UNIVERSITÀ
DEGLI STUDI
DI MILANO

Measurement of the jets faking photons background for the Dark Matter search in the Mono-Photon channel with the ATLAS detector

Relatore interno:
Prof. Leonardo Carminati

Correlatori:
Dott.sa Silvia Resconi
Dott. Davide Pietro Mungo

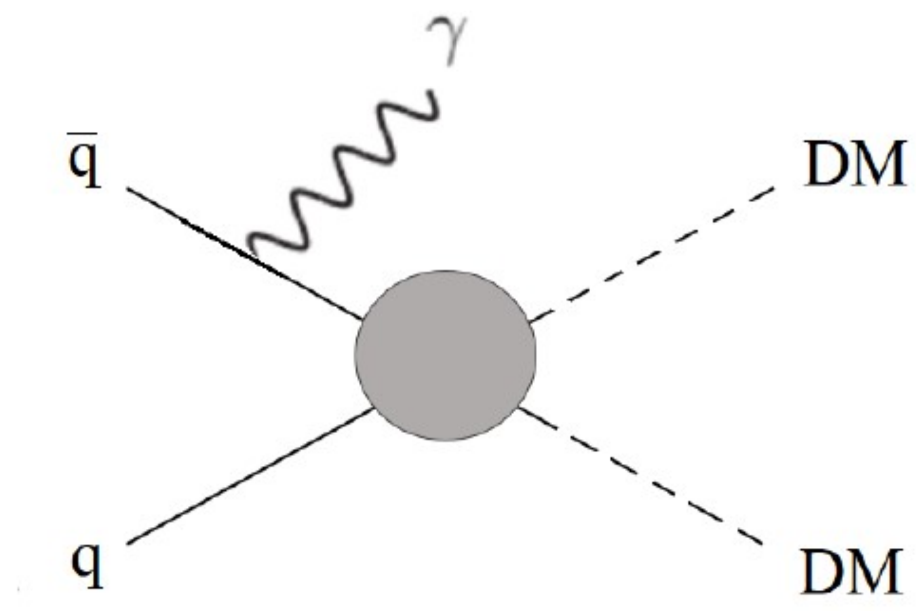
Candidato:
Gianluca Fugante

Anno Accademico 2018/2019

Dark Matter at LHC

- Detection at colliders
  - The WIMPs (Weakly Interacting Massive Particles) hypothesis for Dark Matter is tested at LHC
  - Dark Matter particles production revealed as large missing momentum in the transverse plane (E_T^{miss})
  - Need a SM object associated to the event (γ, e^-, jet, Z or W Boson, Higgs boson)
    - Mono-X searches at LHC $\longrightarrow$ SM object + large E_T^{miss} events

- Mono-Photon analysis
  - Signal Region (SR): Photon + large E_T^{miss}
    - E_T^{miss} > 150 GeV
    - Photon with p_T > 150 GeV



Mono-Photon analysis

- Standard Model (SM) backgrounds:
  - Z($\nu\nu$) + γ, irreducible background (~60%)
  - Z(ll) + γ, where both leptons are not reconstructed
  - W($l\nu$) + γ, where the lepton is not reconstructed
  - Z($\nu\nu$) + jet, where the jet fakes a photon
  - W($l\nu$) + jet, where the jet or the lepton fakes a photon

Jets faking photons:

expected ~8% of the total background

Mono-Photon analysis

The Signal Region (SR) is divided in 5 regions with different $E_T{}^{miss}$:

	ISR1	ISR2	ISR3	ESR1	ESR2
$E_T{}^{miss}$ [GeV]	> 150	> 225	> 300	150 - 225	225 - 300

For each of this regions 3 control regions are defined, each enriched with different processes:

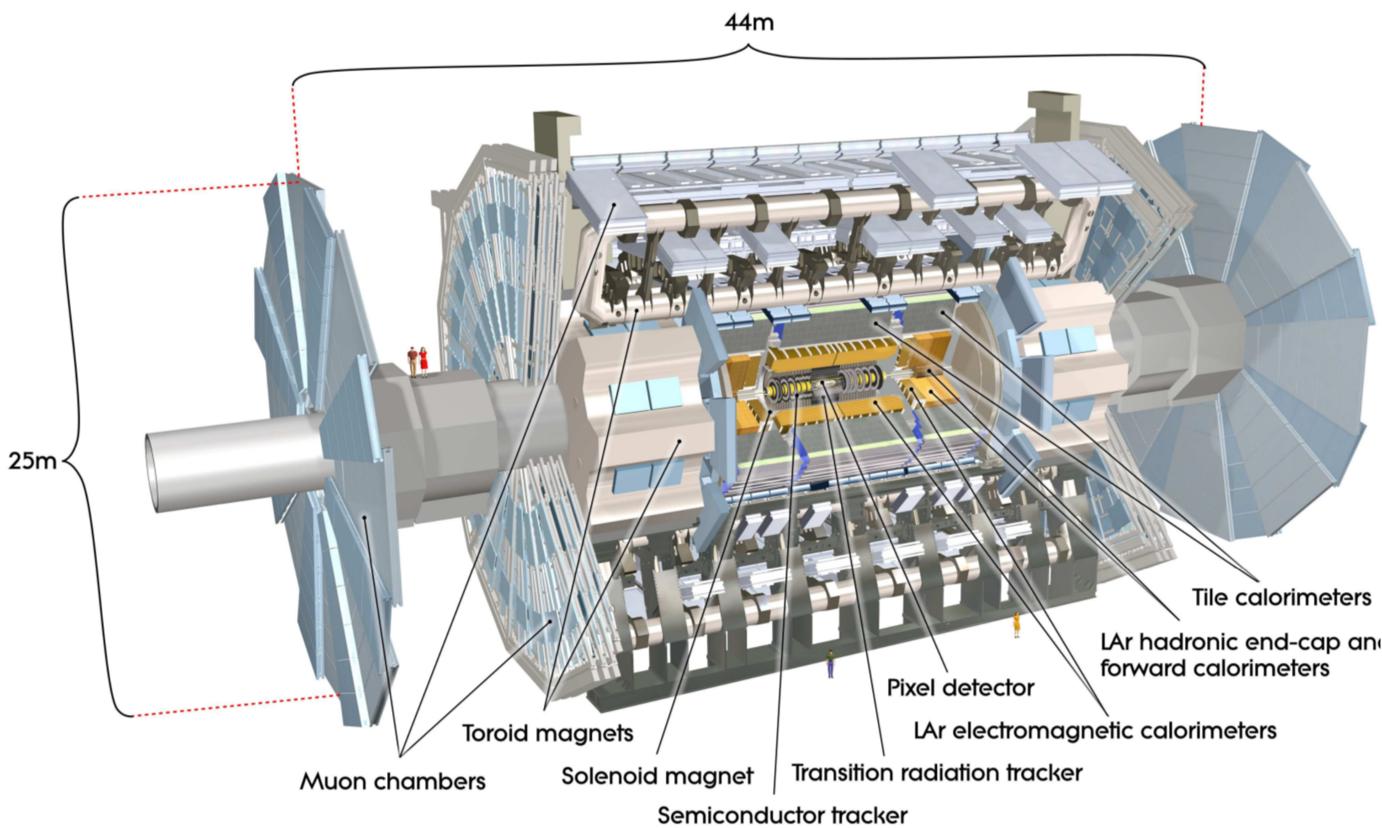
- one muon CR (1muCR) $\longrightarrow$ W $(\mu\nu)$ + γ
- two muon CR (2muCR) $\longrightarrow$ Z $(\mu\mu)$ + γ
- two electron CR (2eCR) $\longrightarrow$ Z (ee) + γ

One more control region, the gammajetCR, dominated by jet + γ process (85 GeV < $E_T{}^{miss}$ < 110 GeV).

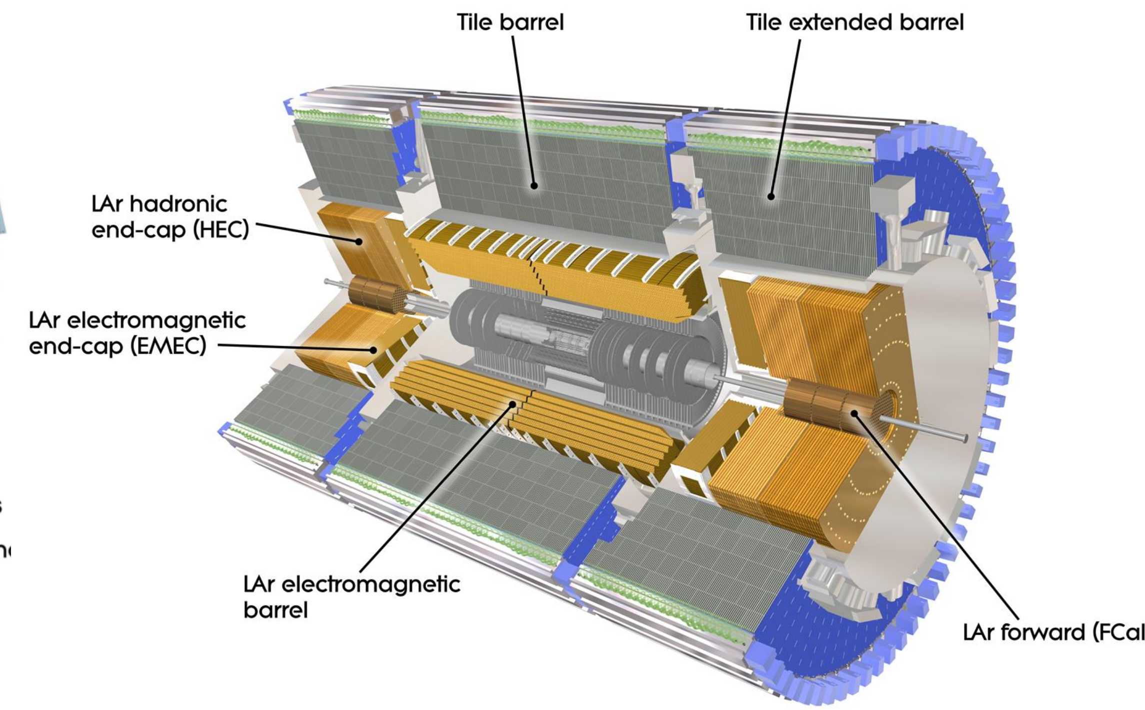
Normalization factors are extracted comparing data to MCs estimates in the CRs.

ATLAS

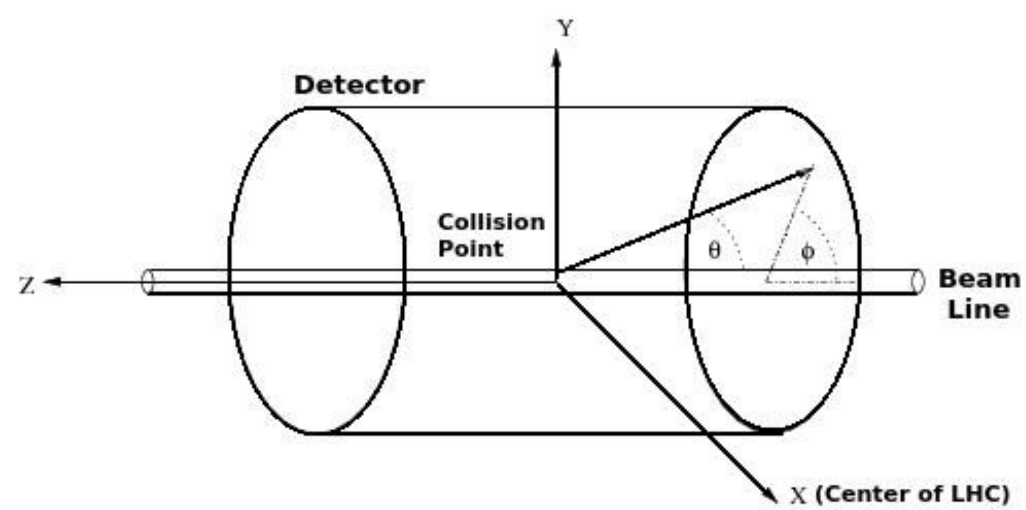
Detector



Calorimeters



Coordinate system



$$\eta = -\ln \tan \theta/2$$

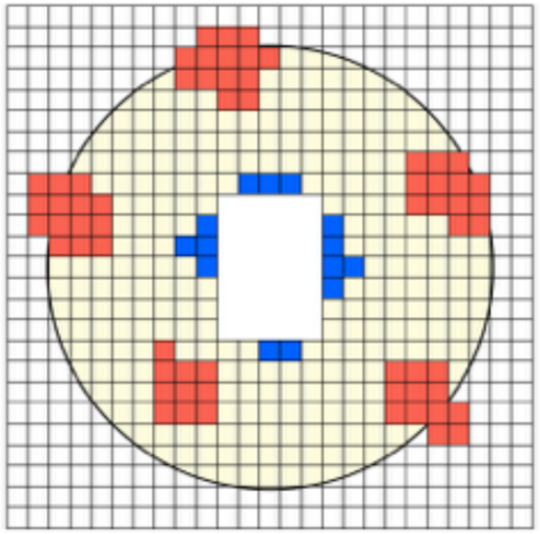
$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

Photons Reconstruction

Photons are reconstructed from energy clusters in the EM calorimeter and tracks in the inner detector:

- Photon Isolation
  - Inside the EM calorimeter
  - Energy deposits summed up in a cone of radius $\Delta R = 0.4$
  - The contribution of the photon is subtracted
  - The result is saved in the *TopoEtCone40* variable
  - Isolation energy:

$$E_{iso} = TopoEtCone40 - 0.022 p_T - 2.45 \text{ GeV}$$



- Photon Identification
  - Selection based on shower shapes variables

	Loose	Tight
R_{had1}	✓	✓
R_{had}	✓	✓
R_η	✓	✓
$w_{\eta,2}$	✓	✓
R_ϕ		✓
$w_{s,3}$		✓
w_{stot}		✓
f_{side}		✓
ΔE_s		✓
E_s		✓
f_1		✓

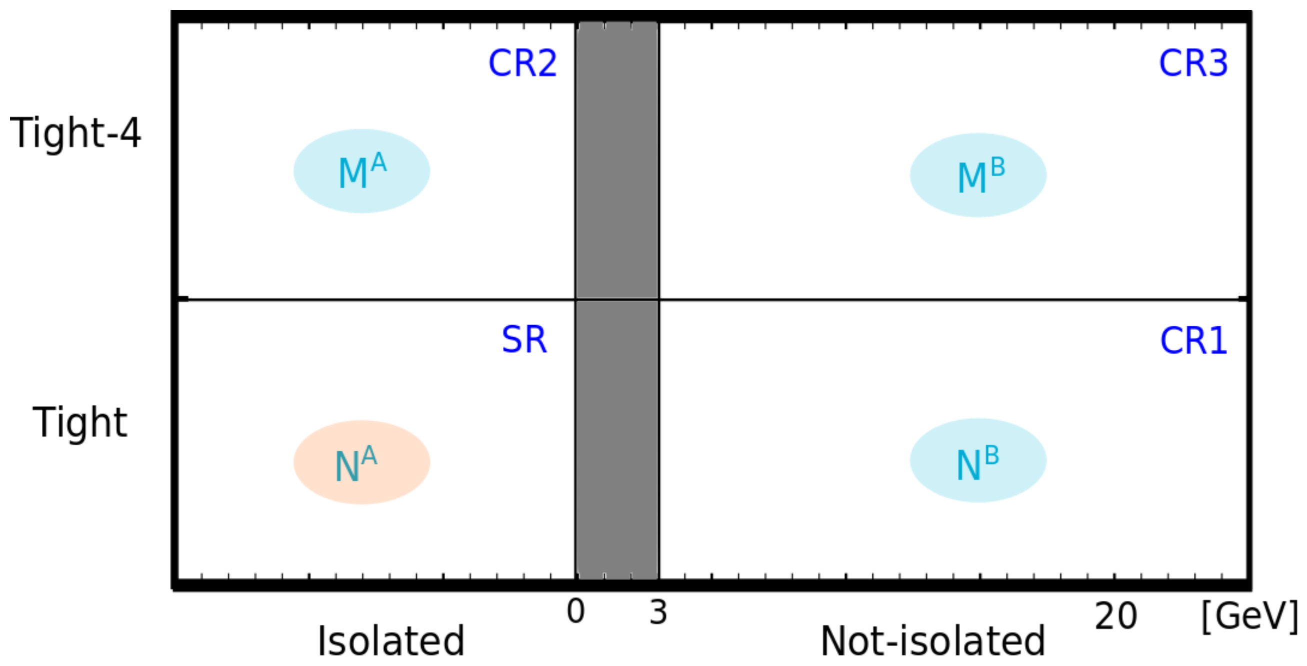
  - Fail one or more selection:
    - $f_{side}, \Delta E_s, w_{s,3}$ ⟶ Tight-3
    - $f_{side}, \Delta E_s, w_{s,3}, E_{ratio}$ ⟶ Tight-4
    - $f_{side}, \Delta E_s, w_{s,3}, E_{ratio}, w_{stot}$ ⟶ Tight-5

The 2D sideband method

MC samples cannot describe accurately the jets faking photons background. A data-driven technique, known as the 2D sideband method, is needed.

The method defines a 2D plane (x,y) formed by a photon isolation variable x and an identification variable y.

- Identification:
  - Photons that pass the Tight selection are considered as identified
  - Photons that pass the Tight-4 selection but not the Tight one are considered as not-identified
- Isolation:
  - $E_{iso} < 0$ GeV : isolated photon
  - $E_{iso} > 3$ GeV : not-isolated photon
  - A gap of 3 GeV is excluded to prevent signal leakage



The 2D sideband method

Two simplifying hypotheses:

1. The correlation between tightness and isolation is negligible for the background
2. The number of signal photons in the control regions is negligible compared to the number of fake photons

The method provides a fully <u>data-driven</u> formula to estimate the number of signal photons in the SR:

$$N^A_{\text{sig}} = N^A - N^A_{\text{bkg}} = N^A - N^B \frac{M^A}{M^B}$$

the purity is then:

$$P = N^A_{\text{sig}}/N^A = 1 - \frac{N^B}{N^A}\frac{M^A}{M^B}$$

The 2D sideband method

The method can be corrected by means of Monte Carlo (MC) simulations:

1. Non negligible correlation in the background between tightness and isolation

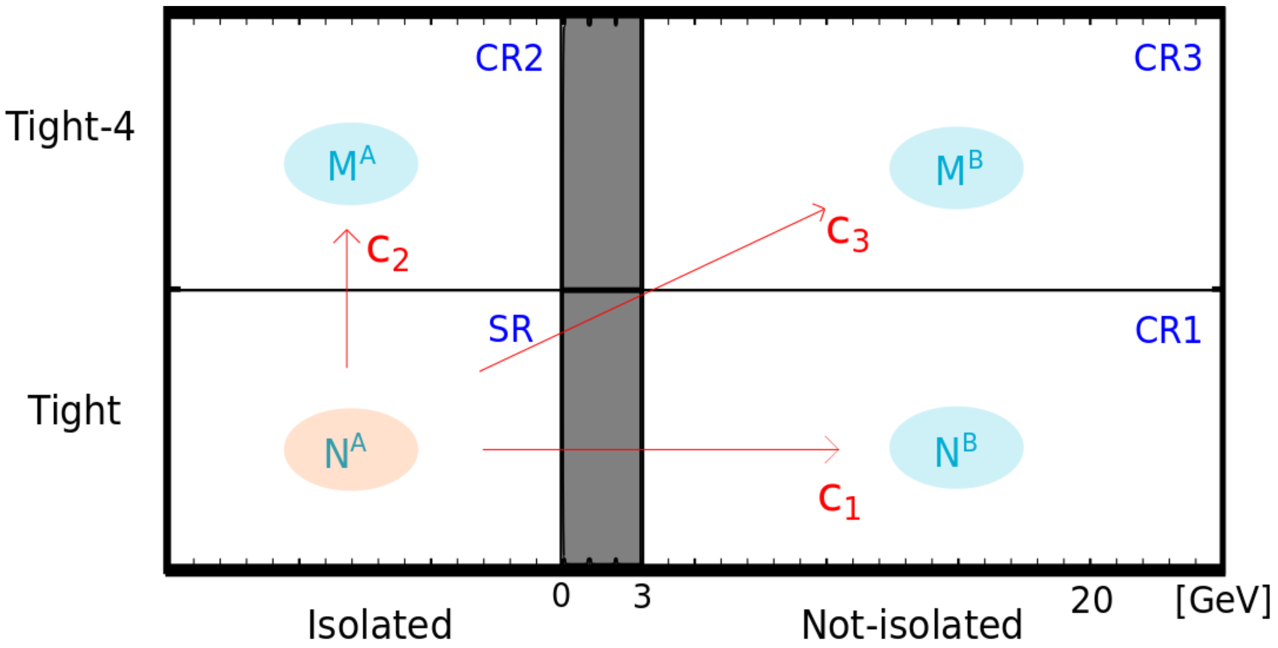
$$R_{MC} = \frac{N^A_{bkgMC}}{N^B_{bkgMC}} \frac{M^B_{bkgMC}}{M^A_{bkgMC}}$$

2. Signal leakage in background control regions

$$c_1 = \frac{N^B_{sigMC}}{N^A_{sigMC}}$$

$$c_2 = \frac{M^A_{sigMC}}{N^A_{sigMC}}$$

$$c_3 = \frac{M^B_{sigMC}}{N^A_{sigMC}}$$



The 2D sideband method

The number of signal photons in the SR is then:

$$N^A_{sig} = \frac{(M^B + N^A c_3 - N^B c_2 R_{MC} - M^A c_1 R_{MC})(-1 + \sqrt{1 + \frac{4(c_1 c_2 R_{MC} - c_3)(N^A M^B - N^B M^A R_{MC})}{(M^B + N^A c_3 - N^B c_2 R_{MC} - M^A c_1 R_{MC})^2}})}{2(c_1 c_2 R_{MC} - c_3)}$$

and the purity:

$$P = \frac{(M^B + N^A c_3 - N^B c_2 R_{MC} - M^A c_1 R_{MC})(-1 + \sqrt{1 + \frac{4(c_1 c_2 R_{MC} - c_3)(N^A M^B - N^B M^A R_{MC})}{(M^B + N^A c_3 - N^B c_2 R_{MC} - M^A c_1 R_{MC})^2}})}{2N^A(c_1 c_2 R_{MC} - c_3)}$$

Validation

To validate the method we used a mixed Monte Carlo sample of W($\mu\nu$) + γ and W + jets, with known purity:

- The correlation factor and the signal leakage coefficients are calculated on the same sample
- Focus on the SR - ISR1

In this region the coefficients are:

R_{MC}	c_1 [%]	c_2 [%]	c_3 [%]
2,76 ± 0,49	7,08 ± 0,80	4,50 ± 0,63	0,49 ± 0,21

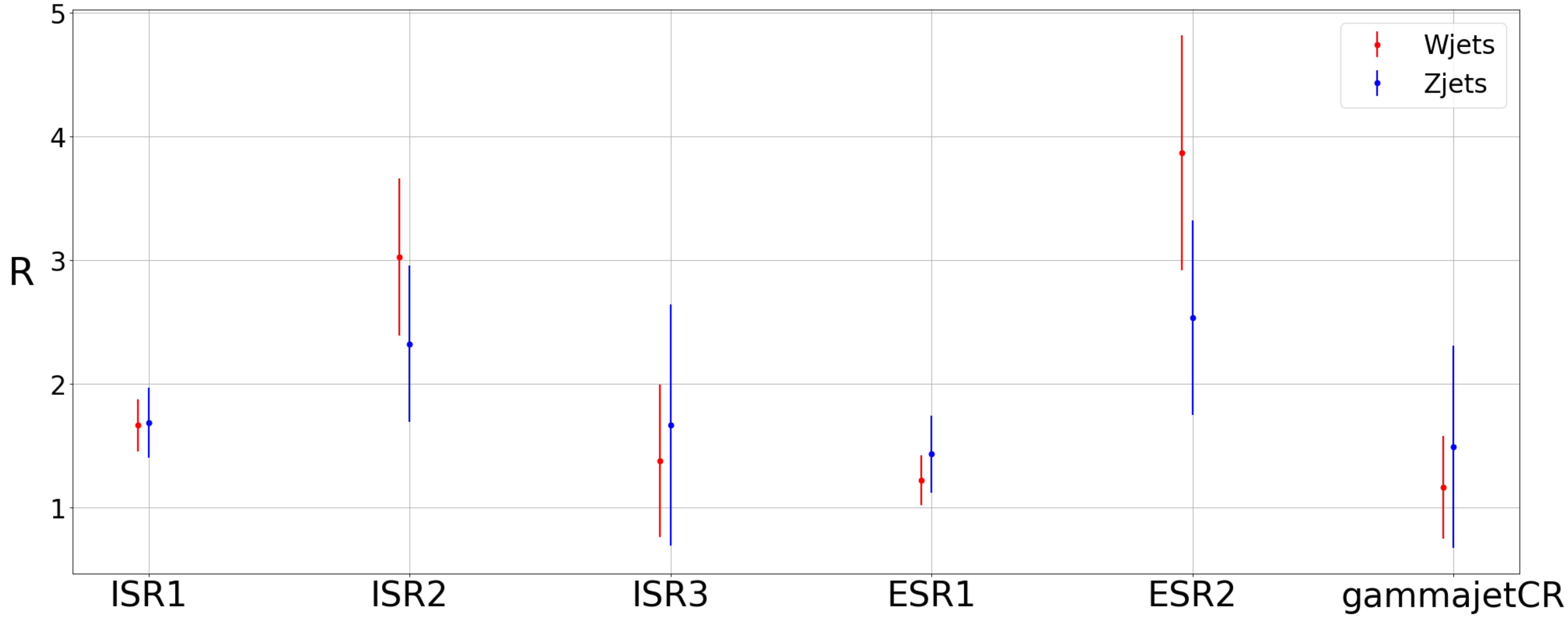
And the calculated purity is compatible with what expected within the statistical uncertainty:

	Purity [%]
Expected	84,3
Calculated	86,4 ± 5,9 ± 7,3

The method is validated.

Correlation Factor

The correlation factor has been calculated in each region on two MC samples of W + jets and Z + jets, selecting from these samples only reconstructed photons matched with true level jets. The results for the two samples are the following:

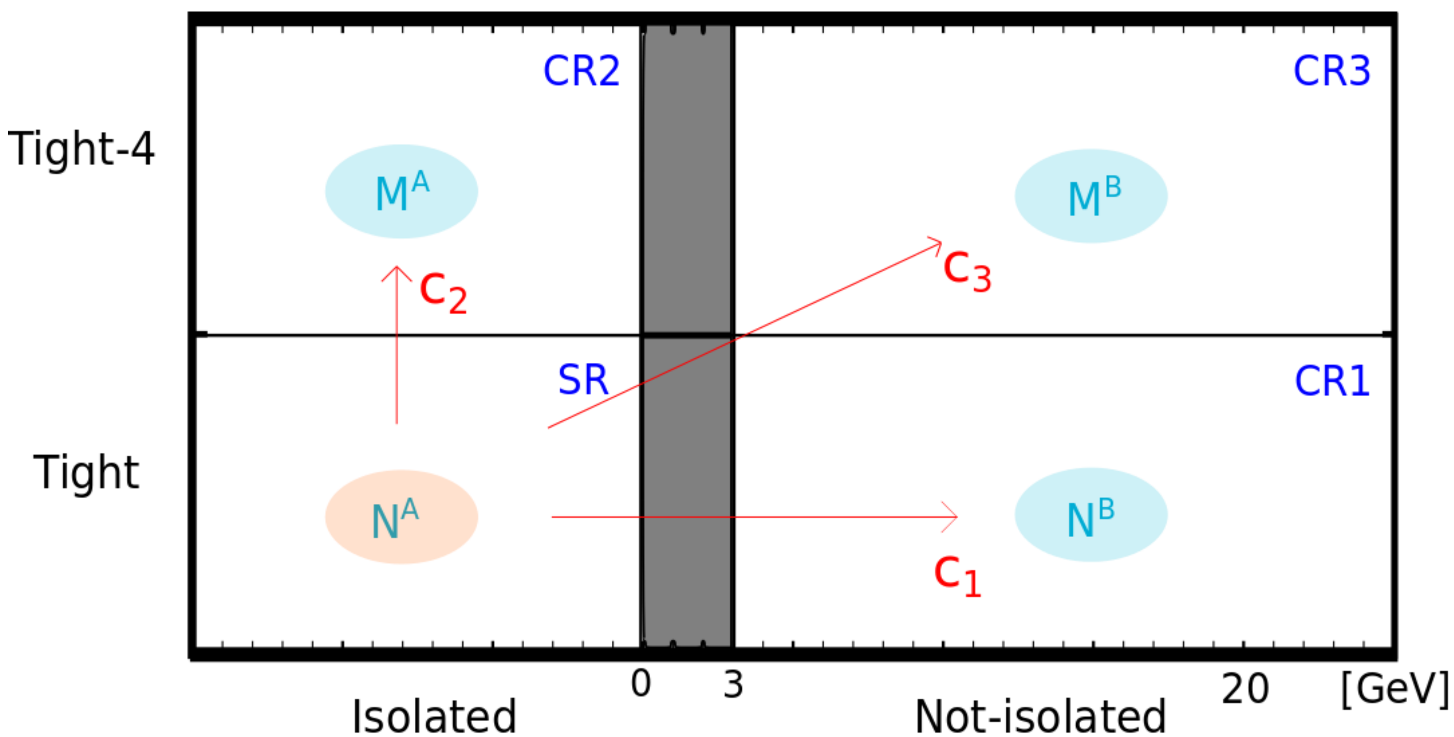


- Good compatibility between the two samples everywhere, no systematic uncertainty assigned
- To evaluate the compatibility between 2015-16 data and MC samples on the prediction of R_{MC} I computed also a correlation factor in a completely not-isolated region.

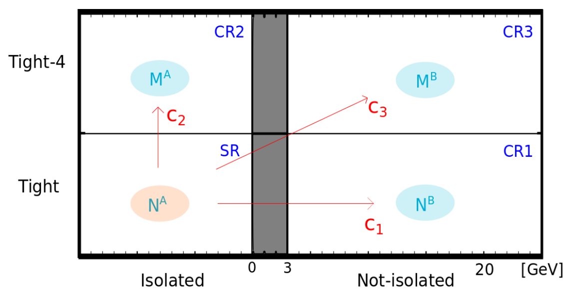
Signal Leakage

- The signal leakage coefficients have been calculated separately on three MC samples of:
  - W + γ
  - Z + γ
  - Z $(\nu\nu)$ + γ

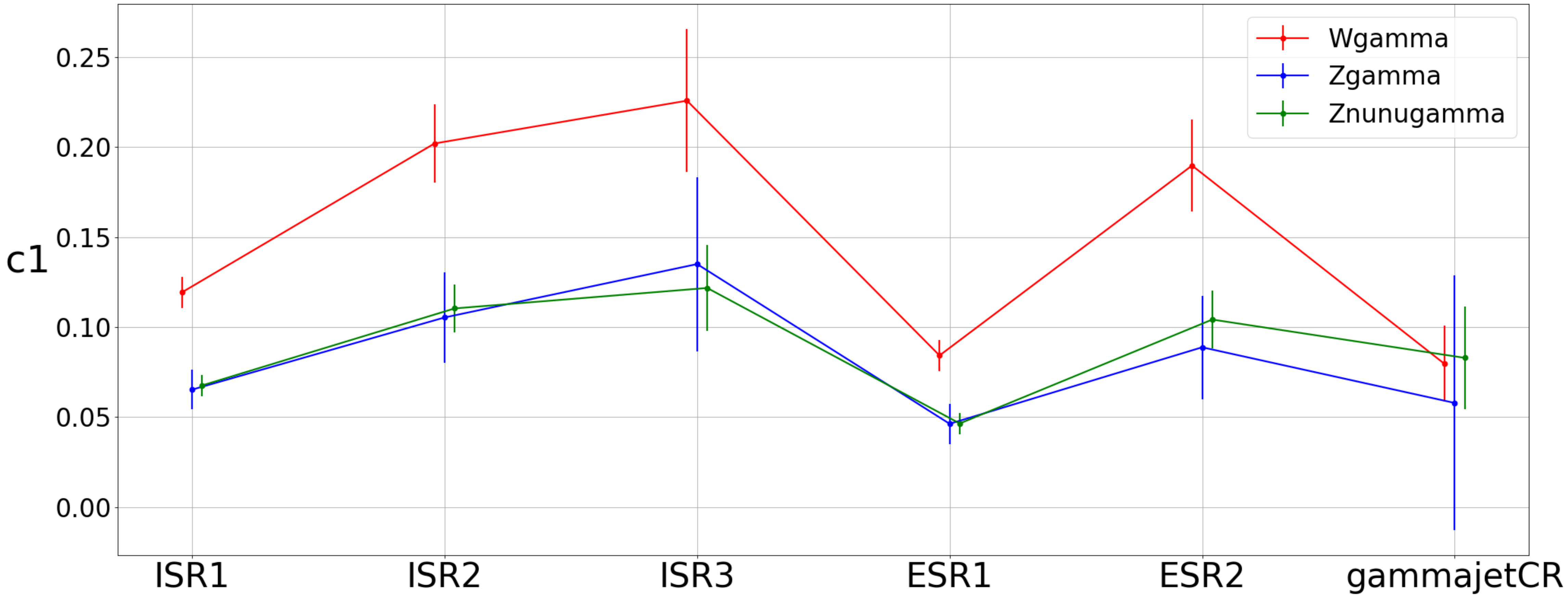
  selecting from this samples only reconstructed photons matched with true level photons.



Signal Leakage

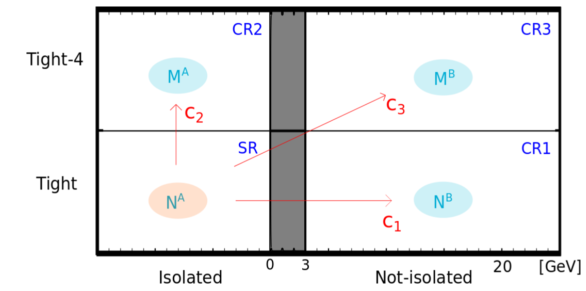


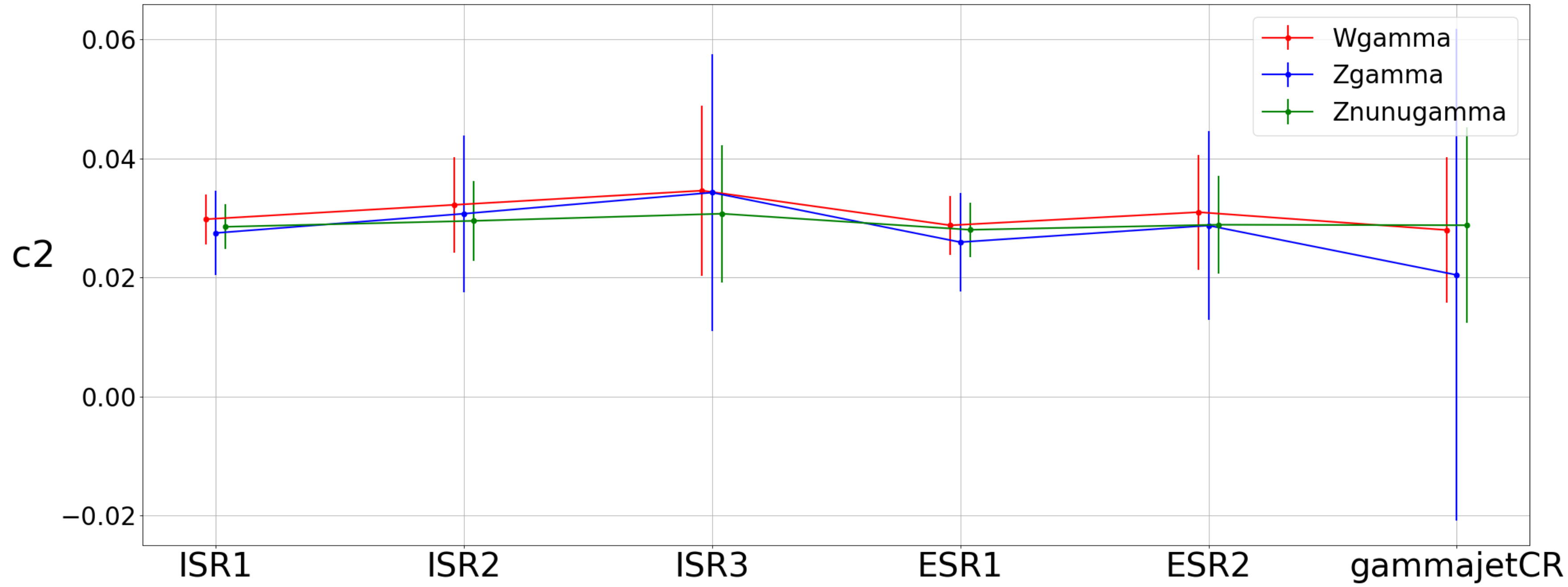
Results for the three samples are the following:

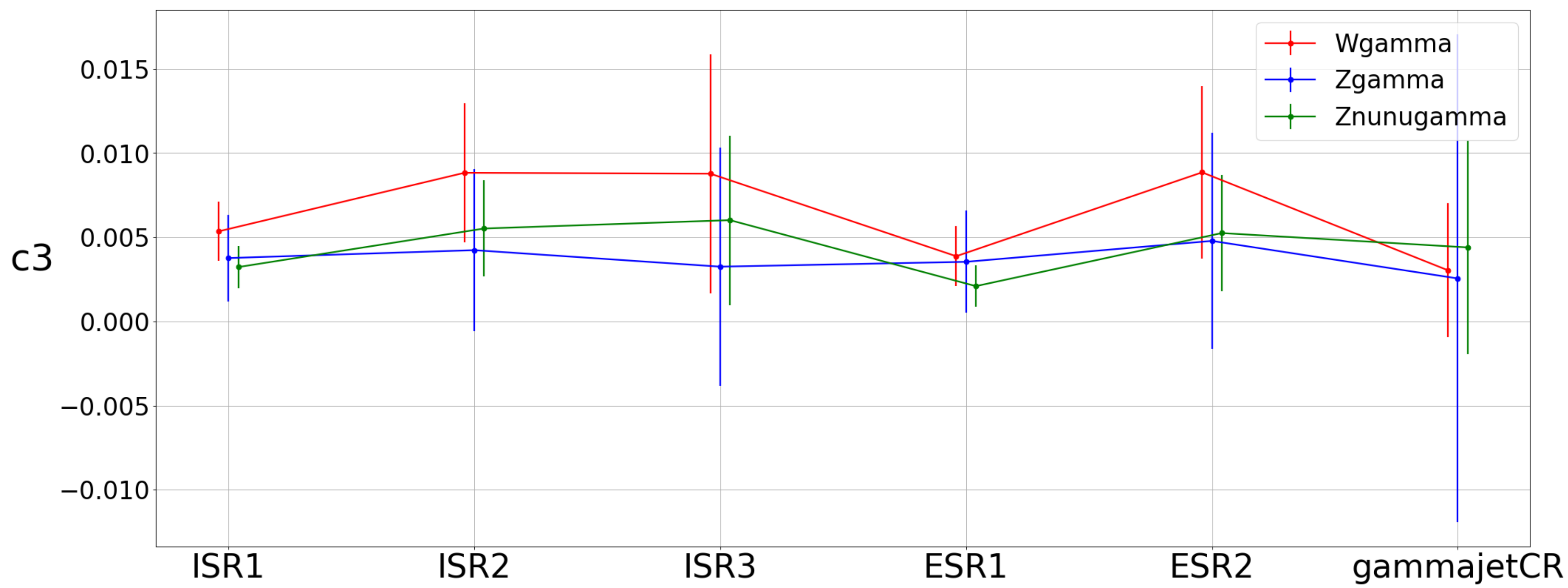


- Systematic difference between W + γ and Z + γ samples

Signal Leakage







Data analysis

- All the coefficients (c_1, c_2, c_3 and R_{MC}) are treated in the same way:
  - In each of the 5 + 1 regions we take the weighted mean of the 2 (R_{MC}) or 3 (c_1, c_2, c_3) values from different samples
  - Each of these 4 values is used in a specific SR and in all the corresponding CRs, except for the gammajetCR coefficients that are used only in this CR

- 2015-16 data analysis:
  - Nominal purities calculated in each region of the analysis
  - Statistical errors propagated only from data statistics (not MC statistics)
  - Systematic uncertainties evaluated changing the parameters of the method:
    - Tightness

      Tight-4 $\longrightarrow$ Tight-3, Tight-5
    - Isolation, not-isolated photon definition

      3 GeV $\longrightarrow$ 2 GeV, 4 GeV
    - Coefficients (c_1 and c_3 only)

      $c_{1,3} \longrightarrow c_{1,3} \pm$ RMS

  the maximum variation is quoted as systematic uncertainty.

Results

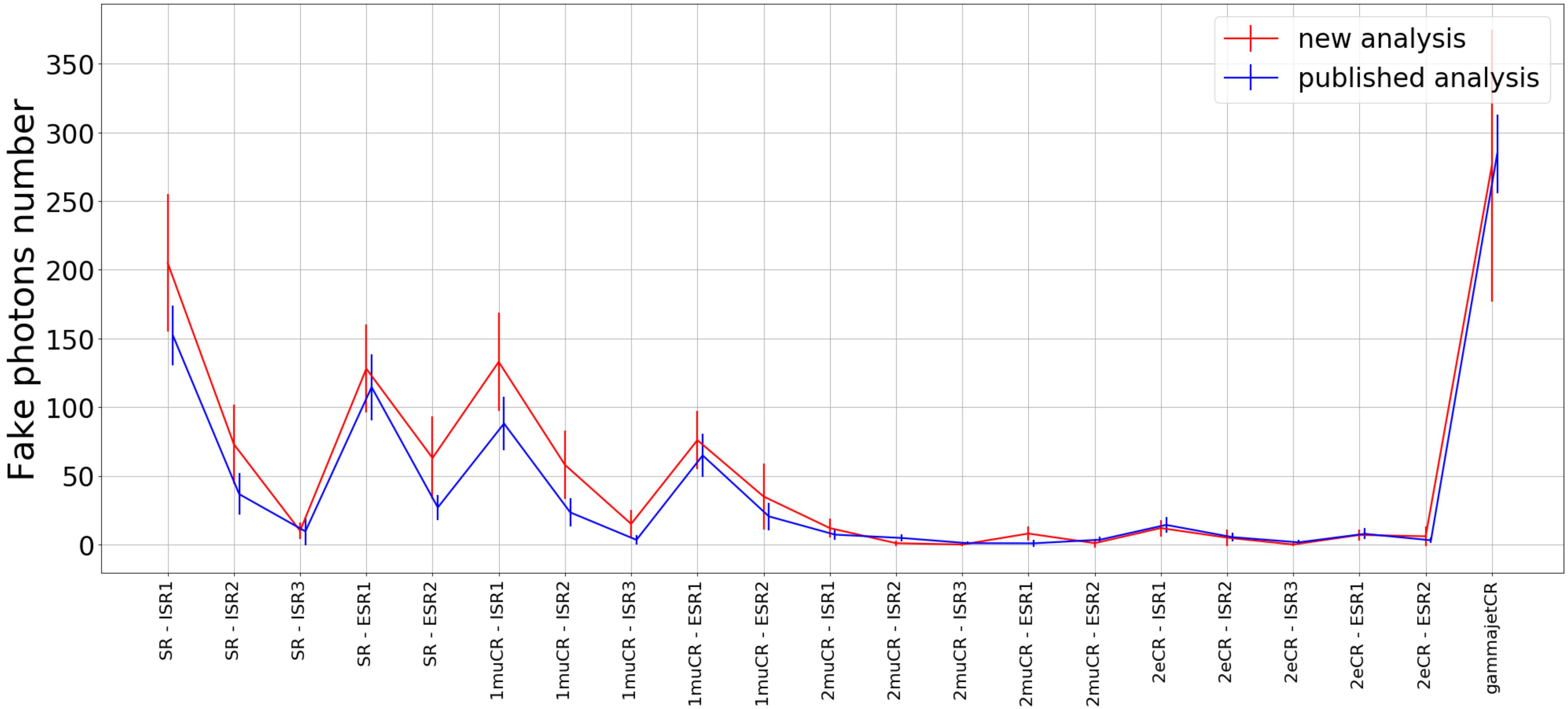
Number of fake photons in each region of the analysis for 2015-16 data.

regions	mean	stat.	tightness syst.	isolation syst.	c1 stat.	c2 stat.	c3 stat.	R stat.	c1 syst.	c3 syst.	total syst.	total error
SR - ISR1	205	21	36	3	3	8	0	22	13	0	45	50
SR - ISR2	73	15	19	2	2	6	0	13	6	0	25	29
SR - ISR3	10	4	2	0	1	2	0	4	1	0	5	6
SR - ESR1	128	15	20	5	2	6	0	17	7	0	29	32
SR - ESR2	63	16	20	3	2	6	0	14	5	0	26	30
1muCR - ISR1	133	15	28	5	1	3	0	14	4	0	32	36
1muCR - ISR2	58	12	19	3	1	2	0	10	2	0	22	25
1muCR - ISR3	15	5	6	1	0	1	0	6	0	0	8	10
1muCR - ESR1	76	10	14	2	1	2	0	10	2	0	18	21
1muCR - ESR2	35	9	21	1	1	2	0	7	2	0	22	24
2muCR - ISR1	12	4	5	0	0	1	0	1	1	0	6	7
2muCR - ISR2	1	2	1	0	0	1	0	0	0	0	1	2
2muCR - ISR3	0	0	0	0	0	0	0	0	0	0	0	1
2muCR - ESR1	8	3	3	1	0	0	0	1	1	0	4	5
2muCR - ESR2	1	3	1	0	0	1	0	0	0	0	1	3
2eCR - ISR1	12	5	3	1	0	0	0	1	1	0	4	6
2eCR - ISR2	5	4	4	0	0	0	0	1	0	0	4	6
2eCR - ISR3	0	1	0	0	0	0	0	0	0	0	0	1
2eCR - ESR1	7	3	3	0	0	0	0	1	1	0	3	4
2eCR - ESR2	6	5	5	1	0	0	0	1	0	0	5	7
gammajetCR	276	18	26	20	8	32	1	85	5	0	97	99

- Errors dominated by systematic uncertainties, mainly from tightness and the R statistics.

Results

Results are compatible within the total errors with the analysis published in "The ATLAS collaboration, *Search for dark matter at $\sqrt{s}$ = 13 TeV in final states containing an energetic photon and large missing transverse momentum with the ATLAS detector*. CERN, Geneva, Eur. Phys. J. C 77 (2017) 393."



Conclusions

- I employed the 2D sideband method to estimate the jets faking photons background for the Mono-Photon analysis

- The method has been validated on a mixed MC sample

- The method has been applied to 2015-16 data and different MC samples
  - Compatible results with the published analysis
  - Errors dominated by systematic uncertainties

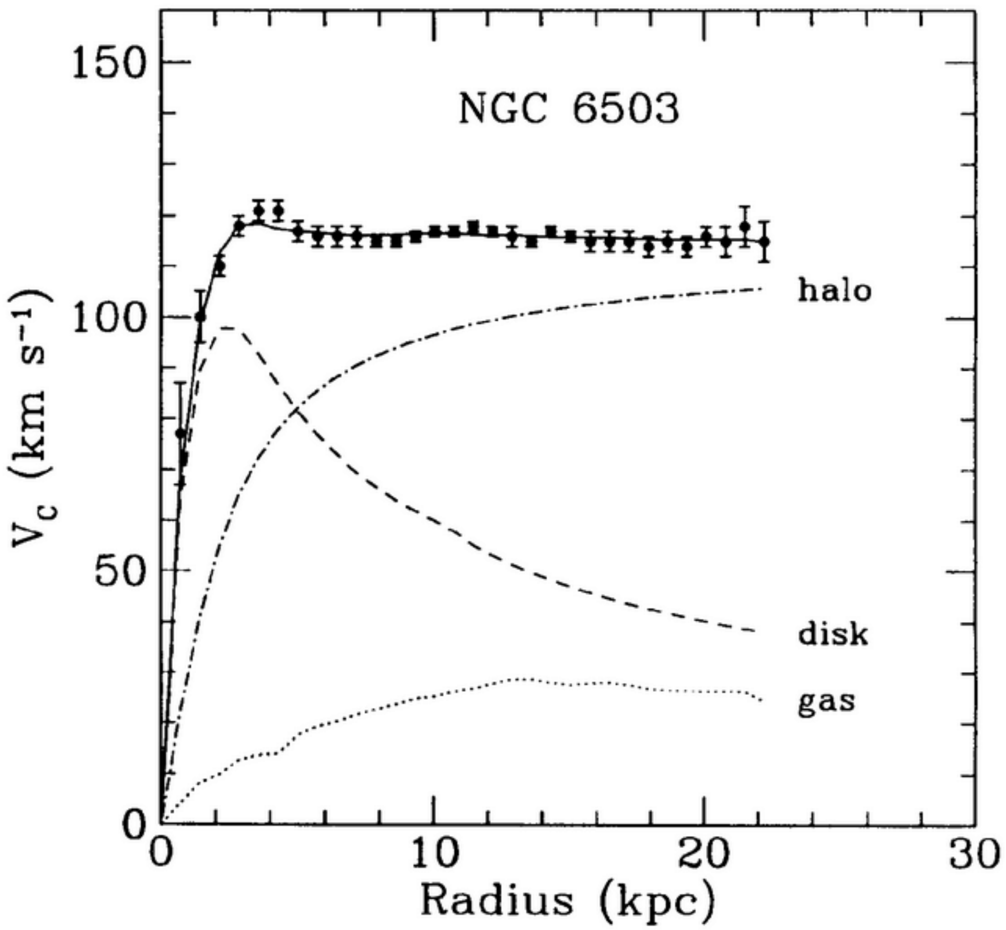
- Preliminary results on full Run-2 data (2015-18) have been produced
  - The analysis is still blinded
  - Ongoing consolidation of systematic uncertainties

Thank you for the attention!

Backup

Dark Matter observations

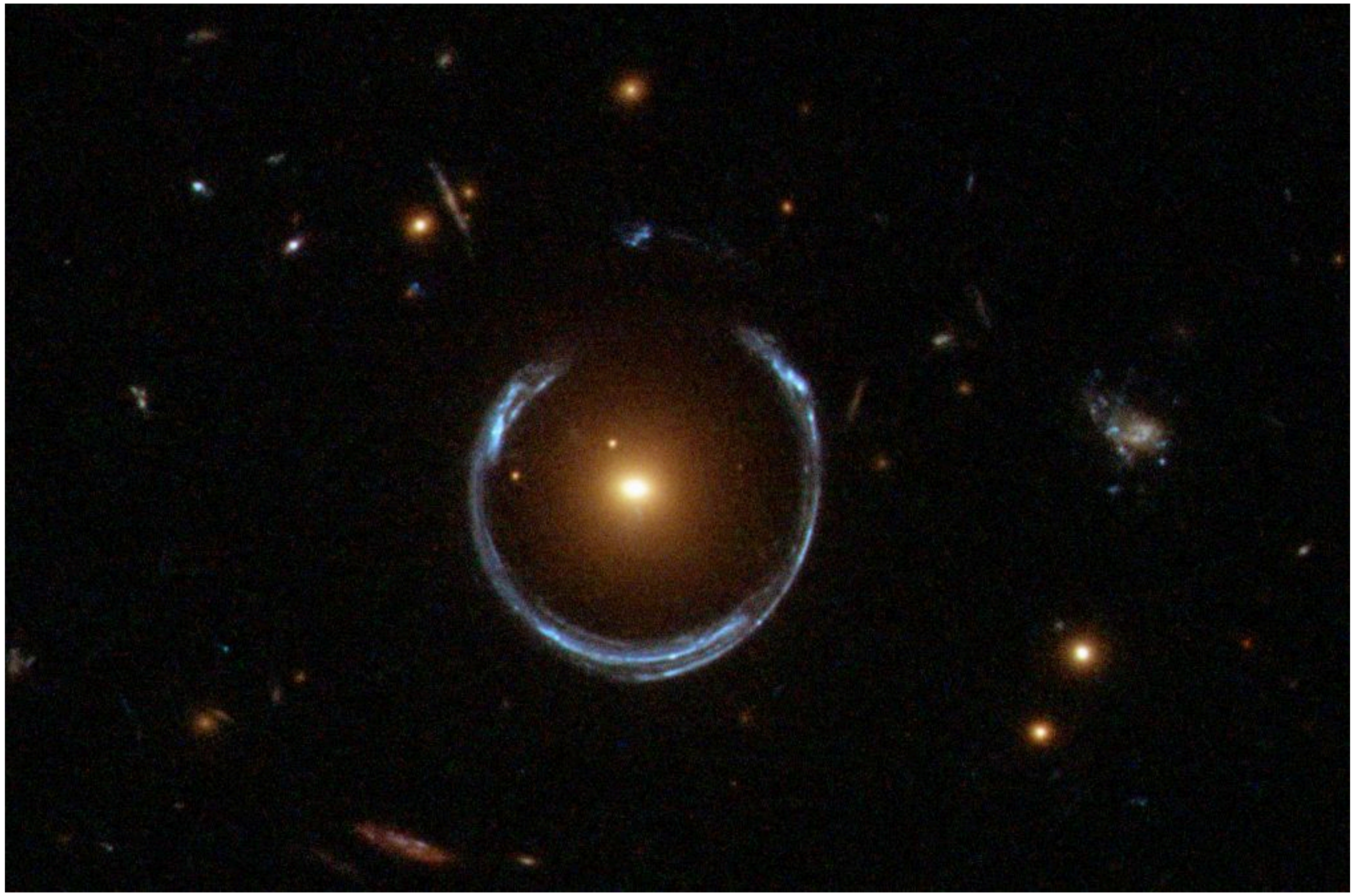
- Rotation curve of galaxy NGC 6503. It is evident that the halo contribution is necessary to justify experimental observations.



From: Doroshkevich, A., Lukash, Vladimir, Mikheeva, Elena, *A solution to the problems of cusps and rotation curves in dark matter halos in the cosmological standard model*. Phys. Usp. 55 3-17 (2012).

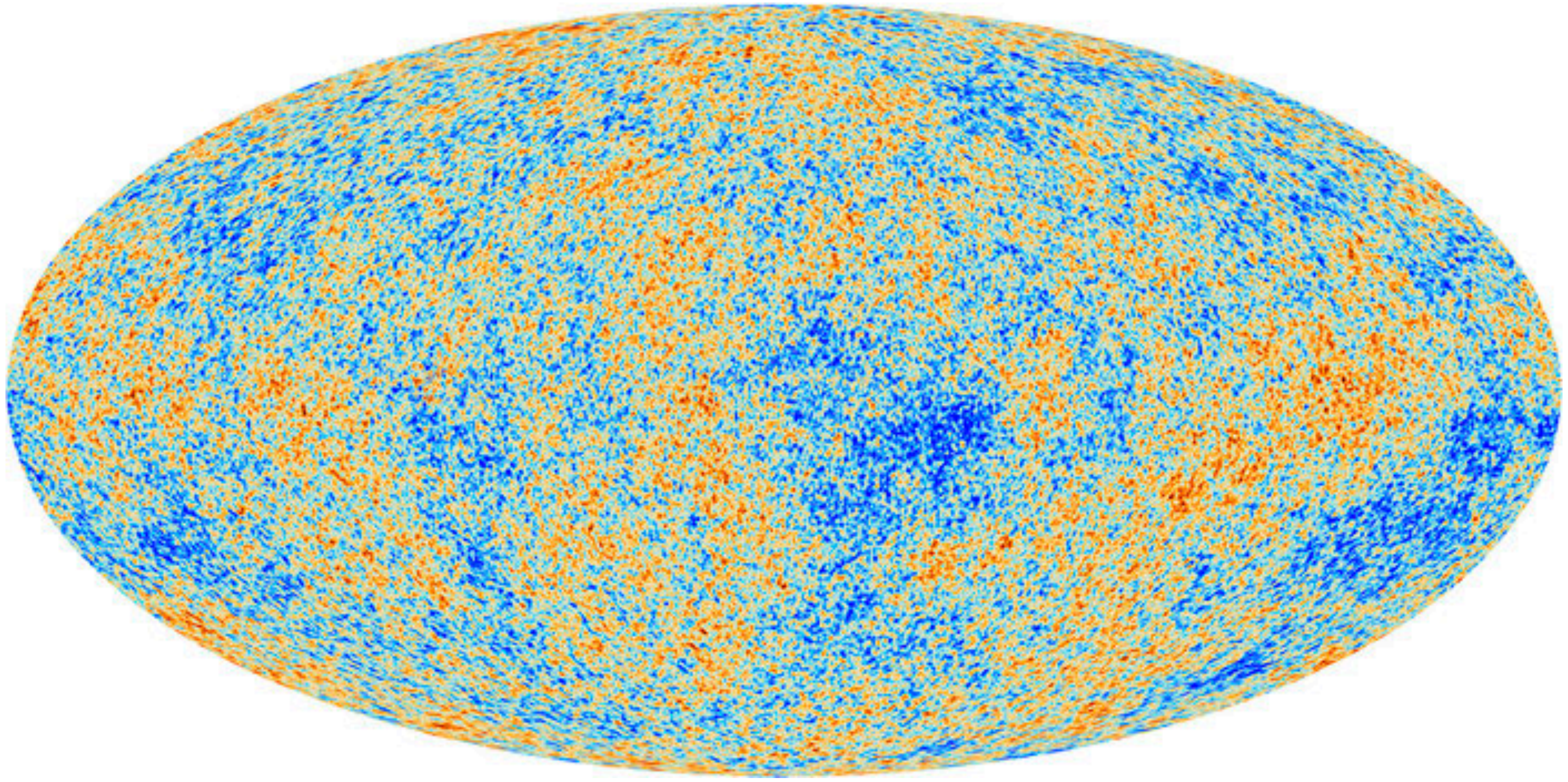
Dark Matter observations

- Gravitational lensing: the source of light is a blue galaxy in the background, while the object deviating the light is the yellow galaxy (LRG 3-757).



Dark Matter observations

- Cosmic Microwave Background (CMB) measured by PLANK experiment.



From: Planck collaboration: P. A. R. Ade et al, *Planck 2013 results. I. Overview of products and scientific results*.

Dark Matter candidates

- WIMPs are predicted by a number of consistent theories extending the SM, such as Supersymmetry and Extra Dimensions.

- The abundance of Dark Matter is consistent with a Dark Matter particle of mass in the WIMP mass range (10 GeV - 1 TeV). These particles are expected to interact with ordinary matter by means of the weak force and they would have the same abundance of DM, compatible with the weak cross section. This is known as the WIMP miracle, which assumes that Dark Matter was created thermally in the early stages of the Universe and can now only annihilate.

- Alternative proposals:
  - WISP, Weakly Interacting Sub-eV Particles;
  - Modification of Newton dynamics (MOND theory) and General Relativity, however these theories are still controversial and don't really explain all the cosmological observations.

E_T^{miss}

To reconstruct the E_T^{miss} an algorithm sums up all contributions from all the reconstructed objects. A Soft Term is also included to account for energy clusters not topologically connected to any of the reconstructed objects and account for low pt objects, electronic noise, pile-up and cosmic rays background.

In the x (y) direction the sum of the missing energies is:

$$E_{x(y)}^{miss} = E_{x(y)}^{miss,e} + E_{x(y)}^{miss,\gamma} + E_{x(y)}^{miss,\tau} + E_{x(y)}^{miss,jets} + E_{x(y)}^{miss,mu} + E_{x(y)}^{miss,SoftTerm}$$

Where $E_{x(y)}{}^{miss,k}$ is the negative sum of the momenta of all reconstructed object ok tipe k.

The E_T^{miss} is then:

$$E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2}$$

Shower Shapes

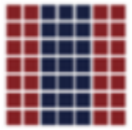
Variables and Position

	Strips	2nd	Had.
Ratios	f_1, f_{side}	R_η*, R_ϕ	$R_{Had.}$*
Widths	$w_{s,3}$, $w_{s,tot}$	$w_{\eta,2}$*	-
Shapes	ΔE, E_{ratio}		

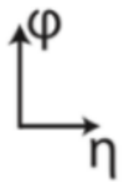
* Used in PhotonLoose.

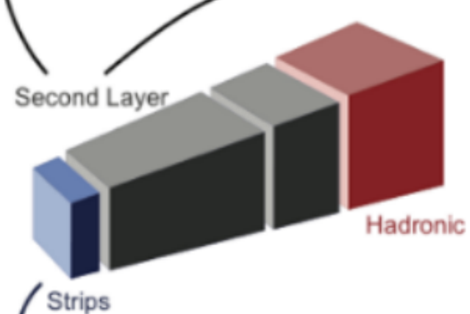
Energy Ratios

$$R_\eta = \frac{E^{S2}_{3\times7}}{E^{S2}_{7\times7}}$$



$$R_\phi = \frac{E^{S2}_{3\times3}}{E^{S2}_{3\times7}}$$





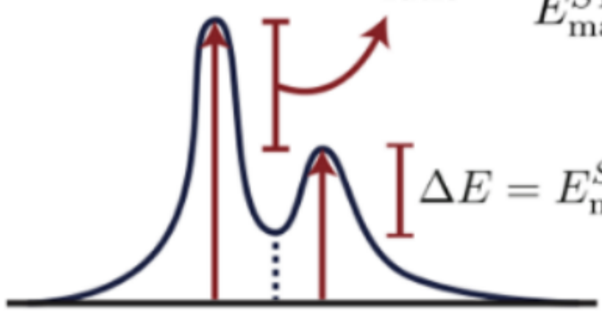
$$R_{\mathrm{Had}} = \frac{E_T^{\mathrm{Had}}}{E_T}$$

$$f_1 = \frac{E_{S1}}{E_{\mathrm{Tot.}}}$$

$$f_{\mathrm{side}} = \frac{E_7^{S1} - E_3^{S1}}{E_3^{S1}}$$

Shower Shapes

$$E_{\mathrm{ratio}} = \frac{E^{S1}_{\mathrm{max},1} - E^{S1}_{\mathrm{max},2}}{E^{S1}_{\mathrm{max},1} + E^{S1}_{\mathrm{max},2}}$$

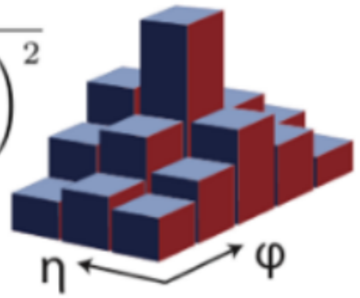


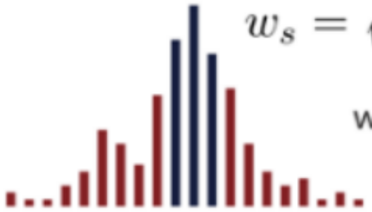
$$\Delta E = E^{S1}_{\mathrm{max},2} - E^{S1}_{\mathrm{min}}$$

Widths

$$w_{\eta,2} = \sqrt{\frac{\sum E_i \eta_i^2}{\sum E_i} - \left(\frac{\sum E_i \eta_i}{\sum E_i}\right)^2}$$

Width in a 3×5 (Δη×Δφ) region of cells in the second layer.





$$w_s = \sqrt{\frac{\sum E_i (i - i_{\mathrm{max}})^2}{\sum E_i}}$$

ws3 = w_1 uses 3 strips in η; wstot is defined similarly, but uses 20 strips.

The 2D sideband method

We start with two simplifying hypothesis:

1. The correlation between x and y is negligible for the background
2. The number of signal candidates is negligible compared to the number of fake candidates is negligible in the control regions:

$$N^B_{\mathrm{bkg}} \gg N^B_{\mathrm{sig}}$$
$$M^A_{\mathrm{bkg}} \gg M^A_{\mathrm{sig}}$$
$$M^B_{\mathrm{bkg}} \gg M^B_{\mathrm{sig}}$$

As a consequence of the first assumption we can assume that:

$$N^A_{\mathrm{bkg}}/N^B_{\mathrm{bkg}} = M^A_{\mathrm{bkg}}/M^B_{\mathrm{bkg}}$$

and as a consequence of the second assumption:

$$N^B = N^B_{\mathrm{bkg}}$$
$$M^A = M^A_{\mathrm{bkg}}$$
$$M^B = M^B_{\mathrm{bkg}}$$

The 2D sideband method

Therefore combining the two hypothesis:

$$N^A_{\text{bkg}} = N^B_{\text{bkg}} \times M^A_{\text{bkg}}/M^B_{\text{bkg}} = N^B \times M^A/M^B$$

which provides a fully data-driven technique to estimate the background.

The signal in region A is therefore:

$$N^A_{\text{sig}} = N^A - N^A_{\text{bkg}} = N^A - N^B \frac{M^A}{M^B}$$

so that the purity is:

$$P = N^A_{\text{sig}}/N^A = 1 - \frac{N^B}{N^A}\frac{M^A}{M^B}$$

which is valid only if the two assumptions are satisfied.

The 2D sideband method

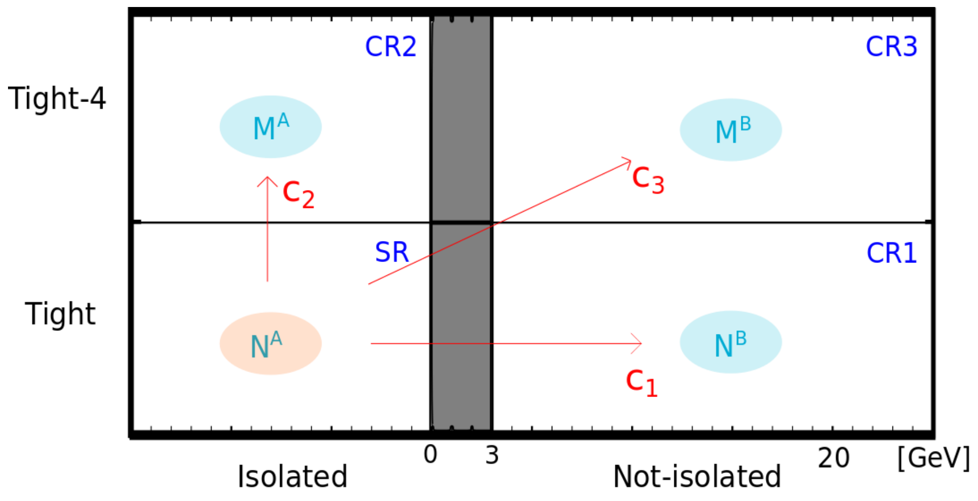
If one or both of the two hypothesis are not satisfied we can correct the formula for the purity using Monte Carlo

1. Non negligible correlation in the background between x and y

$$
\begin{aligned}
N^A_{\rm sig} &= N^A - N^A_{\rm bkg} = N^A - N^A_{\rm bkg}\frac{N^B_{\rm bkg}\times M^A_{\rm bkg}/M^B_{\rm bkg}}{N^B_{\rm bkg}\times M^A_{\rm bkg}/M^B_{\rm bkg}} \\
&\approx N^A - \left(N^B\frac{M^A}{M^B}\right)\left(\frac{N^A_{\rm bkgMC}}{N^B_{\rm bkgMC}}\frac{M^B_{\rm bkgMC}}{M^A_{\rm bkgMC}}\right) \\
&\approx N^A - \left(N^B\frac{M^A}{M^B}\right)R_{\rm MC}
\end{aligned}
$$

2. Signal leakage in background control regions

$$
\begin{aligned}
N^B &= N^B_{\rm bkg} + N^B_{\rm sig} = N^B_{\rm bkg} + N^A_{\rm sig}\frac{N^B_{\rm sig}}{N^A_{\rm sig}} \\
M^A &= M^A_{\rm bkg} + M^A_{\rm sig} = M^A_{\rm bkg} + N^A_{\rm sig}\frac{M^A_{\rm sig}}{N^A_{\rm sig}} \\
M^B &= M^B_{\rm bkg} + M^B_{\rm sig} = M^B_{\rm bkg} + N^A_{\rm sig}\frac{M^B_{\rm sig}}{N^A_{\rm sig}}
\end{aligned}
\qquad \text{where} \qquad
\begin{aligned}
\frac{N^B_{\rm sig}}{N^A_{\rm sig}} &= c_1 \approx \frac{N^B_{\rm sigMC}}{N^A_{\rm sigMC}} \\
\frac{M^A_{\rm sig}}{N^A_{\rm sig}} &= c_2 \approx \frac{M^A_{\rm sigMC}}{N^A_{\rm sigMC}} \\
\frac{M^B_{\rm sig}}{N^A_{\rm sig}} &= c_3 \approx \frac{M^B_{\rm sigMC}}{N^A_{\rm sigMC}}
\end{aligned}
$$



The 2D sideband method

Substituting we obtain:

$$N^A_{\text{sig}} \quad = \quad \frac{(M^B + N^A c_3 - N^B c_2 R_{\text{MC}} - M^A c_1 R_{\text{MC}})(-1 + \sqrt{1 + \frac{4(c_1 c_2 R_{\text{MC}} - c_3)(N^A M^B - N^B M^A R_{\text{MC}})}{(M^B + N^A c_3 - N^B c_2 R_{\text{MC}} - M^A c_1 R_{\text{MC}})^2}}}{2(c_1 c_2 R_{\text{MC}} - c_3)}$$

and the purity straightforward.

We can also approximate this formula if:

$$\left| \frac{4(c_1 c_2 R_{\text{MC}} - c_3)(N^A M^B - N^B M^A R_{\text{MC}})}{(M^B + N^A c_3 - N^B c_2 R_{\text{MC}} - M^A c_1 R_{\text{MC}})^2} \right| \ll 1$$

so that:

$$N^A_{\text{sig}} \quad = \quad \left(N^A - N^B \frac{M^A}{M^B} R_{\text{MC}} \right) \frac{1}{1 + \frac{c_3 N^A - c_2 N^B R_{\text{MC}} - c_1 M^A R_{\text{MC}}}{M^B}}$$

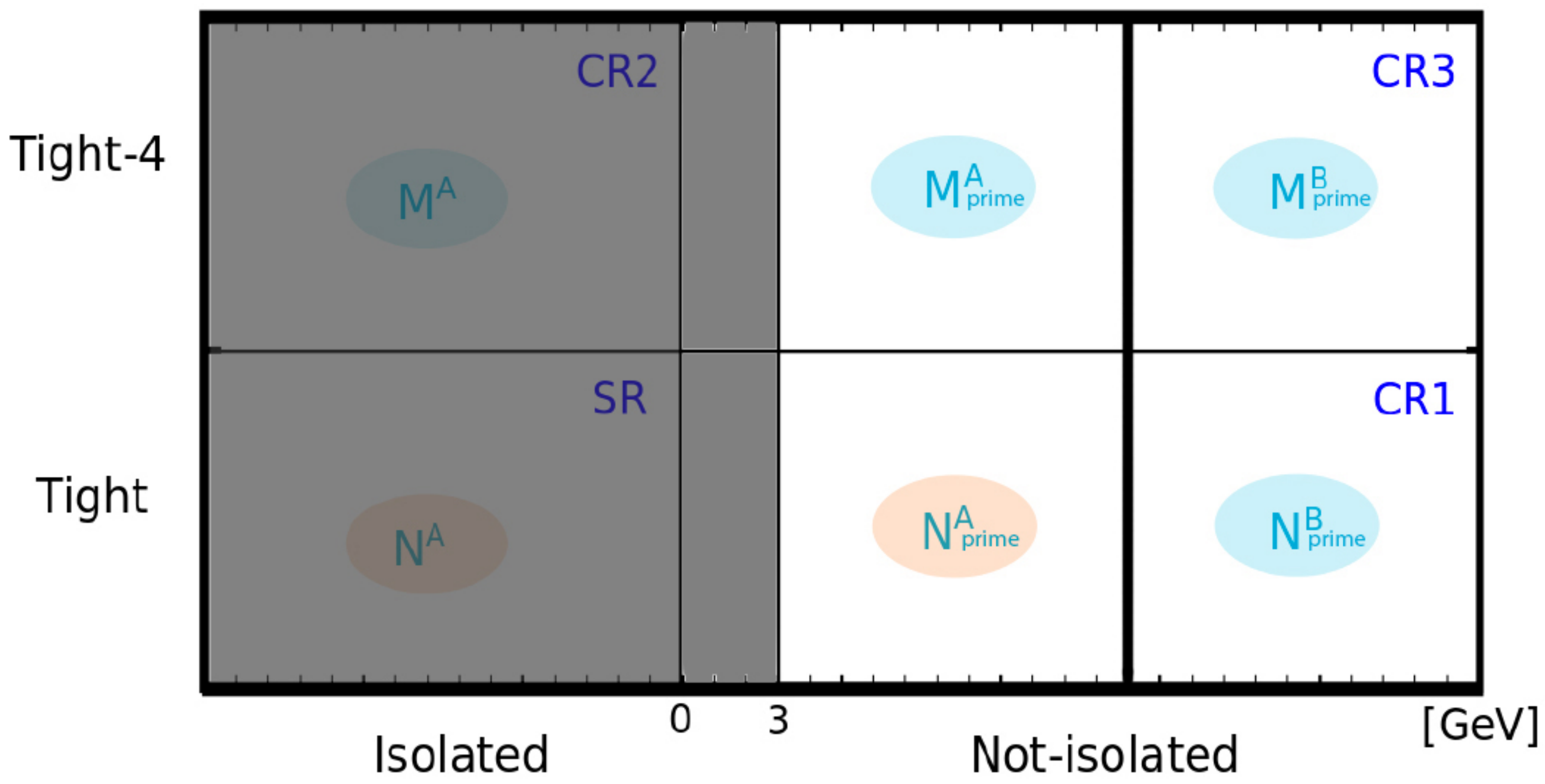
more reliable when we compute the derivatives for the propagation of uncertainty.

R prime

To evaluate the compatibility on the predictions on R_{MC} I calculated a correlation factor in a completely not-isolated region, where no signal is expected, both for Monte Carlo samples and for 2015-16 data.

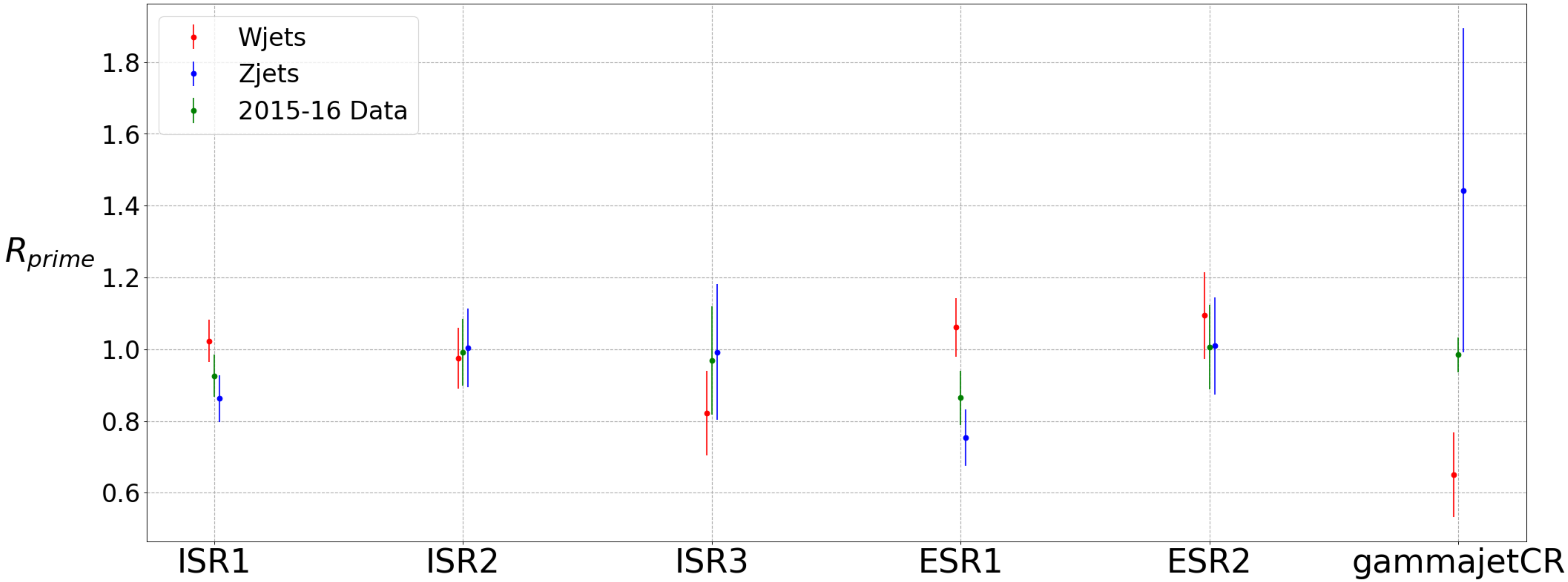
The rectangular division in the not-isolated region is defined by the following parameters:

- Calorimetric isolation (E_{iso}): 50 GeV
- Track isolation: 0.2



R prime

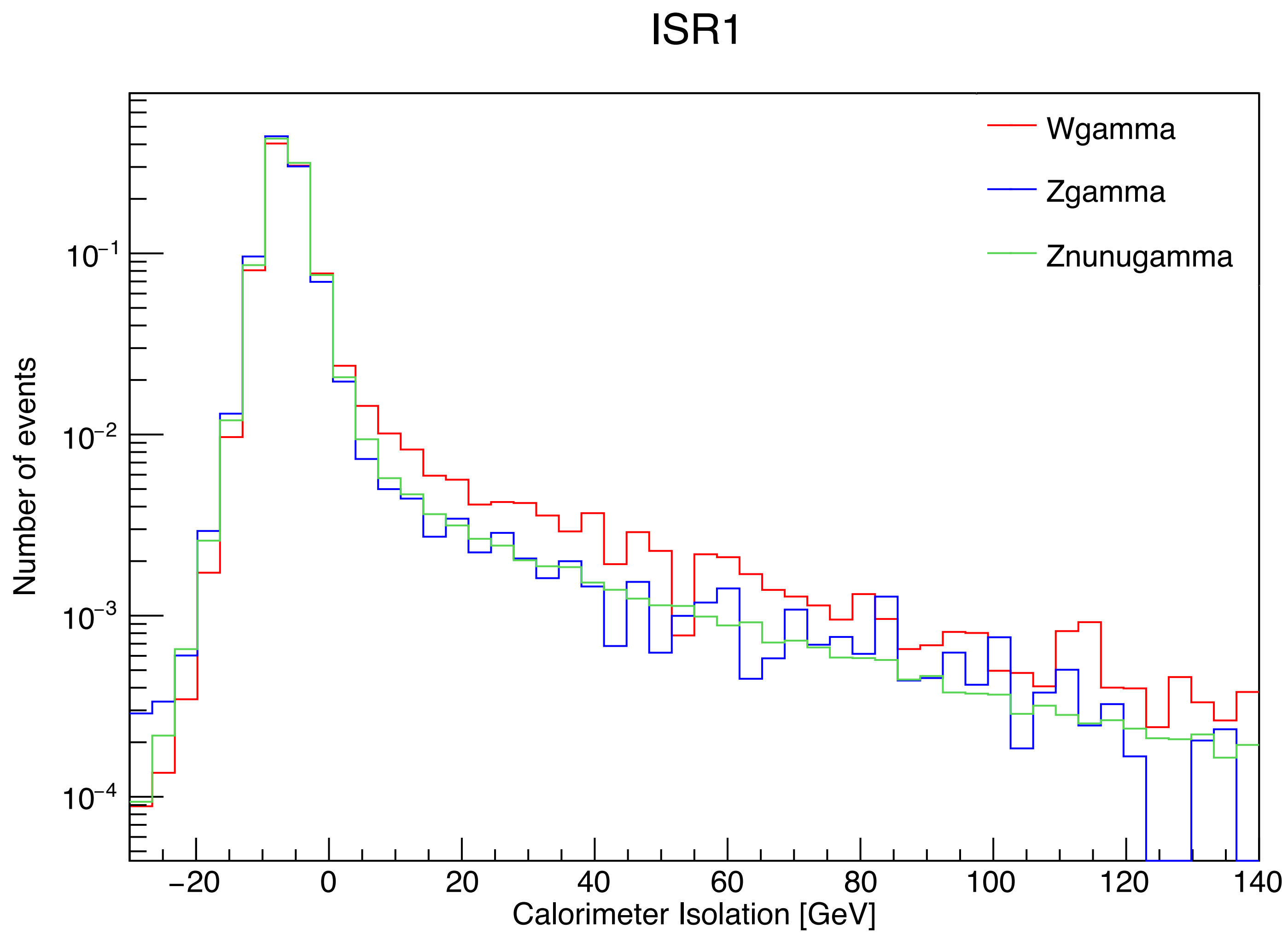
Results in the different regions are the following:



- Fair compatibility between MCs and 2015-16 data

Signal Leakage

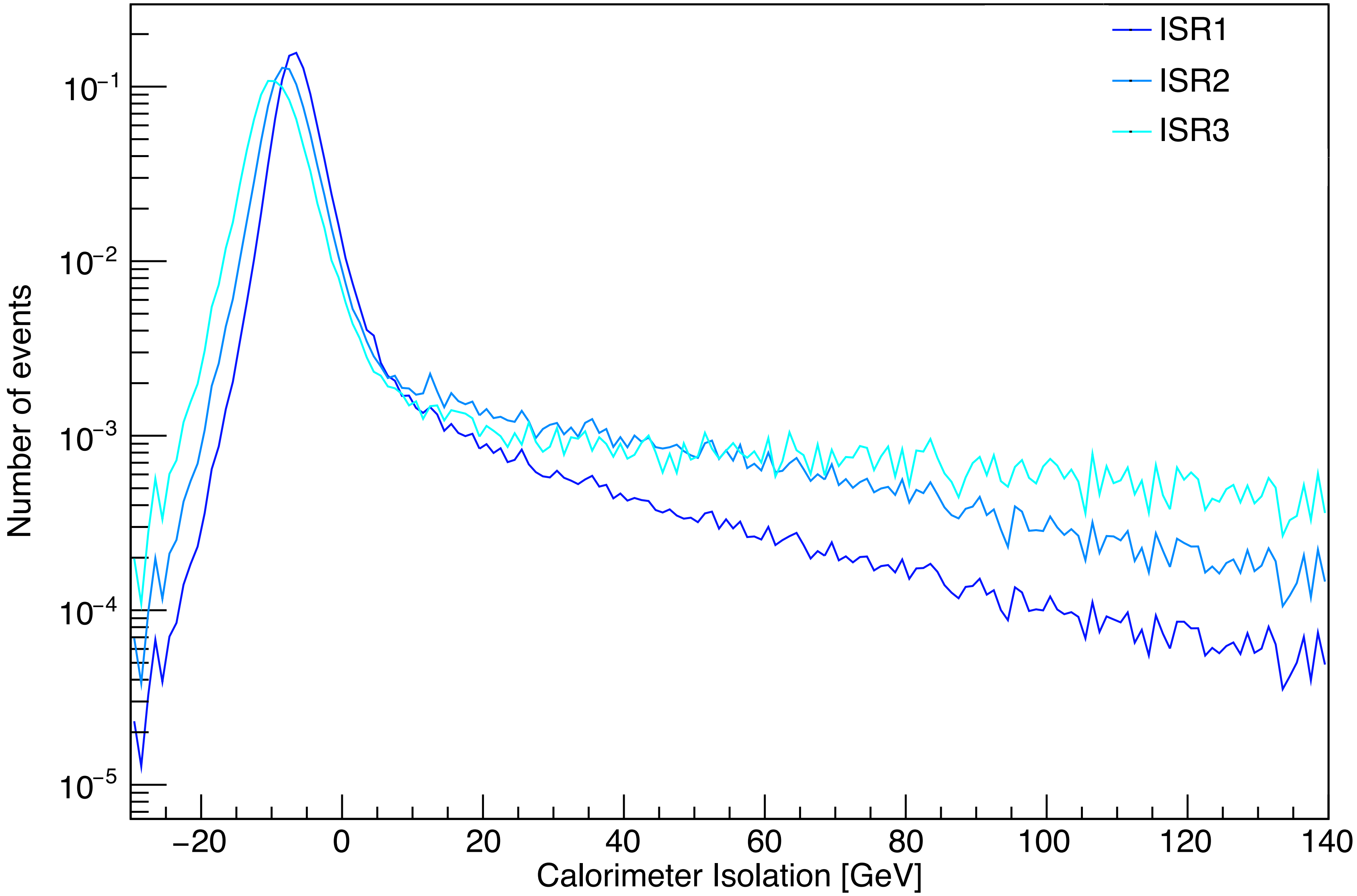
- Results for Z + γ and Z ($\nu\nu$) + γ compatible everywhere
- c_1 systematically higher for W + γ
  - Different tails in the calorimeter isolation profile



Signal Leakage

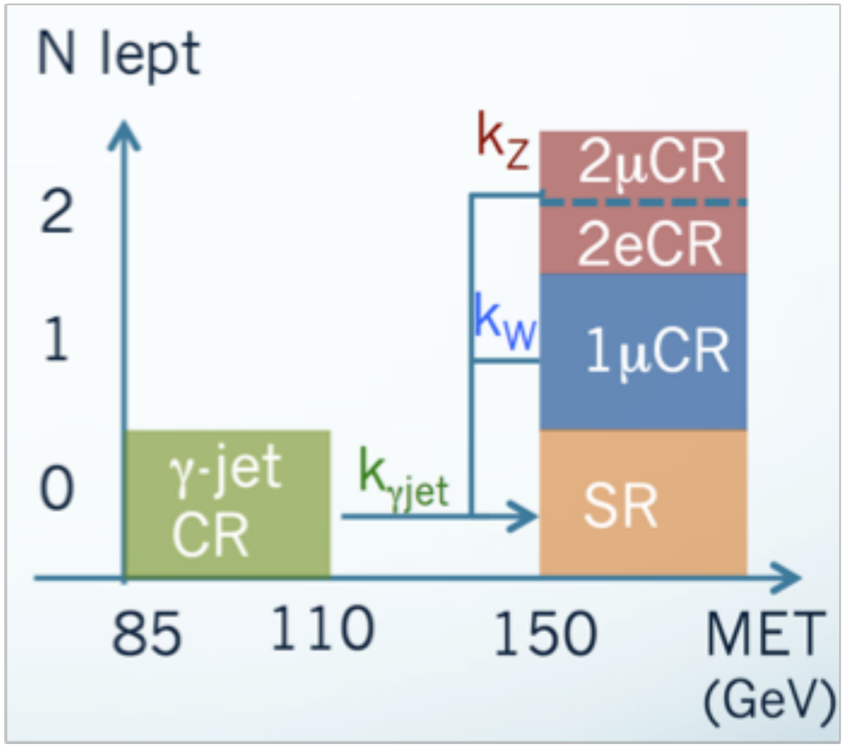
- E_T^{miss} dependence, difference in the tails of the isolation profile

Z(νν) + γ



Mono-Photon CRs

- Scheme of the Control Regions of the Mono-Photon analysis and corresponding k-factors needed to the normalization of the different backgrounds in the Signal Region.



Signal region	E_T^{miss} [GeV]	$k_{W\gamma}$	$k_{Z\gamma}$	$k_{\gamma+jets}$	$k'_{W\gamma}$	$k'_{Z\gamma}$	$k'_{\gamma+jets}$
SRI1	> 150	1.05 ± 0.09	1.10 ± 0.09	1.07 ± 0.25			
SRI2	> 225	1.04 ± 0.11	1.14 ± 0.13	1.06 ± 0.25			
SRI3	> 300	1.04 ± 0.15	1.27 ± 0.23	1.06 ± 0.24	1.03 ± 0.14	1.27 ± 0.23	
SRE1	150–225				1.06 ± 0.10	1.10 ± 0.10	1.07 ± 0.25
SRE2	225–300				1.02 ± 0.12	1.09 ± 0.14	

Mono-Photon Results

- Mono-Photon exclusion plots: 95% CL exclusion contours, where the free parameters are the particle mass m_χ, the mediator mass m_{med}, the width of the mediator Γ_{med}, and the couplings g_q, g_χ and g_l.