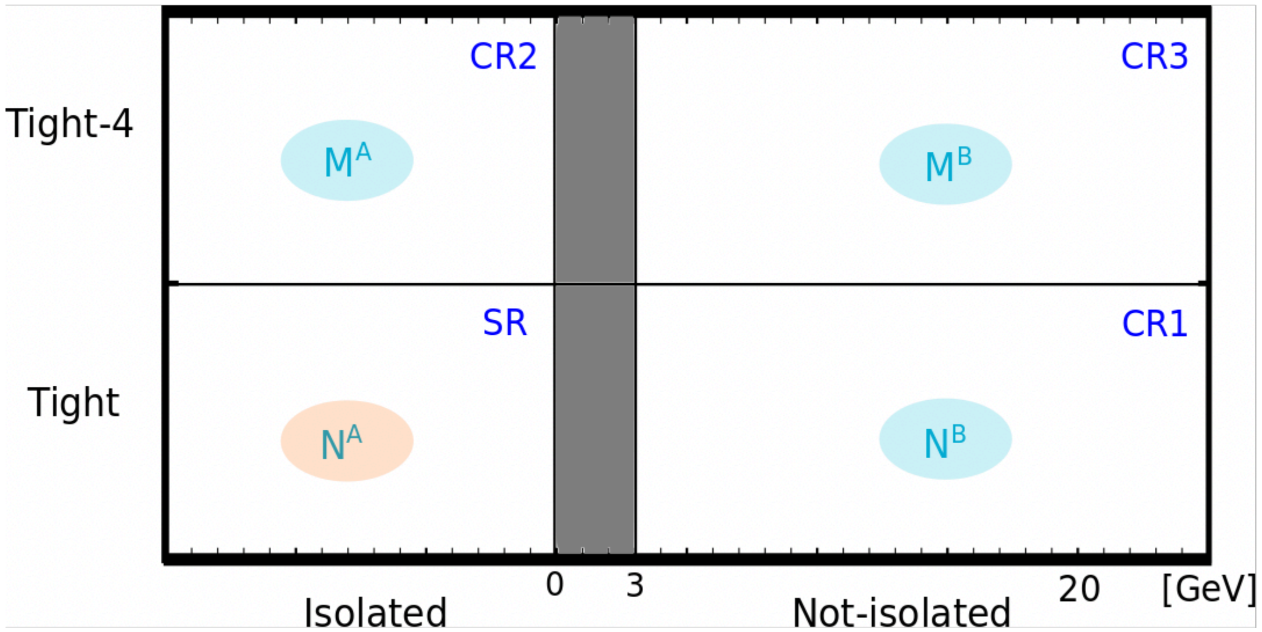
Measurement of the jets faking photons background for the Dark Matter search in the mono photon channel with the ATLAS detector

The 2D sideband method

In the 2D sideband method we consider candidates in a 2D plane (x,y) formed by a photon identification variable x and an isolation 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
  - Tight-3 and Tight-5 selections are used to estimate the systematic uncertainty
- Isolation:
  - E_{iso} = TopoEtCone40 - 0.022pt - 2.45 GeV
    - $E_{iso} < 0$ GeV : isolated photon
    - $E_{iso} > 3$ GeV : not-isolated photon
  - A Gap of 3 GeV is excluded to prevent signal leakage.

- The plane is so divided in 4 regions. The number of photons in each region is:
  - N^A : tight isolated (signal region)
  - N^B : tight not-isolated (control region 1)
  - M^A : not-tight isolated (control region 2)
  - M^B : not-tight not-isolated (control region 3)

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:

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

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

The signal in region A is therefore:

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

so that the purity is:

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

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:

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

and the purity straightforward.

We can also approximate this formula if:

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

so that:

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

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

The systematic uncertainties are evaluated moving the calorimeter isolation threshold from 3 to 2 and 4, and the not-tight region from Tight-4 to Tight-3 and Tight-5, than taking the maximum difference from the original purity. The total systematic uncertainty is computed as the square sum of the two uncertainties.

Control Regions

The Signal Region (SR), is divided in 5 regions with different MET cuts:

	ISR1	ISR2	ISR3	ESR1	ESR2
MET [GeV]	> 150	> 225	> 300	150 - 225	225 -300

For each of this regions we define 3 control regions, each enriched with different processes:

- one muon CR (1muCR) $\longrightarrow$ W (μν) + γ
- two muon CR (2muCR) $\longrightarrow$ Z (μμ) + γ
- two electron CR (2muCR) $\longrightarrow$ Z (ee) + γ

We define one more control region, the gammajetCR, dominated by jet + γ process. This region can't be divided as the other because it has a different MET cut (85 GeV < E_T^{miss} < 110 GeV).

Validation

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

- The correlation factor (R_{MC}) and the signal leakage coefficients (c1, c2, c2) are calculated on the same sample
- Focus on the ISR1 of the SR
- Two different calculi of the purity, taking and not taking into account Monte Carlo weights

In this region the coefficients are:

	R_{MC}	c_1	c_2	c_3
Not Weighted	1,60 ± 0,24	6,98 ± 0,42 %	2,94 ± 0,27 %	0,31 ± 0,09 %
Weighted	2,76 ± 0,49	7,08 ± 0,80 %	4,50 ± 0,63 %	0,49 ± 0,21 %

At first the not weighted purity wasn't compatible with the result we expected, the reason is that the approximation made is not satisfied. We then used the complete formula calculating all the derivatives for the propagation of the uncertainty with Mathematica.

The results with propagated and systematic uncertainties respectively are:

	Not weighted purity	Weighted purity
Expected	96,96%	84,27%
Calculated	97,05 ± 0,91 ± 1,56 %	86,41 ± 5,87 ± 7,26%

Compatible with the expected purities.

Correlation factor

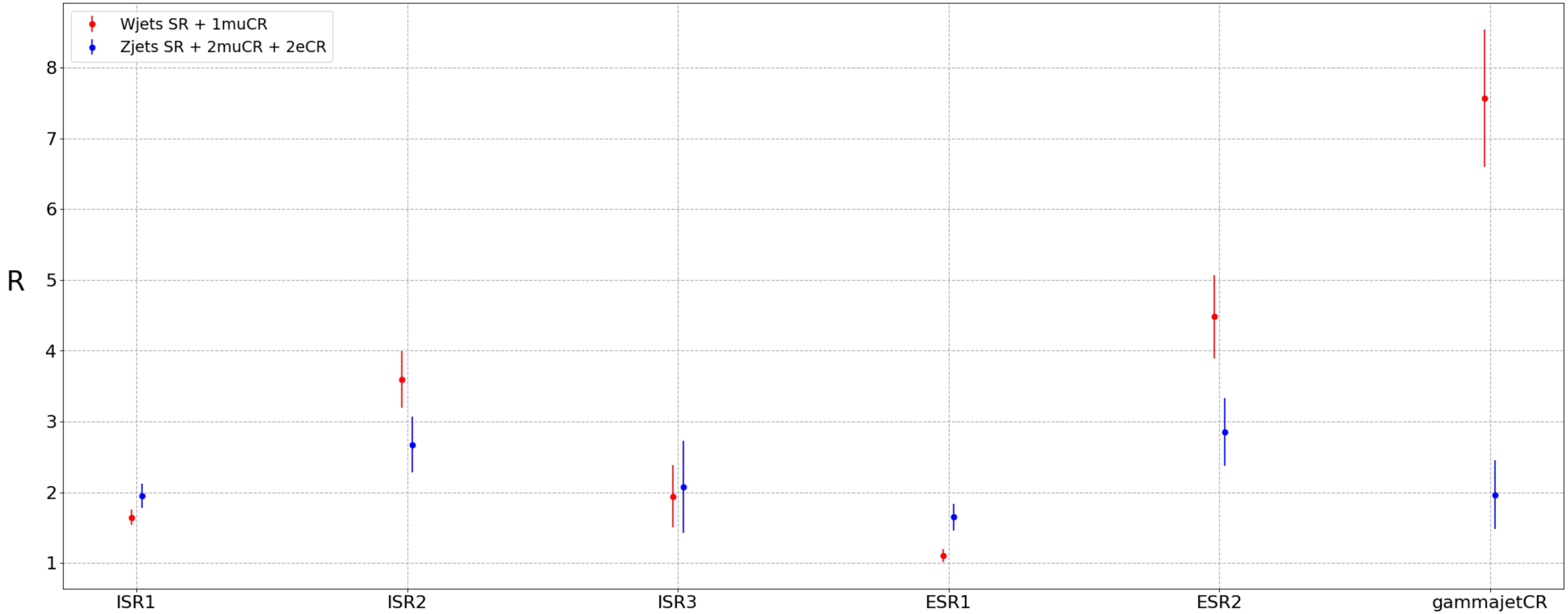
The correlation factors, one for each region of the analysis, have been calculated separately on two samples of Z + jets and W + jets.

Very strange results because of Monte Carlo weights. We managed this weights as follow:

- If |MC weight| > 100, it is rescaled to 1
- If an event has pt < 140 GeV at truth level we excluded it
  - we already select photons with pt > 150 GeV, but some of them may have been reconstructed with higher pt, possibly gaining a very high cross section
- The regions where a certain sample is dominant have been merged:
  - W + jets $\longrightarrow$ SR + 1muCR
  - Z + jets $\longrightarrow$ SR + 2muCR + 2eCR
  - gammajetCR treated separately
- To improve statistics the cut on track isolation (pt_{cone20}/pt) has been released in the control regions
  - we consider not isolated the events that fail the calorimetric isolation OR the track isolation
  - A gap of 0.05 on the track isolation variable has been excluded to prevent signal leakage

Correlation factor

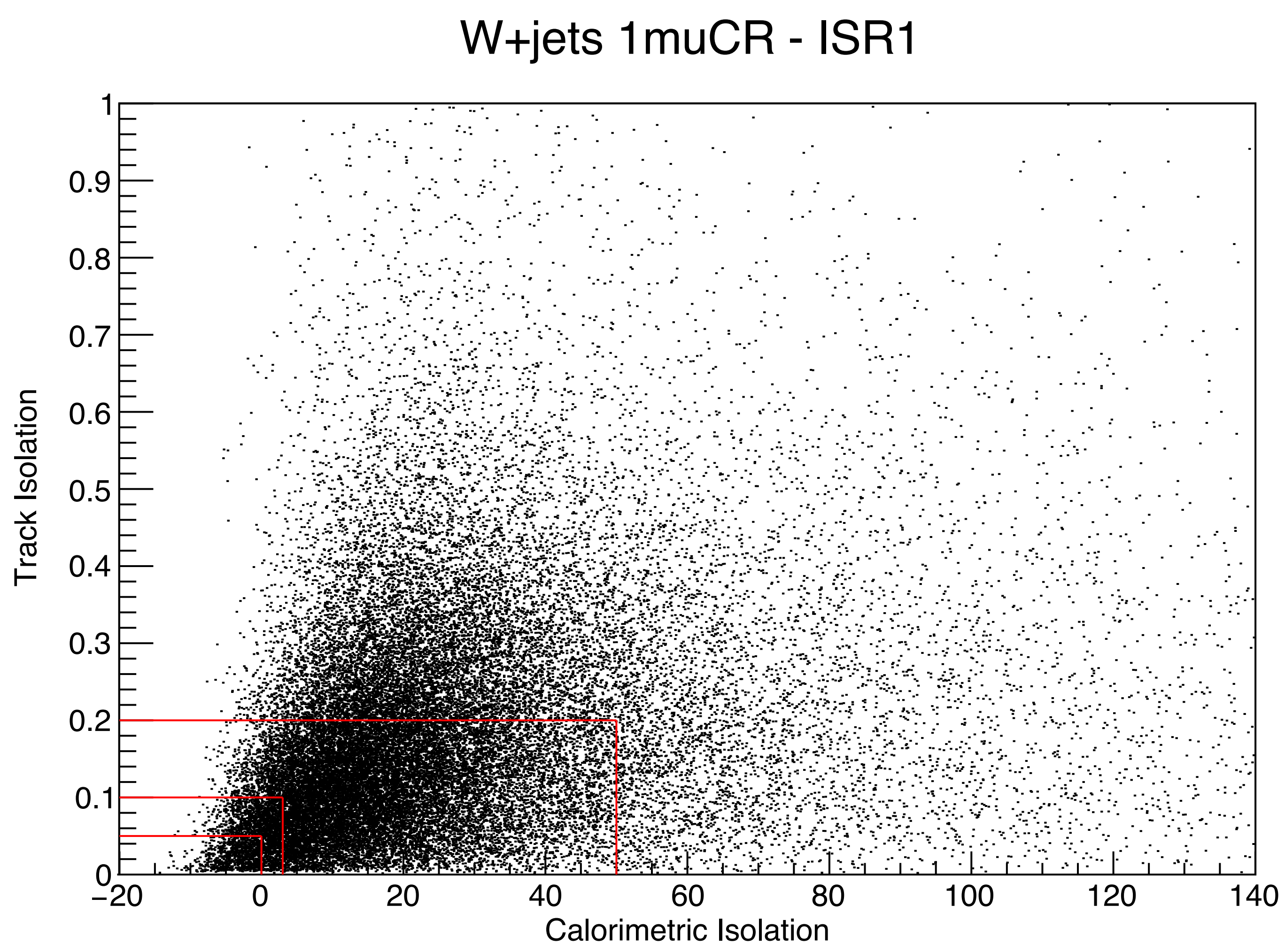
The results in the different regions are the following:



R prime

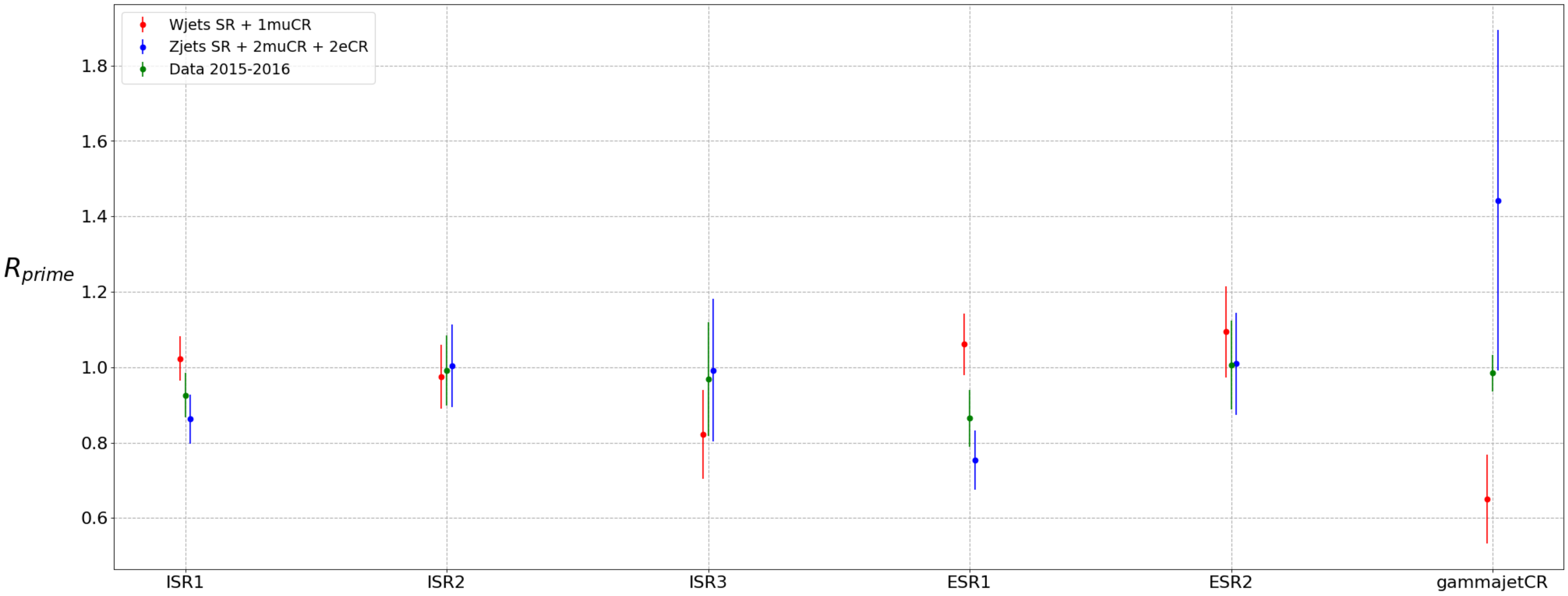
To evaluate the accuracy on the predictions on R we calculated a correlation factor in a completely non-isolated region, both for Monte Carlo samples and for 2015-2016 data. The new number of candidates are as follow:

- N^A_{prime}: 3 GeV < E_{iso} < 50 GeV and $pt_{cone20}/pt < 0.2$ OR E_{iso} < 3 GeV and $0.1 < pt_{cone20}/pt < 0.2$
- N^B_{prime}: 50 GeV < E_{iso} < 100 GeV and $pt_{cone20}/pt < 1.0$ OR E_{iso} < 50 GeV and $0.2 < pt_{cone20}/pt < 1.0$
- Same for M^A_{prime}, M^B_{prime} but in the Tight-4 region



R prime

The results in the different regions are the following:



Signal Leakage

The signal leakage coefficients have been calculated separately on four samples of W + γ, Z + γ, Z (νν) + γ and γ + jets, selecting from this samples only reconstructed photons matched with true photons.

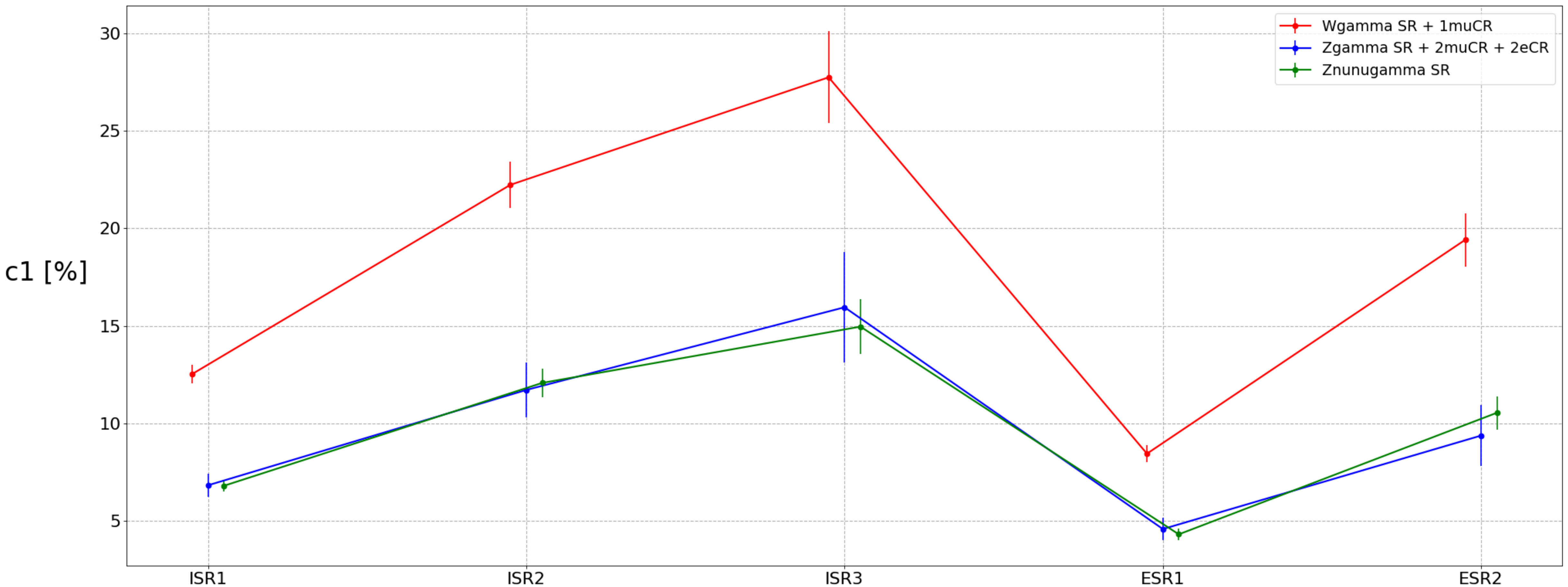
As we did for the correlation factors the track isolation has been released in the control regions and we merged the regions in which a certain process is dominant. The gammajetCR is treated separately.

We expect more leakage with respect to the previous analysis because of three factors:

- The track isolation has been released in the CRs
- The different calorimetric isolation cut definition
- The different Monte Carlo generation, which has a looser isolation cut at truth level

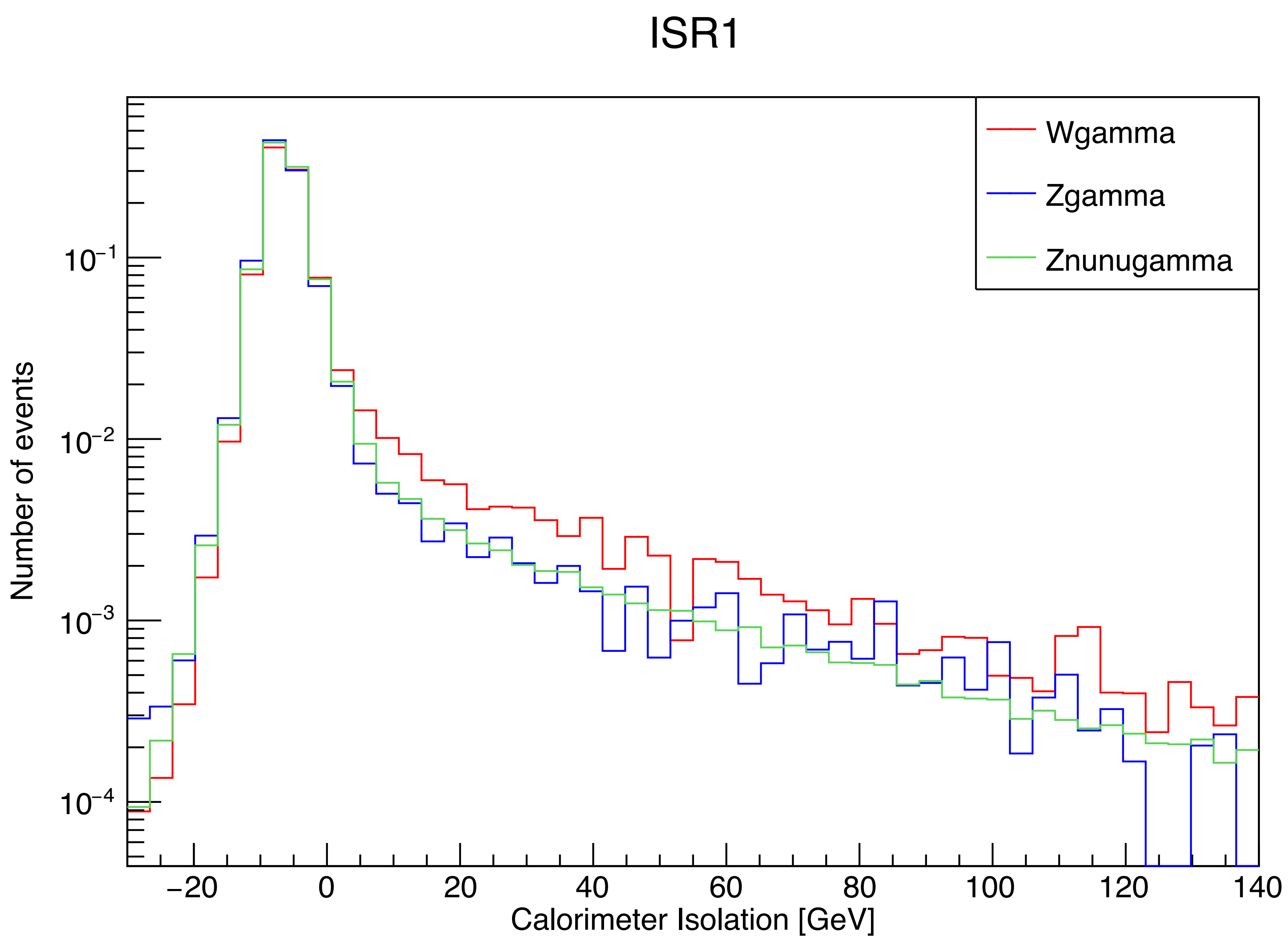
Signal Leakage (c1)

The results in the different regions for the first three samples (W + γ, Z + γ, Z (νν) + γ) are the following:



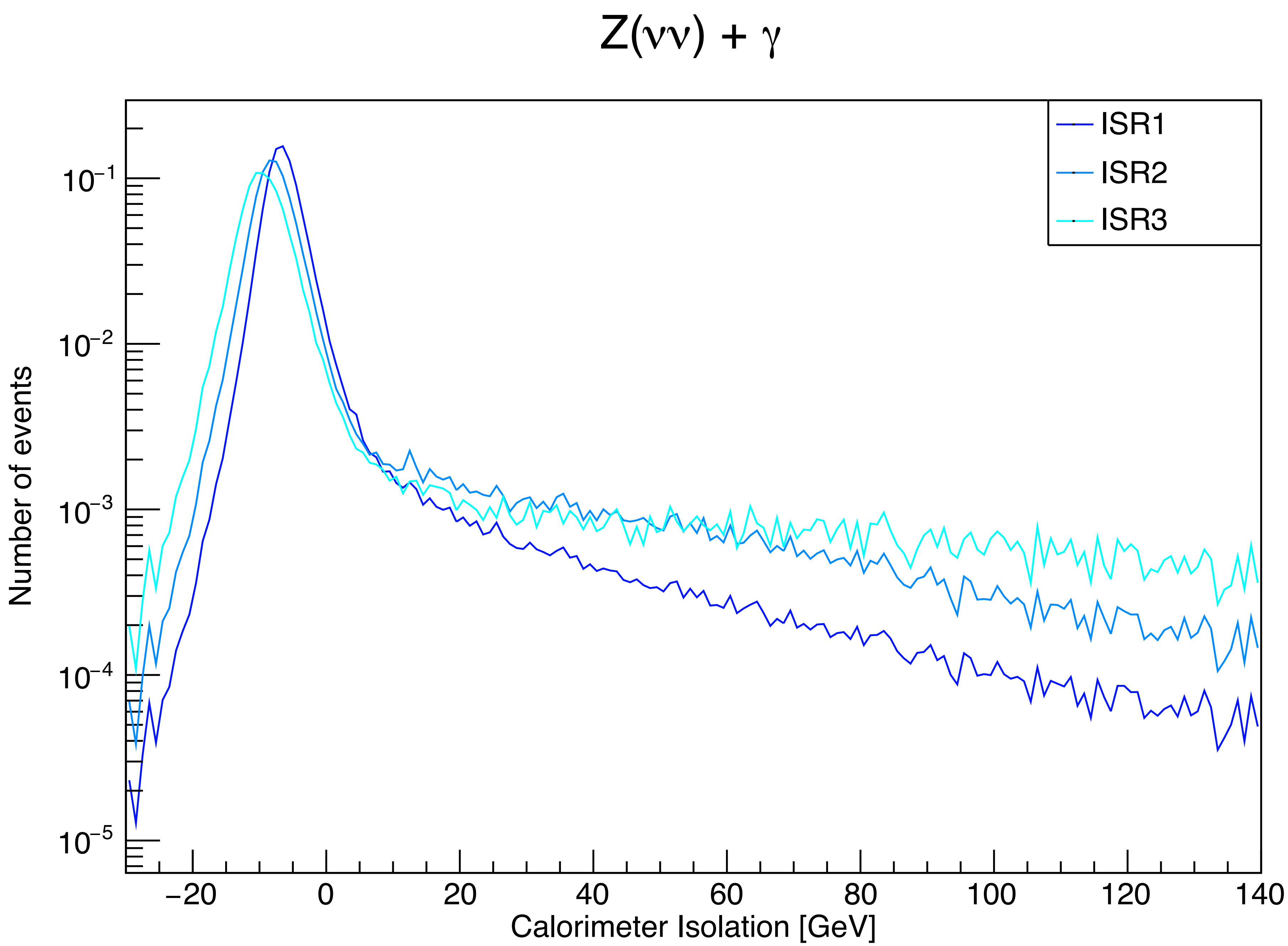
Signal Leakage

- Results for Z + γ and Z (vv) + γ compatible everywhere
- c1 and c3 systematically higher for W + γ
  - Different tails in the calorimeter isolation profile

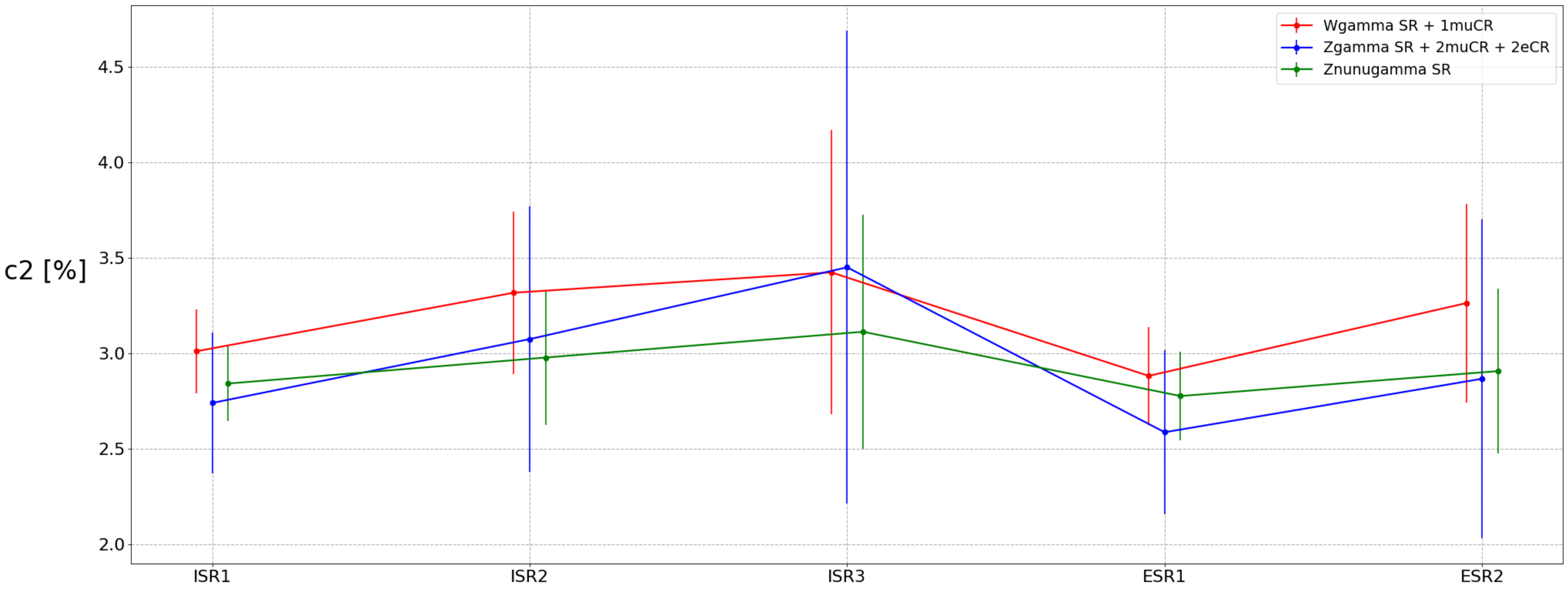


Signal Leakage

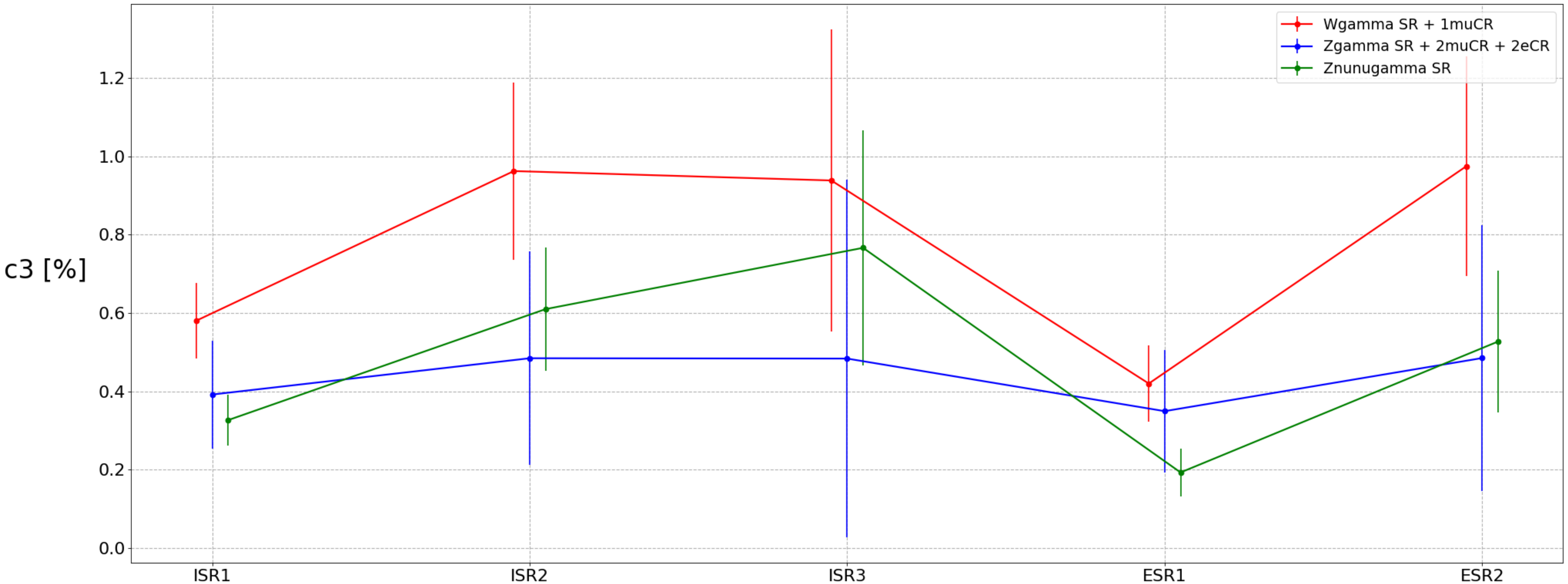
- MET dependence



Signal Leakage (c2)

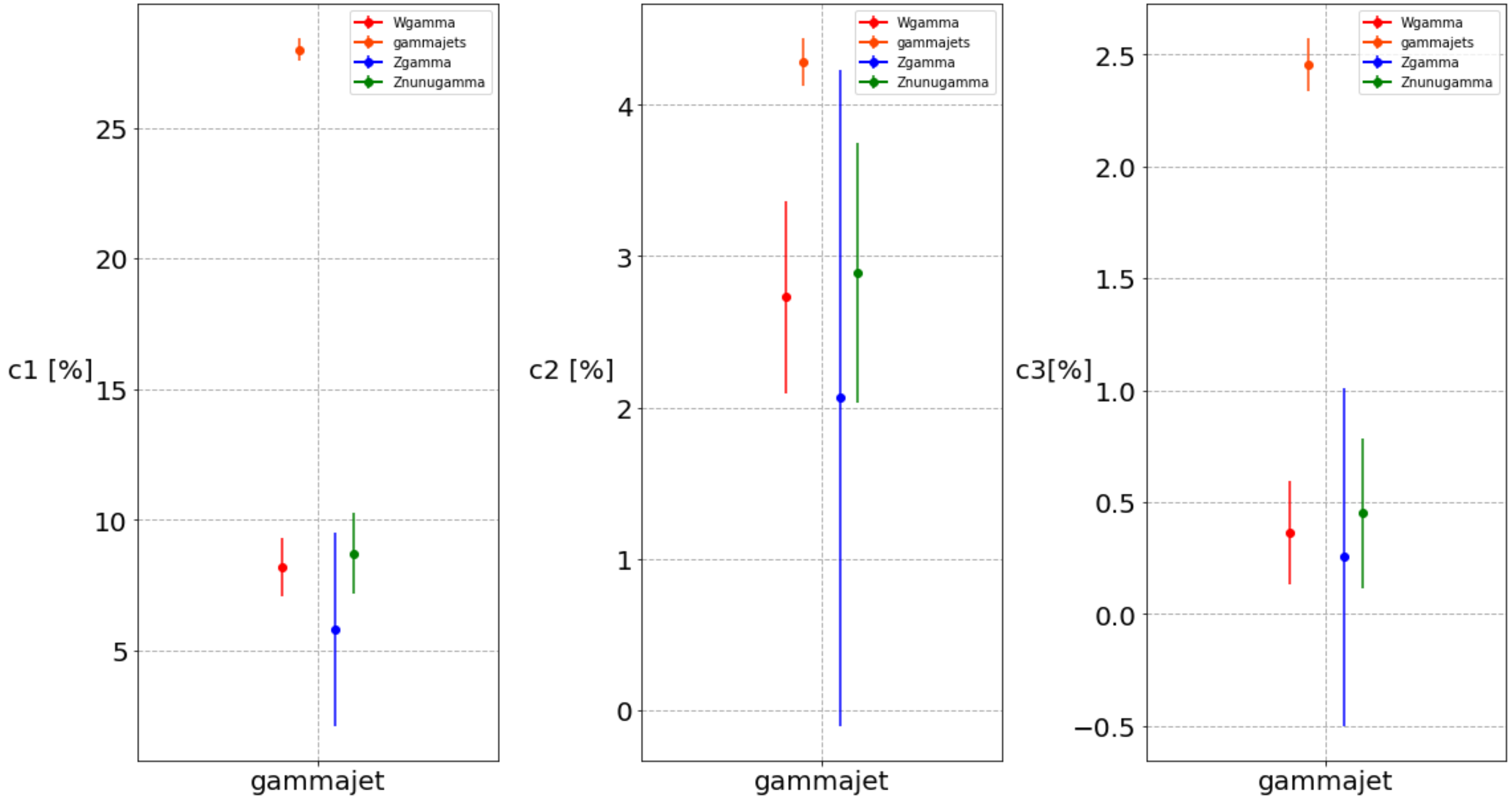


Signal Leakage (c3)



Signal Leakage

The results in the gammajetCR are the following:



The results from the γ + jets sample are much higher the the others, but should be the most precise due to the greater population in this region.

Conclusions

- Correlation factor
  - R prime : good compatibility between MC16a and 2015-2016 data (can trust the MC)
  - R : where compatible between processes (Z/W + γ) we take the weighted mean, otherwise we will take the mean and quote the uncertainty as the maximum variation
  - Very different results in the gammajetCR still to be understood
- Signal leakage
  - Systematic differences of c1 and c3 between W + γ and Z + γ samples not fully understood
  - Increase of the c1 coefficient with the MET threshold
  - We can define an upper limit for the calorimetric isolation to reduce the difference between W and Z and have a smaller systematic uncertainty
  - Conservative : use the central value from the dominant process in each region and take as a systematic the difference with respect to the subdominant process.
  - Large differences in the gammajetCR (not yet understood)
- Almost all ingredients ready for a preliminary evaluation of the jets faking photons contamination in all regions in the analysis (5 x 4 + 1 = 21 regions)