HH→γγbb

γγ background decomposition studies and
Jet pT reweighting with BTagging Efficiencies for the background templates

Adele D'Onofrio, Shuo Han, Yanping Huang, Shan Jin

22nd August 2019

NANJING UNIVERSITY 1902

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

Motivations

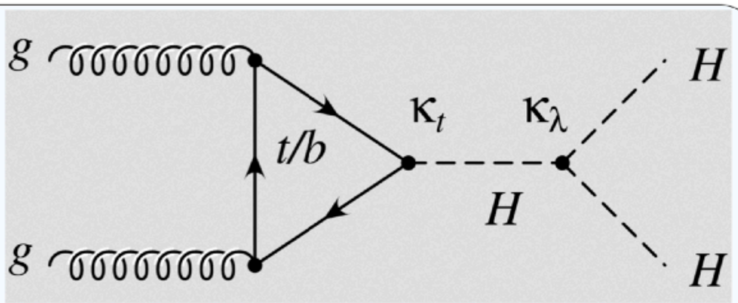
$$\mathcal{L}_H = \frac{1}{2}(\partial^\mu H)(\partial_\mu H) - \boxed{\lambda v^2} H^2 - \underbrace{\lambda v H^3 - \frac{1}{4}\lambda H^4}$$

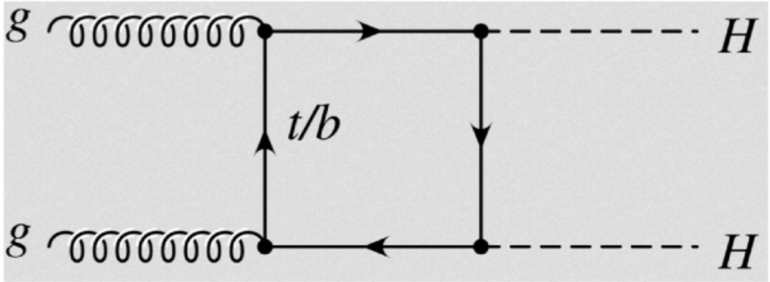
Mass term: ~125 GeV
(measured at the LHC)

Higgs vacuum expectation value: υ~ 246 GeV
(known from Fermi constant)

There are vertices with 3 and 4 Higgs bosons

Therefore in the SM we expect multi-Higgs production at tree-level with already known couplings





However, still very small cross section:
σ(HH) = 34 fb (c.f. σ(gg→H) ~50 pb)
due to negative interference with box diagram

HH→ bbγγ: JHEP 11 (2018) 040

- High BR from H->bb
- Clean signature from γγ

	bb	WW	ττ	ZZ	γγ
bb	33%				
WW	25%	4.6%			
ττ	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
γγ	0.26%	0.10%	0.029%	0.013%	0.0005%

Direction of largest BR

Outline

- γγ background decomposition studies
- Reweighting technique to reduce statistical fluctuation in the di-photon background MC
- Samples: skimmed mc16a/d/e di-photon Sherpa MC MxAOD
- Full Run 2 simulation studies

Object Reconstruction and Event Selection

Object Reconstruction

Photon Objects

- Loose identification criteria
- $p_T > 25$ GeV
- $p_T/m_{\gamma\gamma} > 0.35$ and 0.25, respectively for leading and subleading γ
- Overlap removal (electrons and jets $\Delta R < 0.4$)

Jet Objects

- Jets are reconstructed using the anti-k_t algorithm with radius parameter 0.4
- $|y| < 4.4$
- $p_T > 25$ GeV
- Jets with $|\eta| < 2.5$ and $p_T < 120$ GeV originating from pile-up collisions are suppressed by JVT

B-jets

- MV2c10, AntiKt4EMTopoJets, FixedCutBEff_85, FixedCutBEff_70

2H γγbb analysis

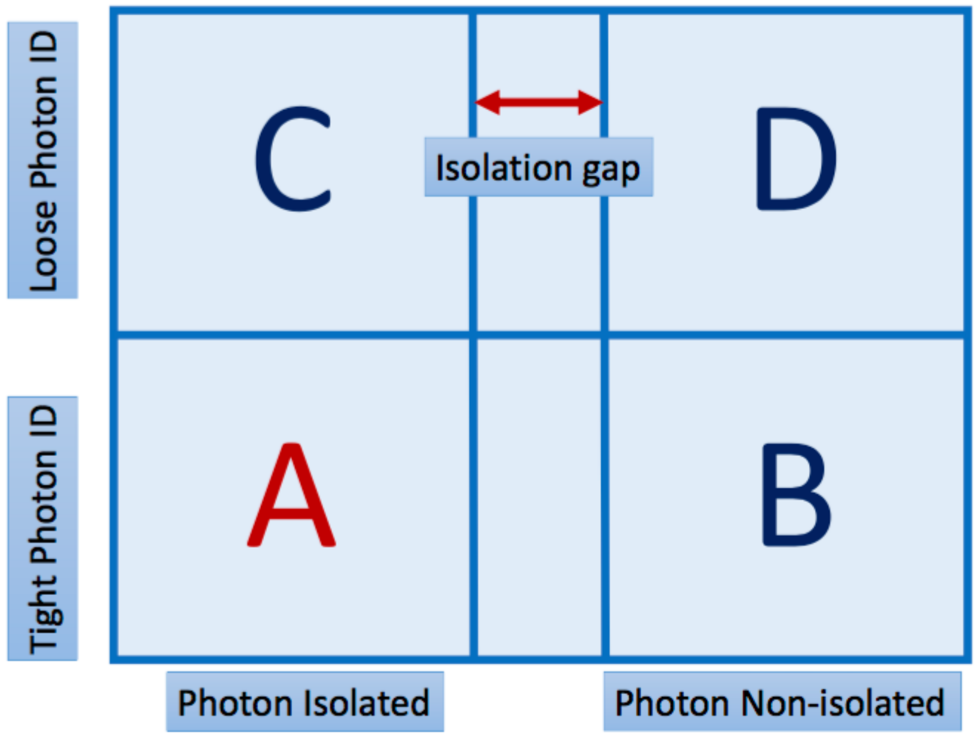
Event Selection

- Trigger Selection:
  - diphoton trigger with p_T thresholds of 35 GeV and 25 GeV
- Jet cleaning
- N_j >2
- N_j_central >= 2 ($|\eta| < 2.5$)
- N_lep = 0
- N_btag70<3
- 2 tight and isolated photons
- $m_{\gamma\gamma} \in (105, 160)$ GeV

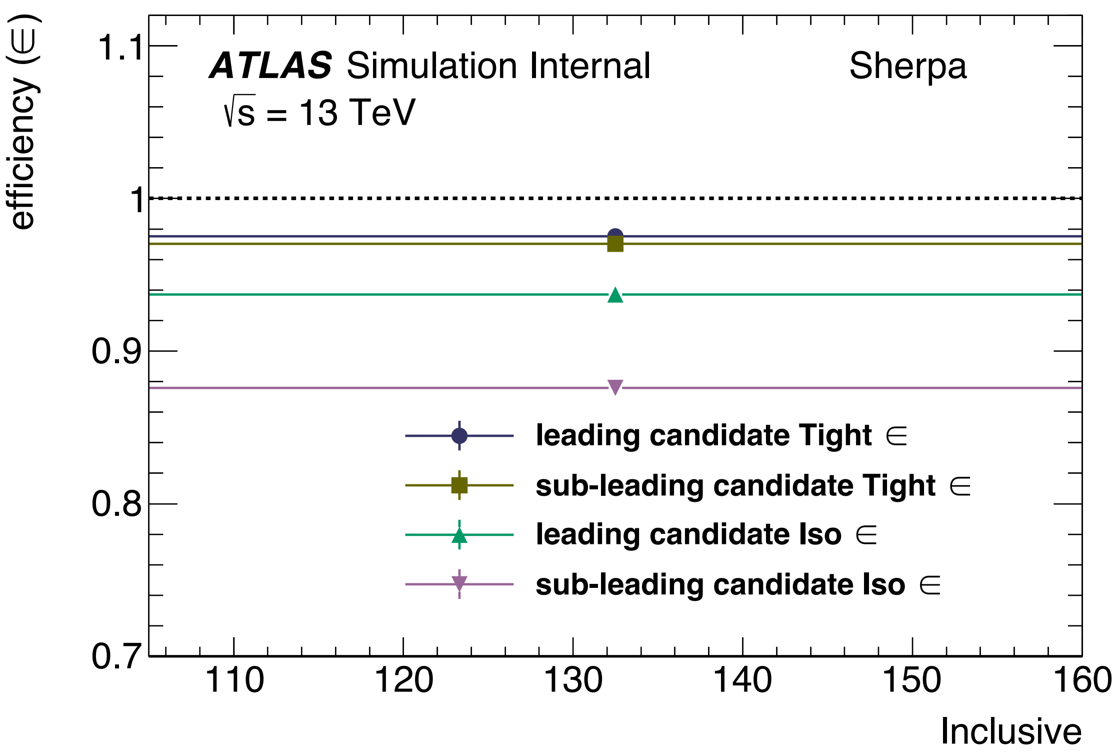
Sherpa γγ sample

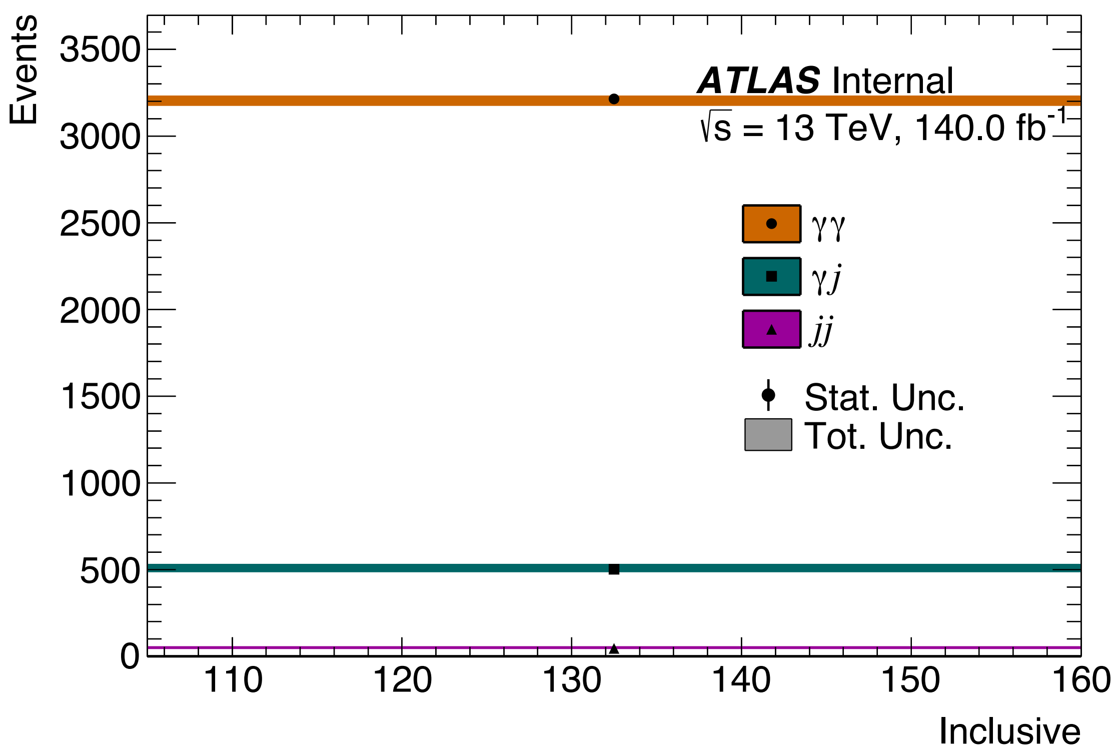
Sideband Method

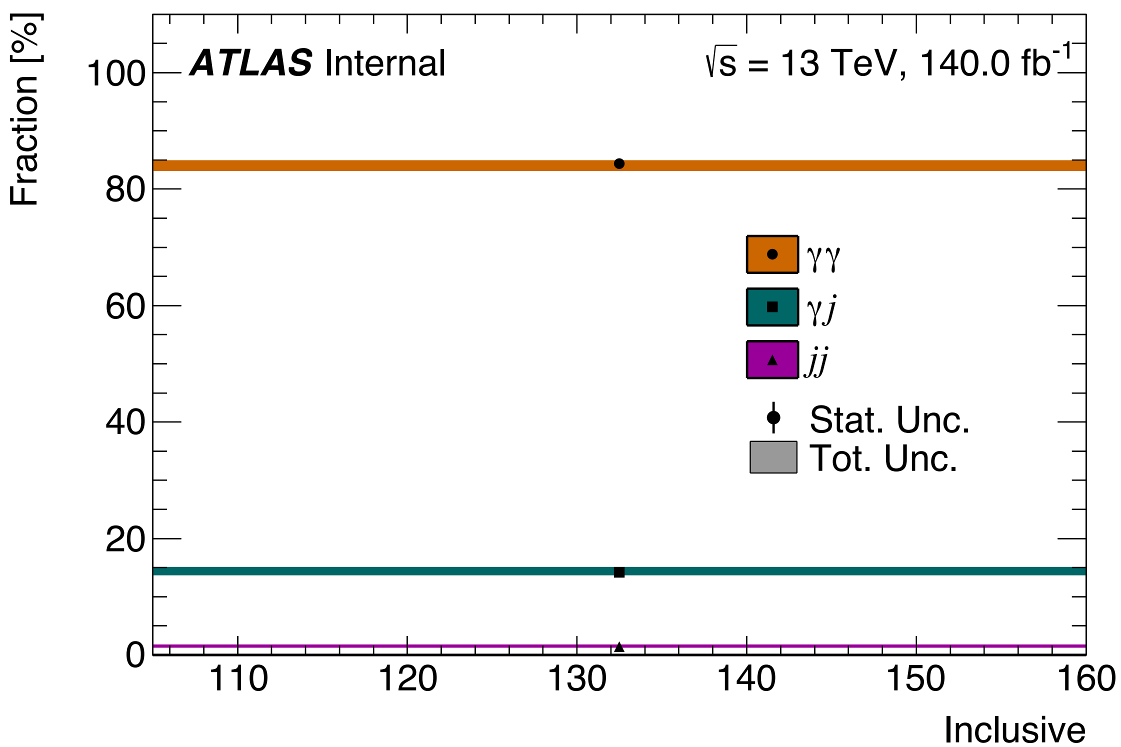
- H→γγ signal region: two Tight (T) identified and isolated (I) photon candidates
- Dominant SM background processes: γγ, γ-jet, and jet-jet
- Data-driven background decomposition using the 2x2D sideband method:
  - 16 regions (15 background control region and 1 signal region)
  - each region categorised based on photon ID and isolation
  - required inputs from simulation: photon efficiencies
  - solve the 16 equations corresponding to the TT region to extrapolate rates of the different bkg event processes in the background control regions into the signal region



Inclusive region

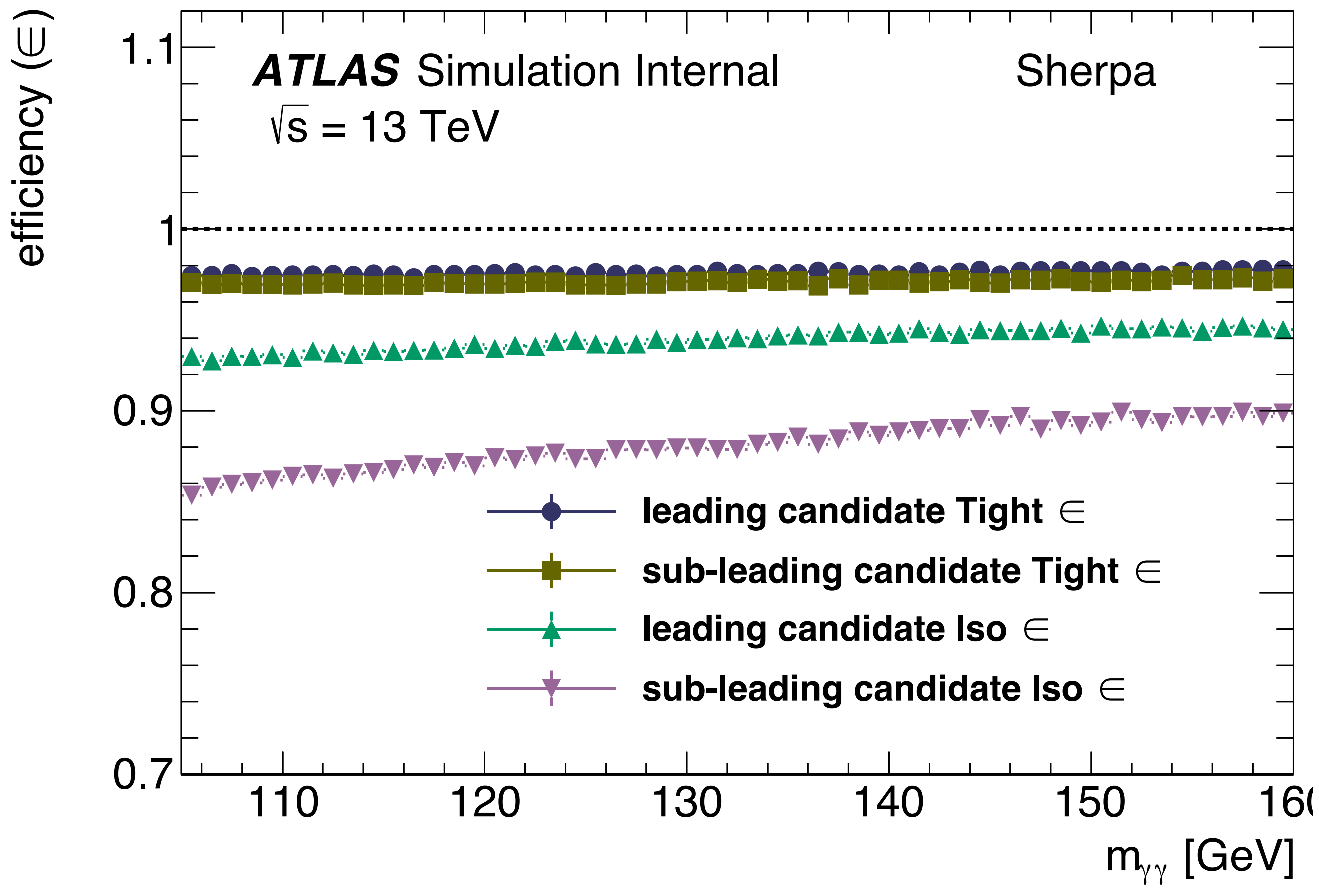


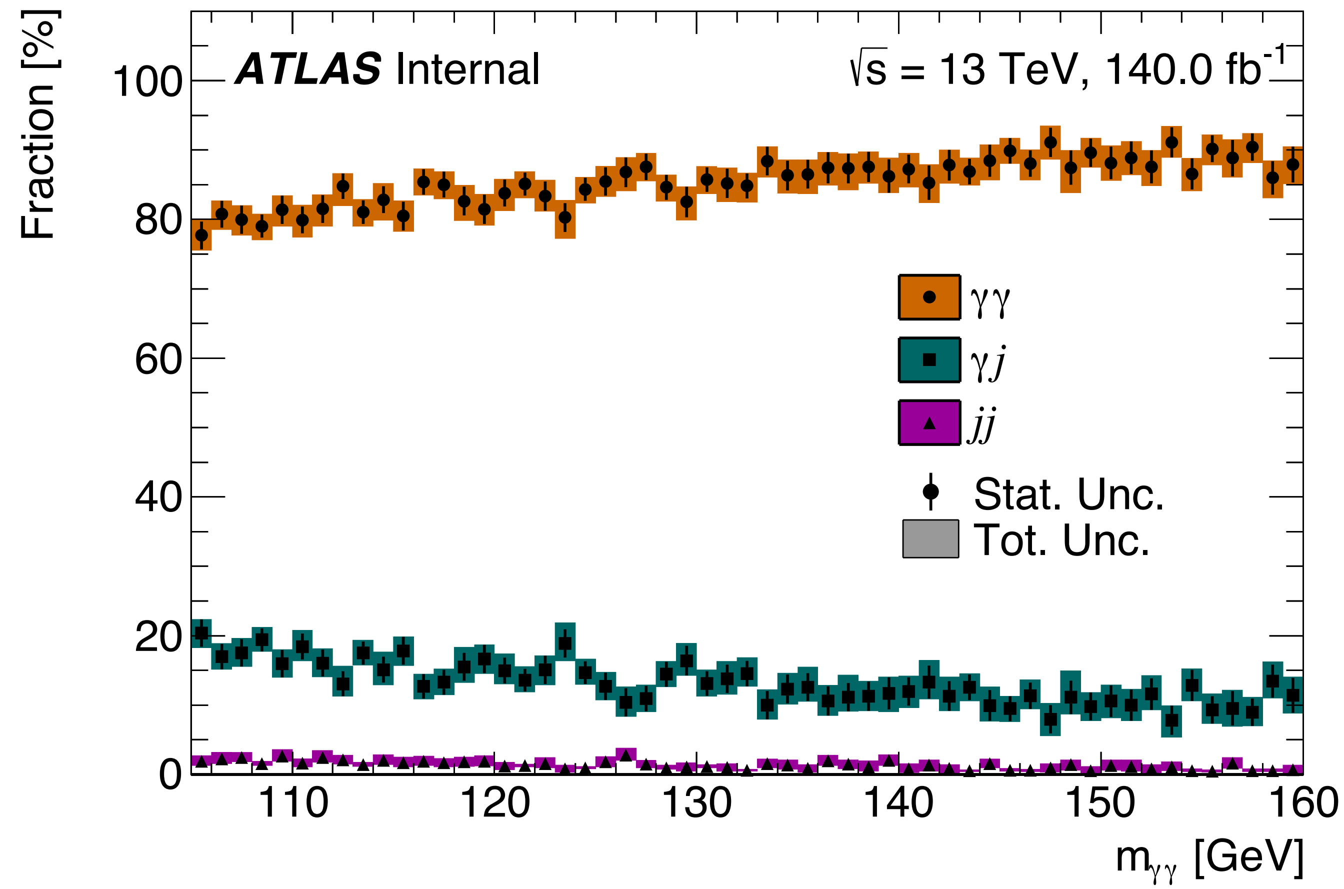




Region	Fraction (%)	Error (%)
$\gamma\gamma$	84	$\pm$ 1
$\gamma\, jet$	15	$\pm$ 0.5
$jet\; jet$	1	$\pm$ 0.1

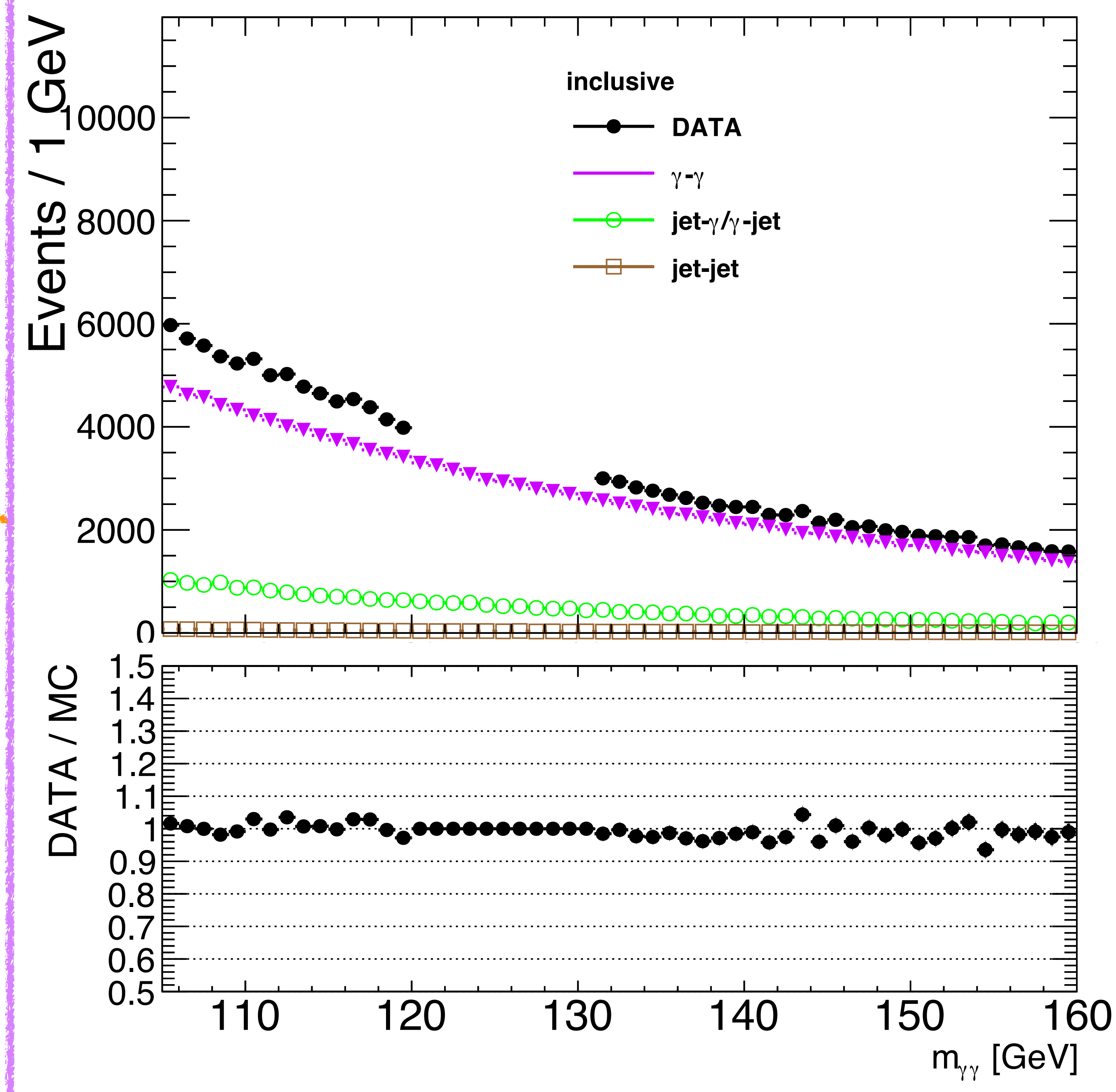
γγ mass spectrum

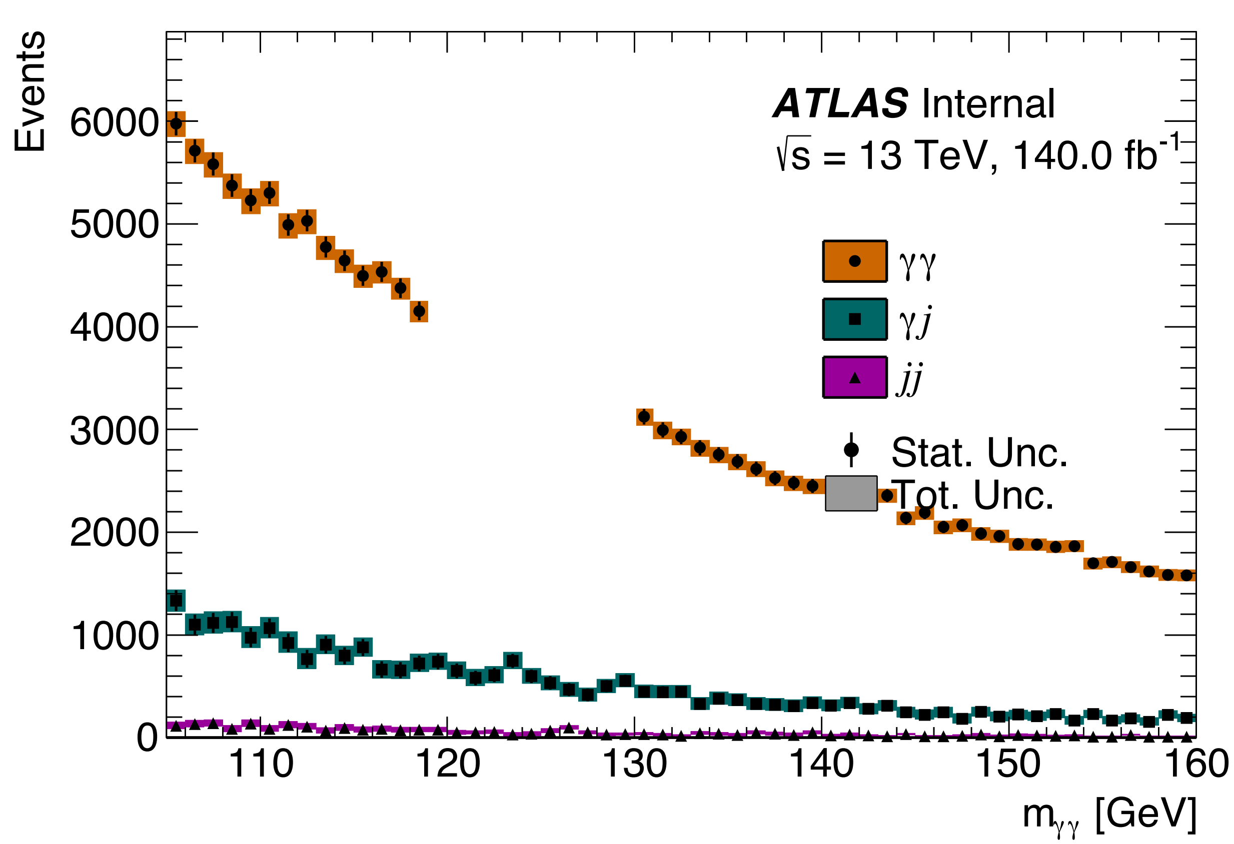




γγ mass spectrum

- Data driven technique to estimate γγ,γj,jj contribution
- Shape given by MC distribution





Signal Region Categories

Inclusive Region

Common Selection	N_{cen} jets < 6, $N_{lep} = 0$, $N_{70\% b-tags} < 3$

Cut based selection for the non resonant analysis

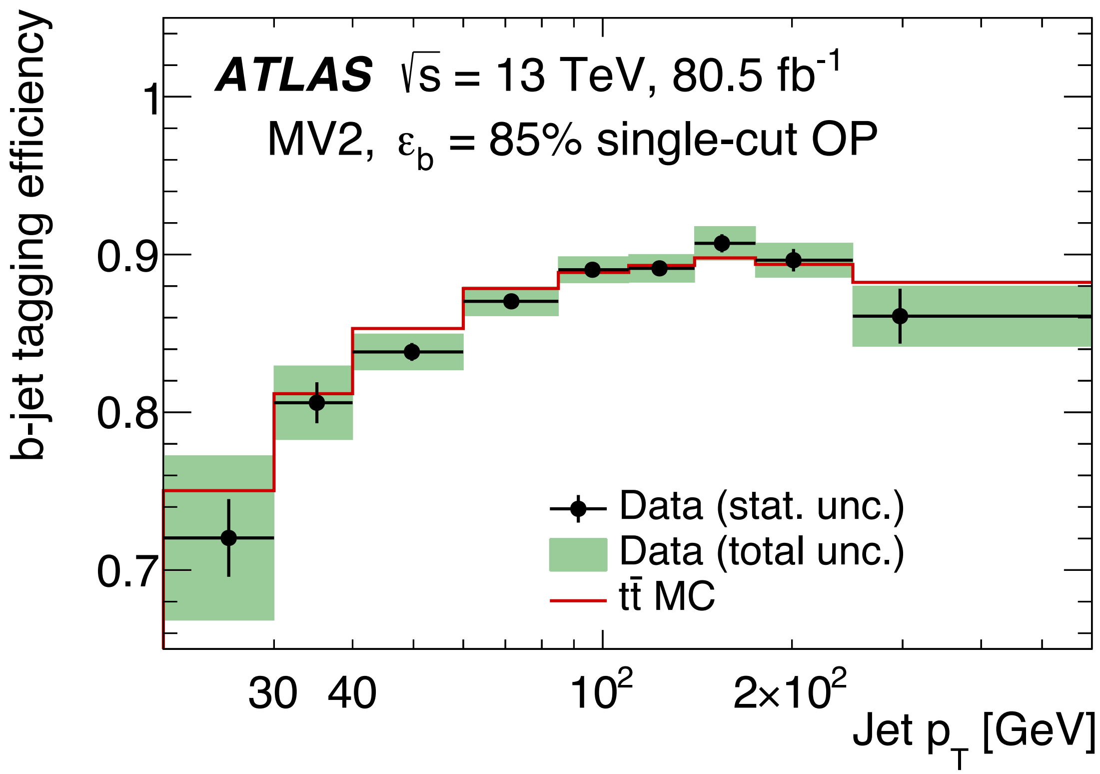
	Common Selection	N_{cen} jets < 6, $N_{lep} = 0$, $N_{85\% b-tags} \geq 2$, $N_{70\% b-tags} < 3$, $m_{bb} \in [90, 140]$ GeV
	Category	Further Selections
Cat-4	b-tag tight, $M_x^* \geq 350$ GeV	$M_x^* \geq 350$ GeV, $N_{70\% b-tags} = 2$, $\Delta R(\gamma, \gamma) < 2.0$, $\Delta R(b, b) < 2.0$, $\Delta R(bb, \gamma, \gamma) < 3.4$
Cat-3	b-tag loose, $M_x^* \geq 350$ GeV	$M_x^* \geq 350$ GeV, $\Delta R(\gamma, \gamma) < 2.0$, $\Delta R(b, b) < 2.0$, $\Delta R(bb, \gamma, \gamma) < 3.4$
Cat-2	b-tag tight, $M_x^* < 350$ GeV	$M_x^* < 350$ GeV, $N_{70\% b-tags} = 2$
Cat-1	b-tag loose, $M_x^* < 350$ GeV	$M_x^* < 350$ GeV

$$M_x^* = M_{b\bar{b}\gamma\gamma} - M_{b\bar{b}} - M_{\gamma\gamma} + 250.0\,\text{GeV}$$

Reference for this categorisation can be found here: *link*

Signal Region: Combination of MV2@85 && MV2@70

Comparing Tagging Efficiencies

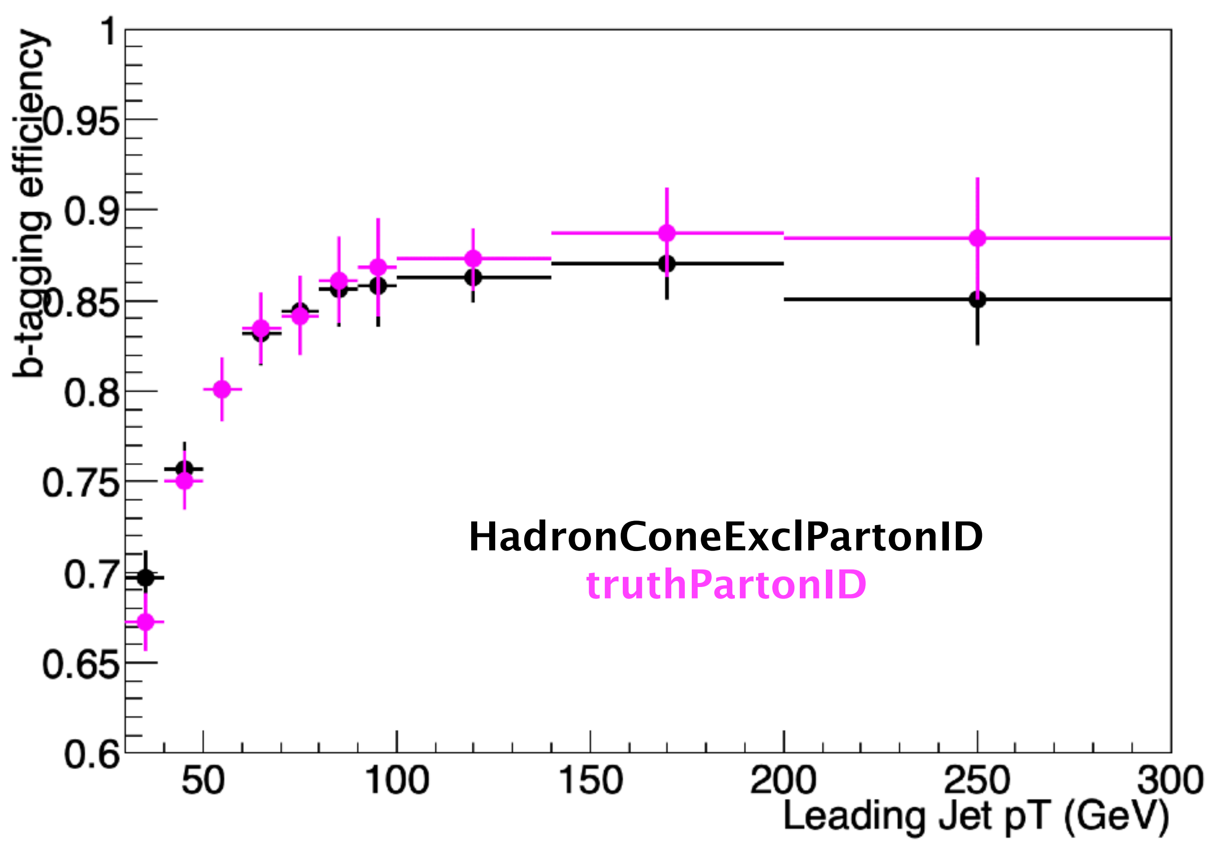


MV2c10@85% WP

Given the above Object reconstruction and Event Selection

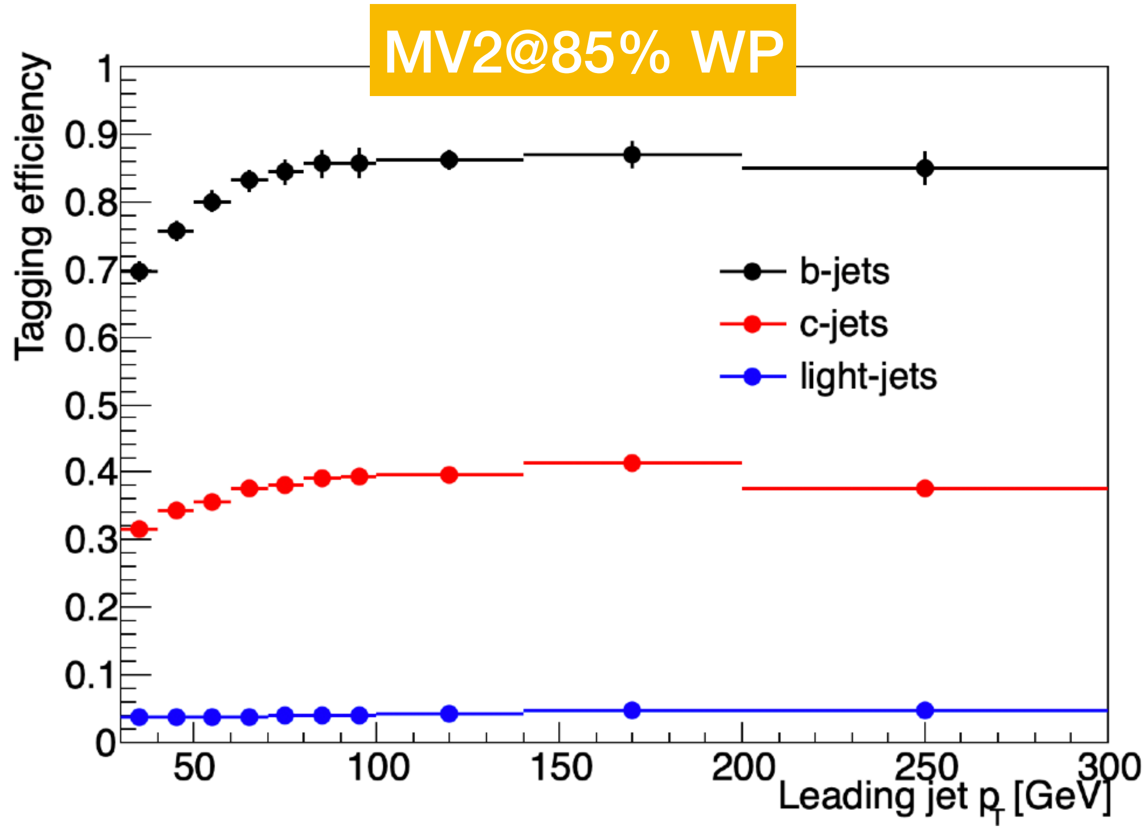
Efficiency definition: Numerator/Denominator

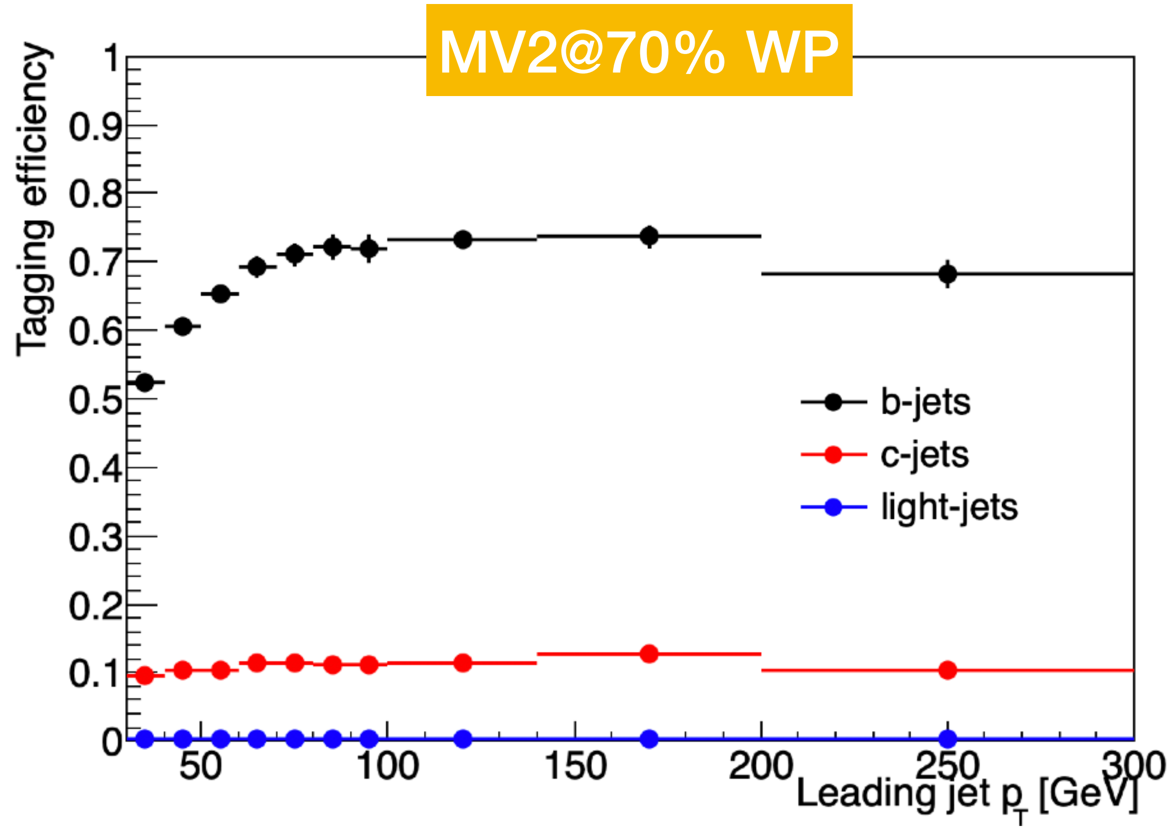
Denominator is: the number of jet that pass the b-truth matching and in the central region
Numerator is: the number of jets that pass MV2 selection on top of the denominator

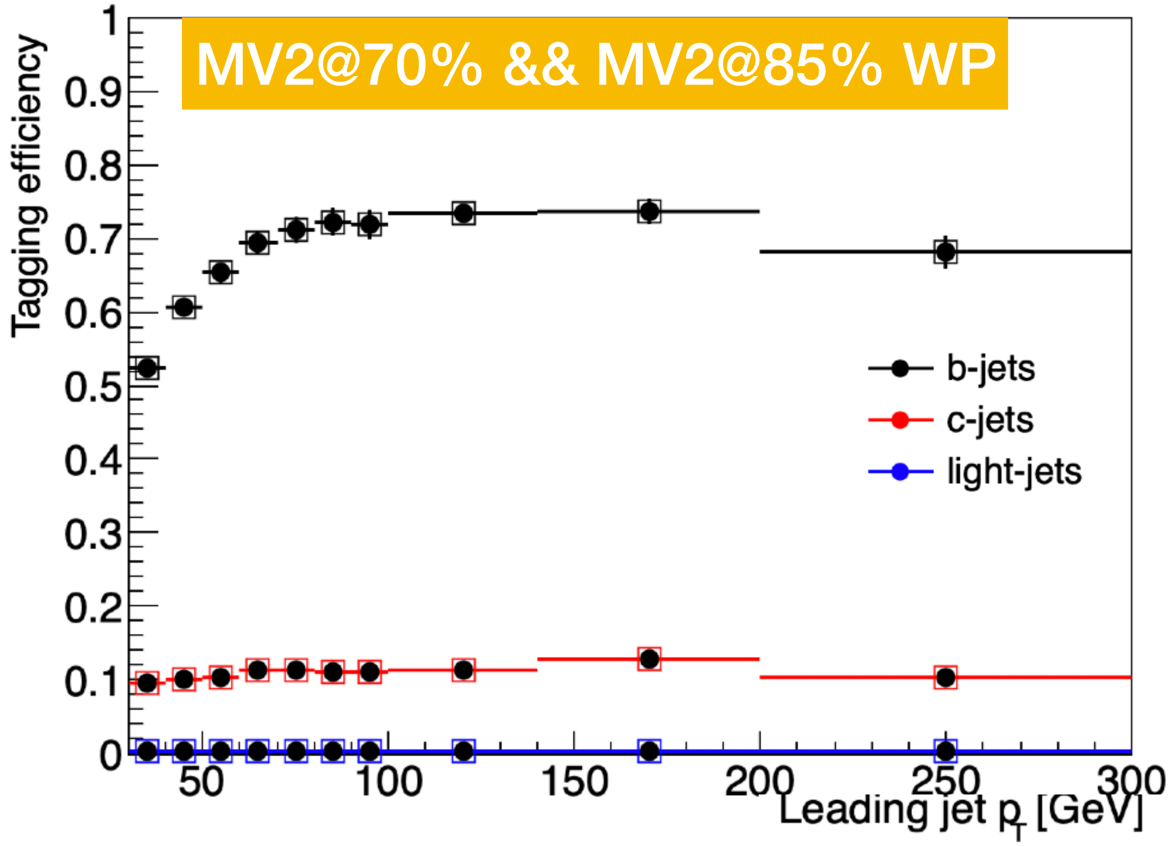


- **Updated truth parton definition to the into account gluon splitting in bottom quark pairs**
- https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTagCalib2017
- **The b-tagging efficiency at low pT has improved of a factor ~5%**
- **There is still a ~13% difference wrt the performance group**

Local tagging efficiency results



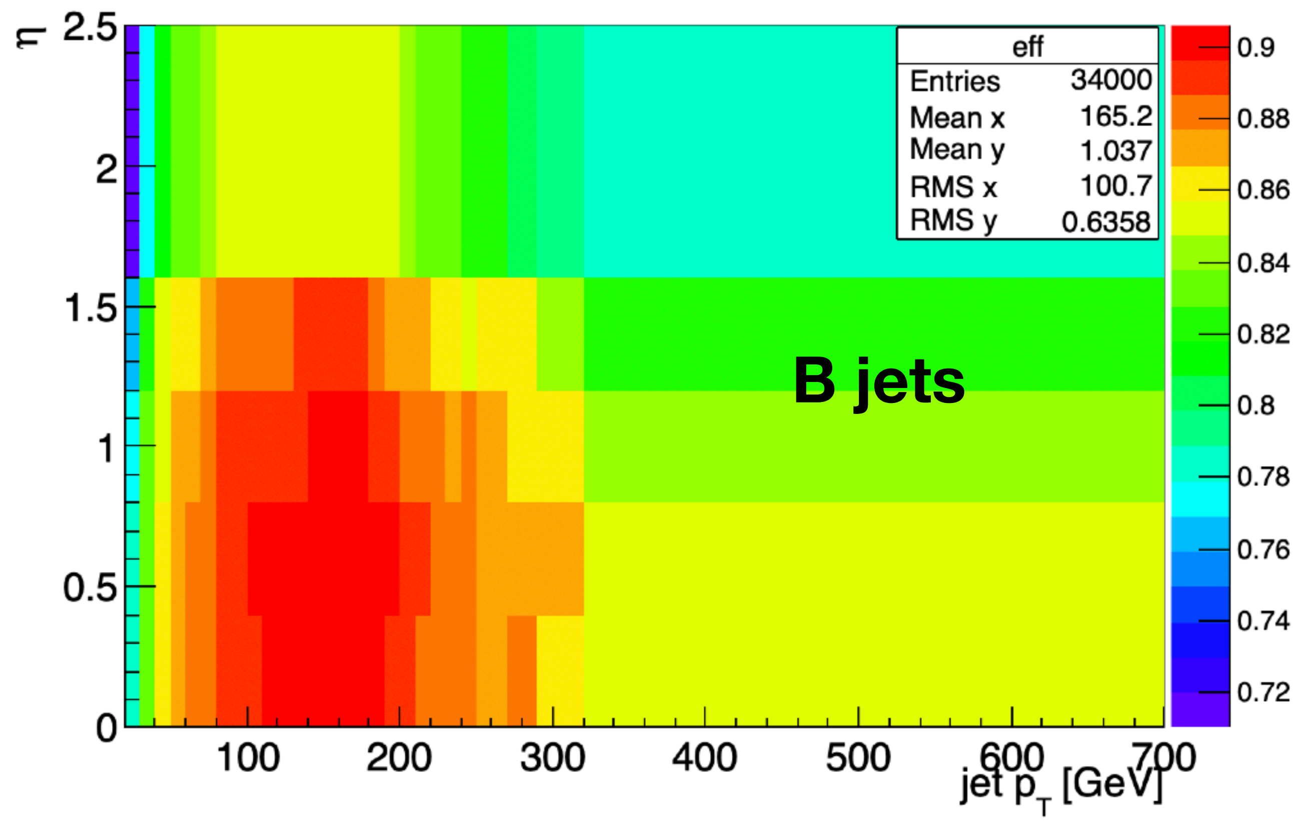


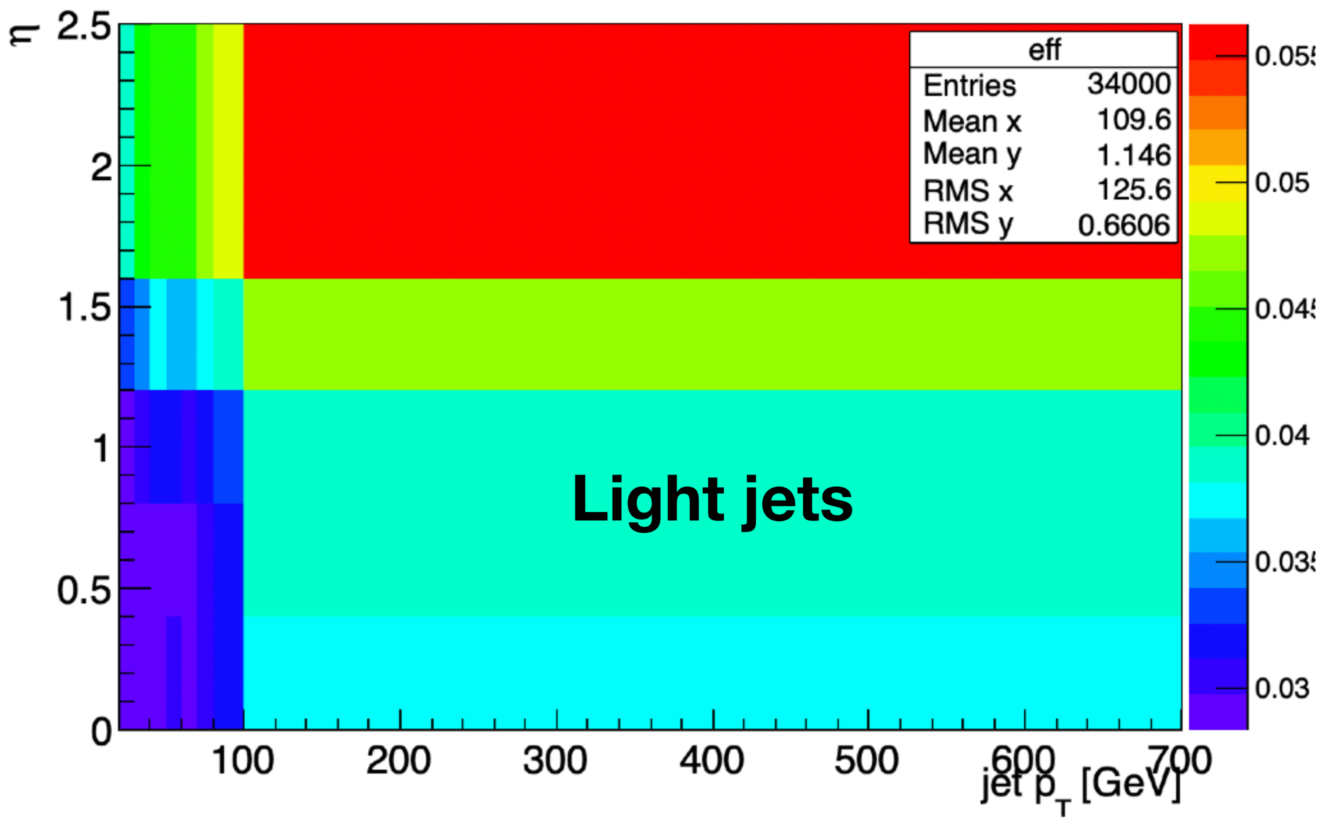


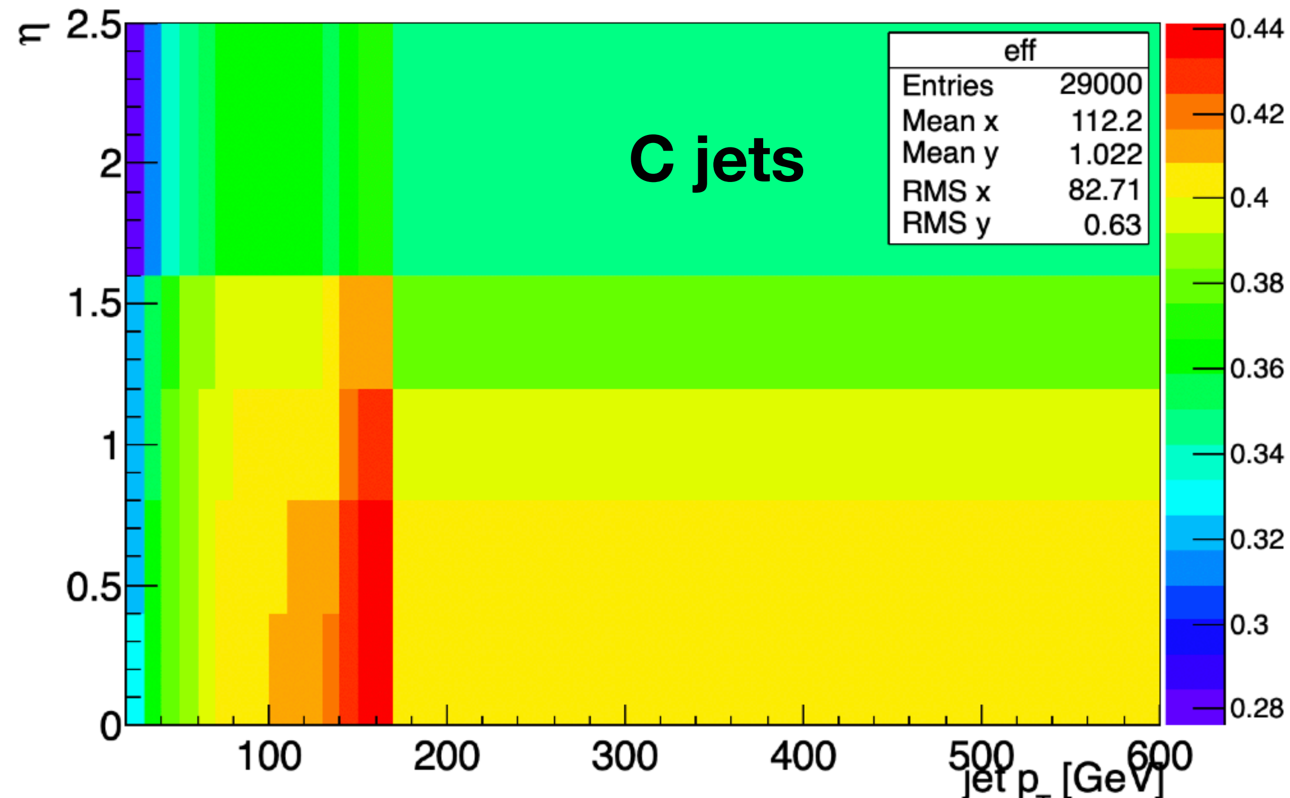
- Our signal region is a combination of the Mv2c10@70% and MV2c10@85% working point
- Logical AND of these conditions
- The resulting efficiency is equivalent to the Mv2c10@70% one

Tagging Efficiencies, CDI File

- CDI File from July 2019 /cvmfs/atlas.cern.ch/repo/sw/database/GroupData/xAODBTaggingEfficiency/13TeV/2017-21-13TeV-MC16-CDI-2019-07-30_v1.root
- Pythia 8 sample
- ttbar events (object reconstruction and event selection in the back-up)
- MV2c10, AntiKt4EMTopoJets, FixedCutBEff_85



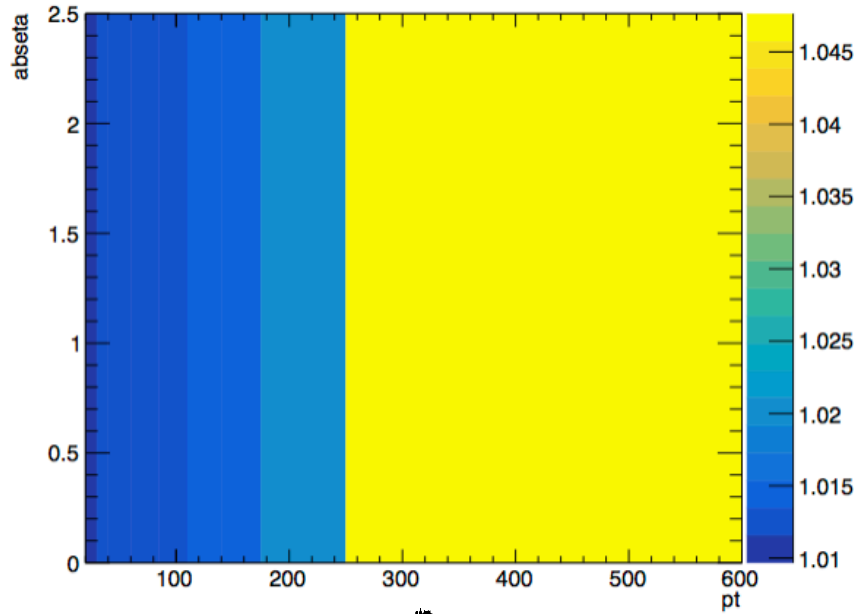




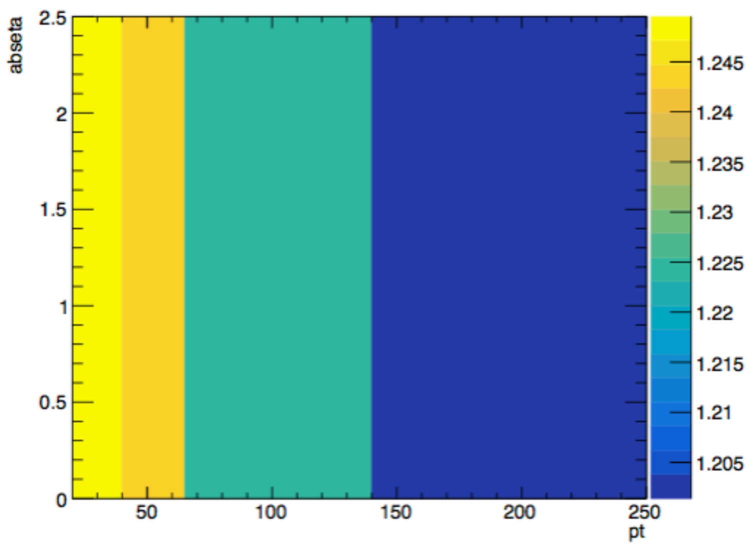
Pythia8/Sherpa Scale Factors

July 2019: MC/MC Pythia8 / Sherpa MV2c10@85% EMTopoJets

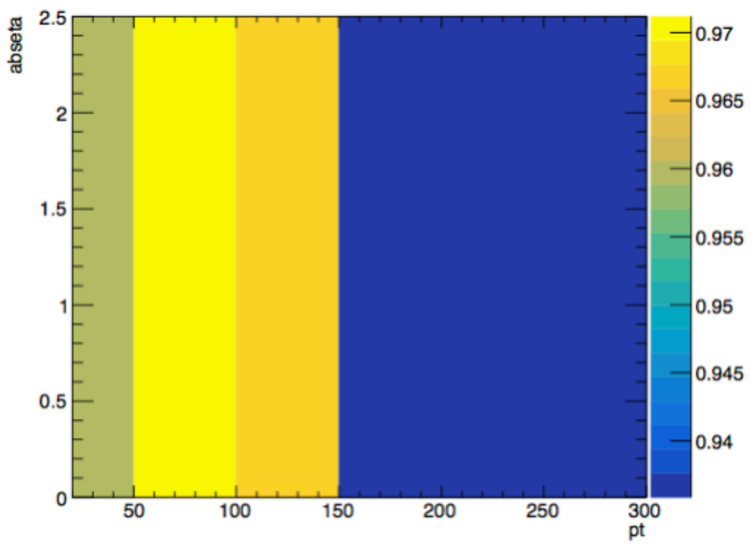
b-jets



c-jets



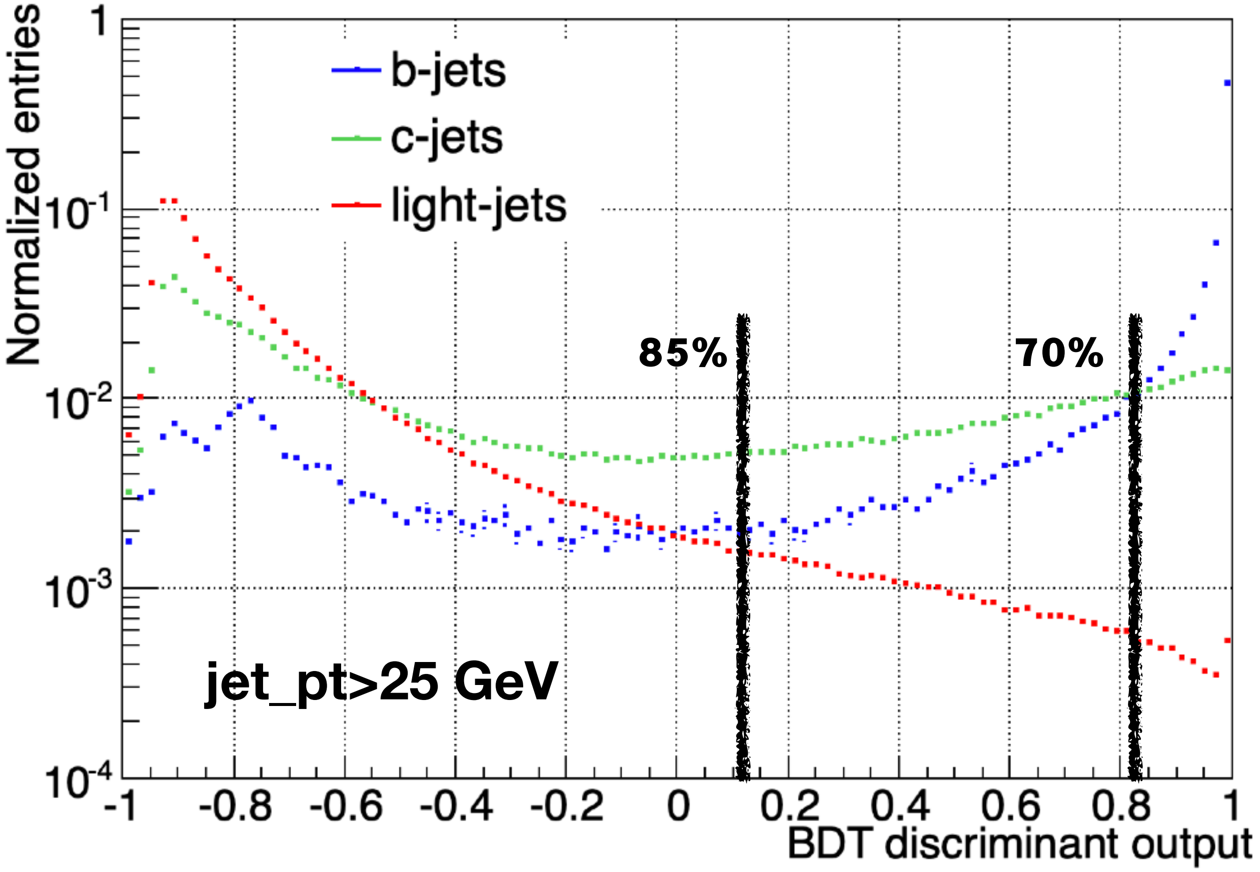
light-jets



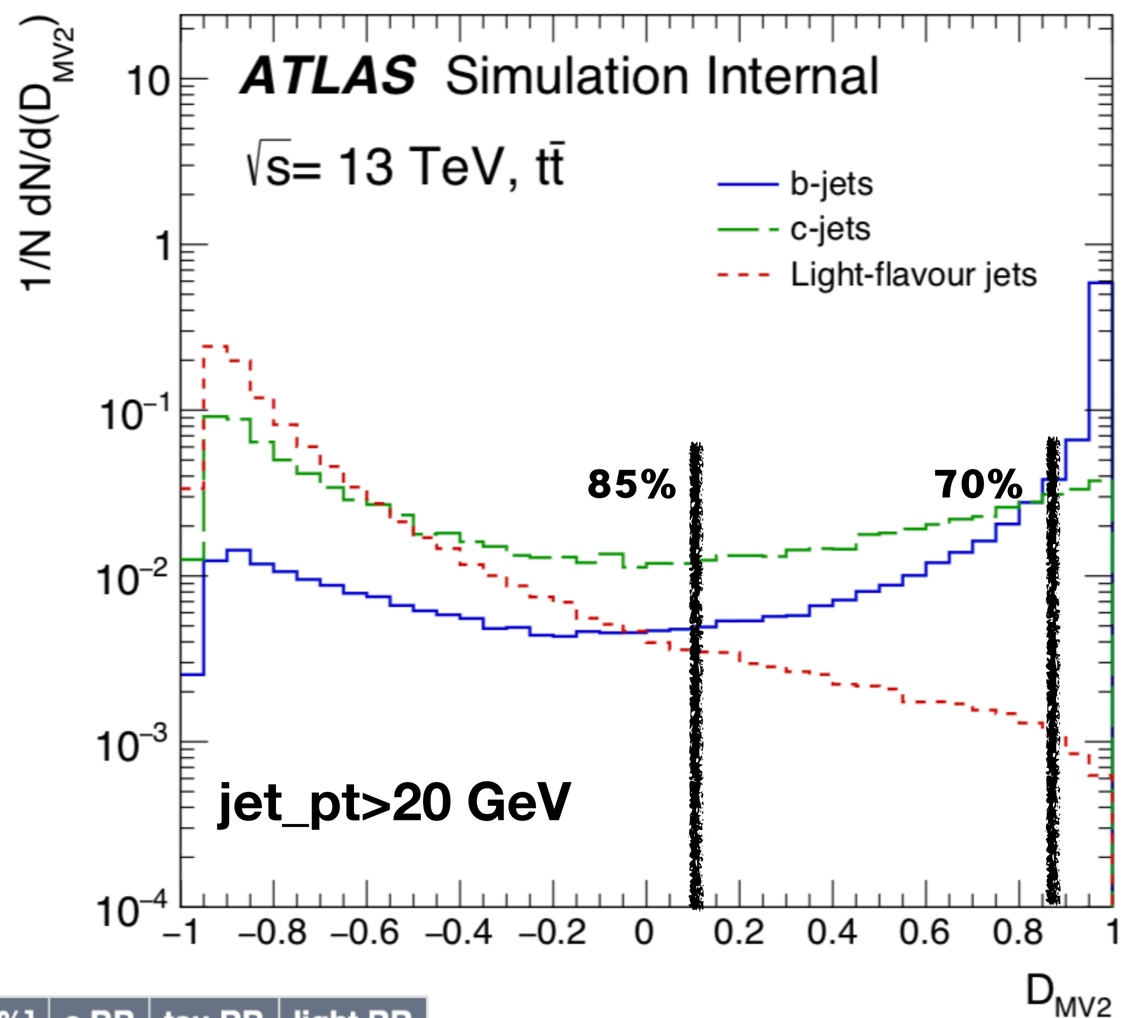
These SFs can justify up to a ~2.5% difference
(Thanks Marko and the flavtag-btagging algorithms mailing list)

Crosscheck on BDT discriminant

Our result



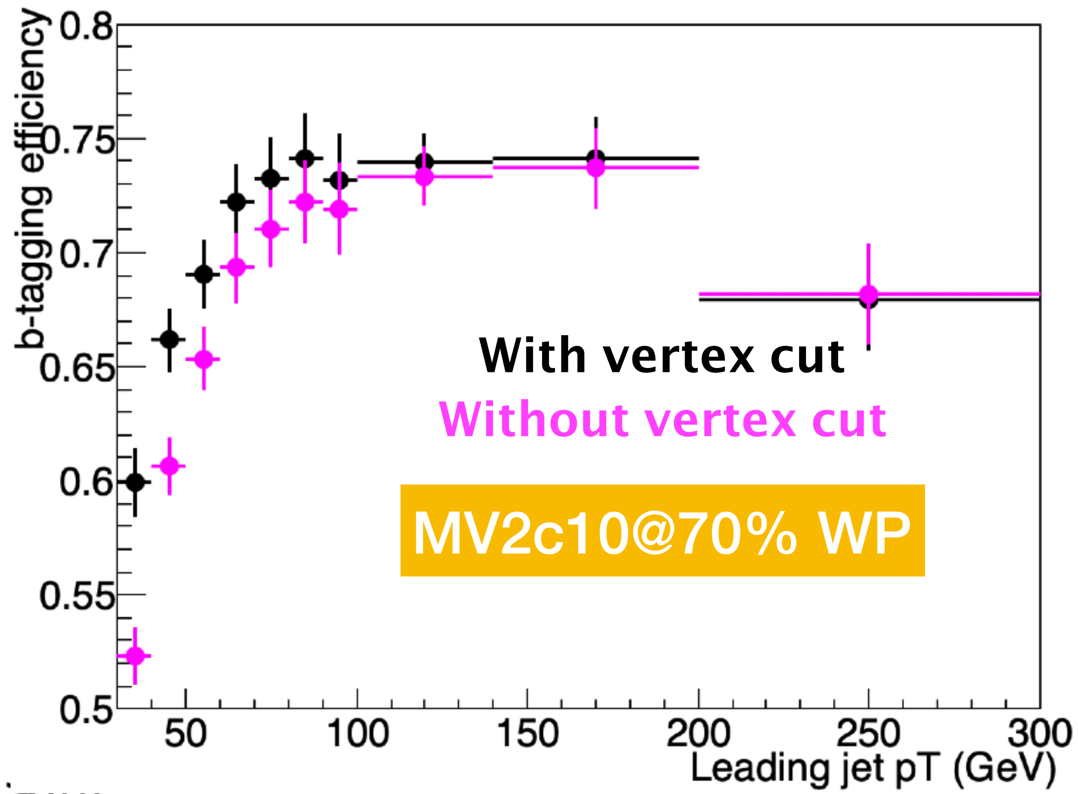
Performance group

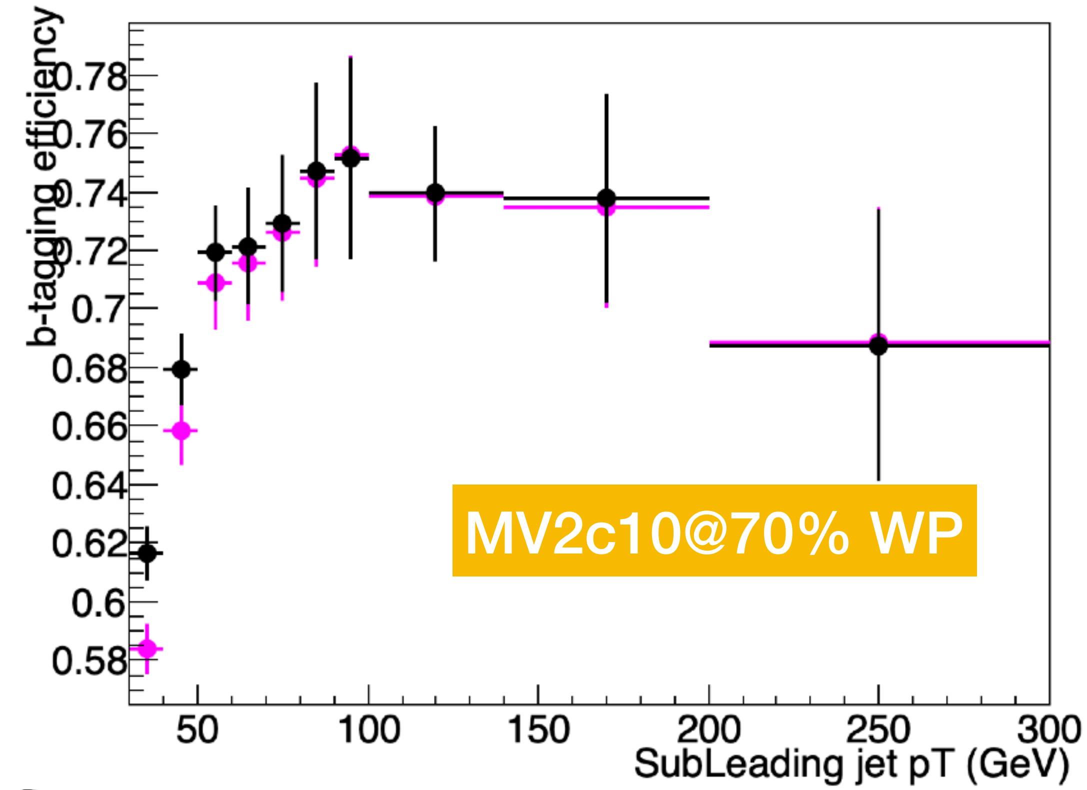


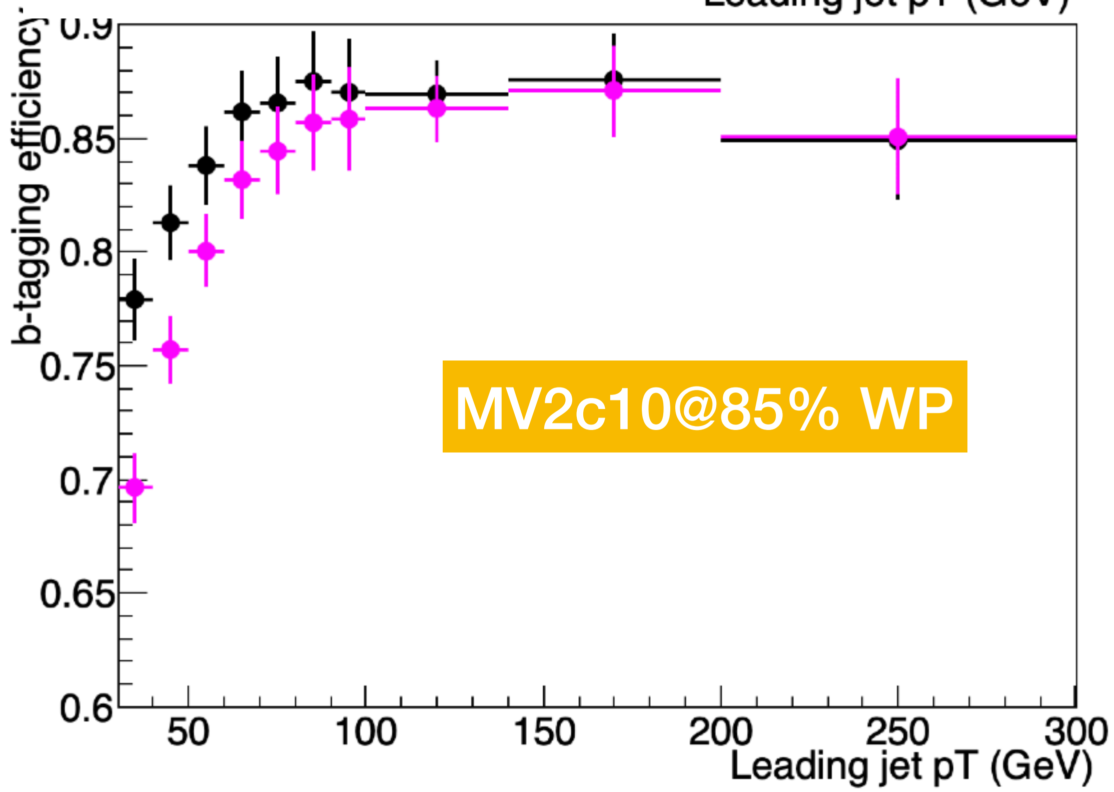
name	cut value	b-jet efficiency [%]	c RR	tau RR	light RR
HybBEff_60	-	61.11	22	151	1150
HybBEff_70	-	70.88	8	38	301
HybBEff_77	-	77.60	4	15	109
HybBEff_85	-	85.27	2	6	27
FixedCutBEff_60	0.94	61.14	22	150	1204
FixedCutBEff_70	0.83	70.84	8	39	313
FixedCutBEff_77	0.64	77.53	4	16	113
FixedCutBEff_85	0.11	85.23	2	6	28

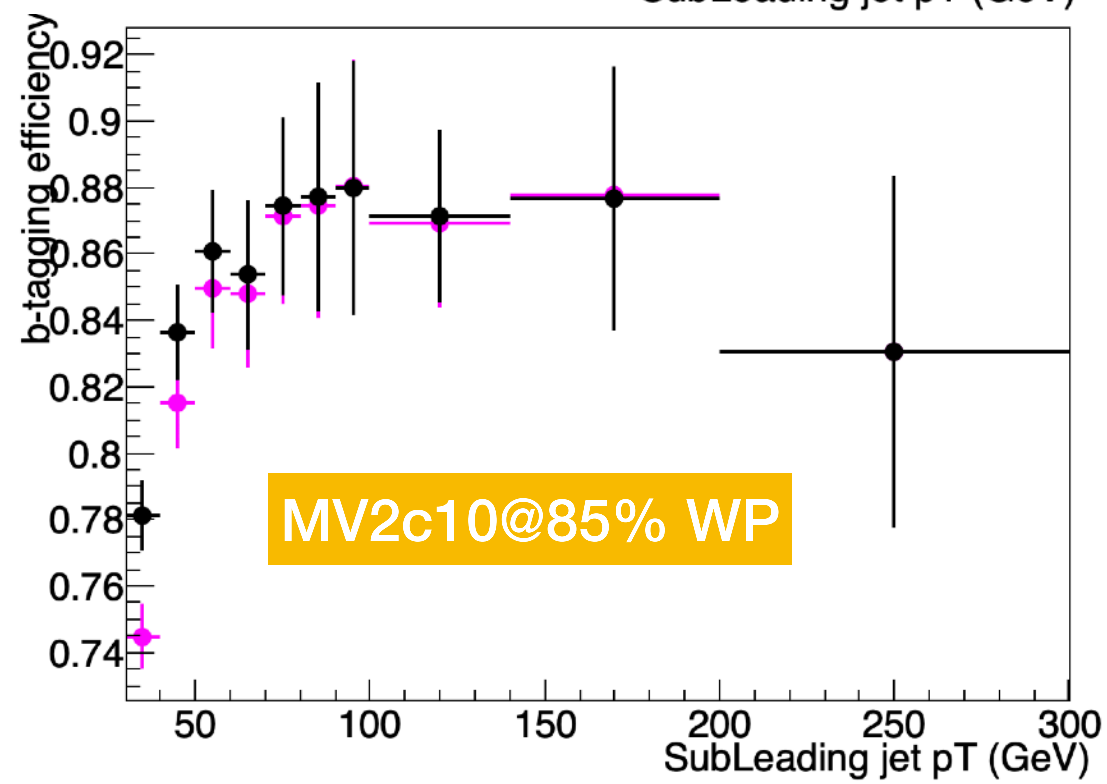
Check on the vertex variables

Further cut on the events when the hardest vertex does not correspond to the selected vertex





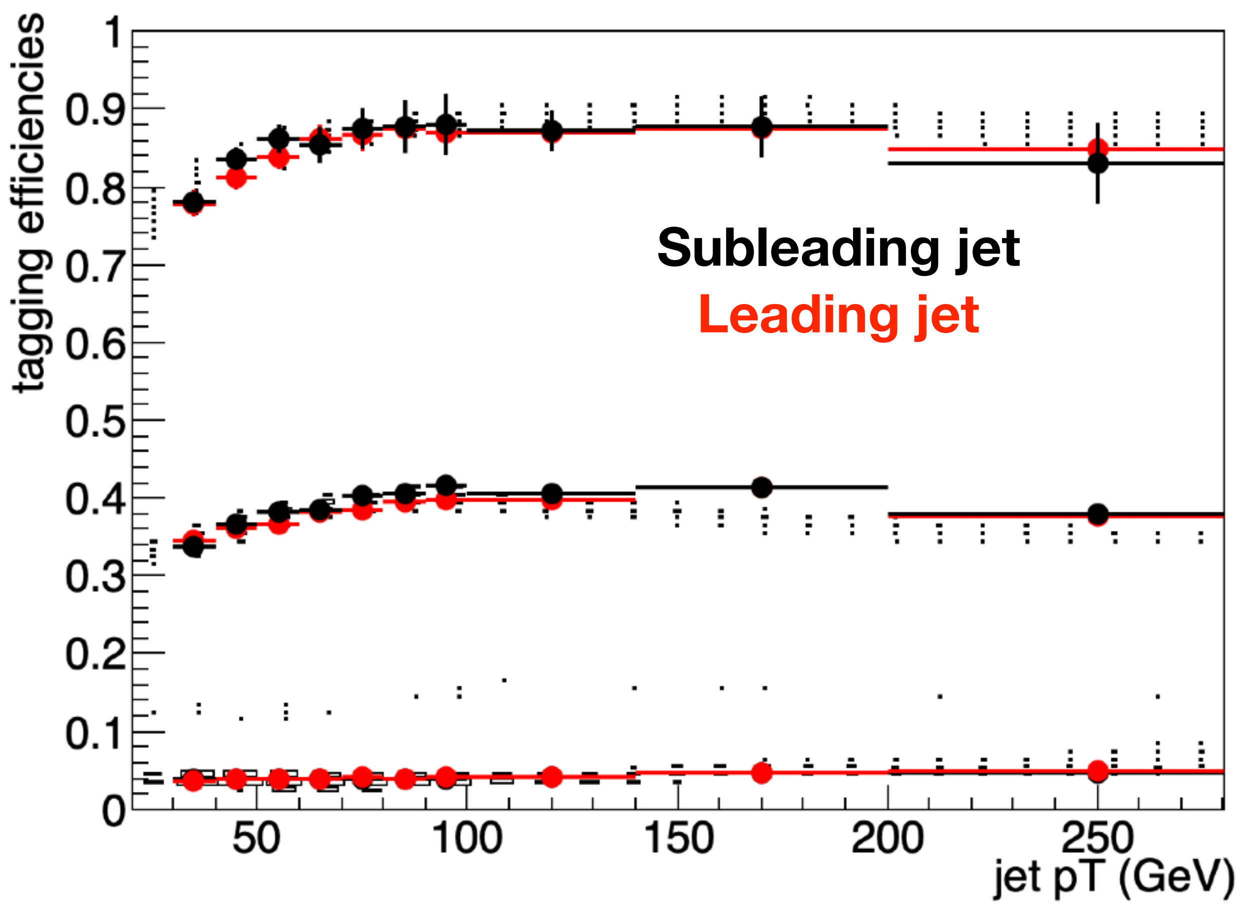




Conclusions on tagging efficiencies

- Our tagging efficiency is lower wrt the performance group result
  - light jets efficiency: in agreement
  - c-jets efficiency: compatible
  - b-jets efficiency: 13% difference at low jet pT
  - these differences appear in all η bins
  - the subheading jet tagging efficiencies have a better agreement
- We updated the jet truth matching according to the recommendations
- The MC/MC scale factors (Pythia8/Sherpa) are not enough to justify the difference
- We checked the MV2 BDT output distribution and the cuts
- The vertex related variables cut is important at low pT

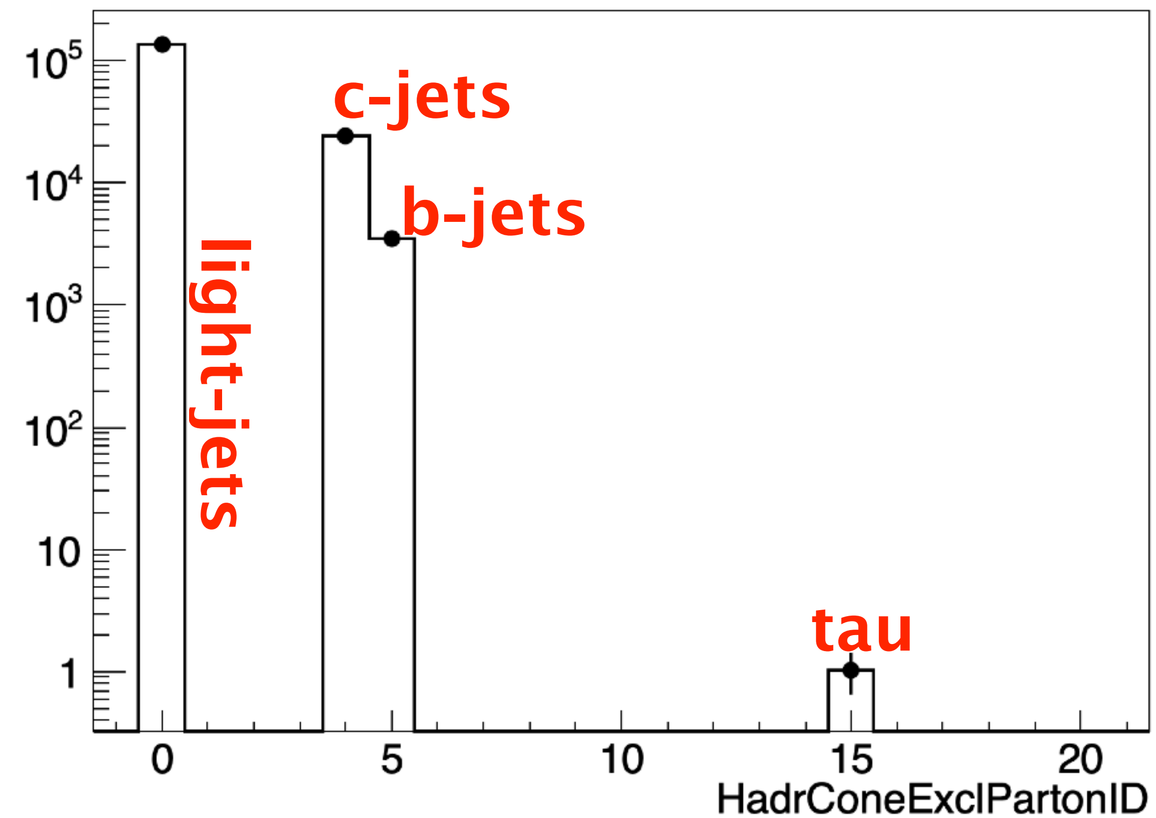
Our best result
I think we have quite good agreement

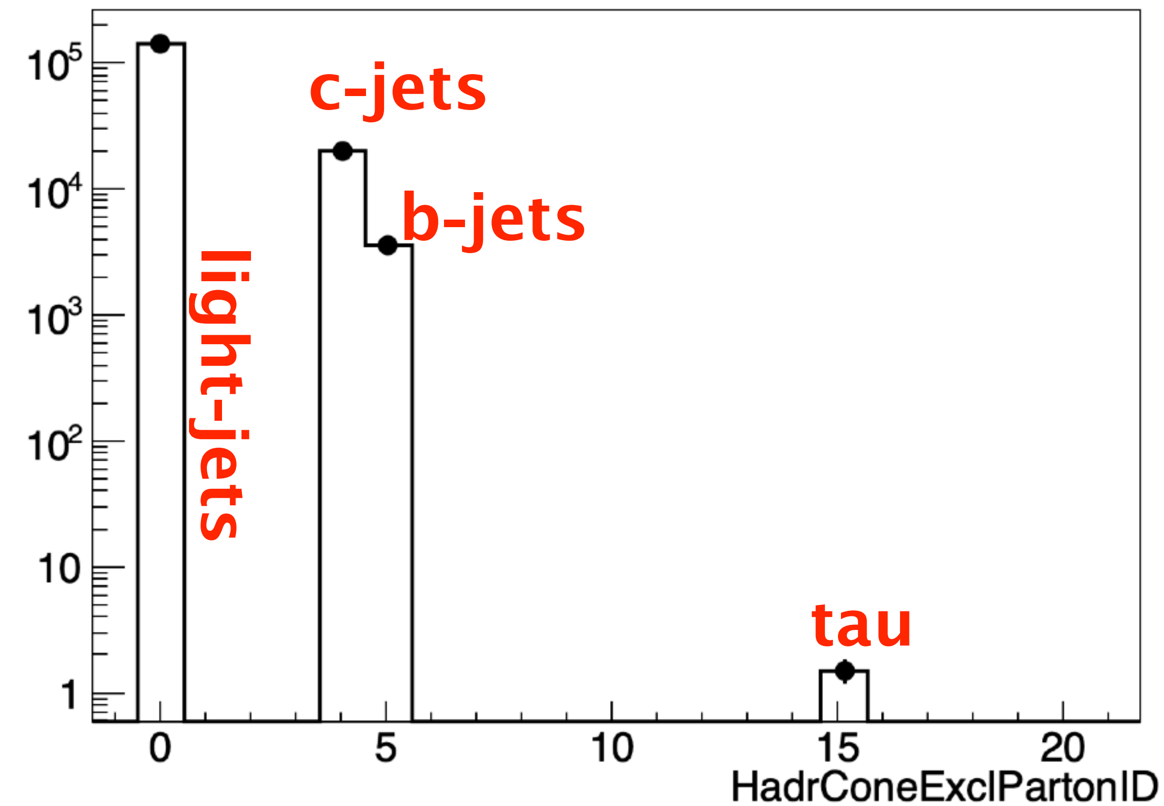


Truth jets flavour fractions

Inclusive Region

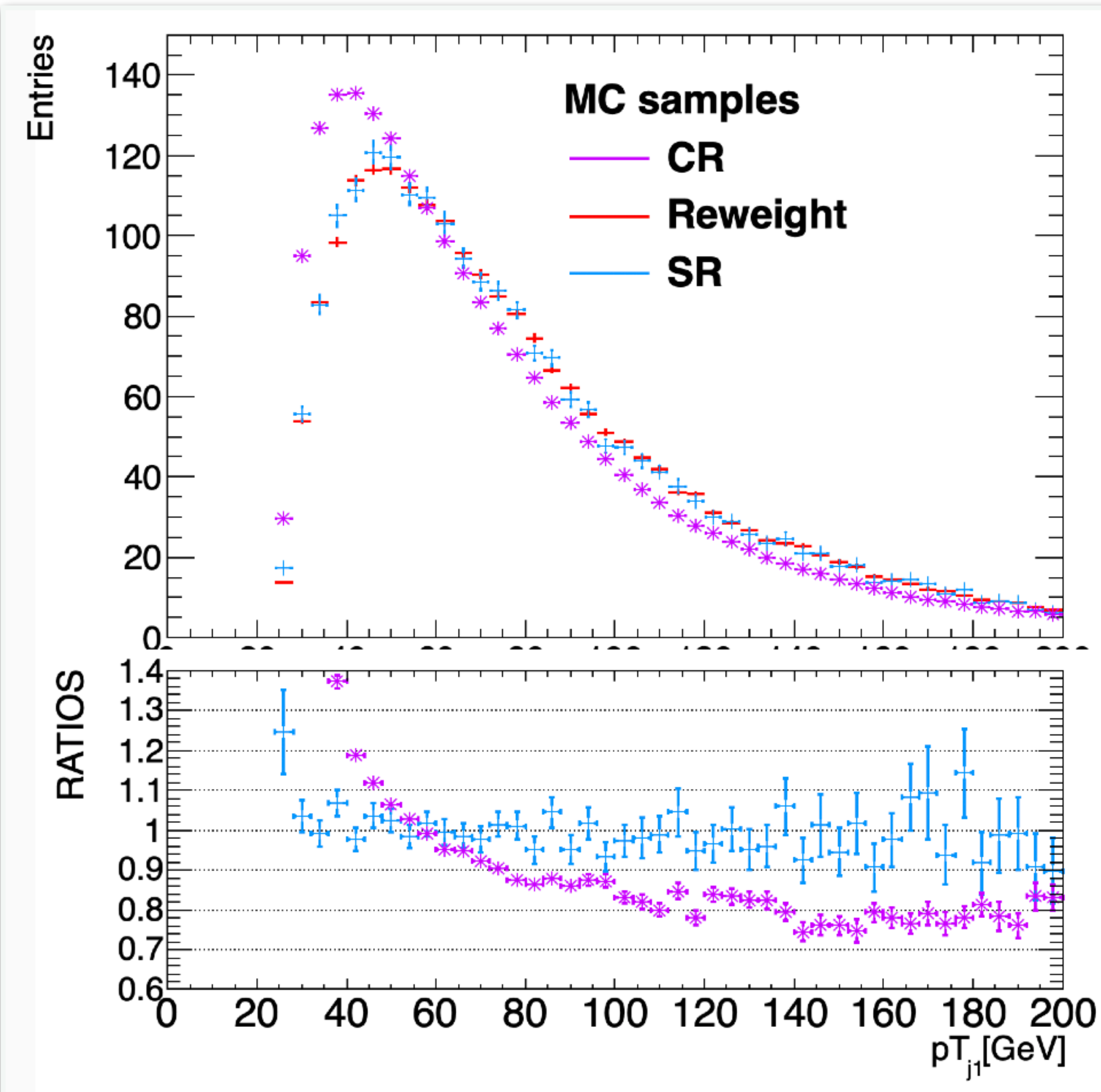
	Leading jet	SubLeading jet
light-jets %	83.2	85.8
c-jets %	14.6	12.1
b-jets %	2.1	2.2
tau %	0.0006	0.0009

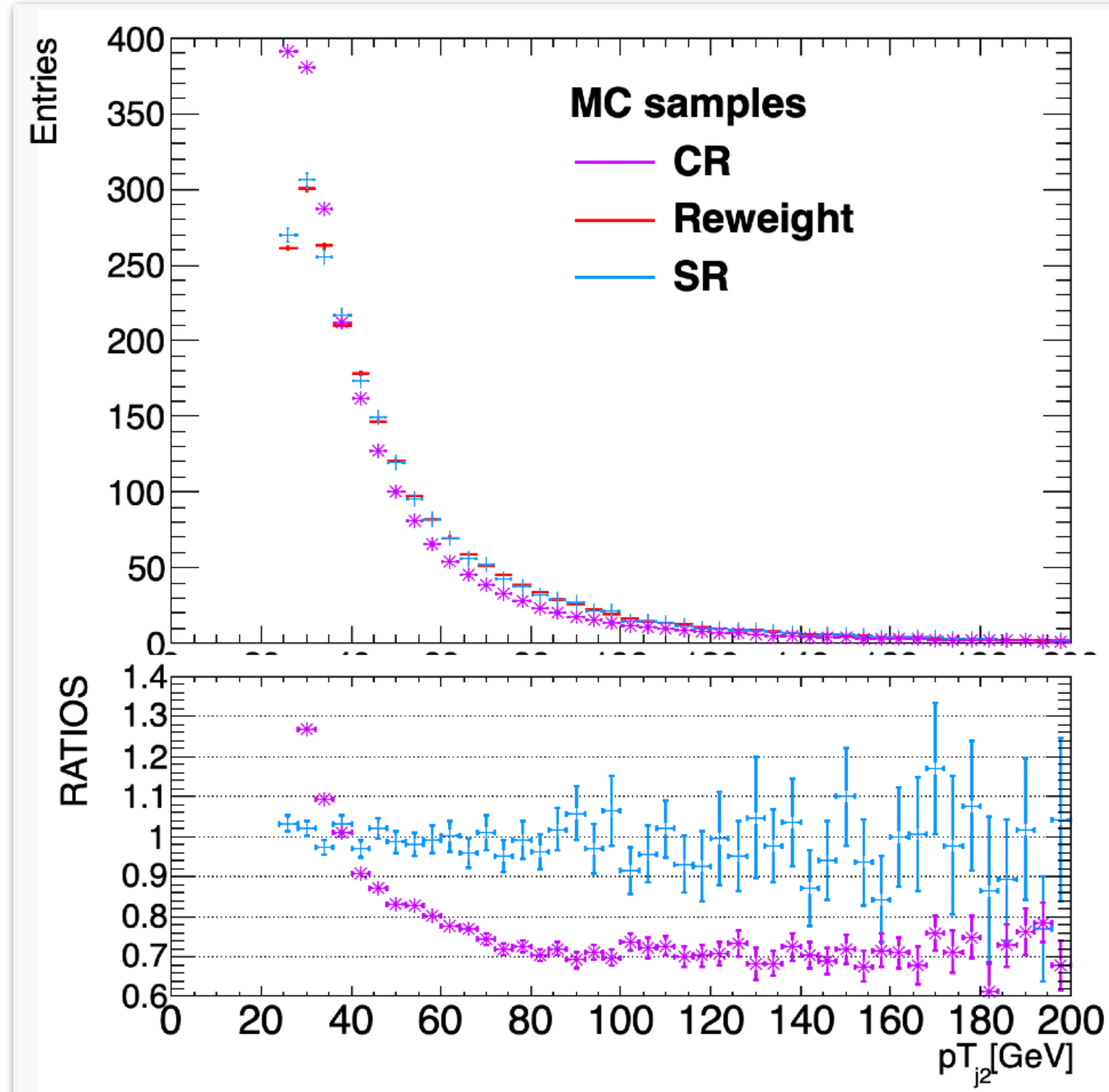




Closure test, pT distributions

Inclusive region

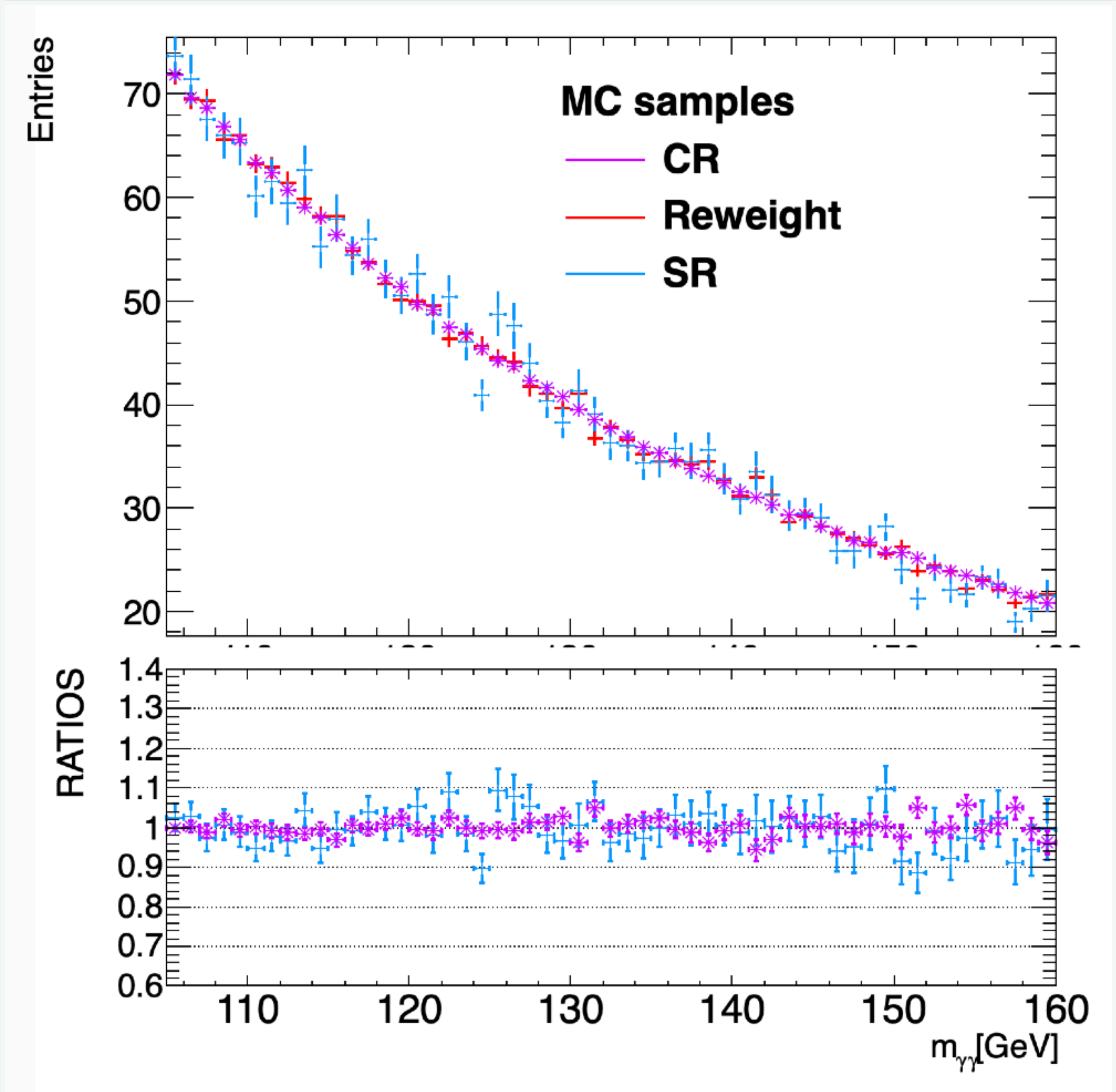




Reweighting done with the new efficiency definition

Closure tests, invariant mass spectra

Inclusive region



MV2c10@85% WP

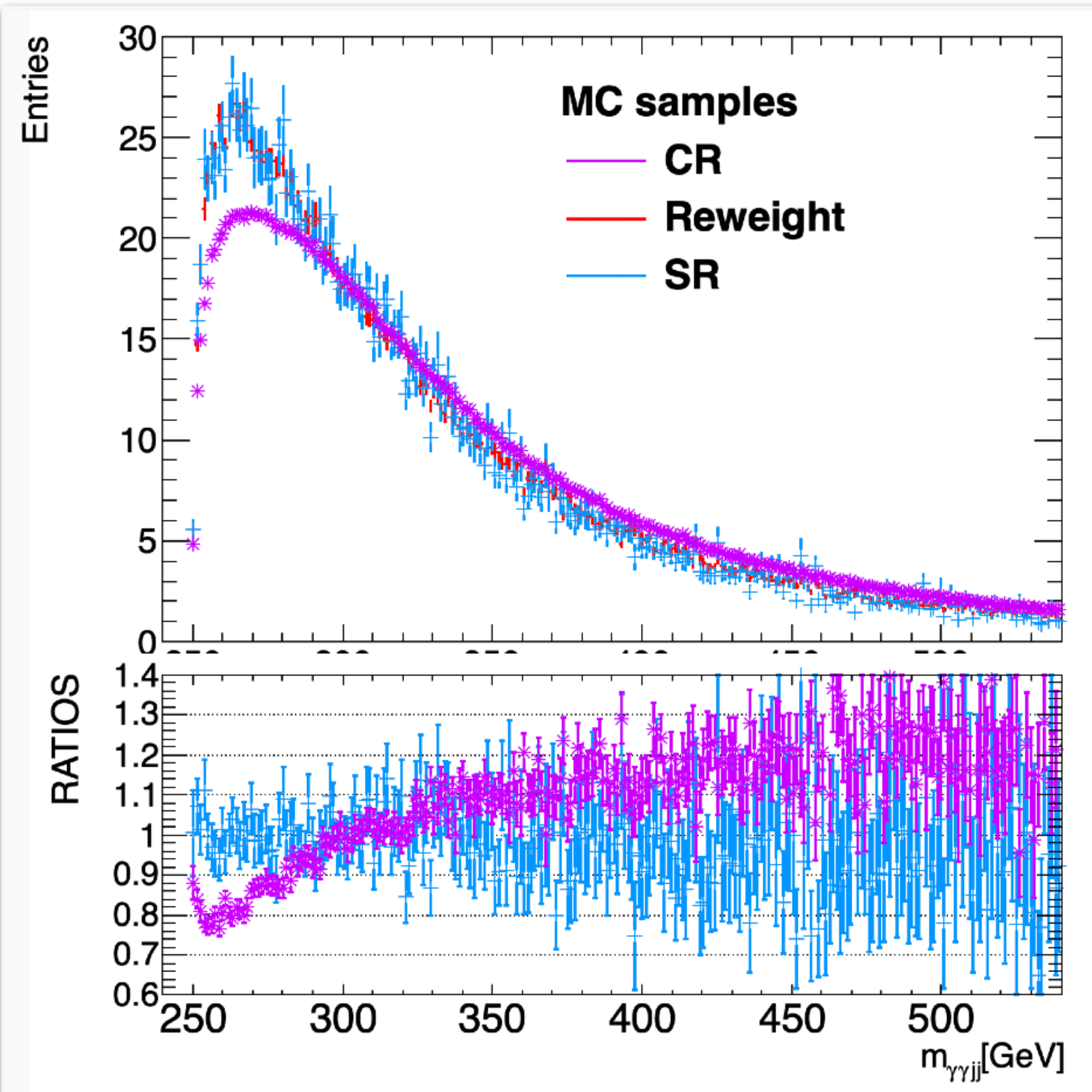
- Application with>=2 central jets
- Signal Region: events with 2 Btag85 jets
- Control Region: events with >=2 central jets
- 2-D reweighting on the pT of the 2 central leading jets
- Assuming third/fourth jets will not pass the Btagging

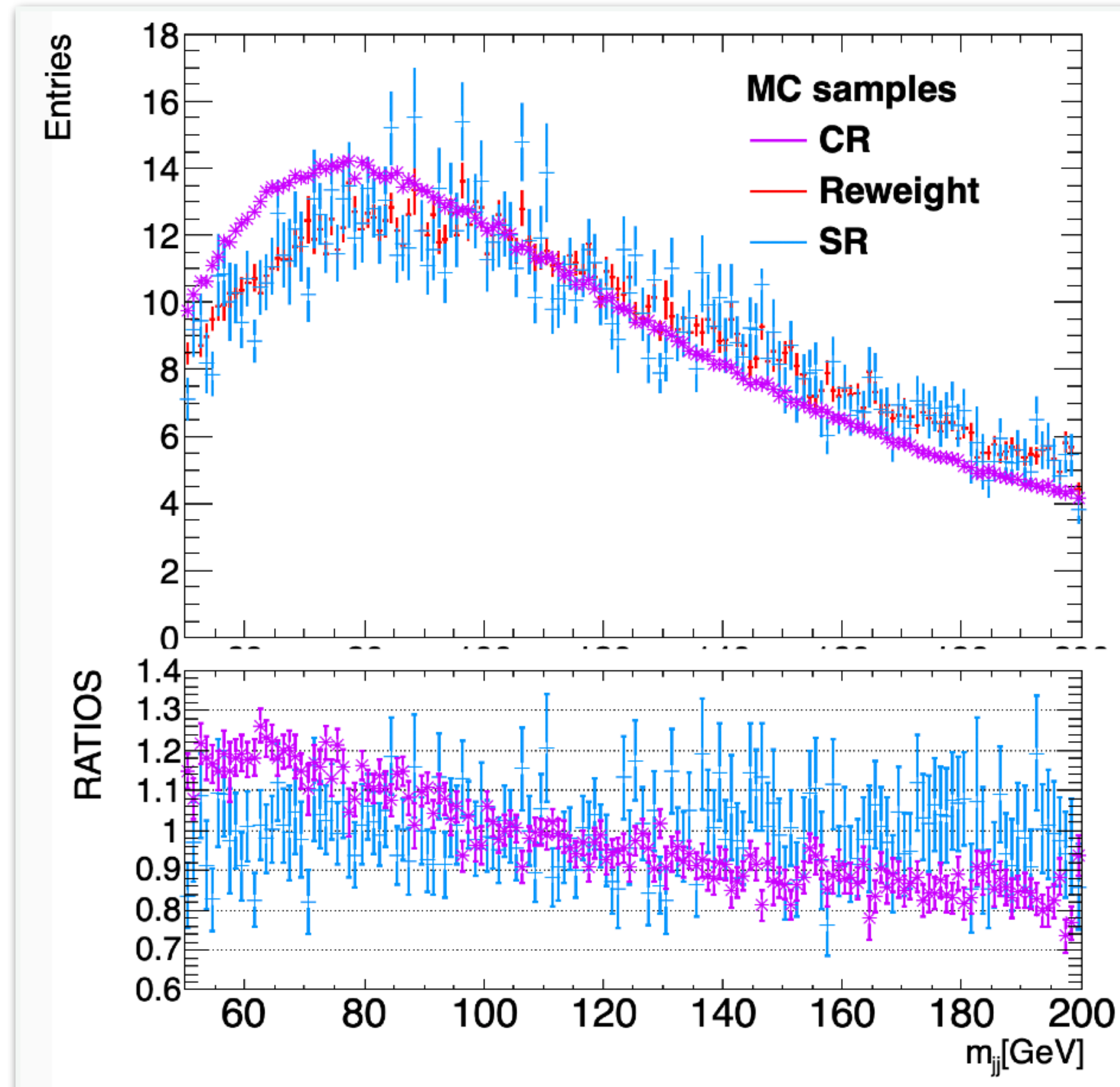
	Stat	MC Yield	(sum weight)^2/ (sum weight^2)
SR	74551	2242.5	28738.6
Reweight	5505400	2121.1	136965
CR	5505400	162668	2026610

- Statical amount is enlarged by ~70 times for Btag85=2
- Effective statistics Reweight/SR ~ 5
- Effective statistics CR/Reweight ~ 15

Closure tests, invariant mass spectra

Inclusive region

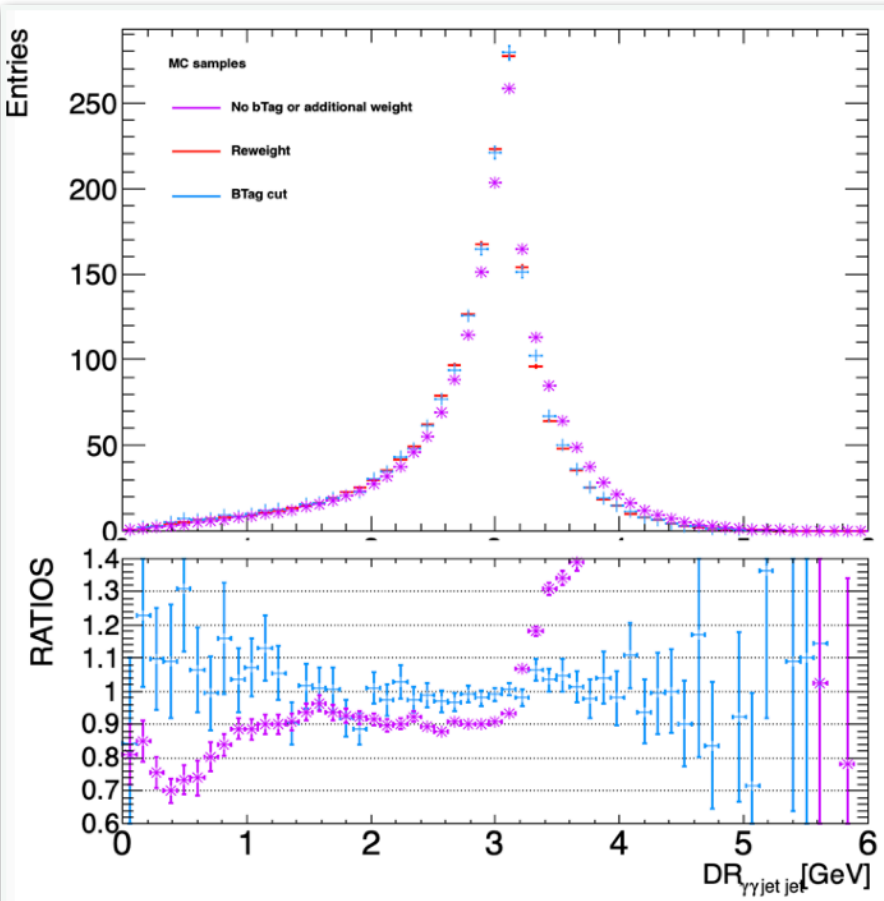




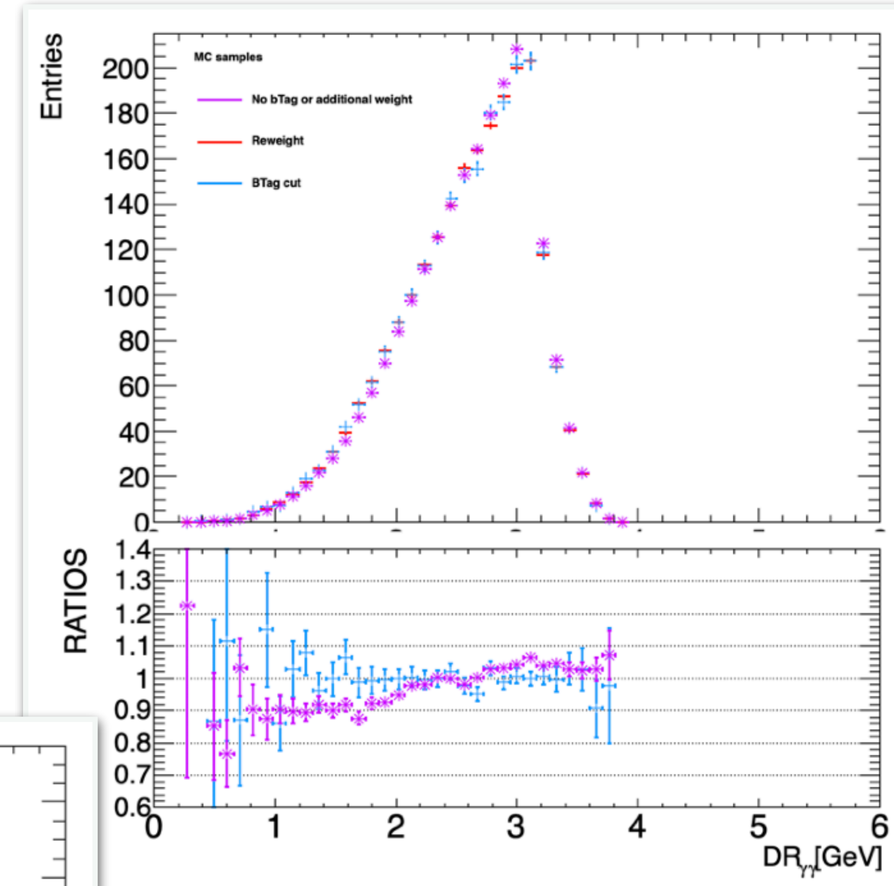
$$M_x^* = M_{b\bar{b}\gamma\gamma} - M_{b\bar{b}} - M_{\gamma\gamma} + 250.0\,\mathrm{GeV}$$

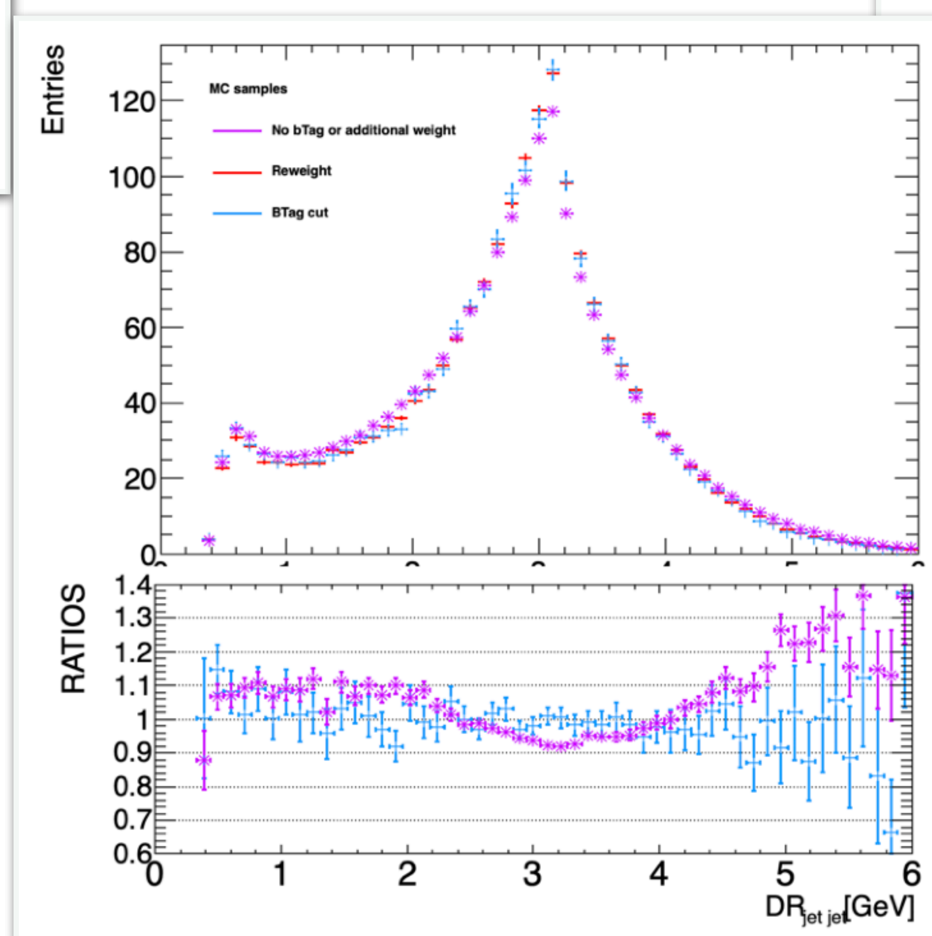
Closure tests-More kinematic variables

Inclusive region



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





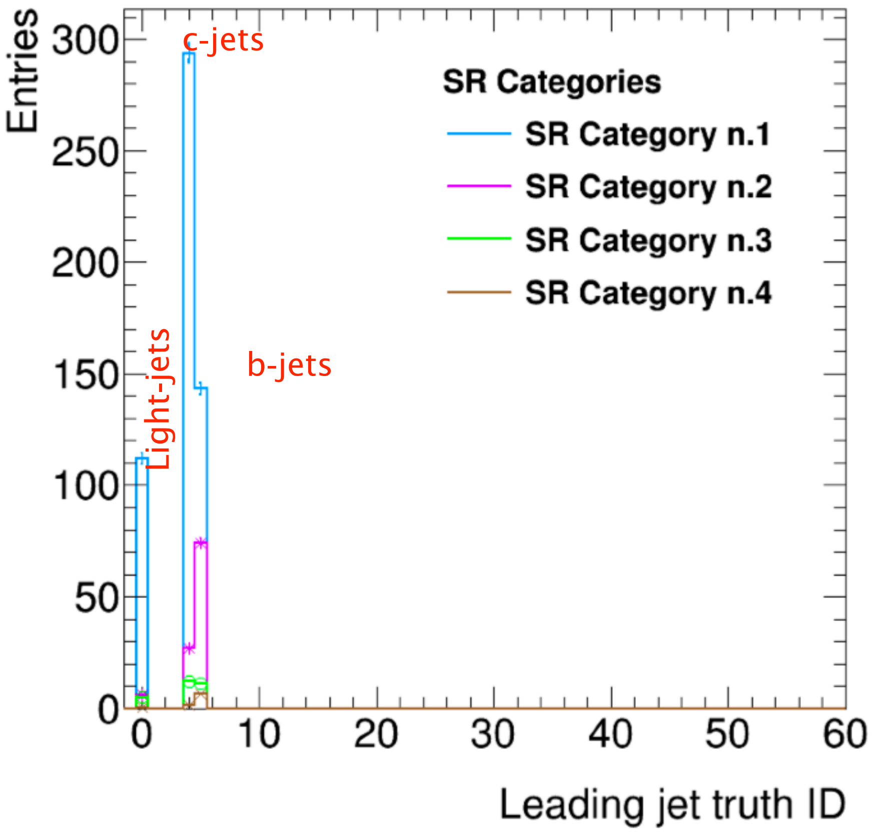
Conclusions

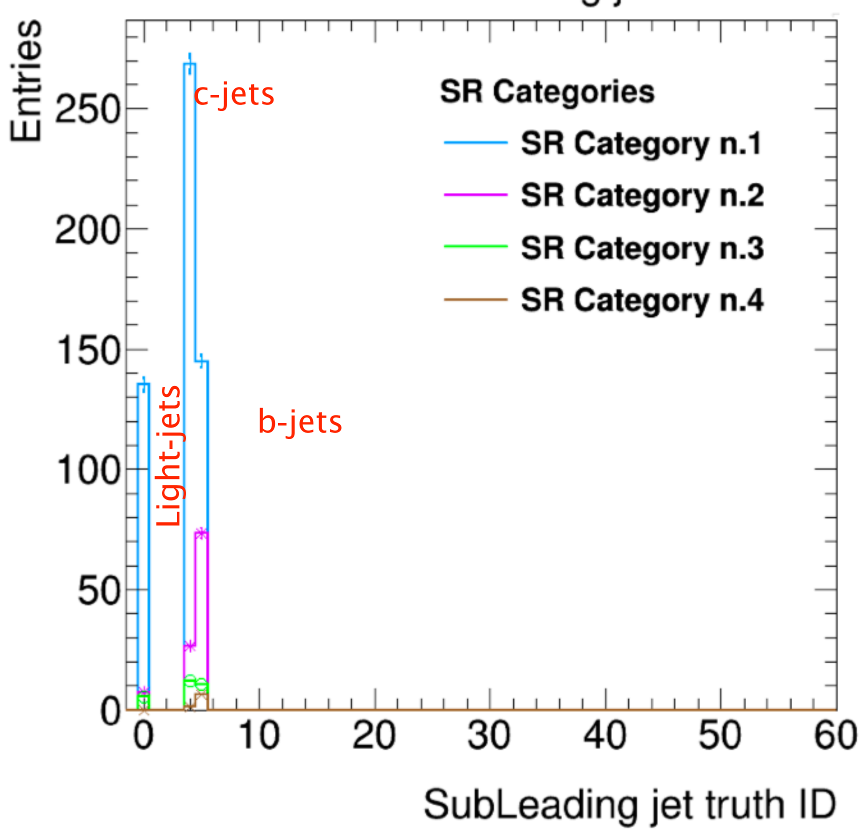
- 2x2D sideband Method to estimate the γγ, γ-jet, jet-γ and jet-jet components
- We did many checks on the tagging efficiency while comparing with the performance group, where the largest difference comes from the vertex
- We also checked the η-dependence, Sherpa-Pythia8 difference and those differences are small
- The reweighted shapes are well modeled not only on mjj and myyjj, but also on the dR variables, which may be helpful in the categorization optimization
- We estimated the effective statistics in the inclusive sample and in the 4 SR categories
    - There is a gain using the reweighted technique wrt the SR (a factor of 4)
    - But we have too few truth b-jet which driven the statistic in the re-weighted shape

Back-up

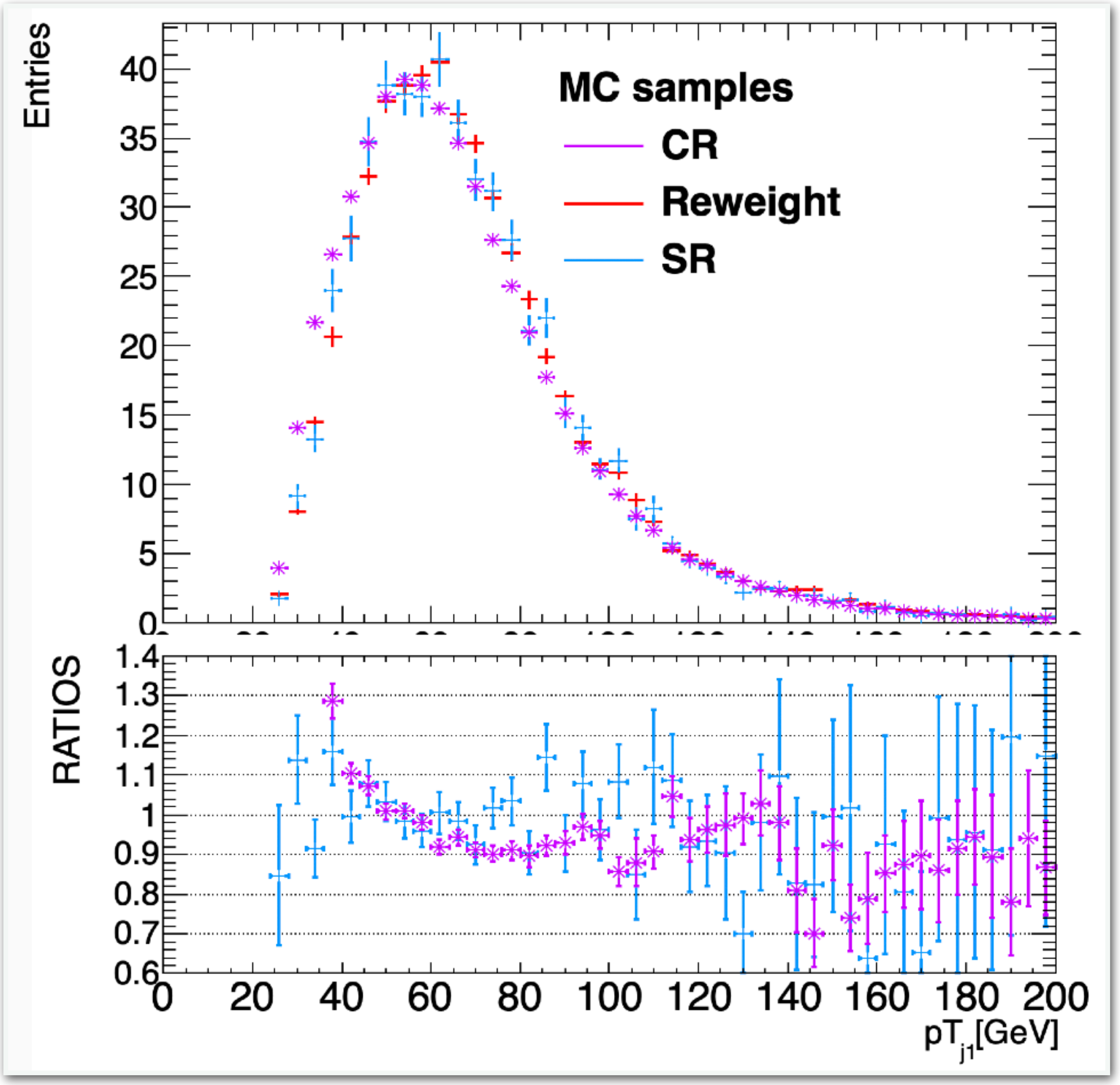
Truth Level Particle ID, SR categories

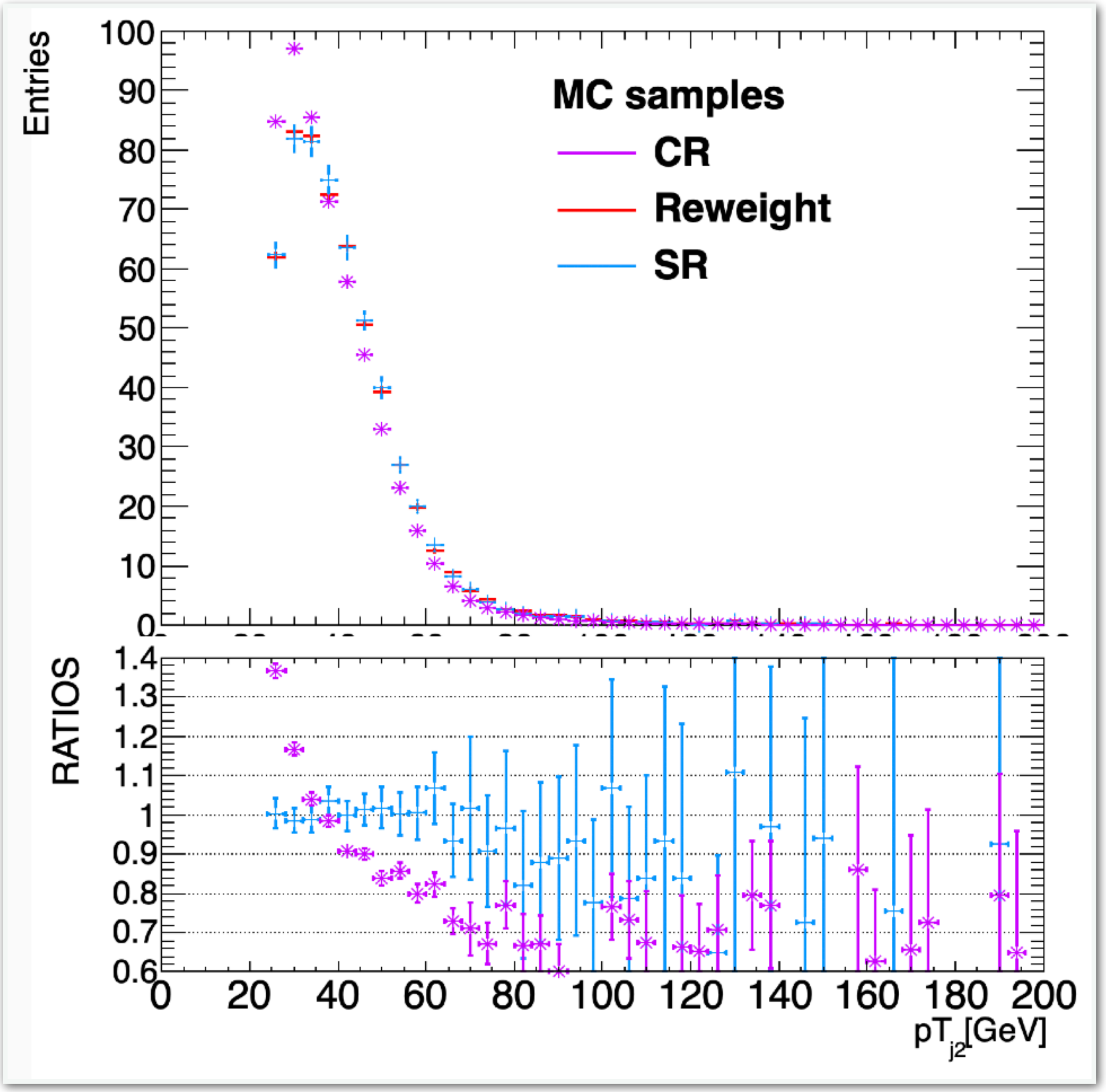
		Cat. 1	Cat. 2	Cat. 3	Cat. 4
Leading jet	Light-jets %	20.4	5.6	17.4	1.6
	C-jets %	53.5	25.2	42.6	14.4
	B-jets %	26.1	69.2	39.9	84.0
SubLeading jet	Light-jets %	24.6	6.8	19.7	0.8
	C-jets %	48.9	24.7	42.8	17.3
	B-jets %	26.4	68.5	37.4769	81.9



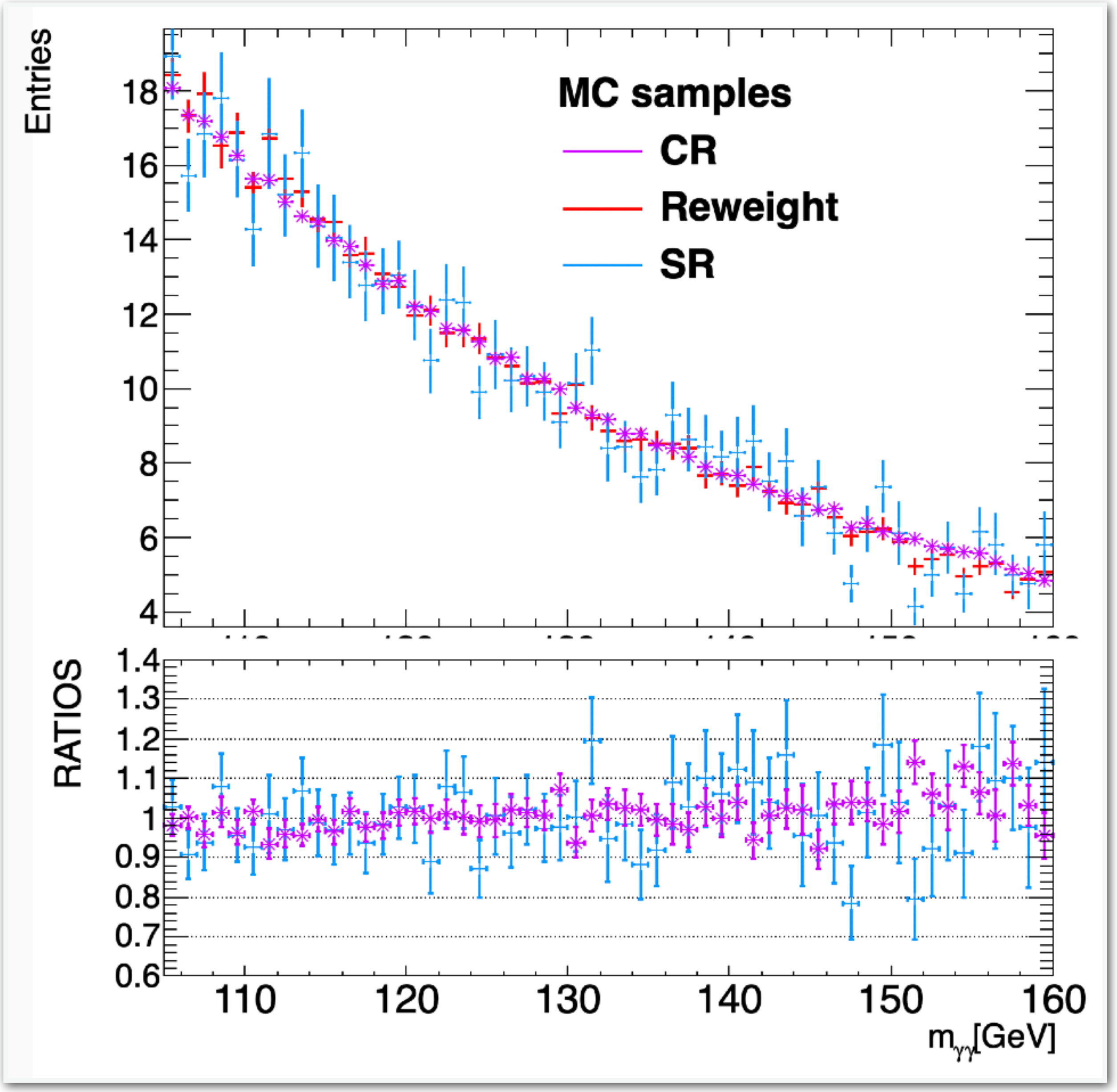


Closure test, pT distributions-SR1





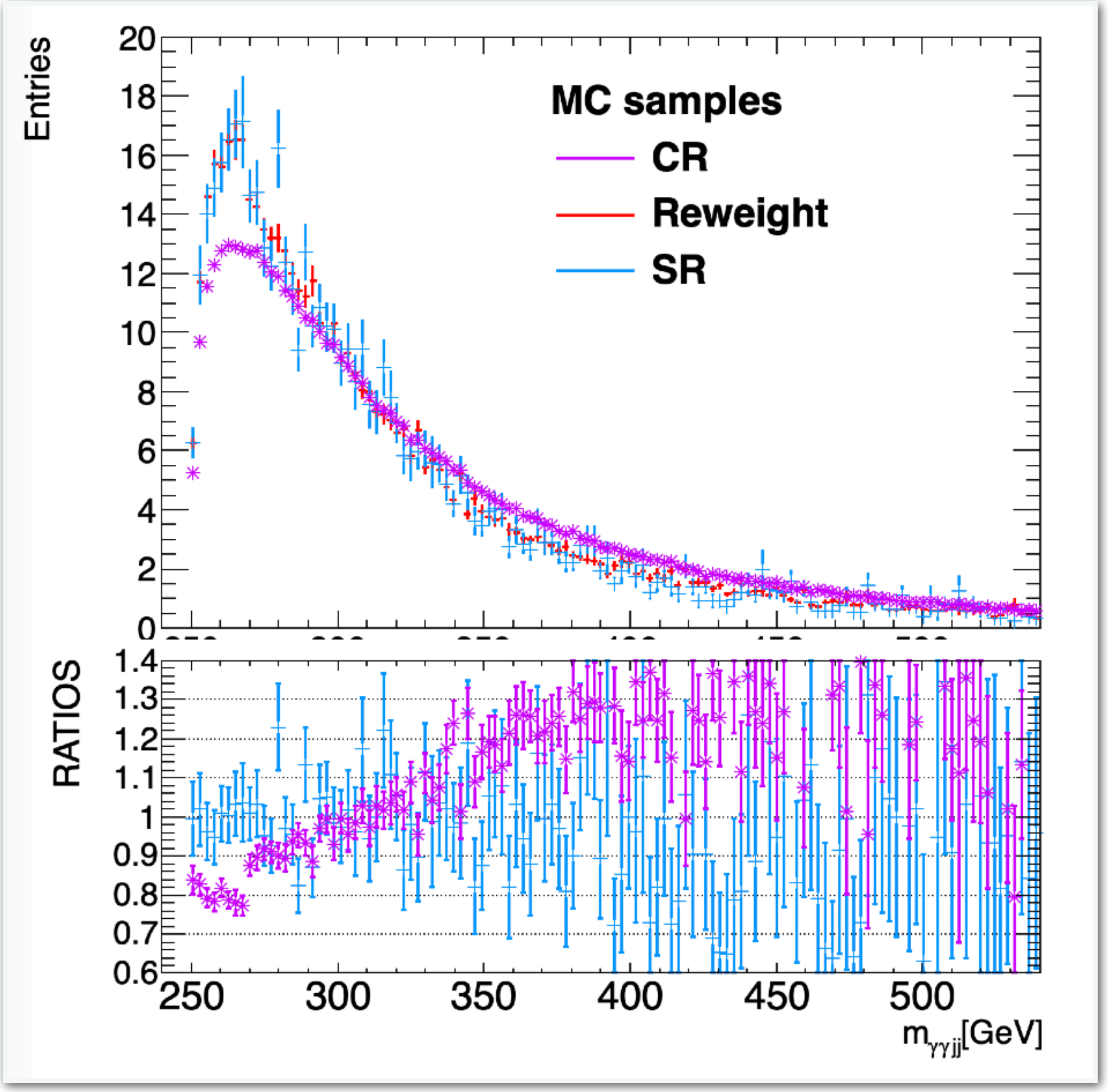
Closure test, mass spectra-SR1



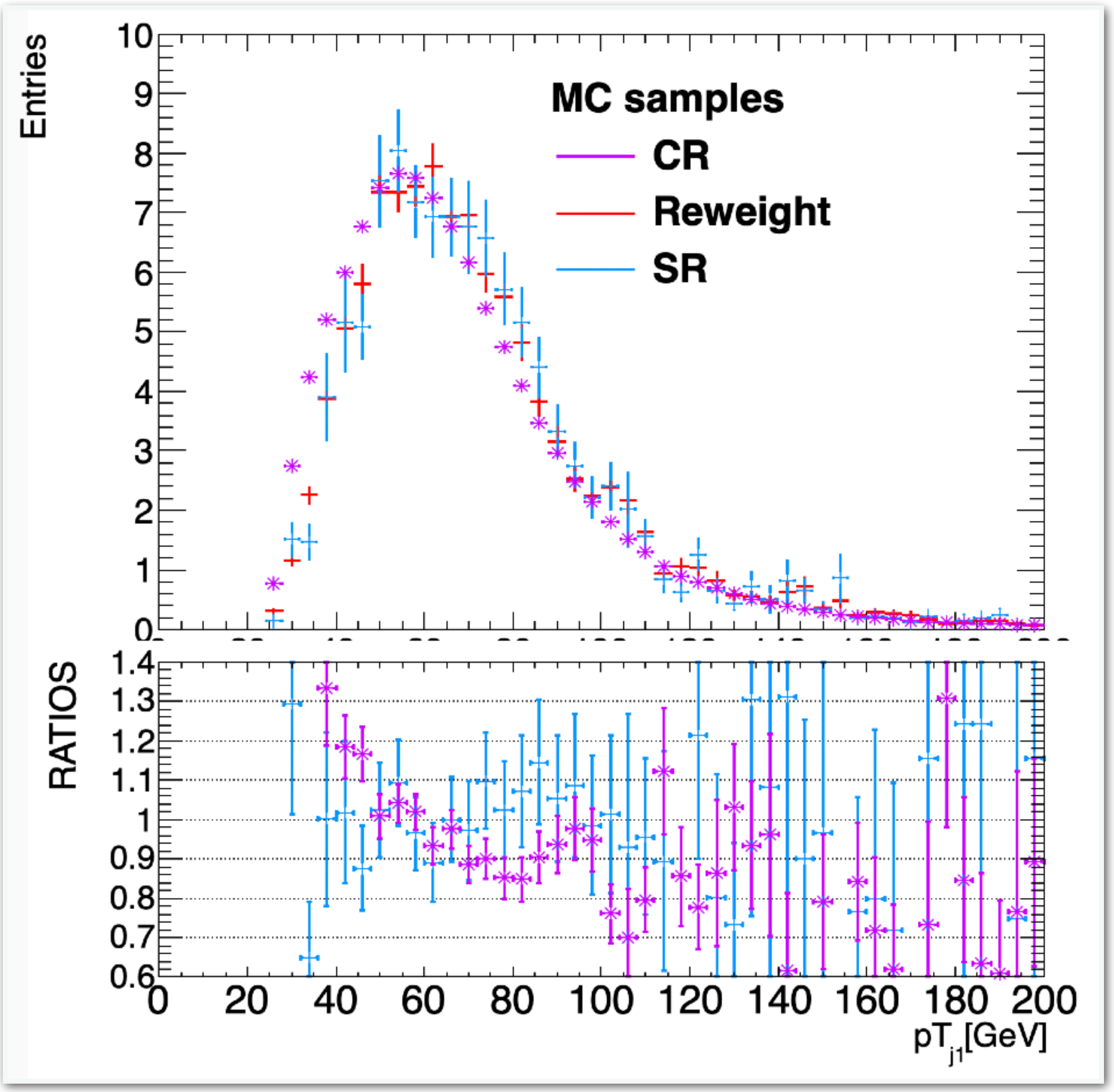
	Stat	MC Yield	(sum weight)^2/ (sum weight^2)
SR	18044	548.6	7058.3
Reweight	1376260	529.2	35913.8
CR	1376260	40977.6	504060

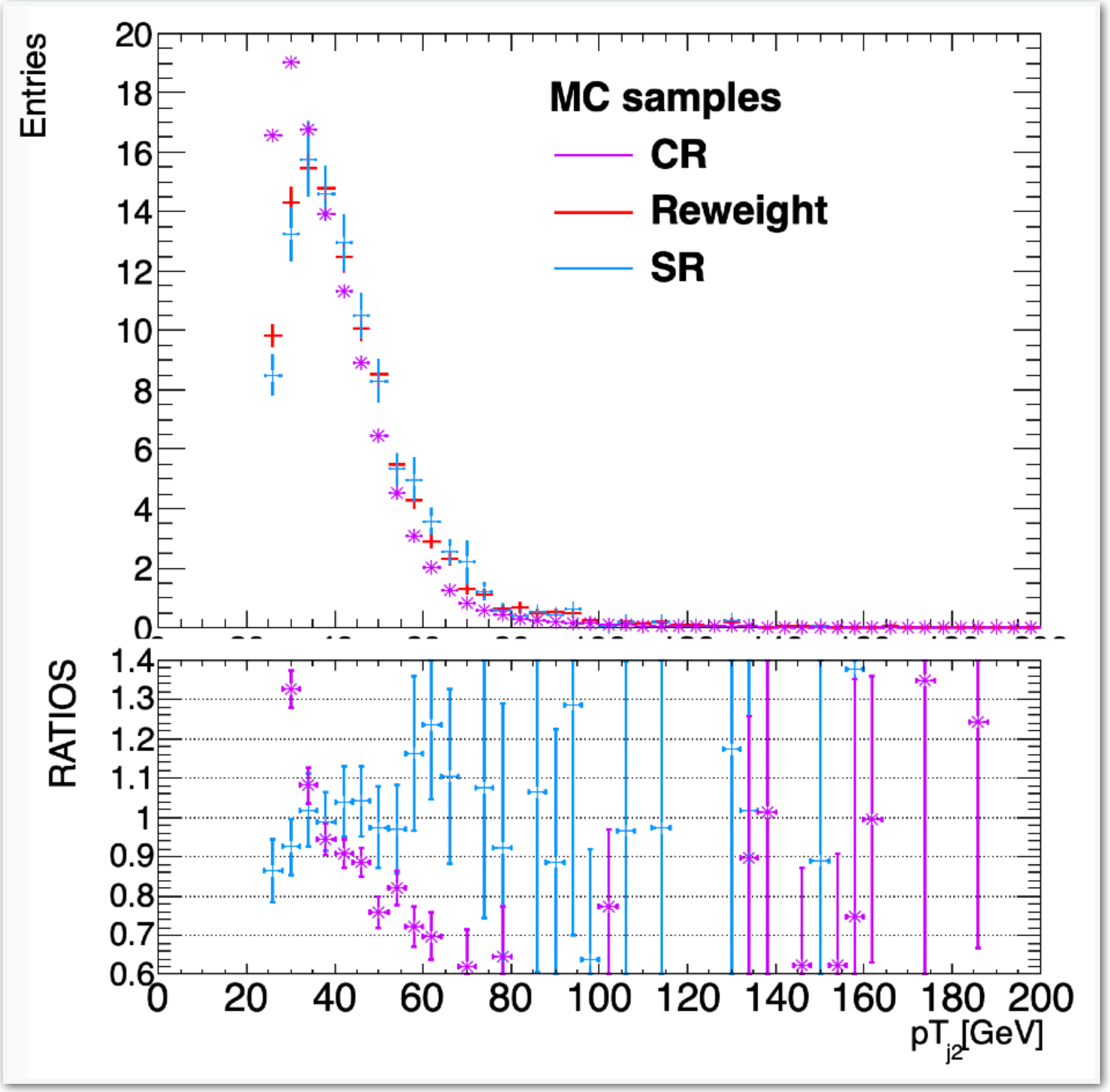
- In the reweighted case, the statical amount is enlarged by ~76 wrt SR
- Effective statistics Reweight/SR ~ 5
- Effective statistics CR/Reweight ~ 15

Closure test, mass spectra-SR1

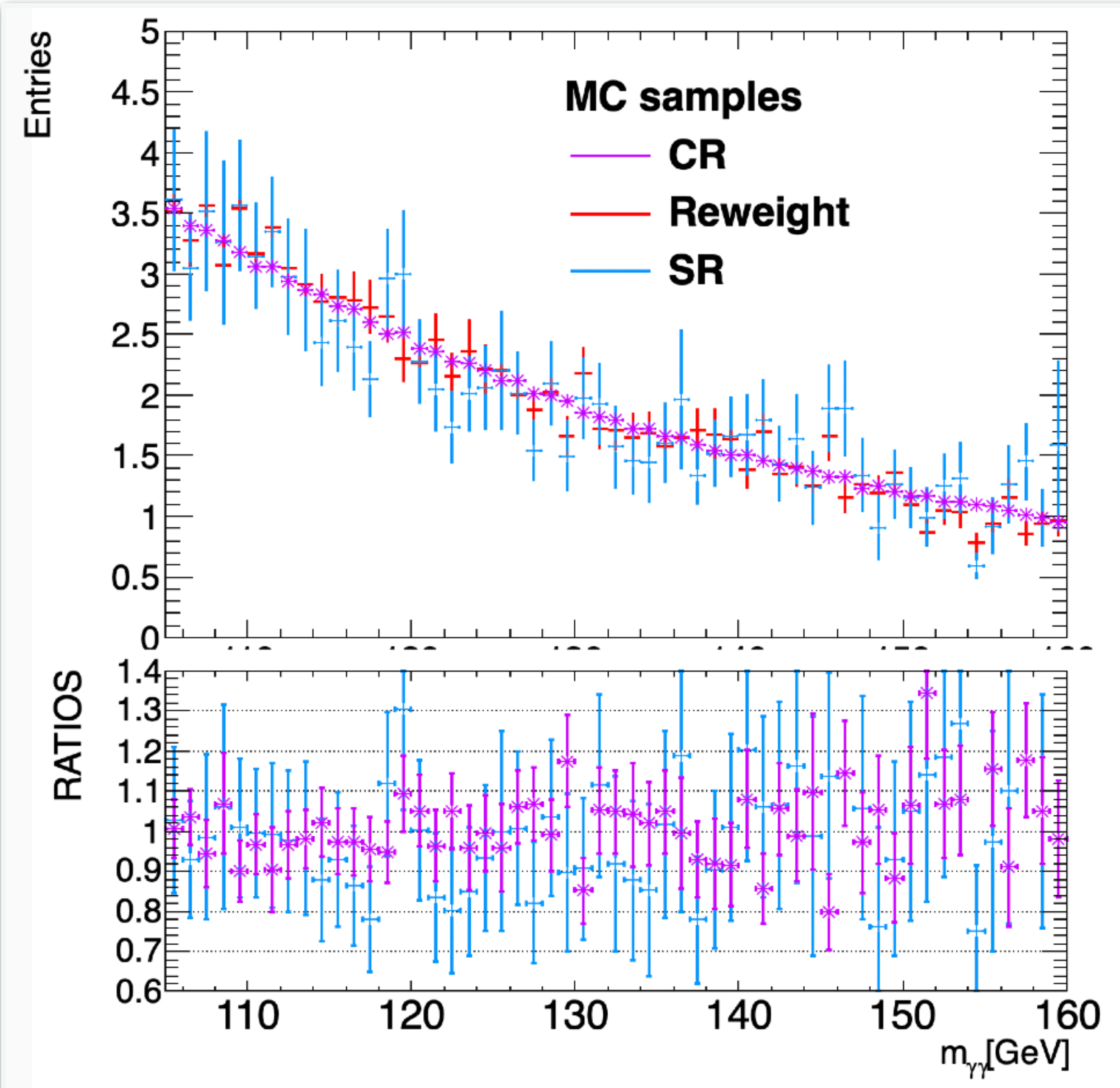


Closure test, pT distributions-SR2





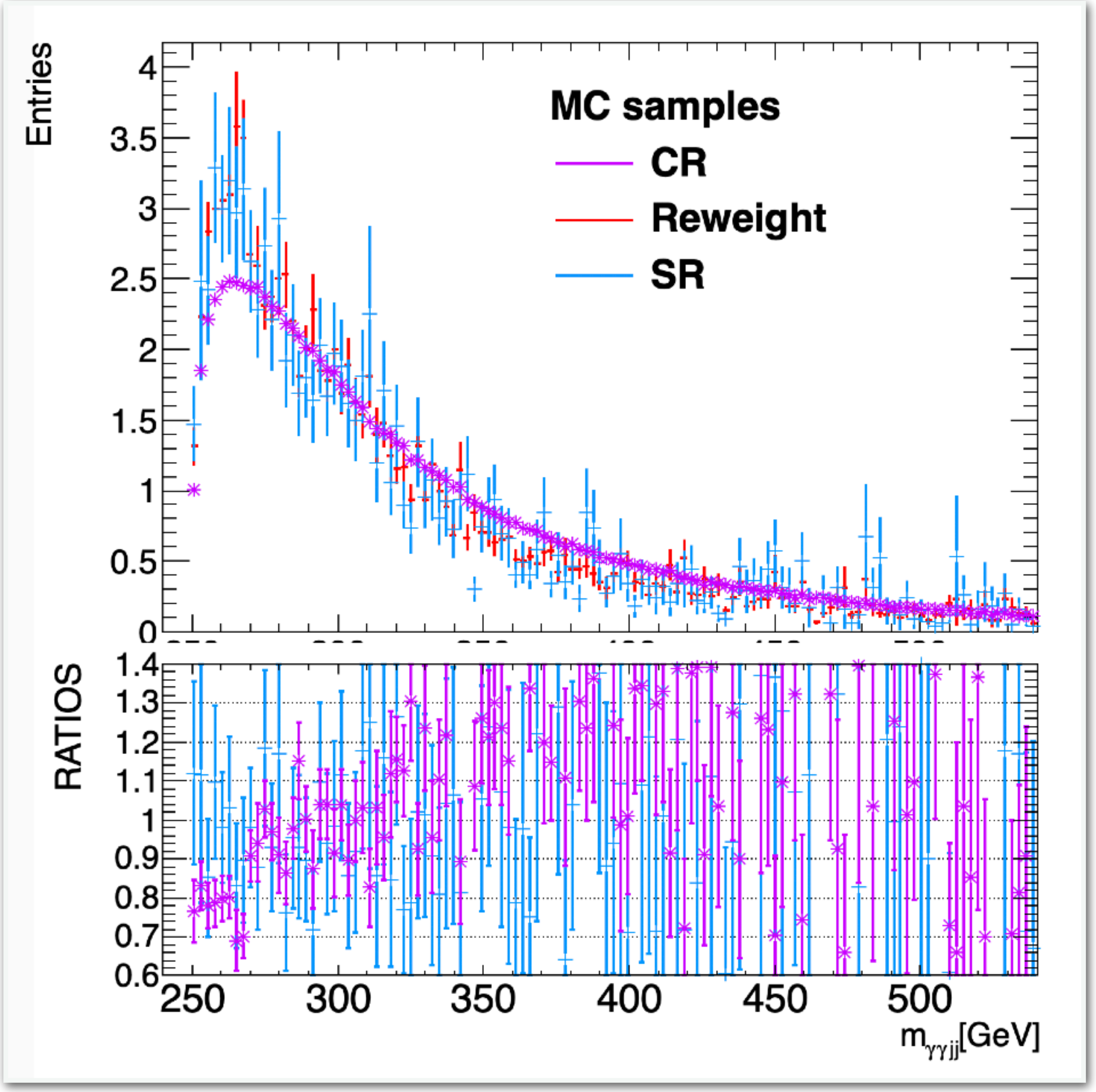
Closure test, mass spectra-SR2



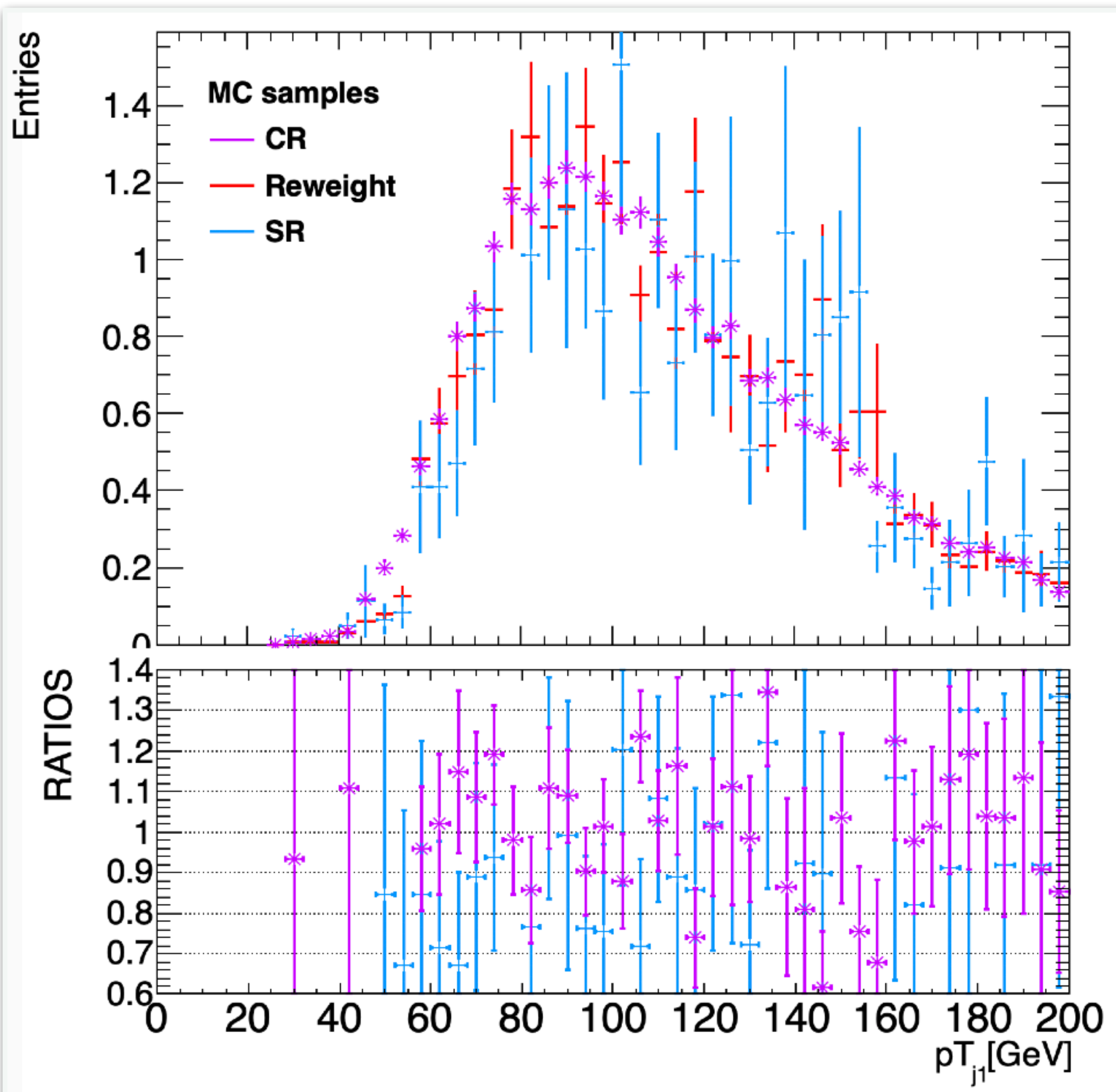
	Stat	MC Yield	(sum weight)^2/ (sum weight^2)
SR	3497	107.4	1403.6
Reweight	1376260	109.1	5015.3
CR	1376260	40977.6	504060

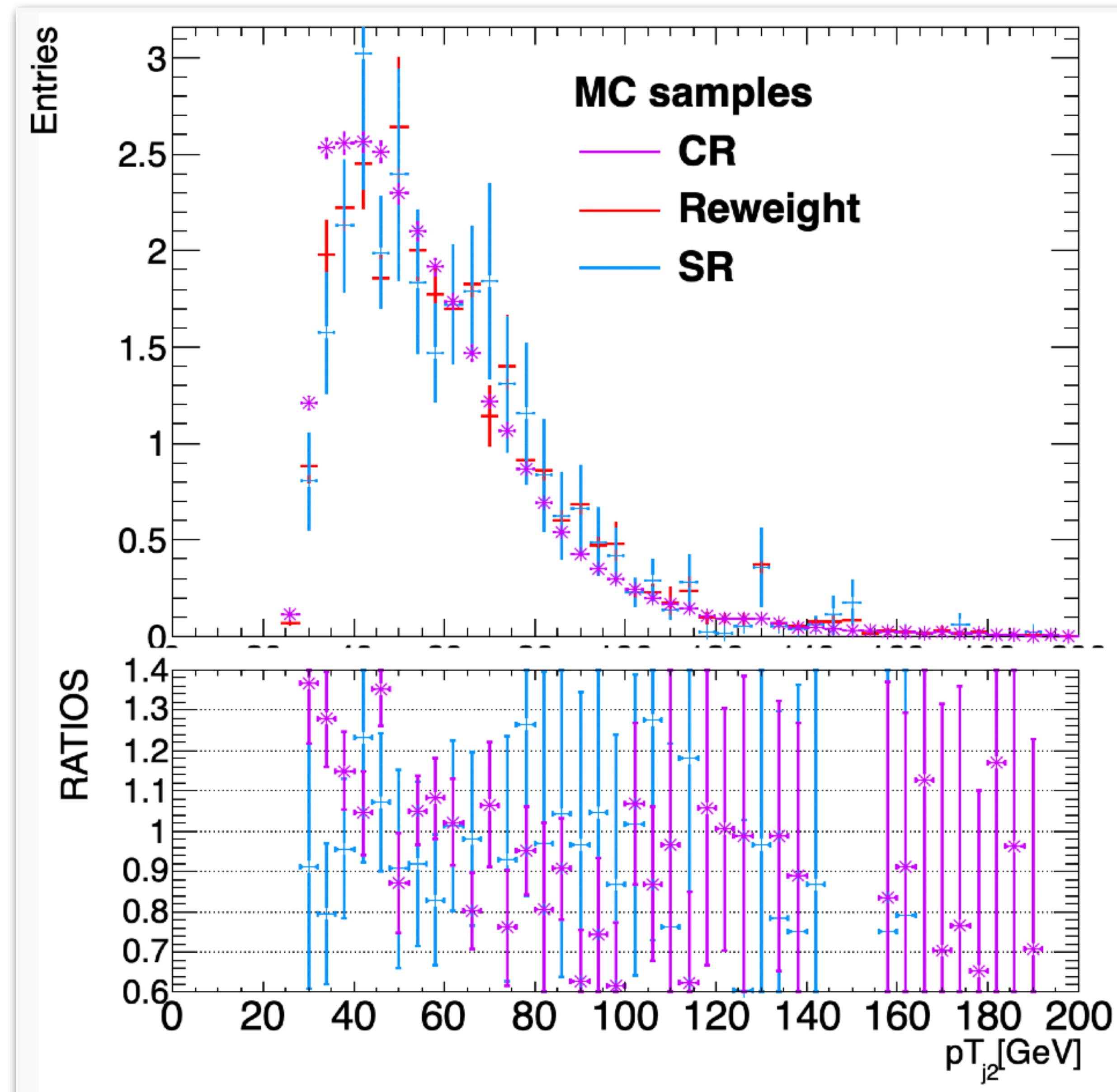
- In the reweighted case, the statical amount is enlarged by ~390 wrt SR
- Effective statistics Reweight/SR ~ 3.5
- Effective statistics CR/Reweight ~ 100

Closure test, mass spectra-SR2

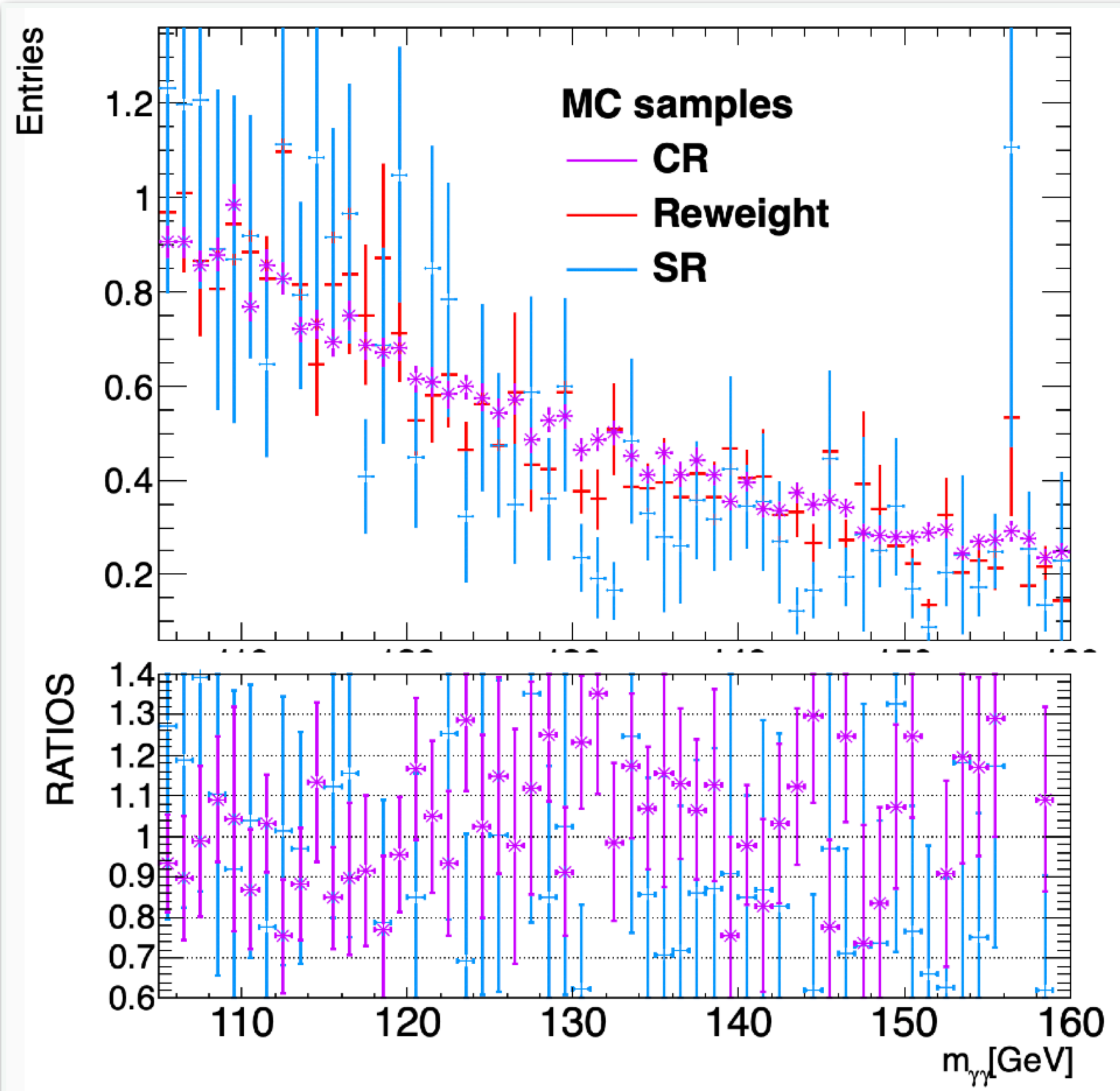


Closure test, pT distributions-SR3





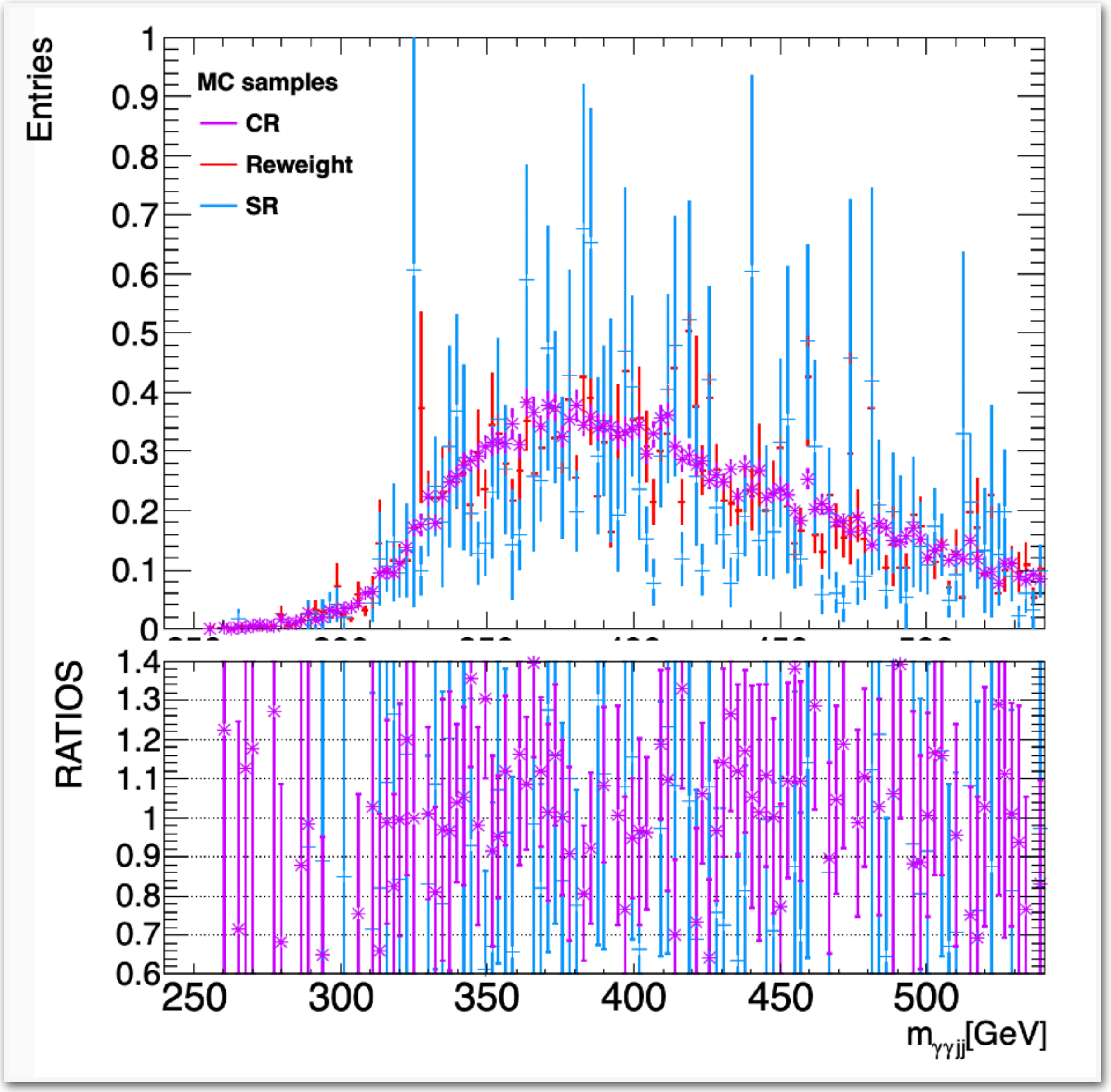
Closure test, mass spectra-SR3



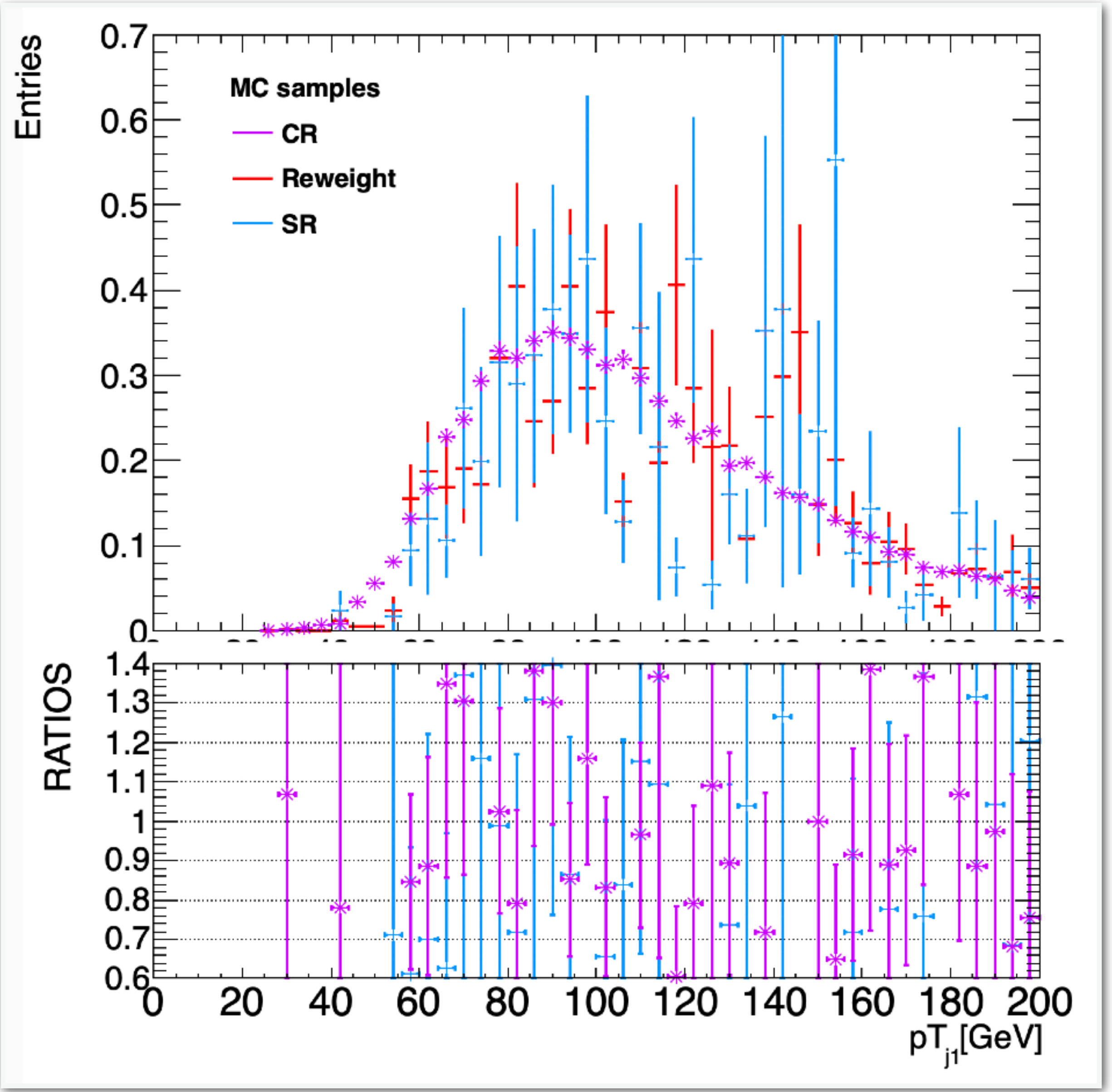
	Stat	MC Yield	(sum weight)^2/ (sum weight^2)
SR	828	28.0	316.3
Reweight	51683	27.3	1190.6
CR	51683	1730.6	20639.3

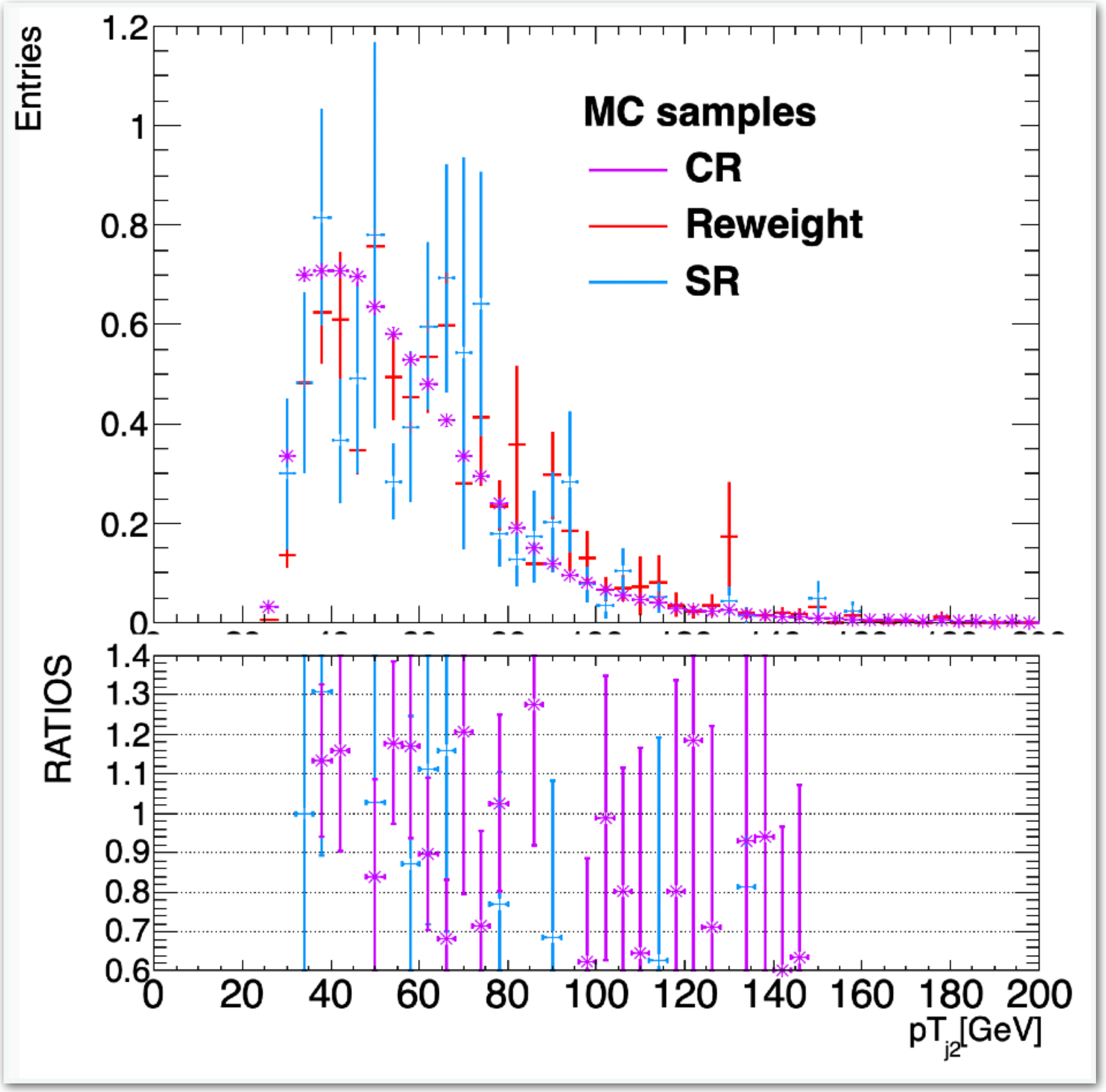
- In the reweighted case, the statical amount is enlarged by ~62 wrt SR
- Effective statistics Reweight/SR ~ 3.8
- Effective statistics CR/Reweight ~ 17

Closure test, mass spectra-SR3



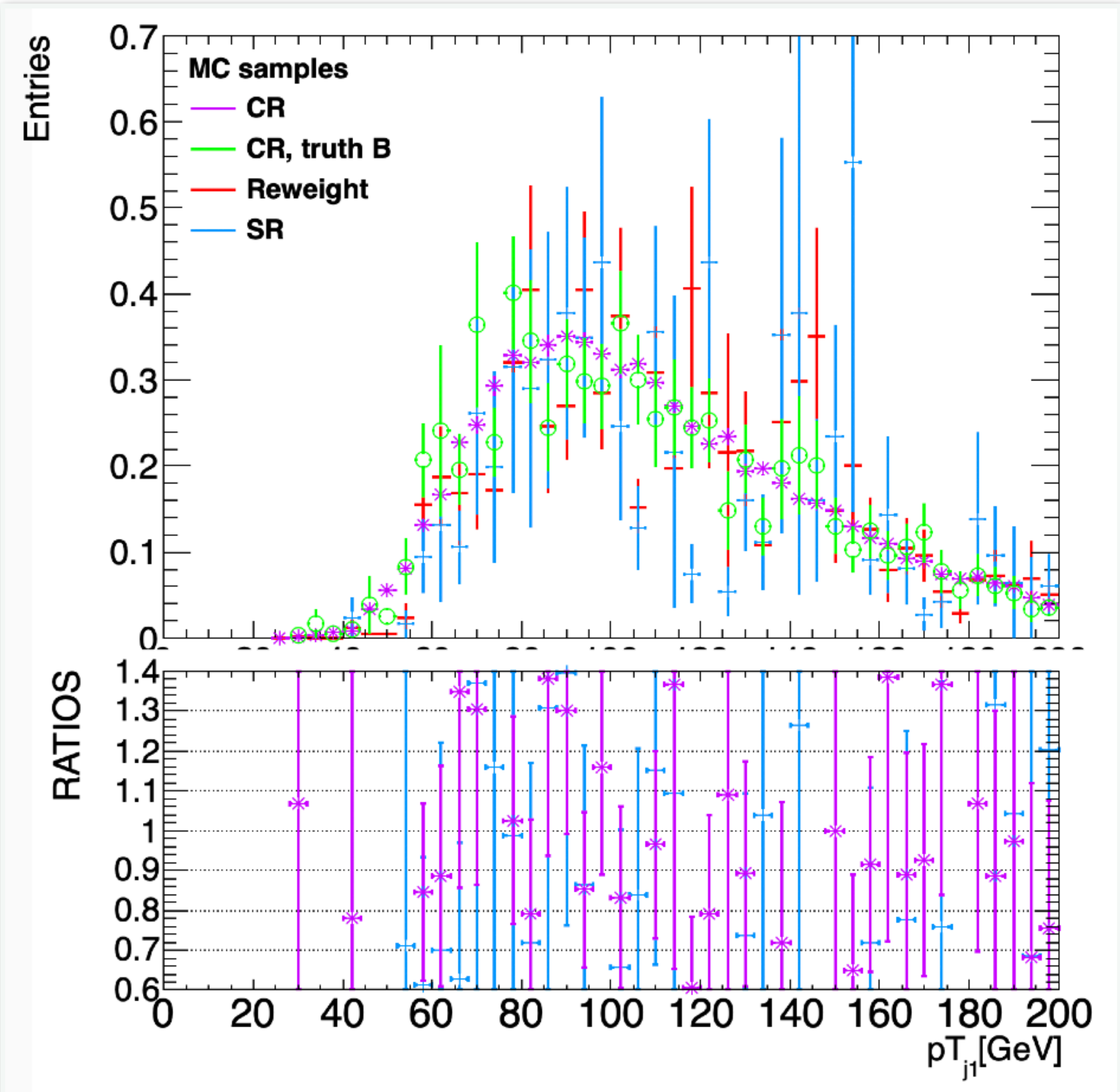
Closure test, pT distributions-SR4

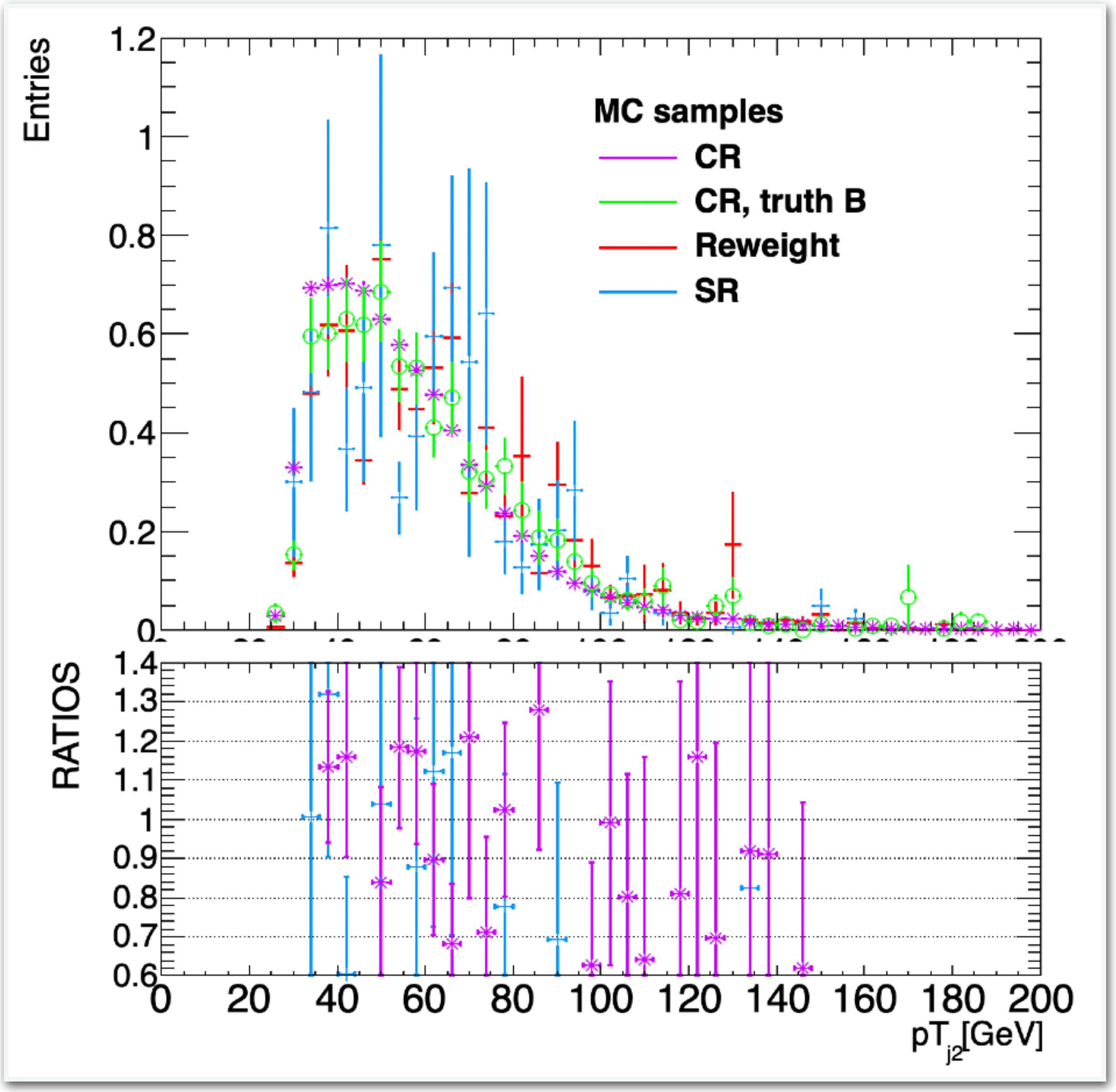




Closure test, pT distributions-SR4

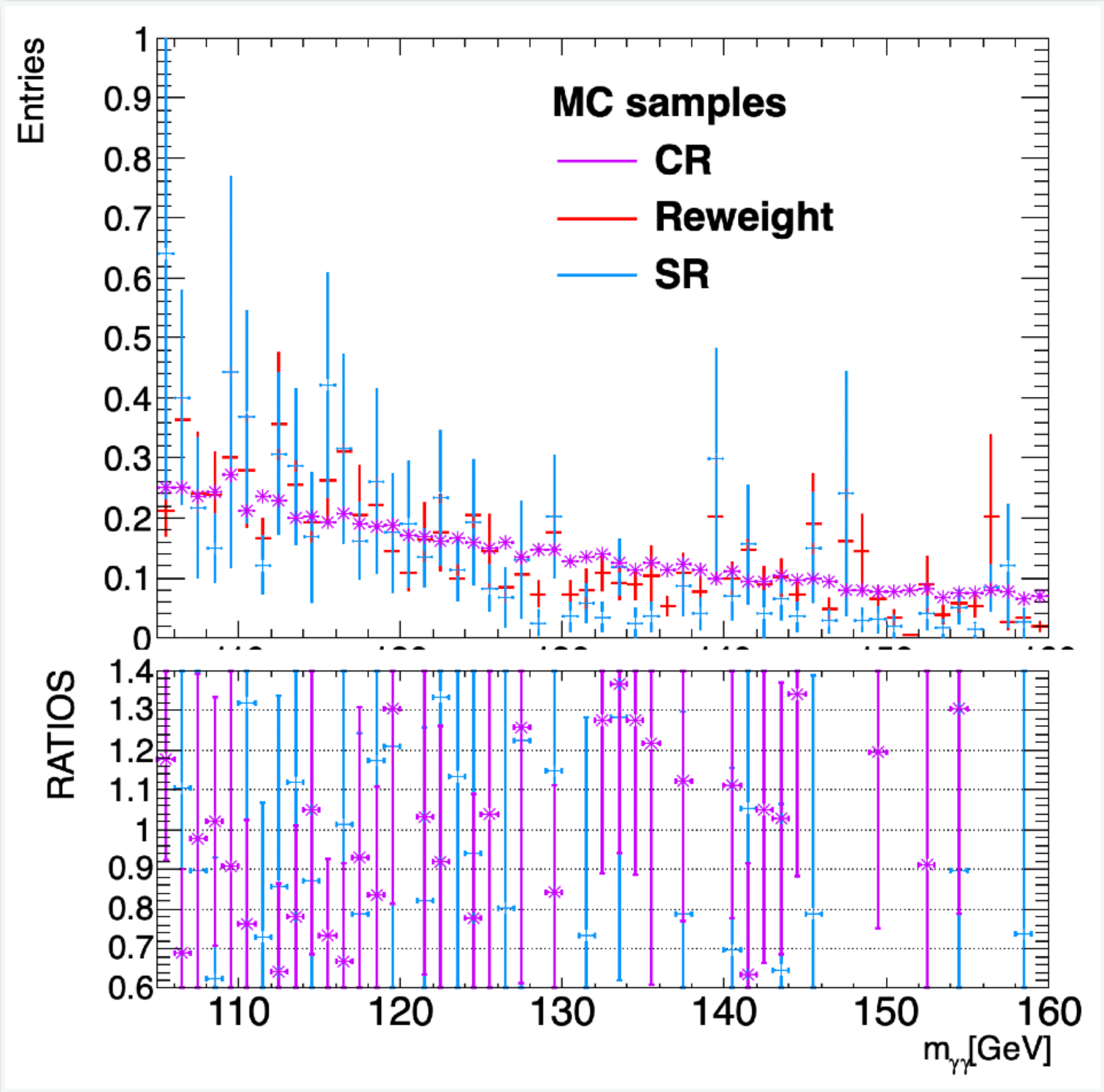
Why is the reweighed distribution not smooth?





Because the b-jets distribution itself is not smooth

Closure test, mass spectra-SR4



	Stat	MC Yield	(sum weight)^2/ (sum weight^2)
SR	228	7.8	86.1
Reweight	51683	7.7	265.1
CR	51683	1730.6	20639.3

- In the reweighted case, the statical amount is enlarged by ~226 wrt SR
- Effective statistics Reweight/SR ~ 3
- Effective statistics CR/Reweight ~ 78

Closure test, mass spectra-SR4