150]$GeV;
- ptbin08: $[150, 250]$GeV;

Analysis trigger

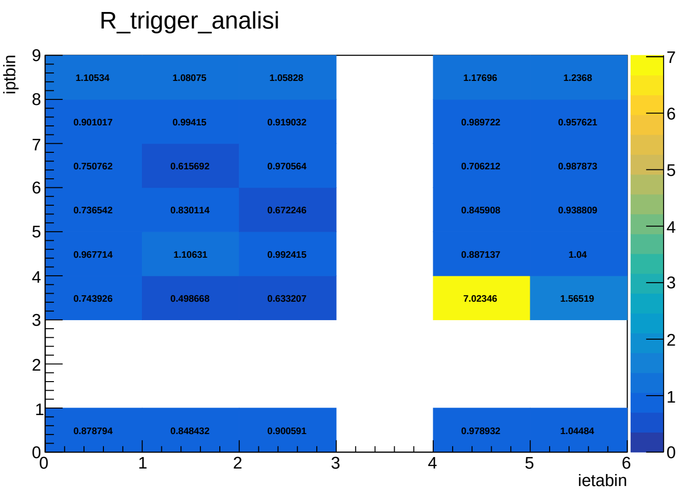


Figure 17: R, analysis trigger

Leptonic trigger

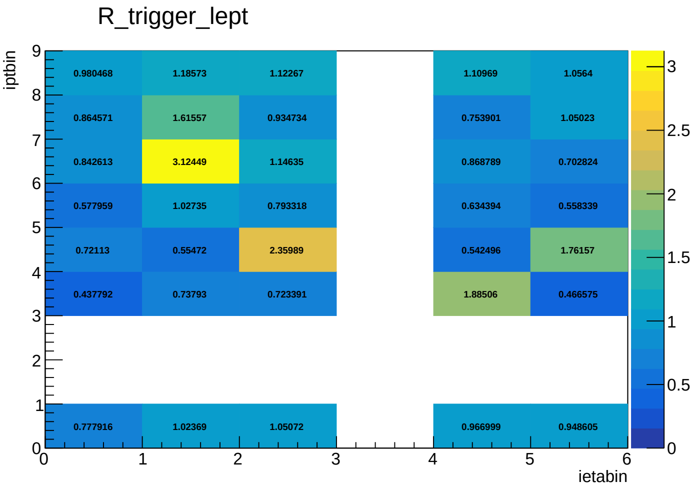


Figure 18: R, leptonic trigger

MET trigger

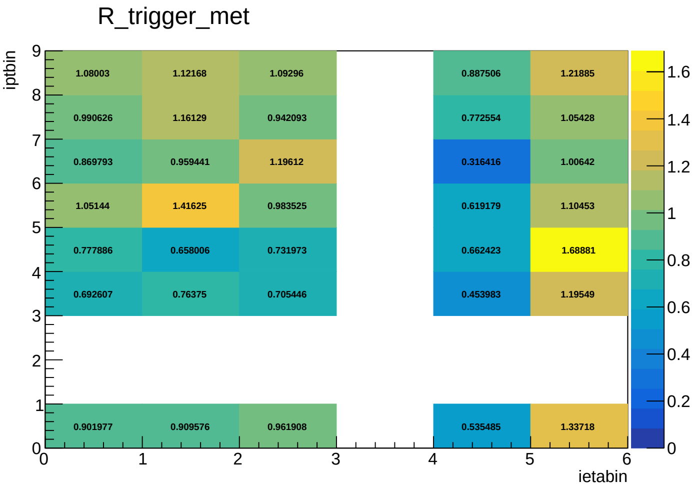


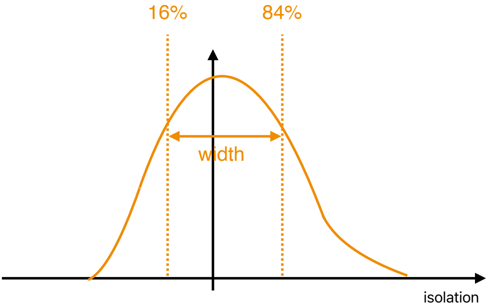
Figure 19: R, MET trigger

Validation on loose5

Let's do the same validation using median and width, two indicators less sensitive to outliers.

$$\frac{med_{L5}^{data} - med_{L}^{data}}{w_L^{data}} = R\frac{med_{L5}^{MC} - med_{L}^{MC}}{w_L^{MC}}$$

where *med* is the median and *w* is calculated as:



Analysis trigger

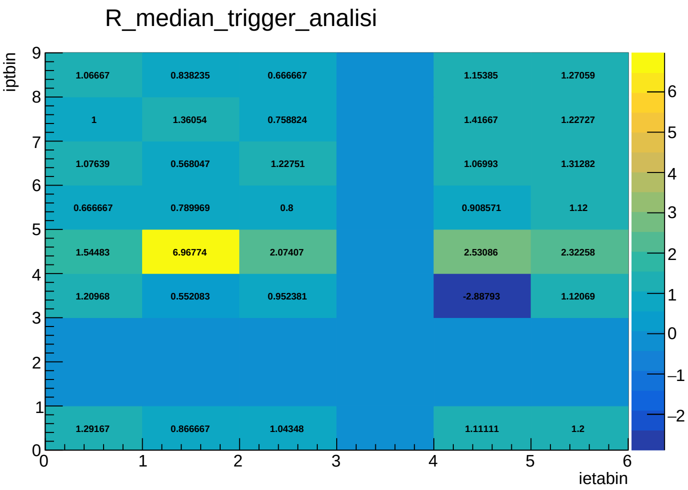


Figure 20: R, analysis trigger

Leptonic trigger

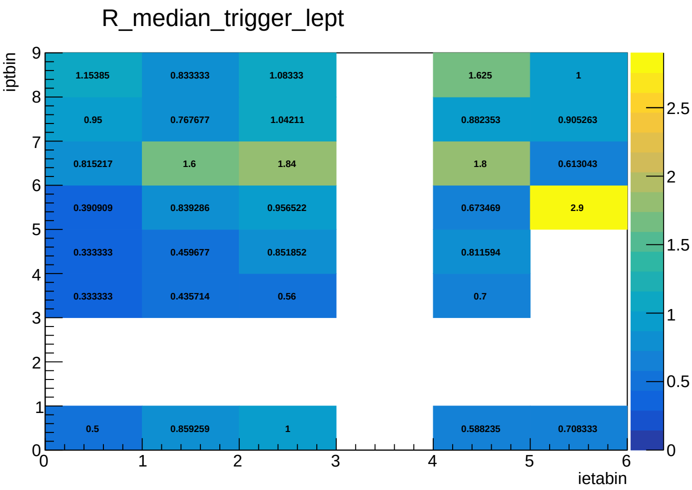


Figure 21: R, leptonic trigger

MET trigger

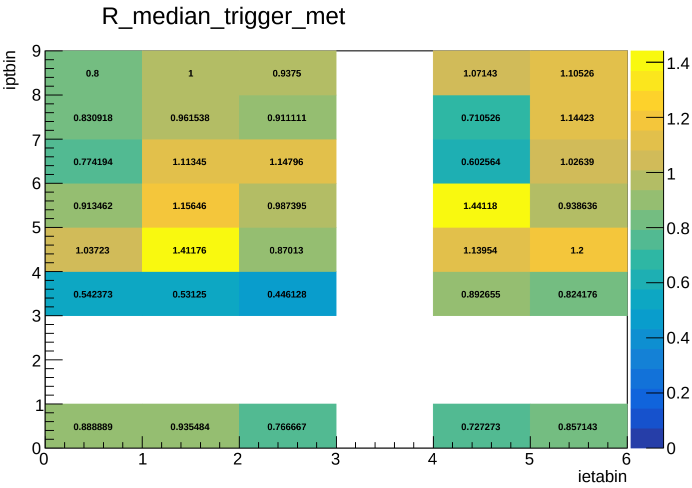


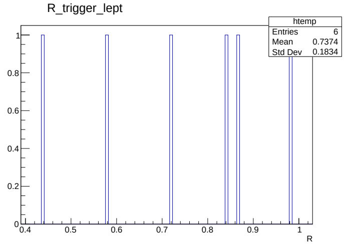
Figure 22: R, MET trigger

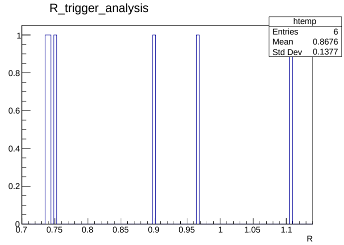
Validation on loose5

Problem: R is very unstable. It is not possible to perform the extrapolation in an exclusive regions in p_T, η.

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

R comparison for different trigger, inclusive in η

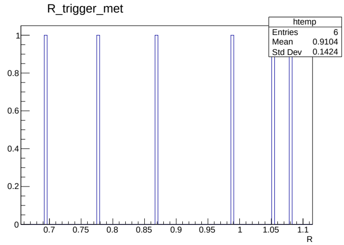




Leptonic Trigger

Mean: 0.74

Spread: 0.18



MET Trigger

Mean: 0.94

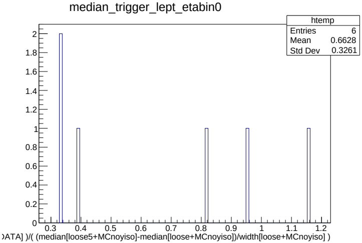
Spread: 0.14

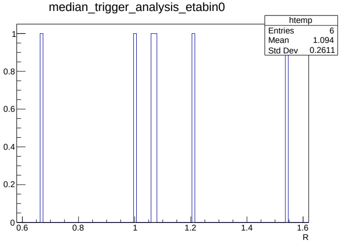
Analysis Trigger

Mean: 0.87

Spread: 0.14

R comparison for different trigger, median and width, inclusive in η

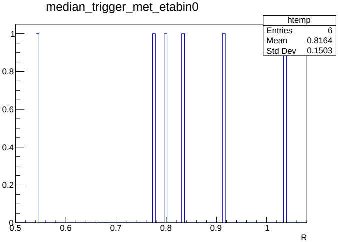




Leptonic Trigger

Median: 0.66

Width: 0.32



MET Trigger

Median: 0.81

Width: 0.15

Analysis Trigger

Median: 1.1

Width: 0.26

η distribution, MET trigger

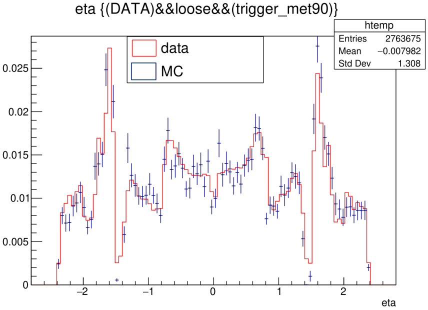


Figure 23: Loose photons

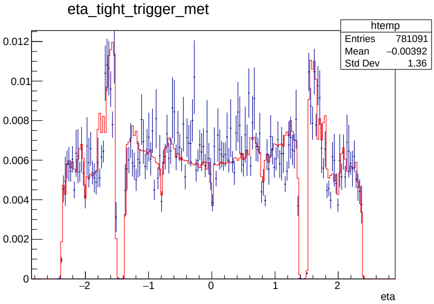


Figure 24: Tight photons

$\Longrightarrow$ good agreement ✔

η distribution, analysis trigger

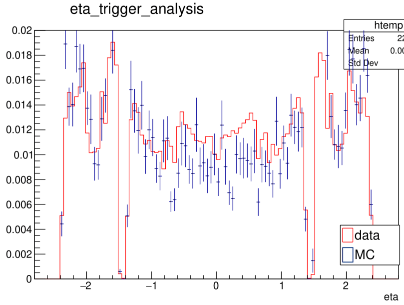


Figure 25: Loose photons

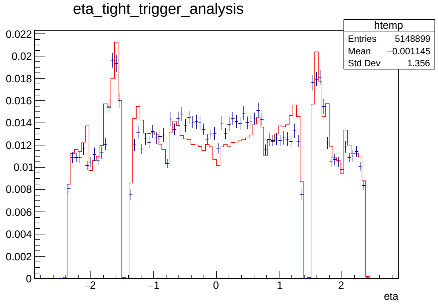
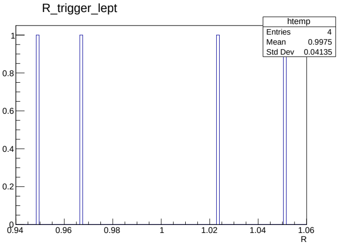
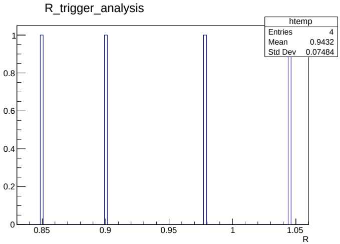


Figure 26: Tight photons

$\Longrightarrow$ good agreement ✔

R comparison for different trigger, inclusive in p_T

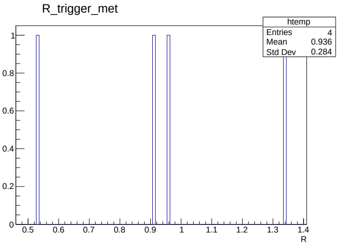




Leptonic Trigger

Mean: 0.99

Spread: 0.04



MET Trigger

Mean: 0.94

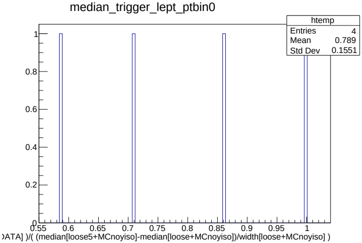
Spread: 0.26

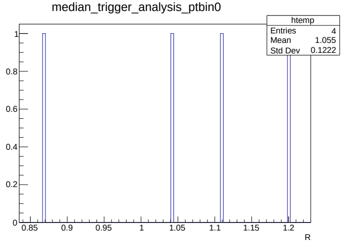
Analysis Trigger

Mean: 0.94

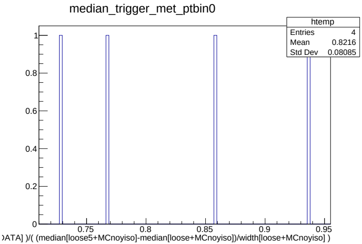
Spread: 0.07

R comparison for different trigger, median and width, inclusive in p_T





Leptonic Trigger
Median: 0.79
Width: 0.15



MET Trigger
Median: 1.0
Width: 0.12

Analysis Trigger
Median: 0.82
Width: 0.08

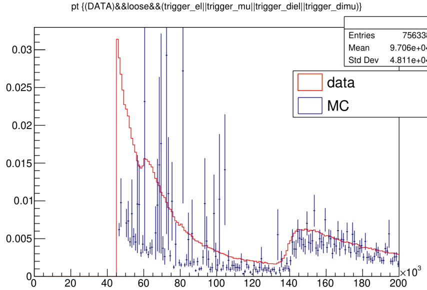


Figure 27: Loose photons

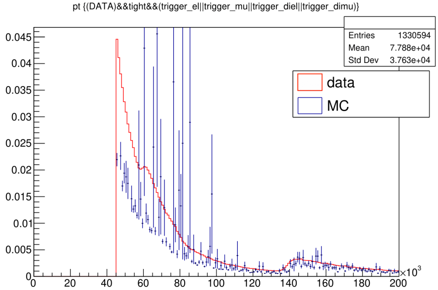


Figure 28: Tight photons

$\Longrightarrow$ bad agreement ✘

p_T distribution, analysis trigger

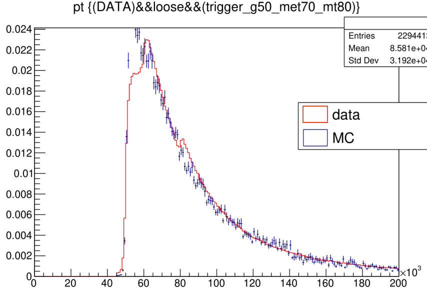


Figure 29: Loose photons

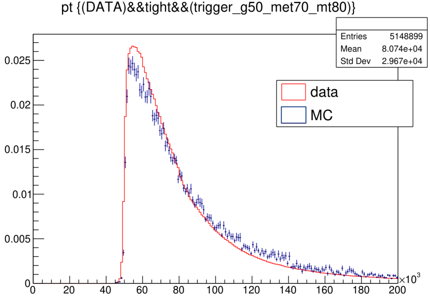


Figure 30: Tight photons

$\Longrightarrow$ good agreement ✔

p_T distribution, MET trigger

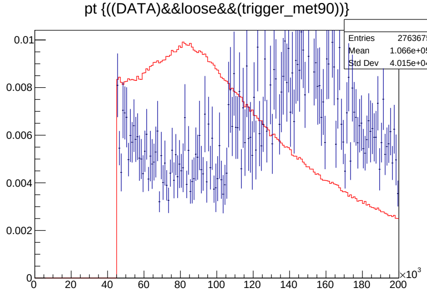


Figure 31: Loose photons

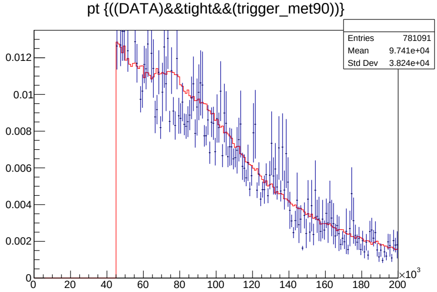


Figure 32: Tight photons

$\Longrightarrow$ bad agreement ✘

Elected trigger

We are left with these possibilities:

MET Trigger	**Analysis Trigger**	**MET Trigger**	**Analysis Trigger**
Mean: 0.94	Mean: 0.87	Median: 0.81	Median: 1.1
Spread: 0.14	Spread: 0.14	Width: 0.15	Width: 0.26
inclusive in η	inclusive in η	inclusive in η	inclusive in η
mean, sigma	mean, sigma	median, width	median, width

Analysis Trigger	**Analysis Trigger**
Mean: 0.94	Median: 0.82
Spread: 0.07	Width: 0.08
inclusive in p_T	inclusive in p_T
mean, sigma	median, width

Elected trigger

We are left with these possibilities:

MET Trigger	Analysis Trigger	MET Trigger	Analysis Trigger
Mean: 0.94	Mean: 0.87	Median: 0.81	Median: 1.1
Spread: 0.14	Spread: 0.14	Width: 0.15	Width: 0.26
inclusive in η	inclusive in η	inclusive in η	inclusive in η
mean, sigma	mean, sigma	median, width	median, width

Analysis Trigger	Analysis Trigger
Mean: 0.94	Median: 0.82
Spread: 0.07	Width: 0.08
inclusive in p_T	inclusive in p_T
mean, sigma	median, width

Conclusions and next steps

Summary

- We presented this **new method** to estimate the jet faking photons background based on **extrapolation** of fake tight photons distributions in data from loose photons distributions.
- We tried to **validate** the method comparing loose5 distribution with the extrapolated one.
- Ratio R in exclusive regions in η, p_T was found to suffer from **fluctuations**, hence the extrapolation should be done in a region inclusive in η or p_T only;
- Different **triggers** were explored in order to have $R \sim 1$ and a small spread in either η or p_T: the two best options is Analysis Trigger inclusive in p_T, using as indicators median and width.

Conclusions and next steps

Next steps

- **Extrapolate** fake tight photons isolation distribution;
- Calculate **fake factors**;
- How to treat fake factors **uncertainties**: need to propagate mean and rms errors, envelope method?
- Calculate **purity** from the extrapolated distribution of tight fake photons in DATA: normalizing the tail and subtracting;